



HAL
open science

Synthèse et étude d'ADN et d'ARN G-quadruplexes à topologies contrôlées. Applications pour la caractérisation et la sélection de ligands

Laureen Bonnat

► **To cite this version:**

Laureen Bonnat. Synthèse et étude d'ADN et d'ARN G-quadruplexes à topologies contrôlées. Applications pour la caractérisation et la sélection de ligands. Chimie inorganique. Université Grenoble Alpes, 2017. Français. NNT : 2017GREAV081 . tel-01744347

HAL Id: tel-01744347

<https://theses.hal.science/tel-01744347>

Submitted on 27 Mar 2018

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 LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Chimie et Science du Vivant**

Arrêté ministériel : 25 mai 2016

Présentée par

Lauren BONNAT

Thèse dirigée par **Eric DEFRANCQ**
et codirigée par **Corinne RAVELET**

préparée au sein des laboratoires **DCM** et **DPM**
dans **l'École Doctorale Chimie et Science du Vivant**

Synthèse et étude d'ADN et d'ARN G- quadruplexes à topologies contrôlées. Applications pour la caractérisation et la sélection de ligands

Thèse soutenue publiquement le **19 décembre 2017**,
devant le jury composé de :

Mme Florence MAHUTEAU-BETZER

Chargée de recherche CNRS, Institut Curie d'Orsay (Rapporteur)

M. François MORVAN

Directeur de recherche INSERM, Université de Montpellier (Rapporteur)

M. Didier GASPARUTTO

Ingénieur de recherche CEA, COMUE Grenoble-Alpes (Président du jury)

M. Pierre-Yves RENARD

Professeur, Université de Rouen (Examineur)

M. Eric DEFRANCQ

Professeur, Université Grenoble-Alpes (Membre)

Mme Corinne RAVELET

Maître de conférences, Université Grenoble-Alpes (Membre)

M. Thomas LAVERGNE

Chargé de recherche CNRS, Université Grenoble-Alpes (Invité)



Remerciements

Pour commencer, je remercie le labex arcane qui a financé cette thèse.

Je remercie également les membres du jury qui ont évalué ces travaux de thèse :

- Mme Florence Mahuteau-Betzer, Chargé de recherche CNRS, Institut Curie d'Orsay
- M. François Morvan, Directeur de recherche INSERM, Université de Montpellier
- M. Didier Gasparutto, Ingénieur de recherche CEA, COMUE Grenoble-Alpes
- M. Pierre-Yves Renard, Professeur, Université de Rouen

Je remercie aussi le Dr. Serge Cosnier, directeur du Département de Chimie Moléculaire pour m'avoir accueilli dans le laboratoire ainsi que le Dr. Didier Boturyn en tant que responsable de l'équipe I2BM.

Il est aujourd'hui temps pour moi de refermer ce merveilleux chapitre de ma vie que fut ma thèse. Ces remerciements sont alors l'occasion de dire un grand « MERCI » à tous ceux qui m'ont apporté tant de joie et de soutien durant ces trois années.

Tout d'abord, je souhaite remercier mes trois encadrants, le Pr. Eric Defrancq, le Dr. Corinne Ravelet et le Dr. Thomas Lavergne, qui ont énormément participé à mon épanouissement durant cette thèse grâce à leur écoute, leur soutien et leur bonne humeur permanente.

Eric, il y a 6 ans tu m'as accordé ta confiance en acceptant de me prendre en stage de L3 pendant trois semaines. Tu m'as fait découvrir tes recherches et un labo que je n'ai plus voulu quitter. Je te remercie d'avoir cru en moi, de n'avoir jamais cessé de m'encourager et de m'épauler. Tu as toujours été très disponible et à l'écoute, je pourrai en écrire encore bien plus, mais pour résumer tu as tenu ton rôle de directeur de thèse à merveille et pour cela, merci infiniment.

Thomas, je te remercie d'avoir été autant investi dans ma thèse, ton exigence m'a permis de me surpasser et quand je vois tout ce qui a été fait, je ne peux que te remercier. Au-delà de ça, tu as toujours su m'épauler, me rassurer et être disponible aussi bien pour les problèmes de manips que pour des choses plus personnelles. Il y aurait également encore bien plus à dire, mais pour résumer mille fois merci pour le soutien que tu m'as apporté durant ces trois ans.

Corinne, j'ai énormément apprécié travailler avec toi au DPM, tu m'as fait découvrir de nouvelles recherches, de nouvelles manips avec toute ta bonne humeur. Je te remercie également pour ta gentillesse et l'implication que tu as eu dans ma thèse.

J'espère qu'avec ces quelques mots vous aurez compris tout le bonheur que ça a été pour moi de travailler avec vous durant ces trois années.

Je souhaite également remercier les membres du DPM avec qui j'ai travaillé et notamment Emmanuelle pour sa gentillesse, sa bienveillance et sa joie de vivre communicative.

Je souhaite maintenant remercier tous les membres de l'équipe I2BM grâce à qui le labo est devenu durant trois ans ma deuxième maison. Vous avez tous à un moment donné eu la gentillesse de m'aider dans mes manips ou de me donner quelques conseils, vous avez su égayer mon quotidien et ça a été un plaisir pour moi de travailler avec vous tous.

Je remercie plus particulièrement les membres de l'équipe avec qui j'ai étroitement travaillé. Nico, même si tu as mis un grand nombre de fois le « bazar » dans le labo, je te remercie pour tous tes

précieux conseils et pour la bonne humeur que tu fais régner dans le labo. Je ne suis pas prête d'oublier nos discussions interminables et tous nos moments de rire.

Jérôme je te remercie pour toute l'aide et toutes les manip que tu as réalisées en SPR et en BLI. J'ai énormément apprécié travailler avec toi mais j'ai encore plus apprécié ton vin et ton fromage !!!! Surtout ne me supprime pas de ta mail liste, je compte bien continuer à t'en commander.

Laure, je te remercie pour toutes les manip que tu as réalisées en SPR et en BLI, mais encore plus pour tes conseils, ton aide et ta gentillesse.

Je souhaite maintenant remercier ma petite Régine, ma maman du labo pour toute la gentillesse et le soutien qu'elle m'a apporté durant toutes ces années. Il y aurait tellement à dire pour te remercier, tu as été mon pilier au labo, ma confidente, tu m'as rassuré, écouté, épaulé, aidé, conseillé, consolé... Un énorme MERCI ma Régine d'avoir autant été là pour moi.

Je souhaite également remercier ma petite Sabinette pour toute l'aide et tout le soutien qu'elle m'a apporté, mais surtout pour son immense gentillesse et son sourire quotidien. Tu m'as énormément conseillé et aidé dans mes enseignements, dans mes manip et dans ma recherche de travail post-thèse alors mille mercis pour tout cela.

Les filles, je n'oublierais jamais nos footings du midi, c'était notre petit moment de décompression, notre petite pause rigolade, « papotage ». De vrais moments de joie qui vont énormément me manquer. Encore mille mercis pour tous les fous rires et pour tous les bons moments que nous avons partagé ensemble au labo et en dehors du labo et pour tous ceux qui sont à venir.

Je remercie maintenant mes amies Doti et Céline. Les filles ça a été un bonheur de partager avec vous ces années de master et de thèse. Nous avons tout partagé ensemble, le stress des partiels, le stress des concours, le stress des manip qui ne marche pas, nous nous sommes aidées, comprises, soutenues. Mais au-delà de ça nous avons surtout partagé de superbes moments d'amitié, de gros fous rire et de longs moments de bavardage dans les couloirs du labo (au grand désespoir de certains^^).

Je remercie maintenant tous les copains du labo avec qui j'ai tant rigolé et passé de super moments : Carlo mon chouchou italien, merci d'avoir partagé ces trois années de thèse avec moi, Fabien, merci de m'avoir fait autant rire (sans oublié Justine, merci pour tes délicieux gâteaux), Eugénie, merci d'avoir été ma super copine de labo et d'avoir sauvé mon impression de thèse, Benjamin, merci pour tous les bons moments que l'on a partagé, David, merci également pour tous ces bons moments et pour ta gentillesse, Melissa, merci pour tous nos petits moments à papoter, Claire, merci pour nos discussions interminables dans le bureau, Alex, merci pour ta bonne humeur et toutes tes blagues qui m'ont tant fait rire, Hugues, merci pour tous les supers moments que l'on a partagé au labo et en dehors, Fabien G, merci pour nos superbes photos sur lesquelles on est trop « beaux gosses », Estelle, merci pour ta bonne humeur, Noémie, merci pour tous nos fous rires, Fatima, merci pour tes « # » mémorables. Un grand merci à tous pour tous les supers moments que nous avons partagé au labo, ou autour d'un verre, ou en congrès, ou encore en soirée. Vous êtes devenus bien plus que de simples collègues de travail, vous allez tous énormément me manquer et j'espère que malgré les différents chemins professionnels que nous empruntons nos liens resteront toujours aussi forts.

Parce que durant ces trois années je n'ai pas non plus fait que travailler, je souhaite maintenant remercier ma famille et mes amis d'avoir vécu avec moi cette aventure.

Je remercie tout d'abord mes Corky d'amour : Clémounette, Natachou, Audette, Eléna et Adeline de m'avoir encouragé durant ces trois ans. Tibo, Antho, Michou et Barbu, je vous associe

également aux filles pour tous vous dire un grand MERCI d'être des amis au top et d'avoir réussi à me sortir de ma thèse pendant toutes nos soirées, tous nos week-end et tous les moments que l'on a partagé.

Ma Clémounette, durant ces trois années tu n'as jamais cessé de m'encourager, de me soutenir, de me rassurer, de croire en moi et je ne pourrais jamais assez te remercier pour ça. Nos petits thés le soir, nos moments shopping, nos journées burger ciné, nos week-end tous les 4 avec Tibo et Adri m'ont permis de décompresser, de m'évader durant ces trois années très intenses. Il y aurait encore des milliers de choses à dire, mais pour résumer je te remercie d'être une amie en or.

Je souhaite maintenant remercier mon petit papoune et ma petite mamounette. Il n'y aurait pas assez d'une thèse pour vous dire tout ce que je ressens et tout le bonheur que c'est d'être votre fille. Depuis toujours vous me dites de suivre mon instinct, mes envies et de faire ce que j'aime, j'ai suivi votre conseil et grâce à vous je suis totalement épanouie. Vous m'avez toujours soutenu, aidé et surtout vous avez toujours cru en moi. C'est en très grande partie grâce à vous que j'ai pu réussir à faire cette thèse, alors MERCI d'être des parents aussi géniaux.

Je souhaite également remercier mon keukeu (alias Bernard). L'aide et le soutien que tu m'apporte son très précieux à mes yeux, merci.

Je vais finir ces remerciements en remerciant celui qui est mon pilier depuis de nombreuses années et qui a été d'un soutien sans faille durant ces trois ans. Mon amour, merci d'avoir toujours cru en moi, merci de m'avoir donné confiance en moi, merci d'avoir été si patient et d'avoir enduré durant trois ans des horaires de travail imposants. Il n'y a pas de mots assez forts pour te remercier comme il se doit du soutien que tu ne cesses de m'apporter, alors je te dis tout simplement merci de rendre ma vie plus belle chaque jour.

Sommaire

Abréviations.....	1
Introduction.....	3
I. Présentation des acides nucléiques.....	4
I.1. Découverte des acides nucléiques	4
I.2. Structure des acides nucléiques.....	6
I.2.a. Structures de base des acides nucléiques : les nucléotides.....	6
I.2.b. Structures secondaires des acides nucléiques.....	7
<i>i) La double hélice.....</i>	<i>7</i>
<i>ii) Autres structures secondaires.....</i>	<i>8</i>
II. Quadruplexes d'acides nucléiques.....	9
II.1. Structures et topologies des G-quadruplexes d'acides nucléiques.....	10
II.1.a. Structures des G-quadruplexes (G4).....	10
<i>i) Formation des motifs G-quadruplexes.....</i>	<i>10</i>
<i>ii) Stabilisation des G-quadruplexes par des cations métalliques.....</i>	<i>10</i>
II.1.b. Topologies des G-quadruplexes (G4).....	11
<i>i) Les différentes topologies adoptées par les G-quadruplexes ADN.....</i>	<i>11</i>
<i>ii) Particularité des G-quadruplexes ARN.....</i>	<i>15</i>
II.1.c. Détermination de la topologie des G-quadruplexes par absorption UV (TDS : Thermal Difference Spectra) et par Dichroïsme Circulaire (CD).....	16
<i>i) TDS : Thermal Difference Spectra.....</i>	<i>16</i>
<i>ii) Dichroïsme circulaire.....</i>	<i>16</i>
<i>iii) Résonance magnétique nucléaire.....</i>	<i>17</i>
II.2. Structures des G-triplexes d'acides nucléiques.....	18
II.3. Structures des i-motifs d'acides nucléiques.....	19
III. Localisation des G-quadruplexes.....	20
III.1. G-quadruplexes localisés dans la région télomérique.....	21
III.1.a. G-quadruplexes et télomérase.....	22
III.1.b. Les différents types de G-quadruplexes dans la région télomérique.....	22
III.2. G-quadruplexes localisés dans les génomes bactériens et viraux.....	23
III.3. Les autres G-quadruplexes génomiques.....	23
IV. Localisation des i-motifs.....	24

V. Ligands de G-quadruplexes.....	24
V.1. Les différents ligands de G-quadruplexes étudiés.....	26
V.1.a. Les dérivés de salophen.....	27
V.1.b. Les dérivés de porphyrines.....	27
V.1.c. Les dérivés d'acridine.....	28
V.1.d. Les dérivés de bisquinolinium.....	28
VI. Ligands des i-motifs.....	29
VII. Concept TASQ.....	31
VIII. Ligations chimiosélectives.....	33
IX. Présentation du projet de thèse.....	37
Chapitre 1 : Synthèse et caractérisation de mimes de motifs G-quadruplexes, G-triplexe et de i-motif.....	40
Introduction.....	41
I. Généralités sur la synthèse peptidique.....	41
II. Généralités sur la synthèse oligonucléotidique.....	42
II.1 Généralités sur la synthèse d'ADN.....	42
II.2. Généralités sur la synthèse d'ARN.....	43
Partie I : Mimes d'ARN G-quadruplexes de la région télomérique.....	45
I. Introduction.....	45
I.1. Découverte du G-quadruplexe de l'ARN télomérique (séquence TERRA) et intérêt de son ciblage.....	46
I.2. Découverte du G-quadruplexe hybride ADN/ARN de la séquence télomérique et intérêt de son ciblage.....	47
II. Synthèse et caractérisation d'édifices mimant les G-quadruplexes de l'ARN télomérique humain.....	48
II.1. Synthèse et caractérisation des mimes du motif G-quadruplexe parallèle.....	48
II.1.a. Stratégie oxime.....	49
<i>i) Synthèse du châssis peptidique.....</i>	<i>49</i>
<i>ii) Synthèse des brins d'ARN.....</i>	<i>52</i>
<i>iii) Assemblage de l'édifice.....</i>	<i>53</i>

II.1.b. Stratégie CuAAC.....	55
i) Synthèse du châssis peptidique 4	55
ii) Synthèse des brins d'ARN.....	57
iii) Assemblage de l'édifice.....	58
II.2. Synthèse et caractérisation du mime du G-quadruplexe hybride ADN/ARN E.....	59
II.2.a. Synthèse du châssis peptidique 8.....	59
II.2.b. Synthèse des brins d'ADN et d'ARN.....	61
i) Synthèse des brins d'ADN.....	61
ii) Synthèse des brins d'ARN 3'alcyne.....	62
II.2.c. Assemblage de l'édifice.....	62
II.3. Synthèse et caractérisation d'un édifice mimant un motif G-quadruplexe antiparallèle, dérivant de l'ARN télomérique.....	63
II.3.a. Synthèse du châssis peptidique.....	64
II.3.b. Synthèse des brins d'ARN.....	64
i) Synthèse des brins d'ARN 3'aldehyde.....	64
ii) Synthèse des brins d'ARN 5'alcyne.....	64
II.3.c. Assemblage de l'édifice.....	65
III. Etudes de la topologie des édifices mimant les G-quadruplexes de l'ARN télomérique humain.....	66
III.1. Etudes d'absorption UV : "thermal difference spectrum" (TDS).....	66
III.2. Etudes de dichroïsme circulaire.....	67
III.2.a. Etude de la topologie des motifs G-quadruplexes.....	67
III.2.b. Etude de la stabilité des motifs G-quadruplexes.....	69
i) Evaluation de la stabilité des édifices.....	69
ii) Influence de la concentration sur la température de dénaturation	70
iii) Influence de la cinétique de repliement sur la température de dénaturation.....	71
III.3. Etudes par résonance magnétique nucléaire.....	72
IV. Bilan.....	73
Partie II : Mimes d'ADN G-quadruplexe présent dans le génôme du VIH-1.....	74
I. Introduction.....	74
I.1. Découverte du G-quadruplexe de la séquence HIV-PRO1 et intérêt de son ciblage...74	
II. Synthèse et caractérisation de l'édifice HIV-PRO1.....	76
II.1. Synthèse du châssis peptidique.....	77

II.2. Synthèse des brins d'ADN.....	79
II.2.a. Synthèse du brin d'ADN 5'alcyne - 3'aldéhyde 12	79
II.2.b. Synthèse du brin d'ADN 5'thiol - 3'alcyne 13	80
II.3. Assemblage de l'édifice.....	81
III. Etudes de la topologie de l'édifice G mimant le motif G-quadruplexe de la séquence HIV-PRO1.....	87
III.1. Etudes d'absorption UV : "thermal difference spectrum" (TDS).....	87
III.2. Etudes de dichroïsme circulaire.....	88
III.2.a. Etude de la topologie des motifs G-quadruplexes.....	88
III.2.b. Etude de la stabilité des motifs G-quadruplexes.....	89
III.3. Etudes de résonance magnétique nucléaire.....	90
IV. Bilan.....	91
Partie III : Mimes d'ADN G-triplexe de la région télomérique.....	92
I. Introduction.....	92
II. Synthèse et caractérisation d'édifices G-triplexes dérivés de l'ADN télomérique.....	93
II.1. Synthèse du châssis peptidique.....	94
II.1.a. Synthèse et insertion de la fonction oxyamine.....	94
II.1.b. Description de la synthèse du châssis peptidique.....	95
II.2. Synthèse des oligonucléotides.....	97
II.2.a. Synthèse de l'oligonucléotide boucle 5'alcyne-3'aldéhyde 16	97
II.2.b. Synthèse des oligonucléotides 3'-5' court et long.....	98
II.3. Assemblage des édifices.....	99
III. Etudes de la topologie des édifices mimant le motif G-quadruplexe de la séquence télomérique humaine et son intermédiaire triplex.....	100
III.1. Etude de la topologie des édifices.....	101
III.1.a. Etude de la topologie des édifices H et I par dichroïsme circulaire.....	101
III.1.b. Etude de la topologie des édifices H et I par mesure des TDS.....	104
III.2. Etude de la stabilité des édifices.....	105
III.2.a. Evaluation de la stabilité des édifices.....	105
III.2.b. Influence de la concentration sur la stabilité de l'édifice H.....	107

III.3. Etudes de résonance magnétique nucléaire.....	108
IV. Bilan.....	109
Partie IV : Mime d'ADN i-motif de la région télomérique.....	110
I. Introduction.....	110
II. Synthèse de l'édifice i-motif de l'ADN télomérique.....	110
II.1. Synthèse du châssis peptidique.....	113
II.2. Synthèse des brins d'ADN.....	114
II.2.a. Synthèse de l'oligonucléotide 5'alcyne-3'aldéhyde 20	114
II.2.b. Synthèse de l'oligonucléotide 5'alcyne-3'thiol 21	115
II.3. Assemblage de l'édifice J	116
II.3.a. Assemblage de l'intermédiaire J₂	116
II.3.b. Assemblage de l'intermédiaire J₅ à partir de l'intermédiaire J₂	116
II.3.c. Assemblage de l'édifice J à partir de l'intermédiaire J₅	118
III. Bilan.....	119
Chapitre 2 : Utilisation de G-quadruplexes contraints pour la caractérisation et la sélection de ligands.....	121
Introduction.....	122
Partie I : Utilisation de motifs G4 contraints pour évaluer l'affinité et la spécificité de différentes familles de ligands par SPR et par BLI	123
I. Introduction.....	123
I.1. Principe de la SPR.....	124
I.2. Principe du BLI.....	125
I.3. Traitement des données.....	126
II. Evaluation de l'affinité et de la sélectivité de ligands pour les édifices G-quadruplexes ADN parallèle A, ARN parallèle C et hybride ADN/ARN E par SPR.....	127
III. Evaluation de l'affinité et de la sélectivité de ligands pour les édifices G-quadruplexes de type « basket » B et HIV-PRO1 G par SPR.....	129
IV. Evaluation de l'affinité et de la spécificité de ligands pour les édifices G-triplexe H, G-quadruplexe I et duplexe d'ADN, par BLI.....	134

V. Bilan.....	136
Partie II : Développement d'une nouvelle classe de ligands de G-quadruplexe : les aptamères	137
I. Introduction.....	137
I.1. Principe de la SELEX classique.....	137
I.2. Principe de la CE-SELEX.....	138
I.3. Choix de la cible.....	139
II. Synthèse de la cible G-quadruplexe.....	140
II.1. Synthèse du châssis peptidique 22	140
II.2. Synthèse de l'oligonucléotide 16	141
II.3. Assemblage de l'édifice K	142
III. Sélection avec des bases naturelles.....	143
III.1. Détermination des conditions de sélection.....	143
III.1.a. Banque utilisée.....	143
III.1.b. Choix du tampon d'incubation.....	143
III.2. Test d'interaction entre la cible et la banque.....	144
IV. Sélection avec des bases modifiées.....	145
IV.1. Test d'interaction entre la cible et la banque modifiée.....	145
IV.2. Réalisation de la CE-SELEX.....	146
IV.3. Résultat du séquençage.....	147
IV.4. Evaluation de l'affinité des séquences sélectionnées pour la cible.....	150
V. Bilan.....	151
Conclusion et perspectives.....	152
Partie expérimentale.....	160
Références bibliographiques.....	196
Annexes.....	208

Abréviations

A : Adenosine	DMTr : Diméthoxytrityle
Ac ₂ O : Anhydride acétique	DTT: dithiothreitol
ACN : Acétonitrile	
AcOEt : Acétate d'éthyle	EDTA: Ethylenediaminetetraacetic acid
AcOH : Acide acétique	Eei : 1-éthoxyéthylidène
AcONH ₄ : Acétate d'ammonium	Et ₂ O : Ether diéthylique
AELC : Anion Exchange Liquid Chromatography	EtOH : Ethanol
AFM : Atomic Force Microscopy	ESI-MS : Electrospray Ionisation Mass Spectrometry
Alloc : Allyloxycarbonyl	
ADN : Acide désoxyribonucléique	
ARN : Acide ribonucléique	Fmoc : 9-fluorénylméthoxyloxycarbonyl
Biot : Biotine	G : Guanosine
BLI : Bio-Layer Interferometry	G3 : G-triplexe
Boc : <i>Tertio</i> -butyloxycarbonyl	G4 : G-quadruplexe
C : Cytidine	HBTU : <i>O</i> -(benzotriazol-1-yl)- <i>N,N,N',N'</i> -tetramethyluronium hexafluorophosphate
CD : Circular Dichroism	HEPES : Acide 4-(2-hydroxyethyl)-1-piperazineethanesulfonique
CE-SELEX : Capillary Electrophoresis–Systematic Evolution of Ligands by Exponential Enrichment	
CNE : Cyanoéthyle	IC ₅₀ : Concentration inhibant de 50% un paramètre
CuAAC : Copper (I) Catalyzed Alkyne-Azide	IEDDA : Inverse Electron Demand Diels-Alder
CPG : Controlled Pore Glass	ISAHC : Imidazole-1-sulfonyl acide hydrochloride
DBU : 1,8-diazabicyclo[5.4.0]undec-7-ene	K _D : Constante d'équilibre de dissociation
DCC : <i>N,N'</i> -dicyclohexylcarbodiimide	K _{off} : Constante cinétique de dissociation
DCM : Dichlorométhane	K _{on} : Constante cinétique d'association
Dde : 1-(4,4-diméthyl-2,6-dioxocyclohex-1-ylidène)éthyl	MALDI-Tof: Matrix Assisted Laser Desorption Ionisation Mass Spectrometry-Time of flight
DIEA : <i>N,N</i> -diisopropyléthylamine	
DMF : <i>N,N</i> -diméthylformamide	

MeOH : Méthanol

NHS : *N*-hydroxysuccinimide

PCR : Polymerase chain reaction

PivOM : Pivaloyloxyméthyle

PyBop : benzotriazol-1-yl-
oxytripyrrolidinophosphonium

hexafluorophosphate

Pyr : Pyridine

RAFT : Regioselectively Addressable
Functionalized Template

RMN : Résonance Magnétique Nucléaire

RP-HPLC : Reverse-Phase High Performance
Liquid Chromatography

RU : Resonance Unit

SEC: Size Exclusion Chromatography

SPOS : *Solid-Phase Oligonucleotide Synthesis*

SPPS : *Solid-Phase Peptide Synthesis*

SPR : Surface Plasmon Resonance

T : Thymidine

TASQ : Template Assembled Synthetic G-
quadruplex

TBDMS : *tert*-butyldiméthylsilyle

tBu : *tert*-butyl

TDS : Thermal Difference Spectrum

TEA : Triéthylamine

TEAAc : acétate de triéthylammonium

TFA : Acide Trifluoroacétique

TFE : Trifluoroéthanol

THF : Tétrahydrofurane

THPTA : Tris-
(hydroxypropyltriazolylmethyl)amine

TIS : Triisopropylsilane

T_m : Température de dénaturation

Tr : Temps de rétention

TRAP : Telomeric repeat Amplification

TRIS : *Tris*(hydroxymethyl)aminométhane

U : Uridine

UPLC : Ultra Performance Liquid
Chromatography

UV: Ultraviolet

VIH : Virus de l'immunodéficience humaine

Introduction

I. Présentation des acides nucléiques

Depuis plus de 150 ans les acides nucléiques fascinent les chercheurs du monde entier. Ces macromolécules dont il existe deux types, l'ADN et l'ARN sont les piliers de la vie cellulaire. Il fut un temps où il était absolument incroyable de penser que l'enchaînement de ces molécules d'à peine 2 nm de diamètre pourrait être le fondement de la vie. Et pourtant, aujourd'hui grâce à des milliers d'études nous savons que tout cela est bien réel. Ces macromolécules permettent le développement, le fonctionnement et la reproduction des êtres vivants.

L'ADN et l'ARN possèdent des rôles bien distincts dans la cellule, le premier est le porteur de l'information génétique et le deuxième remplit plusieurs fonctions. Il existe ainsi différents types d'ARN. Les ARN messagers qui sont le lien direct entre l'ADN et la production de protéines. Les ARN de transferts qui sont impliqués dans l'adressage des acides aminés vers les ribosomes lors de la traduction, c'est-à-dire lors de la synthèse des protéines à partir des ARN messagers. Il existe également des ARN catalytiques qui sont capables de catalyser des réactions chimiques du métabolisme cellulaire. Mais aussi des ARN guides qui ont la capacité de s'associer à des enzymes protéiques. Au sein des cellules, on retrouve également des ARN régulateurs de l'expression génétique.

D'un point de vue évolutif, il a été proposé que l'ARN soit antérieur à l'ADN comme support de l'information génétique, ce qui expliquerait ses fonctions plus étendues. En 1986 W. Gilbert propose « un monde à ARN », dans lequel, il y a des millions d'années l'ARN aurait été le précurseur de toutes les macromolécules biologiques et particulièrement de l'ADN¹. Ce dernier serait alors apparu plus tard et aurait remplacé l'ARN dans son rôle de stockage de l'information génétique à long terme, en raison de sa plus grande stabilité.

I.1. Découverte des acides nucléiques

C'est en 1869, dans les laboratoires de l'université de Tübingen, dans le Sud-Ouest de l'Allemagne que fut découvert pour la première fois l'ADN². F. Miescher alors jeune chercheur s'intéresse à la physiologie des cellules lymphatiques humaines. Ces cellules difficiles à extraire des ganglions lymphatiques, se trouvent en grande quantité dans le pus des infections, qu'il récolte sur des bandages de l'hôpital voisin. En étudiant les constituants de ces cellules, il isole des noyaux cellulaires et met en évidence une substance inconnue qu'il nomme « nucléine ». Il en est certain, la composition chimique de celle-ci est différente de celle des protéines ou de toute autre molécule connue à cette époque. Il fallut ensuite attendre deux ans après cette découverte pour voir

apparaître la première publication sur la nucléine : temps nécessaire à F. Miescher pour confirmer cette découverte en isolant de nouveau la nucléine à partir d'autres cellules. Il utilisa notamment des spermatozoïdes de saumons, qu'il préféra ensuite au pus comme matériel de départ.

En 1889, R. Altmann un pathologiste et histologiste allemand extrait de la nucléine pure à partir de sperme de saumon et découvre une substance acide. Il l'a renommé alors « acide nucléique ». Quelques années plus tôt, en 1882, W. Flemming, biologiste allemand découvre les chromosomes et examine leur comportement durant la division cellulaire³. Ce n'est que dix ans plus tard, que T. Boveri et W. Sutton identifieront que les chromosomes sont les porteurs de l'information génétique. Ses unités seront nommées « gènes » en 1909 par W. Johannsen. En 1944, l'équipe d'O. T. Avery⁴ identifie l'ADN comme matériel génétique en démontrant qu'il est capable d'engendrer une transformation bactérienne. Cette découverte a marqué un tournant majeur dans l'étude des acides nucléiques. Les chercheurs ont très vite compris l'importance de cette molécule et ont focalisé leurs recherches sur l'étude de sa structure et de son rôle biologique.

A peine dix ans plus tard, dans le numéro du 25 avril 1953 de la revue Nature, deux jeunes chercheurs, le Britannique G. F. Crick et l'Américain J. Watson publient un petit article d'une seule page qui marqua à jamais le monde de la génétique et de la biologie moléculaire. Ils y décrivent leur proposition de structure de l'ADN⁵. Dans leurs recherches, ils se sont inspirés des travaux de P. Levene⁶ qui a identifié en 1929 que l'ADN était formé de « building- blocks » composés de quatre bases : adénine (A), cytosine (C), guanine (G), et thymine (T) (le désoxyribose est identifié et on parlera par la suite d'acide désoxyribonucléique). Ils se sont également servis des travaux d'E. Chargaff⁷ qui a découvert en 1950 que la composition de base de l'ADN varie selon les espèces. Il a également déterminé que dans une espèce, les bases de l'ADN sont toujours présentes dans des rapports fixes, c'est-à-dire qu'il y a toujours le même nombre de A et de T et le même nombre de C et de G. F. Crick et J. Watson aidés par les travaux de M. Wilkins et R. Franklin^{8,9} sur la diffraction des rayons X qui prouvent l'existence d'une double hélice d'ADN, ont alors l'idée d'organiser les bases de manière complémentaire deux à deux. En liant A et T ensemble et C et G ensemble pour former la double hélice. Cette découverte leur valut ainsi qu'à M. Wilkins le prix Nobel de médecine en 1962. R. Franklin décédée avant l'attribution de ce prix ne figure pas parmi les lauréats bien que sa contribution à cette découverte ait été essentielle.

Parallèlement aux études menées sur l'ADN, des analyses chimiques ont révélé au XX^{ème} siècle que les ARN partagent trois bases avec l'ADN (A, C et G). L'uracile a été découverte comme base unique à l'ARN, remplaçant la thymine de l'ADN.

I.2. Structure des acides nucléiques

I.2.a. Structures de base des acides nucléiques : les nucléotides

Les acides nucléiques sont constitués d'un enchaînement de nucléotides reliés par des liaisons phosphodiester. Ces nucléotides sont composés de trois éléments : un dérivé saccharidique, un résidu phosphate et une base hétérocyclique azotée (figure 1).

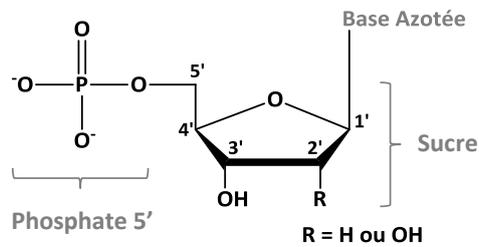


Figure 1 : structure d'un nucléotide

Le dérivé saccharidique est un sucre à cinq carbones dérivé du pentose. Dans l'ADN il s'agit du 2'-désoxy-D-ribose. Dans l'ARN, le sucre est un D-ribose. Le phosphate présent sur le carbone 5' du sucre permet de lier deux nucléotides entre eux, *via* une liaison phosphodiester, en se fixant sur le carbone 3' du nucléotide suivant. Les acides nucléiques sont ainsi formés d'un squelette constitué de l'enchaînement sucre, phosphate, sucre, etc... L'architecture des acides nucléiques ne varie donc que par leurs bases azotées qui sont liées sur le carbone 1' du sucre par une liaison β glycosidique. Les bases nucléiques sont classées en deux familles : les pyrimidines et les purines (figure 2). La première famille regroupe la cytosine, la thymine et l'uracile et la deuxième famille l'adénine et la guanine. Les pyrimidines sont liées par leur N₁ au C_{1'} du ribose ou du déoxyribose et les purines par leur N₉.

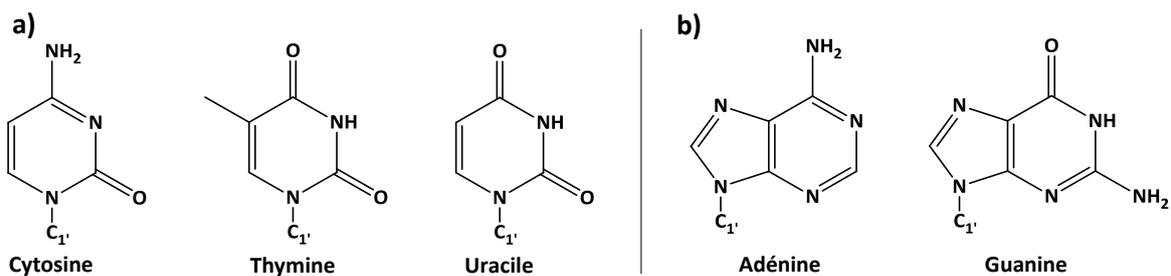


Figure 2 : structure des principales bases azotées de la famille (a) des pyrimidines et (b) des purines

La conformation des nucléotides peut varier suivant la structure secondaire adoptée par les acides nucléiques. Les bases azotées peuvent ainsi adopter les conformations *anti* ou *syn* (figure 3.a) et les sucres les conformations C_{2'} *endo* ou C_{3'} *endo* (figure 3.b).

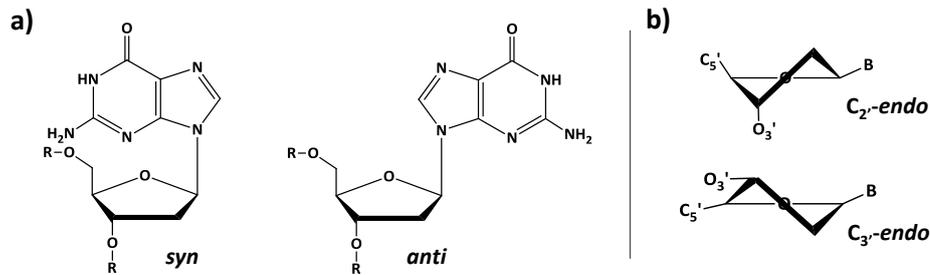


Figure 3 : (a) structure d'un nucléotide avec une guanine en conformation syn et anti, (b) structure d'un sucre en conformation C_{2'} endo et C_{3'} endo

I.2.b. Structures secondaires des acides nucléiques

Les acides nucléiques peuvent en fonction de leur séquence et des conditions environnementales adopter une grande variété de structures secondaires. Ces dernières résultent essentiellement de l'interaction des bases nucléiques entre elles par des liaisons hydrogènes, mais également par des phénomènes d'empilement, aussi appelés « π -stacking ».

i) La double hélice

La structure secondaire la plus connue des acides nucléiques est l'hélice double brin découverte en 1953 par J. Watson et F. Crick grâce aux travaux de cristallographie aux rayons X de M. Wilkins et R. Franklin. Cette structure est formée par l'appariement complémentaire de deux brins d'acide nucléique. Cet appariement résulte de l'association des bases entre elles à l'aide de liaisons hydrogènes de type Watson-Crick entre les couples A:T (ou U) et C:G (figure 4).

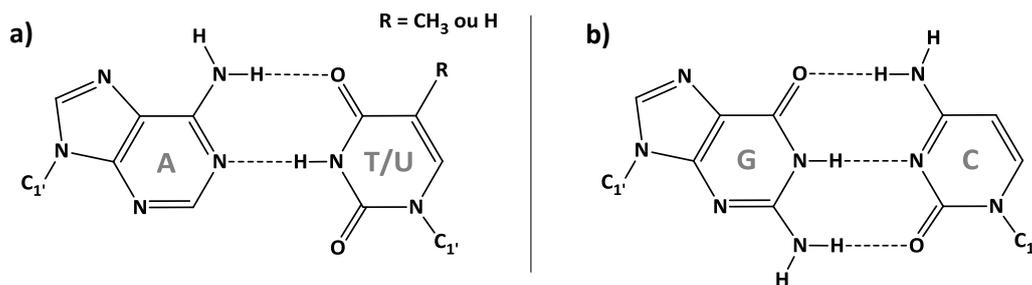


Figure 4 : appariement Watson-Crick des bases azotées pour les couples (a) A :T (ou U) et (b) G :C

Il est à noter que contrairement à l'ADN, l'ARN se trouve principalement sous forme de simple-brin. Il peut cependant se replier sur lui-même en formant des appariements de Watson-Crick intramoléculaires qui conduisent à la formation de régions en hélice adjacentes à des motifs en boucle où les bases ne sont pas appariées.

Au sein de la double hélice, les brins d'ADN ne sont pas positionnés de façon symétrique par rapport à l'axe de la double hélice et définissent donc deux sillons de tailles inégales, un grand sillon

et un petit sillon. La conformation adoptée par l'ADN double brin dépend de son degré d'hydratation, de sa séquence, de son taux de surenroulement, mais également de la nature et de la concentration des ions métalliques. Il existe trois formes d'ADN double brin dont la taille des sillons varie. La forme la plus répandue est l'ADN B (figure 5.a) qui présente un diamètre de 2 nm et qui a été cristallisée pour la première fois en 1980¹⁰. Un tour d'hélice possède dix paires de bases. Ces dernières sont orientées selon la conformation anti et les sucres selon la conformation C_{2'} endo. L'ADN A (figure 5.b) est une forme moins répandue d'ADN double brin et s'observe dans les échantillons d'ADN déshydraté et à haute force ionique. Cette forme est caractérisée par une conformation des sucres en C_{3'} endo. Les bases azotées sont quant à elles toujours orientées en anti. Cette double hélice est plus large que l'ADN B, avec un diamètre de 2,3 nm. Un tour d'hélice représente onze paires de bases. La troisième forme d'ADN double brin est l'ADN Z (figure 5.c). Cette forme est plus contrainte que les autres et s'observe préférentiellement dans les régions riches en paires de bases G-C. Il s'agit d'une double hélice gauche, dont l'axe s'écarte significativement des paires de bases. Cette dernière forme est plus étroite que les deux autres et possède un diamètre d'environ 1,8 nm et douze paires de bases par tour d'hélice. A l'intérieur de l'hélice, les pyrimidines sont orientées en position *anti* et le sucre sur lequel elles sont liées est en conformation C_{2'} endo. A l'inverse, les purines sont orientées en position *syn* et leurs sucres présentent la conformation C_{3'} endo.

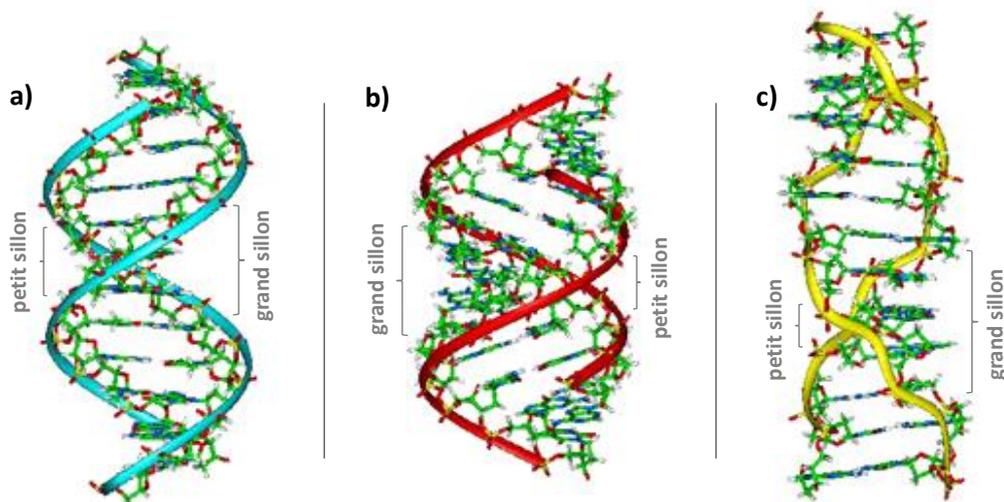


Figure 5 : représentation des différentes structures double brins de l'ADN : (a) ADN-B, (b) ADN-A, (c) ADN-Z

ii) Autres structures secondaires

Bien que la double hélice soit la structure secondaire principale des acides nucléiques, l'environnement et la séquence peut favoriser la formation d'autres structures. Par exemple, des Jonctions de Holliday (figure 6) sont formées lors de recombinaisons homologues entre deux molécules d'ADN.

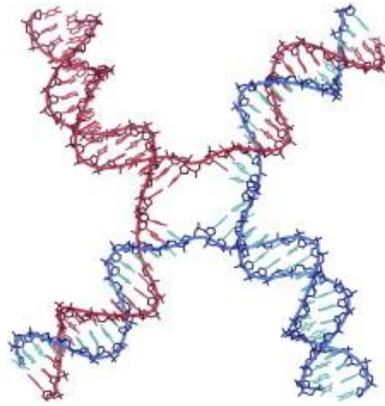


Figure 6 : représentation d'une Jonction Holliday entre deux oligonucléotides

On retrouve également des structures tige-boucle (figure 7), aussi appelées épingle à cheveux. Elles se forment sur un simple brin d'acide nucléique et sont principalement présentes dans l'ARN. Elles résultent de la formation d'acide nucléique double brin par appariement de bases complémentaires suivi par la formation de boucles constituées de bases non appariées. Lorsque seulement une ou deux bases ne sont pas appariées, le motif formé n'est pas une boucle mais un renflement.

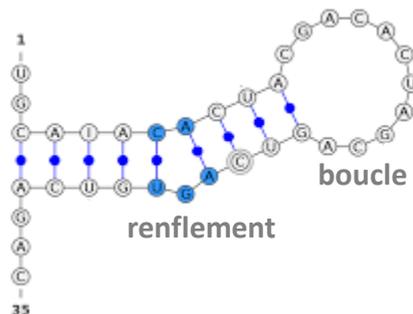


Figure 7 : représentation d'une structure tige-boucle

Des structures issues de la combinaison de liaisons hydrogène et d'empilement « π -stacking » peuvent également se former dans des régions riches en purines. Il s'agit de motifs quadruplexes dont l'étude a été l'objet de cette thèse. La description de leurs structures est représentée dans le paragraphe suivant.

II. Quadruplexes d'acides nucléiques

Les motifs quadruplexes sont des structures secondaires à quatre brins formées par le repliement d'acides nucléiques riches en guanines ou cytosines. Dans le premier cas on parle de G-

quadruplexes (G4) ou quadruplexes de guanines et dans le second cas de i-motifs ou de quadruplexes de cytosines.

II.1. Structures et topologies des G-quadruplexes d'acides nucléiques

II.1.a. Structures des G-quadruplexes (G4)

i) Formation des motifs G-quadruplexes

La formation des motifs G-quadruplexes repose sur l'empilement par π -stacking de plusieurs tétrades de guanines (figure 8). Ces dernières, également appelées G-quartet sont formées de quatre résidus guanines, appariées les unes aux autres par des liaisons hydrogène de type Hoogsteen. La guanine est la seule base azotée capable de former au moins deux liaisons hydrogène sur chacune de ses faces, lui conférant ainsi la possibilité de s'auto-assembler. Cette propriété a été rapportée pour la première en 1962 par M. Gellert et al¹¹. Dans leur étude, ils ont observé que des solutions concentrées en acide guanylique à pH 5 pouvaient être extrêmement visqueuses, allant jusqu'à former des gels à basse température. L'étude aux rayons X de ces gels a révélé l'existence d'une structure hélicoïdale formée par l'acide guanylique. Ils ont alors conclu l'existence de tétramères planaires formant le motif de base de cette structure hélicoïdale. Chaque tétramère est formé par l'auto assemblage de quatre guanines, reliées entre elles par huit liaisons de type Hoogsteen.

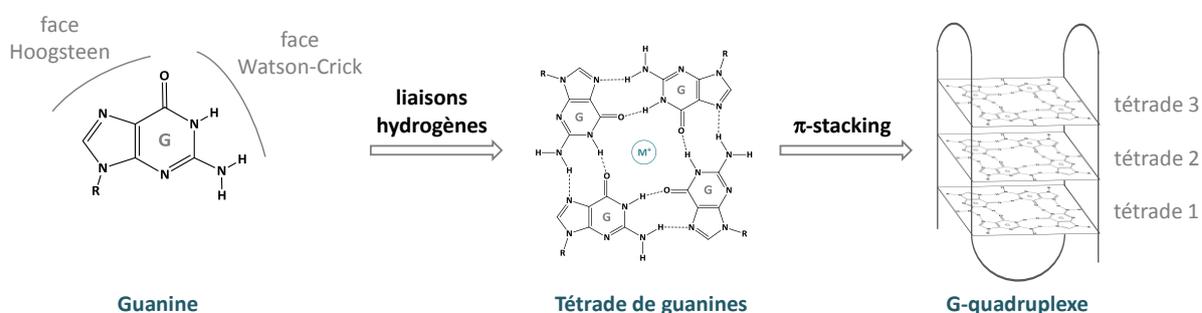


Figure 8 : représentation d'une guanine, d'une tétrade de guanine et d'un G-quadruplexe

ii) Stabilisation des G-quadruplexes par des cations métalliques

Quinze ans après la découverte des tétrades, T. J. Pinnavaia et ses collaborateurs démontrent par des études de RMN que la présence de cations métalliques tels que le sodium ou le potassium stabilise ces auto-assemblages de guanines¹². Les cations stabilisent les G-quartets en se coordonnant aux oxygènes des fonctions carbonyles de chaque guanine. En 1992, l'équipe de C. Bailey publie une liste des cations métalliques stabilisant les G-quadruplexes¹³ qu'ils ont classé par ordre de stabilisation décroissante. Il en résulte les classements suivant : $K^+ > Ca^{2+} > Na^+ > Mg^{2+} > Li^+$ et $K^+ > Rb^+$

> Cs⁺. Il ressort de cette étude le fort pouvoir stabilisant du potassium. Ceci s'explique par la position du cation au sein du motif G-quadruplexe. Avec son rayon atomique de 220 pm le potassium est trop volumineux pour s'insérer au cœur d'une tétrade, il se place donc entre deux tétrades et interagit avec chacune, ce qui stabilise fortement le G-quadruplexe (figure 9). A l'inverse, le sodium qui possède un rayon atomique de 180 pm est assez petit pour se placer au centre d'une tétrade, ce qui stabilise la structure dans une moindre mesure que le potassium (figure 9). Les cations Li⁺ et Cs⁺ sont quant à eux trop petits pour interagir efficacement à l'intérieur d'une tétrade, ils ont donc un pouvoir stabilisant quasiment nul.

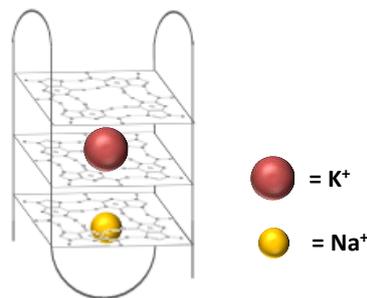


Figure 9 : position du potassium et du sodium au sein d'un motif G-quadruplexe

Les cations métalliques, en plus de stabiliser les motifs G-quadruplexes influencent leurs topologies. En effet, pour une même séquence, des topologies différentes de G-quadruplexe peuvent-être adoptées en fonction du cation métallique utilisé.

II.1.b. Topologies des G-quadruplexes (G4)

i) Les différentes topologies adoptées par les G-quadruplexes ADN

Les acides nucléiques capables de former des motifs G-quadruplexes possèdent des séquences du type : X_n G_p X_n G_p X_n... où X représente les différentes bases azotées de nombre n variable et G les guanines dont le nombre varie généralement entre 2 et 4. Les motifs G-quadruplexes peuvent être formés par quatre, deux ou un seul brin d'acides nucléiques. Dans le premier cas, il s'agit de G-quadruplexes tétramoléculaires (figure 10.a), dans le second cas de G-quadruplexes bimoléculaires (figure 10.b) et enfin dans le dernier cas de G-quadruplexes intramoléculaires (figure 10.c).

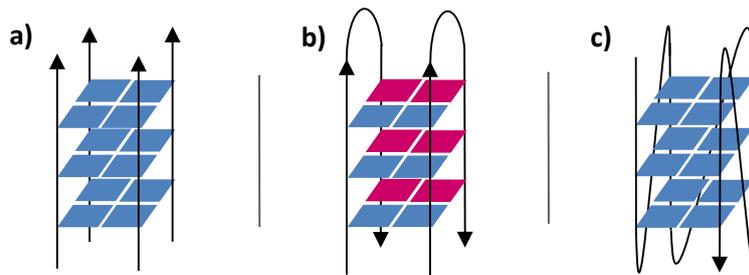


Figure 10 : exemples de G-quadruplexes (a) tétramoléculaire, (b) bimoléculaire, et (c) intramoléculaire

Quel que soit le nombre de brins qui les composent, les motifs G-quadruplexes peuvent adopter une grande variété de topologies. Ces dernières sont influencées par certains paramètres comme par exemple la séquence du brin d'acide nucléique, la nature du cation métallique utilisé comme stabilisant, la concentration et la température du milieu, ou encore les conditions d'hybridation du motif G-quadruplexe. La topologie adoptée par les G-quadruplexes est fonction de l'orientation des brins dans la structure. Lorsque ces derniers sont tous orientés dans le même sens, le motif G-quadruplexe adopte une conformation parallèle. Les G-quadruplexes peuvent également se replier en topologie antiparallèle lorsque les brins s'associent avec une polarité opposée, deux brins sont orientés dans le même sens et les deux autres dans le sens opposé. Une troisième topologie à pouvoir être adoptée par les G-quadruplexes est la conformation hybride, présentant trois brins orientés dans le même sens et le quatrième brin orienté dans le sens opposé. Dans la figure 11 est représenté un exemple de ces trois topologies pour chacune des structures tétramoléculaire (figure 11.a), bimoléculaire (figure 11.b) et intramoléculaire (figure 11.c).

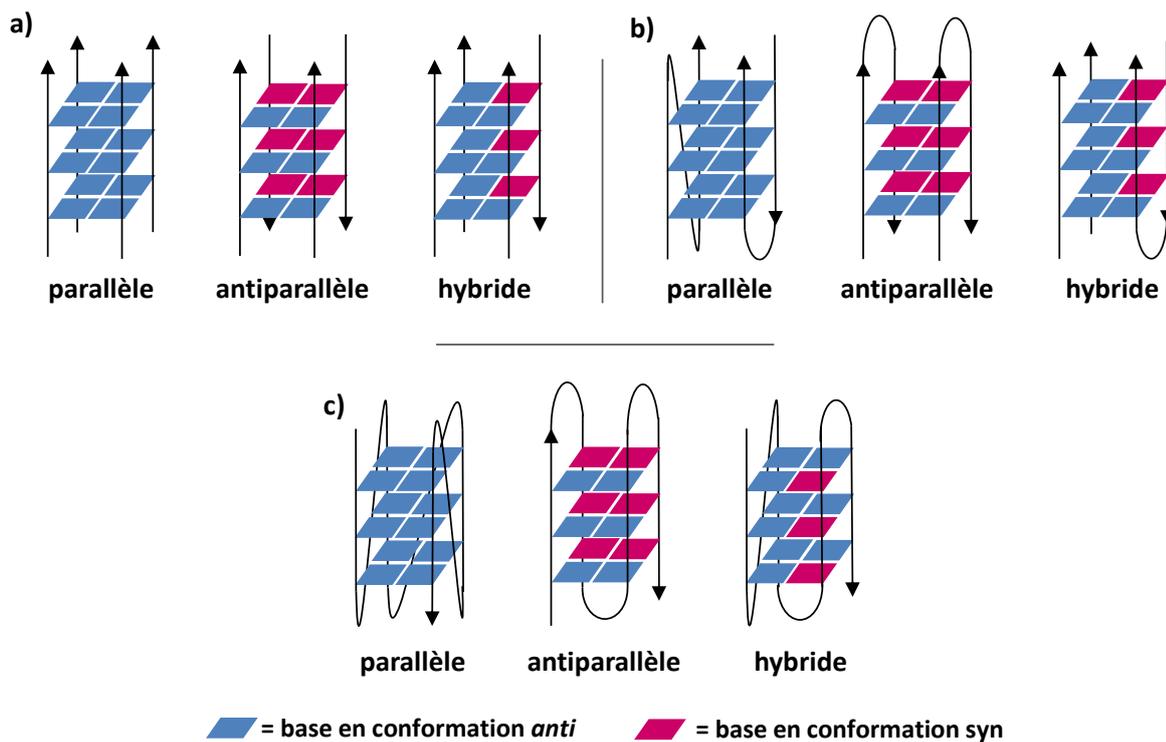


Figure 11 : exemples de topologies parallèles, antiparallèles et hybrides pour des G-quadruplexes (a) tétramoléculaire, (b) bimoléculaire et (c) intramoléculaire

Il est important de noter qu'il n'existe pas une, mais une multitude de topologies parallèles, antiparallèles et hybrides qui diffèrent par l'orientation des bases, par la nature des boucles et par la nature des sillons.

- **Orientation des bases nucléiques au sein des motifs G-quadruplexes :**

Au sein d'un motif G-quadruplexe, l'orientation des bases azotées qui dépend de l'angle de torsion glycosidique (*syn* ou *anti*) n'est pas due au hasard. Elle est régit par l'orientation des brins et permet de positionner les guanines de façon à favoriser leur appariement par liaisons hydrogène. Ainsi deux brins adjacents parallèles présentent leurs bases azotées selon des angles de torsion glycosidique similaires (*syn/syn* ou *anti/anti*) alors que deux brins adjacents antiparallèles les présentent selon des angles opposés (*syn/anti*). Dans les figures 10 et 11 les bases *anti* sont représentées en bleu et les bases *syn* en rose.

- **Nature des boucles au sein des motifs G-quadruplexes :**

La structuration des motifs G-quadruplexes bimoléculaires et intramoléculaires entraîne la formation de boucles dont il existe plusieurs sortes. Certaines peuvent se former au-dessus ou au-dessous des tétrades de guanines en reliant soit des brins adjacents (boucles latérales) soit des brins opposés (boucles diagonales) et d'autres peuvent être positionnées à l'extérieur de la structure (boucles externes) (figure 12). Ces boucles varient d'une structure G-quadruplexe à l'autre par leurs

positions au sein du G-quadruplexe et par la nature et le nombre de nucléotides dont elles sont constituées. Un motif G-quadruplexe peut contenir une à trois boucles qui sont généralement spécifiques d'une séquence donnée. Elles représentent donc des éléments de reconnaissance important dans le ciblage spécifique des G-quadruplexes.

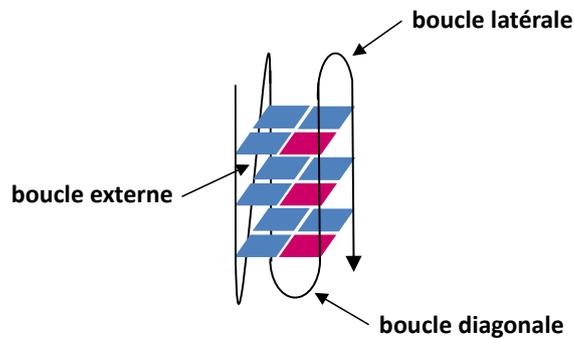


Figure 12 : représentation des différentes boucles pouvant se former au sein d'un motif G-quadruplexe

- **Nature des sillons au sein des motifs G-quadruplexes :**

Les sillons d'un motif G-quadruplexe varient en fonction de l'orientation des brins et donc de l'orientation des guanines au sein de la structure. Trois types de sillons sont recensés : étroit, moyen et large. Les combinaisons de guanines variant par leurs angles glycosidiques (*syn* ou *anti*) donnent accès à seize tétrades théoriques, représentées dans la figure 13. La forme et la largeur des sillons varient en fonction de la séquence du motif G-quadruplexe mais également en fonction de sa topologie. Chaque motif G-quadruplexe possède donc des sillons qui lui sont spécifiques, ce qui en font, tout comme les boucles, d'intéressants éléments de reconnaissance.

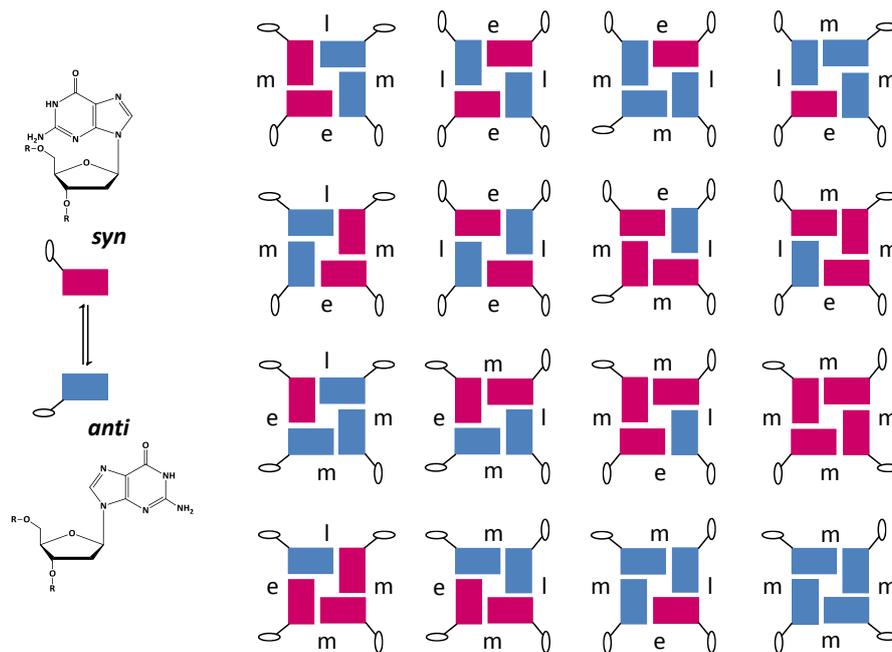


Figure 13 : tétrades théoriques suivant l'orientation des bases *anti* ou *syn* avec la représentation de leurs sillons suivant trois tailles : étroit (e), moyen (m) et large (l)

ii) Particularité des G-quadruplexes ARN

Les G-quadruplexes en série ARN apparaissent plus monomorphe que les G-quadruplexes en série ADN. En effet, depuis plusieurs années, de nombreuses études ont démontré que les G-quadruplexes d'ARN se replient préférentiellement en topologie parallèle^{14,15,16,17}. Une explication à ce phénomène, serait la déstabilisation de la conformation *syn* de la guanine par l'hydroxyle en C₂. Or cette conformation est nécessaire à la formation de G-quadruplexes antiparallèles ou hybrides. En 2006, l'équipe de R. H. Shafer¹⁸ a démontré une importante différence d'énergie entre les conformations *syn* et *anti* pour un G-quadruplexe d'ARN par rapport à un motif ADN. Ils concluent que cette différence d'énergie pourrait expliquer que la topologie parallèle soit largement favorisée par rapport à la topologie antiparallèle, qui d'après eux a besoin de beaucoup plus d'énergie pour se former.

Malgré toutes ces études, il a été découvert en 2014 un ARN G-quadruplexe¹⁹ de conformation antiparallèle, dans un aptamère ARN sélectionné en tant que ligand de la protéine GFP. Très récemment, l'équipe de Y. Xu a réussi à former un motif G-quadruplexes ARN antiparallèle dérivé de la séquence télomérique humaine²⁰. Pour ce faire, ils ont remplacé les guanines par des bases modifiées, les 8-bromoguanosines. Ils ont démontré que cette modification permettait d'orienter les guanines aussi bien en conformation *anti* que *syn*, rendant ainsi possible la formation de G4 ARN antiparallèle.

II.1.c. Détermination de la topologie des G-quadruplexes par absorption UV (TDS : Thermal Difference Spectra) et par Dichroïsme Circulaire (CD)

Au cours de cette thèse, la topologie des G-quadruplexes étudiée a été déterminée par trois techniques : l'absorption UV, la résonance magnétique nucléaire (RMN) et le dichroïsme circulaire (CD). Les deux premières techniques nous permettent de valider la formation d'un motif G-quadruplexe et la dernière de déterminer sa topologie. D'autres méthodes d'analyses qui n'ont pas été utilisées durant cette thèse sont également employées pour des études topologiques comme la RMN en deux dimensions²¹, la spectrométrie de masse²² ou encore des études de FRET pour « Fluorescence Resonance Energy Transfert »²³.

i) TDS : Thermal Difference Spectra

La technique « thermal difference spectra » (TDS) développée par J. L. Mergny permet de définir rapidement et avec de faible quantité de produit la structure adoptée par un acide nucléique. Elle consiste à enregistrer les spectres d'absorbances entre 200 et 350 nm d'un acide nucléique dans ses états repliés et dénaturés²⁴. En général, les spectres sont enregistrés à des températures largement inférieures et largement supérieures à la température de dénaturation (T_m). Le spectre TDS est obtenu par soustraction du spectre à basse température (état replié) au spectre enregistré à haute température (état dénaturé). Le TDS a une forme unique pour chaque type de structure d'acides nucléiques qui reflète les interactions d'empilement de base. La figure 14 présente l'exemple d'un TDS de duplexe d'ADN et d'un TDS de G-quadruplexe caractérisé par deux maximums aux alentours de 245 et 275 nm et d'un minimum vers 295 nm.

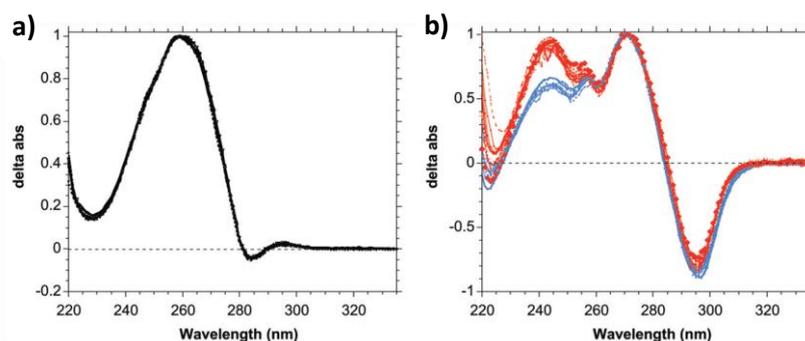


Figure 14 : exemples de TDS de (a) duplexe et de (b) G-quadruplexe d'acide nucléique source : Mergny, J.L. *et al Nucleic Acids Res.* **33**, 1-6 (2005)

ii) Dichroïsme circulaire

Le dichroïsme circulaire est une méthode d'analyse permettant la caractérisation des biopolymères et essentielle à l'étude des G-quadruplexes. Elle permet d'identifier la topologie

formée par un motif G-quadruplexe grâce aux signaux caractéristiques de chaque conformation. Dans la figure 15 sont représentés les signaux des trois topologies adoptées par les G-quadruplexes : parallèle, antiparallèle et hybride. La conformation parallèle est caractérisée par un maximum à 260 nm et un minimum à 240 nm (figure 15.a). Le signal spécifique d'une topologie antiparallèle présente deux maximums à 240 et 290 nm et un minimum à 260 nm (figure 15.b). Enfin, la structuration des G-quadruplexe en conformation hybride est caractérisée par deux maximums à 260 et 290 nm et un minimum à 240 nm (figure 15.c). Cette différence de signal provient de la modification de l'interaction π -stacking entre les différentes tétrades, d'une topologie G-quadruplexe à l'autre, du fait de la variation des angles glycosidiques (*syn* ou *anti*)^{25,26}. Il est à noter que le signal CD est un signal moyen, en cas de mélange de plusieurs topologies G-quadruplexe, la nature de ces dernières est impossible à déterminer. Par exemple, le mélange de structures parallèles (50%) et antiparallèles (50%) pourrait ressembler à un signal hybride. Le CD est donc souvent associé à une autre méthode pour déterminer avec plus de certitude la structure d'un G-quadruplexe.

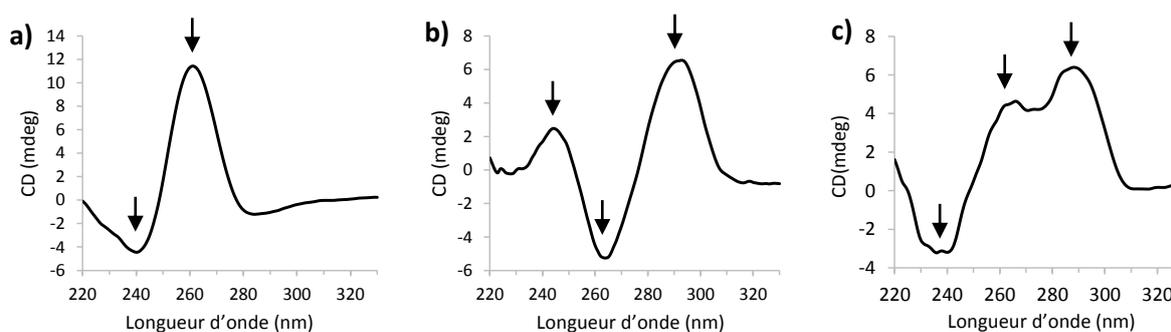


Figure 15 : exemples de spectres CD de G-quadruplexe parallèle (a) antiparallèle (b) et hybride (c)

iii) Résonance magnétique nucléaire

La RMN en 1D permet de valider la structuration d'un motif G-quadruplexe. Les signaux caractéristiques des G-quadruplexes sont retrouvés aux alentours des 10-12 ppm. Ils correspondent au déplacement chimique des protons-iminos non-échangeables des motifs G-quadruplexes (figure 16).

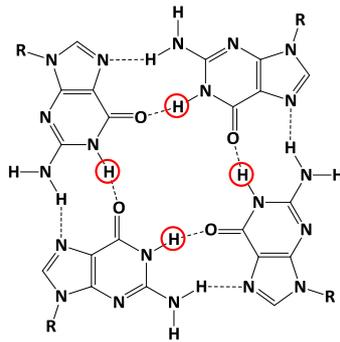


Figure 16 : schéma d'une tétrade de guanine avec les protons iminos entourés en rouge

II.2. Structures des G-triplexes d'acides nucléiques

Il a été mis en évidence depuis peu que lors de leur repliement en motifs G-quadruplexes, les acides nucléiques pourraient adopter une forme intermédiaire nommée G-triplexe^{27,28}.

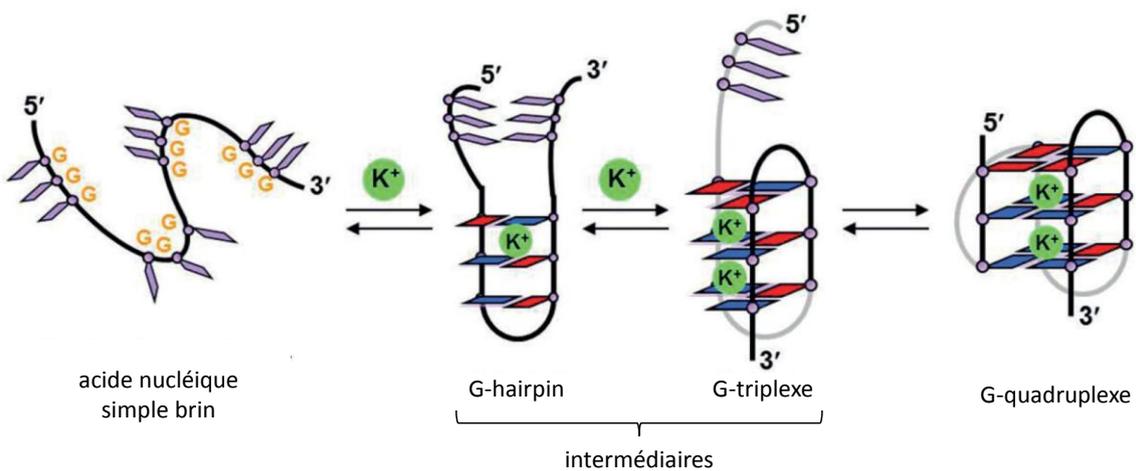


Figure 17 : processus de formation des G-quadruplexes source : Rajendran, A. *et al Angew. Chem. Int. Ed.* **53**, 4107-4112 (2014)

La figure 17 schématise le processus suggéré par l'équipe d'H. Sugiyama pour le repliement d'un G-quadruplexe à partir d'un acide nucléique simple brin. Il s'agit d'un processus en plusieurs étapes qui fait intervenir deux états intermédiaires, un G-hairpin et un G-triplexe. Les G-triplexes sont formés de l'empilement par π -stacking de triades de guanines (figure 18.b). Ces dernières sont composées d'une guanine centrale appariée à deux autres par des liaisons hydrogène de type Hoogsteen (figure 18.a).

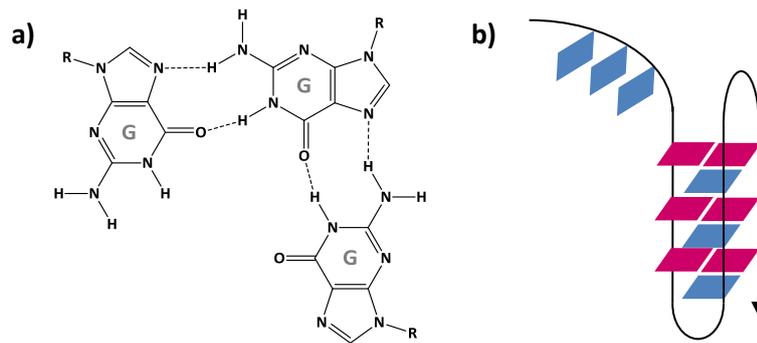


Figure 18 : (a) triade de guanine et (b) exemple de G-triplexe

Ces structures, tout comme les motifs G-quadruplexes sont également stabilisées par des cations métalliques. Une étude récente²⁹ a mis en évidence la stabilisation des G-triplexe par les cations classiques K^+ et Na^+ et par deux cations bivalents Ca^{2+} et Mg^{2+} selon l'effet stabilisant suivant : $Ca^{2+} > K^+ > Mg^{2+} > Na^+$.

II.3. Structures des i-motifs d'acides nucléiques

Il a été démontré en 1993 par l'équipe de M. Guéron³⁰ que dans un double brin d'acide nucléique, la séquence complémentaire à celle des G-quadruplexes, par conséquent riche en cytosines, était capable de former une structure secondaire particulière, appelée i-motif (figure 19).

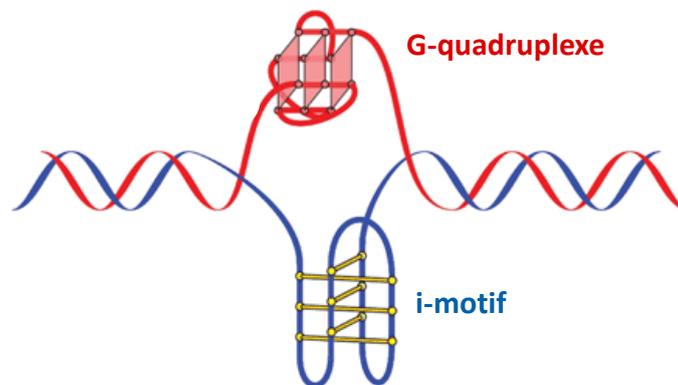


Figure 19 : exemple de double brin d'acide nucléique structuré en G-quadruplexe et en i-motif

Cette structure possède quatre brins appariés deux à deux, dont deux sont orientés dans un sens et les deux autres dans le sens opposé. Les i-motifs sont formés par l'intercalation de paires de cytosines les unes aux dessus des autres (figure 20.b et 20.c). L'appariement des pyrimidines deux à deux est régi par trois liaisons hydrogène et est favorisé à pH acide puisqu'il nécessite la protonation d'une des deux cytosines (figure 20.a).

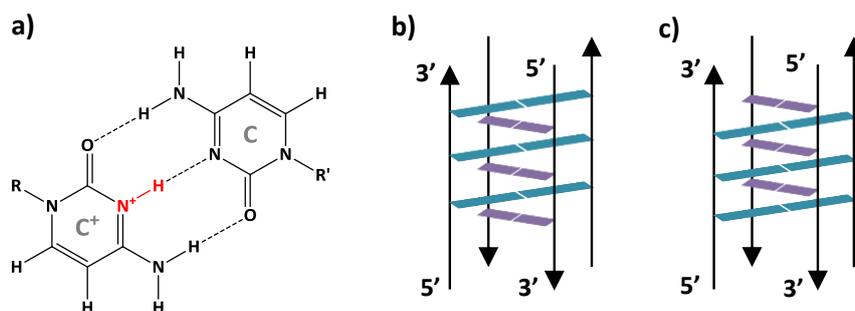


Figure 20 : (a) couple de cytosine, (b) i-motif 3'E et (c) i-motif 5'E

Sept ans après la découverte des i-motifs, l'équipe de M. Guéron³¹ démontre que d'une séquence à l'autre ces structures adoptent toujours la même topologie antiparallèle, mais différent par le nombre de paires de bases, par la structure des boucles et par la façon dont s'intercale les couples de cytosines. Pour une même séquence, cette intercalation des paires de bases peut conduire à la formation de deux i-motifs différents. En effet, si le dernier couple de cytosine est placé en 3' des brins, il s'agit de i-motif 3'E (figure 20.b) et s'il est placé en 5' des brins, le i-motif est qualifié de 5'E (figure 20.c). En 2012, l'équipe d'H. Schwalbe³² a montré que la conformation 3'E était la plus favorable cinétiquement, par rapport à la forme 5'E qui semble être la plus stable.

En raison de leur repliement dépendant du pH, il a longtemps été pensé que les i-motif étaient instables au pH physiologique. De récentes études ont montré que ce n'était pas toujours le cas et que la stabilité de la structure pouvait également dépendre d'autres facteurs, tels que la séquence et les conditions environnementales. Il a été prouvé que dans certaines séquences (par exemple la séquence télomérique humaine) la présence d'interactions entre des bases non cytosines (par exemple A-T) pouvait augmenter la stabilité du i-motif³³. Il a également été mis en évidence que de fortes conditions d'encombrement moléculaire ou « molecular crowding », favoriseraient la stabilité des i-motifs à pH neutre^{34,35}. En 2009, J. Zhou et *al.*³⁶ ont mis en évidence la formation de i-motifs à pH neutre, en diminuant la température du milieu (4°C). Tous ces travaux démontrant que les i-motifs peuvent être formés aussi bien à pH physiologique, qu'à pH acide laissent entrevoir un possible rôle biologique de ces structures.

III. Localisation des G-quadruplexes

Plusieurs équipes de recherche ont récemment mis en évidence, directement ou indirectement, l'existence des G-quadruplexes dans les organismes vivants.

L'équipe d'A. Plückthu a découvert en 2005 la présence de G-quadruplexes chez les ciliés, organismes unicellulaires appartenant à la famille des eucaryotes³⁷. Les G-quadruplexes ont été

détectés à l'aide d'anticorps fluorescents ciblant les G4 télomériques de *Styloynchia lemnae*. Quelques années plus tard, l'équipe d'H. J. Lipps³⁸ a démontré le rôle de certaines protéines télomériques dans la formation de ces G-quadruplexes, évoquant ainsi l'implication des G-quadruplexes dans le métabolisme des télomères chez les Ciliés.

En 2013, l'équipe de S. Balasubramanian³⁹ a mis en évidence la présence de G-quadruplexes ADN dans des cellules humaines. Pour ce faire, les chercheurs ont sélectionnés *in-vitro* un anticorps (BG4) ayant une très forte affinité pour les G-quadruplexes d'ADN et l'ont ensuite utilisés pour visualiser des structures G-quadruplexes dans l'ADN génomique de cellules humaines. Dans cette étude, ils ont également mis en évidence la relation entre le cycle cellulaire et la formation des G4. En effet, il semblerait que la formation des G-quadruplexes soit plus importante au moment de la réplication. La séparation de l'ADN duplexe en deux monobrin, faciliterait le repliement des motifs G-quadruplexes. Un an plus tard, la même équipe⁴⁰ a montré que l'anticorps BG4 pouvait également se lier au G4 ARN et a permis de visualiser de l'ARN structuré en G-quadruplexe dans le cytoplasme de cellules humaines. En 2014, l'équipe de P. M. Lansdorp⁴¹ a produit le deuxième anticorps spécifique des G-quadruplexes, le 1H6, avec lequel, ils ont visualisé l'abondance des G-quadruplexes dans les cellules humaines.

L'équipe de D. Monchaud travaille également sur la détection et la visualisation de G-quadruplexes à l'intérieur de la cellule. Leur approche n'utilise pas d'immunodétection, mais un ligand de G4, le NaphtoTASQ, qui agit comme une sonde fluorescente à activation multiphotonique. Dans leurs récents travaux^{42,43}, D. Monchaud et ses collaborateurs rapportent la visualisation en direct d'ARN et d'ADN G-quadruplexes dans des cellules vivantes *via* des méthodes de spectroscopie.

Toutes ces études fournissent des preuves substantielles de la formation de structures G-quadruplexes dans le génome humain. Elles complètent et renforcent les travaux effectués par un grand nombre d'équipes visant à identifier dans quelle partie du génome se trouvent les G-quadruplexes et à comprendre leurs rôles dans la régulation cellulaire. Selon des études^{44,45} publiées en 2005, il y aurait plus de 370000 séquences capables de former des G-quadruplexes dans le génome humain. Les paragraphes suivants présentent des exemples de localisation des G-quadruplexes.

III.1. G-quadruplexes localisés dans la région télomérique

Les chromosomes possèdent en leur extrémité une région non codante et hautement répétitive, appelé télomère (figure 21), qui les protègent en empêchant notamment deux chromosomes de fusionner. Chez l'humain la séquence répétitive est d(TTAGGG) sur une longueur de

La région télomérique est ainsi une région riche en G-quadruplexes, dont leurs rôles dans la régulation des télomères et de la télomérase ont été démontrés, faisant de ces structures d'intéressantes cibles biologiques. Durant cette thèse nous nous sommes intéressés à l'étude et au ciblage de quelques motifs G-quadruplexes : le motif G4 de la séquence d'ADN télomérique, le motif G4 de la séquence d'ARN télomérique et le motif G4 hybride ADN/ARN.

III.2. G-quadruplexes localisés dans les génomes bactériens et viraux

De récentes études ont montré que les G-quadruplexes n'étaient pas seulement présents dans le génome humain, mais qu'ils pouvaient être retrouvés dans les génomes de protozoaires⁵⁶, de bactéries⁵⁷ ou encore de virus⁵⁸. En ce qui concerne la dernière famille d'agents pathogènes, des séquences G-quadruplexes ont été localisées dans les génomes du virus de l'immunodéficience humaine (VIH-1)^{59,60}, du papillomavirus⁶¹ et du virus de l'Herpes⁶². Très récemment, des G-quadruplexes ont également été découverts dans les génomes des virus à l'origine des dernières épidémies, telles que le virus Ebola⁶³ et le virus Zika⁶⁴. Tout comme les stratégies anticancéreuses visant les G-quadruplexes télomériques, des études suggèrent que les motifs G-quadruplexes présents dans l'ADN ou l'ARN viral pourraient présenter d'intéressantes cibles biologiques. Leur ciblage par des ligands spécifiques est donc pertinent et pourrait permettre de réguler ou encore d'inhiber par la stabilisation des G-quadruplexes, certaines étapes clés dans le développement du virus. Ces études ouvrent la voie au développement de nouvelles stratégies de lutte contre les maladies virales : stratégies qui sont actuellement explorées dans le contexte de l'infection par le VIH-1^{65,66}. Durant cette thèse, nous nous sommes intéressés à l'étude et au ciblage du motif G-quadruplexe localisé dans le génome du VIH-1, au même titre que les G-quadruplexes télomériques.

III.3. Les autres G-quadruplexes génomiques

Dans le génome humain, les G-quadruplexes sont également présents dans certains promoteurs d'oncogènes comme c-kit⁶⁷, c-myc⁶⁸ ou encore KRAS⁶⁹. Ces motifs interviennent également dans les stratégies anticancéreuses puisqu'ils sembleraient avoir une influence sur la régulation des gènes. En effet, leur stabilisation au niveau de la partie promotrice des oncogènes inhiberait la transcription de ces derniers⁷⁰.

Des G-quadruplexes ont aussi été retrouvés dans les régions non codantes des ARN messagers, 5' UTR^{71,72} et 3' UTR^{73,74} ainsi que dans certaines régions codantes (ORFs)⁷⁵. La stabilisation de motifs G4 dans ces régions pourrait permettre l'inhibition de la traduction de l'ARNm

en protéines^{76,77,78}. Ces motifs pourraient représenter des cibles intéressantes dans le cas de fabrication de protéines cancéreuses.

Des motifs G-quadruplexes ont de même été retrouvés dans les régions ILPR (insulin-linked polymorphic region), connues pour réguler la transcription du gène codant pour l'insuline. En 2009, l'équipe de H. Mao⁷⁹ a révélé par des études de dichroïsme circulaire la coexistence de G4 parallèle et antiparallèle dans cette région, en présence de 100 mM KCl. Les chercheurs ont mis en évidence que les forces liées à la structuration des motifs G-quadruplexes étaient plus élevées que les forces liées à leur déroulement qui sont provoquées par des enzymes ayant des propriétés hélicase. Grâce à ces données ils ont émis l'hypothèse d'une inhibition de la transcription et donc de la production de molécules d'insuline par la stabilisation des G-quadruplexes.

IV. Localisation des i-motifs

Les i-motifs sont moins étudiés que les motifs G-quadruplexes mais semblent présenter un intérêt biologique tout aussi important. Par exemple, il a été démontré par l'équipe de L. H. Hurley que la formation d'une structure en i-motif dans la région du promoteur des gènes *BCL2*⁸⁰ et *Myc*⁸¹ participait à la régulation de ces gènes.

En 2002, l'équipe de J.-L. Mergny⁸² a démontré qu'à pH acide la séquence complémentaire à celle des G-quadruplexes pouvait former un i-motif et que cette structure pourrait tout comme les G-quadruplexes être impliquée dans la régulation de la télomérase.

L'équipe d'H. Mao qui a mis en évidence la présence de G-quadruplexes dans la région ILPR en 2009 a identifié un an plus tard l'existence de i-motif dans cette même région⁸³. Ils ont démontré que tout comme les G4, les i-motifs stabilisés provoquaient l'inhibition de la transcription et donc de la production de molécules d'insuline.

Dans l'ensemble, les i-motifs, séquences complémentaires aux G-quadruplexes sont retrouvés aux mêmes endroits et sembleraient présenter les mêmes rôles biologiques. Tout comme les G4, les i-motifs sont devenus de véritables cibles biologiques. La conception et l'identification de ligands spécifiques capables de les stabiliser, constitue une piste thérapeutique prometteuse.

V. Ligands de G-quadruplexes

Il a été montré précédemment que les G-quadruplexes sont largement répandus dans le génome humain et qu'ils participent à la régulation d'un grand nombre de processus biologiques. De ce fait, la conception de molécules pouvant interagir efficacement et sélectivement avec ces motifs est en pleine expansion. L'objectif des équipes de recherches est de trouver des molécules ayant une très forte affinité pour les G-quadruplexes (certains ligands présentent une affinité au nanomolaire) et une très forte sélectivité par rapport au duplex d'ADN.

Les ligands sont bien souvent de petites molécules organiques, ayant des propriétés qui leur permettent d'interagir avec les G-quadruplexes. Au sein d'un motif G-quadruplexe, les ligands peuvent interagir avec les tétrades de guanines, avec le squelette sucre-phosphate des sillons ou encore avec les bases nucléiques des boucles (figure 22).

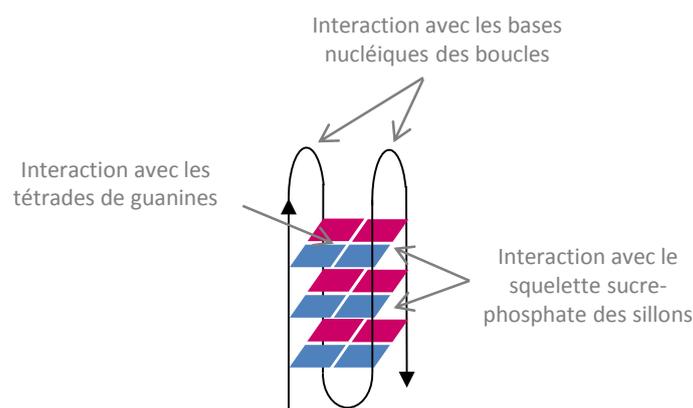


Figure 22 : sites d'interactions des ligands sur un G-quadruplexe

La plupart des ligands possèdent un noyau aromatique capable de se lier par π -stacking sur les plateaux supérieurs des G-quadruplexes. Ce mode de liaison par empilement externe, représente une caractéristique unique dans la reconnaissance des G4 par rapport au duplex d'ADN. Pour favoriser les interactions π -stacking et assurer la sélectivité par rapport au duplex d'ADN, les ligands doivent comporter une grande surface aromatique. En plus de ce noyau aromatique, ils présentent très souvent des charges positives qui leur permettent d'interagir avec les G-quadruplexes *via* des liaisons électrostatiques. Certains ligands sont également dotés de charge ponctuelle cationique (un métal de transition par exemple), capable de stabiliser le centre du G4. La conception rationnelle de ligands, guidée par ces deux critères : interactions π -stacking et interactions électrostatiques, a conduit à la synthèse d'un très grand nombre de petites molécules organiques. On retrouve ainsi des complexes métalliques, des dérivés de télomestatine, des dérivés tricycliques, des dérivés d'acridines, des dérivés de naphthalènes ou encore des dérivés de macrocycles^{84,85,86}. Bien qu'un très

grand nombre de ligands ait été synthétisés, aucun n'a encore trouvé d'application thérapeutique. Ceci révèle la difficulté de synthétiser un ligand spécifique et sélectif répondant à toutes les obligations thérapeutiques. Le Quarfloxin⁸⁷ (figure 23), ciblant le motif G4 du promoteur d'oncogène *c-myc* a par exemple été arrêté en phase clinique II pour des problèmes de pharmacocinétique.

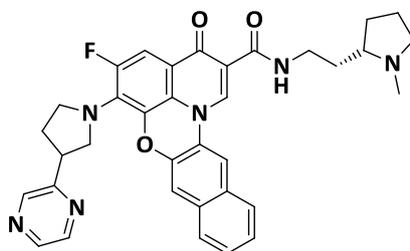


Figure 23 : structure du Quarfloxin

V.1. Les différents ligands de G-quadruplexes étudiés

Cette partie présente une liste non-exhaustive des ligands de G-quadruplexes actuellement développés. En effet, sont décrites uniquement les petites molécules qui ont été étudiées durant cette thèse.

V.1.a. Les dérivés de salophen

Un certain nombre de dérivés salophen ont déjà été décrit^{88,89}, parmi tous ces ligands de G-quadruplexes existant, deux ont été étudiés dans cette thèse. Il s'agit de salophen métalliques (L8 et L14) (figure 24) synthétisé par l'équipe de F. Thomas^{90,91}. Ces complexes possèdent un atome de Nickel et des bras d'alkyl-imidazolium orientés dans différentes directions. Ces ligands ont été conçus pour interagir par π -stacking et par liaisons électrostatiques avec les motifs G-quadruplexes. L'équipe a démontré par modélisation moléculaire que les interactions π -stacking ont lieu entre le noyau aromatique et les tétrades de guanines et les interactions électrostatiques entre les bras chargés et les sillons des motifs G-quadruplexes. Il a également été démontré que l'orientation des bras d'alkyl-imidazolium jouait un rôle important dans la liaison aux G-quadruplexes. Il semblerait donc que ces bras puissent induire une discrimination des motifs G-quadruplexes en fonction de leur topologie. Ces deux composés présentent une grande sélectivité des G4 par rapport au duplex d'ADN. De par leur forte affinité pour les G4 et notamment pour les G4 télomériques ces ligands sont capables d'inhiber *in vitro* la télomérase (IC_{50} L8 = 1 μ M, IC_{50} L14 = 70 nM) dans un test d'inhibition de la télomérase nommé TRAP.

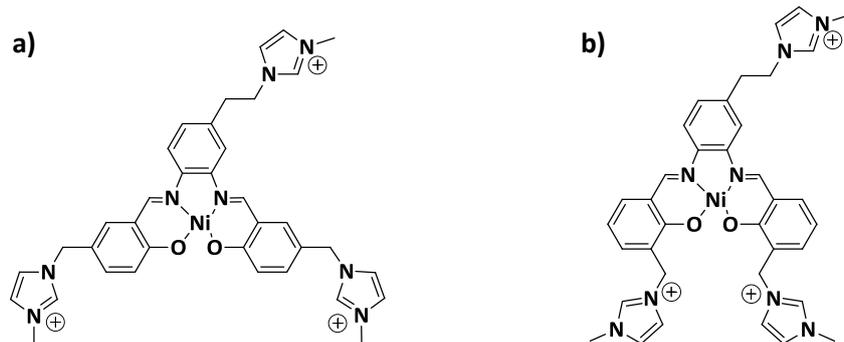


Figure 24 : structure des ligands salophen métalliques (a) L8 et (b) L14

V.1.b. Les dérivés de porphyrines

De nombreux ligands sont dérivés de la porphyrine, ce qui leur confère une grande surface aromatique pouvant interagir par π -stacking avec les tétrades de guanines. Certains de ces composés possèdent des substituants chargés sur le noyau porphyrine, leur permettant d'interagir avec les sillons des motifs G-quadruplexes. C'est le cas de TMPyP4 (figure 25) qui a été développé en 1998 par l'équipe de L. H. Hurley⁹². Ce ligand possède une forte affinité pour les G4, de l'ordre du μM . Il est le premier à avoir été testé sur la séquence télomérique humaine et est capable d'inhiber efficacement la télomérase avec un IC_{50} proche des $10 \mu\text{M}$ dans le test d'inhibition *in-vitro* TRAP. Il est également capable de réduire l'expression de certains oncogènes tel que *c-myc*, ou encore de convertir des G-quadruplexes antiparallèles en G-quadruplexes parallèles. Il est intéressant de noter que malgré son manque de sélectivité vis-à-vis des duplexes d'ADN⁹³, l'intérêt pour cette molécule particulière n'a jamais diminué.

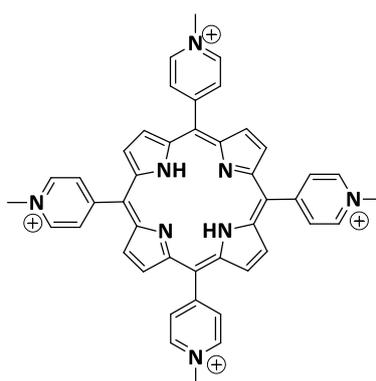


Figure 25 : structure de TMPyP4

V.1.c. Les dérivés d'acridine

Les dérivés d'acridine possèdent un noyau aromatique composé de trois cycles, très souvent substitués par des bras chargés. BRACO-19 est le plus étudié de ces dérivés et possède trois fonctions amines capables d'interagir avec les sillons des motifs G4 (figure 26). Ces substituants sont reliés sur un corps aromatique pouvant interagir par π -stacking avec les tétrades de guanines des motifs G4. Ce composé possède une bonne affinité pour les G4, de l'ordre du micromolaire ainsi qu'une très bonne sélectivité par rapport au duplexe. Des études ont également révélé une activité anticancéreuse par la stabilisation des G4 dans la région télomérique⁹⁴ ainsi qu'une activité antivirale due à la stabilisation des G4 du génome du VIH⁶⁵.

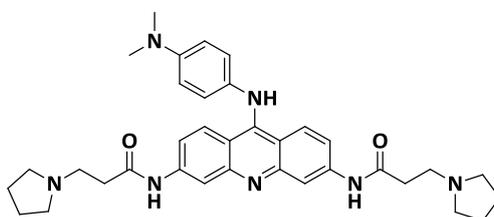


Figure 26 : structure de BRACO-19

Un second dérivé d'acridine et plus précisément de quinacridine a été étudié durant cette thèse, il s'agit de MMQ1⁹⁵ (figure 27). Ce composé dont l'interaction avec les G4 est dominée par le π -stacking présente une bonne stabilisation des motifs G-quadruplexes et agit comme inhibiteur de la télomérase.

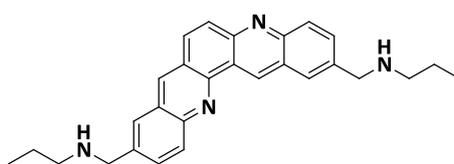


Figure 27 : structure de MMQ1

V.1.d. Les dérivés de bisquinolinium

Parmi les composés comportant deux fonctions quinoliniums, deux composés se distinguent : pyridostatine (PDS) (figure 28.a) et PhenDC3 (figure 29). Le premier possède un fort pouvoir stabilisant et une bonne sélectivité des G-quadruplexes vis-à-vis des duplexes. Il a été l'un des premiers ligands étudiés comme anti-cancéreux⁹⁶. En se liant au G4 télomérique, PDS entraîne l'altération de certaines protéines shelterins (POT1 et TRF1) déclenchant ainsi le mécanisme de réponse au dommage de l'ADN conduisant à la sénescence cellulaire.

Une étude récente a montré qu'avec un substituant acide carboxylique en plus, le ligand carboxyPDS (figure 28.b) est capable de reconnaître des G-quadruplexes ARN présents dans le cytoplasme des cellules humaines⁴⁰.

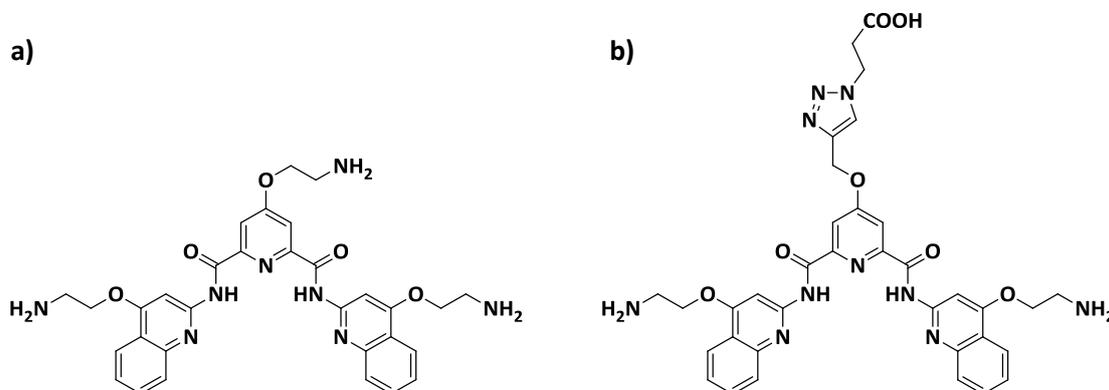


Figure 28 : (a) structure de PDS, (b) structure de carboxyPDS

Le ligand PhenDC3 (figure 29) possède un très fort pouvoir stabilisant et une très forte sélectivité pour les G-quadruplexes, ce qui confirme le grand potentiel des dérivés de bisquinolinium. Ce ligand a été synthétisé par l'équipe de M.P. Teulade-Fichou avec l'idée de créer une correspondance géométrique parfaite avec les tétrades de guanines⁹⁷. Il a été utilisé pour identifier les motifs G-quadruplexes au sein de cellules mammifères. Son interaction avec un motif G-quadruplexe formé dans le promoteur de l'oncogène *c-myc* a été résolue par des études RMN à haute résolution révélant une interaction π -stacking entre les ligands et la tétrade externe⁹⁸.

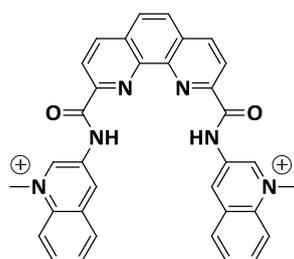


Figure 29 : structure de PhenDC3

VI. Ligands des i-motifs

De manière générale, il existe très peu de ligands capables de cibler les i-motifs. Le premier exemple de composé stabilisant un i-motif a été publié en 2000 par l'équipe de L. H. Hurley⁹⁹. En utilisant les séquences télomériques humaines d(CCCAAT)₄ et d(AATCCC)₄, les chercheurs ont étudié les propriétés de liaison du ligand porphyrine TMPyP4 sur les i-motifs. Ils ont démontré que ce

composé était capable de favoriser la formation de i-motif à pH 4,5. Des expériences de modélisation ont suggéré une liaison de TMPyP4 par π -stacking sur l'extrémité des motifs.

Les nanotubes de carbone (SWNT) comportant des modifications de type carboxy figurent parmi les ligands les plus étudiés (figure 30). Ils sont capables de stabiliser une structure i-motif à pH 5,5¹⁰⁰ en solution et à pH physiologique dans des conditions d'encombrement moléculaire (« molecular crowding »)¹⁰¹. En raison de leur taille, ils se lient dans le sillon principal du i-motif, à l'extrémité 5' par interactions électrostatiques entre les cytosines protonées et leurs groupements carboxy. Les SWNT sont spécifiques aux i-motifs et n'interagissent ni avec les G-quadruplexe ni avec les duplexes d'ADN. Il a été démontré que ces composés pourraient inhiber l'activité de la télomérase et interférer avec les fonctions des télomères dans les cellules cancéreuses. L'inhibition observée n'est en réalité pas directement liée à la formation de structure i-motif. En effet, en stabilisant les i-motifs, les ligands favorisent la dissociation du duplexe d'ADN, ce qui induit la formation de G-quadruplexes, connus pour inhiber la télomérase¹⁰².

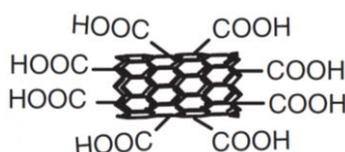


Figure 30 : représentation des SWNT

L. H. Hurley et ses collaborateurs ont récemment montré que la séquence BCL-2 présente un équilibre entre une forme i-motif et une structure hairpin⁸⁰. Ils ont criblé une bibliothèque de 1990 composés contre le i-motif BCL-2. A partir de quatorze succès initiaux ils ont identifié deux composés, IMC-48 (figure 31.a) comme ligand stabilisant du i-motif et IMC-76 (figure 31.b) en tant que ligand déstabilisant. Des études plus poussées sur les effets biologiques de ces ligands ont montré que leurs interactions avec BCL-2 modulent la quantité de i-motif disponible pour se lier à hnRNP LL, qui régule la transcription du gène¹⁰³.

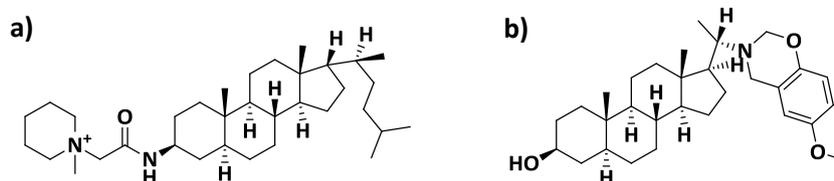


Figure 31 : (a) structure de IMC-48, (b) structure de IMC-76

Très récemment, l'équipe de Z. A. E. Waller¹⁰⁴ a montré que la mitoxantrone (figure 32) était capable d'induire le repliement et la stabilisation de séquences d'ADN formant des i-motifs. Cette molécule est à l'origine utilisée dans le traitement de certains cancers, principalement le cancer du

sein métastatique, la leucémie myéloïde aiguë et le lymphome non-hodgkinien, mais aussi dans le traitement des formes agressives de sclérose en plaques. Ce médicament est un inhibiteur de la topoisomérase II, une enzyme qui contrôle la structure topologique de l'ADN. L'inhibition de son activité entraîne l'arrêt de la synthèse et de la réparation de l'ADN dans les cellules cancéreuses, mais également dans les cellules saines. La mitoxantrone agit sur l'ADN via son effet intercalant, qui lui permet de se lier aux structures i-motifs. Les résultats d'études d'interactions par SPR ont montré que la mitoxantrone avait une affinité modérée pour les i-motifs de la séquence télomérique et de *c-myc* à pH 5,5, aux alentours des 10 μ M. Ce ligand est cependant sélectif aux i-motifs par rapport au duplexe d'ADN.

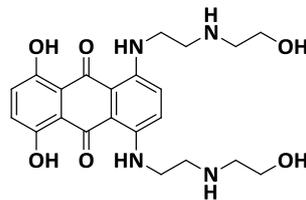


Figure 32 : structure de la mitoxantrone

VII. Concept TASQ

L'une des difficultés majeures dans la conception de ligands ciblant les motifs G-quadruplexes est la grande variété de topologies que peuvent adopter ces structures en solution. Ce polymorphisme inhérent aux motifs G-quadruplexes complique le développement et l'étude de ligands spécifiques et affins. Un exemple de ce polymorphisme est la séquence de l'ADN télomérique qui peut adopter différentes conformations. Sept topologies ont à ce jour été isolées et caractérisées, elles sont représentées dans la figure 33¹⁰⁵.

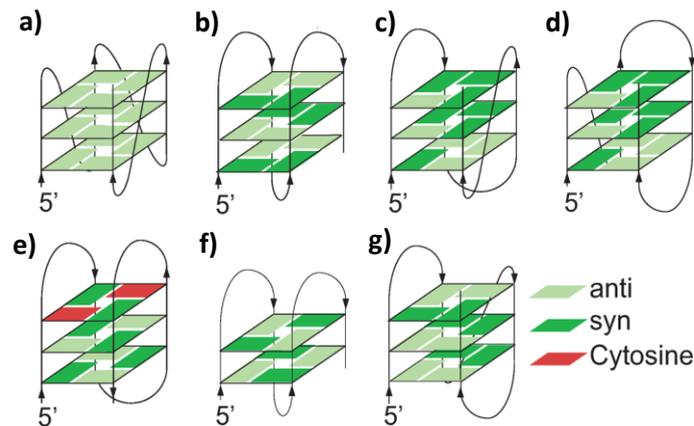


Figure 33 : représentations schématiques des différentes topologies G-quadruplexes adoptées par la séquence télomérique humaine : (a) conformation parallèle, (b) conformation basket antiparallèle, (c) conformation hybride 1 [3 + 1], (d) conformation hybride 2 [3 + 1], (e) conformation chaise, (f) deux tétrades en conformation antiparallèle et (g) conformation antiparallèle [2 + 2], source : Noer, S. L. *et al. Nucleic Acids Res.* **44**, 464-471 (2016)¹⁰⁵

Cette diversité structurale est exacerbée *in vitro* par les conditions expérimentales utilisées, telles que la concentration, la température ou encore la nature du cation métallique. Il a par exemple été démontré que le sodium favorise la conformation basket antiparallèle, alors que le potassium engendre un repliement des G-quadruplexes en topologies parallèle ou hybrides. Ce polymorphisme rend la conception rationnelle de ligands extrêmement difficile et malgré la synthèse de plusieurs centaines de petites molécules, aucune d'entre-elles n'a abouti à une molécule thérapeutique. Afin de pallier les phénomènes de polymorphisme des motifs G-quadruplexes, le laboratoire a développé il y a quelques années un système biomoléculaire permettant le contrôle de la topologie des G-quadruplexes. Ce concept nommé TASQ pour « Template Assembled Synthetic G-Quadruplex » consiste à utiliser une plateforme peptidique rigide et soluble dans l'eau pour la stabilisation et la préorganisation des motifs G-quadruplexes. Cette plateforme est un cyclodécapeptide connu sous le nom de « RAFT » pour « Regioselectively Addressable Fonctionnalised Template »¹⁰⁶.

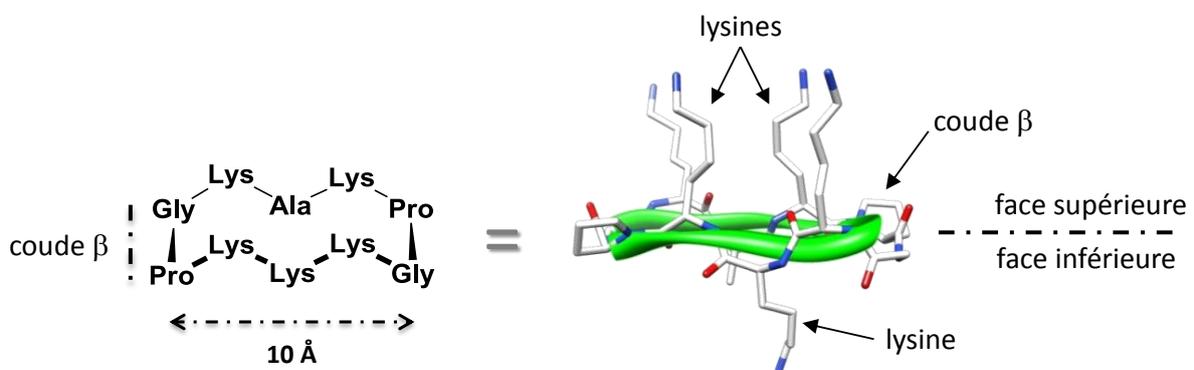


Figure 34 : structure chimique et modélisation du RAFT

Sa rigidité est apportée par la présence de deux coudes β proline-glycine qui induisent une structuration en feuillet β (figure 34). Les chaînes latérales des six résidus centraux, qui sont essentiellement des résidus lysines ou des alanines, sont orientées de part et d'autre du plan moyen du peptide cyclique, ce qui crée deux faces indépendantes d'adressage chimique. Cette orientation spécifique des chaînes latérales des lysines naturelles ou modifiées permet le greffage distinct d'éléments chimiques ou de molécules sur chaque face. Dans le cas du concept TASQ, la face supérieure permet de lier par différentes méthodes de bioconjugaisons des oligonucléotides formant le motif G-quadruplexe et la face inférieure permet l'immobilisation de l'édifice *via* une biotine lors d'étude d'interactions par résonance de plasmon de surface (SPR) (figure 35). Au-delà du concept TASQ, cette plateforme peptidique d'adressage est largement utilisée dans notre équipe pour l'assemblage d'agents de ciblage et de reconnaissance multivalents^{107,108}.

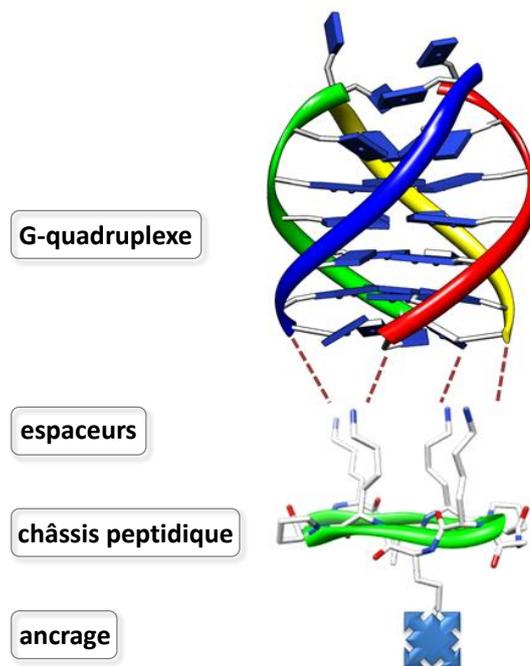


Figure 35 : représentation schématique du concept TASQ

VIII. Ligations chimiosélectives

La liaison des oligonucléotides sur le châssis peptidique s'effectuent par des ligations chimiosélectives. Ces réactions mettent en jeu deux fonctions chimiques ayant une réactivité réciproque et spécifique. Au-delà de l'impératif d'orthogonalité chimique associée à la synthèse de tout bioconjugué, l'assemblage d'édifice à base d'acide nucléique structuré, nécessite l'utilisation de ligations chimiques qui respectent des critères supplémentaires. Tout d'abord, l'encombrement

stérique des liaisons chimiques formées doit être pris en compte pour ne pas affecter le repliement des oligonucléotides. Pour les mêmes raisons, les fonctions chimiques complémentaires doivent être installées sur des espaceurs de longueurs comparables. Enfin, pour réaliser des cyclisations intramoléculaires permettant de former des structures à boucles, la ligation chimique utilisée doit pouvoir être réalisée avec un rendement élevé dans des conditions stœchiométriques.

En théorie, il est possible d'utiliser quatre ligations chimiosélectives différentes, mais en pratique il est compliqué de combiner quatre ligations orthogonales répondant à toutes les conditions que nécessitent la chimie des oligonucléotides et des peptides. Deux ligations chimiosélectives étaient déjà utilisées aux laboratoires pour l'assemblage de conjugués peptide-oligonucléotides. Il s'agit de la ligation oxime et de la CuAAC qui sont orthogonales entre elles et parfaitement compatible avec la chimie des oligonucléotides et des peptides.

- **Ligation oxime :**

Cette ligation est une réaction d'addition nucléophile, d'une fonction nucléophile (oxyamine) sur une fonction comportant un carbone électrophile (aldéhyde) (schéma 1)¹⁰⁹. La présence d'un hétéroatome donneur d'électrons (l'oxygène) en position α de l'amine permet de stabiliser le lien éther d'oxime formé dans une large gamme de pH allant de 3 à 9. La formation du lien est favorisée à pH acide (pH optimal 4-5). De par sa forte réactivité et la formation d'un lien stable et peu encombrant cette ligation est aujourd'hui classiquement utilisée pour l'obtention de conjugués oligonucléotidiques.

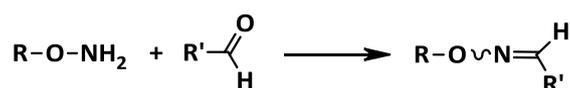


Schéma 1 : ligation oxime

- **CuAAC :**

La cycloaddition 1,3-dipolaire de Huisgen catalysée par le Cu^I (CuAAC pour Copper catalyzed Azide Alkyne Cycloaddition) a été développée par R. Huisgen en 1963¹¹⁰. Cette réaction est une cycloaddition [2+3] entre une fonction alcyne et une fonction azoture, qui permet l'obtention d'un lien triazole 1,4 disubstitué (schéma 2). En présence d'oligonucléotide, elle est essentiellement réalisée avec du CuSO₄ donc du Cu^{II}, réduit in-situ par de l'ascorbate de sodium. Un ligand du cuivre est très souvent utilisé pour faciliter la solubilité du métal et stabiliser sa forme réduite Cu^I. De nombreux ligands présentant généralement des fonctions amines substituées ou cycliques ont été décrits. Durant cette thèse, nous avons utilisé l'un des plus courants, le THPTA. La CuAAC peut être réalisée avec de très bons rendements dans des conditions douces (milieu aqueux, 37°C et pH 7,4).

Elle est également orthogonale vis-à-vis de très nombreuses fonctions chimiques. De plus, grâce à la formation d'un lien triazole peu encombrant elle est couramment utilisée pour la bioconjugaison de biomolécules.

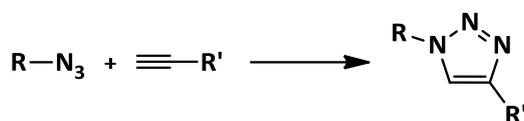


Schéma 2 : réaction CuAAC

L'utilisation de seulement deux ligations pour la conception de bioconjugué limite les possibilités de représentation des motifs G-quadruplexes. Afin d'arriver à contrôler totalement la structure des G-quadruplexes, il serait nécessaire d'utiliser quatre ligations chimiosélectives différentes. Avant d'arriver à combiner orthogonalement quatre ligations, nous nous sommes appliqués à trouver une troisième ligation compatibles avec la chimie des acides nucléiques et des peptides, mais également compatible chimiquement avec la ligation oxime et la CuAAC.

Parmi toutes les ligations utilisées à ce jour dans la bioconjugaison de biomolécules et plus précisément d'oligonucléotides¹¹¹, certaines pourraient potentiellement être utilisées en tant que troisième ligation.

- **Ligation de Staudinger :**

La réaction de Staudinger se produit entre une phosphine, bien souvent un dérivé de triphenylphosphine et une fonction azoture (schéma 3). Cette réaction réalisable en milieu aqueux à pH neutre présente une forte sélectivité et une accessibilité facile des « building blocks » (commerciaux ou synthèse facile). Elle présente néanmoins quelques désavantages. Le premier est sa faible réactivité, le second est la formation d'un lien encombrant qui pourrait gêner la structuration des G-quadruplexes. Le dernier désavantage de cette réaction est la stabilité limitée des phosphines qui s'oxydent rapidement au contact de l'air.

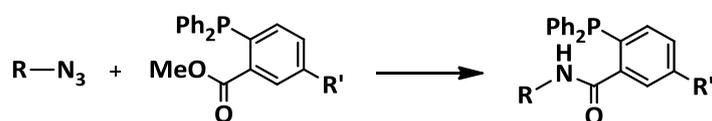


Schéma 3 : ligation de Staudinger

- **Cycloaddition inverse de Diels-alder (IEDDA : Inverse Electron Demand Diels-Alder):**

La réactivité de cette cycloaddition [4+2] entre un alcène et diène dépend beaucoup des partenaires utilisés. En présence d'un norbornène et d'une tétrazine la réaction est très réactive et sélective (schéma 4). De plus, elle peut être réalisée avec de très bons rendements en présence de

tampon phosphate salin. L'équipe d'A. Jäschke a montré en 2010 et 2011 la compatibilité de cette réaction avec les oligonucléotides^{112,113}. Néanmoins, cette ligation présente de gros désavantages avec la formation d'un lien très encombrant et pour la plupart des substrats, une absence de régiosélectivité.

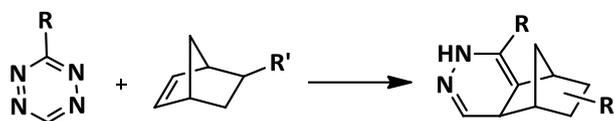


Schéma 4 : cycloaddition IEDDA

- **Réactions organométalliques :**

Depuis peu, des réactions organométalliques de type Suzuki-Miyaura (schéma 5.a), Sonogashira (schéma 5.b) et Stille-Migita (schéma 5.c) sont utilisées pour l'assemblage de conjugués oligonucléotides-sondes fluorescents¹¹⁴. Ces couplages sont catalysés par des dérivés du palladium en milieu aqueux avec des conditions réactionnelles compatibles à la stabilité des oligonucléotides. Par contre ces couplages semblent manquer de réactivité, un des partenaires étant généralement introduit en large excès.

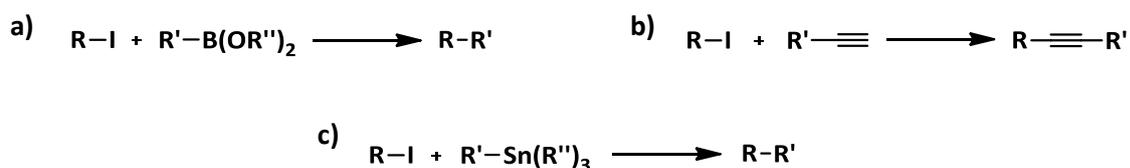


Schéma 5 : réactions organométalliques de type (a) Suzuki-Miyaura, (b) Sonogashira et (c) Stille-Migita

- **Couplage thiol-maléimide :**

La réaction de thiol-maléimide est une addition de Michael entre un groupement donneur de Michael, le thiol et un groupement maléimide comportant une double liaison accepteur de Michael (schéma 6). Très récemment, l'équipe de P.Y. Renard a démontré que ce couplage réalisé dans des conditions douces (tampon phosphate salin, pH 8) était compatible avec la ligation oxime et la CuAAC¹¹⁵. Cette réaction présente une bonne réactivité et une bonne sélectivité. Un autre avantage est la stabilité de la liaison C-S formée. En revanche les dérivés maléimides s'hydrolysent facilement en milieu basique. Une alternative à ce problème d'instabilité est l'utilisation d'une fonction halogénoacétamide à la place de la fonction maléimide.

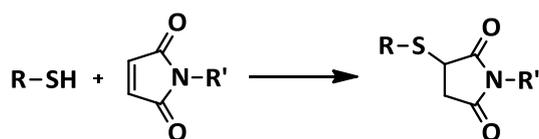


Schéma 6 : couplage thiol-maléimide

- **Ligation thioéther :**

Tout comme le couplage thiol-maléimide, cette réaction présente une forte réactivité ainsi qu'une forte sélectivité et permet de s'affranchir de dérivés maléimides instables. Elle a lieu entre une fonction thiol et une fonction halogénoacétamide pour former une liaison C-S très stable (schéma 7). Un autre avantage de cette ligation est le très faible encombrement stérique du lien thioéther généré. De plus, ce couplage peut être réalisé en milieu aqueux avec des co-solvants organiques compatibles avec les oligonucléotides (acétonitrile, méthanol, DMF). L'orthogonalité de cette réaction avec la chimie des biomolécules ainsi que sa compatibilité avec la ligation oxime et la CuAAC ont été décrites par P.Y. Renard¹¹⁵ en 2013. La même année, l'équipe de G. Gasser a utilisé la ligation thioéther pour la construction d'oligomères PNA multimétalliques¹¹⁶. Un an plus tard, F. Morvan et ses collaborateurs ont également utilisés cette ligation pour la synthèse de conjugués glycocluster-oligonucléotides¹¹⁷.

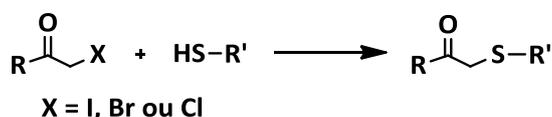


Schéma 7 : ligation thioéther

Parmi toutes ces ligations, pour des raisons de disponibilité des « building blocks », de réactivité, de sélectivité et d'encombrement stérique du lien, nous avons décidé d'utiliser comme troisième ligation la ligation thioéther entre une fonction thiol et une fonction chloroacétamide.

IX. Présentation du projet de thèse

Depuis plusieurs années, en appliquant le concept TASQ, le laboratoire vise à concevoir des mimes de motifs G-quadruplexes et de i-motifs, représentant des cibles biologiques intéressantes. Les premiers travaux ont permis de réaliser l'assemblage et la stabilisation de plusieurs motifs G-quadruplexes de conformation définie *via* l'utilisation d'une ou deux ligations chimiosélectives (ligation oxime et CuAAC). Il s'agit des édifices **A** et **B** mimant les motifs G4 parallèle¹¹⁸ (figure 36.a) et antiparallèle de type basket¹¹⁹ (figure 36.b) de la séquence télomérique. Dans ces deux cas, les G-quadruplexes obtenus sont fortement stabilisés et monomorphes. L'utilisation de ces G-

quadruplexes contraints a permis de déterminer par résonance de plasmon de surface (SPR), l’affinité et la sélectivité de différentes familles de ligands pour ces G-quadruplexes¹²⁰.

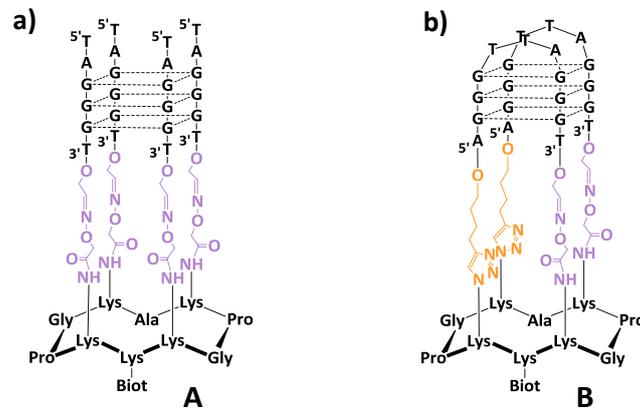


Figure 36 : (a) édifice G4 parallèle et (b) édifice G4 antiparallèle de type basket dérivés de la séquence télomérique

Suite à ces travaux encourageants, l’objectif de cette thèse est **de développer de nouveaux édifices contraints pour l’indentification de ligands sélectifs de G-quadruplexes par SPR et pour le développement d’une nouvelle classe de ligands de G-quadruplexes, les aptamères.**

Ce travail de thèse se divise en deux parties. La première partie est consacrée au développement de nouvelles structures quadruplexes (G4 et i-motif) et triplexes dérivant de la séquence télomérique (édifices **C-F** et **H-J**, figure 37) et de la séquence promotrice (HIV-PRO1) du VIH (édifice **G**, figure 37). L’assemblage des édifices **G** à **J** a nécessité l’utilisation d’une troisième ligation chimiosélective, le couplage thiol-chloroacétamide, orthogonale à la ligation oxime et à la CuAAC. L’optimisation de cette réaction ainsi que la synthèse et la caractérisation (TDS, CD, RMN) de chaque édifice sont présentées dans le chapitre I. Dans ce dernier, est également démontré la structuration et la stabilité de chaque motif.

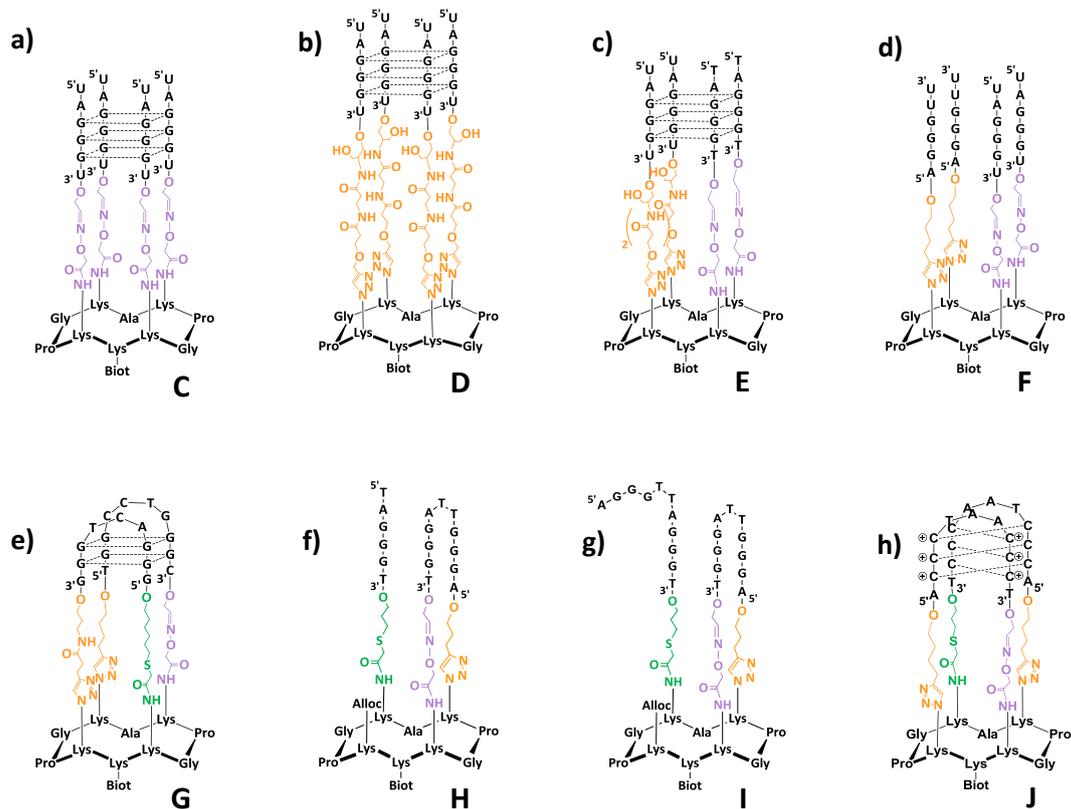


Figure 37 : édifices G-quadruplexes, G-triplexes et i-motif synthétisé durant cette thèse, dérivant de la séquence télomérique : (a) et (b) édifice ARN parallèle, (c) édifice hybride, (d) édifice antiparallèle, (f) édifice triplex, (g) édifice G-quadruplexe, (h) édifice i-motif et de la séquence HIV-PRO1 (e)

L'utilisation de ces nouveaux édifices dans l'identification et le développement de ligands sélectifs des G-quadruplexes est décrite dans la deuxième partie de ce manuscrit. Les édifices **C**, **E**, **G** et **H**, ont été immobilisés sur surface *via* le groupe biotine présent sur la face inférieure des châssis, pour évaluer, par SPR et par BLI, l'affinité de différents ligands de G-quadruplexes pour ces structures. Les constantes d'affinité de ces ligands pour les motifs G-quadruplexes ainsi que leur mode d'interaction sont rapportés dans cette deuxième partie. L'utilisation de l'édifice **B**, pour le développement d'une nouvelle classe de ligands est également décrite dans ce deuxième chapitre. Il s'agit d'aptamères obtenus par une méthode de sélection *in vitro*, la SELEX pour Systematic Evolution of Ligands by EXponential enrichment.

Chapitre 1

Synthèse et caractérisation de mimes de motifs G-
quadruplexes, G-triplexe et de i-motif.

Introduction

Ce premier chapitre est consacré à la description de la synthèse de bioconjugués oligonucléotides-peptide, mimant des motifs particuliers d'ADN et d'ARN adoptés *in-vivo* par les acides nucléiques. Ces motifs, biologiquement intéressants sont les G-quadruplexes, les G-triplexes et les i-motifs. L'assemblage de ces bioconjugués nécessite en premier lieu la synthèse indépendante de châssis peptidiques et d'oligonucléotides fonctionnalisés de façon complémentaire. Le greffage des oligonucléotides sur le châssis est réalisé dans un second temps avec l'utilisation de ligations chimiosélectives. Ce chapitre est composé de quatre parties et décrit la synthèse des édifices **C-J** présentés en introduction. La conception de chaque édifice est présentée en détail, incluant la synthèse des oligonucléotides et des châssis peptidiques ainsi que l'optimisation des ligations chimiosélectives. La structuration des édifices en conformation G-quadruplexe, G-triplexe ou i-motif est également discutée, de même que leur stabilité.

I. Généralités sur la synthèse peptidique

De manière générale, la synthèse des châssis peptidiques est réalisée par synthèse supportée en phase solide (SPPS), selon les principes développés par R. B. Merrifield¹²¹ en 1960 qui lui ont valu le prix Nobel de chimie en 1984. La SPPS repose sur l'ajout successif d'acides aminés par formation de lien peptidique. Elle consiste à effectuer une succession d'étapes de couplage et de déprotection, permettant l'élongation d'une chaîne peptidique fixée de manière covalente à une bille de résine. L'élongation de la chaîne peptidique se fait du C-terminal vers le N-terminal et chaque étape de couplage correspond à l'ajout d'un nouvel acide aminé en présence d'un agent activateur (par exemple HBTU) et de DIEA. Dans la stratégie de synthèse utilisée « Fmoc/tBu », les acides aminés ajoutés à la séquence en élongation possèdent un groupement protecteur Fmoc sur leur amine en α . Cette protection permet de ne coupler qu'un seul acide aminé à la fois. Le groupement Fmoc est ensuite éliminé par β -élimination en présence de 10% de pipéridine dans du DMF, pour permettre l'ajout de l'acide aminé suivant. Cette stratégie de synthèse nécessite également la protection des fonctions réactives présentes sur les chaînes latérales des acides aminés. Les groupements protecteurs utilisés doivent être stables dans les conditions de couplage et de déprotection utilisées pour l'élongation de la chaîne peptidique. La résine qui a été utilisée pour la synthèse de tous les châssis peptidiques décrits dans ce manuscrit est la 2-chlorotriptyl. Cette dernière présente l'avantage de pouvoir libérer le peptide dans des conditions acides douces, 1% de TFA ou 10% d'acide acétique, compatibles avec les groupements protecteurs présents sur les différentes chaînes peptidiques

couleur rouge-orange, permet d'évaluer par simple dosage l'efficacité de chaque cycle de synthèse. La deuxième étape permet d'introduire sur l'oligonucléotide en construction, dont l'hydroxyle en 5' vient d'être déprotégé, un phosphoramidite protégé par un groupement cyanoéthyle. Cette réaction est réalisée en présence de tétrazole, un activateur permettant le couplage et la formation d'un phosphite triester. La troisième étape est le coiffage, plus connue sous le nom de « capping ». A chaque cycle, le couplage n'est jamais total, certains hydroxyles en 5' restent libres. S'ils réagissent lors des prochains cycles de synthèse, des oligonucléotides tronqués de séquences différentes seront formés. Pour éviter cela, un mélange de deux réactifs, « cap A » (Ac₂O/pyridine/THF 10/10/80) et « cap B » (N-méthylimidazole/THF 10/90) est utilisé pour acétyler l'hydroxyle 5'. Les oligonucléotides acétylés seront plus faciles à séparer et donc à éliminer que les oligonucléotides tronqués. Et enfin, la dernière étape permet de générer une liaison phosphodiester stable, par oxydation en présence d'iode et d'eau d'une liaison triester.

II.2. Généralités sur la synthèse d'ARN

Les brins d'ARN sont également synthétisés par SPOS, en utilisant la stratégie PivOM¹²³ à l'échelle d'une micromole. A cause de son groupement hydroxyle en 2' du ribose, l'ARN est très sensible aux conditions basiques utilisées classiquement pour déprotéger les bases et qui permet de libérer le brin du support. En effet, en milieu basique, l'hydroxyle en 2' peut réagir sur le phosphate en 3' induisant alors une coupure de brin. Lors de la synthèse d'ARN il est donc primordial de protéger le groupement hydroxyle en C2', pour éviter sa dégradation lors du traitement ammoniacal final. Les voies de synthèses les plus communes utilisent un groupement protecteur non-basolabile, le tertbutyldiméthyle silyl (TBDMS) (figure 38.a) éliminé à la fin de la synthèse par un traitement avec des ions fluorures. Au laboratoire, une autre stratégie de synthèse est utilisée. Développée par l'équipe de Françoise Debart¹²³, elle est basée sur l'utilisation de phosphoramidites dont l'hydroxyle en C2' du ribose est protégé par le groupement pivaloyloxymethyl (PivOM) basolabile (figure 38.b). Ce dernier est alors éliminé en milieu basique, lors du traitement de déprotection. La stabilité de l'ARN est maintenue grâce à la formation d'un groupement hémiacétal sur l'hydroxyle en C2', qui protège l'ARN de la dégradation durant le traitement. L'hémiacétal est ensuite éliminé lors de l'évaporation par déplacement de l'équilibre vers le formaldéhyde.

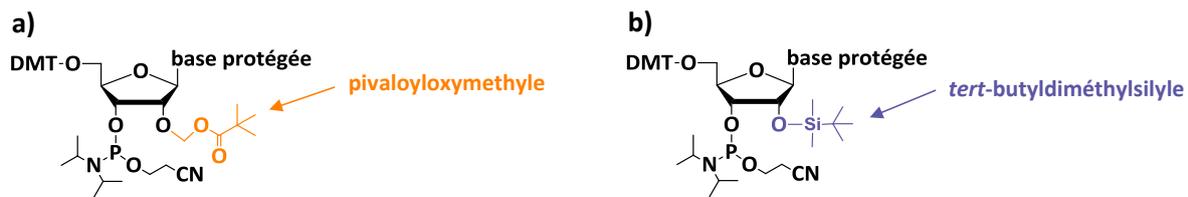


Figure 38 : (a) phosphoramidite PivOM, (b) phosphoramidite TBDMS

Le groupement cyanoéthyle protégeant le phosphate durant la synthèse est normalement éliminé lors du traitement final de déprotection. Dans cette stratégie de synthèse, il est éliminé dans un premier temps par un traitement au DBU, pour éviter que le groupe hydroxyle de l'hémiacétal ne réagisse sur le phosphotriester plus réactif qu'un phosphodiester.

Partie I : Mimes d'ARN G-quadruplexes de la région télomérique

I. Introduction

Cette première partie est consacrée à la description de la synthèse de mimes de G-quadruplexes d'ARN, présents au sein de la région télomérique humaine. Deux mimes parallèles ont été assemblés, l'un mimant le motif G-quadruplexe de la séquence TERRA (figure 39.a), et l'autre le motif G4 hybride ADN/ARN (figure 39.b).

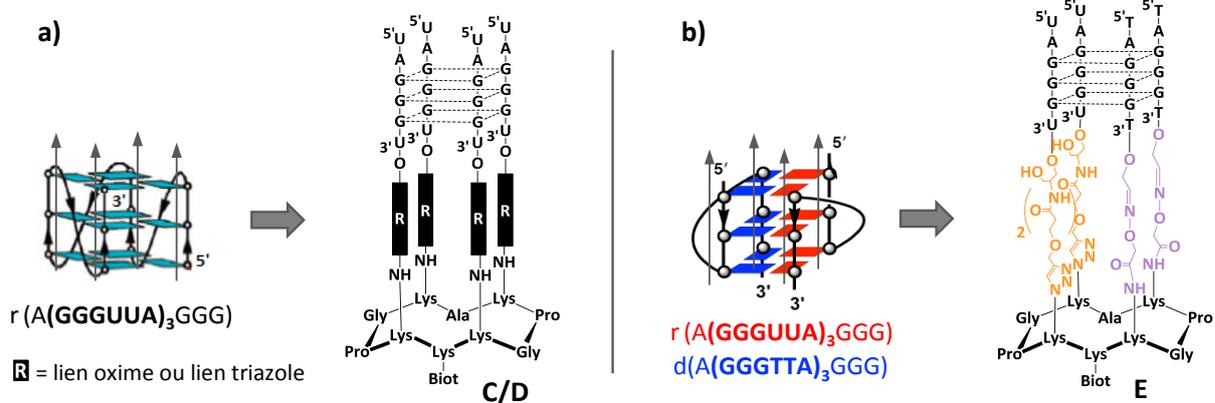


Figure 39 : représentation des édifices **C/D** et **E** mimant : (a) le motif G4 parallèle de l'ARN télomérique et (b) le motif G4 parallèle hybride ARN/ADN de la région télomérique

Comme évoqué en introduction (p 15), les motifs G-quadruplexes d'ARN adoptent préférentiellement une topologie parallèle, contrairement aux G-quadruplexes ADN qui peuvent se structurer dans différentes conformations. Il a néanmoins été découvert un motif G-quadruplexe¹⁹ de conformation antiparallèle dans un aptamère ARN, sélectionné en tant que ligand de la protéine GFP. Afin de mieux comprendre si cette structuration est une exception dans le cas de cet aptamère, ou si elle est également envisageable pour d'autres motifs G-quadruplexes d'ARN, nous avons étudié la possibilité de former un motif G-quadruplexe de conformation antiparallèle dérivant de l'ARN télomérique (figure 40).

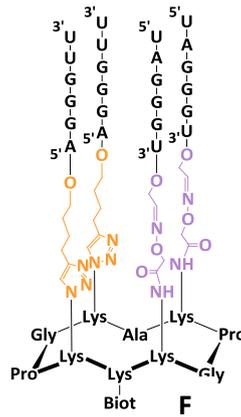


Figure 40 : structure chimique de l'édifice **F** représentant un G4 antiparallèle dérivant de l'ARN télomérique

I.1. Découverte du G-quadruplexe de l'ARN télomérique (séquence TERRA) et intérêt de son ciblage

En 2007, l'équipe de J. Lingner¹²⁴ a découvert dans des fractions nucléaires d'extrait cellulaire l'existence de TERRA, une séquence d'ARN conservée dans plusieurs lignées cellulaires et comportant des motifs répétés r(UUAGGG). Pour confirmer l'hypothèse de séquences provenant de la transcription de la région subtélomérique, il a été démontré la présence d'hybridation entre cette nouvelle séquence d'ARN et une sonde dérivant de la région subtélomérique. Par la suite, ces résultats ont été appuyés par des analyses de localisation cellulaire qui ont révélé la présence de la séquence TERRA au sein de l'hétérochromatine télomérique humaine. Ces résultats ont été renforcés quelques mois plus tard par l'équipe de M. A. Blasco¹²⁵ dont les travaux ont prouvé la formation d'ARN télomérique non codant et de mêmes séquences répétées lors de la transcription de l'ADN télomérique par l'ARN polymérase II. Un an plus tard l'équipe de M. Komiyama¹⁴ a identifié par différentes méthodes d'analyses (RMN, MALDI-TOF MS, dichroïsme circulaire et gel d'électrophorèse) le repliement de l'ARN télomérique de séquence r(UUAGGG)₄, sous forme d'un G-quadruplexe stable ($T_m = 51^\circ\text{C}$) de conformation parallèle.

L'équipe de A. T. Phan¹⁵ a ensuite mis en évidence par des études structurales (RMN et de dichroïsme circulaire) le repliement de l'ARN télomérique en configuration parallèle, mais cette fois-ci en présence de potassium. Ses travaux ont également révélé qu'en fonction de la séquence utilisée plusieurs motifs de G-quadruplexes peuvent former des structures d'ordre supérieur. Par exemple, la séquence r(GGGUUAGGGU) favorise le « π -stacking » de deux motifs G-quadruplexes tandis que la séquence r(UAGGGUUAGGGU) comportant deux nucléotides de plus le défavorise. Cette étude démontre d'une part l'existence des motifs d'ARN G-quadruplexes au sein de la région télomérique et d'autre part, qu'en présence de cations métalliques et en fonction de la séquence utilisée des structures formées par l'empilement de plusieurs G-quadruplexes peuvent exister.

Depuis ces premières études, les travaux visant à mieux comprendre le rôle de TERRA dans la région télomérique et sa potentielle implication dans la régulation de la télomérase ne cessent de croître. En 2010¹²⁶ et 2013¹²⁷, l'équipe de J. Lingner a identifié TERRA comme un ligand de la télomérase et un inhibiteur direct de la télomérase humaine. Pour comprendre l'interaction entre TERRA et la télomérase, l'équipe de P. Chartrand¹²⁸ a lié une sonde fluorescente sur une séquence d'ARN télomérique afin de suivre par microscopie son évolution dans la cellule. Il a été découvert que la production de TERRA est déclenchée lorsque le télomère portant son gène est tronqué. Il a ensuite été observé que les molécules de TERRA s'accumulent en un point précis et au même moment, recrutent la télomérase qui est ensuite acheminée vers le télomère tronqué ayant déclenché la production de TERRA. Plus récemment, les travaux de H. Seimiya¹²⁹ ont révélé une corrélation directe entre la régulation de l'expression de gènes immunitaires innés et la présence de motifs G4 au sein de TERRA. Il a été démontré, que l'allongement des télomères régule la transcription de TERRA et l'expression de gènes immunitaires innés. En d'autres termes leurs travaux *in-vivo* sur des cellules cancéreuses humaines ont démontré d'une part que l'élongation des télomères provoque l'augmentation de la transcription de TERRA. Et d'autre part que les motifs G4 présents au sein de TERRA suppriment l'expression de certains gènes impliqués dans le système immunitaire inné.

Bien que le rôle exact de TERRA ne soit pas encore clairement défini, ces travaux et d'autres soulignent l'implication de TERRA dans la régulation de la longueur des télomères et dans la régulation de la télomérase. Bien qu'elle n'ait pas toujours été directement démontrée, il n'est pas faux de penser qu'il pourrait exister une corrélation directe entre l'implication de TERRA dans ces deux phénomènes et la présence de motifs G-quadruplexes. TERRA pourrait ainsi comme la séquence de l'ADN télomérique devenir un nouveau candidat dans la recherche d'un traitement contre le cancer. Par conséquent, créer un modèle de ce G-quadruplexe dans le but de trouver des ligands spécifiques à cette structure peut s'avérer pertinent.

I.2. Découverte du G-quadruplexe hybride ADN/ARN de la séquence télomérique et intérêt de son ciblage

Certaines études^{130,131} cherchant à comprendre le rôle de TERRA ont mis en évidence son association avec l'ADN télomérique pour protéger l'extrémité des télomères. Plusieurs équipes ont révélé l'existence d'un motif G-quadruplexe hybride ARN/ADN de configuration parallèle. De 2012 à 2014 avec l'utilisation d'oligonucléotides modèles d'ADN et d'ARN télomérique, l'équipe de M. Komiyama^{54,132} a mis en évidence la formation de motifs G-quadruplexes hybrides dans les cellules. Dans ces études, l'équipe a également montré les effets du G-quadruplexe hybride sur l'inhibition de la sénescence cellulaire et de la fusion finale des chromosomes, contribuant donc à la protection des

télomères. L'équipe de Z. Than a également démontré par des études *in-vitro*, en utilisant notamment le FRET, la formation de G-quadruplexes hybride ADN/ARN dérivant de la séquence télomérique^{133,134}.

Ces travaux effectués *in-vitro* sont renforcés par ceux de Z. Tan¹³⁵ qui a prouvé l'existence de G4 télomérique hybride *in-vivo*, au sein de la bactérie *Escherichia Coli*. Dans cette étude, il a également démontré que lors de la transcription, la formation du G4 hybride ARN/ADN prédomine sur celle du G4 ADN. Cette observation pourrait suggérer que la compétition structurelle et la conversion entre les G-quadruplexes télomériques ADN et hybride ADN/ARN pourrait jouer un rôle dans la régulation de la transcription et par conséquent, devenir une cible potentielle pour des applications thérapeutiques. Mimer ce G-quadruplexe hybride afin de trouver des ligands spécifiques à cette nouvelle cible devient également pertinent.

II. Synthèse et caractérisation d'édifices mimant les G-quadruplexes de l'ARN télomérique humain

L'assemblage des trois édifices, ARN parallèle **C/D**, hybride parallèle **E** et ARN antiparallèle **F** a nécessité la synthèse des châssis peptidiques et d'oligonucléotides fonctionnalisés. D'une manière générale, les quatre lysines supérieures des châssis comportent les fonctions complémentaires à celles présentes à l'extrémité des oligonucléotides, permettant ainsi leur assemblage par ligations chimiosélectives.

II.1. Synthèse et caractérisation des mimes du motif G-quadruplexe parallèle

Le motif G-quadruplexe de l'ARN télomérique se structure en topologie parallèle, caractérisée par une orientation identique des quatre brins d'ARN. Afin de mimer précisément les différents constituants de cette structure (brins riches en guanines et boucles) (figure 41.a), quatre ligations orthogonales successives devraient être utilisés. Or, à ce stade de nos travaux, seuls deux réactions chimiosélectives sont disponibles au laboratoire. Nous avons donc décidé d'assembler un modèle simplifié sans boucle. Ceci équivaut à greffer quatre brins d'ARN analogues, de séquence 5'UAGGGU3' dérivant de l'ARN TERRA, simultanément sur le châssis peptidique *via* une seule ligation chimiosélective (figure 41).

Pour rappel, les oligonucléotides sont reliés au châssis peptidique par des liens espaceurs qui permettent au système de conserver une certaine flexibilité. Dans le but d'évaluer l'effet de ces liens sur la stabilité du motif G4, nous avons assemblé deux édifices possédant des espaceurs de longueur

variables. Pour des raisons de disponibilité des « building blocks » et de facilité de synthèse, les édifices ont été construits par des ligations chimiosélectives différentes, la ligation oxime et la CuAAC. L'édifice **C** (figure 41.b) comporte des liens éthers d'oximes deux fois plus courts que les liens triazoles de l'édifice **D** (figure 41.c).

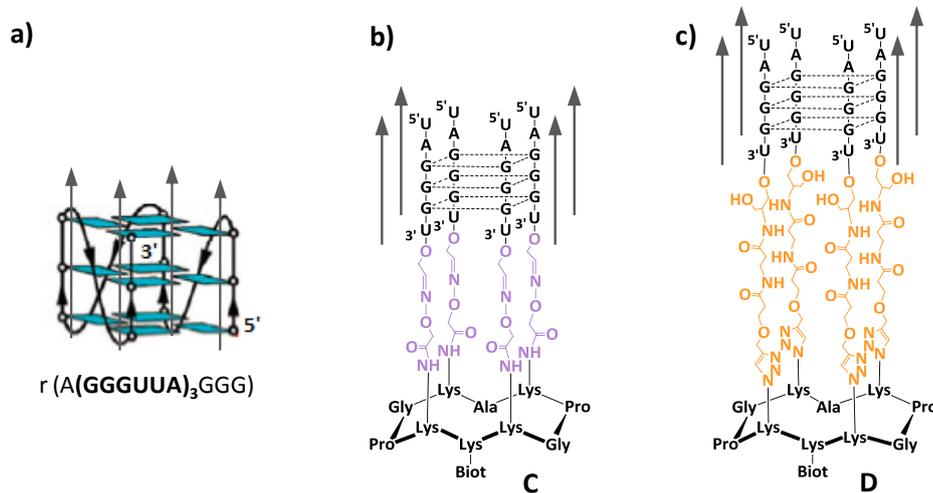


Figure 41 : (a) motif G-quadruplexe ARN parallèle de la région télomérique, structure chimique des édifices **C** et **D** mimant respectivement le motif G4 parallèle de l'ARN télomérique avec (a) la stratégie oxime et (b) la stratégie CuAAC

II.1.a. Stratégie oxime

Cette stratégie de synthèse consiste à greffer les oligonucléotides sur le châssis peptidique en utilisant la ligation oxime. Cette réaction est réalisée entre une fonction oxyamine et une fonction aldéhyde. Dans le cas de l'édifice **C**, les fonctions oxyamines sont portées par les chaînes latérales des lysines et les fonctions aldéhydes complémentaires sont présentes sur les brins d'ARN.

i) Synthèse du châssis peptidique

Le châssis peptidique **1** (figure 42) comporte une fonction biotine sur sa lysine inférieure, essentielle pour les études d'interactions par SPR et quatre fonctions oxyamines sur ses lysines supérieures, nécessaires à l'immobilisation des oligonucléotides. Deux stratégies ont été utilisées pour l'introduction de ces fonctions : l'ajout d'acides-aminés modifiés (« building-blocks ») lors de la SPPS, pour la fonction biotine et la post-fonctionnalisation des chaînes latérales des lysines, en utilisant des esters actifs, pour les fonctions oxyamines.

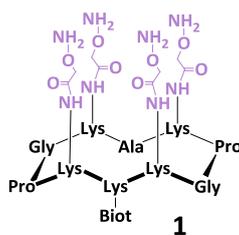


Figure 42 : structure chimique du châssis peptidique 1

- Synthèse et insertion de la fonction biotine

Durant la synthèse supportée, la fonction biotine est introduite sur le châssis peptidique sous forme d'un « building block », la Fmoc-Lys(biot)-OH. Cette dernière est formée d'une lysine comportant sur sa fonction amine en ϵ une biotine, introduite *via* un lien amide. La synthèse de cet acide-aminé modifié (schéma 9) nécessite au préalable l'activation de la fonction acide carboxylique de la biotine par un groupement N-hydroxysuccinimide, en présence de DCC comme agent de couplage. Cette réaction est réalisée dans du DMF durant 12h. La biotine activée **2a** est ensuite introduite sur une Fmoc-Lys-OH, en présence de DMF et d'une base, la DIEA (pH 8-9). Après 2h de réaction la lysine fonctionnalisée est obtenue par précipitation avec un rendement global de 80% et peut-être utilisé en SPPS après contrôle de sa pureté par RMN.

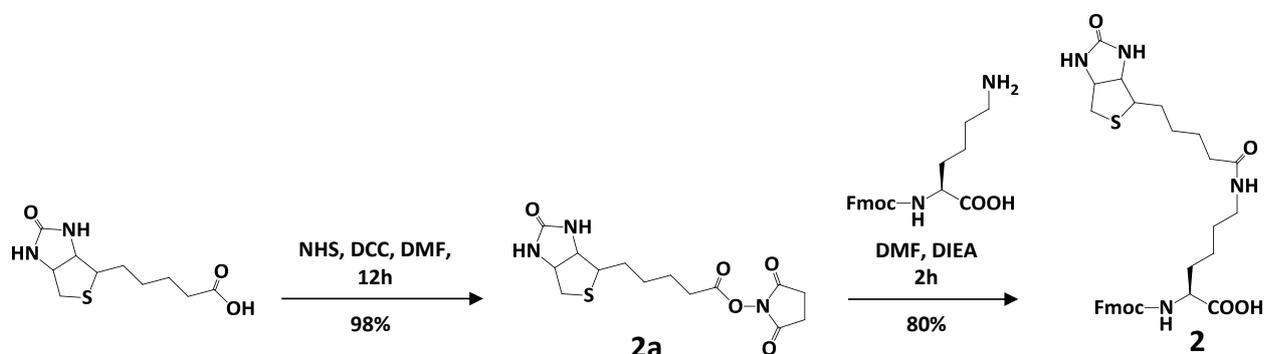


Schéma 9 : synthèse de la Fmoc-Lys(Biot)-OH 2

- Description de la synthèse du châssis peptidique

Le châssis peptidique est synthétisé sur support solide en utilisant la stratégie de synthèse au Fmoc/tBu et la résine 2-chlorotriptyl.

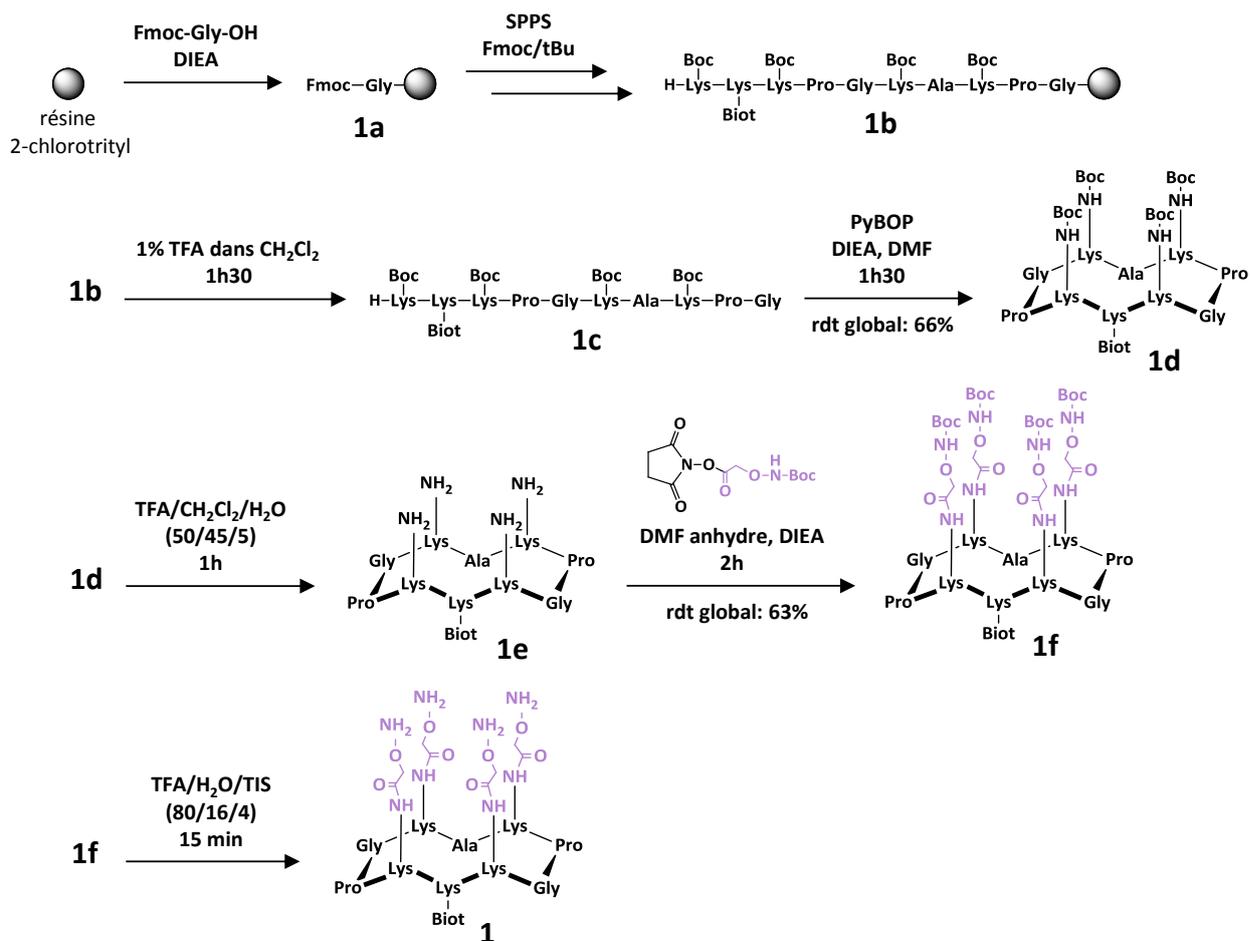


Schéma 10 : synthèse du châssis peptidique **1**

L'ajout du premier acide aminé glycine (schéma 10) est réalisé manuellement en présence de DIEA. L'élongation de la chaîne peptidique est ensuite opérée en utilisant un synthétiseur de peptide. En fin d'élongation, le peptide linéaire est détaché de la résine en milieu acide par une solution de TFA à 1% dans du DCM. La cyclisation du peptide linéaire est ensuite effectuée en milieu basique (pH 8-9 avec DIEA) et en présence de PyBOP comme agent de couplage. Cette réaction est réalisée en milieu dilué, à 10^{-3} M, afin de favoriser la cyclisation intramoléculaire du châssis peptidique et éviter tout risque de polymérisation du peptide linéaire. Il est également à noter que la cyclisation s'effectue sur une glycine en C-terminal. De par son absence de chaîne latérale, la glycine permet d'éviter tout risque de racémisation. Le châssis peptidique intermédiaire **1d** est obtenu après purification RP-HPLC avec un rendement global de 66%. L'étape suivante consiste à déprotéger la fonction amine en ϵ des lysines. L'élimination des groupements protecteurs Boc est réalisée en milieu acide fort, en présence d'une solution TFA/CH₂Cl₂/H₂O (50/45/5). L'eau est utilisée pour piéger les carbocations libérés lors de la déprotection. Après précipitation dans l'éther, l'intermédiaire déprotégé **1e** est directement engagé pour greffer sur les fonctions amines en ϵ des lysines, *via* un lien amide, l'acide Boc-aminoxyacétique activé sur sa fonction acide carboxylique par un

groupement N-hydroxysuccinimide. Cette étape de post-fonctionnalisation qui permet d'introduire les quatre fonctions oxyamines protégées a lieu en présence de DMF anhydre et de DIEA, à pH 8-9 pendant 2h. Il est important de noter qu'un temps de réaction trop long entraîne la formation d'un produit secondaire provenant du couplage de l'acide activé sur la fonction oxyamine protégée. Bien qu'elle soit protégée par un groupement *tert*-butoxycarbonyle (Boc) la fonction oxyamine conserve une certaine réactivité qui peut conduire à des acylations secondaires. Le châssis protégé **1f** est obtenu après purification avec un rendement global de 63%. La dernière étape permet de déprotéger les fonctions oxyamines en éliminant le groupement protecteur Boc. Cette déprotection a lieu en présence d'une solution d'acide fort, TFA/H₂O/TIS (80/16/4). A cette étape, l'eau ne suffira pas à piéger les carbocations libérés, il est donc nécessaire d'utiliser un piègeur de carbocations (triisopropylsilane). Le châssis final fonctionnalisé **1** est obtenu après précipitation dans l'éther et est caractérisé par LC-MS (figure 43).

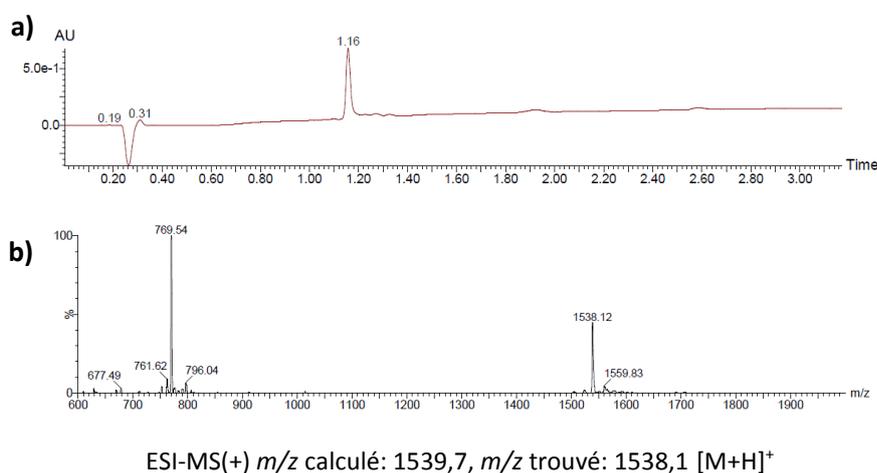


Figure 43 : (a) chromatogramme UPLC et (b) spectre ESI-MS du châssis peptidique **1**

ii) Synthèse des brins d'ARN

Afin de mimer le motif G-quadruplexe TERRA, de séquence r(A(GGGUUA)₃GGG), les brins d'ARN ont été synthétisés avec la séquence suivante : 5'UAGGGUX^{3'} où X représente la fonction aldéhyde, nécessaire à l'immobilisation des oligonucléotides sur le châssis. La séquence fonctionnalisée (schéma 11) est obtenue en utilisant une résine 3'glyceryl CPG, qui permet l'obtention d'une fonction 1,2 diol qui sera ensuite oxydée en aldéhyde¹³⁶. La libération et la déprotection des oligonucléotides sont réalisées par un traitement ammoniacal à 28% à température ambiante. Il est à noter qu'avant cette étape de déprotection un pré-traitement de l'oligonucléotide en présence de DBU est effectué. Cette dernière est une base organique, non nucléophile qui élimine par β -élimination la protection cyanoéthyle des phosphates, sans déprotéger les bases azotées. Le groupement phosphotriester est alors transformé en groupement phosphodiester, beaucoup moins

réactif. L'objectif est d'empêcher une éventuelle réaction de cyclisation de la fonction diol du support glyceryl sur le phosphotriester, ce qui pourrait entraîner la formation d'un isomère 1,3 diol¹³⁷. Une troisième étape de déprotection est nécessaire avant l'obtention de l'ARN **3a**. Il s'agit d'un traitement avec 80% d'acide acétique permettant d'éliminer le groupement diméthoxytrityl présent sur l'hydroxyle en 5' du dernier nucléotide incorporé. Après purification par chromatographie échangeuse d'ions l'oligonucléotide **3b** est obtenu avec un rendement de 32%. La fonction 3'aldehyde, est ensuite générée par coupure oxydative de la fonction diol, en présence de périodate de sodium. Le sel de périodate et le formaldéhyde formés lors de la réaction sont éliminés sur colonne d'exclusion de taille. La réaction est considérée comme quantitative et l'oligonucléotide **3** obtenu est caractérisé par RP-HPLC et par spectrométrie de masse (figure 44).

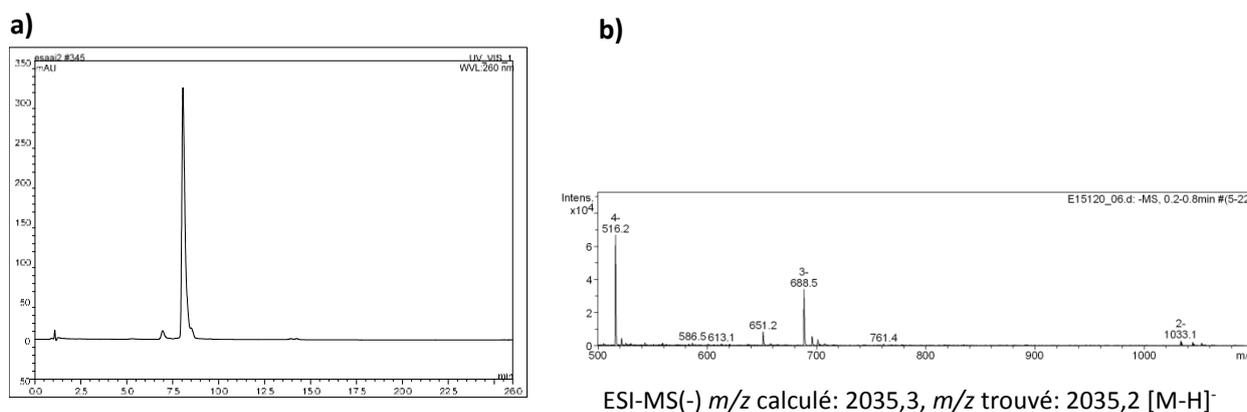
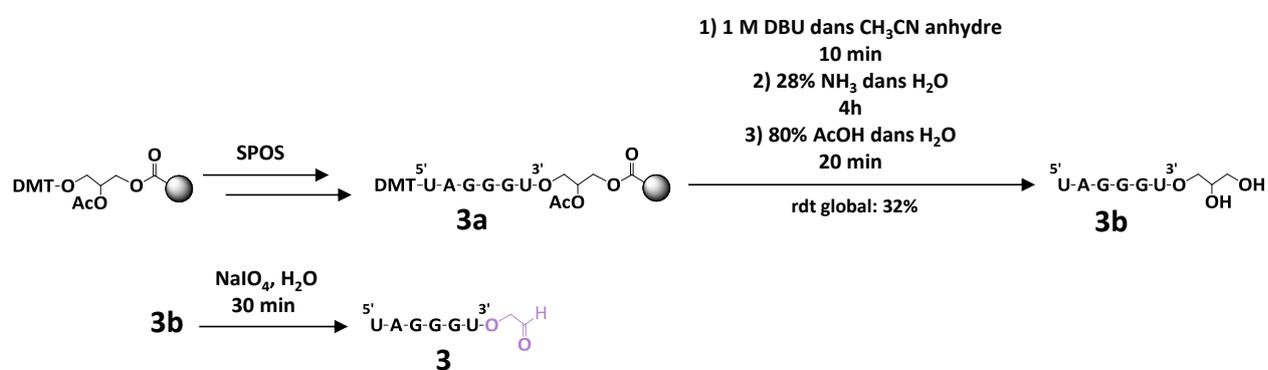


Figure 44 : (a) chromatogramme AELC et (b) spectre ESI-MS de l'ARN 3'aldehyde **3**

iii) Assemblage de l'édifice

La formation des liens éther d'oxime entre l'oligonucléotide **3** et le châssis peptidique **1** (schéma 12) est réalisée en présence d'une solution tamponnée d'acétate d'ammonium (0,4 M à pH 4,5), pendant 45 min à 55°C. 1,2 eq d'oligonucléotides 3'aldehyde **3** sont ajoutés par fonction oxyamine. Le brut réactionnel est purifié par chromatographie échangeuse d'ions et les sels résiduels

de purification sont éliminés sur colonne d'exclusion de taille. L'édifice final **C** est obtenu avec un rendement de 48%.

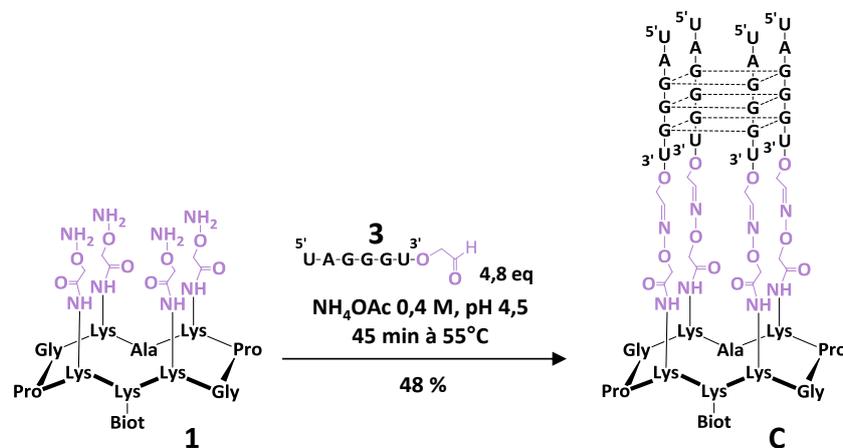


Schéma 12 : synthèse de l'édifice ARN parallèle **C**

L'utilisation de l'HPLC en phase inverse n'est pas appropriée à la purification de cet édifice, ainsi que de tous les systèmes d'ARN G-quadruplexes. Ces conditions de purification ne permettent pas une bonne séparation des différents produits contenus dans le brut réactionnel. Dans ce dernier, sont présent l'édifice final **C** mais également l'oligonucléotide **3** de départ ainsi qu'une faible quantité d'édifice n-1, c'est-à-dire un édifice comportant trois brins d'ARN au lieu de quatre. Les motifs G-quadruplexes d'ARN étant fortement stabilisés, ces produits peuvent alors se structurer ensemble pour former des motifs G-quadruplexes intermoléculaires stables, visibles en RP-HPLC sous forme d'un amas de pics* (figure 45.a). Afin d'améliorer la séparation des produits, nous utilisons une colonne échangeuse d'ions (AELC), chauffée à 75°C avec une phase mobile de type perchlorate de lithium à pH 7. Ces conditions de purification possèdent un fort pouvoir dénaturant et nous permettent d'obtenir des chromatogrammes parfaitement définis, rendant ainsi possible l'isolation de l'édifice **C** (figure 45.b), qui a été caractérisé par AELC et spectrométrie de masse (figure 46).

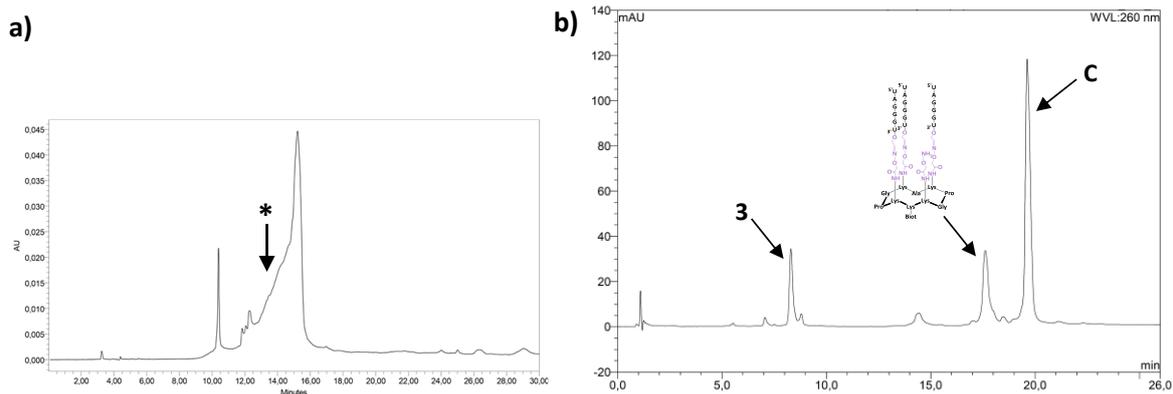


Figure 45 : (a) chromatogrammes RP-HPLC et (b) AELC de l'édifice brut C

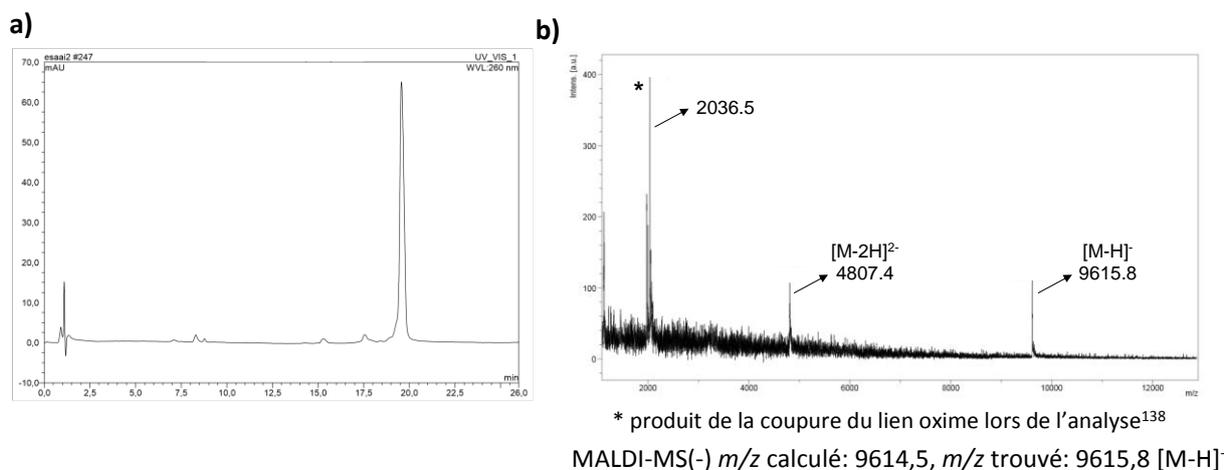


Figure 46 : (a) chromatogramme AELC et (b) spectre MALDI-ToF de l'édifice ARN G-quadruplexe parallèle C

II.1.b. Stratégie CuAAC

Dans la stratégie CuAAC, les oligonucléotides sont liés au châssis en utilisant la cycloaddition (1,3) dipolaire catalysée au cuivre (CuAAC), entre une fonction azoture et une fonction alcyne. Dans le cas de l'édifice **D** les fonctions azotures sont portées par les chaînes latérales des lysines et les fonctions alcynes complémentaires sont présentes sur les brins d'ARN.

i) Synthèse du châssis peptidique **4**

Dans cette stratégie de synthèse, le châssis peptidique (figure 47) présente quatre fonctions azotures qui sont introduites sous forme de « building-block » durant la synthèse supportée.

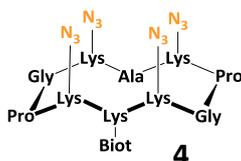


Figure 47 : structure chimique du châssis peptidique **4**

- Synthèse et insertion de la fonction azoture

La fonction azoture est introduite sur le châssis peptidique durant la synthèse supportée, sous forme d'un « building block », la Fmoc-azidonorleucine. Cet acide aminé modifié est parfaitement compatible avec les différentes étapes de couplage et de déprotection opérées en SPPS. Il est synthétisé à partir de la Fmoc-Lys-OH en réalisant une réaction de diazo-transfert¹³⁹ sur l'amine en ϵ (schéma 13). Cette réaction a lieu en milieu basique, en présence d'un réactif de diazo-transfert **5**, l'ISAHC et de sulfate de cuivre. Après une nuit à température ambiante, une extraction acide permet d'obtenir la Fmoc-azidonorleucine **6** avec un rendement de 98%. Après contrôle de sa pureté par RMN, l'acide-aminé modifié peut être utilisé en SPPS.

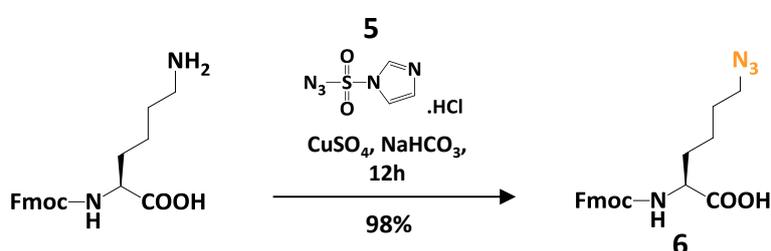


Schéma 13 : synthèse de la Fmoc-azidonorleucine **6**

- Description de la synthèse du châssis peptidique

Le châssis peptidique est synthétisé sur support solide en utilisant la stratégie de synthèse au Fmoc/tBu et la résine 2-chlorotrityl. La voie de synthèse de ce châssis peptidique étant similaire à celle du châssis **1** présentée p 53, elle ne sera pas décrite (schéma 14). Le châssis peptidique **4** est obtenu après purification HPLC avec un rendement global de 66% et a été caractérisé par LC-MS (figure 48).

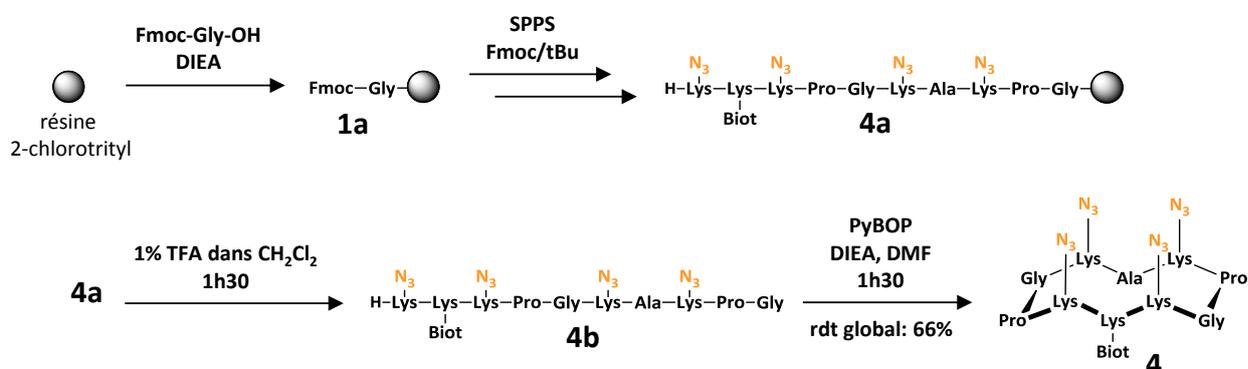
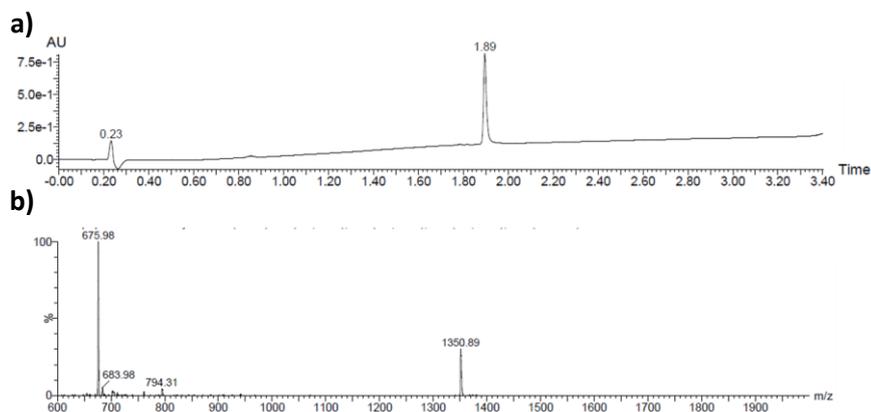


Schéma 14 : synthèse du châssis peptidique **4**



ESI-MS(+) m/z calculé: 1351,6, m/z trouvé: 1350,9 [M+H]⁺

Figure 48 : (a) chromatogramme UPLC et (b) spectre ESI-MS du châssis peptidique **4**

ii) Synthèse des brins d'ARN

Dans le cas de la stratégie CuAAC, la séquence d'ARN (5'UAGGGUY^{3'}, Y = 3'alcyne) comporte une fonction alcyne en son extrémité 3'. Cette fonction est obtenue en utilisant une résine 3'Alkyne-Modifier Serinol CPG. La voie de synthèse de cet oligonucléotide étant similaire à celle de l'ARN **3** présentée p 52, elle ne sera pas décrite (schéma 15).

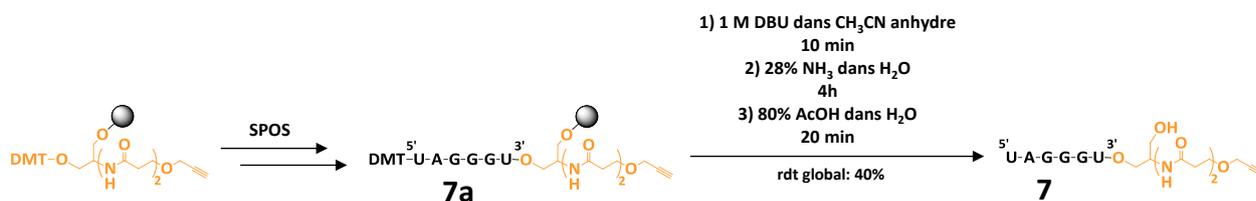
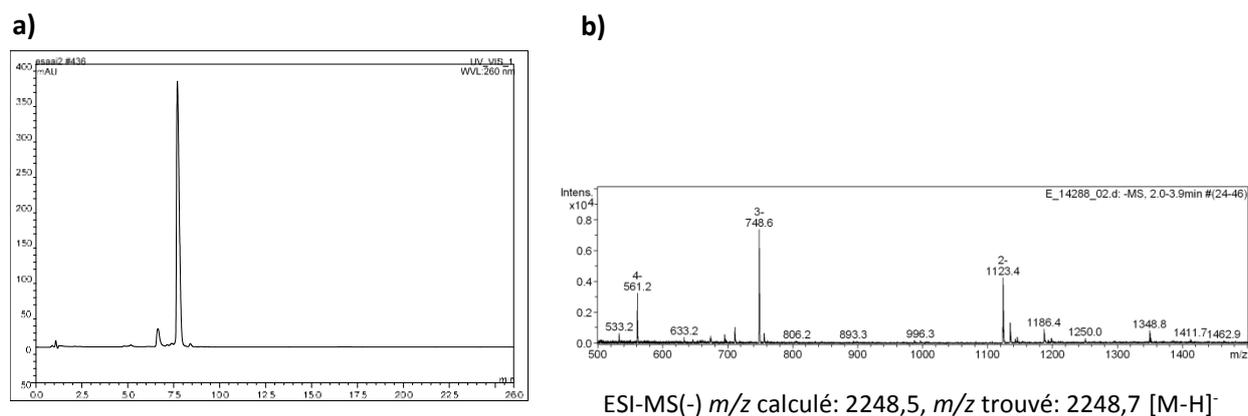


Schéma 15 : synthèse de l'ARN 3'alcyne **7**



ESI-MS(-) m/z calculé: 2248,5, m/z trouvé: 2248,7 [M-H]⁻

Figure 49 : (a) chromatogramme AELC et (b) spectre ESI-MS de l'ARN 3'-alcyne **7**

iii) Assemblage de l'édifice

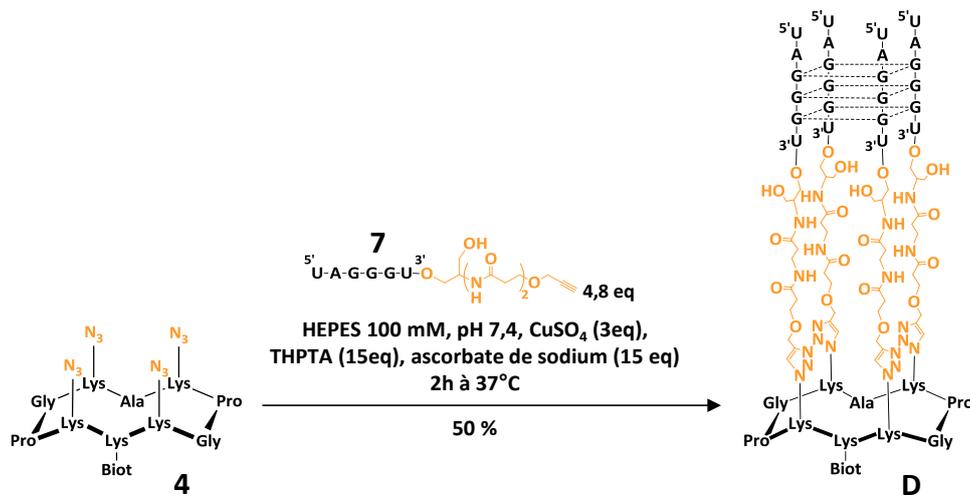
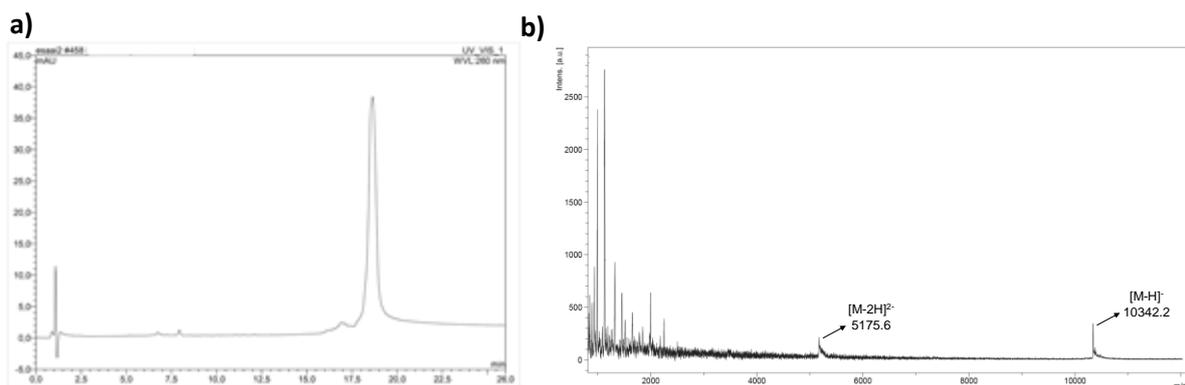


Schéma 16 : synthèse de l'édifice ARN G-quadruplexe parallèle **D**

L'assemblage de l'édifice est réalisé dans un tampon HEPES (100 mM, pH 7,4), en présence de sulfate de cuivre (3 eq/alcyne), de THPTA (15 eq/alcyne) et d'ascorbate de sodium (15 eq/alcyne) durant 2h à 37°C. L'ascorbate de sodium est utilisé pour réduire la forme Cu^{II} du sulfate de cuivre en Cu^{I} , nécessaire à la réaction. Le Cu^{II} en solution est connu pour induire des coupures oxydantes sur les oligonucléotides¹⁴⁰, afin d'éviter cela, la réaction est réalisée en présence de THPTA, un ligand du Cu^{II} . Il est à noter qu'en fin de réaction, le cuivre est piégé par un excès d'EDTA (50 eq par azoture) puis éliminé sur colonne d'exclusion stérique. Après purification sur colonne échangeuse d'ions l'édifice final **D** est obtenu avec un rendement de 50% et est caractérisé par AELC et spectrométrie de masse (figure 50).



MALDI-MS(-) m/z calculé: 10342,2, m/z trouvé: 10342,2 [M-H]⁻

Figure 50 : (a) chromatogramme AELC et (b) spectre MALDI-ToF de l'édifice ARN parallèle **D**

II.2. Synthèse et caractérisation du mime du G-quadruplexe hybride ADN/ARN E

Le motif G-quadruplexe hybride ADN/ARN de la région télomérique présente une topologie parallèle. Le mimer en ne représentant ni les boucles supérieures, ni les boucles inférieures (figure 51.a) équivaut à greffer deux brins d'ARN identiques dérivés de la séquence TERRA et deux brins d'ADN identiques dérivés de la séquence d'ADN télomérique (figure 51.b). Pour greffer indépendamment les brins d'ARN et les brins d'ADN, deux ligations chimiosélectives sont nécessaires. Les deux ligations utilisées sont la ligation oxime et la CuAAC. Comme évoquée en introduction, de précédentes études¹¹⁹ ont montré la compatibilité de ces deux réactions.

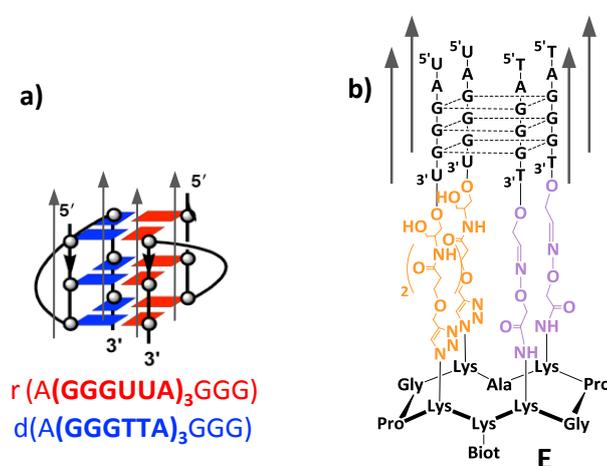


Figure 51: (a) motif G-quadruplexe hybride ADN/ARN de la région télomérique et (b) structure chimique de l'édifice E mimant le motif G4 hybride ADN/ARN

Pour des raisons de simplicité de purification, nous avons choisi de greffer en premier lieu les brins d'ADN. A cause de leur grande réactivité, il est préférable de faire réagir les fonctions oxyamines juste après leur déprotection, afin d'éviter la formation d'adduits covalents non désirés. La ligation oxime doit de ce fait être effectuée avant la CuAAC. Par conséquent les brins d'ADN ont été liés au châssis par la ligation oxime et les brins d'ARN *via* la CuAAC.

II.2.a. Synthèse du châssis peptidique 8

Le châssis peptidique **8**, présentant deux fonctions oxyamines et deux fonctions azotures est préparé selon la même procédure qu'auparavant (voir synthèse châssis **1** et **4** p 49 et 55) (schéma 17). Les fonctions azotures sont introduites par l'utilisation du « building-block » Fmoc-azidonorleucine **6** et les fonctions oxyamines par une post-fonctionnalisation des amines en ϵ des lysines avec l'acide Boc-aminoxyacétique activé. La plateforme peptidique **8** est caractérisée par UPLC-MS (figure 52).

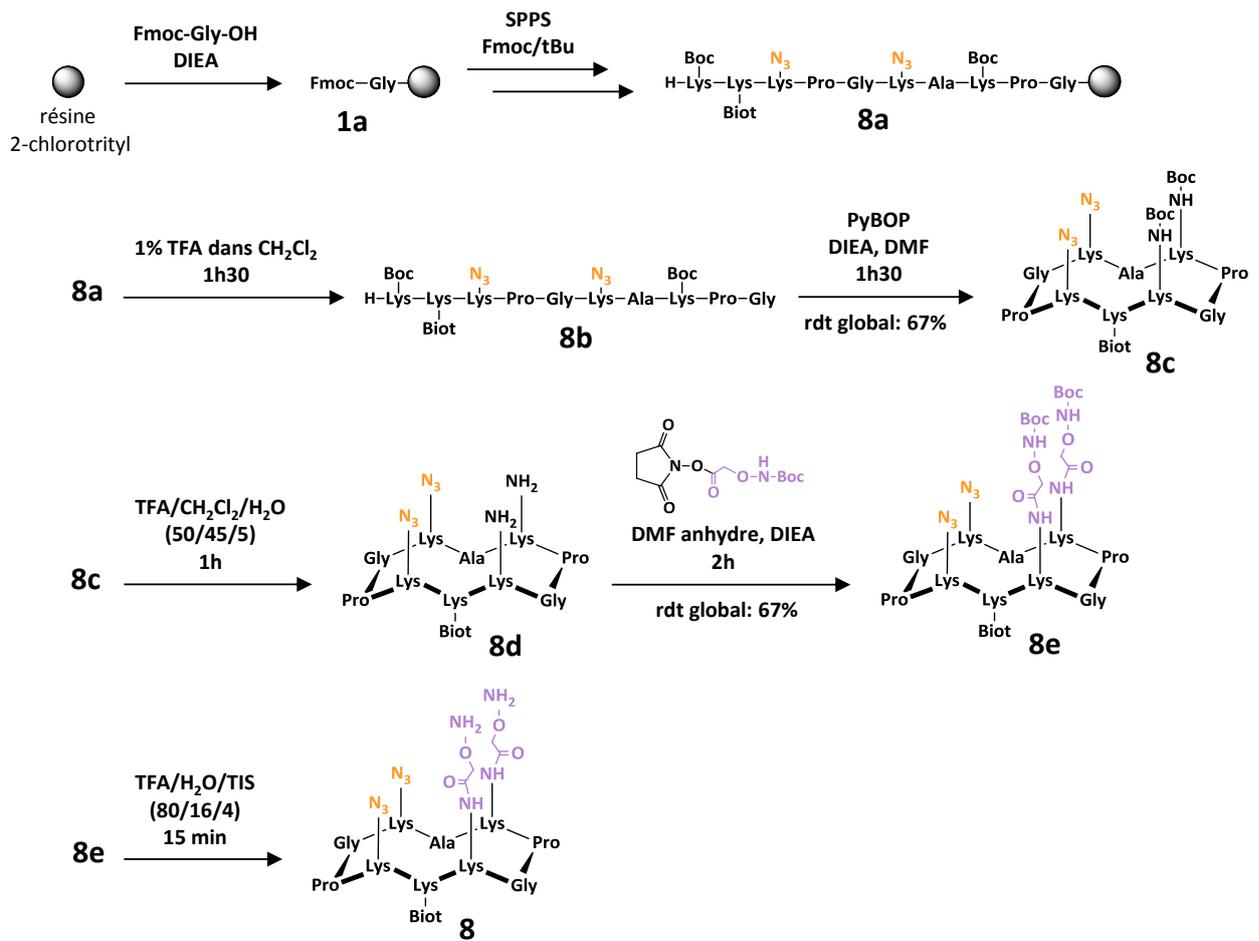
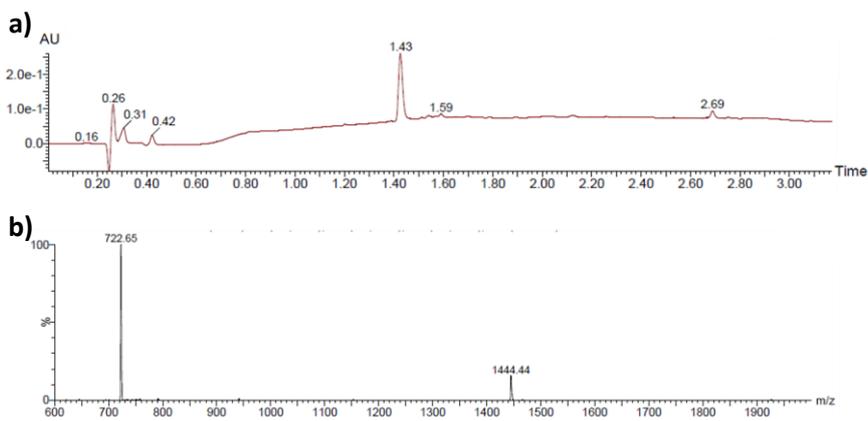


Schéma 17 : synthèse du châssis peptidique



ESI-MS(+) m/z calculé: 1445,7, m/z trouvé: 1444,4 $[\text{M}+\text{H}]^+$

Figure 52 : (a) chromatogramme UPLC et (b) spectre ESI-MS du châssis peptidique **8**

II.2.b. Synthèse des brins d'ADN et d'ARN

i) Synthèse des brins d'ADN

La séquence d'ADN utilisée pour la fabrication de l'édifice hybride ADN/ARN est la suivante : 5'TAGGGTX^{3'}, X = 3'aldehyde. Après synthèse supportée sur la résine 3' glyceryl CPG, les oligonucléotides ont subi un premier traitement basique en présence de DBU (voir p 52) avant d'être libéré du support par une solution d'ammoniaque et de méthylamine (50/50) à 65°C. Ces conditions de déprotection diffèrent de celle utilisées en présence d'ARN, car contrairement à ces derniers, les brins d'ADN sont plus stables en milieu basique. L'ADN 3'diol **9c** est obtenu après purification par RP-HPLC avec un rendement global de 45%. Le groupement DMT conservé après la SPOS sur le dernier nucléotide facilite la séparation de l'oligonucléotide voulu des oligonucléotides tronqués (produits secondaires de la synthèse) lors de la purification. Ce groupement est ensuite éliminé avec 80% d'acide acétique. L'ADN 3'aldehyde **9** est obtenu après coupure oxydante de l'ADN **9c** par du périodate de sodium.

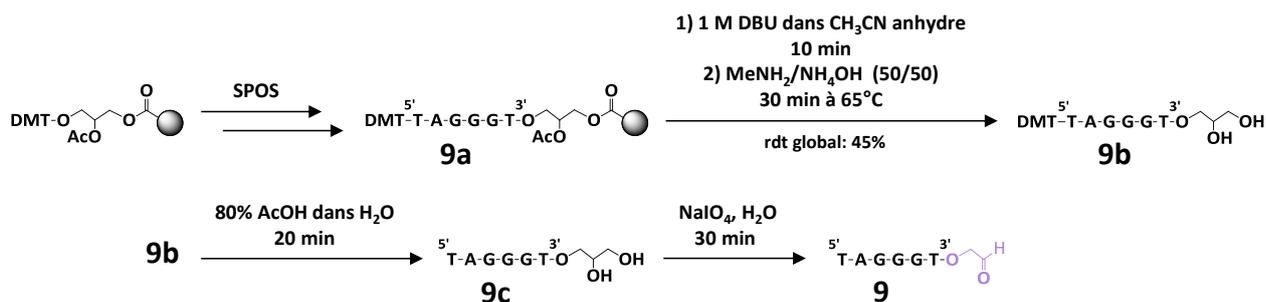


Schéma 18 : synthèse de l'ADN 3'aldehyde **9**

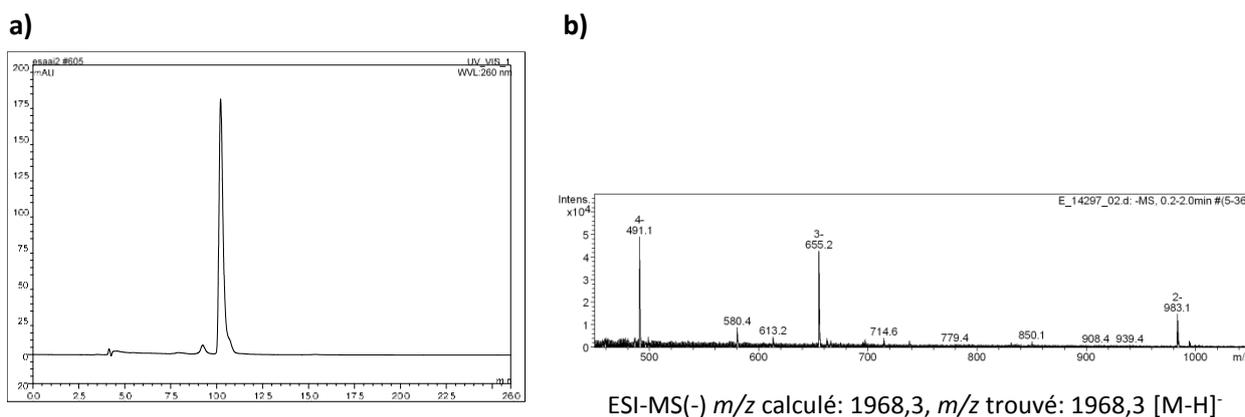


Figure 53 : (a) chromatogramme AELC et (b) spectre ESI-MS de l'ADN 3'aldehyde **9**

ii) Synthèse des brins d'ARN 3'alcyne

L'oligonucléotide utilisé pour l'assemblage de l'édifice hybride ADN/ARN est l'ARN 3'alcyne **7** qui a servi à la synthèse de l'édifice **D**. Sa synthèse est décrite dans la partie II.1.2.b, p 57.

II.2.c. Assemblage de l'édifice

L'assemblage de l'édifice hybride ADN/ARN est réalisé en deux étapes (schéma 19). La première étape consiste à greffer les brins d'ADN 3'aldéhyde **9** sur le châssis et la deuxième étape à lier les brins d'ARN 3'alcyne **7**.

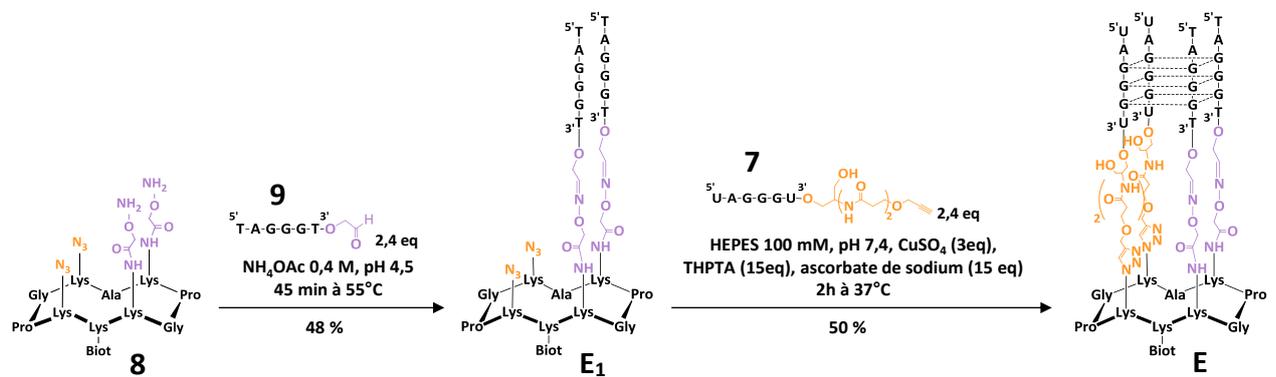


Schéma 19 : synthèse de l'édifice hybride ADN/ARN **E**

La formation des liens éther d'oxime, est réalisée dans les mêmes conditions qu'utilisées précédemment pour l'édifice **C** (tampon acétate d'ammonium 0,4 M, pH 4,5, 45 min à 55°C). L'intermédiaire **E₁** est obtenu après purification par RP-HPLC avec un rendement de 48% et caractérisé par RP-HPLC et spectrométrie de masse (figure 54). Il est à noter que dans le cas de motif d'ADN G-quadruplexe, la purification par chromatographie échangeuse d'ions n'est pas nécessaire, aucune formation de G4 intermoléculaires n'a été observée.

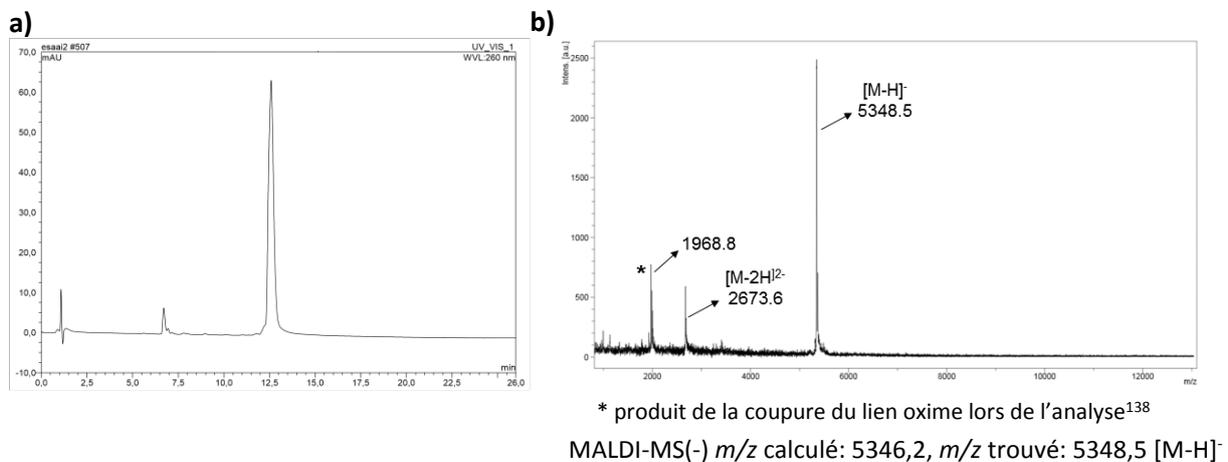
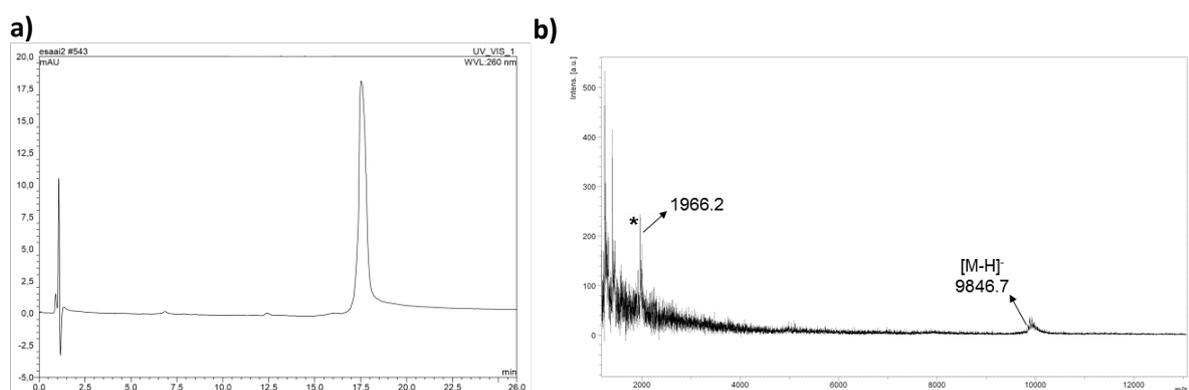


Figure 54 : (a) chromatogramme AELC et (b) spectre MALDI-ToF de l'édifice intermédiaire **E₁**

Les brins d'ARN sont ensuite greffés sur l'intermédiaire **E₁** *via* la CuAAC, en utilisant les mêmes conditions que pour l'édifice **D** (tampon HEPES 100 mM, pH 7,4, 3 eq de CuSO₄, 15 eq de THPTA et 15 eq d'ascorbate de sodium). Après 2 h de réaction à 37°C, un traitement EDTA est effectué ainsi qu'un dessalage par colonne d'exclusion de taille. Le brut réactionnel est ensuite purifié par AELC et l'édifice hybride ADN/ARN **E** est obtenu avec un rendement de 50%. Cet édifice a ensuite été caractérisé par AELC et spectrométrie de masse (figure 55).



* produit de la coupure du lien oxime lors de l'analyse¹³⁸
MALDI-MS(-) m/z calculé: 9844,0, m/z trouvé: 9846,7 [M-H]⁻

Figure 55 : (a) chromatogramme AELC et (b) spectre MALDI-Tof de l'édifice hybride ADN/ARN **E**

II.3. Synthèse et caractérisation d'un édifice mimant un motif G-quadruplexe antiparallèle, dérivant de l'ARN télomérique

Comme évoqué en introduction (p 15), les motifs G-quadruplexes à topologie antiparallèle ne sont pas favorisés, contraindre cette structuration par une pré-orientation des brins constitue donc un challenge. Nous avons voulu vérifier si cela était possible en utilisant l'approche TASQ. Pour ce faire, nous avons essayé d'assembler un motif tétramoléculaire dérivant de la séquence télomérique. Afin d'orienter les brins d'ARN en conformation antiparallèle, deux brins ont été greffés sur le châssis par leur extrémité 3' et les deux autres, de séquences différentes ont été attachés par leur extrémité 5' (figure 56). Pour pouvoir contrôler cet assemblage, nous avons utilisé deux ligations chimiosélectives différentes, la ligation oxime et la CuAAC.

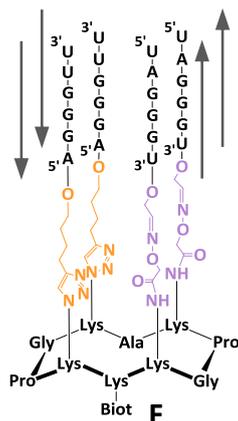


Figure 56 : structure chimique de l'édifice **F** représentant un G4 antiparallèle dérivant de l'ARN télomérique

II.3.a. Synthèse du châssis peptidique

Le châssis peptidique utilisé dans l'assemblage de l'édifice **F** est le châssis **8** (schéma 17), employé dans la synthèse de l'édifice hybride **E**. La synthèse de ce châssis bi-fonctionnalisé, par deux fonctions azotures et deux fonctions oxyamines est décrite p 59.

II.3.b. Synthèse des brins d'ARN

i) Synthèse des brins d'ARN 3'aldéhyde

L'oligonucléotide utilisé est l'ARN 3'aldéhyde **3**. Sa synthèse est décrite dans la partie II.1.1.b, p 55.

ii) Synthèse des brins d'ARN 5'alcyne

L'oligonucléotide greffé par son extrémité 5' est fonctionnalisé par un groupement alcyne ($5'YAGGGUU3'$, Y = 5'alcyne). Cette fonction est introduite en fin d'élongation, par l'ajout d'un phosphoramidite modifié par un groupement hexynyl (schéma 20). L'ARN 5'alcyne **10** est obtenu après déprotection en milieu basique et purification avec un rendement de 31%, il a été caractérisé par AELC et spectrométrie de masse (figure 57).

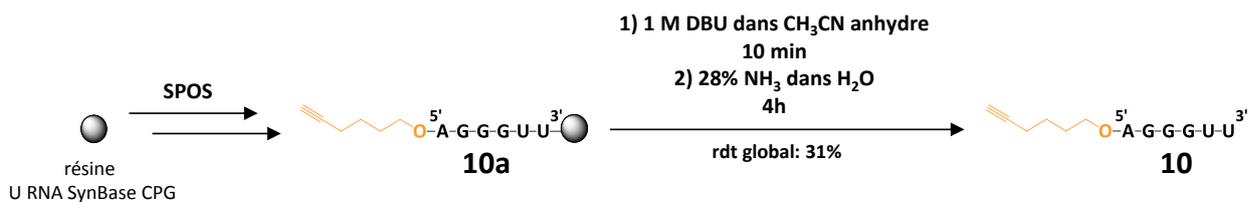


Schéma 20 : synthèse de l'ARN 5'alcyne **10**

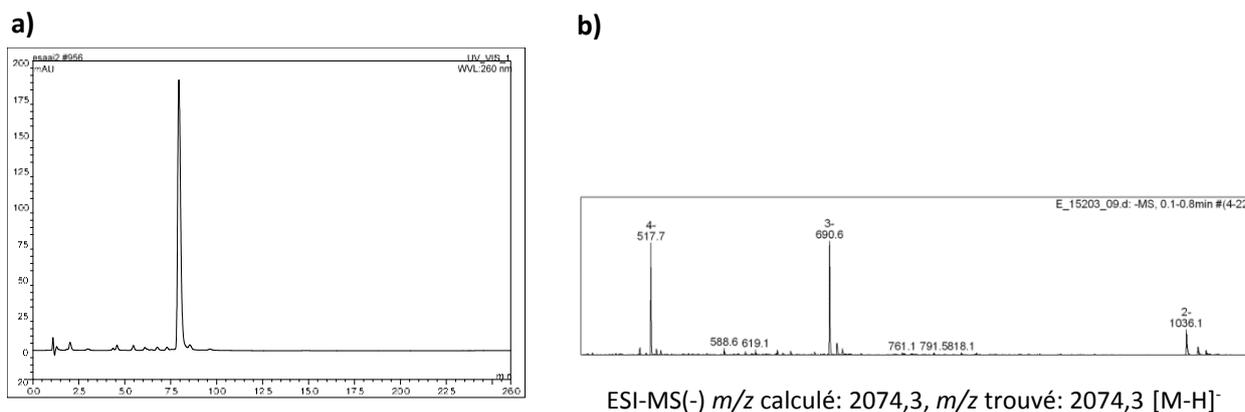


Figure 57: (a) chromatogramme AELC et (b) spectre ESI-MS de l'ARN 5'alcyne **10**

II.3.c. Assemblage de l'édifice

L'assemblage de l'édifice **F** est réalisé en deux étapes (schéma 21).

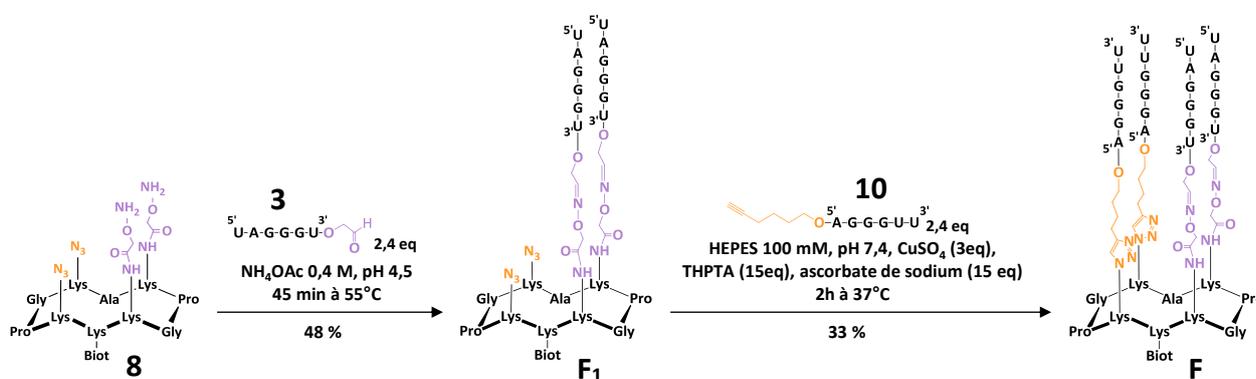
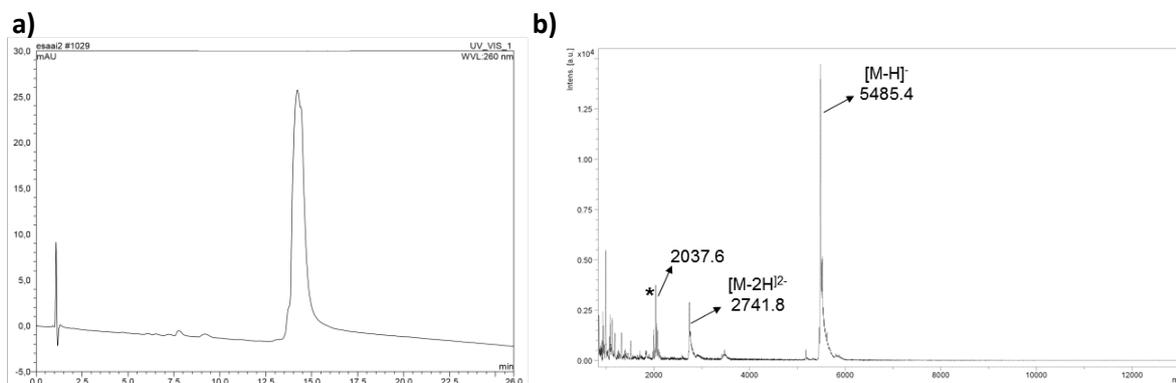


Schéma 21: synthèse de l'édifice ARN G4 antiparallèle **F**

Pour la même raison que l'édifice hybride, il est nécessaire d'effectuer la ligation oxime en premier. Cette dernière réalisée dans une solution tamponnée d'acétate d'ammonium 0,4 M à pH 4,5 permet de greffer les brins d'ARN 3'aldéhyde **3** sur le châssis peptidique. L'intermédiaire **F₁** est obtenu avec un rendement de 49% après purification par AELC (figure 58).



* produit de la coupure du lien oxime lors de l'analyse¹³⁸

MALDI-MS(-) m/z calculé: 5483,0, m/z trouvé: 5485,4 [M-H]⁻

Figure 58 : (a) chromatogramme AELC et (b) spectre MALDI-Tof de l'intermédiaire **F₁**

La deuxième étape consiste à lier les brins d'ARN 5' alcyne sur l'intermédiaire **F**₁ en utilisant la CuAAC, suivant le protocole mise au point p 55. Le brut réactionnel est purifié par AELC et l'édifice ARN G4 **F** est obtenu avec un rendement de 33 %. Cet édifice a ensuite été caractérisé par AELC et spectrométrie de masse (figure 59).

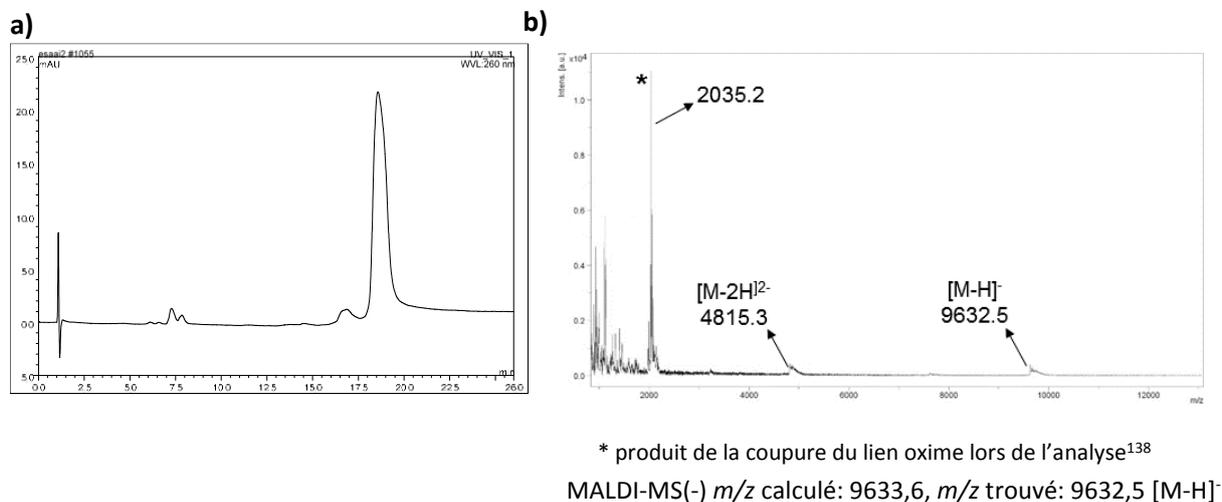


Figure 59 : (a) chromatogramme AELC et (b) spectre MALDI-ToF de l'édifice ARN antiparallèle **F**

III. Etudes de la topologie des édifices mimant les G-quadruplexes de l'ARN télomérique humain

Les caractérisations des différents édifices par AELC et spectrométrie de masse ont permis de valider leurs formations d'un point de vue chimique, mais ne prouvent pas la structuration des motifs G-quadruplexes. Il est donc nécessaire d'utiliser d'autres méthodes d'analyses, permettant de vérifier que les oligonucléotides présents sur les édifices se structurent en motif G-quadruplexe. Il est également important de déterminer la topologie adoptée par les G-quadruplexes et d'évaluer leur stabilité. Nous avons utilisé la RMN, le dichroïsme circulaire et l'absorption UV pour réaliser ces études.

III.1. Etudes d'absorption UV : "thermal difference spectrum" (TDS)

La technique « thermal difference spectrum » (TDS) utilise la différence d'absorbance d'un motif G-quadruplexe à deux températures²⁴. Dans le cas des motifs G-quadruplexes d'ARN les absorbances sont mesurées entre 220 et 320 nm à 5 et 80°C. Le spectre de différence d'absorbance est obtenu par soustraction du spectre à 5°C au spectre enregistré à 80°C et normalisation. Pour les quatre édifices **C-F**, les données ont été collectées dans du tampon Tris 10 mM à pH 7,4 en présence de cations métalliques (10 mM NaCl ou 100 μM KCl) ou en absence de cations ajoutés. Les spectres

obtenus, présentés dans la figure 60, confirment la structuration des motifs G-quadruplexes. En effet, ils présentent tous les signaux caractéristiques d'un G-quadruplexe avec deux maxima aux alentours de 243 et 273 nm et un minimum proche de 295 nm.

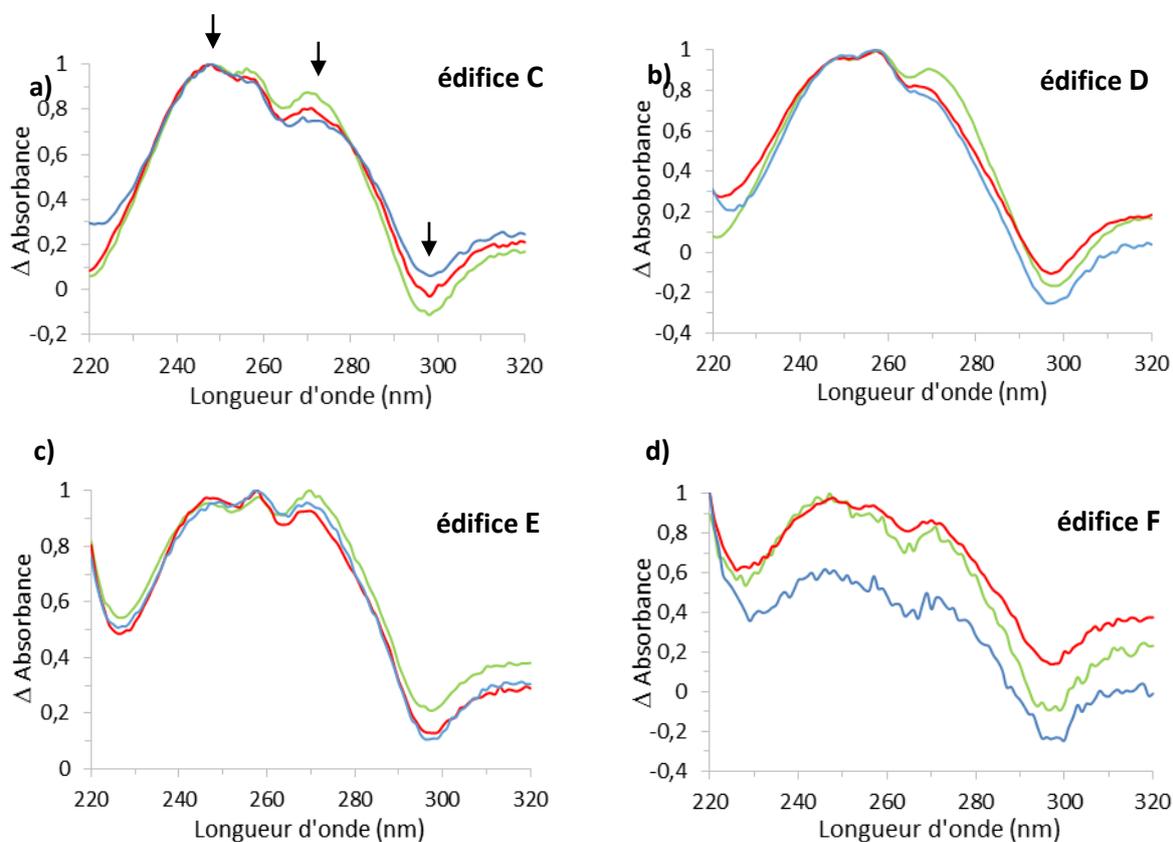


Figure 60 : spectres normalisés représentant la différence d'absorbance entre 80 et 5°C pour les edifices C-F, conditions : 1,6 μM d'édifice dans du tampon Tris-HCl 10 mM à pH 7,4 en présence de 10 mM NaCl (bleu) ou de 100 μM KCl (rouge) ou en absence de cation ajouté (vert)

III.2. Etudes de dichroïsme circulaire

Les spectres « TDS » ont permis de confirmer la structuration des motifs G-quadruplexes. Les analyses de dichroïsme circulaire permettent quant à elles de déterminer la topologie adoptée par les motifs G-quadruplexes et d'évaluer leur stabilité. Elles sont réalisées soit en présence de cations métalliques (10 mM NaCl ou 100 μM KCl) soit en absence de cation ajouté.

III.2.a. Etude de la topologie des motifs G-quadruplexes

Les spectres CD ont été enregistrés à 20°C et un contrôle ne comportant pas de châssis peptidique a été utilisé (séquence TERRA : 4 x r(UAG₃U)). Quel que soit l'édifice étudié, les spectres CD présentent tous le même profil avec un maximum à 260 nm et un minimum à 240, caractéristique d'une topologie parallèle (figure 61).

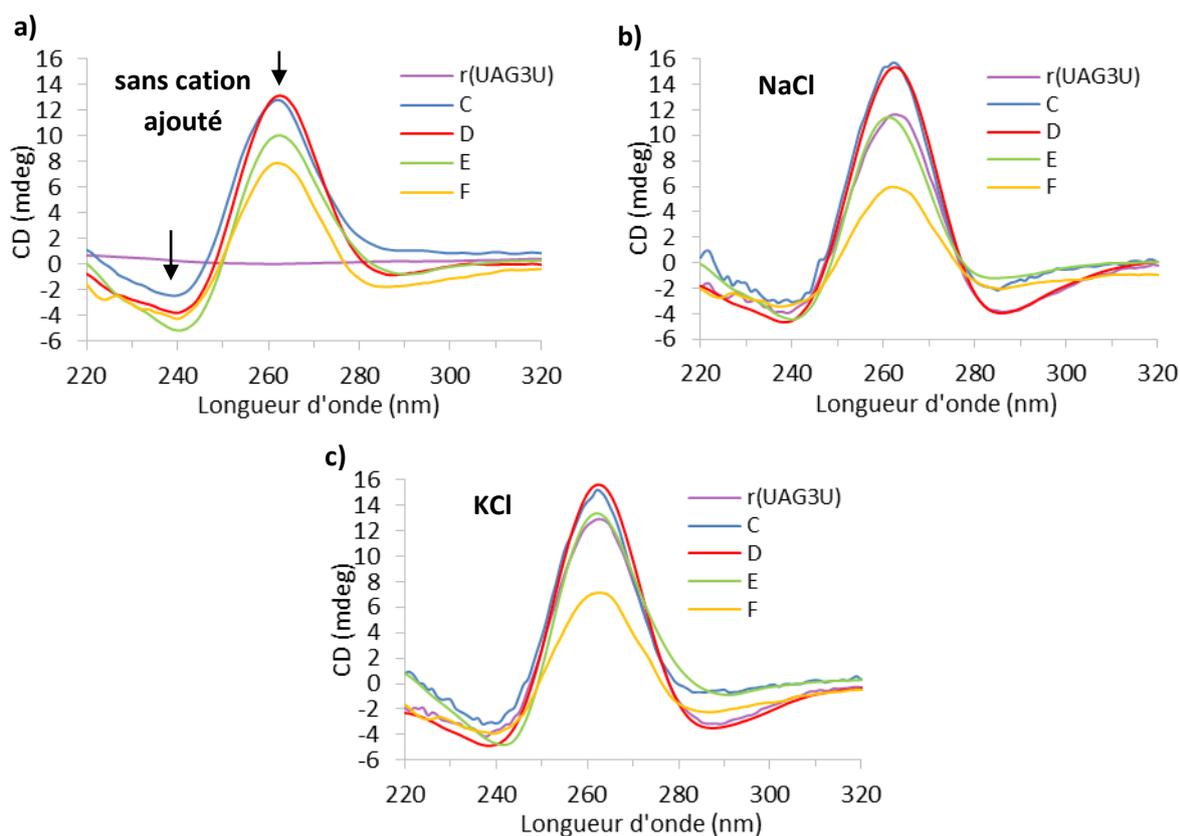


Figure 61 : spectres CD à 20°C de la séquence r(UAG₃U) (6,4 μM) et des édifices **C-F**, conditions : 1,6 μM d'édifices dans 10 mM de tampon Tris-HCl (a) en absence de cation ajouté ou en présence de (b) 10 mM NaCl ou (c) 100 μM KCl

Il est à noter qu'en absence de cation métallique ajouté (figure 61.a), le contrôle sans châssis peptidique ne se structure pas en G-quadruplexe. A l'inverse, dans les mêmes conditions, les brins oligonucléotidiques de tous les édifices forment un motif G-quadruplexe parallèle. La présence du châssis peptidique permet donc la formation de motif G-quadruplexe ARN en l'absence de cations métalliques ajoutés. La structuration des motifs G-quadruplexe en conformation parallèle était attendue pour les édifices ARN parallèle **C**, **D** et hybride **E**, mais elle est surprenante pour l'édifice **F**. En effet, bien que les G-quadruplexes d'ARN forment préférentiellement des motifs parallèles, il était envisageable que la pré-orientation de brins d'ARN à polarité opposés, induite par le châssis peptidique apporte suffisamment de contraintes pour favoriser la formation d'un motif antiparallèle. Pour expliquer les signaux CD obtenus, nous avons émis l'hypothèse d'un repliement des brins ARN dans une ou plusieurs conformations parallèles, plus stables thermodynamiquement. En effet, la longueur des espaceurs peut permettre une telle flexibilité des brins. De plus, l'intensité des signaux CD et moins importante pour l'édifice **F**, ce qui met en évidence la formation d'une structure G-quadruplexe moins stable ou formée de moins de tétrades (par exemple figure 62).

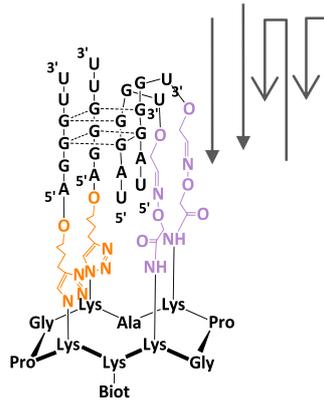


Figure 62 : hypothèse de formation d'un motif G-quadruplexe parallèle sur l'édifice **F**

III.2.b. Etude de la stabilité des motifs G-quadruplexes

La stabilité des motifs G-quadruplexes est évaluée en mesurant leur température de dénaturation (T_m) de 5 à 80°C en présence de cations métalliques (Na^+ ou K^+) ou en absence de cation métallique ajouté (tableau 1). Pour cette étude, nous avons dû diminuer la concentration saline utilisée (10 mM de NaCl et 100 μM de KCl) car les édifices sont trop stables dans les concentrations habituelles de 100 mM.

i) Evaluation de la stabilité des édifices

Motifs G4 Conditions: Tris 10 mM pH 7,4	4 x 5' UAG ₃ U 3'					
		C	D	E	F	A
NaCl 10 mM	45°C	66°C	55°C	51°C	44°C	51°C
KCl 100 μM	46°C	58°C	47°C	47°C	43°C	48°C
Sans sel ajoutés		45°C	34°C	34°C	27°C	32°C

Tableau 1 : température de dénaturation (T_m) de la séquence r(UAG₃U) (6,4 μM) et des édifices **A**, **C-F**, conditions : 1,6 μM d'édifices dans 10 mM de tampon Tris-HCl pH 7,4 en absence de cation ajouté ou en présence de 10 mM NaCl ou 100 μM KCl

En présence de cations métalliques, les édifices contraints **C** et **D** présentent des T_m supérieurs à ceux du brin contrôle (augmentation d'au moins 10°C). Alors qu'aucun signal de G-

quadruplexe n'est présent pour le brin d'ARN seul sans ajout de sel, des T_m de 45°C et 34°C sont respectivement obtenus pour les édifices **C** et **D**. Tous ces éléments démontrent que le châssis peptidique stabilise le motif G-quadruplexe parallèle.

Quelques que soient les conditions salines utilisées, l'édifice **D** montre des T_m de 10°C inférieurs à ceux de l'édifice **C**. Ce résultat indique que la longueur des espaceurs joue un rôle important dans la stabilisation des motifs G-quadruplexes. Une chaîne trop longue apporte plus de flexibilité au système mais entraîne une diminution de la stabilité du G-quadruplexe. Il est toutefois nécessaire de laisser une certaine longueur à l'espaceur, afin que le système ne soit pas totalement figé.

Les valeurs de T_m obtenues pour l'édifice **E** sont similaires à celles de l'édifice **A**. Ces résultats soulignent que la formation d'édifice hybride ADN/ARN est plausible et que le motif G-quadruplexe formé est aussi stable qu'un motif G-quadruplexe comportant quatre brins d'ADN. Ceci renforce donc l'hypothèse de son existence *in-cellulo* et permet d'envisager pour la première fois des études d'interactions sur cette structure.

Enfin, les T_m obtenus pour l'édifice **F** sont d'au moins 15°C inférieurs à ceux de l'édifice **C**. Cette différence de stabilité semble confirmer l'hypothèse d'un réarrangement des brins d'ARN pour se replier en conformation parallèle, malgré la contrainte imposée en faveur d'un repliement en conformation antiparallèle.

ii) Influence de la concentration sur la température de dénaturation

Afin de s'assurer que le signal obtenu en CD ne soit pas dû à la formation de motifs G-quadruplexes intermoléculaires formés par l'assemblage de brins d'ARN présents sur deux ou plusieurs édifices différents, les températures de dénaturations de l'édifice hybride **E** ont été mesurées à trois concentrations différentes : 0,8, 1,6 et 3,2 μM . Il est à noter que l'obtention de T_m différents à différentes concentrations est caractéristique d'interactions intermoléculaire dont la cinétique est régit par la concentration. Les analyses ont été réalisées pour chaque concentration en présence de NaCl (figure 63.a) et en présence de KCl (figure 63.b). L'obtention de T_m identiques démontre que la concentration n'influence pas la stabilité du motif G4, ce qui confirme une structuration intramoléculaire du G4 hybride ADN/ARN **E**.

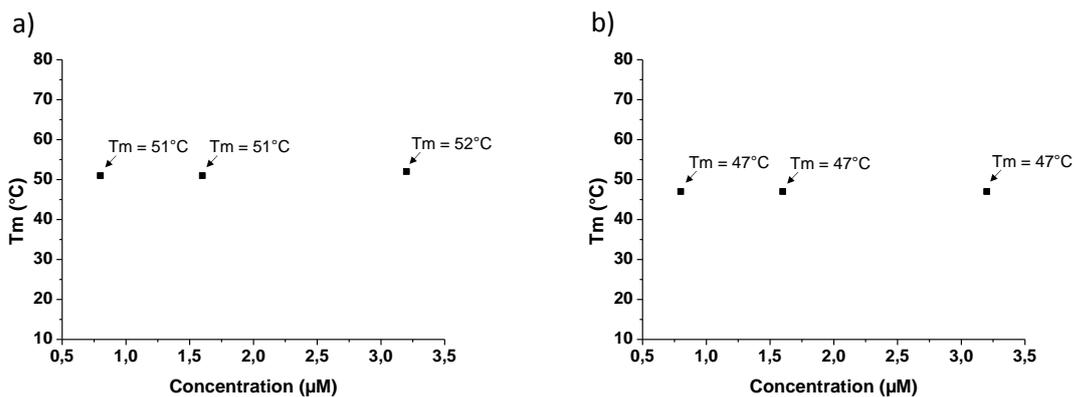


Figure 63 : température de dénaturation (T_m) de l'édifice hybride ADN/ARN **E** à 0,8, 1,6 et 3,2 μM , dans Tris-HCl 10 mM et NaCl 10 mM (a) ou KCl 100 μM (b)

Dans le cas de l'édifice **F**, il est également nécessaire d'effectuer ce type d'analyse, afin de vérifier que le motif G-quadruplexe parallèle est formé par le réarrangement des brins d'ARN entre eux et non pas par l'assemblage de deux édifices. La formation du motif G4 intramoléculaire a été validée par l'obtention de T_m très fortement similaires pour les trois concentrations utilisées (figure 64).

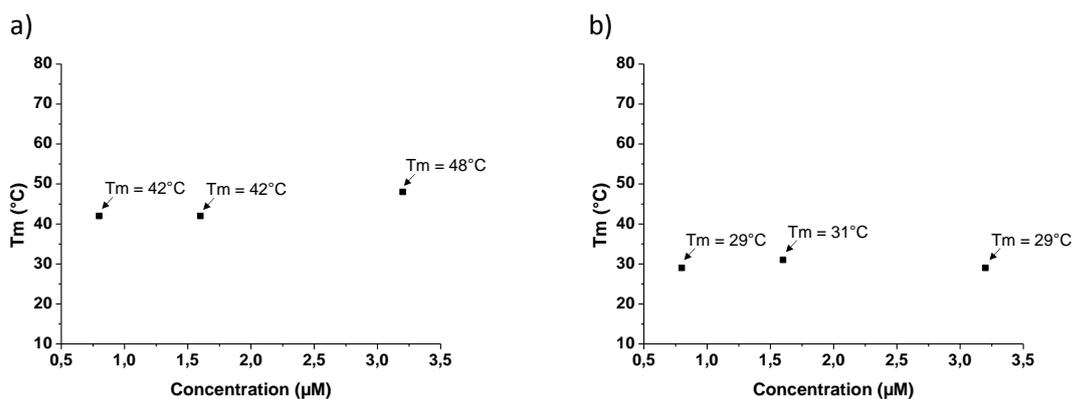


Figure 64 : température de dénaturation (T_m) de l'édifice ARN G4 **F** à 0,8, 1,6 et 3,2 μM , dans Tris-HCl 10 mM et (a) NaCl 10 mM ou (b) KCl 100 μM

iii) Influence de la cinétique de repliement sur la température de dénaturation

Pour l'édifice **F**, contrairement aux trois autres, nous avons observé une hystérésis entre les courbes de dénaturation (5 à 80°C) et les courbes d'hybridation (80 à 5°C) (figure 65).

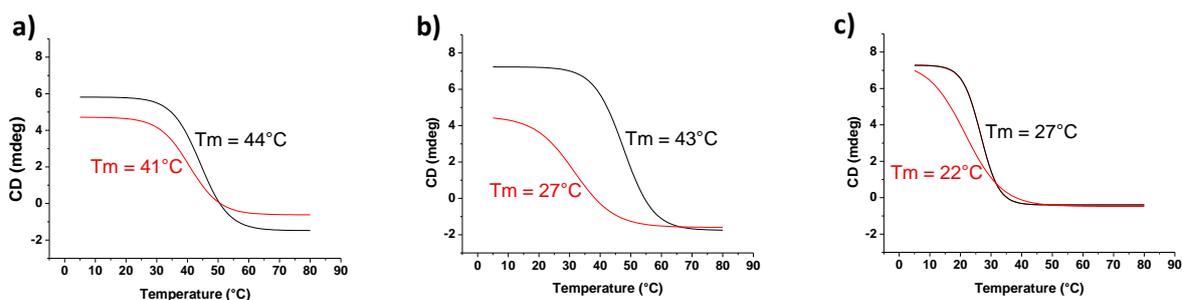


Figure 65 : courbes de dénaturation de 5°C à 80°C (noir) et d'hybridation de 80°C à 5°C (rouge) de l'édifice **F**, conditions : 1,6 μM d'édifice dans Tris-HCl pH 7,4, avec (a) 10 mM NaCl, ou (b) 100 μM KCl, ou (c) sans sel ajouté

Alors que l'hystérésis est modérée en présence de Na⁺ (figure 65.a) ou en l'absence de cations ajoutés (figure 65.b) (ΔT_m observé de 3°C et 5°C, respectivement), il est beaucoup plus marqué en présence de K⁺ (figure 65.c) (ΔT_m observé de 16°C), ce qui suggère une association plus lente des brins dans cette condition saline particulière. Globalement, ces données soutiennent l'idée que les brins d'ARN du conjugué **3** s'engagent dans une voie de repliement complexe à cinétique lente, pour adopter une architecture G4 parallèle stable thermodynamiquement.

III.3. Etudes par résonance magnétique nucléaire

Les études par RMN permettent de confirmer la structuration des oligonucléotides en motif G-quadruplexes (introduction, p 18). Pour les quatre édifices **C-F**, des signaux caractéristiques des G4 aux alentours des 10-12 ppm sont observés. Les profils des édifices G-quadruplexes parallèles **C** et **D** sont similaires (figure 66), ce qui confirme la structuration du même motif G-quadruplexe. Ce dernier comporte trois tétrades et est parfaitement symétrique, ce qui se traduit par la présence de trois signaux regroupant chacun les quatre protons imino d'une tétrade.

Le spectre de l'édifice hybride ADN/ARN **E** ne comporte pas de signaux parfaitement définis, ce qui est dû à la non-symétrie du motif G-quadruplexe. En effet, au sein d'une même tétrade, deux guanines proviennent d'un brin d'ARN et les deux autres d'un brin d'ADN, les quatre protons imino ne possèdent donc pas tout à fait le même environnement. Par conséquent ils ne peuvent pas avoir un déplacement chimique identique.

Pour l'édifice **F**, le signal obtenu est un signal hétérogène, qui traduit la coexistence de plusieurs espèces de motif G-quadruplexe en solution. Au vu des données de dichroïsme circulaire (III.2. p 68) qui nous laissent penser à un repliement de deux brins pour former un motif G4 parallèle, ces espèces pourraient être le fruit de toutes les combinaisons possibles dans le repliement des

brins. De plus, les guanines en 3' pourraient être impliquées dans ce repliement, défavorisant ainsi la stabilisation de la tétrade et entraînant la formation de motifs G4 ne possédant que deux tétrades.

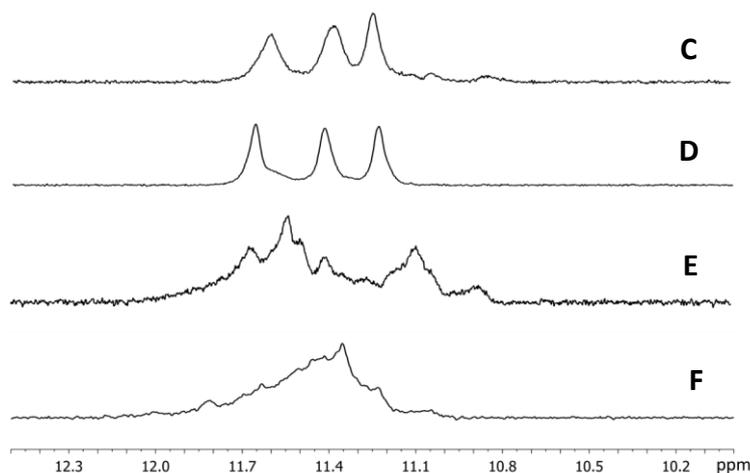


Figure 66 : spectre RMN ^1H (500 Mz) à 25°C, représentant la région des protons imino des édifices **C**, **D**, **E** et **F**, conditions : 80 μM d'édifices en présence de 10 mM de tampon phosphate sodium à pH 7,5, de 10 mM de NaCl et de 10 % de D_2O

IV. Bilan

Ces travaux qui ont fait l'objet d'une publication parue en 2016¹⁴¹ (L. Bonnat, et *al.* *Chem Eur. J.* **2016**, *22*, 3139-3147) ont permis d'isoler et de stabiliser un motif ARN G4 parallèle, ainsi qu'un motif hybride ADN/ARN. Ces motifs constituent des cibles potentielles pour l'identification et la conception de ligands affins et spécifiques de G-quadruplexe. Au cours de ce travail, nous avons pu également confirmer que la formation d'une topologie antiparallèle est fortement défavorisée pour les structures G-quadruplexe d'ARN. Malgré une orientation forcée des brins d'ARN, nous n'avons pas pu surmonter la barrière thermodynamique qui empêche la formation de la topologie antiparallèle.

Bien que ces systèmes ne miment pas l'intégralité des motifs G-quadruplexes, les boucles étant absentes des mimes synthétisés, ils constituent néanmoins une plateforme de référence pour commencer à évaluer l'affinité et la sélectivité de ligands de G4. De plus, ces systèmes rendent maintenant possible le ciblage sélectif de certains motifs G-quadruplexe et notamment les structures hybrides ADN/ARN. Des études de SPR permettant d'évaluer l'affinité de certains ligands G4 ont notamment été réalisés avec les édifices **C** et **E**, les résultats sont décrits en partie I du chapitre 2 p 127.

Partie II : Mîme d'ADN G-quadruplexe présent dans le g nome du VIH-1

I. Introduction

Tout comme les strat gies anticanc reuses visant les G-quadruplexes de la r gion t lom rique ou encore des promoteurs d'oncog nes, les G-quadruplexes des g nomes viraux, pourraient  galement  tre cibl es par des ligands s lectifs des G-quadruplexes afin de bloquer la prolif ration du virus. Dans ce contexte, cette deuxi me partie est consacr e   la description de la synth se et de la caract risation d'un m me de G-quadruplexe (HIV-PRO1) pr sent au sein du g nome du VIH-1 (figure 67).

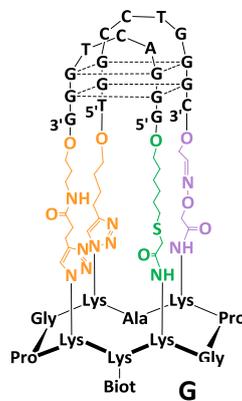


Figure 67 : structure chimique de l' difice **G** mimant le motif G-quadruplexe HIV-PRO1

I.1. D couverte du G-quadruplexe de la s quence HIV-PRO1 et int r t de son ciblage

Le virus de l'immunod ficiency humaine, abr g  sous le nom de VIH (HIV en anglais) est un r trovirus responsable du syndrome d'immunod ficiency acquise, plus commun ment appel  SIDA. La paternit  de ce virus d couvert dans les ann es 80 a  t  longtemps controvers e, mais fini par  tre octroy e   une  quipe de l'institut Pasteur, ce qui valut   L. Montagnier et F. Barr -Sinoussi le prix Nobel de m decine en 2008. Ce virus qui a touch  plus de 36 millions de personnes en 2016 selon l'ONU, affaibli le syst me immunitaire de son h te, le rendant ainsi plus vuln rable aux infections. Il est aujourd'hui le virus le plus  tudi , mais son importante variabilit  g n tique complique le d veloppement de traitements et de vaccins. Le VIH p n tre dans les cellules cibles (en particulier les lymphocytes T) pr sentant   leur surface des r cepteurs CD4, reconnu par les prot ines gp120 pr sentes sur la surface virale. Ce processus de reconnaissance permet la lib ration du mat riel viral (ARN viral et prot ines virales) au sein de la cellule cible. Le VIH est un r trovirus, c'est- -dire qu'il poss de pour g nome de l'ARN et non de l'ADN. Pour assurer la r plication du virus,

le génome viral doit être intégré au génome de la cellule cible, nécessitant donc la reconversion de l'ARN viral en ADN. Ce processus est assuré par la transcriptase inverse, une enzyme de type ADN-polymérase ARN-dépendante, qui ne possède pas de mécanisme de détection des erreurs de transcription. Les erreurs sont donc fréquentes et c'est une des raisons pour lesquelles le VIH possède une très grande variabilité génétique. Deux types de VIH ont été décrits, le VIH-1 le plus répandu et le VIH-2 qui touche essentiellement le continent Africain. Dans le cas du VIH-1, l'ADN viral nouvellement formé migre dans le noyau et est intégré dans le génome de la cellule infectée. A ce stade, le provirus intégré du VIH-1 est composé de deux régions régulatrices (5' LTR et 3' LTR) et de neuf gènes codant pour des protéines virales. La machinerie cellulaire de la cellule hôte transcrit les gènes viraux et produit de nouvelles protéines virales, permettant ainsi l'assemblage de nouveaux virus. Une étape clé du cycle viral est la régulation de la transcription du provirus par un promoteur viral situé à l'extrémité 5' LTR. La région U3 de ce promoteur contient cinq sites de liaison du facteur de transcription (NF- κ B-1 et 2, SP1-1, 2 et 3) et possède une séquence riche en guanines de cinquante nucléotides. Une première étude biophysique réalisée par l'équipe de S. N. Ritcher⁶⁰ a montré la formation de deux motifs G-quadruplexes dans cette séquence riche en guanine. Ces travaux ont également mis en évidence l'inhibition de l'activité du promoteur lors de la stabilisation des G-quadruplexes par des ligands spécifiques de G4. Cette découverte met en avant l'activité antivirale exercée par les G-quadruplexes, qui représentent ainsi de nouvelles cibles pour le développement de médicaments antiviraux. Un an plus tard, D. Piekna-Przybylska et ses collaborateurs¹⁴² ont confirmé la formation de G-quadruplexes au niveau des sites de liaison du facteur de transcription, cruciaux pour l'initiation de la transcription.

La même année, l'équipe de S. N. Ritcher⁵⁹ décrit l'existence de motifs G-quadruplexes, mais cette fois si au sein de la région régulatrice 3' LTR, plus précisément dans la partie codant pour le gène Nef. La stabilisation de ces motifs par des ligands de G-quadruplexe réprime l'expression de Nef et diminue la réplication virale. Nef présente plusieurs rôles durant l'infection virale, il est notamment impliqué dans l'évasion immunitaire et il a été démontré que l'absence de Nef semble corrélérer avec une faible charge virale et une inhibition de la maladie. Cibler les G4 situés dans la séquence de codage de Nef présente donc une nouvelle piste thérapeutique intéressante.

Suite à la découverte de G-quadruplexes dans la région U3 de l'extrémité 5' LTR par S.N. Ritcher, S. Amrane et ses collaborateurs¹⁴³ ont analysé en 2014 la conservation de cette séquence riche en guanine. Pour cela ils ont comparé 1684 séquences 5' LTR fournies par la base de données du VIH-1 et ont montré un niveau élevé de conservation. Ils ont également mis en évidence la capacité de toutes ces séquences à former des structures G-quadruplexes stables *in-vitro* (T_m compris entre 30 et 60°C). A partir de ses données, ils se sont focalisés sur l'étude du motif G-

quadruplexe le plus représenté parmi toutes les sous-espèces du VIH-1 entre les positions -57 et -78. Il s'agit du motif G4 de la séquence $5'\text{TGGCCTGGGCGGGACTGGG}3'$ nommé HIV-PRO1. Grâce à des études de dichroïsme circulaire et de RMN, ils ont obtenu la structure de ce motif G-quadruplexe (figure 68).

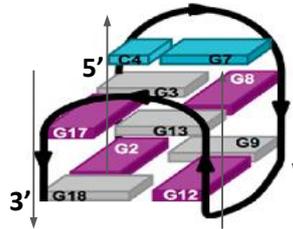


Figure 68 : structure du motif G-quadruplexe de la séquence HIV-PRO1

Ce dernier adopte une conformation antiparallèle et présente deux boucles latérales situées du même côté du G-quadruplexe, qui diffèrent par la nature et le nombre de leurs nucléotides (séquence $5'\text{ACT}3'$ et séquence $5'\text{CCTG}3'$).

D'après les travaux de S. N. Ritcher, la stabilisation de G-quadruplexes dans la région U3 entraînerait l'inhibition de l'activité virale; ce motif pourrait donc constituer une nouvelle cible pour le développement de ligands d'intérêt biologique voire thérapeutique. Dans ce contexte, nous avons voulu mimer ce motif G4 par notre approche TASQ pour en étudier la reconnaissance par de petites molécules.

II. Synthèse et caractérisation de l'édifice HIV-PRO1

Pour mimer le motif G4 antiparallèle de la séquence HIV-PRO1, nous avons décidé de reproduire ses deux tétrades de guanines, son plateau C-G (représenté en bleu sur la figure 69) ainsi que ses deux boucles supérieures (figure 69). Cet édifice présenterait alors les trois éléments de reconnaissance permettant l'identification de ligands spécifiques de cette structure G-quadruplexe. Pour reproduire ses trois éléments, deux brins d'ADN bi-fonctionnalisés qui diffèrent par la nature et le nombre de leurs nucléotides doivent être utilisés. Leur assemblage sur le châssis peptidique nécessite l'utilisation de trois ligations chimiosélectives différentes. La troisième ligation utilisée, orthogonale à la ligation oxime et à la CuAAC est le couplage thiol-chloroacétamide. Cette ligation présentée en introduction (p 38) est réalisée entre une fonction thiol et une fonction chloroacétamide. Pour des raisons de simplicité de synthèse, la fonction thiol est portée par l'oligonucléotide et la fonction chloroacétamide par le châssis peptidique. La stratégie utilisée consiste dans un premier temps à greffer indépendamment chaque oligonucléotide sur le châssis

peptidique, par la ligation oxime et couplage thiol-chloroacétamide, puis dans un second temps à former simultanément les deux boucles *via* la CuAAC.

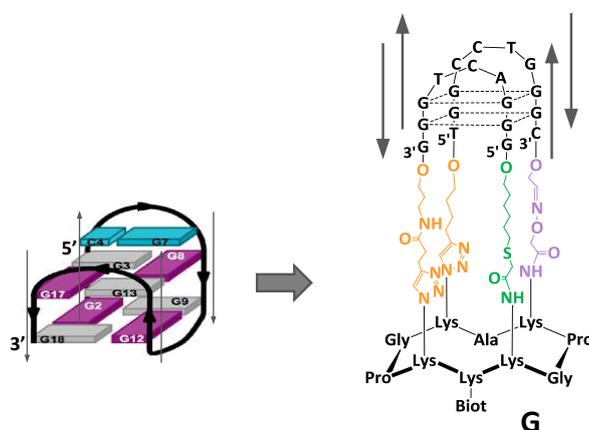


Figure 69 : représentation de l'édifice **G**, motif G4 antiparallèle de la séquence HIV-PRO1

II.1. Synthèse du châssis peptidique

Pour pouvoir réaliser les trois ligations, le châssis peptidique doit être tri-fonctionnalisé (figure 70) et comporter une fonction oxyamine, une fonction chloroacétamide et deux fonctions azotures.

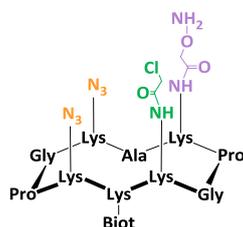


Figure 70 : structure chimique du châssis peptidique **11**

L'incorporation d'une troisième fonctionnalisation, nécessite la protection d'une des lysines par un groupement protecteur Dde, orthogonal au groupement Boc et aux fonctions azotures. Les premières étapes de synthèse (schéma 22), jusqu'à l'obtention de l'intermédiaire cyclisé **11c** sont similaires à celle du châssis **8** présentée p 59. La fonctionnalisation du châssis cyclisé nécessite ensuite plusieurs étapes de synthèse. Dans un premier temps, le groupe protecteur Dde est éliminé par traitement avec de l'hydrazine et la fonction chloroacétamide est introduite par action de l'anhydride chloroacétique. Dans un second temps, le groupement protecteur Boc est éliminé par une solution de TFA à 50% pour permettre l'introduction d'une fonction oxyamine protégée. Il est important de noter que la fonction oxyamine, même protégée par un groupement Boc reste assez réactive pour réagir avec l'anhydride chloroacétique. Il est donc impératif d'effectuer l'ajout de la fonction chloroacétamide avant celle de la fonction oxyamine protégée. Le peptide final **11** est

obtenu après libération de la fonction oxyamine en milieu acide. Il a été caractérisé par UPLC et spectrométrie de masse (figure 71).

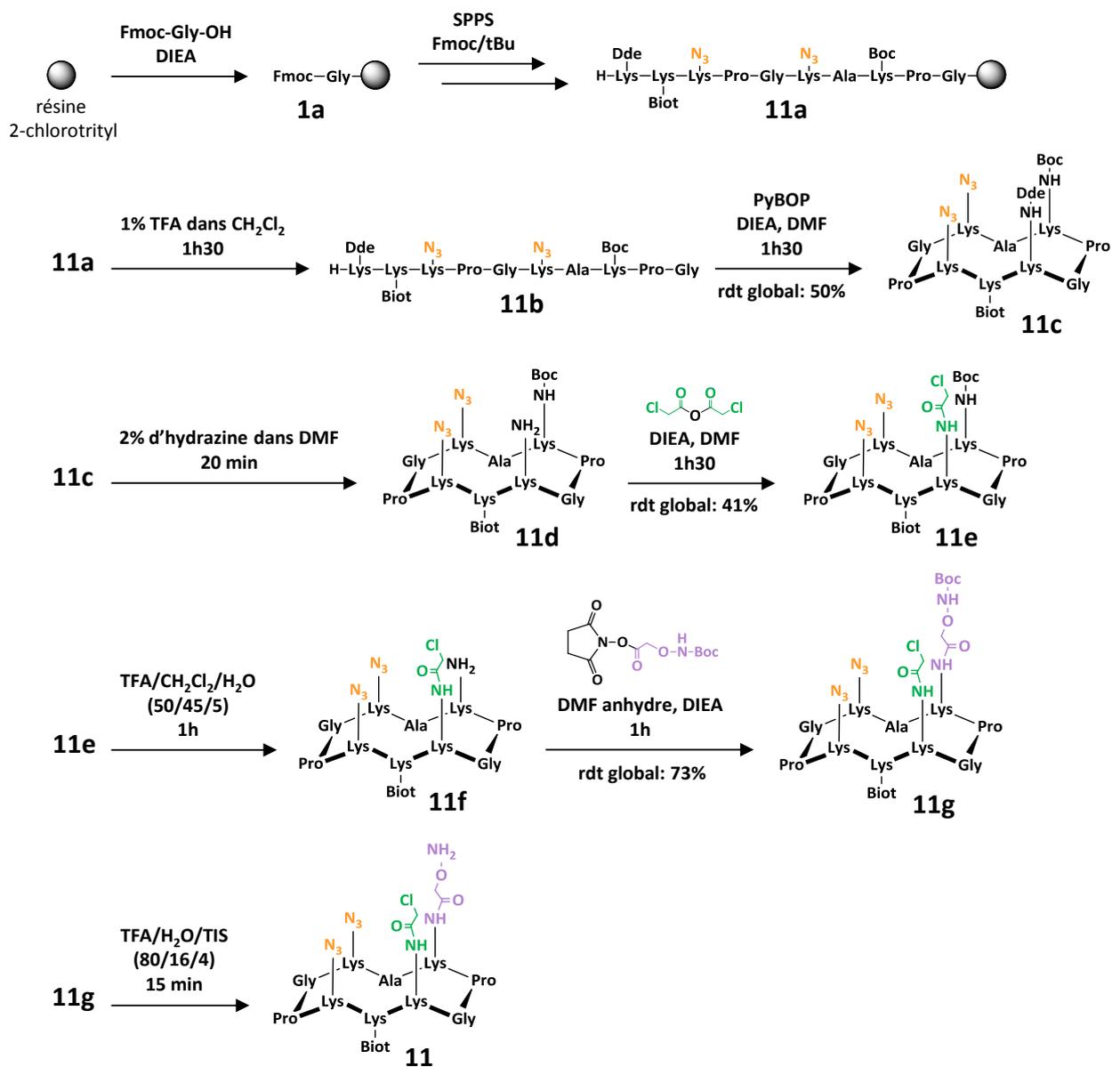
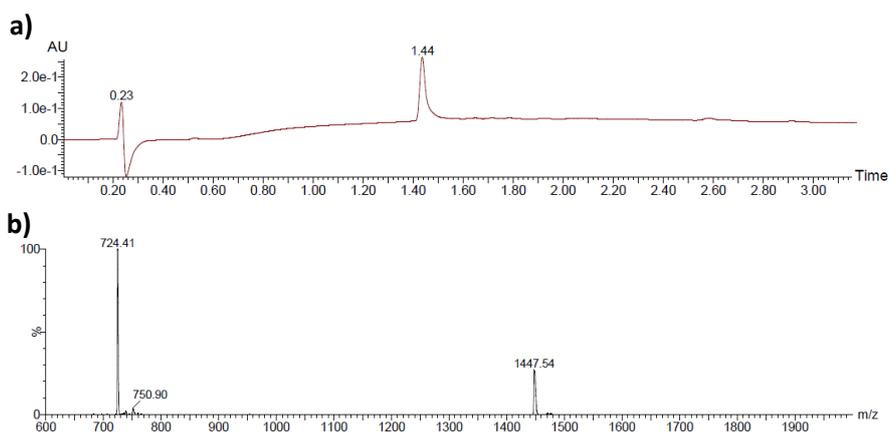


Schéma 22 : synthèse du châssis peptidique 11



ESI-MS(+) m/z calculé: 1447,7, m/z trouvé: 1447,5 $[M+H]^+$

Figure 71: (a) chromatogramme UPLC et (b) spectre ESI-MS du châssis peptidique **11**

II.2. Synthèse des brins d'ADN

Les deux brins d'ADN sont fonctionnalisés en leur extrémité 5' et 3' par les fonctions complémentaires à celles du châssis peptidique. Le premier est formé de dix nucléotides et présente la séquence suivante : $5'YTGGCCTGGGCX^{3'}$, Y représente la fonction 5'alcyne et X la fonction 3'aldehyde. Le deuxième, qui comporte un nucléotide de moins possède une fonction thiol en son extrémité 5' et une fonction alcyne en son extrémité 3' ($5'ZGGGACTGGGY^{3'}$, Y = 3' alcyne, Z = 5' thiol).

II.2.a. Synthèse du brin d'ADN 5'alcyne - 3'aldéhyde **12**

La séquence **12** est obtenue comme précédemment (p 64) en utilisant le support 3'glyceryl pour l'obtention de la fonction 3' aldehyde et le phosphoramidite « modifier 5' hexynyl » pour insertion de la fonction 5' alcyne (schéma 23).

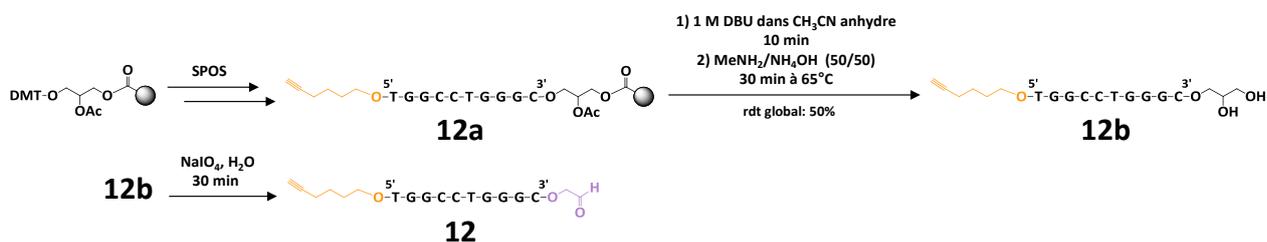


Schéma 23 : synthèse de l'ADN 5'alcyne - 3'aldéhyde **12**

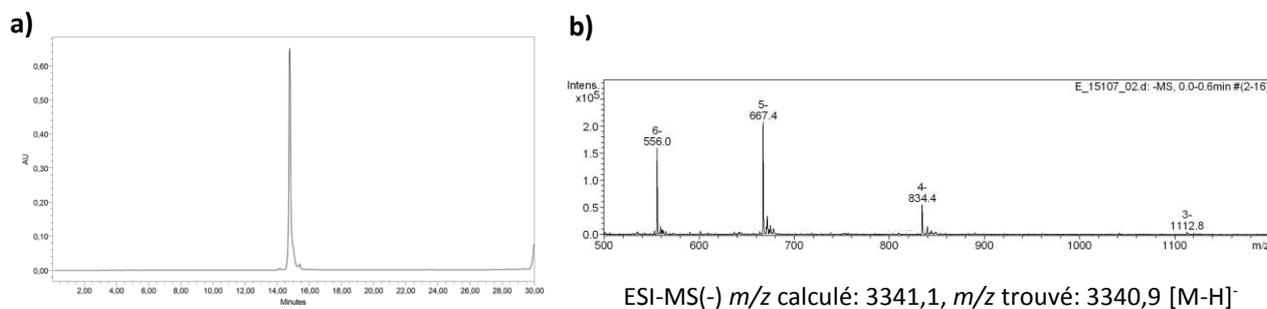


Figure 72 : (a) chromatogramme RP-HPLC et spectre (b) ESI-MS de l'ADN 5'alcyne - 3'aldehyde **12**

II.2.b. Synthèse du brin d'ADN 5'thiol - 3'alcyne **13**

Le brin d'ADN bifonctionnalisé **13** est synthétisé sur un support « 3'-PT-amino-modifier C3 CPG », qui permet l'obtention d'une fonction amine en position 3' après coupure du support en milieu basique, sur laquelle est ensuite additionnée la fonction alcyne (schéma 24). Nous avons choisi cette stratégie de fonctionnalisation plutôt que l'utilisation du support « 3'alkyne-modifier Serinol CPG » employé dans la synthèse des édifices ARN, pour conserver une longueur d'espaceur similaire. La fonction thiol est incorporée en position 5' sous forme protégée par un trityl, en utilisant un phosphoramidite modifié, le « 5'-thiol-modifier C6 ». Après élongation de la chaîne, déprotection et élimination du support, l'oligonucléotide **13b** est purifié par RP-HPLC et obtenu avec un rendement de 41%. L'acide 4-pentynoïque activé sous forme d'ester de succinimidyl est couplé sur l'amine en 3' pour obtenir la fonction 3'alcyne. Le couplage est considéré quantitatif et l'intermédiaire **13c** est utilisé dans la réaction suivante après élimination des réactifs sur colonne d'exclusion de taille. L'oligonucléotide final **13** est obtenu après libération de la fonction thiol en 5'. L'élimination du groupement trityl est réalisée en deux temps. La première étape consiste à mettre en présence l'oligonucléotide **13c** avec une solution aqueuse de nitrate d'argent et 0,1 M de TEAA à pH 6,5. Après 45 min à température ambiante, une solution aqueuse de DTT est ajoutée au brut réactionnel. Celui-ci est ensuite centrifugé, le surnageant est prélevé et dessalé sur colonne d'exclusion de taille. La réaction est considérée quantitative, l'oligonucléotide final **13** a été caractérisé par RP-HPLC et spectrométrie de masse (figure 73) et est prêt à être greffé sur le châssis peptidique.

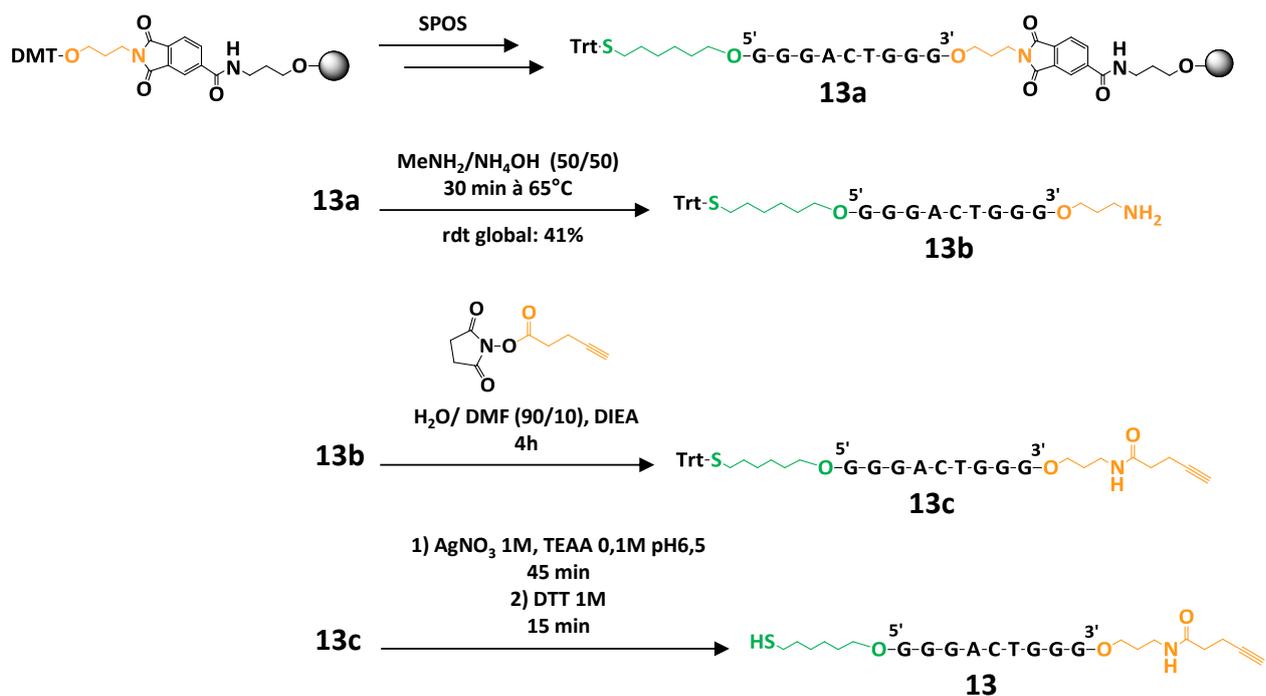


Schéma 24 : synthèse de l'ADN 5'thiol - 3'alcyne **13**

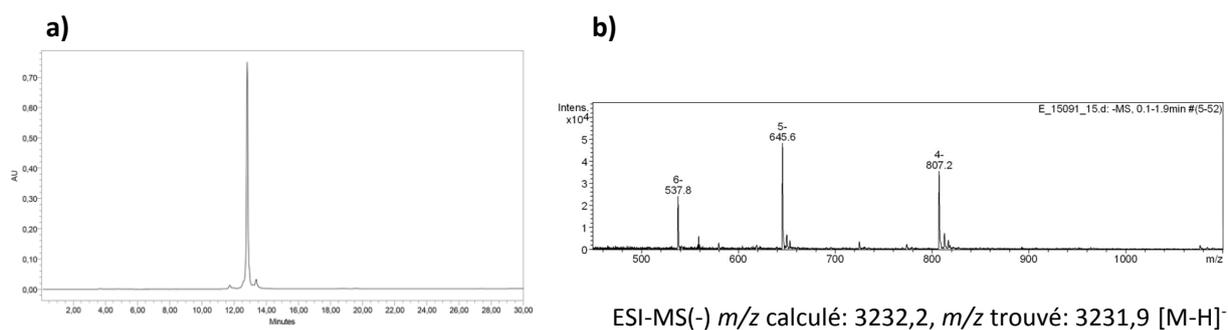


Figure 73 : (a) chromatogramme RP-HPLC et (b) spectre ESI-MS de l'ADN 5'thiol - 3'alcyne **13**

II.3. Assemblage de l'édifice

L'assemblage de l'édifice **G**, présenté dans le schéma 25 est réalisé en trois étapes.

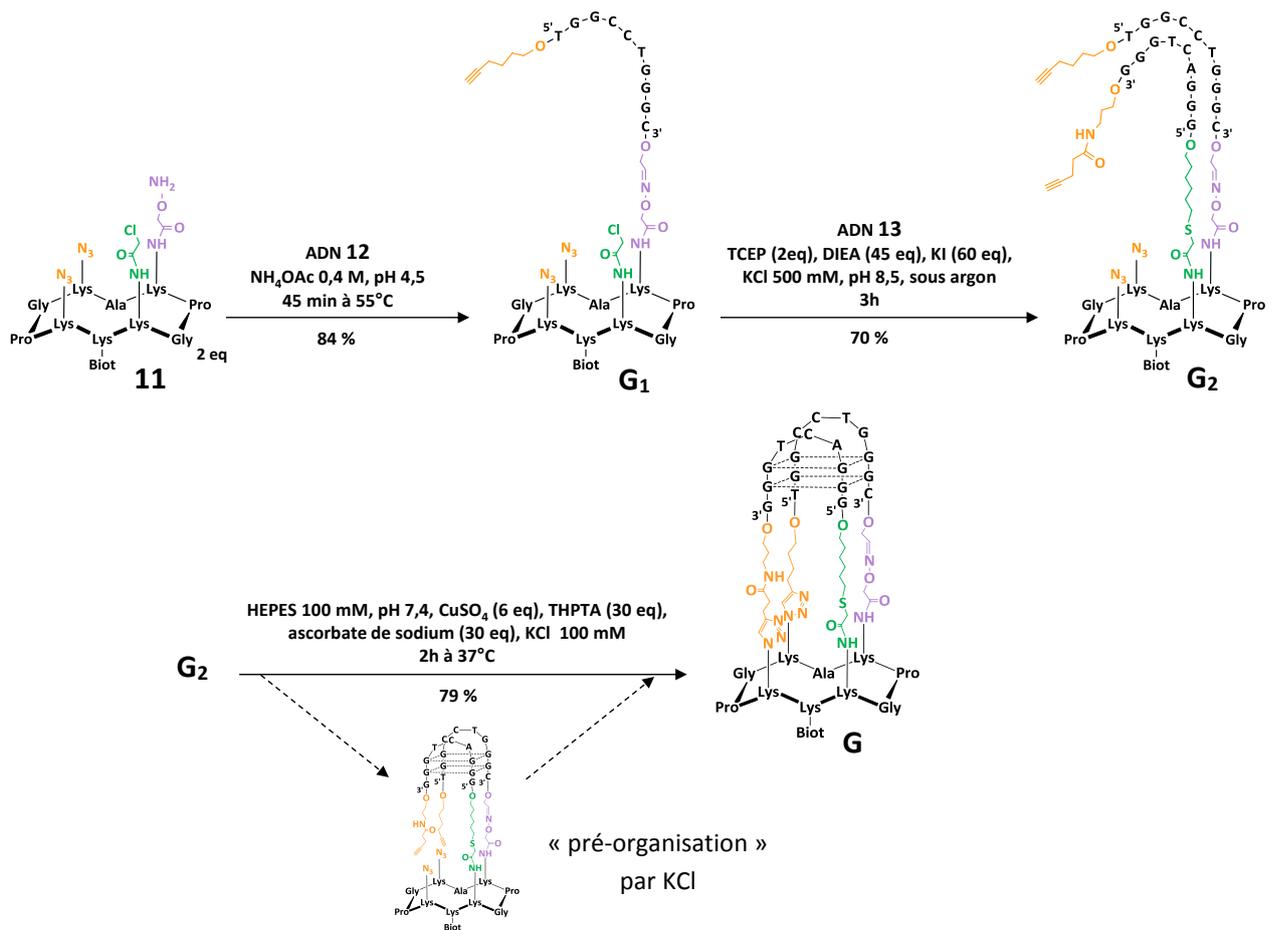


Schéma 25 : synthèse de l'édifice HIV-PRO1 **G**

La première étape consiste à greffer par son extrémité 3', le brin d'ADN **12** sur le châssis peptidique. Comme pour les édifices ARN **E** et **F** (p 56), la ligation oxime est effectuée en premier. Elle est réalisée avec 2 eq de peptide, en présence d'une solution tamponnée d'acétate 0,4 M à pH 4,5. Après 45 min de réaction à 55°C le brut réactionnel est purifié par RP-HPLC. L'intermédiaire **G₁** obtenu avec un rendement de 84% a été caractérisé par RP-HPLC et spectrométrie de masse (figure 74).

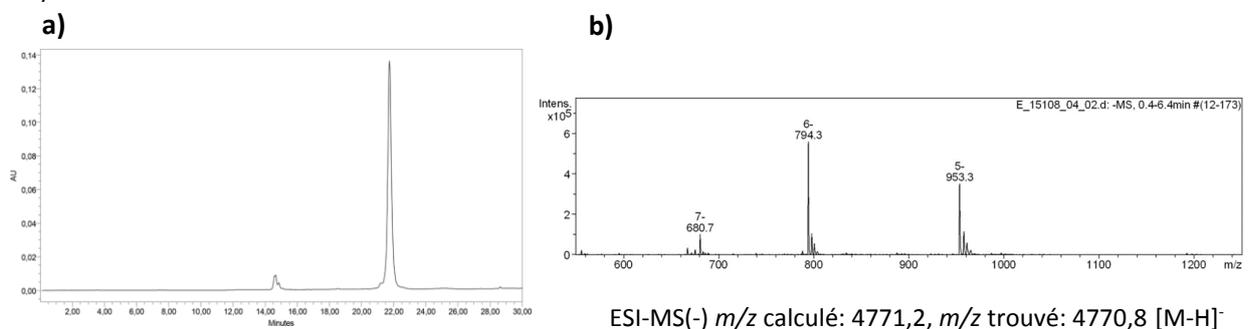


Figure 74 : (a) chromatogramme RP-HPLC et (b) spectre ESI-MS de l'édifice intermédiaire **G₁**

La deuxième étape consiste à immobiliser sur l'édifice intermédiaire **G₁** le deuxième brin d'ADN (l'oligonucléotide **13**) *via* un couplage thiol-chloroacétamide. Bien que cette réaction ait déjà été utilisée avec des biomolécules par différentes équipes (introduction p 33), nous avons dû optimiser les conditions réactionnelles. D'après les différents protocoles trouvés dans la littérature, la réaction est habituellement effectuée dans des mélanges solvants aqueux/solvants organiques avec un excès de 2 à 10 équivalents d'un réactif par rapport à l'autre. La réaction est également effectuée dans des conditions réductrices (tris (2-carboxyéthyl) phosphine (TCEP)) pour empêcher la formation d'espèces disulfures. L'iodure de potassium (KI) est fréquemment utilisé pour l'activation *in-situ* de la réaction, par formation transitoire d'un groupe réactif iodoacétamide. Enfin, la réaction est réalisée à pH 8,5 par addition d'amines tertiaires ou en présence d'un agent tampon organique. Une première série de tests reprenant les conditions les plus utilisées a été opérée et est présentée dans le tableau 2. Le rendement a été évalué par HPLC après 3h de réaction.

	Test 1	Test 2	Test 3	Test 4	Test 5
TCEP		2 eq	2 eq	10 eq	10 eq
Température	30°C	30°C	TA	30°C	TA
Rendement HPLC	4%	5,5%	7%	0%	0%

Tableau 2 : récapitulatif des premières conditions de réaction testées pour le couplage thiol-chloroacétamide, conditions : 1 eq de châssis **G₁**, 2 eq d'ADN **13**, 60 eq de DIEA, 20 eq de KI, H₂O/CH₃CN 90/10, concentration finale en **G₁** de 0,5 mM, 3 h de réaction

En absence de TCEP (test 1), nous obtenons un rendement très faible de 4% ainsi qu'une quantité élevée d'oligonucléotides sous forme de dimères (20%). Cette dimérisation est due à la formation de pont-disulfures entre deux brins d'ADN.

Nous avons ensuite ajouté un agent réducteur, le TCEP pour limiter la formation de pont-disulfures. Nous avons fait varier la quantité de TCEP ajouté, 2 ou 10 eq ainsi que la température de la réaction, température ambiante ou 30°C. Avec 10 eq de TCEP, que ce soit à 30°C (test 4) ou à température ambiante (test 5), aucun produit n'est formé, tandis qu'avec 2 eq de TCEP, 5,5% de produit est formé à 30°C (test 2) et 7% à température ambiante (test 3). Suite à ces résultats, le pH des différents milieux réactionnels a été mesuré, il est de 4 en présence de 10 eq de TCEP et de 8 avec 2 eq. A plus forte concentration, le TCEP acidifie donc le milieu réactionnel empêchant ainsi la réaction de couplage de se produire. En effet, le pKa des thiols aliphatiques étant aux alentours de 8 il est nécessaire que le pH de la réaction soit plus élevé pour qu'elle ait lieu. En utilisant 10 eq de TCEP, il faut en contrepartie augmenter la quantité de DIEA (200 eq au lieu de 60) pour obtenir un mélange réactionnel à pH 8,5.

L'analyse par spectrométrie de masse des produits obtenus lors des tests 2 et 4 révèle la présence de conjugués porteurs de fonctions amines issues de la réduction des fonctions azotures du châssis **G₁**. En revanche, l'analyse des produits obtenus lors des tests 3 et 5 n'a révélé la présence d'aucun produit secondaire. La formation de ce produit secondaire est sûrement due à une réaction de Staudinger entre les phosphines du TCEP et les fonctions azotures du châssis, qui semblerait être favorisée par l'augmentation de la température.

Pour résumer, la présence de 2 eq de TCEP empêche la formation de pont-disulfure mais il doit être utilisé à température ambiante, pour éviter tout risque de réaction secondaire avec les fonctions azotures du châssis peptidique. Il est également primordial que le pH du milieu réactionnel soit supérieur à 8 pour que la réaction de couplage ait lieu.

Pour justifier de la très faible réactivité observée, nous avons émis l'hypothèse de la présence de répulsions de charge entre l'oligonucléotide **13** et le brin d'ADN du conjugué **G₁**, ce qui pourrait contribuer à défavoriser la formation de l'édifice intermédiaire **G₂**. La répulsion de sondes d'ADN densément immobilisées a été étudiée et l'écrantage cationique a été identifié comme une approche efficace pour stabiliser ces oligonucléotides étroitement espacés¹⁴⁴. Nous avons donc exploré l'addition de grandes concentrations de cations métalliques ou organiques (tableau 3).

	Test 6	Test 7	Test 8	Test 9
Sels ajoutés	NaCl 500 mM	KCl 500 mM	LiCl 500 mM	NH ₄ Cl 500 mM
Concentration finale en G₁	0,5 mM	0,5 mM	0,5 mM	0,5 mM
Rendement HPLC	40%	55%	20%	5%

Tableau 3 : récapitulatif des secondes conditions de réaction testées pour le couplage thiol-chloroacétamide, conditions : 1 eq de châssis **G₁**, 2 eq d'ADN **13**, 2 eq de TCEP, 60 eq de DIEA, 20 eq de KI, H₂O/CH₃CN 90/10, 3h de réaction à température ambiante

Les différents tests réalisés confirment notre hypothèse. L'écrantage cationique a permis d'obtenir des rendements HPLC convenables en présence de 500 mM de NaCl (test 6) et de 500 mM de KCl (test 7). Les tests 8 et 9 ont quant à eux révélé une moins bonne neutralisation des charges par Li⁺ et NH₄⁺. Suite à ces résultats encourageants, les conditions utilisées pour le test 7 ont été sélectionnées et optimisées pour arriver à un rendement HPLC de 76%. Durant cette phase d'optimisation nous avons observé que l'addition de 4 eq d'ADN-SH **13** n'augmente pas le taux de conversion. Ce dernier a cependant été fortement augmenté par l'ajout d'une plus grande quantité de KI (60 eq). Nous avons également observé que l'ajout de 45 eq de DIEA est suffisant pour

maintenir le pH à 8,5. Nous avons aussi essayé d'optimiser la concentration en KCl, avec 100 mM nous avons obtenu un faible rendement HPLC de 20%, tandis que l'ajout de 1 M de KCl n'a rien changé par rapport à nos conditions initiales de 500 mM. Les conditions finales du couplage thiol-chloroacétamide sont donc les suivantes (tableau 4) :

	Conditions finales
Châssis G₁	1 eq
ADN-SH 13	2 eq
TCEP	2 eq
DIEA	45 eq
KI	60 eq
H ₂ O/CH ₃ CN	(90/10)
Sel ajoutés	KCl 500 mM
Concentration finale en G₁	0,5 mM
Température	TA
Temps de réaction	3h
Rendement HPLC	76%

Il est à noter que l'ordre d'ajout des réactifs est important. A cet effet, l'oligonucléotide **13** dissous dans le mélange H₂O/CH₃CN (90/10) est tout de suite mis en présence de TCEP pour éviter tout risque de dimérisation. A cette solution est ensuite ajouté dans l'ordre : le châssis peptidique, les 500 mM de KCl, les 45 eq de DIEA et enfin les 60 eq de KI. Le pH du milieu réactionnel est systématiquement contrôlé et doit être compris entre 8,5 et 9. Pour finir, la réaction est réalisée sous argon.

Tableau 4 : récapitulatif des conditions finales de réaction pour le couplage thiol-chloroacétamide

En utilisant ces conditions de couplage, l'édifice intermédiaire **G₂** est obtenu avec un rendement de 70% après purification par RP-HPLC. Les analyses par spectrométrie de masse et RP-HPLC ont confirmé la formation de l'édifice intermédiaire **G₂** (figure 75).

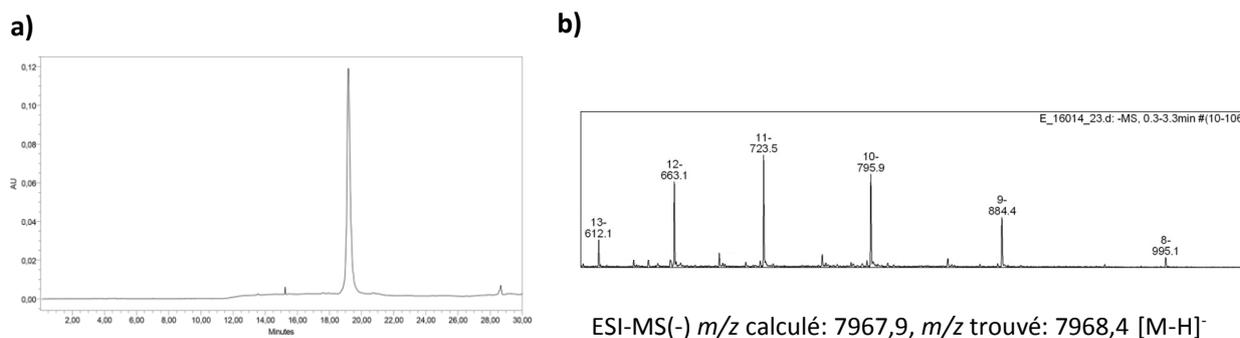


Figure 75 : (a) chromatogramme RP-HPLC et (b) spectre ESI-MS de l'édifice intermédiaire **G₂**

La dernière étape de synthèse permet de refermer simultanément les deux boucles en utilisant la CuAAC. Les oligonucléotides **12** et **13** sont respectivement liés par leur extrémités 5'alcyne et 3'alcyne sur l'intermédiaire **G₂** comportant les deux fonctions azotures complémentaires. Afin de favoriser la fermeture intramoléculaire la réaction est réalisée en milieu dilué (10^{-4} M), en présence de 100 mM d'HEPES à pH7,4, de CuSO₄ (6 eq), de THPTA (30 eq) et d'ascorbate de sodium (30 eq). Dans le cas de ce motif G-quadruplexe et contrairement aux édifices ARN, nous ajoutons 100

mM de KCl dans le mélange réactionnel. Cet ajout d'ion K^+ permet de préformer le motif G-quadruplexe en conformation antiparallèle afin de favoriser la formation de l'édifice désiré. Des cations stabilisants ont déjà été utilisés avec succès il y a quelques années dans la synthèse d'un édifice similaire¹¹⁹. Il a été mis en évidence qu'en préformant le motif G-quadruplexe, ils permettent le maintien des fonctions azotures et alcynes proches les unes des autres, abaissant ainsi le coût entropique et empêchant la formation de produits ramifiés. La stabilisation par K^+ du G4 HIV-PRO1 en conformation antiparallèle, a été mis en évidence dans l'étude de S. Amrane¹⁴³. Le spectre CD en figure 76.a confirme qu'en présence de K^+ , l'intermédiaire **G₂** se structure en G4 antiparallèle stable ($T_m = 53^\circ C$).

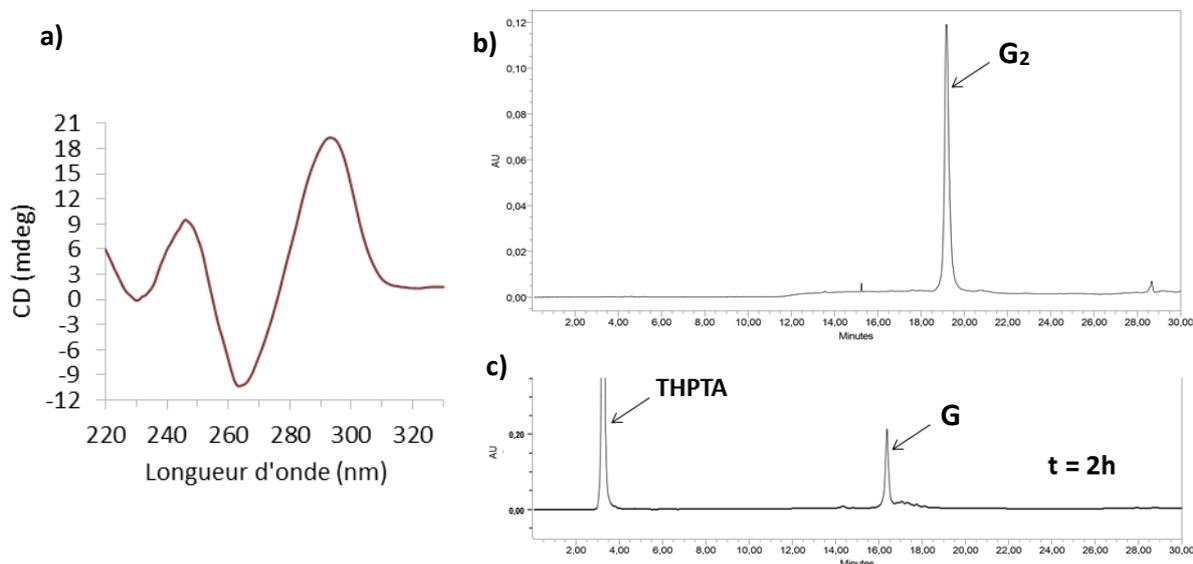


Figure 76 : (a) spectre CD de l'intermédiaire **G₂** en présence de Tris 10 mM pH 7,4 et KCl 100 mM, (b) chromatogramme RP-HPLC de l'intermédiaire **G₂** et (c) chromatogramme RP-HPLC du brut réactionnel de l'édifice **G**

Ainsi, dans ces conditions de CuAAC, la formation de l'édifice **G** est totale après 2 h de réaction, comme en témoigne les chromatogrammes HPLC. La conversion de l'intermédiaire **G₂** (figure 76.b) en édifice **G** (figure 76.c) est confirmée par l'apparition d'un pic ayant un temps de rétention inférieur à celui de l'intermédiaire **G₂**. L'édifice HIV-PRO1 est obtenu avec un rendement de 79%, après élimination du cuivre par colonne d'exclusion de taille et purification par RP-HPLC. Il a ensuite été caractérisé par RP-HPLC et spectrométrie de masse (figure 77).

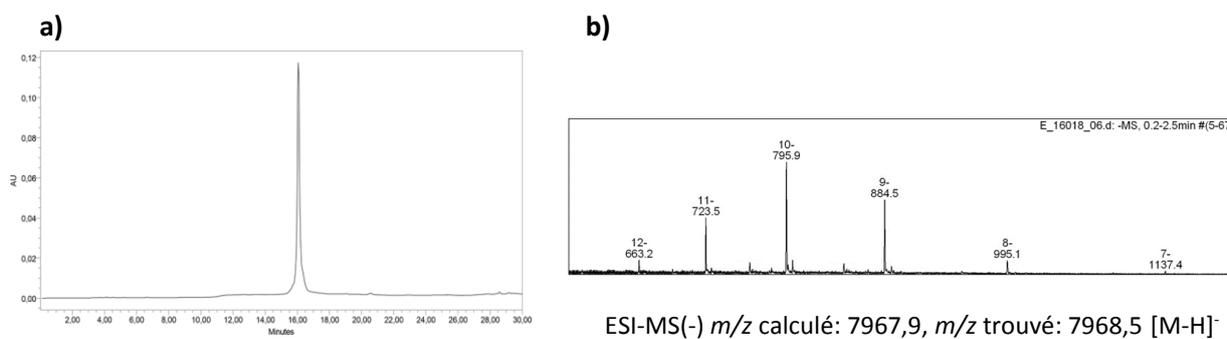


Figure 77 : chromatogramme RP-HPLC (a) et spectre ESI-MS (b) de l'édifice HIV-PRO1 **G**

III. Etudes de la topologie de l'édifice **G** mimant le motif G-quadruplexe de la séquence HIV-PRO1

Tout comme pour les édifices ARN, nous avons utilisé trois techniques d'analyses pour valider la formation du motif G-quadruplexe ainsi que pour déterminer sa topologie et sa stabilité. Il s'agit de la RMN, du dichroïsme circulaire et de l'UV-visible. Deux contrôles ont été utilisés pour ces analyses, l'intermédiaire **G₂** ainsi que la séquence non contrainte HIV-PRO1.

III.1. Etudes d'absorption UV : "thermal difference spectrum" (TDS)

Le principe de cette méthode a été décrit dans la première partie de ce chapitre (p 70) . Dans le cas de l'édifice HIV-PRO1, les absorbances sont mesurées entre 225 et 320 nm à 5 et 90°C. Les données ont été collectées pour l'édifice final **G** (figure 78.a), pour l'intermédiaire **G₂** (figure 78.b) ainsi que pour la séquence non contrainte HIV-PRO1 (figure 78.c), dans du tampon Tris à 10 mM à pH 7,4 en présence de 100 mM NaCl ou 100 mM KCl. En présence de 100 mM KCl, les spectres obtenus pour les trois composés confirment leur structuration en motifs G-quadruplexes. En effet, ils présentent tous le signal caractéristique d'un G-quadruplexe avec un maximum aux alentours de 273 nm et un minimum proche de 295 nm. En revanche, en présence de 100 mM de NaCl, il n'y a que l'intermédiaire **G₂** et l'édifice **G** qui conservent des signaux caractéristiques d'une topologie G-quadruplexe. En absence de contrainte apportée par le châssis peptidique, la séquence libre ne se structure pas en G4.

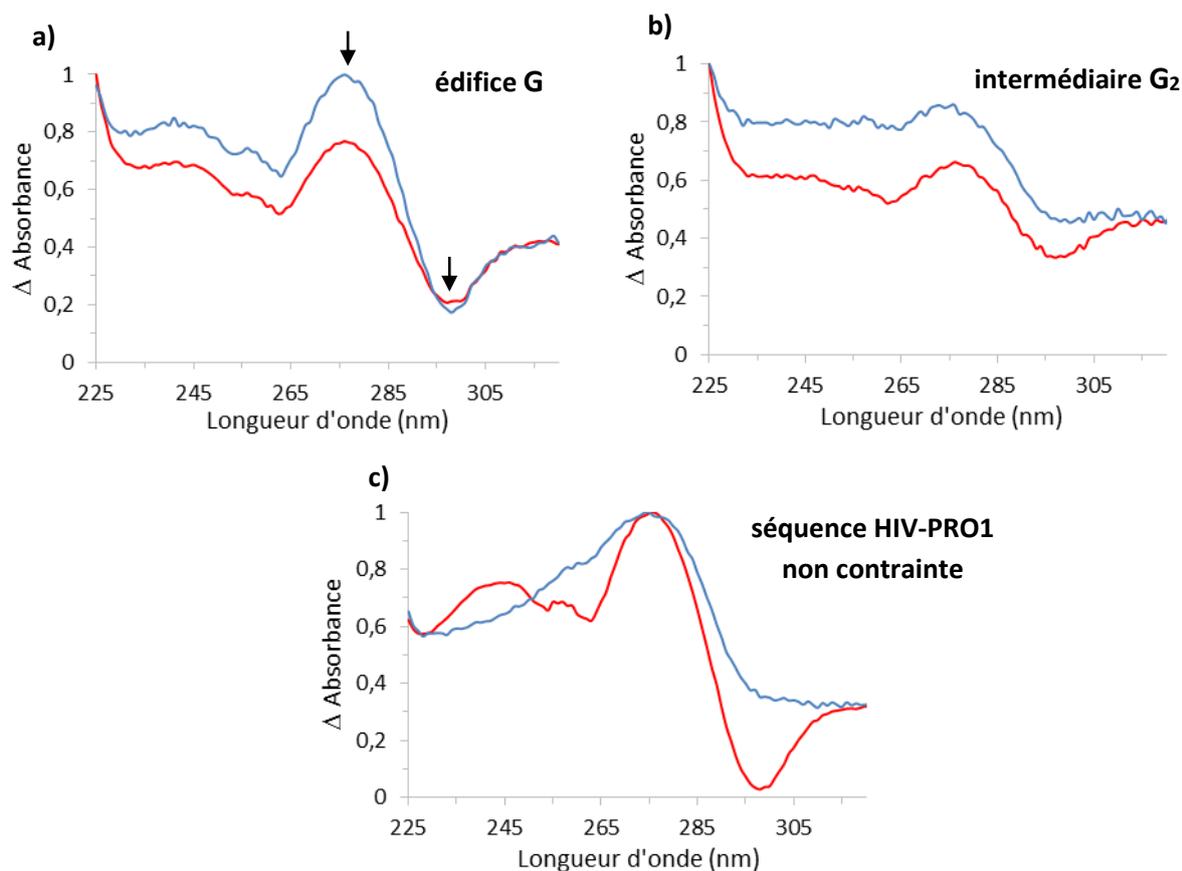


Figure 78 : spectres normalisés représentant la différence d'absorbance entre 90 et 5°C pour les édifices (a) **G** et (b) **G₂**, et (c) pour la séquence non contrainte HIV-PRO1, conditions : 1,6 μM d'édifices **G** ou **G₂** ou de séquence non contrainte, dans du tampon Tris-HCl 10 mM pH 7,4 en présence de 100 mM NaCl (bleu) ou de 100 mM KCl (rouge)

III.2. Etudes de dichroïsme circulaire

III.2.a. Etude de la topologie des motifs G-quadruplexes

Les topologies des motifs G-quadruplexes de l'édifice **G**, de l'intermédiaire **G₂** et de la séquence non contrainte ont été déterminées en enregistrant leurs spectres CD à 20°C. Ces mesures ont été réalisées dans un tampon Tris-HCl 10 mM à pH 7,4 en présence de KCl 100 mM ou de NaCl 100 mM. Avec K⁺ comme cation stabilisant (figure 79.a), les motifs G-quadruplexes des édifices **G** et **G₂** sont structurés en conformation antiparallèle (maxima à 240 et 290 nm et minimum à 260 nm) tout comme la séquence HIV-PRO1 libre. Nous avons donc réussi à reproduire un motif G-quadruplexe présentant une topologie identique à celle du G-quadruplexe naturel de la séquence HIV-PRO1.

Le spectre CD de la séquence non contrainte enregistré dans NaCl 100 mM (figure 79.b) confirme les données TDS en mettant en évidence sa non-structuration en G-quadruplexe. En

revanche, les édifices **G₂** et **G** sont structurés en G-quadruplexes et la topologie adoptée est la même qu'en présence de KCl (figure 79.b).

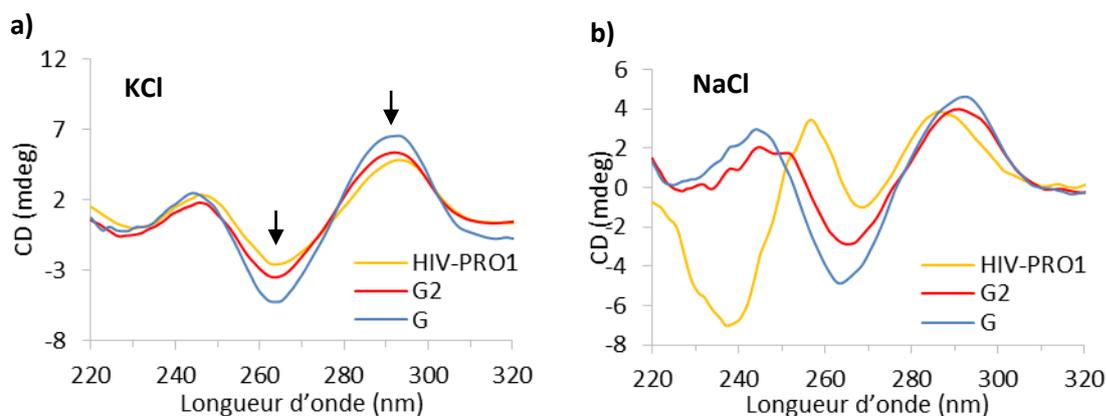


Figure 79 : spectres CD à 20°C de la séquence HIV-PRO1 (5′TGGCCTGGGCGGGACTGGG3′), conditions : 1,6 μM d’édifices **G** ou **G₂** ou de séquences non contrainte, dans 10 mM de tampon Tris-HCl en présence (a) de 100 mM KCl ou (b) de 100 mM NaCl

III.2.b. Etude de la stabilité des motifs G-quadruplexes

La stabilité du motif G-quadruplexe de l’édifice **G** a été évaluée en mesurant sa température de dénaturation de 5 à 90°C en présence de cations métalliques (100 mM NaCl ou 100 mM KCl) (tableau 5). A titre de comparaison, les stabilités des motifs G-quadruplexe de l’intermédiaire **G₂** et de la séquence HIV-PRO1 non contrainte ont également été mesurées.

<p>Motifs G4</p> <p>Conditions: Tris 10 mM pH 7,4</p>	<p>Séquence HIV-PRO1 non contrainte</p>			
	<p>KCl 100 mM</p>	<p>56°C</p>	<p>53°C</p>	<p>76°C</p>
	<p>NaCl 100 mM</p>	<p></p>	<p>51°C</p>	<p>67°C</p>

Tableau 5 : température de dénaturation (*T_m*) de la séquence HIV-PRO1 (5′TGGCCTGGGCGGGACTGGG3′), conditions : 1,6 μM des édifices **G** ou **G₂** ou de la séquence non contrainte dans 10 mM de tampon Tris-HCl pH 7,4 en présence de 100 mM KCl ou de 100 mM NaCl, suivi à 260 nm

Que ce soit en présence de sodium ou de potassium, la contrainte imposée par le châssis peptidique augmente considérablement la stabilité du motif G-quadruplexe. En présence de KCl, le motif G-quadruplexe contraint est 20°C plus stable que la séquence naturelle, tandis qu’avec NaCl la

contrainte apportée par le châssis permet de former un motif G4 stable ($T_m = 67^\circ\text{C}$). Il est également intéressant de noter qu'en présence de KCl, seule la contrainte totale des oligonucléotides influence la stabilisation du système. En effet, avec des T_m très proches, respectivement 53°C et 56°C , le motif G4 de l'intermédiaire **G**₂ n'est pas plus stable que celui de la séquence naturelle.

Tout comme il est important d'évaluer la stabilité du motif G4, il est également primordial de confirmer sa structuration intramoléculaire et non intermoléculaire. Dans ce dernier cas, il s'agirait de l'autoassemblage de deux édifices **G**. Afin d'écarter cette hypothèse, des T_m ont été mesurés à trois concentrations différentes : 0,8, 1,6 et 3,2 μM et les valeurs obtenues (figure 80) à savoir respectivement 77, 76 et 76°C confirme la formation d'un motif intramoléculaire.

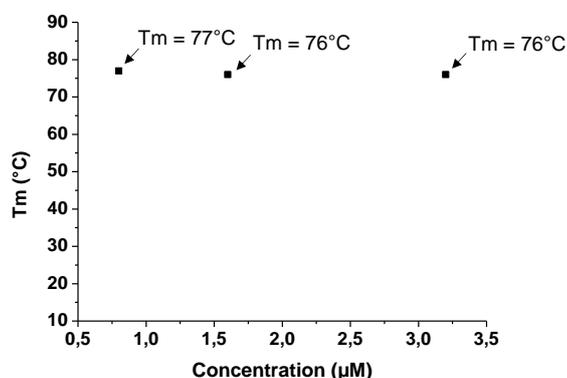


Figure 80 : température de dénaturation (T_m) de l'édifice **G** dans Tris-HCl 10 mM et KCl 100 mM à 0,8, 1,6 et 3,2 μM

III.3. Etudes de résonance magnétique nucléaire

Des analyses RMN ont été réalisées sur le composé **G** en présence de sodium et de potassium. Les spectres obtenus témoignent de la formation d'un motif G-quadruplexe dans les deux conditions salines. La présence de signaux correspondant aux protons imino aux alentours des 10-12 ppm confirme la formation du motif G-quadruplexe. En présence de KCl (figure 81.b), un large pic est observé aux alentours de 13,7 ppm, il révèle la présence d'une paire de bases de type Watson-Crick supplémentaire. Ceci est cohérent avec la formation de la paire de bases CG dans la boucle la plus longue (boucle $5'\text{CCTG}3'$ porté par l'oligonucléotide **13**) de HIV-PRO1, décrit par Samir Amrane¹⁴³.

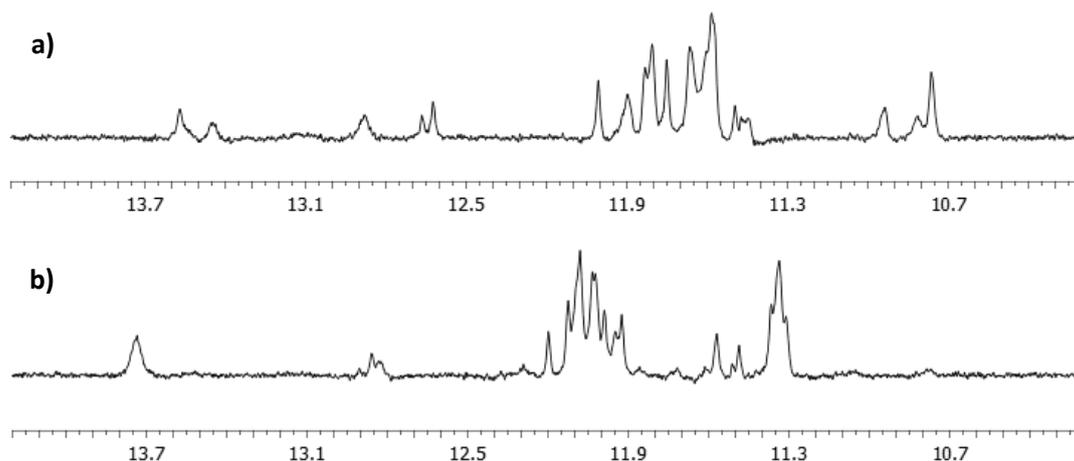


Figure 81: spectre RMN ^1H (500 Mz) à 25°C, représentant la région des protons imino de l'édifice HIV-PRO1 à 80 μM en présence (a) de 20 mM de tampon phosphate sodium à pH 7,5, de 70 mM de NaCl et de 10 % de D_2O ou (b) de 20 mM de tampon phosphate potassium à pH 7,5, de 70 mM de KCl et de 10 % de D_2O

IV. Bilan

Dans ces travaux qui ont fait l'objet d'une publication parue en 2017¹⁴⁵ (L. Bonnat, et *al.* *Chem Eur. J.* **2017**, *23*, 5602-5613), nous avons réussi à mimer et stabiliser le motif G-quadruplexe antiparallèle de la séquence HIV-PRO1. La synthèse de ce bioconjugué a permis de mettre au point au laboratoire une méthodologie permettant de réaliser trois ligations d'oligonucléotides sur un même châssis peptidique. Les études de dichroïsme circulaire effectuées à partir de cet édifice ont mis en évidence la formation d'un motif G-quadruplexe adoptant la même structuration antiparallèle que la séquence naturelle, en présence de potassium. Grâce à ces systèmes biomoléculaires, nous avons pu stabiliser le motif G4 en présence de sodium, ce qui n'est pas réalisable pour la séquence naturelle non contrainte et qui pourrait être intéressant au vu d'éventuelles études biophysiques. Ce G-quadruplexe issue d'une région régulatrice pourrait être une cible thérapeutique intéressante, l'édifice **G** représente donc un mime très intéressant pour l'identification et la conception de ligands visant à inhiber l'activité virale. Par conséquent, nous l'avons utilisé dans des études SPR visant à évaluer l'affinité et la spécificité de différents ligands pour ce motif G-quadruplexe. Les résultats sont décrits dans la première partie du chapitre 2, p 129.

Partie III : Mime d'ADN triplexe de la région télomérique

I. Introduction

Comme évoqué en introduction p10, des études suggèrent que la formation des G-quadruplexes passe par un état intermédiaire appelé G-triplexe. L'étude de cette structure, formée par l'empilement de triades de guanines, pourrait nous aider à comprendre les différents aspects de la formation des G-quadruplexes. L'identification de ligands capables de reconnaître et d'interagir spécifiquement avec ces structures pourrait permettre d'étudier leurs rôles dans la structuration des G-quadruplexes ainsi que leur pertinence comme cible thérapeutique. Il serait donc intéressant d'identifier des ligands spécifiques à ces structures, ou du moins présentant une différence d'affinité significative entre les G-quadruplexes et les G-triplexes. Il est cependant difficile d'isoler les structures G-triplexes, ces dernières étant moins stables que les motifs G-quadruplexes. Les premières études ont été basées sur des prédictions dynamiques pour valider la possibilité de former ces structures intermédiaires. V. Limongelli²⁷ et ses collaborateurs, ont visualisé en 2013 un intermédiaire G-triplexe lors du repliement en G-quadruplexe de l'aptamère à thrombine (TBA), par simulation métadynamique. Un an plus tard, en utilisant la séquence 11-mer-3'-t-TBA, cette même équipe a confirmé la formation de ce G-triplexe en solution, en présence de potassium¹⁴⁶.

La même année, l'équipe d'H. Sugiyama²⁸ a utilisé une méthode mettant en jeu des origamis d'ADN pour évaluer la possibilité de former des G-triplexes à partir d'une séquence riche en guanine. A l'aide de l'imagerie AFM, ils ont réussi à visualiser des motifs G-triplexes conduisant à la formation de G-quadruplexes. Deux ans plus tard, toujours avec l'utilisation d'origami d'ADN pour contrôler la stœchiométrie, le nombre de répétition de guanines ou encore le nombre de brins d'ADN, ils ont réalisé la première étude de ciblage de G-triplexe avec un ligand de G4¹⁴⁷. Cette étude révèle que le ligand, en l'occurrence le ligand PDC, se lie initialement à l'intermédiaire G-triplexe et induit un repliement en motif G-quadruplexe.

Très récemment, l'équipe d'A. Randazzo¹⁴⁸ a publié des travaux qui offrent de nouvelles perspectives dans l'étude des G-triplexes, avec l'identification du premier ligand possédant une certaine spécificité pour les G-triplexes. Ils ont combiné le criblage virtuel de plus de 58 000 petites molécules avec des analyses expérimentales visant à évaluer l'affinité des potentiels ligands avec différentes séquences de G-triplexes et de G-quadruplexes. Le motif triplexe choisi dérive de la séquence 11-mer-TBA déjà étudiée et les motifs G-quadruplexes proviennent des séquences *c-kit* et *c-myc* de promoteurs d'oncogènes, connus pour former des G-quadruplexes. Ils ont trouvé une

molécule (figure 82) pouvant stabiliser le motif G-triplexe sans changer sa conformation. Toutefois, il s'avère que ce ligand se lie et stabilise également les deux motifs G-quadruplexes contrôlés.

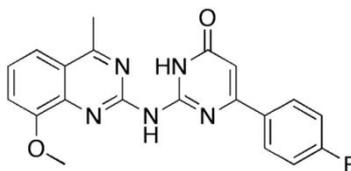


Figure 82 : structure chimique du ligand G3/G4, source : Amato, J. *et al. Biochim. Biophys. Acta - Gen. Subj.* **1861**, 1271 - 1280 (2017)¹⁴⁸

Afin d'étudier la formation des motifs G-triplexes et leur ciblage, nous avons choisi d'assembler un nouveau bioconjugué mimant un intermédiaire G-triplexe du motif G-quadruplexe de l'ADN télomérique humain. La contrainte imposée par le châssis peptidique devrait favoriser la stabilisation du motif G-triplexe et ainsi faciliter les études de sélection et d'identification de ligands. Cette cible a fait l'objet d'une récente étude menée par l'équipe de H. Mao²⁹, dans laquelle ils ont étudié les conditions de stabilisation du G-triplexe. Les analyses de dichroïsme circulaire ont mis en évidence les topologies adoptées par le G-triplexe en présence de différents cations métalliques et pourraient ainsi nous servir d'élément de comparaison pour valider la structuration de notre motif G-triplexe.

II. Synthèse et caractérisation d'édifices G-triplexes dérivés de l'ADN télomérique

Nous avons décidé de nous inspirer de la conformation « hybride » de l'ADN télomérique humain (figure 83.a) afin de préparer deux conjugués susceptibles de former une structure triplex stable ou transitoire. En mimant la structure adoptée par la boucle supérieure (coloré en violet sur la figure 83.a) et le troisième brin d'ADN (coloré en rose sur la figure 83.a) nous avons souhaité assembler un motif G-triplexe stable (figure 83.b). Afin d'étudier l'équilibre entre les formes G-triplexe et G-quadruplexe, l'édifice **I** (figure 83.c) porteur d'un troisième brin plus long que l'édifice **H** a été synthétisé.

L'édifice **H** (figure 83.b) comporte donc un brin court de séquence 5'TAGGGT^{3'} qui devrait contraindre le conjugué à s'associer en G-triplexe. L'édifice **I** (figure 83.c) comporte quant à lui un troisième brin plus long comportant deux répétitions du motif AGGGT (séquence : 5'AGGGTTAGGGT^{3'}) pour permettre une éventuelle dynamique entre le G-triplexe et le G-quadruplexe. Pour assembler ces deux bioconjugués, les trois ligations utilisées précédemment ont de nouveaux été exploitées : la

ligation oxime et la CuAAC, qui permettent de greffer la boucle supérieure du motif et le couplage thiol-chloroacétamide, employé pour greffer les brins d'ADN de longueur variable.

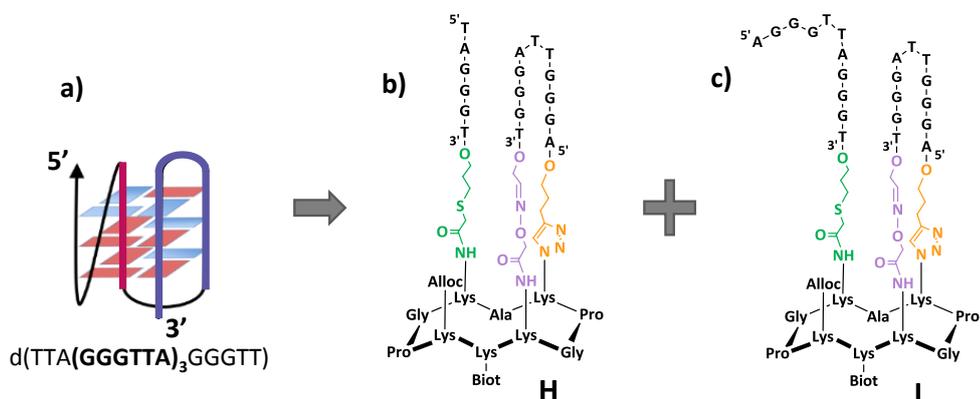


Figure 83 : (a) motif G4 hybride de l'ADN télomérique, (b) édifices **H** et (c) **I**

II.1. Synthèse du châssis peptidique

Pour pouvoir réaliser le greffage des oligonucléotides en utilisant trois ligations chimiosélectives orthogonales, le châssis peptidique doit être trifonctionnalisé (figure 84). Ce dernier comporte donc une fonction azoture, une fonction chloroacétamide et une fonction oxyamine. Dans ces travaux, et contrairement aux précédents, cette dernière fonction a été incorporée sous forme de « building block » durant la synthèse supportée. La fonction biotine est présente sur la face inférieure du châssis. La quatrième lysine du châssis est protégée par un groupement protecteur alloc, ce dernier est inerte vis-à-vis des trois ligations utilisées.

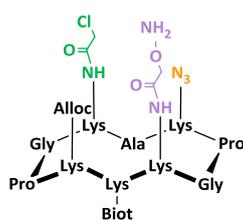


Figure 84 : Structure chimique du châssis peptidique **14**

II.1.a. Synthèse et insertion de la fonction oxyamine

Pour la synthèse du châssis peptidique **14**, nous avons utilisé le « building block » **15**, pour introduire les fonctions oxyamines durant la SPPS¹⁴⁹. Les fonctions sont ainsi protégées par un groupement 1-éthoxyéthylidène (Eei) qui bloque totalement la réactivité des oxyamines et qui est parfaitement compatible avec les conditions de couplage et de déprotection de la SPPS. De plus, ce groupement présente l'avantage de pouvoir être éliminé dans des conditions acides plus douces que le groupement Boc. Il est également à noter que l'utilisation de « building block » réduit le nombre

d'étapes de fonctionnalisation après la SPPS, augmentant ainsi le rendement global final. Le « building block » est obtenu par synthèse du groupement protecteur *via* la substitution nucléophile de l'éthyl-N-hydroxyacétimide sur l'acide iodoacétique, puis par activation de la fonction acide carboxylique du composé **15a** par une fonction *N*-hydroxysuccinimide. Pour finir, le groupement protecteur activé a été ajouté sur la fonction amine ε d'une Fmoc-Lys-OH (schéma 26).

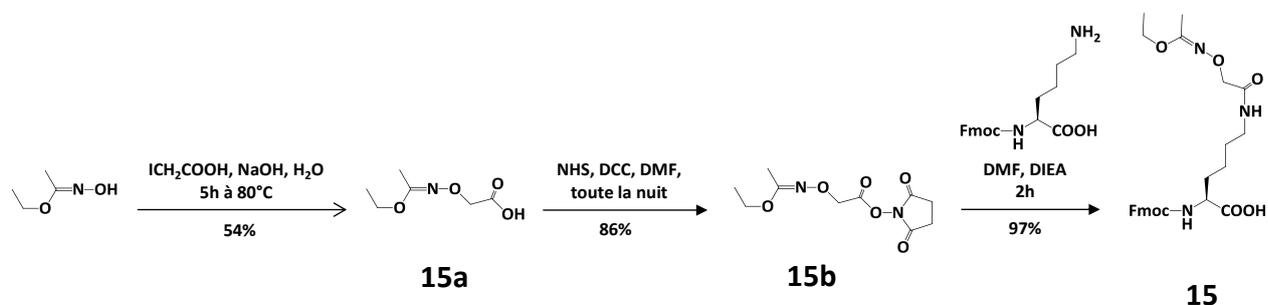


Schéma 26: Synthèse de la Fmoc-Lys(Eei-Aoa)-OH **15**

II.1.b. Description de la synthèse du châssis peptidique

Le châssis peptidique est synthétisé sur support solide en utilisant la stratégie de synthèse au Fmoc/tBu et la résine 2-chlorotrityl, en utilisant le même protocole qu'auparavant (p 77).

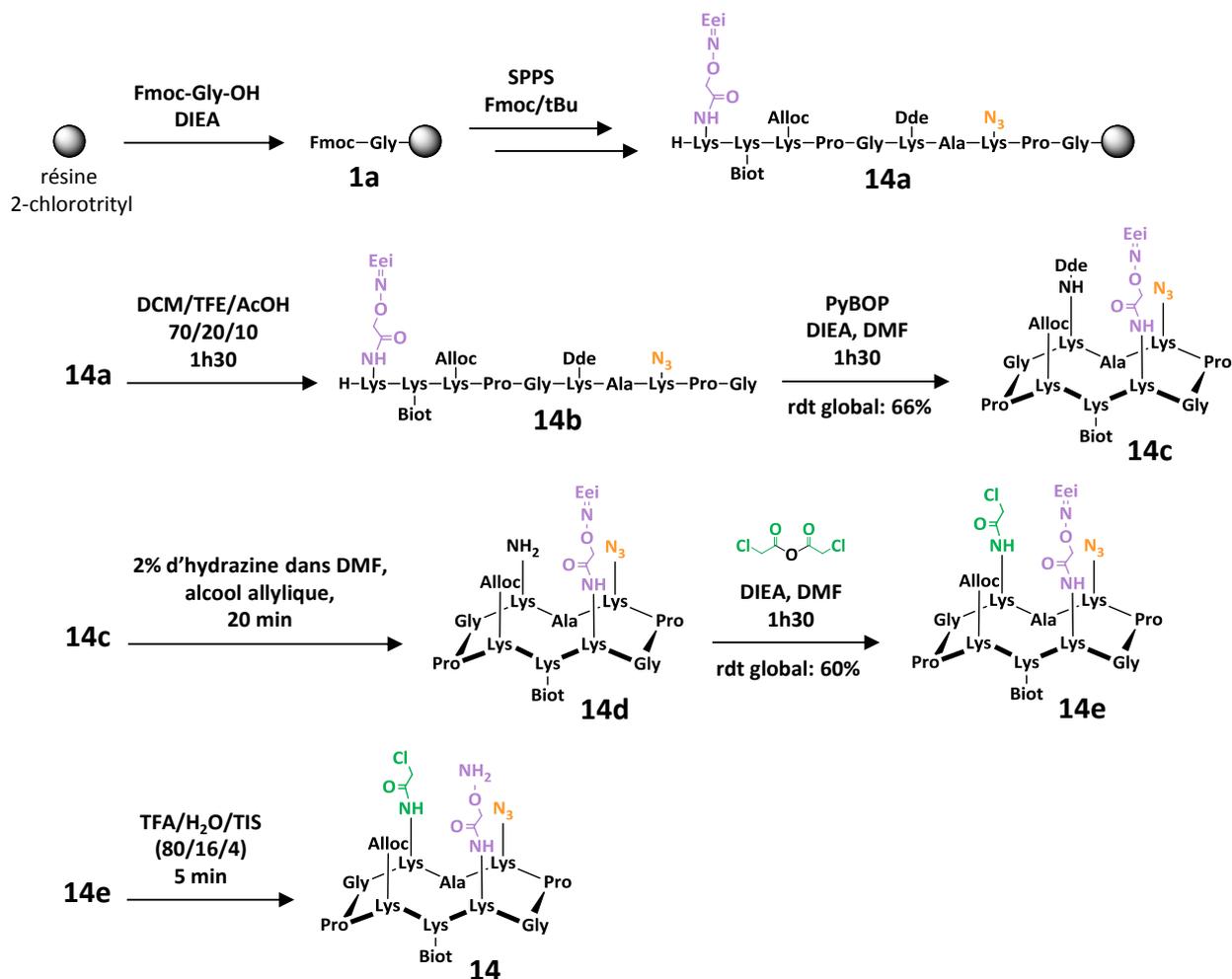
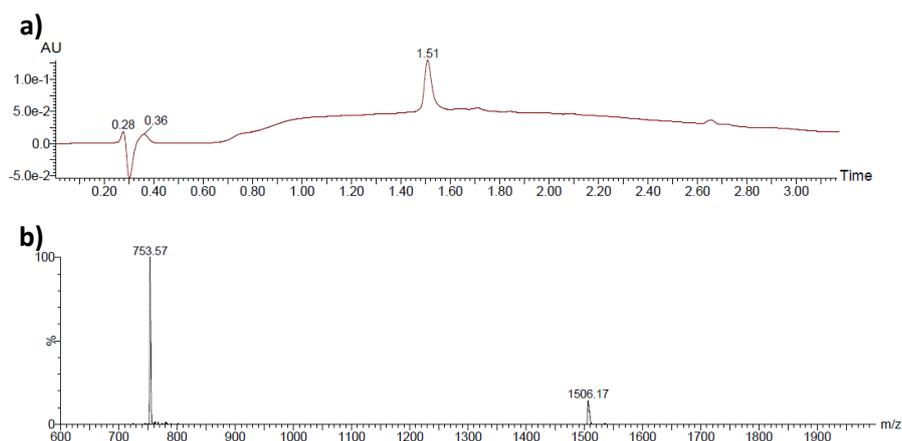


Schéma 27 : Synthèse du châssis peptidique **14**

Après la SPPS, le peptide linéaire est libéré du support en milieu acide faible par une solution DCM/TFE/AcOH (70/20/10). Il est important de noter que le groupement protecteur Eei est très sensible au TFA. Les conditions de coupure des précédents peptides, à savoir 1% de TFA dans le dichlorométhane ne peuvent donc pas être utilisées. Après cyclisation, le groupement protecteur Dde est éliminé en présence d'une solution d'hydrazine à 2% et d'alcool allylique. Ce dernier permet d'éviter la réduction de la double liaison du groupe Alloc. L'intermédiaire **14e** est ensuite obtenu avec un rendement global de 60% après ajout de la fonction chloroacétamide dans les mêmes conditions que décrit précédemment (p 82). Pour éviter la coupure du groupement protecteur Eei lors de la purification, le TFA est remplacé par un acide moins fort, l'acide formique. Le châssis trifonctionnalisable est obtenu après libération de la fonction oxyamine en milieu acide. Le peptide final a été caractérisé par UPLC-MS (figure 85).



ESI-MS(+) m/z calculé: 1507,1, m/z trouvé: 1506,2 $[M+H]^+$

Figure 85 : (a) chromatogramme UPLC et (b) spectre ESI-MS du châssis peptidique **14**

II.2. Synthèse des oligonucléotides

Le brin d'ADN formant la boucle est bi-fonctionnalisé, il présente une fonction aldéhyde à son extrémité 3' et une fonction alcyne à son extrémité 5' ($5'YAGGGTTAGGGTX^3'$, Y = 5'alcyne, X = 3' aldéhyde). Les troisièmes brins sont tous deux fonctionnalisés en 3' par un groupement thiol (séquence du brin court : $5'TAGGGTZ^3'$, séquence du brin long : $5'AGGGTTAGGGTZ^3'$, Z = 3' thiol).

II.2.a. Synthèse de l'oligonucléotide boucle 5'alcyne-3'aldéhyde **16**

L'oligonucléotide **16** a été préparé selon le même protocole qu'auparavant (p 79) et obtenu avec un rendement de 45% (schéma 28). Il a été caractérisé par RP-HPLC et spectrométrie de masse (figure 86).

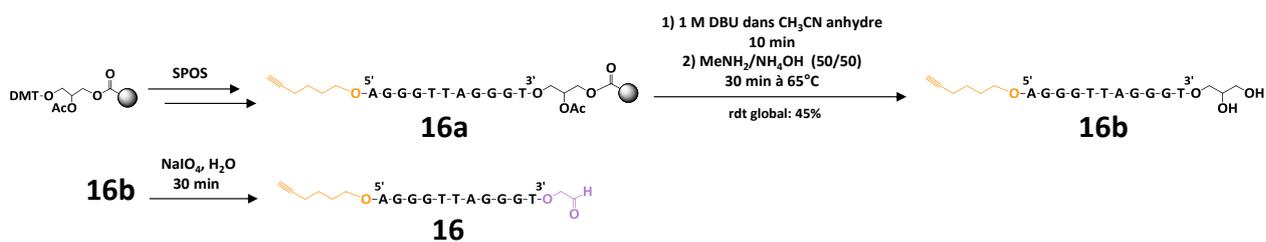


Schéma 28 : Synthèse de l'ADN 5'alcyne - 3'aldéhyde **16**

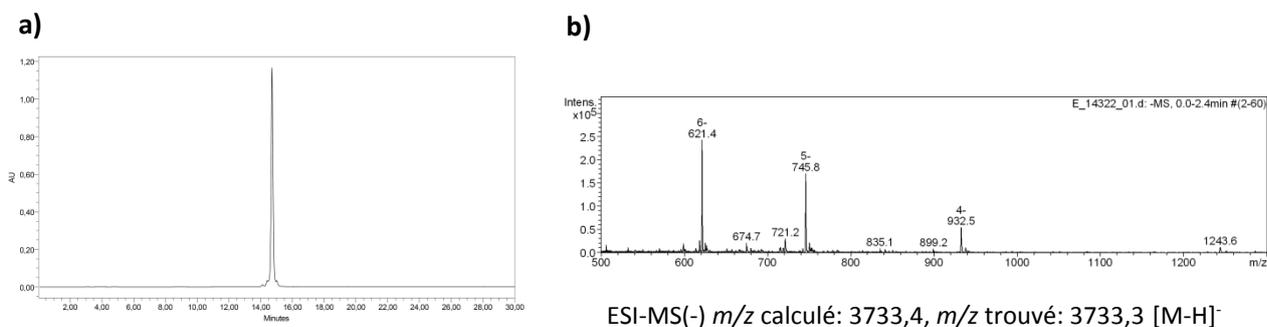


Figure 86 : (a) chromatogramme RP-HPLC et (b) spectre ESI-MS de l'ADN 5'alcyne - 3'aldehyde **16**

II.2.b. Synthèse des oligonucléotides 3'-5' court et long

Les troisièmes brins, court et long sont tous deux synthétisés sur support solide en utilisant une résine 3' thiol (« 3'-thiol-modifier C3 S-S CPG ») (schéma 29). Sa coupure en milieu basique, permet la libération en 3' d'une fonction thiol protégée par un pont disulfure. Les oligonucléotides **17** et **18** ont été respectivement obtenus avec des rendements de 74% et 72% après réduction du pont disulfure par une solution aqueuse de DTT (100 mM) dans 1M de tampon Tris pH 8,5 et purification par RP-HPLC.

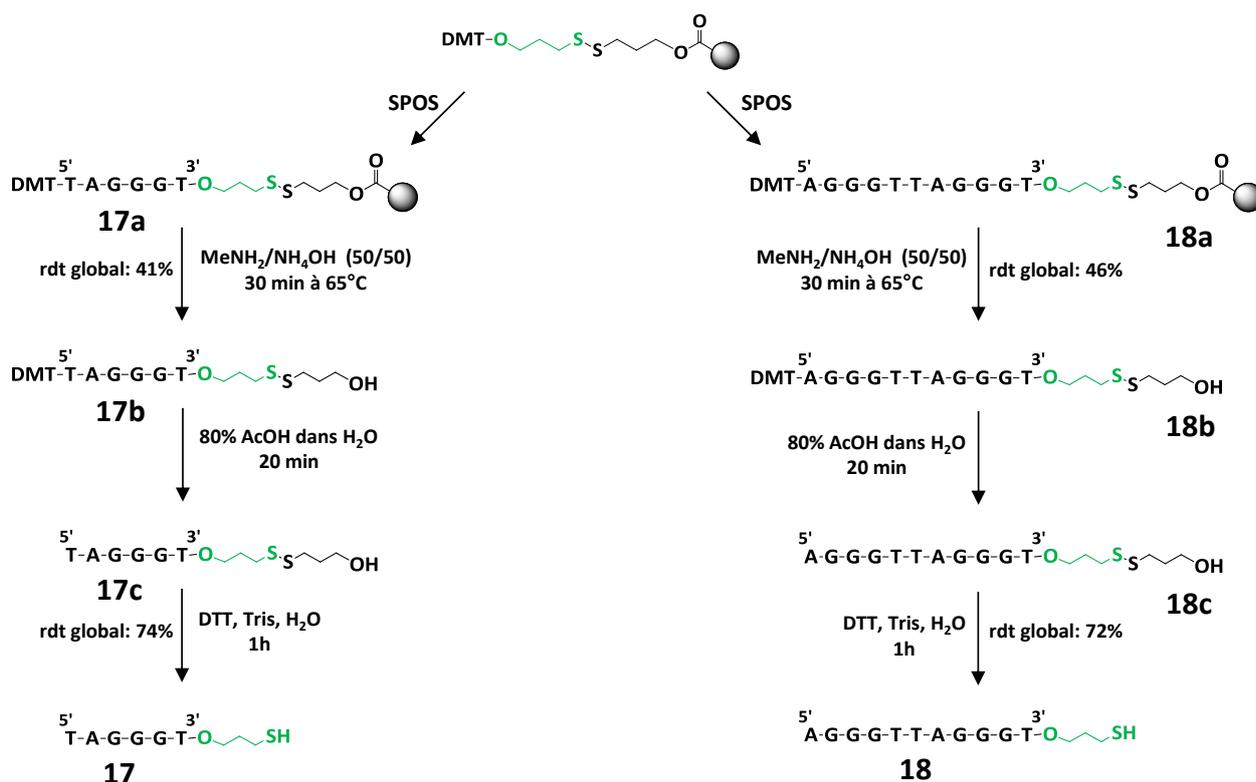
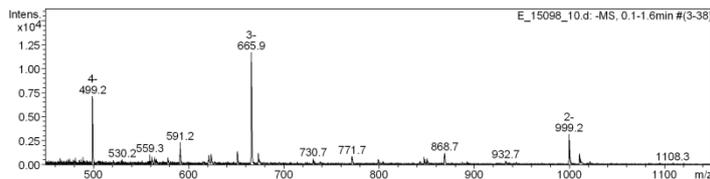
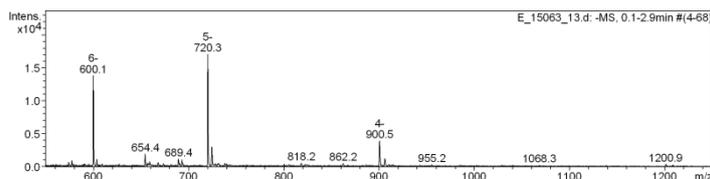


Schéma 29 : Synthèse des brins d'ADN 3'thiol **17** et **18**



ESI-MS(-) m/z calculé: 2000,4, m/z trouvé: 1999,9 [M-H]⁻

Figure 87 : spectre ESI-MS de l'ADN court 3'thiol **17**



ESI-MS(-) m/z calculé: 3605,4, m/z trouvé: 3605,4 [M-H]⁻

Figure 88 : spectre ESI-MS de l'ADN long 3'thiol **18**

II.3. Assemblage des édifices

L'assemblage des édifices **H** et **I** présenté dans le schéma 30 est réalisé en trois étapes.

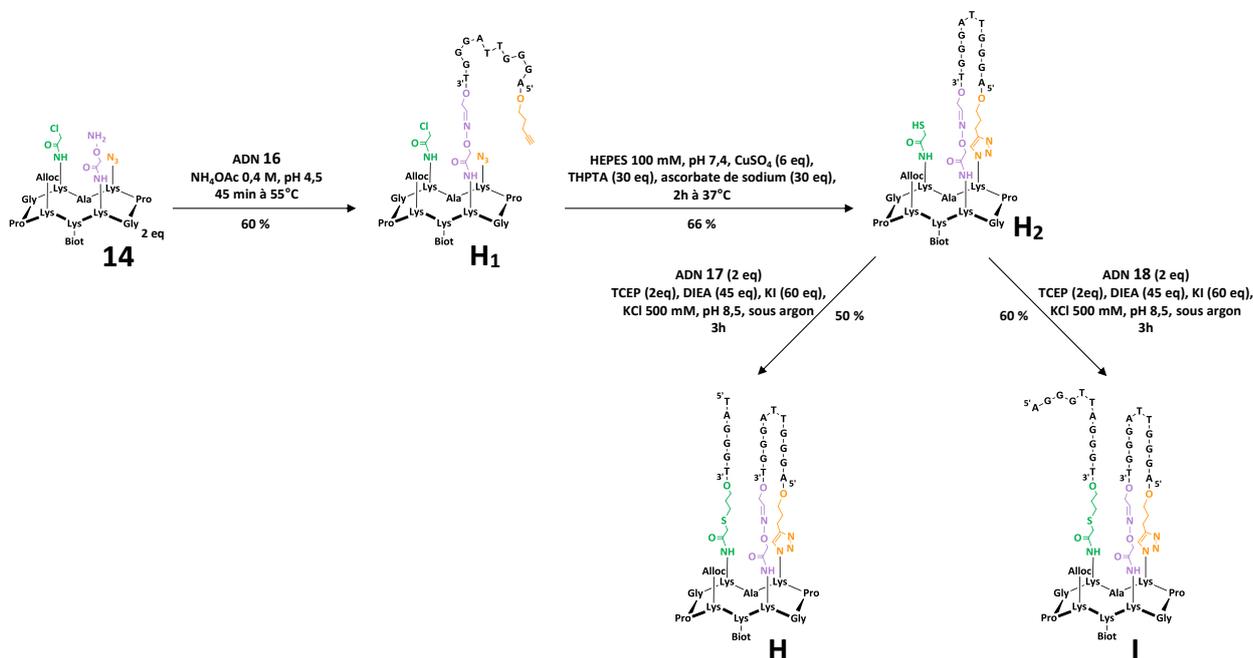


Schéma 30: Synthèse des édifices **H** et **I**

Les deux premières étapes de synthèse (schéma 30) sont communes aux deux édifices puisqu'elles permettent de greffer l'oligonucléotide qui forme la boucle supérieure du motif. Celui-ci est d'abord lié sur le châssis peptidique par son extrémité 3', en utilisant la ligation oxime. La réaction est réalisée avec 2 eq de peptide par rapport à l'oligonucléotide, afin de consommer

entièrement le brin d'ADN. Après 45 min de réaction à 55°C dans une solution tamponnée d'acétate d'ammonium 0,4 M à pH 4,5 et purification, l'intermédiaire **H₁** est obtenu avec un rendement de 60%. Il a ensuite été caractérisé par RP-HPLC et spectrométrie de masse (figure 89).

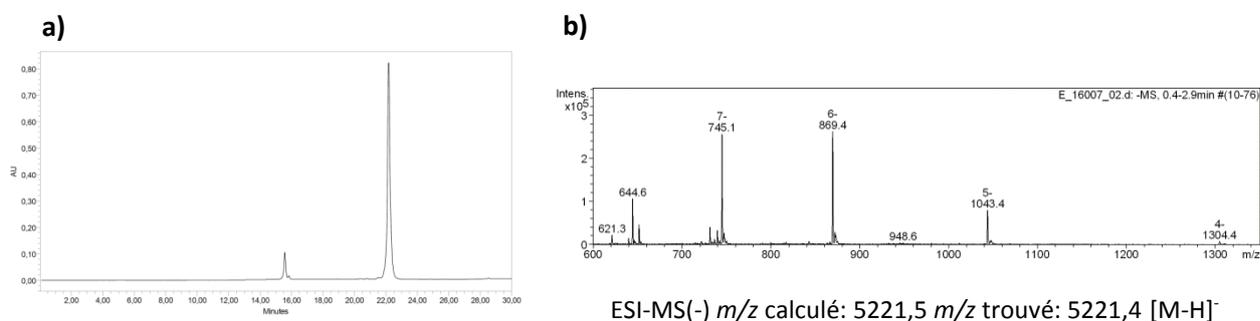


Figure 89 : (a) chromatogramme RP-HPLC et (b) spectre ESI-MS de l'édifice intermédiaire **H₁**

Pour former la boucle supérieure, la réaction CuAAC est ensuite réalisée en présence d'HEPES 100 mM à pH 7,4, de CuSO₄ (6 eq), de THPTA (30 eq) et d'ascorbate de sodium (30 eq) pendant 2 h à 37°C. En fin de réaction, le cuivre est chélaté par de l'EDTA (50 eq) puis éliminé sur colonne d'exclusion de taille. Le brut réactionnel est ensuite purifié par RP-HPLC et l'édifice intermédiaire **H₂** est obtenu avec un rendement de 66%. Il a ensuite été caractérisé par RP-HPLC et spectrométrie de masse (figure 90).

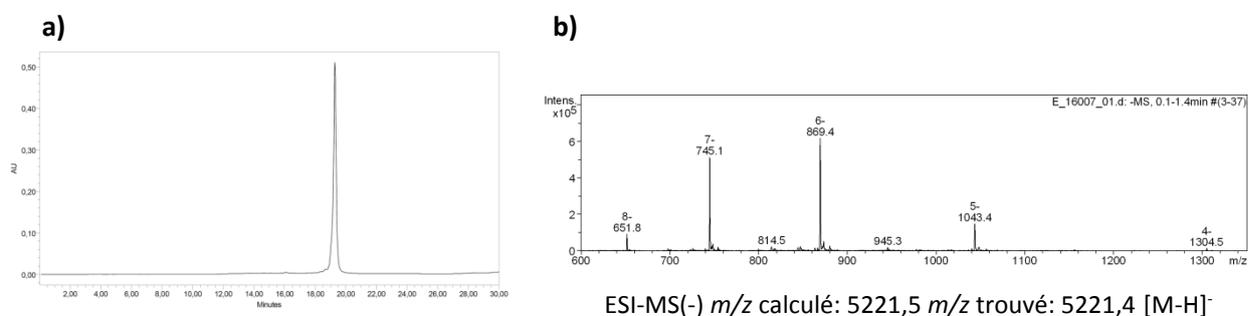


Figure 90 : (a) chromatogramme RP-HPLC et (b) spectre ESI-MS de l'édifice intermédiaire **H₂**

La dernière étape de synthèse consiste à utiliser le couplage thiol-chloroacétamide pour greffer les oligonucléotides **17** ou **18** sur l'intermédiaire **H₂**. La réaction est réalisée dans les mêmes conditions que pour l'assemblage de l'édifice HIV (p 81), à savoir 2 eq d'ADN **17** ou **18**, 2 eq de TCEP, 45 eq de DIEA, 60 eq de KI et 500 mM de KCl. Après 3h de réaction les bruts réactionnels sont purifiés par RP-HPLC et les édifices **H** et **I** sont respectivement obtenus avec un rendement de 50% et 60%. Les deux édifices ont été caractérisés par RP-HPLC et spectrométrie de masse (figure 91 et 92).

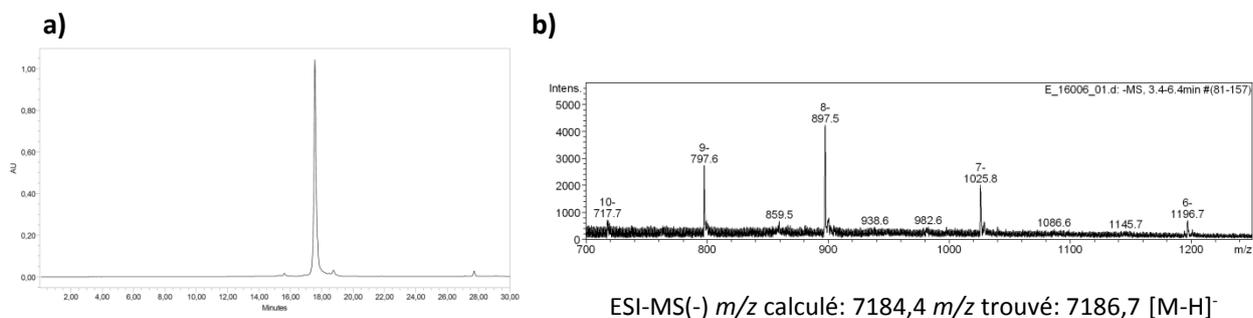


Figure 91 : (a) chromatogramme RP-HPLC et (b) spectre ESI-MS de l'édifice H

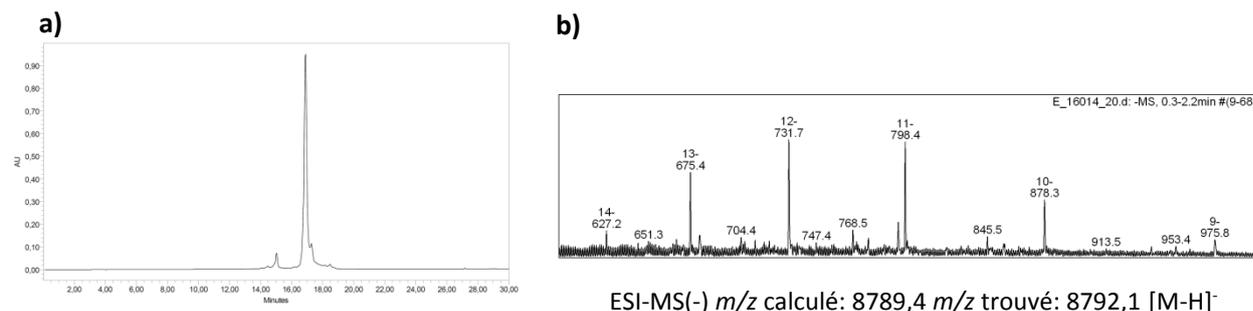


Figure 92 : (a) chromatogramme RP-HPLC et (b) spectre ESI-MS de l'édifice I

III. Etudes de la topologie des édifices mimant le motif G-quadruplexe de la séquence télomérique humaine et son intermédiaire triplex

Nous avons utilisés des analyses d'UV-visible avec la mesure des TDS et de dichroïsme circulaire, pour définir les topologies G-triplexe ou G-quadruplexe adoptées par les édifices. Toutes ces analyses ont été réalisées dans un tampon tris 10 mM à pH 7,4 et en présence de sel ajouté. Quatre cations métalliques ont été utilisés, K^+ et Na^+ largement utilisés pour la stabilisation de tétrades et donc susceptibles de stabiliser également les triades, ainsi que Mg^{2+} et Ca^{2+} . Ces deux derniers cations divalents ont été utilisés par l'équipe de H. Mao²⁹ pour stabiliser hum15 (5'GGGTTAGGGTTAGGG3'), une séquence d'ADN susceptible de former un intermédiaire triplex de la région télomérique. Cette étude est la seule à avoir décrite par dichroïsme circulaire les topologies G-triplexes adoptées par la séquence télomérique en présence de différents cations métalliques. Elle présente donc un élément de comparaison important pour la validation de nos travaux.

III.1. Etude de la topologie des édifices

III.1.a. Etude de la topologie des édifices H et I par dichroïsme circulaire

Pour identifier la topologie formée par les édifices **H** et **I**, des spectres CD ont été enregistrés à 20°C, en milieu tamponné (Tris 10 mM, pH 7,4) et en présence de 100 mM de NaCl, de KCl, de MgCl₂ ou de CaCl₂. Les différents spectres obtenus sont représentés dans la figure 93. De façon générale, quel que soit le sel utilisé, les deux édifices montrent des spectres totalement différents : ce qui témoigne d'une structuration distincte.

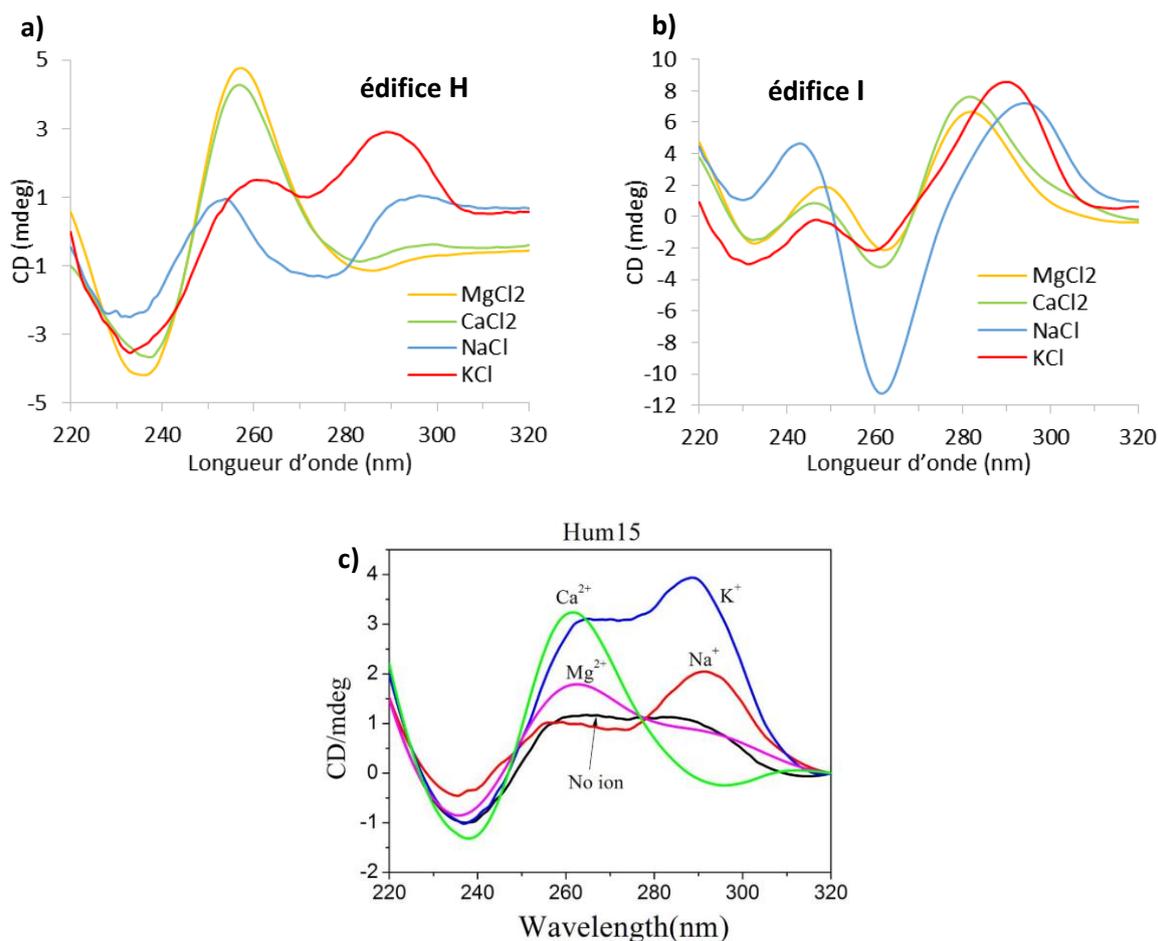


Figure 93: spectres CD à 20°C (a) des édifices **H** et (b) **I**, conditions : 1,6 μM d'édifices **H** ou **I** dans 10 mM de tampon Tris-HCl pH 7,4 en présence de 100 mM de différents cations métalliques ajoutés, (c) spectre CD de la séquence Hum 15, source : Jiang, H.-X. *et al. Sci. Rep.* **5**, 9255 (2015).

Le spectre CD de l'édifice **I** (figure 93.b) montre en présence de NaCl le signal caractéristique d'un motif G4 antiparallèle (figure 94.a), topologie préférentiellement adoptée par la séquence télomérique dans ces conditions salines. En présence de KCl, nous observons la formation d'une conformation G-quadruplexe hybride (figure 94.b). Dans CaCl₂ et MgCl₂, les signaux sont comparables à ceux obtenus en présence de KCl, avec un léger effet hypsochrome à 290 nm. Ces résultats semblent indiquer la formation de motifs G-quadruplexes distincts à partir de l'édifice **I**, en fonction

des conditions salines utilisées cet édifice permet d'obtenir des structurations différentes du G-quadruplexe en fonction des cations métalliques utilisés.

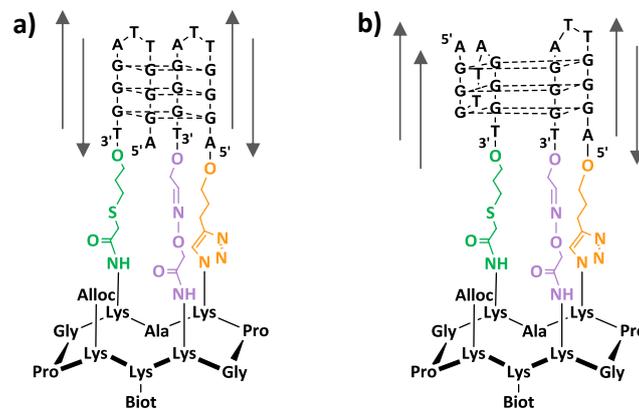


Figure 94 : structure chimique de l'édifice **I** en conformation antiparallèle (a) et en conformation hybride (b)

Pour l'édifice **H**, en présence de K^+ , le signal CD correspond à une topologie hybride (minimum à 235 nm et maximums à 255 et 290 nm), qui pourrait correspondre à la formation d'un G-triplexe, tandis qu'en présence de Na^+ nous observons un signal CD moins intense et la structuration du motif n'est pas clairement définie.

En présence de Mg^{2+} et Ca^{2+} , l'existence d'un signal CD de forte intensité, présentant un maximum à 260 nm suggère la formation d'un G-triplexe. Ces cations sont en effet connus pour stabiliser les triplexes. En comparant les spectres obtenus et ceux publiés par H. Moa²⁹ (figure 93.c), nous observons des correspondances qui semblent indiquer la formation d'un motif G-triplexe. Cependant, pour expliquer les profils CD obtenus par son équipe, H. Moa émet l'hypothèse de la formation d'une conformation parallèle pour Hum 15 en présence de $MgCl_2$ et $CaCl_2$. Hors, dans notre cas, la présence de la boucle contrainte complique fortement la formation d'une topologie parallèle. Cependant, au vu de la ré-organisation observée lors de la synthèse du motif ARN G-quadruplexe antiparallèle **F**, nous ne pouvons pas totalement exclure une ré-organisation structurale majeure donnant lieu à la formation d'une structure parallèle. Une autre hypothèse pourrait être plus simplement la formation d'un G-triplexe de topologie hybride, bien que le profil CD soit typique d'une organisation des parallèles des brins. Il a déjà été démontré par le passé, par l'équipe de K. Weisz¹⁵⁰ que des G-quadruplexes pouvaient présenter un profil CD ne correspondant pas à la conformation classiquement associée à ce profil. Ce phénomène pourrait se produire dans le contexte de systèmes moléculaires contraints impliquant des motifs G-triplexes pour lesquels peu de corrélations entre CD et RMN ont été réalisées. La ré-orientation des guanines nécessaire à un tel

phénomène semble plus probable au sein de triade à l'intérieure desquelles les guanines possèdent moins de contraintes que lorsqu'elles sont impliquées dans des tétrades.

Pour résumer, ces études semblent démontrer la formation du motif G-triplexe. Il reste cependant envisageable que plusieurs G-triplexes coexistent au sein de cette structure. Par ailleurs quelques soient les conditions salines utilisées deux structures distinctes représentées en figure 95 sont susceptibles de se former.

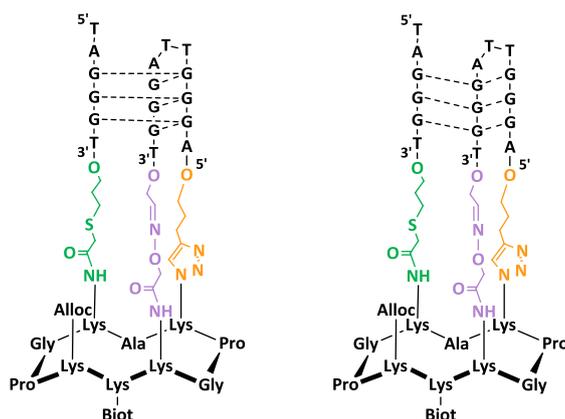


Figure 95 : structure chimique de l'édifice **H** adoptant différentes conformations hybrides

III.1.b. Etude de la topologie des édifices **H** et **I** par mesure des TDS

Des analyses TDS ont été réalisées dans du tampon Tris à 10 mM à pH 7,4 en présence de 100 mM KCl, NaCl, MgCl₂ ou CaCl₂. Les absorbances ont été mesurées entre 225 et 320 nm à 5 et 90°C. Pour l'édifice **I** (figure 96.a), les spectres obtenus en présence des quatre cations métalliques confirment la structuration des motifs G-quadruplexes. En effet, ils présentent avec diverses intensités les signaux caractéristiques d'un G-quadruplexe avec deux maximums aux alentours de 243 et 273 nm et un minimum proche de 295 nm. Pour l'édifice **H** (figure 96.b), les quatre spectres enregistrés sont également comparables, mais distincts de ceux obtenus pour l'édifice **I**. Ce résultat confirme que la topologie observée est distincte d'un G-quadruplexe. Aucune référence n'existe pour le profil TDS d'un motif triplex, ces données pourront donc servir de référence pour de futures études de motifs G-triplexes.

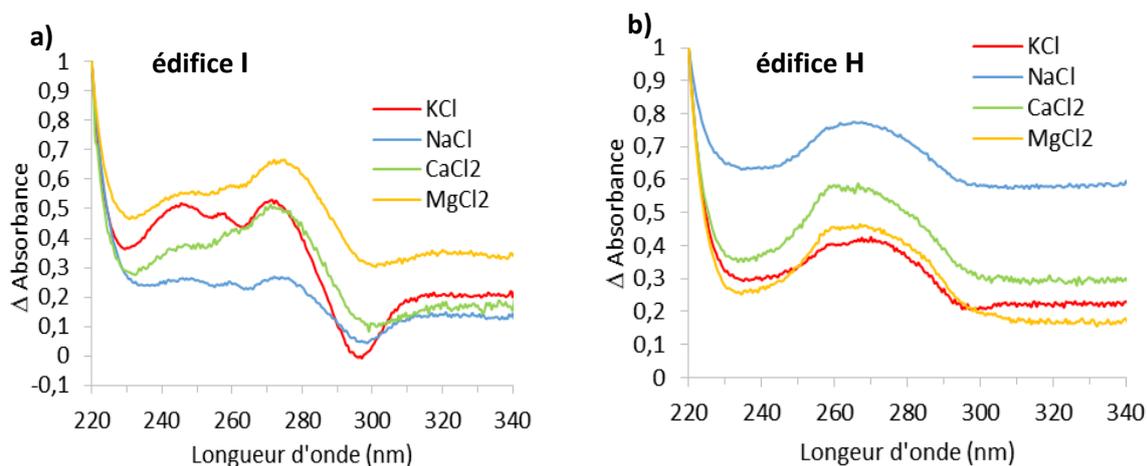


Figure 96 : spectres normalisés représentant la différence d'absorbance entre 90 et 5°C pour les édifices (a) **I** et (b) **H**, conditions : 1,6 μ M d'édifice dans du tampon Tris-HCl 10 mM pH 7,4

III.2. Etude de la stabilité des édifices

III.2.a. Evaluation de la stabilité des édifices

La stabilité des édifices **H** et **I** a été étudiée en mesurant leur température de dénaturation entre 5 et 90°C en présence de 100 mM de K^+ , cation physiologique permettant la formation des motifs G-triplexes et G-quadruplexes. La dénaturation de la séquence hum 15, qui présente le même signal CD à 20°C que l'édifice **H** (figure 97), a également été mesurée.

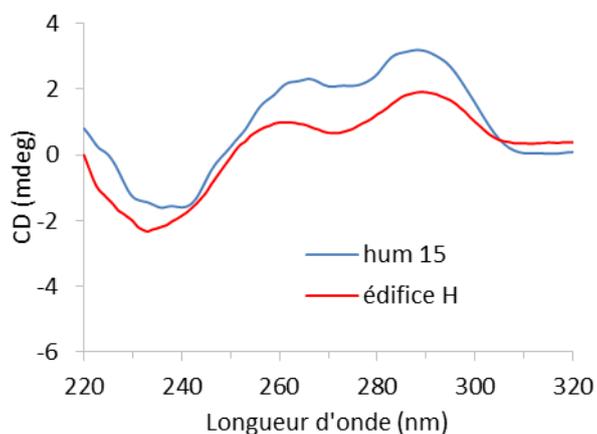


Figure 97 : spectres CD à 20°C de la séquence hum 15 (bleu) et de l'édifice **H** (rouge) (1,6 μ M) dans 10 mM de tampon Tris-HCl pH 7,4 et 100 mM KCl

Pour l'édifice **I**, une seule température de dénaturation est rapportée puisqu'il présente seulement une transition à 290 nm (figure 93.b. III.1.1 courbe rouge). Le T_m mesuré est égal 70°C et témoigne d'une forte stabilité du motif G-quadruplexe hybride (tableau 6).

La figure 98, qui représente la superposition de tous les spectres entre 5 et 90°C de l'édifice **H** et de la séquence hum 15, met en évidence la présence de deux transitions aux alentours des 255 nm et des 290 nm et permet de mesurer deux températures de dénaturations distinctes.

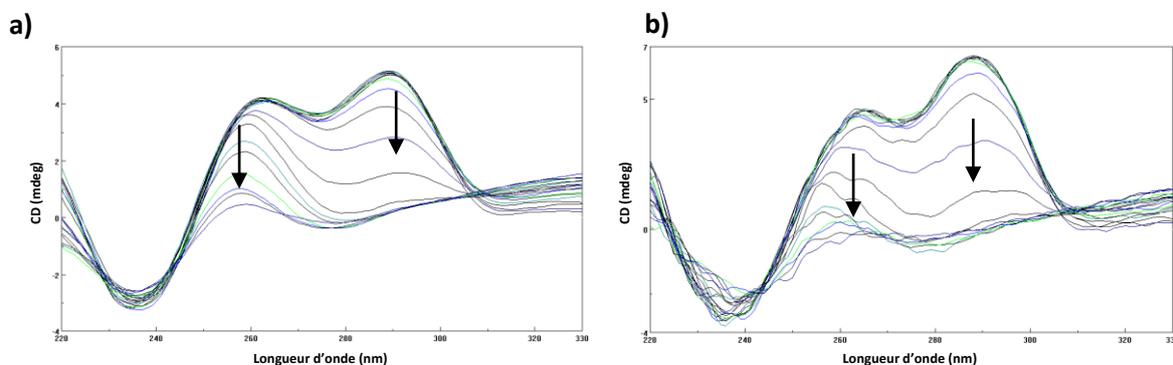


Figure 98 : spectres CD de dénaturation du motif G-triplexe (a) de l'édifice **H** et (b) de la séquence hum 15 entre 5 et 90°C, en présence de 10 mM de tampon Tris-HCl pH 7,4 et 100 mM KCl

Pour la transition à 290 nm, la séquence hum 15 et l'édifice **H** présentent des températures de dénaturation similaires, à savoir respectivement 49°C et 50°C (tableau 6). La structure G-triplexe formée par ces deux édifices semble donc posséder la même stabilité thermodynamique. Cette similitude n'est toutefois pas retrouvée pour la transition à 255 nm. A cette longueur d'onde, l'édifice **H** possède un T_m de 67°C, et hum 15 un T_m égal à 54°C. Cette deuxième transition pourrait correspondre à la dissociation du motif G-hairpin formé par le brin connecté par les liens oxime et triazole. La formation de motifs G-hairpin a également été discutée dans la littérature et isolé expérimentalement¹⁵¹. Ces résultats semblent donc montrer un processus de dénaturation en deux étapes, le brin court du motif G-triplexe de **H** se dissocie dès 50°C alors que persiste un motif G-hairpin à des températures plus élevées.

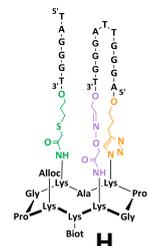
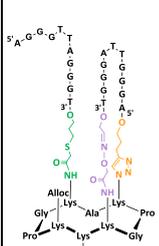
<p>Motifs G3</p> <p>Conditions: Tris 10 mM pH 7,4</p>	<p>hum 15</p> <p>5'GGGTTAGGGTTAGGG3'</p>		 <p>H</p>		 <p>I</p>	
	<p>255 nm</p>	<p>290 nm</p>	<p>255 nm</p>	<p>290 nm</p>	<p>290 nm</p>	
<p>KCl 100 mM</p>	<p>54°C</p>	<p>49°C</p>	<p>67°C</p>	<p>50°C</p>	<p>70°C</p>	

Tableau 6 : température de dénaturation (T_m) de la séquence hum 15 et des édifices **H** et **I** (1,6 μ M) dans 10 mM de tampon Tris-HCl pH 7,4 et 100 mM de KCl

Nous avons également observé lors de ces études CD que les spectres de dénaturation (figure 99.a) et d'association (figure 99.b) de l'édifice **H** ne sont pas identiques. En effet, bien que la transition à 255 nm soit similaire entre les deux processus ($T_{m\text{dénaturation}} = 67^\circ\text{C}$ et $T_{1/2\text{association}} = 68^\circ\text{C}$), une grande différence est observée pour la transition à 290 nm ($T_{m\text{dénaturation}} = 50^\circ\text{C}$ et $T_{1/2\text{association}} = 25^\circ\text{C}$). Ce résultat indique que la structuration des oligonucléotides en motif G-triplexe présente une cinétique lente. Après dénaturation, nous avons finalement observé que plusieurs heures à températures ambiantes étaient nécessaires pour retrouver la structuration initiale.

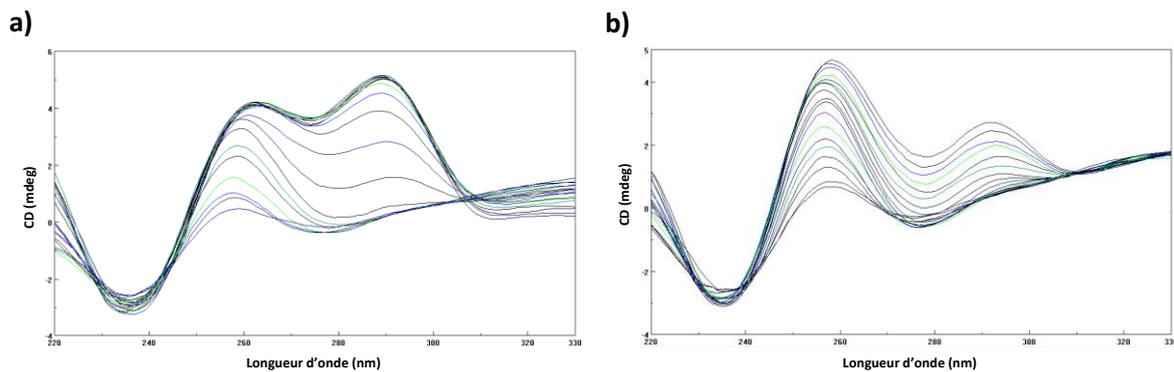


Figure 99 : spectres CD (a) de dénaturation (5 à 90°C) et (b) d'association (90 à 5°C) du motif G-triplexe de l'édifice **H** en présence de 10 mM de tampon Tris-HCl pH 7,4 et 100 mM KCl

III.2.b. Influence de la concentration sur la stabilité de l'édifice **H**

Afin de confirmer que le motif G-triplexe formé soit issu d'un repliement intramoléculaire des oligonucléotides et non d'une structuration intermoléculaire entre plusieurs édifices, des T_m ont été mesurés à différentes concentrations (1,6 μM , 3,2 μM et 6,4 μM) (figure 100).

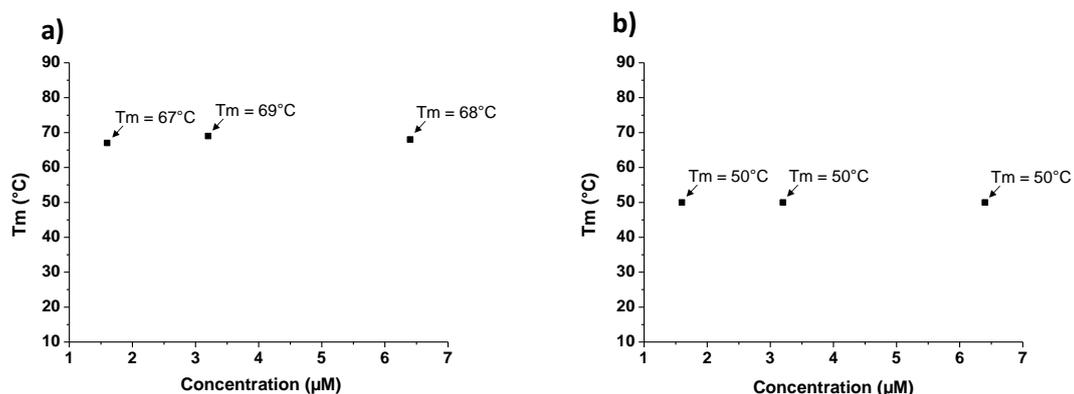


Figure 100 : température de dénaturation (T_m) de l'édifice **H** à (a) 255 nm et (b) 290 nm à 1,6, 3,2 et 6,4 μM dans Tris-HCl 10 mM pH 7,4 et KCl 100 mM

Nous retrouvons des températures de dénaturation similaires aux différentes concentrations, confirmant la structuration intramoléculaire.

III.3. Etude de résonance magnétique nucléaire

Des analyses RMN ont été réalisées sur les composés **H** et **I** en présence de sodium et de potassium. La présence de signal entre 10 et 12 ppm témoigne de la formation d'un motif G-quadruplexe pour le composé **I** dans les deux conditions salines utilisées (figure 101). Les profils obtenus en KCl et NaCl sont distincts, ceci est dû à la structuration du motif G4 dans des topologies différentes. En effet les analyses en dichroïsme circulaire, ont montré un repliement des G-quadruplexes en conformation antiparallèle en présence de NaCl et en topologie hybride en présence de KCl (figure 93.b).

Pour l'édifice **H**, nous observons également la présence de signaux aux alentours des 10-12 ppm (figure 101), comme attendu dans en cas de formation de triades de guanines au sein d'un G-triplexe. Les profils obtenus en KCl et NaCl ne sont pas identiques. Les analyses CD réalisées dans ces mêmes conditions salines (figure 93.a) ont également mis en évidence une structuration différente des brins d'ADN. L'absence de signaux clairement définis renforce l'hypothèse de la formation de plusieurs motifs G-triplexe au sein de l'édifice **H** (voir figure 95). Afin d'étudier plus précisément la topologie adoptée par les brins d'ADN de l'édifice **H**, des méthodes d'analyses plus poussées sont envisagées comme la RMN en deux dimensions ou encore la modélisation moléculaire.

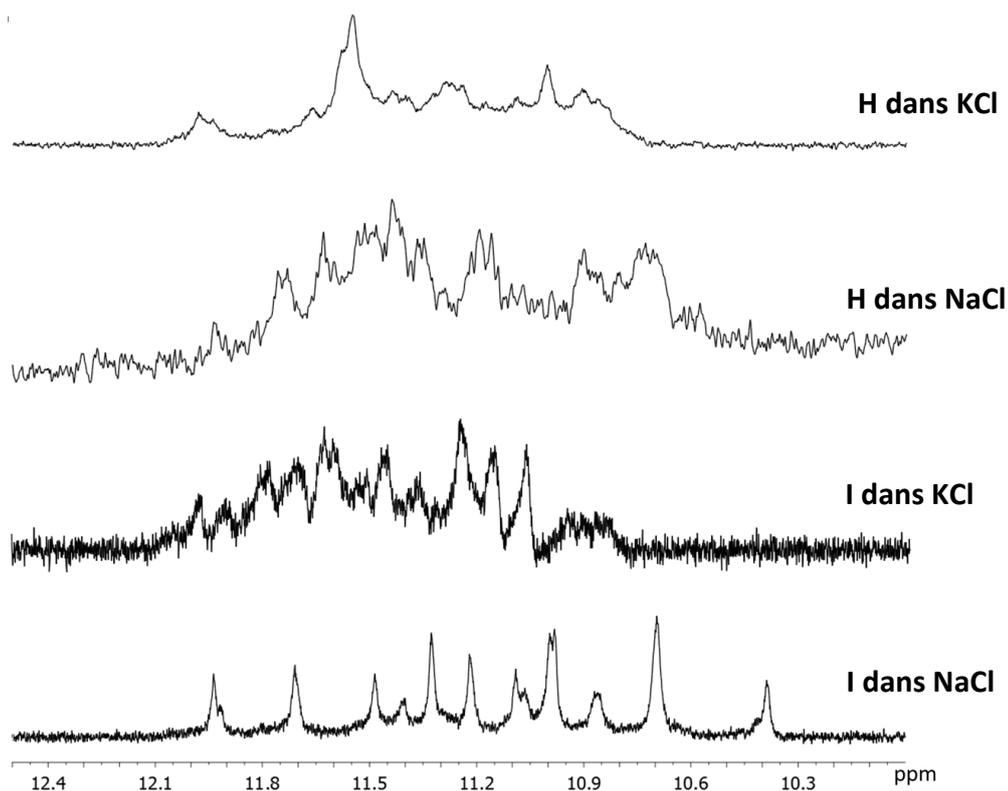


Figure 101 : spectre RMN ^1H (500 Mz) à 25°C, représentant la région des protons imino des édifices **H** et **I** à 80 μM en présence de 10 mM de tampon phosphate sodium à pH 7,5, de 100 mM de NaCl ou 100 mM KCl et de 10 % de D_2O

IV. Bilan

Ces travaux nous ont permis d'assembler deux structures susceptibles de former des G-triplexes. Les résultats CD, TDS et RMN indiquent la formation de G-quadruplexe pour l'édifice **I** et de G-triplexe pour l'édifice **H**. Ce dernier est stable ($T_m = 50^\circ\text{C}$), cependant les analyses de CD et RMN suggèrent la coexistence de plusieurs structures. Des études plus poussées seraient nécessaires pour caractériser cette/ces nouvelle(s) structure(s) G-triplexe(s) et pour observer si un G-triplexe se forme de manière transitoire lors du repliement du G-quadruplexe **I**. Le mime **H** a été utilisé pour étudier la reconnaissance de G-triplexe par des ligands. Cette étude est présentée dans la partie I du chapitre 2, p 134.

Partie IV : Mime d'ADN i-motif de la région télomérique

I. Introduction

Il a été évoqué en introduction, que les i-motifs, structures complémentaires aux G-quadruplexes seraient retrouvés aux mêmes endroits dans le génome humain. Il semblerait également qu'elles possèdent des rôles biologiques similaires dans la régulation cellulaire. Ces structures pourraient donc représenter des cibles biologiques potentielles. Face à cet intérêt, le laboratoire a créé en 2012 un modèle simplifié de i-motif¹⁵². Le modèle présente quatre séquences identiques d'ADN ($5'TCCCT3'$), liées de façon antiparallèle sur le châssis peptidique (figure 102.a). Ces travaux antérieurs ont validé la possibilité de stabiliser un i-motif à l'aide de la plateforme peptidique et ont également permis de montrer que la structure était capable de se former dans une gamme de pH plus large que celle d'un i-motif naturel. Suite à la validation de ce modèle, la stabilisation d'une véritable cible biologique, paraissait pertinente. Dans cette quatrième partie, nous détaillons les différents essais de synthèse de l'édifice représenté dans la figure 102.b, mimant le i-motifs de la région télomérique.

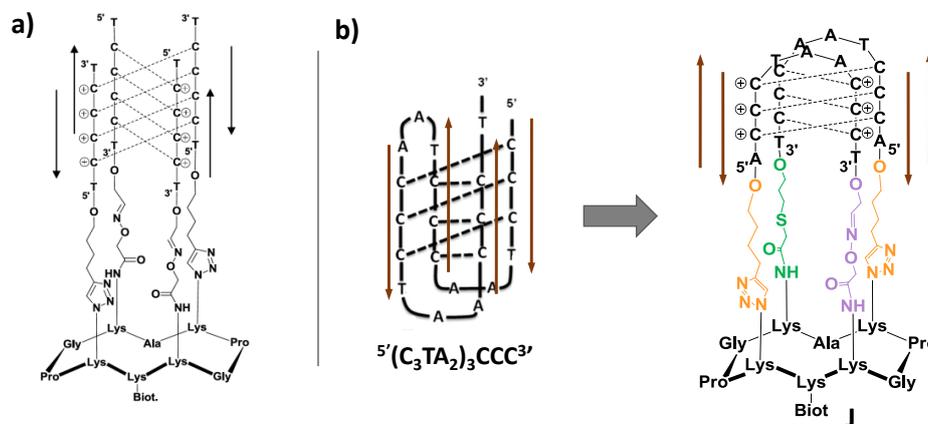


Figure 102 : (a) représentation du modèle de i-motif et (b) de l'édifice J mimant le i-motif de la région télomérique

II. Synthèse de l'édifice i-motif de l'ADN télomérique

Afin de mimer au mieux le i-motif télomérique, nous avons souhaité stabiliser les deux boucles inférieures (figure 102.b). Il est important de noter que les boucles sont des éléments de reconnaissance importants dans les interactions cible-ligand. Elles sont propres à chaque motif et les représenter augmente les chances de sélection de ligands spécifiques. Dans le cas de ce i-motif, les

mimer nécessite un greffage indépendant de chaque extrémité des brins d'ADN. En utilisant l'orthogonalité des ligations chimiosélectives et celles des groupements protecteurs présents sur le châssis peptidique, nous souhaitons utiliser les trois ligations à notre disposition (la ligation oxime, le couplage thiol-chloroacétamide et la cycloaddition CuAAC) pour greffer indépendamment chaque extrémité des deux boucles sur le châssis peptidique. Cette stratégie nécessite donc d'employer deux fois la même ligation chimiosélective. Pour des raisons d'orthogonalité des fonctions réactives nous avons choisi d'utiliser deux fois la CuAAC. Notre stratégie de synthèse présentée dans le schéma 31 est la suivante :

- greffer dans un premier temps un premier oligonucléotide formant une boucle sur le châssis en utilisant la ligation oxime et la CuAAC. Afin de conserver la sélectivité de la CuAAC, seulement une fonction azoture est présente sur le châssis peptidique. Pour pouvoir réaliser la deuxième réaction de cycloaddition, la fonction azoture nécessaire sera générée par la suite à partir de la fonction amine *via* une réaction de diazo-transfert.

- greffer dans un second temps le deuxième oligonucléotide sur le châssis en utilisant le couplage thiol-chloroacétamide et la CuAAC, pour la seconde fois. Dans le but d'optimiser la réaction de diazo-transfert en présence d'un bioconjugué peptide-ADN, deux voies de synthèse ont été envisagées. La voie a (schéma 31, flèche bleu) consiste à greffer dans un premier temps une extrémité du brin d'ADN en utilisant le couplage thiol-chloroacétamide, puis à réaliser la réaction de diazo transfert. Dans la voie b, l'ordre des réactions est inversée (schéma 31, flèche rouge), le couplage thiol-chloroacétamide a lieu cette fois-ci avant la réaction de diazo-transfert. La dernière étape réactionnelle, qui n'est autre que la formation de la deuxième boucle via la CuAAC est commune aux deux voies de synthèse. Il a été choisi d'utiliser cette réaction en dernière étape et non le couplage thiol-chloroacétamide pour plusieurs raisons. La première est tout simplement qu'elle a déjà été utilisée de nombreuses fois avec succès dans des fermetures de boucles intramoléculaires. Sa plus grande réactivité par rapport au couplage thiol-chloroacétamide est également une bonne raison de l'utiliser en dernière étape. En effet, la CuAAC s'est montrée efficace en condition stœchiométrique avec des rendements HPLC quantitatifs, tandis qu'un excès de réactif est nécessaire dans le cadre de la réaction Thiol-SN2 avec des rendements HPLC obtenus de l'ordre de 70%.

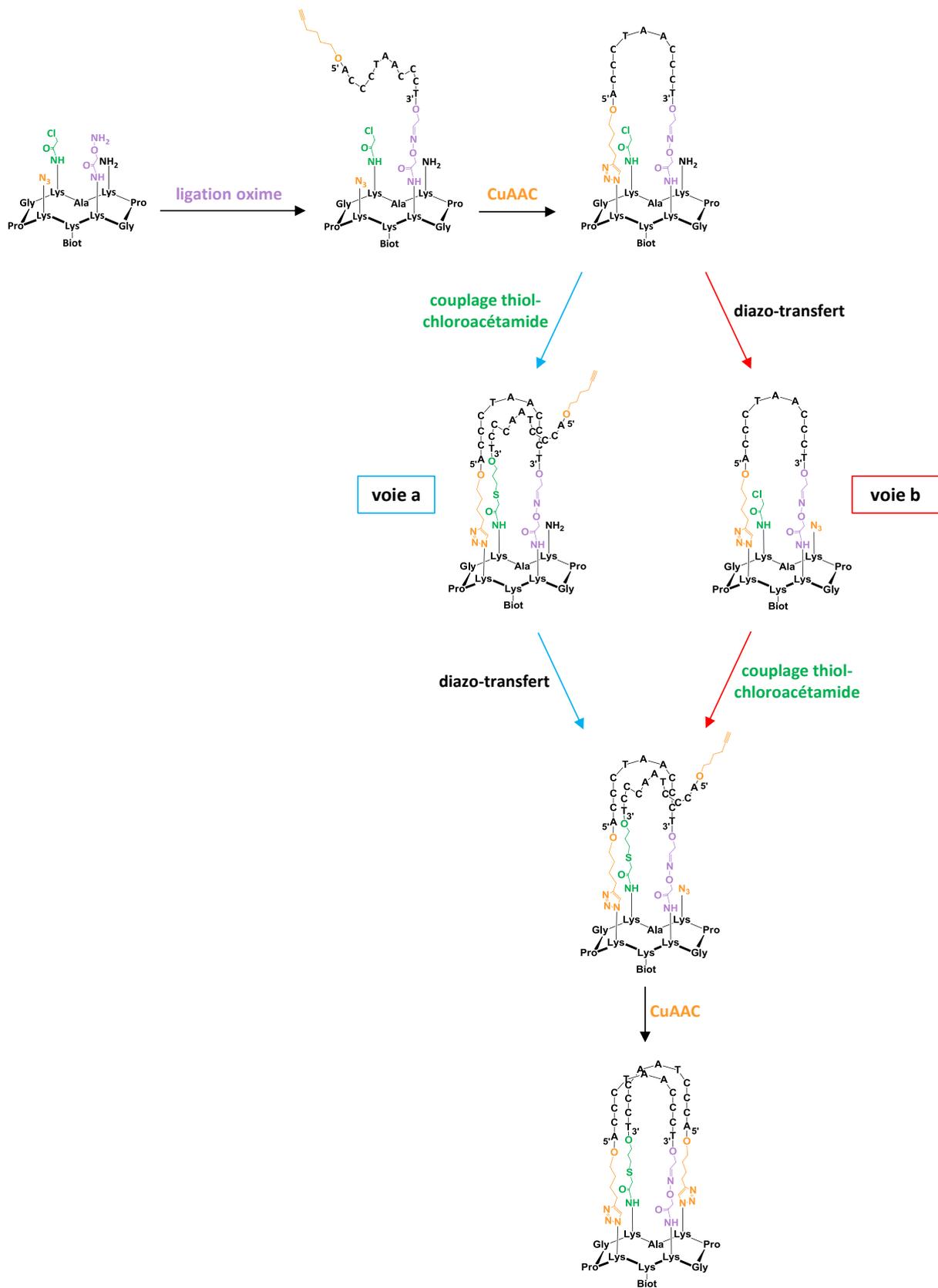


Schéma 31 : Stratégie de synthèse envisagée pour l'assemblage de l'édifice i-motif

II.1. Synthèse du châssis peptidique

Pour pouvoir réaliser consécutivement quatre ligations chimiosélectives, le châssis peptidique doit être tétrafonctionnalisé avec une fonction oxyamine, une fonction azoture, une fonction chloroacétamide et une fonction amine (figure 103).

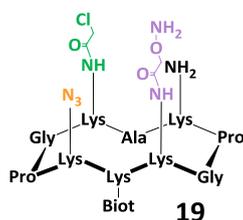


Figure 103 : Structure chimique du châssis peptidique **19**

Le châssis peptidique est synthétisé sur support solide en utilisant la stratégie de synthèse au Fmoc/tBu et la résine 2-chlorotrityl. La synthèse de ce châssis (schéma 32) est similaire à celle du châssis **14** (p 101) elle ne sera donc pas décrite. Le châssis tétrafonctionnalisé a été caractérisé par UPLC et spectrométrie de masse (figure 104).

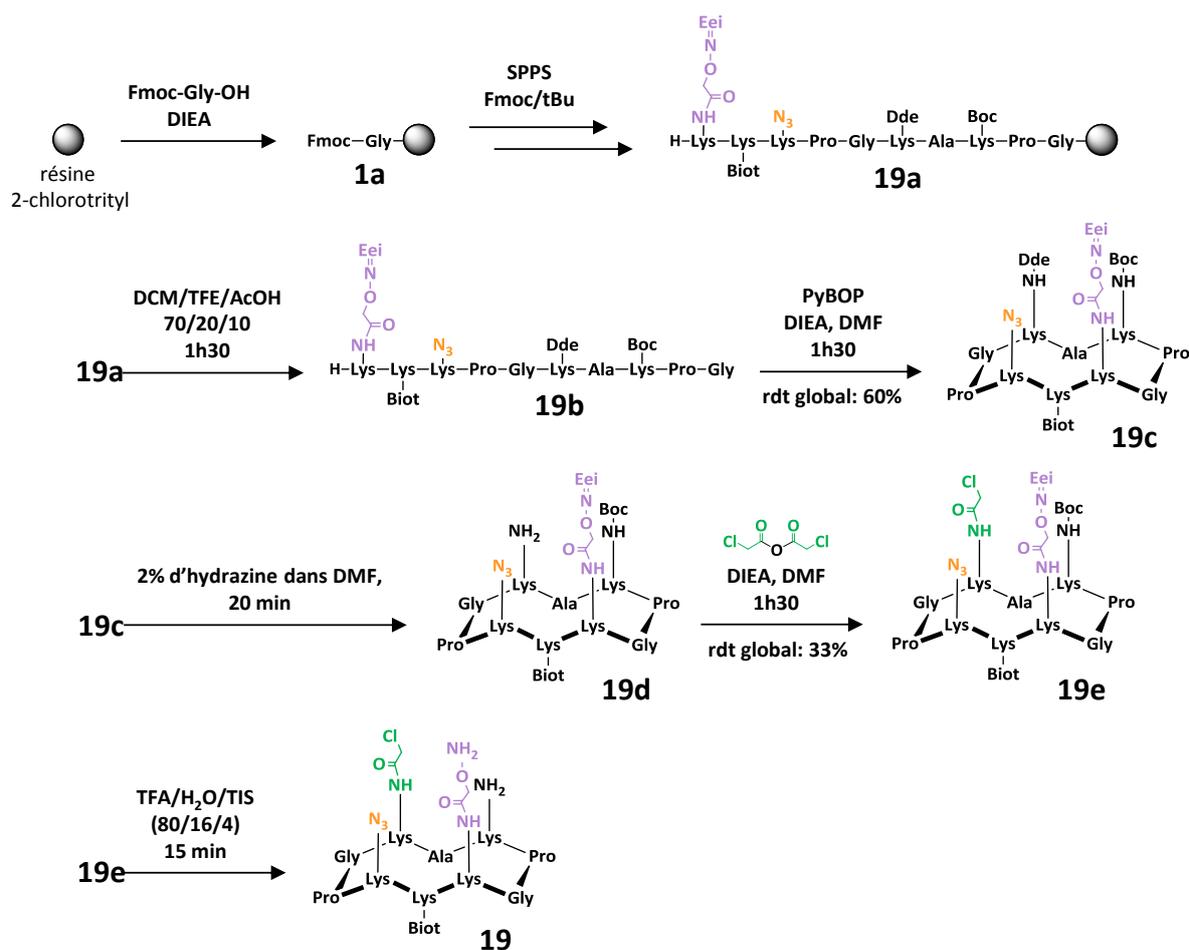
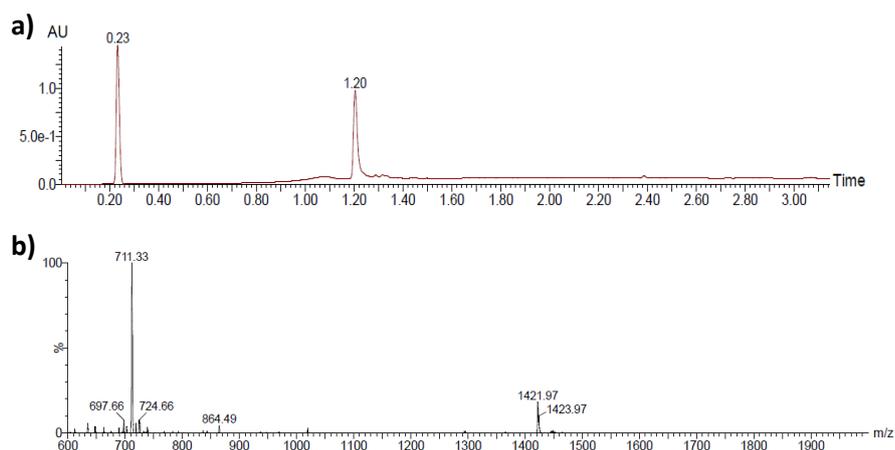


Schéma 32 : Synthèse du châssis peptidique **19**



ESI-MS(+) m/z calculé: 1421,7, m/z trouvé: 1421,9 [M+H]⁺

Figure 104: (a) chromatogramme UPLC et (b) spectre ESI-MS du châssis peptidique **19**

II.2. Synthèse des brins d'ADN

Les deux brins d'ADN présentent la même séquence et la même fonction alcyne en 5', mais sont fonctionnalisés différemment en leur extrémité 3'. L'un comporte une fonction aldéhyde (^{5'}YACCCTAACCTX^{3'}, Y = 5'alcyne, X = 3' aldéhyde) (ADN **20**) et l'autre une fonction thiol (^{5'}YACCCTAACCTZ^{3'}, Y = 5'alcyne, Z = 3' thiol) (ADN **21**).

II.2.a. Synthèse de l'oligonucléotide 5'alcyne-3'aldéhyde **20**

L'oligonucléotide **20** (schéma 33) a été préparé suivant le protocole présenté p 84. Il a été caractérisé par RP-HPLC et par spectrométrie de masse (figure 105).

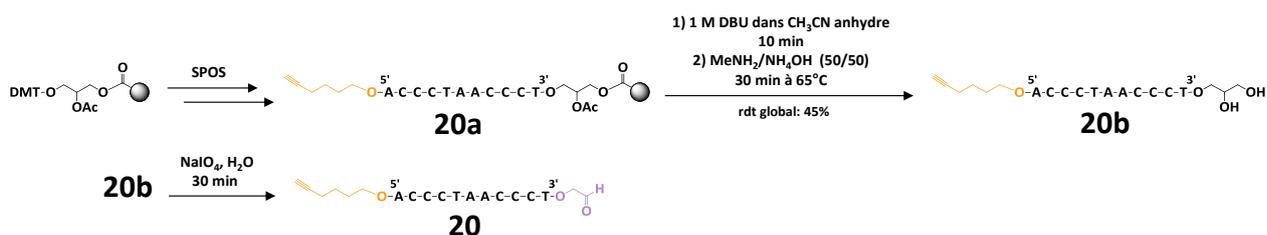


Schéma 33 : Synthèse de l'ADN 5'alcyne - 3'aldéhyde **20**

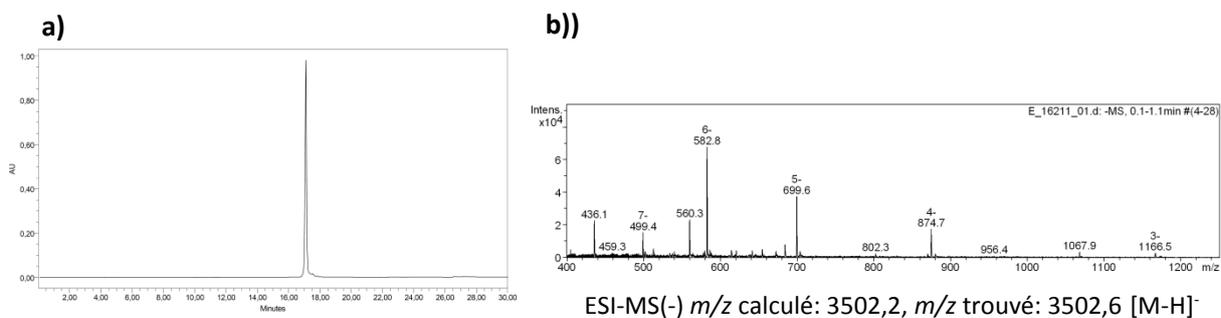


Figure 105: (a) chromatogramme RP-HPLC et (b) spectre ESI-MS de l'ADN 5'alcyne - 3'aldehyde **20**

II.2.b. Synthèse de l'oligonucléotide 5'alcyne-3'thiol **21**

L'oligonucléotide est synthétisé sur support solide en utilisant une résine 3'thiol (3'-thiol-modifier C3 S-S CPG) (schéma 34). Sa coupure en milieu basique, permet la libération en 3' d'une fonction thiol protégée par un pont disulfure. La fonction alcyne en 5' est incorporée sous forme de phosphoramidite modifié, le 5' hexynyl, lors de la synthèse supportée. Après purification par RP-HPLC, l'intermédiaire protégé **21b** est obtenu avec un rendement de 44%. La fonction thiol est libérée par réduction du pont disulfure en présence de 100 mM de DTT et de 1M de tampon Tris. L'ADN **21** bifonctionnel est obtenu avec un rendement de 77% après purification RP-HPLC. Il a ensuite été caractérisé par RP-HPLC et par analyse de masse (figure 106).

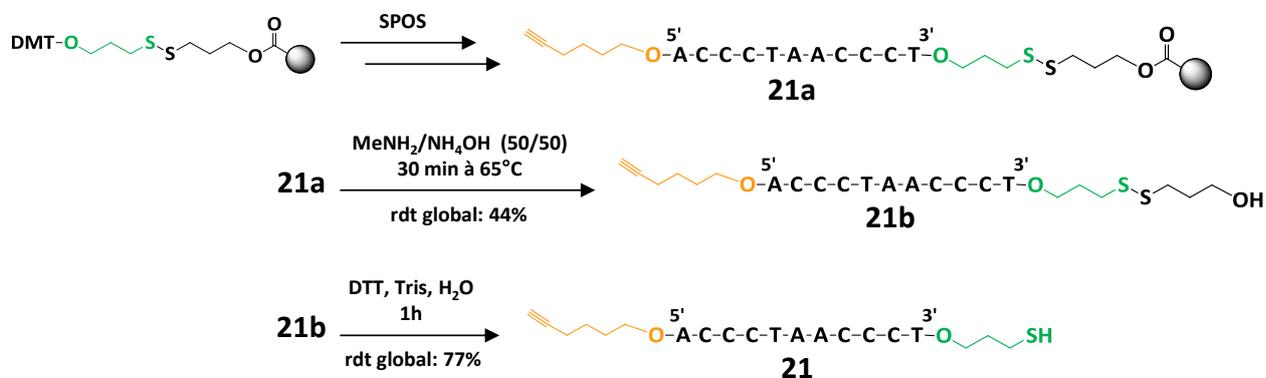


Schéma 34 : Synthèse de l'ADN 5'alcyne - 3' thiol **21**

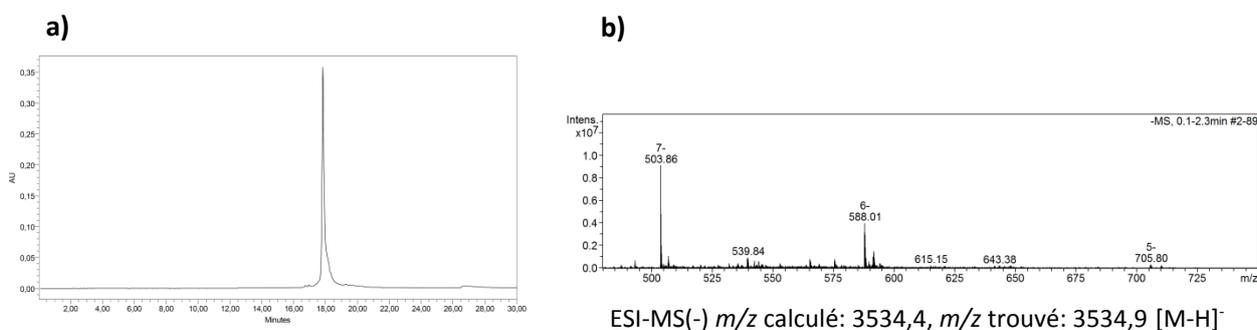


Figure 106 : (a) chromatogramme RP-HPLC et (b) spectre ESI-MS de l'ADN 5'alcyne - 3'thiol **21**

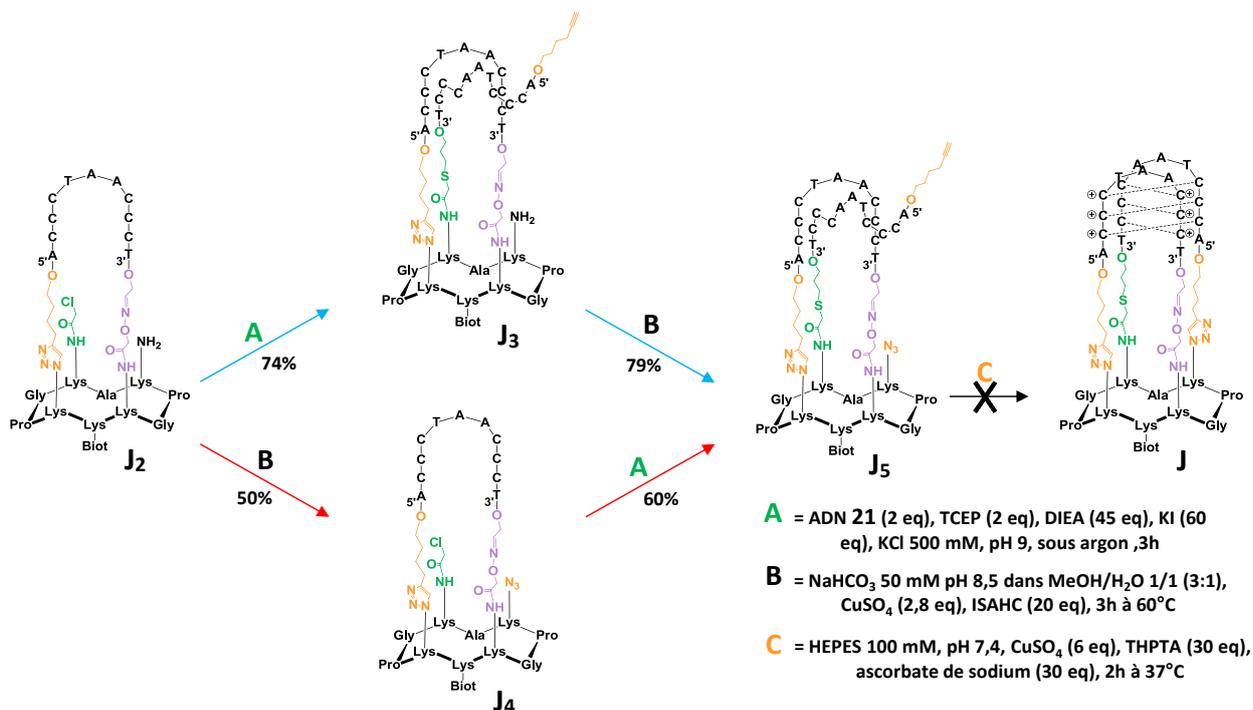


Schéma 36 : Synthèse de l'édifice **J** à partir de l'intermédiaire **J₅**, par deux voies de synthèse : voie a en bleu et voie b en rouge

- **Voie a (flèches bleues)**

Elle consiste à accrocher dans un premier temps l'oligonucléotide **21** *via* le couplage thiol-chloroacétamide, puis à réaliser la réaction de diazo-transfert. Le couplage thiol-chloroacétamide a été réalisé en utilisant les conditions réactionnelles mis au point p 90 et l'intermédiaire **J₃** a été obtenu avec un rendement de 74%. La fonction azoture normalement présente sur le châssis peptidique et nécessaire à la CuAAC est générée à partir de l'intermédiaire **J₃**, par réaction de diazo transfert sur la fonction amine libre. Cette réaction a déjà été étudiée au laboratoire, nous avons utilisé les mêmes conditions réactionnelles¹⁵³, à savoir un mélange méthanol/eau (50/50), 50 mM de NaHCO₃ à pH 8,5, 2,8 eq de CuSO₄ et 20 eq d'ISAHC. L'intermédiaire **J₅** a été obtenu avec un rendement de 79%.

- **Voie b (flèches rouges)**

Dans la voie b l'ordre des réactions est inversé. Le greffage de l'oligonucléotide **21** *via* le couplage-chloroacétamide thiol réaction est effectué après la réaction de diazo transfert. Les conditions réactionnelles employées sont identiques à celles de la première voie de synthèse. Les intermédiaires **J₄** et **J₅** ont respectivement été obtenus avec des rendements de 50% et 60%.

L'édifice intermédiaire **J₅**, caractérisé par RP-HPLC et spectrométrie de masse (figure 108) a été obtenu par les deux voies de synthèse avec des rendements globaux corrects. Dans les deux cas,

l'enchaînement des réactions fonctionne bien. La compatibilité de la réaction de diazo-transfert avec les intermédiaires **J₂** et **J₃** est donc confirmée.

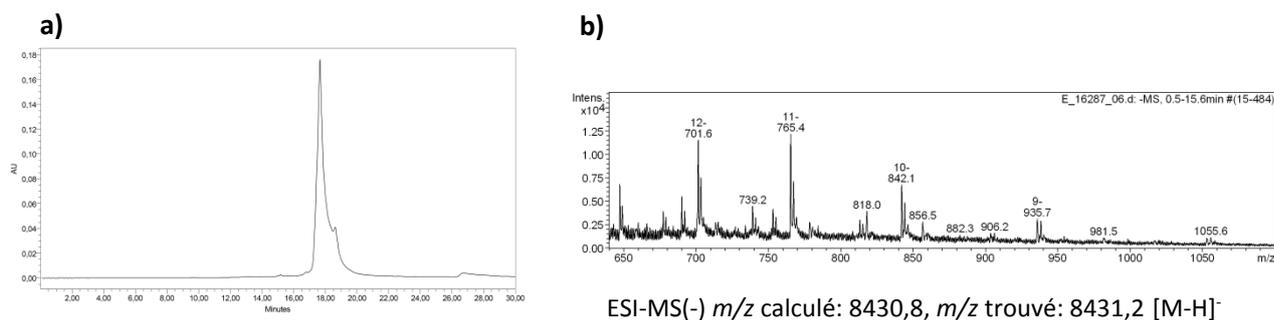


Figure 108 : (a) chromatogramme RP-HPLC et (b) spectre ESI-MS de l'édifice intermédiaire **J₅**

II.3.c. Assemblage de l'édifice **J** à partir de l'intermédiaire **J₅**

La dernière étape de synthèse repose sur l'utilisation de la CuAAC pour former la deuxième boucle du i-motif (schéma 36). Il est à noter que la formation d'un lien triazole intramolécule n'entraîne pas de changement de masse entre les produits de départ et d'arrivée. Un changement de temps de rétention entre **J₅** et **J** lors de l'analyse HPLC permettrait cependant de rendre compte du succès de la réaction.

Nous avons tout d'abord effectué un premier test en utilisant les conditions classiques de CuAAC intramolécule, c'est-à-dire 100 mM d'HEPES à pH 7,4, 6 eq de CuSO₄, 30 eq de THPTA et 30 eq d'ascorbate de sodium. Dans ce premier essai, aucun sel n'a été ajouté, en effet, contrairement aux G-quadruplexes, les i-motifs ne sont pas stabilisés par des cations métalliques, ceci ne permettent donc pas de pré-organisation du motif. Après 2h de réaction, le chromatogramme HPLC montre un pic ayant le même temps de rétention que le produit de départ. Aucun changement n'est visible après des temps plus long (4h et 6h).

Dans un second test, nous avons réalisé la réaction en présence de 500 mM de KCl, dans l'optique d'un écrantage de charge, pour limiter la répulsion électrostatique entre les brins qui pourrait éloigner les deux fonctions réactives (alcyne et azoture). De nouveau, l'analyse par HPLC montre la présence d'un unique pic ayant le même temps de rétention que le produit de départ.

Dans un troisième test, nous avons étudié l'effet du pH sur la réaction. Il est nécessaire de rappeler que la formation des i-motifs est favorisée à pH acide. Or dans les deux tests réalisés, le pH était égal à 7,4. Dans l'objectif de favoriser la réaction par une pré-organisation de l'édifice, nous avons dans un premier temps analysé par CD la topologie adoptée par l'intermédiaire à différents

pH. Les spectres obtenus (figure 109) révèlent la formation d'un i-motif antiparallèle pour des pH inférieurs à 6. A pH 7,4, nous n'avons pas de pré-organisation en i-motif.

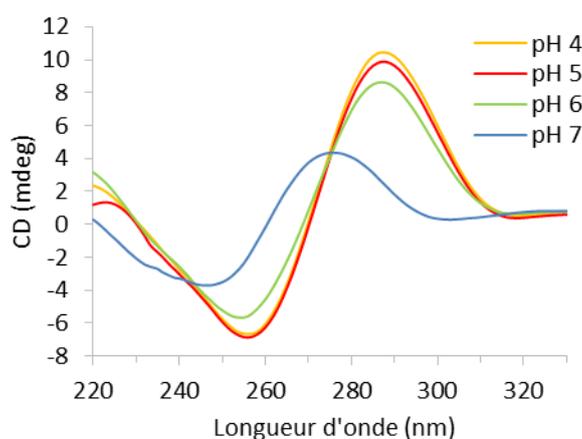


Figure 109 : spectres CD à 20°C de l'intermédiaire **J₅** à différents pH dans 10 mM de tampon PBS

Nous avons donc exploré l'effet du pH sur la réaction. Les tests ont été réalisés en présence de CuSO_4 (6eq) de THPTA (30 eq) et d'ascorbate de sodium (30 eq) en milieu tamponné (100 mM de PBS) à pH 4, 5, 6 ou 7. De nouveau, nous n'avons observé aucun changement dans les chromatogrammes HPLC. Les analyses de masses et les chromatographies HPLC et échangeuse d'ions ne pouvant pas, de manière non ambiguë, confirmer ou infirmer la formation du composé **J**, une solution serait alors d'analyser le brut réactionnel en infrarouge. Dans le but de détecter les bandes correspondantes aux azotures libres ou aux alcynes libres témoignant de la non-formation de l'édifice final. Ceci n'a malheureusement pas pu être réalisé faute d'équipement. Nous pourrions également essayer de mettre au point un test de « quenching » de **J₅** avec un dérivé azoture qui conduirait à un déplacement chromatographique. Par faute de temps cette expérience n'a pas pu être réalisée.

III. Bilan

Durant cette étude, nous avons essayé de greffer de façon indépendante chaque extrémité de deux brins d'ADN sur le châssis peptidique pour former un i-motif. La disponibilité de trois ligations chimiosélectives au laboratoire, nous a contraints d'en utiliser une deux fois. La mise en place de la stratégie de synthèse s'est donc avérée compliquée, car il a été nécessaire de combiner l'orthogonalité des groupements protecteurs avec l'orthogonalité des ligations chimiosélectives tout en préservant la stabilité des oligonucléotides. La stratégie que nous avons choisie ne nous a pas permis d'aller au bout de l'assemblage de l'édifice **J**. Nous avons été bloqués lors de la dernière

étape de synthèse. En finissant par la CuAAC pour la fermeture de la deuxième boucle, nous avons été confrontés à un problème de caractérisation. En effet, la CuAAC n'a entraîné ni changement de temps de rétention, ni changement de masse, ni changement du nombre de charges, ce qui a rendu impossible l'utilisation de la RP-HPLC, de la spectrométrie de masse et de la chromatographie échangeuse d'ions pour la caractérisation de l'édifice. Nous ne sommes donc pas en mesure de savoir si nous avons obtenu ou non l'édifice i-motif.

Chapitre 2

Utilisation de G-quadruplexes contraints pour la caractérisation
et la sélection de ligands.

Introduction générale

Ce deuxième chapitre est consacré à l'étude et au développement de ligands de G-quadruplexe. Comme évoqué en introduction, certaines stratégies anticancéreuses et antivirales visent à concevoir des molécules capables de stabiliser les motifs G-quadruplexes. Cependant, l'importante variété de topologies que ces motifs peuvent adopter en solution complique leur développement et leur étude. L'utilisation des édifices G-quadruplexes synthétisés durant cette thèse, suivant l'approche « TASQ » (chapitre 1), pourrait faciliter les études de reconnaissance ligand-G4. En effet, l'utilisation de G4 contraints, structurés en une seule conformation, devrait nous aider à comprendre les interactions impliquées dans les phénomènes de reconnaissance des G4 par des ligands. Nos édifices contraints pourraient ainsi servir d'outils, pour la sélection et la conception de ligands de G-quadruplexes.

Ce chapitre est composé de deux parties. La première est consacrée à l'étude de l'affinité et de la sélectivité de différentes familles de ligands pour les motifs G-quadruplexes ARN parallèle **C**, hybride ADN/ARN **E**, HIV-PRO1 **G** et G-triplexe **H**, par mesure d'interaction de surface par résonance plasmonique de surface (SPR) ou par interférométrie bio-couche (BLI). Grâce à ces nouveaux édifices qui viennent s'ajouter aux édifices G-quadruplexes ADN parallèle **A** et télomérique « basket » **B** précédemment synthétisés, nous possédons un panel assez important de motifs G4 présentant des séquences et des topologies variées. Il serait donc intéressant de trouver des différences d'interactions et d'affinité entre tous ces motifs, afin d'arriver à cibler spécifiquement une conformation G4. Cette approche, utilisant comme ligands des petites molécules organiques est néanmoins restreinte. En effet, la grande majorité des petites molécules ont une préférence limitée pour une topologie donnée et beaucoup conservent une affinité non négligeable pour le duplex d'ADN. Durant cette thèse nous avons donc essayé de développer une nouvelle classe de ligands, sélectifs d'une topologie définie. Ces ligands sont des aptamères obtenus par une méthode de sélection *in vitro* nommée SELEX pour *Systematic Evolution of Ligands by EXponential enrichment*. Ces travaux sont présentés dans la deuxième partie de ce chapitre où nous décrivons nos différents essais de sélection d'aptamères contre le motif G-quadruplexe de type basket dérivé de la séquence télomérique.

Partie I : Utilisation de motifs G4 contraints pour évaluer l'affinité et la spécificité de différentes familles de ligands par SPR et par BLI

Je tiens ici à associer Laure Bar qui a participé à ces travaux lors de son stage de Master 2 dans notre laboratoire.

I. Introduction

Cette première partie est consacrée à l'étude des interactions entre différentes familles de ligands (les salophens, les porphyrines, les dérivés d'acridine et les dérivés de bisquinolinium) (figure 110) et les motifs G-quadruplexes ARN parallèle **C**, hybride ADN/ARN **E**, HIV-PRO1 **G** et G-triplexe **H**. La caractérisation des interactions a été réalisée par résonance plasmonique de surface (SPR) ou par interférométrie bio-couche (BLI).

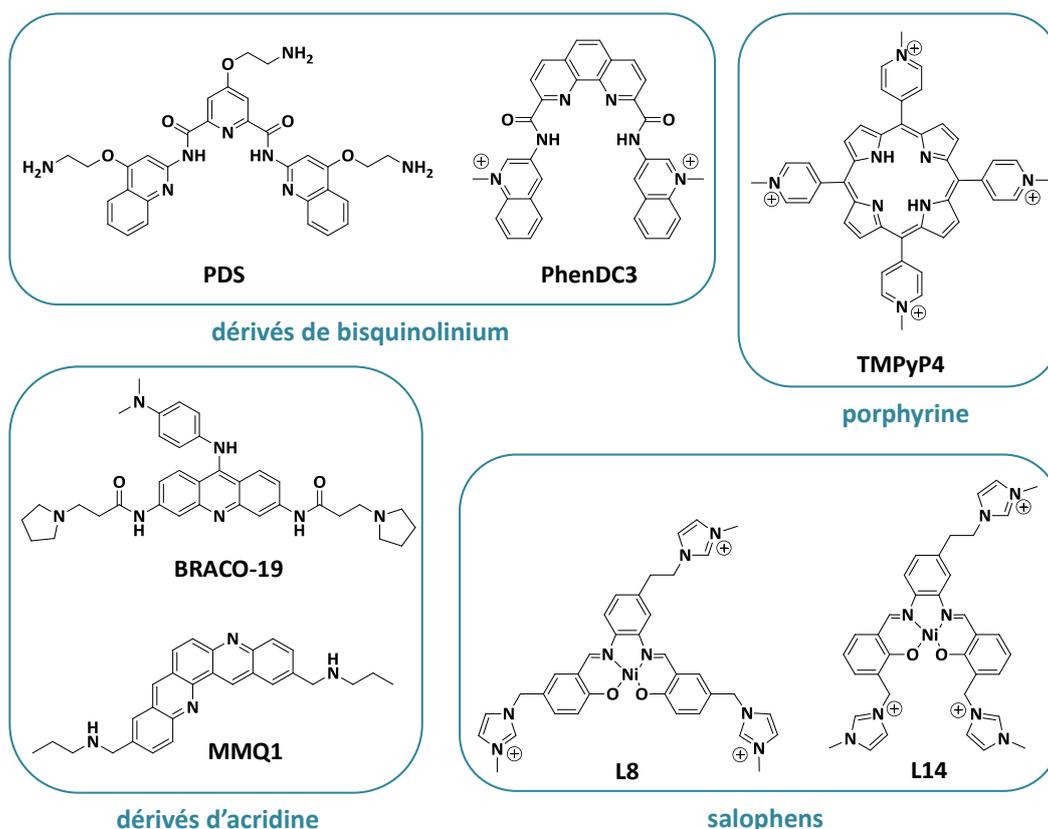


Figure 110 : ligands utilisés pour les études d'interactions

Grâce aux paramètres cinétiques et thermodynamiques obtenus par ces deux techniques, il nous est possible d'identifier des ligands intéressants en terme d'affinité, de sélectivité pour les G4 par rapport au duplexe d'ADN mais aussi par rapport à une topologie donnée.

I.1. Principe de la SPR

La résonance plasmonique de surface a été développée en 1977 par H. Reather¹⁵⁴. Cette technique optique permet de visualiser en temps réel des interactions entre biomolécules non marquées. Pour cela, un premier partenaire (ici un édifice G-quadruplexe) est immobilisé sur une surface constituée de verre et recouverte d'une fine couche de métal fonctionnalisé (ici Au). Un tampon contenant le second partenaire (ici le ligand de G4) est envoyé sur la surface en flux continu. Les mimes de G4 présentés dans le chapitre 1, possèdent tous une fonction biotine sur la face inférieure de leur châssis peptidique. Grâce à cette fonction, ils ont pu être immobilisés de façon non-covalente sur la surface d'or, préalablement fonctionnalisée par de la streptavidine. Le système BIAcore® que nous avons utilisé, permet de suivre en parallèle quatre pistes, la première est non fonctionnalisée et sert de référence, deux autres servent à l'immobilisation des mimes de G-quadruplexes et la dernière à l'immobilisation d'un édifice de duplexe d'ADN. Ce dernier permet d'évaluer la sélectivité des ligands pour les motifs G-quadruplexes par rapport au duplexe d'ADN.

Le principe de la SPR (figure 111) repose sur la capacité d'un métal (ici l'or) à effectuer un plasmon lorsqu'il se situe entre deux milieux d'indice de réflexion différents et qu'il reçoit de la lumière polarisée. Lorsqu'un rayon lumineux se propageant dans un milieu d'indice de réfraction élevé (ici le verre) et arrive à l'interface d'un deuxième milieu d'indice de réfraction plus faible (ici le tampon), une partie de cette lumière est réfractée et une autre est réfléchi. Dans le cas d'une réflexion totale, une partie du rayon totalement réfléchi est transformé en une composante électromagnétique appelé onde évanescente. Cette onde est capable d'interagir avec la couche électronique présente à la surface du métal (plasmon). Cette interaction entraîne la résonance des plasmons de surface qui absorbent alors de l'énergie *via* l'onde évanescente. Ce phénomène engendre une diminution de l'intensité du rayon réfléchi. L'angle correspondant à l'intensité minimale s'appelle angle de résonance et varie en fonction de l'indice de réfraction. C'est cette variation d'angle qui est suivie par SPR. Ainsi, lorsque les ligands circulant à flux constant arrivent à la surface du biocapteur, les changements de masse, mêmes infimes, induits par l'association ou la dissociation des complexes, engendrent une modification de l'indice de réfraction du milieu. Ce qui entraîne des variations de l'angle de résonance, qui sont captées par une photodiode puis enregistrées par un logiciel de contrôle sous forme de sensorgramme (unité de résonance (RU) en fonction du temps).

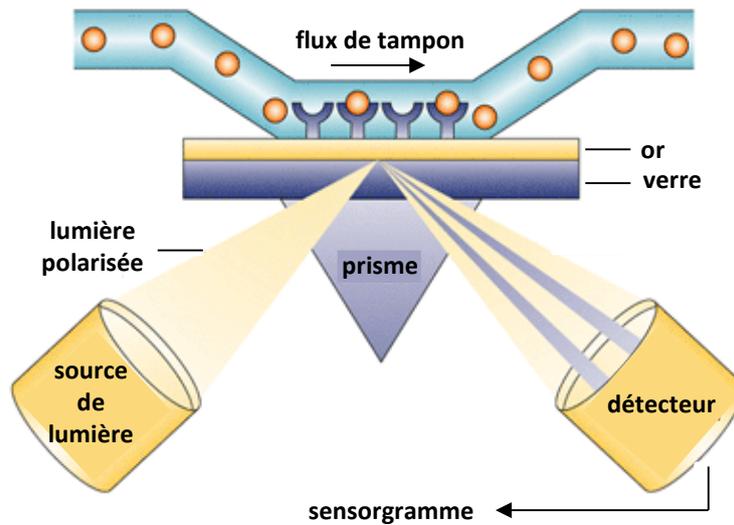


Figure 111 : principe de la SPR

I.2. Principe du BLI

L'interférométrie bio-couche (BLI) est une technique d'analyse optique permettant également de mesurer des interactions biomoléculaires. Contrairement à la SPR, cette technique n'implique pas de fluide, les édifices G-quadruplexes sont immobilisés sur la pointe d'un biocapteur *via* une interaction biotine-streptavidine, la biotine étant présente sur l'édifice G-quadruplexe et la streptavidine sur la surface du capteur. Le biocapteur est ensuite plongé dans des solutions de ligands de G-quadruplexe à différentes concentrations, contenues dans une plaque de 96 puits. Tout comme la SPR, une référence est utilisée, il s'agit dans ce cas d'un capteur non fonctionnalisé. Cette méthode consiste à analyser le diagramme d'interférence de la lumière blanche, réfléchi par deux surfaces : une couche d'édifice G-quadruplexe immobilisé sur le biocapteur et une couche de référence interne (figure 112.a). Toute modification du nombre de molécules liées à la pointe du biocapteur entraîne un changement d'interférence et peut-être mesurée en temps réel. Ainsi, la liaison d'un ligand sur l'édifice G-quadruplexe produit un changement d'épaisseur de la couche à la surface du biocapteur (figure 112.b), ce qui se traduit par un décalage de la longueur d'onde (figure 112.c). Seules les molécules qui se lient ou se dissocient du biocapteur peuvent modifier le diagramme d'interférence et générer un profil de réponse qui est ensuite enregistré par un logiciel de contrôle sous forme de sensorgramme (unité de longueur d'onde en nm en fonction du temps).

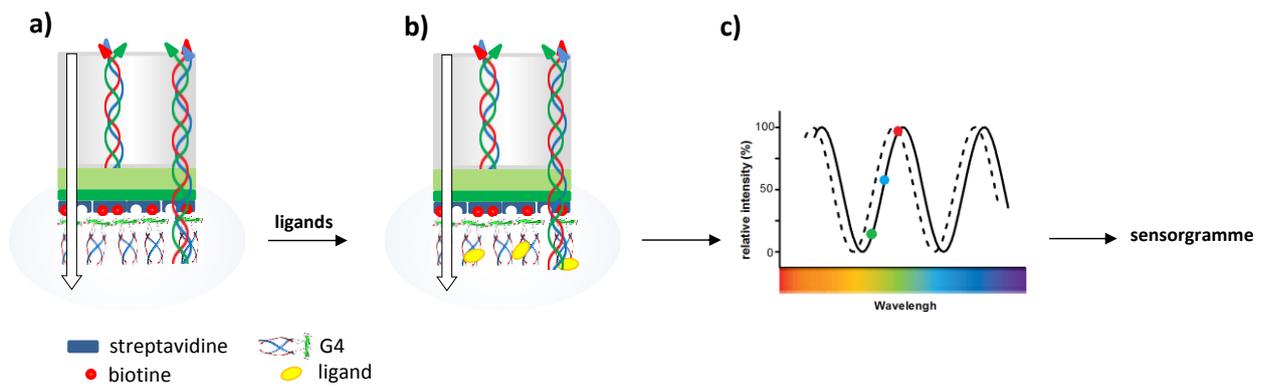


Figure 112 : principe du BLI : (a) pointe d'un biocapteur comportant des G-quadruplexes, (b) pointe d'un biocapteur comportant des G-quadruplexes liés à des ligands, (c) diagramme représentant un décalage de la longueur d'onde

I.3. Traitement des données

Les sensorgrammes enregistrés en SPR ou en BLI présentent trois phases distinctes (figure 113). La première est l'association, phase durant laquelle le ligand va se lier à la cible G-quadruplexe. La seconde est une phase d'équilibre thermodynamique entre le ligand et la cible où les vitesses d'association et de dissociation sont identiques. La troisième phase, permet la dissociation des complexes ligand-cible lors du rinçage.

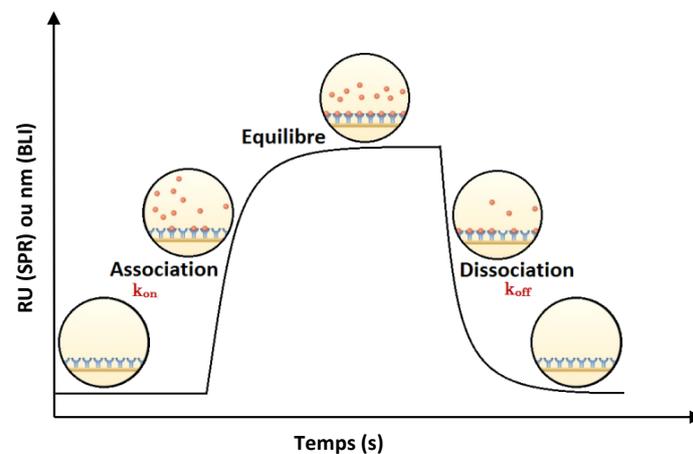


Figure 113 : représentation d'un sensorgramme

Durant cette thèse, l'interprétation des sensorgrammes a été réalisée en collaboration avec les physicochimistes de l'équipe (Laure Bar et Jérôme Dejeu) en utilisant un modèle hétérogène d'interactions. Ce modèle tient compte de la présence de deux interactions, une interaction spécifique correspondant à la liaison du ligand sur la cible G-quadruplexe et une interaction non spécifique, correspondant par exemple à l'association de plusieurs ligands les uns sur les autres par π -stacking. La modélisation des sensorgrammes par les logiciels BIAcore® (pour la SPR) ou Octet®

(pour le BLI) *via* des modèles mathématiques peut être faite par deux approches. La première utilise un modèle cinétique et conduit à l'obtention des constantes cinétiques d'association (k_{on}) et de dissociation (k_{off}), ce qui permet de calculer un K_D . La deuxième consiste en un modèle thermodynamique et permet d'obtenir la constante d'équilibre de dissociation (K_D). Ce paramètre reflète l'affinité d'un ligand pour sa cible. Il correspond à la concentration nécessaire en ligand pour occuper 50% des sites de la cible. Plus la valeur du K_D est faible, meilleure est l'interaction.

Afin d'étudier les interactions des ligands avec des motifs G-quadruplexes provenant de la même séquence ou ayant des topologies semblables, nous avons effectué trois études distinctes. La première a été réalisée avec les édifices ARN contraints **C** et **E**, la seconde avec l'édifice HIV-PRO1 **G** et l'édifice télomérique de type basket **B** et la dernière avec les édifices triplexes **H** et **I**. Pour chaque étude, nous avons utilisé les petites molécules présentées en introduction (p 27) qui sont représentatives des grandes familles des ligands de G-quadruplexes.

II. Evaluation de l'affinité et de la sélectivité de ligands pour les édifices G-quadruplexes ADN parallèle **A**, ARN parallèle **C** et hybride ADN/ARN **E** par SPR

Cette étude vise à comparer les interactions des ligands TmPyP4, MMQ1, Braco-19 et L8 avec les conjugués **A** (G4 ADN parallèle), **C** (G4 ARN parallèle) et **E** (G4 hybride ADN/ARN) issus de la région télomérique (figure 114). Ces édifices ne présentent pas de boucles et sont tous en conformation parallèle. Ils devraient permettre d'examiner l'éventuelle sélectivité entre G-quadruplexes d'ADN, d'ARN et hybride ADN/ARN.

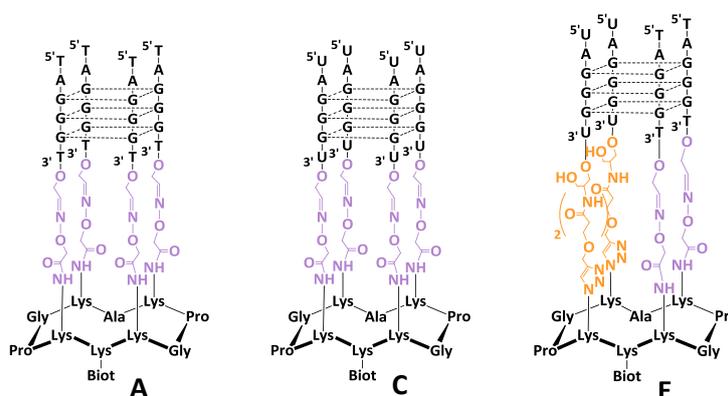


Figure 114 : structure des édifices G-quadruplexes contraints utilisés en SPR

Les constantes de dissociations (K_D) obtenues par SPR, pour chaque édifice avec chacun des quatre ligands sont présentées dans le tableau ci-après (tableau 8). Les analyses ont été réalisées dans un tampon salin (10 mM HEPES à pH 7,4, 50 mM NaCl, 100 mM KCl et 0,005% de P20), qui stabilise les motifs G-quadruplexes dans leur conformation parallèle.

Ligands	G4 ADN parallèle A K_D en nM	G4 ARN parallèle C K_D en nM	G4 hybride ADN/ARN E K_D en nM
TmPyP4	317 ± 34	197 ± 21	348 ± 30
MMQ1	722 ± 63	1449 ± 120	882 ± 70
BRACO-19	8 ± 4	261 ± 31	79 ± 11
L8	290 ± 31	543 ± 58	306 ± 33

Tableau 8 : constante d'équilibre (K_D) des ligands vis-à-vis des édifices G-quadruplexes **A**, **C** et **E**

Les constantes d'équilibre de dissociation obtenues avec TmPyP4 sont du même ordre de grandeur pour les trois cibles et peuvent-être considérées comme similaires. Ce ligand présente donc très peu de sélectivité pour un G-quadruplexe donné. Il est généralement admis que TmPyP4 interagit préférentiellement par π -stacking sur les tétrades extérieures des motifs G-quadruplexes, qui sont fortement similaires d'une cible à l'autre. Ceci pourrait expliquer l'absence de sélectivité entre G-quadruplexes d'ARN, d'ADN et hybride ADN/ARN.

Les trois autres ligands : MMQ1, BRACO-19 et L8, présentent quant à eux des différences d'affinité notables entre les cibles. Ils possèdent tous les trois une meilleure affinité pour les G-quadruplexes d'ADN par rapport aux G-quadruplexes d'ARN, ce qui suggère l'existence d'interactions spécifiques entre les motifs G-quadruplexes et les chaînes latérales des ligands. Il est proposé que ces molécules interagissent par π -stacking avec les tétrades extérieures des G-quadruplexes mais également par liaisons dans les sillons. Ces sillons étant différents d'une cible à l'autre, cela pourrait expliquer les affinités différentes.

Bien que la sélectivité soit relativement faible pour MMQ1 et L8 (K_D multiplié par deux), l'affinité de BRACO-19 semble augmenter significativement avec le nombre de brins d'ADN présent dans les motifs G-quadruplexes. En effet, elle est de 261 nM pour l'édifice ARN, de 79 nM pour l'édifice hybride possédant deux brins d'ADN et de 8 nM pour l'édifice **A** présentant quatre brins d'ADN.

En 2010, l'équipe de G. N. Parkinson¹⁶ a démontré que les groupements 2'hydroxyle des motifs G-quadruplexes d'ARN étaient capables d'interagir avec l'intérieur de la structure G-quadruplexe, modifiant ainsi la taille des sillons. La plus faible affinité de BRACO-19 pour les G-

quadruplexes d'ARN pourrait ainsi être liée aux contraintes stériques imposées par les groupements 2'hydroxyle. Il est possible que les chaînes latérales du ligand ne puissent pas pénétrer entièrement dans les sillons des G-quadruplexes d'ARN par rapport aux G-quadruplexes d'ADN. La plus grande affinité de BRACO-19 pour l'ADN télomérique par rapport à l'ARN a déjà été rapportée par l'équipe de S. Neidle¹⁵⁵. Ils ont démontré par des expériences de dynamiques moléculaires la présence de répulsions électroniques entre l'oxygène de la fonction carbonyle présente sur les deux chaînes latérales du ligand et le groupe 2'hydroxyle d'un motif G-quadruplexe. Ainsi pour des raisons de gêne stérique et de répulsions électrostatiques BRACO-19 possède une plus faible affinité pour l'édifice **C**.

La valeur de K_D obtenue pour l'édifice hybride ADN/ARN est intermédiaire à celle des édifices ADN et ARN. Ceci peut s'expliquer par la présence d'interactions forte entre les chaînes latérales de BRACO-19 et les sillons des brins ADN et d'interactions plus faibles entre les chaînes latérales du ligand et les sillons des brins ARN.

Ces résultats mettent en évidence que l'accessibilité des éléments de reconnaissance des G-quadruplexes est différente suivant les cibles ADN, ARN ou hybride ADN/ARN, et tendent à démontrer que le ciblage spécifique de motif ARN ou hybride est envisageable.

III. Evaluation de l'affinité et de la sélectivité de ligands pour les édifices G-quadruplexes de type « basket » **B et HIV-PRO1 **G** par SPR**

Dans cette deuxième étude, nous avons comparé les interactions des ligands TmPyP4, MMQ1, Braco-19, L8, L14, PDS et PhenDC3 avec les conjugués **B** (G4 télomérique de type basket), et **G** (G4 HIV-PRO1) présentant deux boucles latérales (figure 115). Afin d'évaluer la sélectivité des ligands pour ces deux édifices par rapport au duplexe, un édifice contrôle présentant un duplexe d'ADN a également été utilisé.

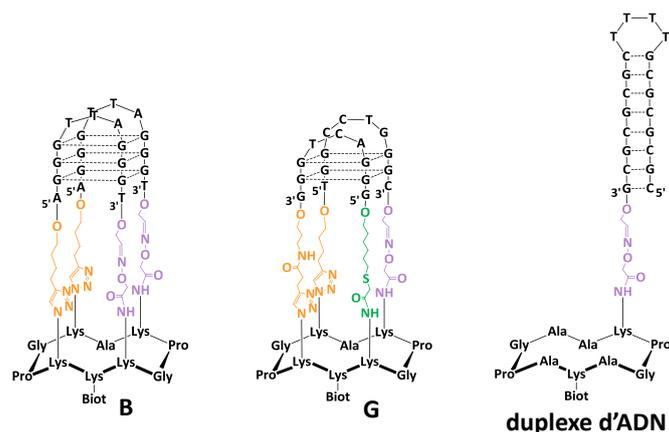


Figure 115 : structure des édifices contraints utilisés en SPR

Les constantes de dissociations (K_D) obtenues par SPR pour chaque édifice avec chacun des sept ligands sont présentées dans le tableau ci-après (tableau 9). Les analyses ont été réalisées dans un tampon salin (10 mM HEPES à pH 7,4, 50 mM NaCl, 100 mM KCl et 0,5% de P20).

Ligands	G4 télomérique basket B K_D en nM	G4 HIV-PRO1 G K_D en nM	duplexe d'ADN K_D en nM
TmPyP4	679 ± 49	2095 ± 361	702 ± 121
MMQ1	4486 ± 1941	1939 ± 820	4886 ± 1293
BRACO-19	307 ± 54	55 ± 9	1169 ± 97
L8	764 ± 84	633 ± 87	2901 ± 971
L14	777 ± 125	1356 ± 415	267 ± 103
PDS	12 ± 6	24 ± 7	nd
PhenDC3	2,4 ± 0,3	2,1 ± 0,1	nd

Tableau 9 : constante d'équilibre (K_D) des ligands vis-à-vis des édifices G-quadruplexes **B**, **G** et du duplexe d'ADN

TmPyP4 et MMQ1 ne présentent pas de sélectivité pour les édifices **B** et **G** par rapport au duplexe d'ADN. Les valeurs de K_D obtenus pour ces deux ligands avec l'édifice G4 télomérique **B** sont en accord avec les valeurs rapportées pour la structure G-quadruplexe télomérique native¹²⁰. TmPyP4 possède environ trois fois moins d'affinité pour l'édifice G4 HIV-PRO1 **G**, que pour l'édifice **B** et le duplexe d'ADN. Ceci s'explique par le fait que ce ligand interagit essentiellement par π -stacking sur les tétrades supérieures des motifs G-quadruplexes. Or l'édifice **G** possède une boucle qui présente une paire de base C-G stable qui pourrait réduire l'accessibilité de la tétrade externe au ligand.

BRACO-19 affiche une très forte sélectivité des motifs G-quadruplexes par rapport au duplexe d'ADN. Il est également celui qui discrimine le mieux les deux édifices G-quadruplexes en présentant respectivement des K_D de 307 nM et de 55 nM pour les G4 télomérique **B** et HIV-PRO1 **G**. Cette affinité, environ six fois supérieure pour l'édifice **G** pourrait être rationalisée par la présence d'interactions favorables entre les deux chaînes flexibles du ligand et les sillons du motif G-quadruplexe. En effet, les deux motifs G-quadruplexes présentent des sillons différents. Le motif G4 HIV-PRO1 **G** comporte deux sillons larges opposés et deux sillons étroits opposés, tandis que le motif G4 télomérique **B** possède deux sillons moyens et un sillon large opposé à un sillon étroit (figure 116).

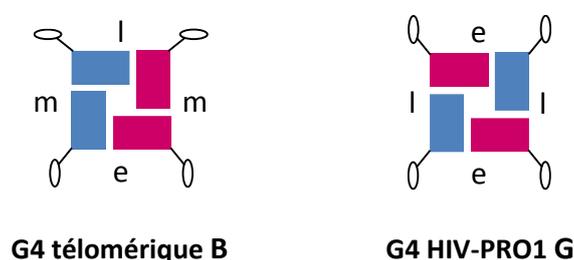


Figure 116 : représentation des sillons des motifs G-quadruplexe télomérique et HIV-Pro1, avec l = large, e = étroit et m = moyen, les rectangles rose et bleu représentent respectivement les guanines en configuration *syn* et *anti*

Même si seule une analyse structurale permettrait de définir le mécanisme de liaison de BRACO-19, nous pouvons émettre l'hypothèse que les chaînes latérales de ce dernier s'adaptent mieux aux sillons larges opposés de la structure G4 HIV-PRO1 **G**. Quel que soit son mode de liaison, BRACO-19 est particulièrement intéressant en termes d'affinité et de sélectivité vis-à-vis du G4 HIV-PRO1, ce qui conforte les études en cours lui attribuant une activité antivirale prometteuse⁶⁵.

L8 comporte des chaînes latérales d'alkylimidazolium pouvant créer des interactions électrostatiques avec les sillons des motifs G-quadruplexes. Malgré la différence de taille des sillons entre les deux motifs G-quadruplexes, ce ligand présente des constantes d'équilibre de dissociation quasiment identiques pour les édifices **B** et **G**, respectivement 764 nM et 633 nM. L8 ne permet pas de discriminer ces deux motifs G-quadruplexes, contrairement à BRACO-19.

Bien que L8 et L14 ne se différencient que par la position ortho ou para de leurs bras alkylimidazolium, ils présentent des interactions différentes vis-à-vis des motifs G-quadruplexes et du duplexe d'ADN. En effet, L8 possède une certaine sélectivité pour les G-quadruplexes par rapport au duplexe d'ADN, tandis que L14 n'est pas du tout sélectif, ce dernier s'associe préférentiellement avec le duplexe d'ADN ($K_D = 267$ nM). La position ortho ou para des chaînes latérales semblent donc jouer

un rôle important dans l'interaction des ligands avec les motifs G-quadruplexes. Il semblerait que les bras de L8 en position para, soient assez espacés pour interagir avec deux sillons des motifs G-quadruplexes. A l'inverse, les chaînes latérales de L14 en position ortho semblent trop resserrées pour pouvoir interagir avec deux sillons.

Les ligands PDS et PhenDC3 sont connus pour être fortement stabilisants et sélectifs pour les motifs G-quadruplexes par rapport au duplexe d'ADN, ce qui a été une nouvelle fois confirmé par cette étude SPR. En effet, dans les concentrations étudiées aucun signal n'a été observé sur le duplexe d'ADN en présence de ces deux ligands. Ces derniers présentent de très forte affinité, de l'ordre du nanomolaire vis-à-vis des édifices G4 **B** et **G**. PhenDC3 est le plus affin avec des K_D de respectivement 2,4 nM et 2,1 nM pour les édifices G4 télomérique **B** et HIV-PRO1 **G**. En revanche aucune discrimination n'est observée entre les deux motifs G-quadruplexes puisque les K_D obtenus sont identiques. Il en est de même pour PDS qui présente une affinité de 12 nM pour l'édifice **B** et de 24 nM pour l'édifice **G**. Bien que PDS et PhenDC3 ne soient pas sélectifs pour une conformation G-quadruplexe donnée, l'analyse de leur mode de liaison révèle des aspects intéressants de leur interaction avec les motifs G-quadruplexes. Ils présentent de grandes divergences dans leur cinétique d'association et de dissociation, ce qui suggère des différences marquées dans leur mode d'interaction.

Les paramètres de dissociation et d'association de chaque ligand pour les édifices **B** et **G** sont représentés dans le diagramme ci-dessous (figure 117).

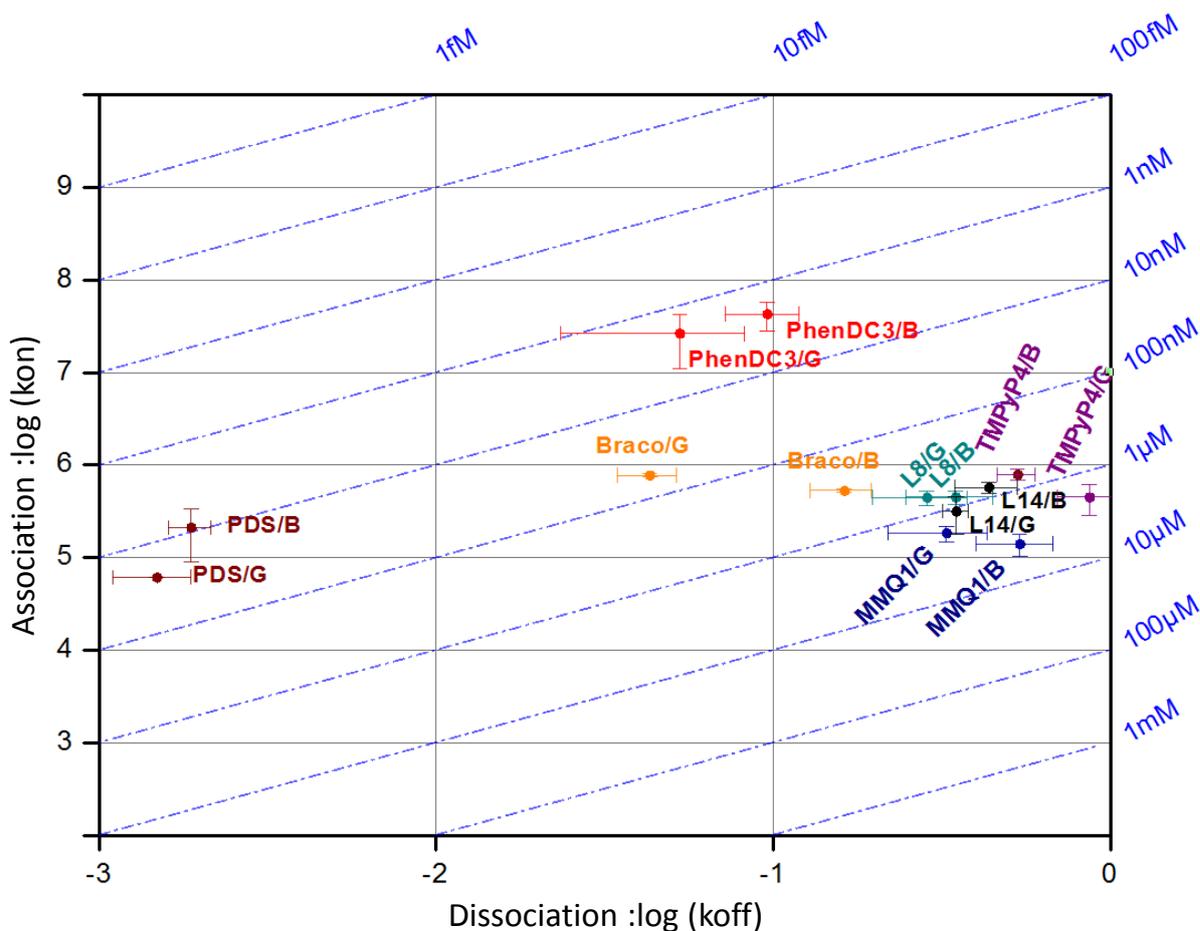


Figure 117 : diagramme des constantes de cinétiques d'interaction entre les édifices **B** et **G** et les différents ligands étudiés

PhenDC3 présente les constantes d'association les plus élevées ($k_{on} \approx 3.10^7 \text{ M}^{-1} \text{ s}^{-1}$). PDS affiche quant à lui des constantes de dissociation très lentes ($K_{off} \approx 2.10^{-3} \text{ s}^{-1}$), qui sont cinquante fois plus faibles que celles de PhenDC3. Cette dissociation très lente de PDS pourrait s'expliquer par la présence d'interactions électrostatiques, fortement stabilisante entre les chaînes protonables du ligand et les sillons des G-quadruplexes chargés négativement. PDS a déjà été utilisé en tant que sonde pour l'indentification de motifs G-quadruplexes d'ADN et d'ARN au sein du génome¹⁵⁶ et du transcriptome¹⁵⁷ et également pour visualiser quantitativement les G-quadruplexes d'ADN dans les cellules humaines³⁹. Il est donc fortement envisageable que ces études aient pu être réalisées grâce à la capacité de PDS à établir des interactions durables avec les structures G-quadruplexes. Ce ligand qui se dissocie très difficilement pourrait ainsi être utilisé comme inhibiteur de certains processus biologiques, *via* une stabilisation durable des quadruplexes.

L'association très rapide de PhenDC3 sur les motifs G-quadruplexes pourrait permettre d'atteindre en très peu de temps des concentrations thérapeutiques (ligands en tant que médicament) ou fonctionnelles (ligand en tant que sonde) significatives. De plus PhenDC3 pourrait

être un très bon candidat pour des activités diagnostiques en permettant une évaluation rapide du taux de G-quadruplexe dans un milieu.

Le diagramme permet également de mettre en évidence la différence de comportement de BRACO-19 vis-à-vis des édifices **B** et **G**. Il semblerait que le ligand se lie avec la même vitesse sur les deux édifices ($k_{on} \approx 6 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$), mais qu'il se dissocie cinq fois moins vite du G4 HIV-PRO1 ($K_{off} \approx 4 \cdot 10^{-2} \text{ s}^{-1}$) que du G-quadruplexe télomérique ($K_{off} \approx 2 \cdot 10^{-1} \text{ s}^{-1}$), ce qui explique sa plus forte affinité vis-à-vis du G4 HIV PRO1 **G**.

Pour résumer, ces résultats révèlent ainsi la sélectivité de BRACO-19 pour le motif G4 HIV-PRO1 **G**, il est le seul des ligands étudiés à discriminer *via* une différence d'affinité significative les deux motifs G4 **B** et **G**. L'étude des constantes cinétiques d'association et de dissociation a mis en évidence des modes d'interactions différents entre les ligands. Cette relation structure-cinétique est très intéressante puisque la reconnaissance moléculaire (décrite par la constante d'association) et la stabilité du complexe (décrite par la constante de dissociation) sont des éléments pertinents dans la prédiction de l'efficacité des médicaments ou des sondes chimiques.

IV. Evaluation de l'affinité et de la spécificité de ligands pour les édifices G-triplexe H, G-quadruplexe I et duplexe d'ADN, par BLI

Cette troisième étude vise à comparer les interactions des ligands TmPyP4, Braco-19, PDS et PhenDC3 avec les conjugués **H** (G3 télomérique), **I** (G4 télomérique) et un duplexe d'ADN (figure 118). Avant la réalisation de cette étude d'interaction, le laboratoire a fait l'acquisition d'un appareil mesurant des interactions de surface entre deux biomolécules par BLI. Des études expérimentales ont démontré que les données obtenues par SPR et BLI sont identiques. La méthode d'analyse par BLI étant plus facile à mettre en œuvre, nous avons décidé de l'utiliser, plutôt que la SPR, pour réaliser l'étude des interactions triplexes-ligands.

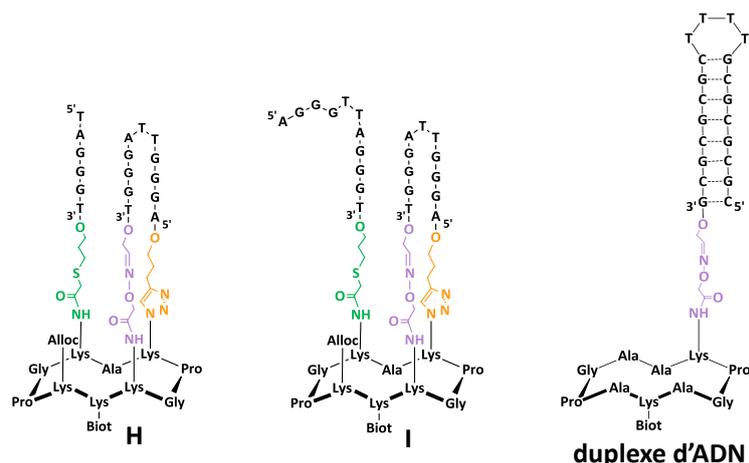


Figure 118 : structure des édifices G-quadruplexes contraints utilisés en BLI

Les constantes de dissociations (K_D) obtenues par BLI pour chaque édifice avec chacun des quatre ligands sont présentées dans le tableau ci-dessous (tableau 10). Les analyses ont été réalisées dans un tampon salin (10 mM HEPES à pH 7,4, 50 mM NaCl, 100 mM KCl et 0,005% de P20).

Ligands	G3 télomérique H K_D en nM	G4 télomérique I K_D en nM	duplexe d'ADN K_D en nM
TmPyP4	139 ± 4	247 ± 43	110 ± 3
BRACO-19	318 ± 9	288 ± 42	823 ± 140
PDS	38 ± 13	96 ± 16	nd
PhenDC3	5,3 ± 0,2	14 ± 2	nd

Tableau 10 : constante d'équilibre (K_D) des ligands vis-à-vis des édifices G3 **H**, et G3 **I** et du duplexe d'ADN

TmPyP4 ne présente pas de sélectivité des édifices **H** et **I** par rapport au duplexe d'ADN, contrairement à BRACO-19 qui est plus affiné pour les édifices **H** et **I** (K_D respectivement égal à 318 nM et 288 nM) que pour le duplexe d'ADN ($K_D = 823$ nM). PDS et PhenDC3 sont quant à eux fortement sélectifs des motifs G3 **H** et G4 **I** télomériques par rapport au duplexe d'ADN. Dans les concentrations étudiées aucun signal n'a été observé sur le duplexe en présence de ces deux ligands. Ils sont également les plus affins pour les deux cibles. PhenDC3 est celui qui présente la meilleure affinité avec des K_D de respectivement 5,3 nM et 14 nM pour les édifices G3 **H** et G4 **I**. Ces constantes de dissociation, témoignent d'une très forte association des motifs G3 **H** et G4 **I** par le ligand.

Bien que les quatre ligands étudiés présentent des affinités différentes pour un édifice donné, nous n'observons pas de différence significative d'interaction entre les cibles **H** et **I** pour un ligand en particulier. Pour chacun, les constantes de dissociation obtenues pour les édifices **H** et **I** sont très proches et peuvent être considérées comme similaire. TmPyP4, BRACO-19, PDS et PhenDC3

se lie donc avec la même affinité au motif G-triplexe télomérique qu'au motif G-quadruplexe télomérique. Cette étude démontre que le ciblage et potentiellement la stabilisation de G-triplexes est envisageable.

V. Bilan

Ces études d'interactions nous ont permis d'étudier l'interaction des édifices contraints avec différentes familles de ligands de G-quadruplexes, les salophens, les porphyrines, les dérivés d'acridine et les dérivés de bisquinolinium. Nous avons montré que quel que soit la famille, les ligands possèdent tous une bonne affinité pour les G-quadruplexes ou G-triplexe contraints. Certains ont cependant montré une affinité pour le duplexe (TmPyP4, MMQ1 et L14), de l'ordre du micromolaire. Les autres ligands, tels que BRACO-19, L8, PDS et PhenDC3 sont peu ou pas affins vis-à-vis du duplexe d'ADN et présentent donc une bonne sélectivité pour les G-quadruplexes et G-triplexe. PDS et PhenDC3 possèdent les meilleures sélectivités et affinités pour les cibles contraintes, de l'ordre du nanomolaire, mais comportent des constantes cinétiques d'interaction totalement différentes. PhenDC3 reconnaît rapidement un motif G-quadruplexe ($k_{on} \approx 3 \cdot 10^7 \text{ M}^{-1} \text{ s}^{-1}$), alors que PDS se dissocie très lentement ($K_{off} \approx 2 \cdot 10^{-3} \text{ s}^{-1}$). Parmi toutes les molécules testées, BRACO-19 est celle qui discrimine le mieux les différentes topologies G-quadruplexes. Il a par exemple démontré une sélectivité pour l'édifice G4 ADN parallèle **A** par rapport à l'édifice G4 ARN parallèle **C**, ou encore pour l'édifice G4 HIV-PRO1 **G** par rapport à l'édifice G4 télomérique basket **B**.

Partie II : Développement d'une nouvelle classe de ligands de G-quadruplexe : les aptamères

Je tiens ici à associer Emmanuelle Fiore, technicienne au DPM qui a largement participé à ces travaux.

I. Introduction

Cette deuxième partie est consacrée au développement d'une nouvelle classe de ligands pour le ciblage d'acides nucléiques structurés en motif G-quadruplexe. L'objectif de cette approche est d'obtenir un ou plusieurs ligands de G-quadruplexes, spécifiques d'une topologie donnée.

L'idée a été d'utiliser comme nouvelle classe de ligands, les aptamères, qui sont des séquences simples brins d'ADN ou d'ARN, modifiées ou non, issues d'un processus de sélection *in vitro*, nommée SELEX (« Systematic Evolution of Ligands by EXponential enrichment »). Ces aptamères sont des acides nucléiques d'environ 50 à 100 nucléotides, capables d'adopter des structures tridimensionnelles variées pour reconnaître spécifiquement leur cible. L'affinité entre l'aptamère et sa cible est généralement très élevée, de l'ordre du nanomolaire, comparable à celle d'un anticorps pour sa cible.

Les cibles des aptamères sont très variées, elles peuvent être des bactéries¹⁵⁸, des virus¹⁵⁹, des protéines¹⁶⁰, des petites molécules¹⁶¹, ou encore des acides nucléiques. Le ciblage de ces derniers est plutôt rare, cependant l'équipe de J. J. Toulmé a identifiée plusieurs aptamères qui interagissent par des interactions de types « kissing » avec leurs cibles ADN ou ARN¹⁶². La sélection d'aptamère contre les cibles acides nucléiques est rendue difficile par l'émergence de séquences présentant la séquence complémentaire à leur cible. L'interaction n'est alors pas due à une reconnaissance structurale mais à une hybridation de type Watson-Crick entre la cible et le ligand. Dans notre cas, l'utilisation d'un édifice G-quadruplexe contraint devrait empêcher la sélection de séquences complémentaires et favoriser l'obtention d'aptamères reconnaissant spécifiquement la structure du motif G-quadruplexe.

I.1. Principe de la SELEX classique

Les aptamères sont obtenus par un processus de sélection *in vitro* d'acides nucléiques par la méthode SELEX. La première sélection d'acides nucléiques a été réalisée en 1990 par L. Gold¹⁶³

contre l'ADN polymérase du bactériophage T4. La méthode SELEX repose sur la sélection des aptamères les plus affins pour une cible, à partir d'une vaste banque d'oligonucléotides. Le principe général, présenté dans le schéma 37, consiste à incuber la cible avec une banque d'oligonucléotides et à réaliser plusieurs cycles de sélection durant lesquels les aptamères les plus affins pour la cible sont isolés, amplifiés par PCR, purifiés puis ré incubés avec la cible. Au fur et à mesure des cycles de sélections, les conditions deviennent de plus en plus drastiques (augmentation de la température, diminution de la concentration en cible), ce qui permet de ne sélectionner que les aptamères les plus affins. A la fin du processus de sélection les potentiels candidats aptamères sont séquencés puis clonés ou simplement séquencés par séquençage haut débit.

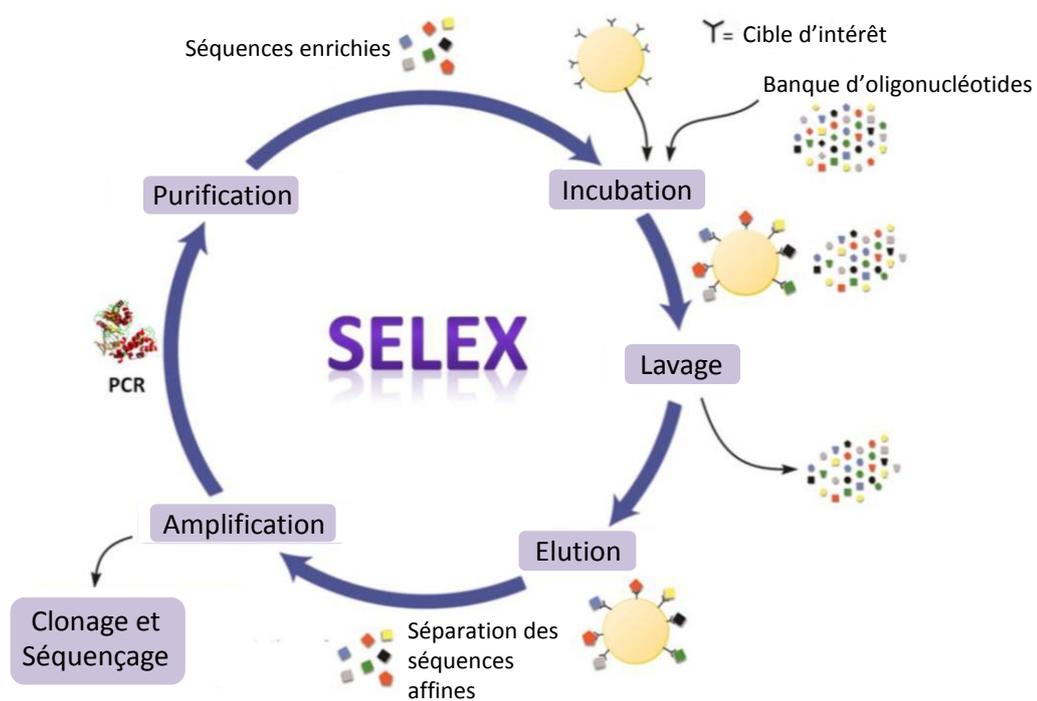


Schéma 37 : principe de la SELEX classique

I.2. Principe de la CE-SELEX

La SELEX classique, présentée dans le schéma 37, nécessite l'immobilisation de la cible sur différents supports. Durant cette thèse, nous avons utilisé une méthode de SELEX par électrophorèse capillaire (CE-SELEX) qui utilise l'électrophorèse capillaire pour séparer les séquences affines pour la cible des séquences non affines. La sélection est réalisée en solution, sans immobilisation de la cible. La séparation des séquences a lieu dans un capillaire recouvert de silanoate et comportant un flux électroosmotique à l'intérieur duquel les molécules migrent de l'anode à la cathode, selon un rapport charge sur masse (schéma 38). Les molécules chargées positivement migrent en premier, suivies des molécules neutres (entraînées par le flux électroosmotique) et enfin des molécules

chargées négativement. Pour pouvoir séparer la cible G-quadruplexe de la banque d'oligonucléotides, toutes deux chargées négativement, le capillaire a été rempli d'un tampon de migration constitué de micelles de triton X100. La cible G-quadruplexe, à laquelle nous avons greffé une chaîne C18 peut s'ancrer à l'intérieur des micelles et être entraînée par le flux électroosmotique, permettant ainsi la séparation de la cible G-quadruplexe (migration avant la banque) et de la banque par électrochromatographie micellaire. La CE-SELEX comporte plusieurs avantages par rapport à une méthode de SELEX classique. Elle permet tout d'abord de s'affranchir de la fixation de la cible, évitant ainsi une contre sélection sur le support. Elle apporte également une plus grande accessibilité à la cible et n'entraîne pas de modification de sa topologie. De plus, la séparation des séquences en solution diminue considérablement la formation de liaisons non spécifiques, ce qui entraîne en taux d'enrichissement plus élevé et par conséquent la réduction du nombre de cycle de sélection à effectuer par rapport à une SELEX classique.

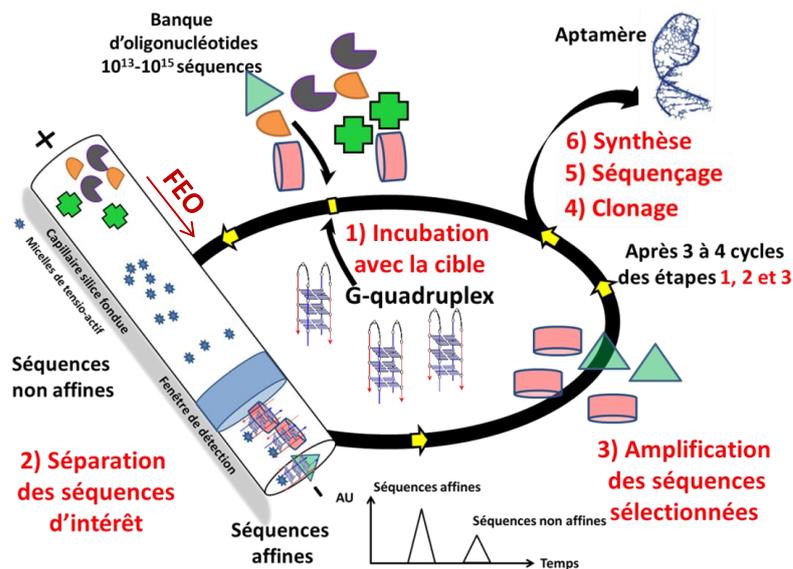


Schéma 38 : principe de la CE-SELEX

La CE-SELEX a été utilisée pour la première fois en 2004 par S. D. Mendonsa et M. T. Bowser contre des anticorps IgE¹⁶⁴. Depuis elle a servi à la sélection d'aptamères contre des cibles plutôt variées, comme par exemple la transcriptase inverse du VIH-1¹⁶⁵, un neuropeptide Y¹⁶⁶ ou encore une petite molécule de la famille des porphyrines¹⁶⁷.

I.3. Choix de la cible

Nous avons choisi comme première cible le motif G-quadruplexe de type « basket » de l'ADN télomérique (figure 119). De par sa topologie et la présence de boucles latérales, qui sont des éléments de reconnaissance important pour la fixation des ligands, ce motif G-quadruplexe est un

candidate idéal pour une sélection d'aptamères. Le choix de ce motif repose également sur sa localisation dans le génome humain. Comme évoqué précédemment, les motifs G-quadruplexes de la région télomérique présentent des cibles intéressantes dans les stratégies anticancéreuses. Il est donc pertinent d'essayer de sélectionner un ligand spécifique d'un motif G-quadruplexe présent dans cette région. De plus, la disponibilité des motifs G-quadruplexes contraints **A**, **C**, **E** et du triplex **H** dérivant également de la région télomérique nous permettra de comparer l'affinité des aptamères sélectionnés pour différents G-quadruplexes de séquences identiques mais de topologies variées.

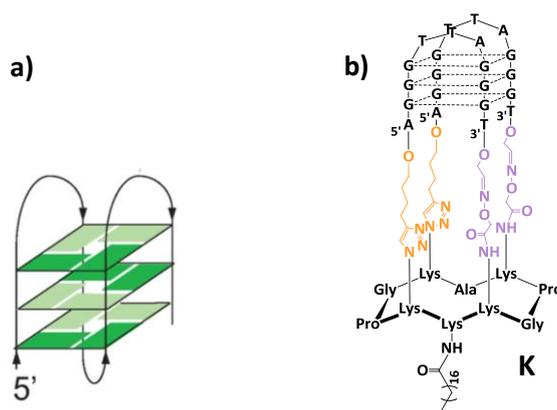


Figure 119 : (a) représentation schématique du motif G-quadruplexe de type « basket » de la région télomérique (séquence : d(A(GGGTTA)₃GGG)) et (b) structure chimique de la cible

II. Synthèse de la cible G-quadruplexe

Le motif G-quadruplexe de type « basket » de la région télomérique (figure 119), a déjà été mimé au laboratoire avec l'édifice **B**¹¹⁹ mais la présence d'une fonction biotine sur le châssis peptidique, ne nous a pas permis de l'utiliser comme cible SELEX. Nous avons donc assemblé l'édifice **K** (figure 119), présentant les mêmes boucles que l'édifice **B** et une chaîne C18 sur la face inférieure du châssis.

II.1. Synthèse du châssis peptidique 22

Le châssis peptidique comporte deux fonctions oxyamines et deux fonctions azotures nécessaires au greffage des oligonucléotides. Il présente également une fonction amine libre sur sa face inférieure, utile à l'ajout de la chaîne C18. Le châssis **22** a été synthétisé sur support solide en utilisant la stratégie de synthèse au Fmoc/tBu et la résine 2-chlorotriptyl (schéma 39). Il a été obtenu selon le même protocole qu'auparavant (p 82).

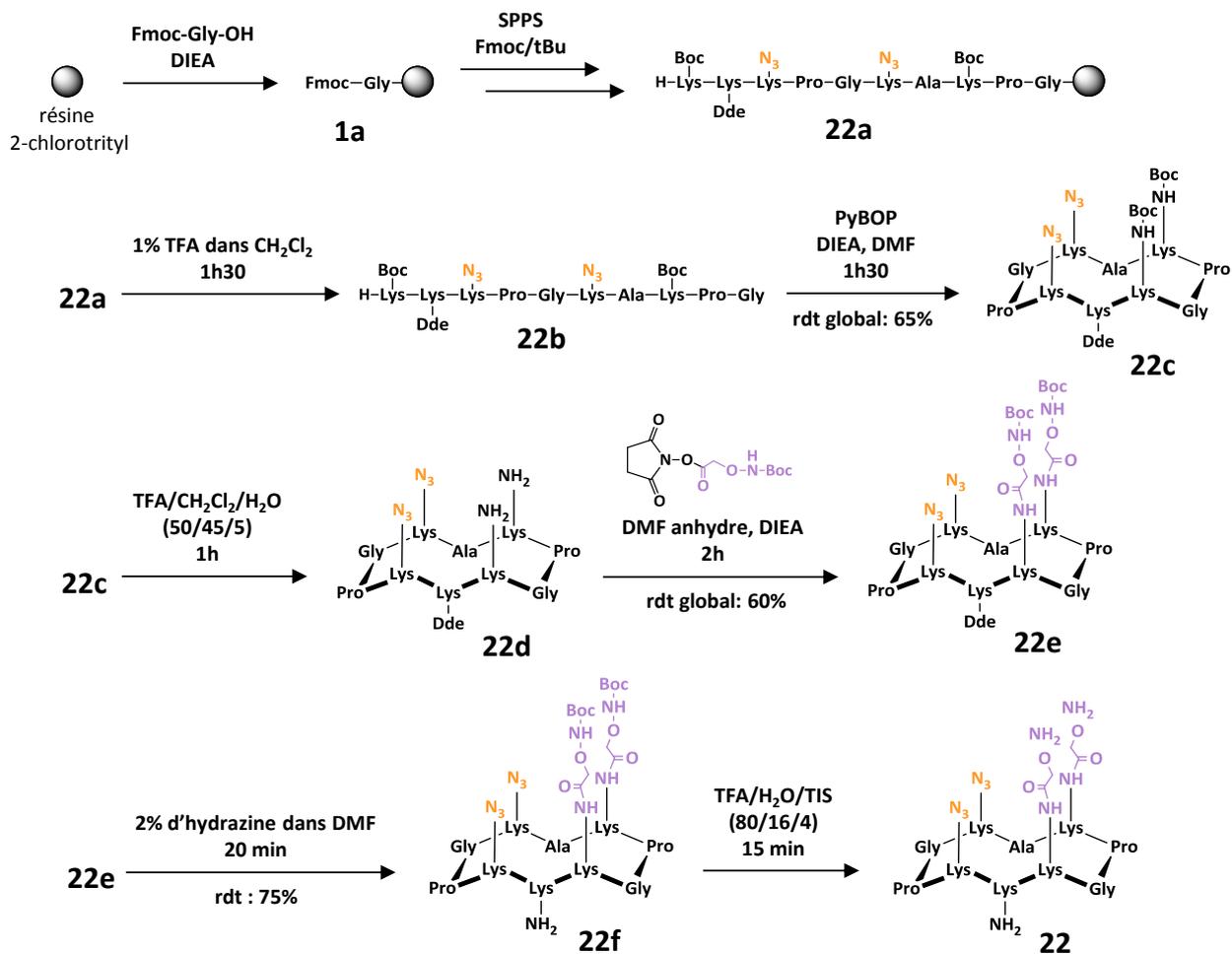
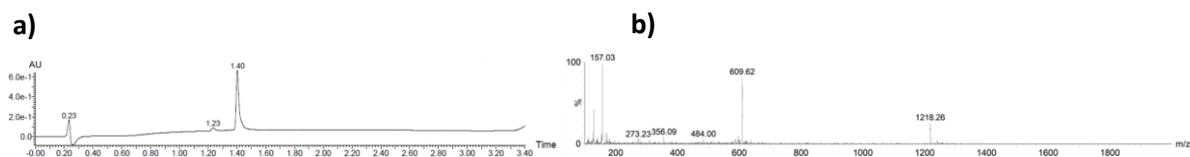


Schéma 39 : synthèse du châssis peptidique **22**

Le peptide final a été caractérisé par UPLC-MS (figure 120).



ESI-MS(+) m/z calculé: 1218,7, m/z trouvé: 1218,3 [M+H]⁺

Figure 120: (a) chromatogramme UPLC et (b) spectre ESI-MS du châssis peptidique **22**

II.2. Synthèse de l'oligonucléotide **16**

L'édifice **K** a été assemblé à partir de l'oligonucléotide **16** (figure 121), utilisé pour la synthèse de l'édifice triplexe. Cet oligonucléotide comporte une fonction alcyne sur son extrémité 5' et une fonction aldéhyde à son extrémité 3' (séquence : ^{5'}YAGGGTTAGGGTX^{3'}, Y = 5'alcyne, X = 3'aldéhyde). Sa synthèse est décrite dans le chapitre 1 p 79.



Figure 121 : structure chimique de l'oligonucléotide **16**

II.3. Assemblage de l'édifice **K**

L'assemblage de l'édifice **K** est réalisé en trois étapes et est présenté dans le schéma 40 ci-dessous.

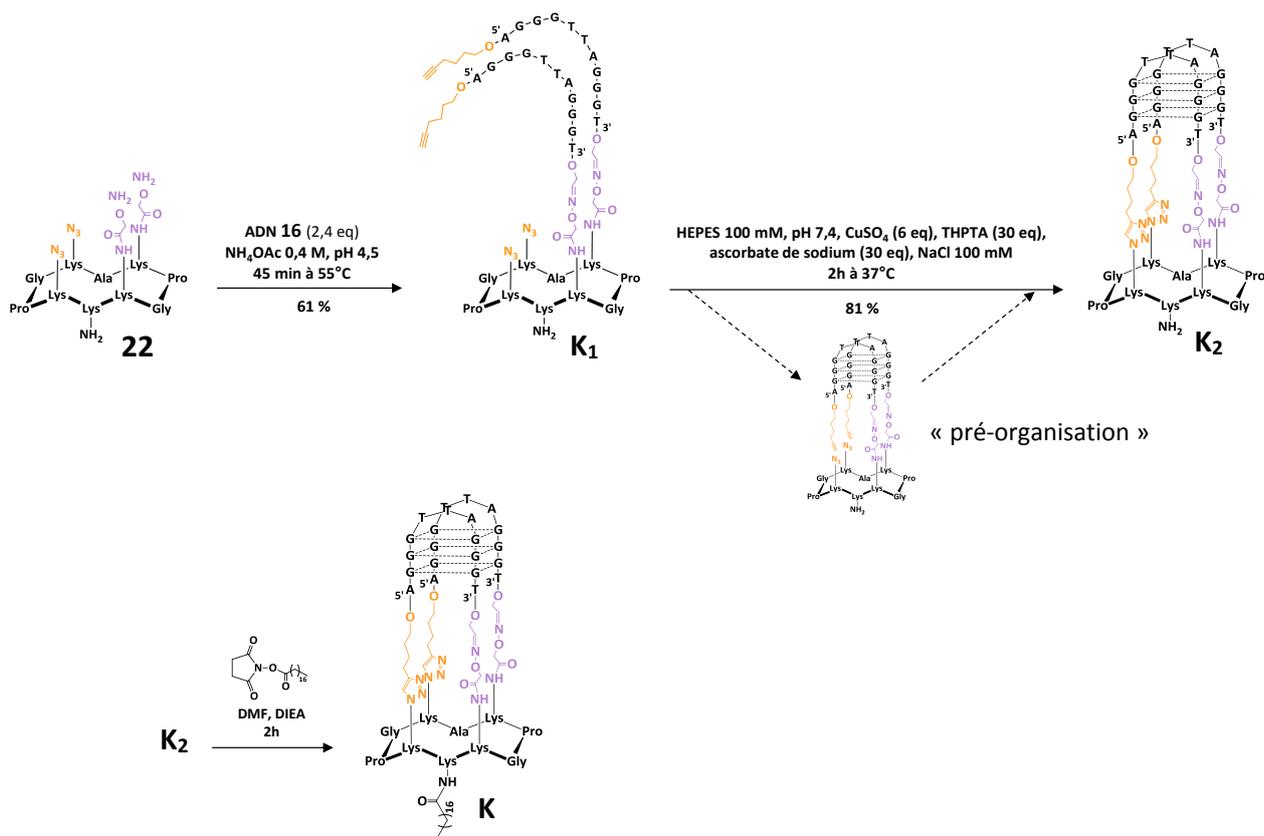


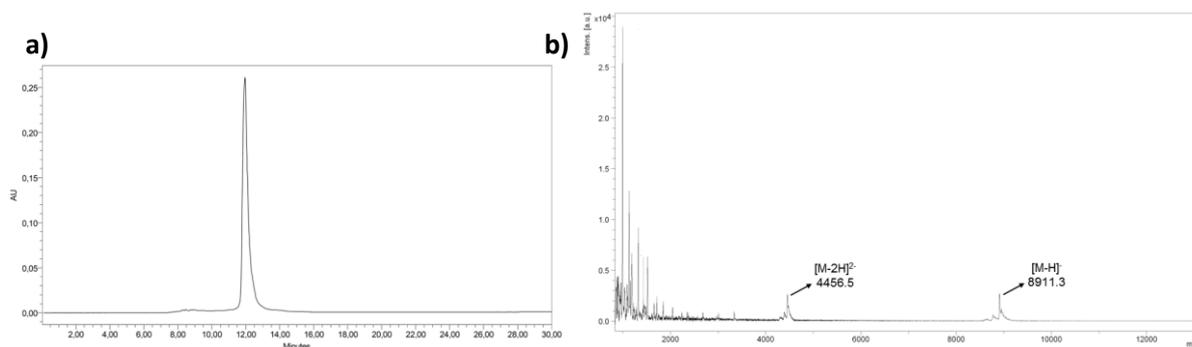
Schéma 40 : synthèse de l'édifice **K**

La première étape consiste à greffer par leurs extrémités 3', les brins d'ADN **16** sur le châssis peptidique. La ligation oxime utilisée pour cette réaction est effectuée dans les mêmes conditions que précédemment (solution tamponnée d'acétate 0,4 M à pH 4,5, 45 min à 55°C). Il est à noter que cette réaction a été réalisée en présence d'une amine libre mais qu'aucune réaction secondaire n'a été observée. La très grande réactivité des oxyamines rend la ligation oxime totalement spécifique.

La deuxième étape permet de refermer simultanément les deux boucles, en liant l'extrémité 5' des oligonucléotides au châssis peptidique. Les conditions réactionnelles classiques sont les mêmes que décrites précédemment (p 91) 100 mM d'HEPES à pH 7,4, CuSO₄ (6 eq), THPTA (30 eq) et ascorbate de sodium (30 eq), concentration à 10⁻⁴ M). Par contre la nature du cation métallique

utilisé pour préformer le motif G-quadruplexe est différente. Dans ce cas, 100 mM de NaCl sont utilisés pour favoriser la conformation « basket antiparallèle » de la séquence télomérique.

La dernière étape consiste à greffer la chaîne C18 activée sous forme d'ester de succinimide sur la fonction amine libre du châssis peptidique. La réaction est considérée quantitative et l'édifice **K** a été caractérisé par HPLC et spectrométrie de masse (figure 122).



MALDI-MS(-) m/z calculé: 8912,1, m/z trouvé: 8911,3 [M-H]⁻

Figure 122 : (a) chromatogramme HPLC et (b) spectre MALDI de l'édifice **K**

III. Sélection avec des bases naturelles

III.1. Détermination des conditions de sélection

III.1.a. Banque utilisée

La banque d'oligonucléotides utilisée comporte 10^{14} séquences différentes, possédant des amorces connues à leurs extrémité 5' et 3' et une partie centrale aléatoire de trente nucléotides.

III.1.b. Choix du tampon d'incubation

Pour faciliter les interactions entre les deux partenaires, la banque est mise en présence de la cible dans du tampon d'incubation, avant la sélection. Celui-ci possède un pH de 7,8 et est constitué de 50 mM de Tris-HCl, de 6 mM de Triton, de 50 mM de NaCl et de 1,5 mM de MgCl₂. La structuration du motif G-quadruplexe dans ces conditions d'incubation a été confirmée par dichroïsme circulaire (figure 123).

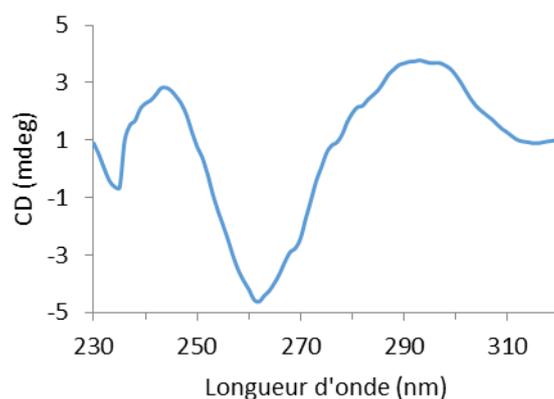


Figure 123 : signal CD de l'édifice K en présence du tampon d'incubation de la SELEX (50 mM de Tris-HCl pH 7,8, 6 mM de Triton, 50 mM de NaCl et 1,5 mM de $MgCl_2$)

III.2. Test d'interaction entre la cible et la banque

Avant de réaliser les cycles de sélection, il est primordial de vérifier s'il y a bien une interaction entre la cible et la banque d'oligonucléotides. Pour quantifier cette interaction, l'électrophorèse capillaire est couplée à un détecteur de fluorescence (EC-LIF) qui mesure la fluorescence émise par la banque marquée par une fluorescéine. La technique consiste à injecter la banque seule, puis la banque en présence de la cible et d'observer un éventuel changement de signal entre les deux injections. La formation d'un complexe oligonucléotide-cible est visualisée par l'obtention d'électrophérogrammes différents. Dans la figure 124 présentée ci-dessous, aucun changement de signal n'est observé entre la courbe rouge représentant la banque seule et la courbe bleue représentant la banque incubée en présence de cible et ce, quel que soit la concentration en cible utilisée (80 μM , 200 μM et 1000 μM de cible testées). Ceci témoigne de l'absence d'interaction entre la cible G-quadruplexe et la banque.

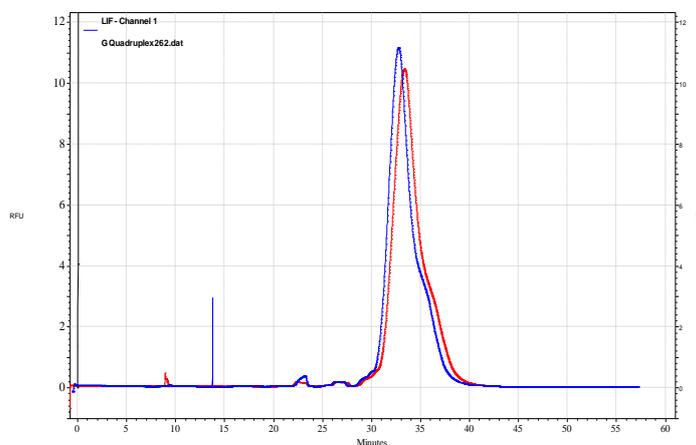


Figure 124 : électrophérogramme EC-LIF de la banque naturelle seule (courbe rouge) et de la banque naturelle incubée en présence de cible à 1000 μM (courbe bleue)

Dans ce contexte, nous avons envisagé l'utilisation d'une banque dans laquelle l'un des nucléotides est remplacé par une base modifiée afin de favoriser les interactions avec le motif G-quadruplexe.

IV. Sélection avec des bases modifiées

L'utilisation de bases modifiées pour favoriser les interactions entre une cible et la banque est de plus en plus employée et permet la sélection d'aptamères très affins^{168,169}. Dans ce travail de thèse, nous avons décidé de remplacer les thymines naturelles par un analogue modifié : la 5-[(3-Indolyl)propionamide-N-allyl] uridine (figure 125). Cette dernière comporte une fonction indole en position C5 qui pourrait créer des interactions hydrophobes supplémentaires avec les oligonucléotides formant le motif G-quadruplexe. Plusieurs sélections d'aptamères, à partir de banques d'oligonucléotides modifiés et comportant des fonctions indoles, ont déjà été réalisées et permettent d'améliorer l'efficacité des aptamères^{170,171}. La banque modifiée a été préparée par amplification de la banque naturelle par PCR, en présence des triphosphates naturels (C, G, A) et modifiés (U) et en utilisant la polymérase Deep-Vent.

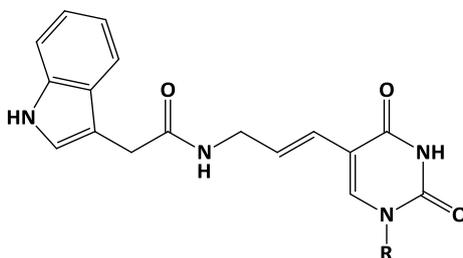


Figure 125 : structure de la 5-[(3-Indolyl)propionamide-N-allyl] uridine

IV.1. Test d'interaction entre la cible et la banque modifiée

Les analyses en électrophorèse capillaire couplée à une détection LIF, présentées dans la figure 126, ont montré la présence d'interactions entre la cible G-quadruplexe et la banque modifiée. Ceci se traduit par un changement de signal de la banque modifiée incubée en présence de la cible (courbe noire) par rapport à la banque seule (courbe rose). En effet, sur l'électrophérogramme, nous observons une augmentation du signal de la courbe noire par rapport à la courbe rose dans la zone 1 et la présence d'une région de désorption au niveau de la zone 2.

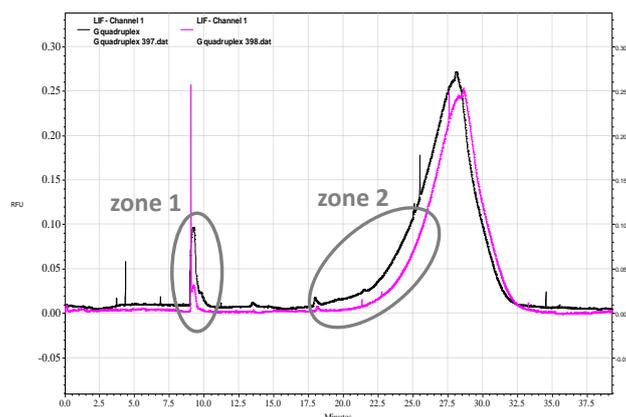


Figure 126: électrophéromme EC-LIF de la banque modifiée seule (courbe rose) et de la banque modifiée incubée en présence de cible (courbe noire)

Suite à ces premiers résultats encourageants, montrant la formation d'un complexe cible-banque modifiée, des cycles de sélections d'aptamères ont été réalisés.

IV.2. Réalisation de la CE-SELEX

Lors du premier cycle de sélection, la banque modifiée a été incubée avec la cible G-quadruplexe d'intérêt et les potentiels candidats aptamères ont ensuite été collectés, amplifiés par PCR, purifiés puis réincubés avec la cible. Au total, trois cycles de sélection ont été réalisés. Pour pouvoir sélectionner les aptamères les plus affins, la concentration de la cible a été divisée par dix à chaque cycle (voir partie expérimentale).

Les oligonucléotides sélectionnés lors de chaque cycle ont été mis en présence de la cible afin de vérifier la formation de complexes banque-cible par EC-LIF. Après le premier cycle de sélection (figure 127.a) une différence de signal est observée entre les oligonucléotides modifiés, en absence (courbe bleue) ou en présence de la cible (courbe rouge). On peut noter que l'amplification de ces candidats aptamères a généré une population très hétérogène (profil chromatographique du SELEX 1 très différent de la banque modifiée de départ (figure 127)). La formation de complexe d'interaction entre la banque et la cible G-quadruplexe est tout de même visible au niveau des pics observés à 10 min.

Après le deuxième cycle de sélection (figure 127.b), nous observons une diminution de la fluorescence de la banque avec l'augmentation de la concentration en cible, les courbes de couleur grise et bleu représentent respectivement la concentration de la cible G-quadruplexe à 40 et 80 μ M. Le phénomène observé pourrait être expliqué par un « quenching » de fluorescence dû à la fixation de la banque sur la cible. En se liant à la cible, les oligonucléotides sélectionnés pourraient adopter une certaine structuration qui entraîne une diminution de la fluorescence.

Pour le dernier cycle de sélection (figure 127.c), nous retrouvons la même tendance, à savoir une diminution de la fluorescence de la banque sélectionnée avec l'augmentation de la concentration en cible.

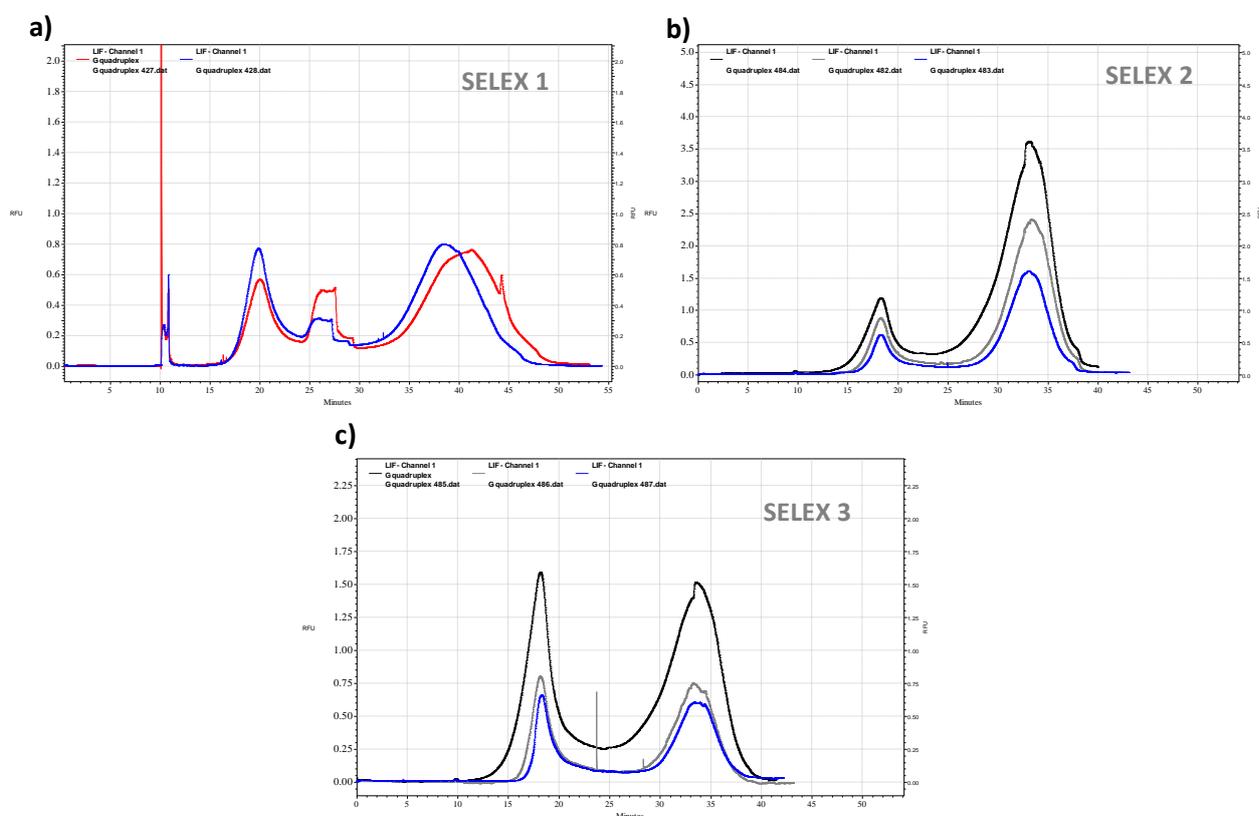


Figure 127: électrophérogramme EC-LIF (a) après le SELEX 1 : population sélectionnée après le SELEX 1 (courbe bleu), population sélectionnée après le SELEX 1 + cible G4 à 40 μM (courbe rouge), (b) après le SELEX 2 : population sélectionnée après le SELEX 2 (courbe noir), population sélectionnée après le SELEX 2 + cible G4 à 40 μM (courbe grise), population sélectionnée après le SELEX 2 + cible G4 à 80 μM (courbe bleue), (c) après le SELEX 3 : population sélectionnée après le SELEX 3 (courbe noir), population sélectionnée après le SELEX 3 + cible G4 à 40 μM (courbe grise), population sélectionnée après le SELEX 3 + cible G4 à 80 μM (courbe bleue)

Suite à ces résultats encourageants, témoignant d'une interaction entre la cible G-quadruplexe et la banque enrichie, les candidats aptamères sélectionnés ont été clonés et séquencés par la société Biofidal (voir partie expérimentale).

IV.3. Résultat du séquençage

Suite au séquençage des trois cycles de sélection, douze, onze et seize séquences issues respectivement des sélections 1, 2 et 3 ont été obtenues. Les trente-neuf séquences issues du séquençage ont ensuite été classées par familles à l'aide de deux logiciels différents, MEME (Multiple Em for Motif Elicitation, dans lequel Em correspond à l'algorithme utilisé pour l'identification des motifs)¹⁷² et Clustal Omega (EBI search)¹⁷³. L'alignement des séquences a été réalisé sur les séquences dépourvues de leurs amorces. Chaque famille comporte des séquences présentant de

fortes similarités. Par exemple, le logiciel MEME a identifié trois familles différentes avec plus de quatre séquences présentant des motifs consensus (tableau 11). Les autres familles sont constituées de séquences présentant des similarités deux à deux.

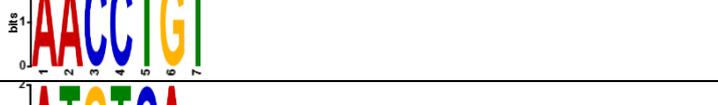
Motifs consensus de chaque famille	Nombre de séquences par famille
	8 séquences
	4 séquences
	4 séquences
	2 séquences
	2 séquences
	2 séquences
	2 séquences

Tableau 11 : résultats obtenus après alignement des séquences avec le logiciel MEME, seuls les motifs consensus de la partie aléatoire de chaque séquence sont représentés

Le logiciel Clustal Omega présentant un algorithme différent, classe les séquences par famille à l'aide d'un dendrogramme (figure 128).

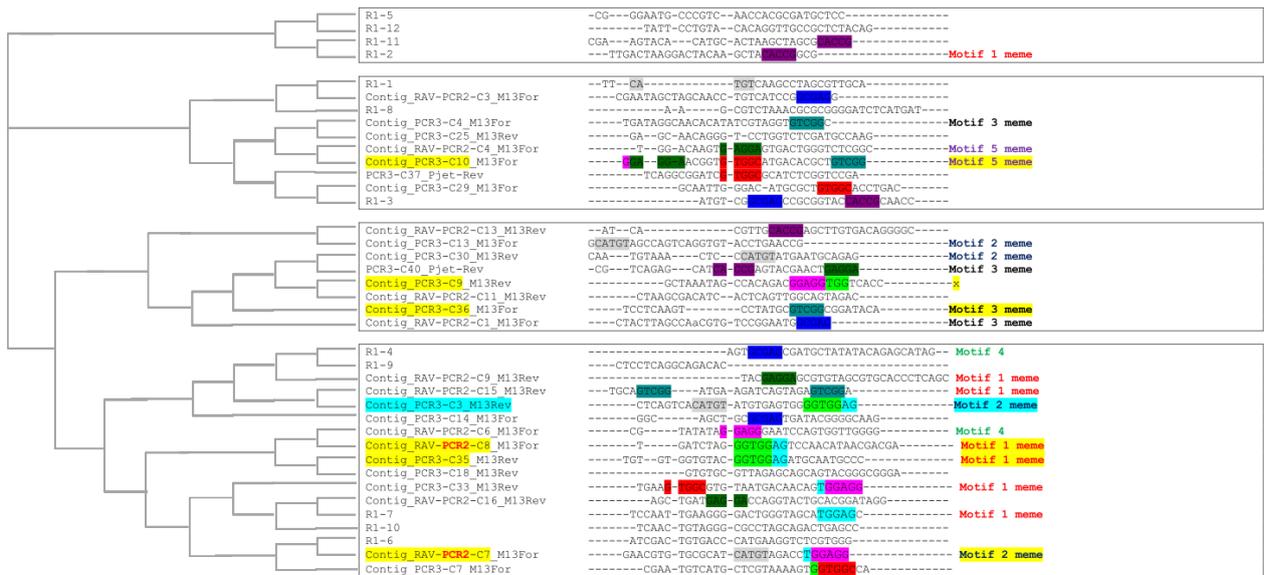


Figure 128: dendrogramme obtenu après alignement des séquences avec le logiciel Clustal Omega, seuls les motifs consensus de la partie aléatoire de chaque séquence sont représentés

Aucune corrélation n'a pu être établie entre ces deux classements. Dix séquences, les plus représentatives des différentes familles observées et issues des différents cycles de sélection ont donc été sélectionnées (tableau 12).

Séquences sélectionnées après séquençage
CGCCGGT GTAGCTTGTAG TCCTTAGTCA
GGAG GAACGGTGTGGCATGACACGCTGTCCG
TCCTCAAGTCCTATGCGTCGG CGGATACA
CTCAGTCA CATGTA TGTGAGTGGG
TGAT CTAGGGTGGAG TCCAACATAA
GTGTGCGCAT CATGTA GACCTGGAGG
CGA ATGTCA TGCTCGTAAA
TGTGTGGT GTACGGTGGAG ATGCAATGCC
ATGT CGGCGAGCCGCGGTACCACCGCAACC
GCTAAATAG CCACAGACGGAGGTGGTCACC

Tableau 12: séquences sélectionnées après le séquençage des SELEX 1, 2 et 3, à l'aide des logiciels MEME et Clustal Omega, seule la partie aléatoire de chaque séquence est représentée

IV.4. Evaluation de l'affinité des séquences sélectionnées pour la cible

Les séquences complémentaires des séquences choisies, comportant le « primer » ont été synthétisées par la société Eurogentec. Une amplification par PCR a ensuite été réalisée de la même manière qu'auparavant, c'est-à-dire en présence de l'uridine modifiée par le groupement indole. Après purification, l'affinité pour la cible G-quadruplexe des dix séquences modifiées a été évaluée par SPR. Nous avons utilisé comme cible l'édifice **B** qui comporte une fonction biotine, nécessaire à la fixation de la cible sur la surface. Cet édifice présente le même motif G-quadruplexe que l'édifice **K**, il a été synthétisé avec les mêmes oligonucléotides et sa topologie est parfaitement identique à celle adoptée par l'édifice qui a servi à la sélection des potentiels aptamères.

Les analyses SPR ont été réalisées dans le tampon d'incubation de la SELEX, afin de conserver la structuration du motif G-quadruplexe, adoptée lors de la sélection. Dans une première série d'analyses, nous avons testé l'interaction de la cible avec chacune des dix séquences à une concentration de 200 nM. Aucune interaction n'a été observée entre la cible G-quadruplexe et les séquences. Une deuxième série d'analyses a donc été effectuée en augmentant la concentration de quatre séquences à 1 μ M. Encore une fois aucune interaction n'a été mesurée.

A ce stade-là plusieurs hypothèses sont envisageables :

- le repliement de l'aptamère induit un signal négatif, opposé au signal de fixation, ce qui l'annule et qui pourrait expliquer l'absence de signal. Ceci est peu probable, car l'absence de signal est observée pour toutes les séquences.
- la piste de SPR est saturée en cible G-quadruplexe, ce qui pourrait gêner la fixation des aptamères, qui sont des molécules de taille importante.
- l'absence de signal est due à l'absence de fixation des oligonucléotides sélectionnés, qui ne sont tout simplement pas affins pour la cible G-quadruplexe.
- un problème d'amplification des candidats aptamères, qui aurait biaisé l'amplification des aptamères à bases modifiées vers les séquences naturelles correspondantes qui ont été clonées.
- un problème lors de l'amplification des séquences choisies pour les transformer en séquences modifiées.

Pour vérifier notre deuxième hypothèse, nous avons mesuré l'interaction entre la cible G-quadruplexe immobilisée à différentes concentrations et deux séquences de potentiels aptamères à 600 nM. Le but étant de ne pas saturer la piste pour laisser la place aux séquences de se fixer sur le G-quadruplexe. Une nouvelle fois aucune interaction n'a été observée. Nos trois dernières

hypothèses sont donc les plus probables, l'absence de signal est sûrement due à un manque d'affinité des séquences sélectionnées pour la cible G-quadruplexe ou à un problème d'amplification.

V. Bilan

Durant cette thèse, nous avons mis à profit le concept TASQ permettant la stabilisation des motifs G-quadruplexes en une conformation définie, pour tenter de sélectionner de nouveaux ligands de G-quadruplexes. Les aptamères qui ont la particularité de se structurer spécifiquement au contact de leur cible ont besoin d'éléments de reconnaissance particuliers pour interagir avec cette dernière. Les tétrades de guanines et les boucles latérales du motif G-quadruplexe choisi semblent être de très bons éléments de reconnaissances. Parmi toutes les séquences sélectionnées, aucune ne possède la séquence complémentaire du motif G-quadruplexe, ce qui démontre que sa stabilisation n'a pas permis l'émergence de séquences s'hybridant par interaction de type Watson-Crick. Ceci valide donc l'utilisation de nos modèles contraints. En revanche, les études d'interactions, pas SPR, entre la cible G-quadruplexe et les dix oligonucléotides issus du séquençage n'ont démontré aucune affinité.

Conclusion et perspectives

Le ciblage spécifique des G-quadruplexes impliqués dans la régulation d'un très grand nombre de systèmes biologiques pourrait avoir des applications dans le traitement de cancers (séquence télomérique, promoteurs *c-myc* et *c-kit*), d'infections virales (HIV, EBV) ou encore de certaines maladies génétiques (diabète de type I). Ce ciblage est néanmoins compliqué par la grande variété de topologies que les G-quadruplexes peuvent adopter en solution. Afin de pallier les phénomènes de polymorphisme inhérent aux G-quadruplexes, le laboratoire a développé il y a quelques années une nouvelle approche permettant de contraindre les motifs G-quadruplexes dans une topologie définie. La méthode consiste à utiliser un gabarit cyclodécapeptidique pour la stabilisation des motifs G-quadruplexes. Ces derniers sont formés par des oligonucléotides greffés sur le châssis peptidique *via* des ligations chimiosélectives. L'application de cette stratégie a permis aux docteurs Pierre Murat et Romaric Bonnet d'assembler durant leur thèse les édifices contraints **A** et **B** mimant des motifs G-quadruplexes présents dans la région télomérique. Dans ce contexte, l'objectif de ces travaux de thèse a été de synthétiser de nouveaux mimes de G-quadruplexes et de les utiliser pour la caractérisation et la sélection de ligands.

Le premier chapitre de ce manuscrit décrit le développement de nouveaux bioconjugués capables de s'auto-assembler en structures quadruplexes (figure 129) et présente leurs synthèses et leurs caractérisations.

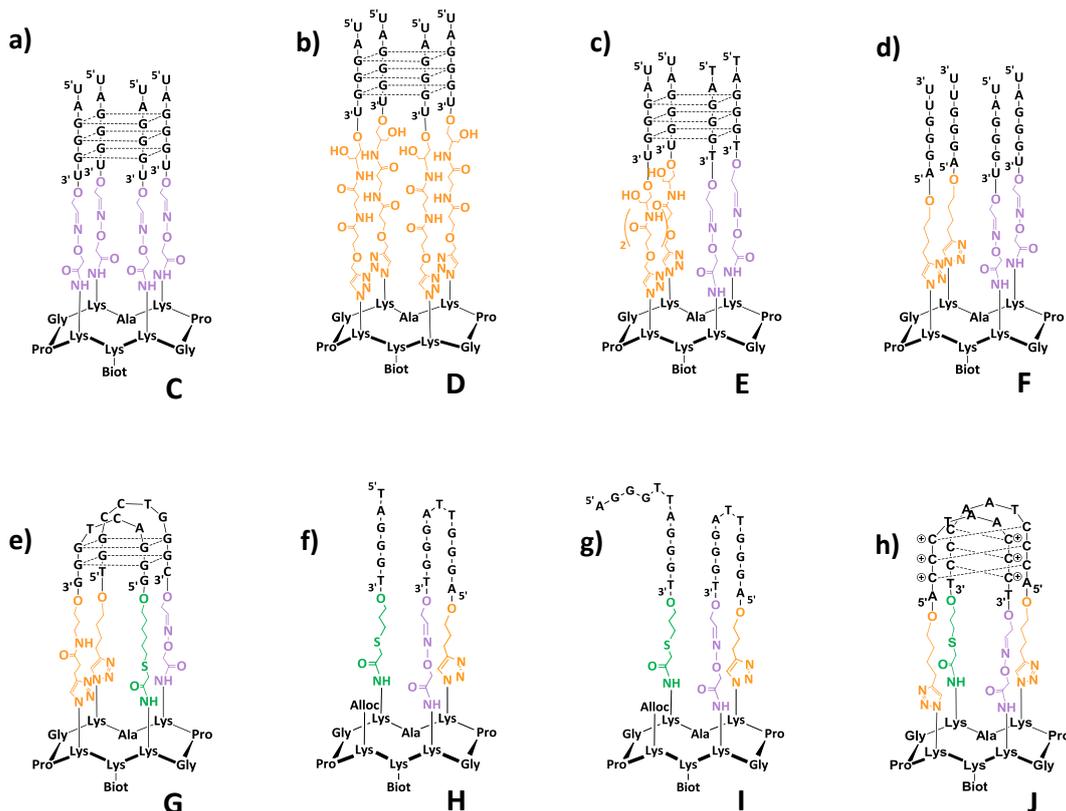


Figure 129 : édifices G-quadruplexes, G-triplexes et i-motif synthétisé durant cette thèse, dérivant de la séquence télomérique : (a) et (b) édifice ARN parallèle, (c) édifice hybride, (d) édifice antiparallèle, (f) édifice triplex, (g) édifice G-quadruplexe, (h) édifice i-motif et de la séquence HIV-PRO1 (e)

Comme évoqué précédemment, les motifs G-quadruplexes présents au sein de la région télomérique sont de prometteuses cibles biologiques. Afin de mieux comprendre leur rôle et étudier leur reconnaissance, nous souhaitons obtenir un panel assez important de motifs G-quadruplexes contraints représentatifs des motifs G4 présents au sein de cette région. Plusieurs édifices mimant des quadruplexes télomériques ont été synthétisés durant cette thèse. L'assemblage de ces édifices *via* l'utilisation de ligations chimiosélectives a nécessité en amont la synthèse et la fonctionnalisation des oligonucléotides et des châssis peptidique. Ces derniers ont été fonctionnalisés sur les fonctions amines ϵ des chaînes latérales de leurs lysines et les oligonucléotides sur leur extrémité 5' et/ou 3'. Les travaux réalisés ont permis d'isoler et de stabiliser le motif G-quadruplexe d'ARN parallèle **C**, ainsi que le motif hybride ADN/ARN **E** coexistant dans la région télomérique. Ils constituent les premiers édifices ARN à avoir été synthétisés. De plus, l'utilisation du concept TASQ a rendu possible la stabilisation du motif hybride ADN/ARN pour la première fois. La synthèse de ces édifices a nécessité l'utilisation d'une ou deux ligations chimiosélectives différentes, la ligation oxime et le CuAAC, largement utilisées au laboratoire.

Toujours dans le but de mimer les structures secondaires dérivant de la séquence télomérique, nous avons étudié la formation d'une structure G-triplexe, susceptible de se former lors

du repliement du motif G-quadruplexe de l'ADN télomérique humain. L'identification de ligands capable de reconnaître et d'interagir spécifiquement avec des structures G-triplexes pourrait aider à mieux comprendre leur rôle et étudier leur pertinence comme cible thérapeutique. L'assemblage de l'édifice triplexe **H** a nécessité l'utilisation d'une troisième ligation chimiosélectives, le couplage thiol-chloroacétamide, orthogonale aux deux premières (ligation oxime et CuAAC) pour le greffage des oligonucléotides sur le châssis peptidique.

Au-delà des G-quadruplexes télomériques, nous nous sommes également intéressés au ciblage d'un motif G-quadruplexe présent dans la séquence promotrice du VIH-1, qui semble être une cible antivirale prometteuse. La stabilisation de ce motif par l'assemblage de l'édifice **G** a également nécessité l'utilisation de trois ligations différentes, la ligation oxime, la CuAAC et le couplage thiol-chloroacétamide.

La structuration des édifices G-triplexe et G-quadruplexes en une topologie définie a été confirmée par RMN et dichroïsme circulaire. Cette dernière méthode d'analyse a également mis en évidence la stabilité des édifices formés grâce à la présence du châssis peptidique.

La synthèse des bioconjugués **H** et **G** a permis de mettre en place au laboratoire une méthodologie permettant de réaliser trois ligations d'oligonucléotides sur un même châssis peptidique. Grâce à cette troisième ligation nous avons pu avoir accès à ces motifs plus complexes dont l'assemblage nécessite trois points d'ancrage différents. Toujours en utilisant trois ligations, nous avons souhaité assembler un mime d'une structure i-motif de la séquence télomérique humaine. Ce motif de séquence complémentaire au G-quadruplexe pourrait contribuer à la régulation cellulaire. L'assemblage de ce mime implique un greffage indépendant de chaque extrémité des brins d'ADN. En utilisant l'orthogonalité des ligations chimiosélectives et celles des groupements protecteurs présents sur le châssis peptidique, nous avons utilisé les trois ligations à notre disposition pour greffer indépendamment chaque extrémité des deux boucles sur le châssis peptidique. Cette stratégie a nécessité d'employer deux fois la même ligation chimiosélective (la CuACC). Toutefois, cette stratégie de synthèse ne nous a pas permis d'aboutir à l'assemblage de l'édifice souhaité. Nous nous sommes notamment heurtés à des problèmes de purifications et de caractérisation.

Pour contourner ces difficultés, une autre stratégie, facilitant notamment la caractérisation en fin de synthèse, a été envisagée. Faute de temps, elle n'a pu peut être réalisée lors de cette thèse. Les oligonucléotides et le peptide misent en jeu sont représentés dans la figure 130.

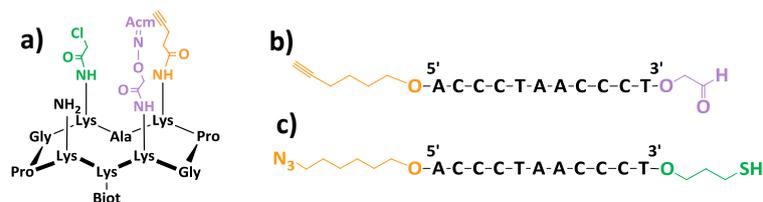


Figure 130 : Structure chimique du peptide **22** (a) et des oligonucléotides **23** (b) et **24** (c)

La première étape (schéma 41) consistera à greffer l'extrémité 3' de l'oligonucléotide **24** en réalisant un couplage thiol-chloroacétamide. La deuxième étape permettra de refermer la boucle *via* la CuAAC. Il est à noter que la fonction alcyne n'est plus présente sur l'oligonucléotide, mais sur le châssis peptide, elle sera incorporée sous forme de building block durant la synthèse. Ce dernier sera formé par l'addition de l'acide pentynoïque sur la fonction amine en ϵ d'une Fmoc-Lys-OH. La fonction azoture présente à l'extrémité 5' du brin d'ADN sera générée par réaction de diazo-transfert sur une fonction amine, obtenue en utilisant un phosphoramidite modifié (5' amino-modifier C6). La troisième étape consistera à générer une fonction azoture sur le châssis peptidique à partir de la fonction amine libre, par réaction de diazo-transfert. L'oligonucléotide **23** pourra ensuite être greffé sur le châssis par son extrémité 5' *via* la CuAAC. La fonction oxyamine présente sur le châssis sera ensuite libérée en milieu acide. Des tests de déprotection ont déjà été réalisés pour évaluer la stabilité des brins d'ADN en milieu acide. Ils ont mis en évidence que l'utilisation de 0,5% de TFA permet d'éliminer le groupement protecteur Eei sans dégrader l'oligonucléotide. Enfin, la dernière étape consistera à utiliser la ligation oxime pour refermer la deuxième boucle. La réalisation de cette ligation en dernière étape devrait permettre un suivi de la réaction en spectrométrie de masse. En effet, la formation du lien éther d'oxime provoque l'élimination d'une molécule d'eau, ce qui entraîne une différence de masse entre le produit de départ et le produit d'arrivée.

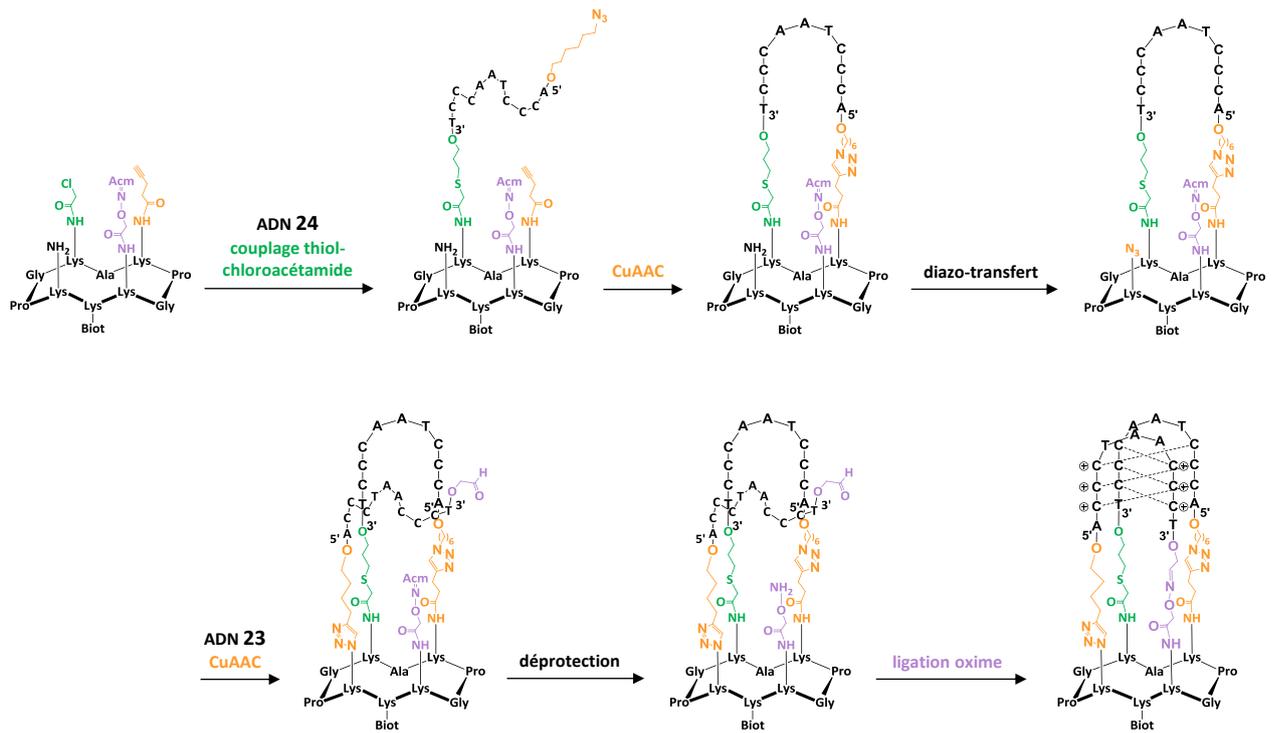


Schéma 41 : Synthèse de l'édifice i-motif avec la nouvelle stratégie de synthèse

Il est important de noter que, si la chimie utilisée pour la formation de ces édifices semble simple et efficace, de nombreuses difficultés sont rencontrées lors de la purification et de la caractérisation de certains édifices. Bien que la synthèse devienne de plus en plus compliquée et délicate à réaliser avec l'augmentation du nombre de ligations chimiosélectives, cela permet de représenter un grand nombre d'éléments structurels pour chaque motif et de se rapprocher au plus près des conformations susceptibles d'exister *in-vivo*.

Afin de mimer avec précision un motif G-quadruplexe ou un i-motif, il est théoriquement nécessaire de représenter tous ces nucléotides et toutes ces boucles. Dans le cadre d'un motif G-quadruplexe de topologie parallèle, cela consisterait à utiliser un seul oligonucléotide présentant une fonctionnalisation à chaque extrémité et deux fonctionnalisations intra-oligonucléotidique. Un exemple d'une telle synthèse, visant à mimer le motif G-quadruplexe parallèle de l'ADN télomérique en représentant toutes les boucles « propeler » est présenté en figure 131.

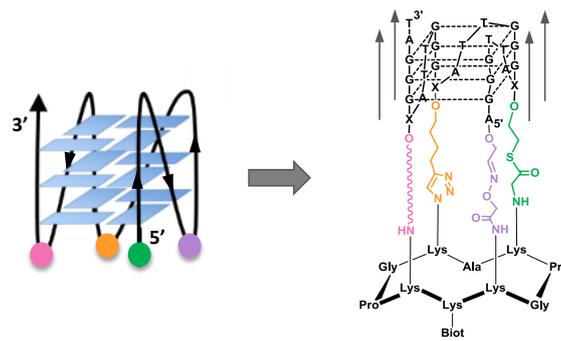


Figure 131 : édifice mimant le motif G-quadruplexe parallèle de l'ADN télomérique humain

En perspective de ce travail de thèse et afin de réaliser ce type de structure, la mise en place d'une quatrième ligation orthogonale et la fonctionnalisation d'ADN intra-séquence est actuellement en cours de développement au laboratoire.

Le second chapitre de ce manuscrit a été consacré à l'étude des interactions entre les édifices contraints présentés dans le chapitre 1 et différents ligands de G-quadruplexes ainsi qu'à la sélection d'aptamères contre un édifice G-quadruplexe.

Les édifices G-quadruplexes et G-triplexe assemblés durant cette thèse constituent des plateformes pertinentes pour commencer à évaluer l'affinité et la sélectivité de ligands de G-quadruplexes. Les études SPR et BLI réalisées sur les édifices contraints nous ont permis d'identifier des ligands intéressants en terme d'affinité, de sélectivité par rapport au duplexe d'ADN et de sélectivité par rapport à une topologie donnée. BRACO-19, L8, PDS et PhenDC3 ont par exemple montré une faible affinité pour le duplexe d'ADN et donc une bonne sélectivité pour les G-quadruplexes et G-triplexe. PDS et PhenDC3 ont présenté une très bonne affinité, de l'ordre du nanomolaire, quel que soit le motif G-quadruplexe étudié, mais n'ont pas été sélectifs d'une topologie quadruplexe donnée. Bien qu'ils présentent des affinités comparables pour une même cible ces deux ligands montrent des constantes cinétiques d'interaction totalement différentes : PhenDC3 reconnaît rapidement sa cible G4 tandis que PDS se dissocie très difficilement. Parmi tous les ligands étudiés, BRACO-19 est celui qui a permis de discriminer le mieux les différentes topologies de G-quadruplexes. Il a par exemple démontré une sélectivité pour l'édifice G4 ADN parallèle **A** par rapport à l'édifice G4 ARN parallèle **C**, ou encore pour l'édifice G4 HIV-PRO1 **G** par rapport à l'édifice G4 télomérique basket **B**.

Ces études d'interactions ont donc démontré que le ciblage sélectif de motif ARN ou encore hybride ADN/ARN est envisageable. Elles ont également mis en évidence la possibilité de cibler des motifs G-triplexes. En effet les ligands étudiés ont été capables de se lier avec une très bonne affinité

à l'édifice G-triplexe mais n'ont en revanche pas permis de discriminer le motif G-triplexe du motif G-quadruplexe. Il serait donc intéressant d'évaluer la sélectivité d'autres ligands, afin d'en identifier un ou plusieurs capables de reconnaître spécifiquement les G-triplexes par rapport aux G-quadruplexes ou vis-versa.

Durant cette thèse, nous avons également mis à profit le concept TASQ permettant la stabilisation des motifs G-quadruplexes en une conformation définie, pour tenter de sélectionner de nouveaux ligands de G-quadruplexes. Les aptamères qui ont la particularité de se structurer spécifiquement au contact de leur cible ont besoin d'éléments de reconnaissance particuliers pour interagir avec cette dernière. Les tétrades de guanines et les boucles latérales de nos motifs G-quadruplexes contraints semblent être de bons éléments de reconnaissance. Nous avons donc choisi comme cible le motif G-quadruplexe de type « basket » de la séquence télomérique présentant ces deux éléments de reconnaissance. La sélection d'aptamères par la méthode CE-SELEX a nécessité trois cycles de sélection. Les potentiels candidats sélectionnés lors de chaque étape ont présenté des interactions avec la cible G-quadruplexe. En effet, la formation de complexe cible G-quadruplexe-aptamère a été visualisée par EC-LIF. En revanche, l'évaluation de l'interaction, par SPR, entre la cible G-quadruplexe et dix oligonucléotides issus du séquençage n'a démontré aucune affinité. Les perspectives sont d'évaluer l'affinité par SPR des candidats aptamères en mélange, sélectionnés à l'issue du troisième cycle de sélection et ce afin de s'assurer qu'il n'y ait pas eu un problème d'amplification lors de la conversion des aptamères modifiés sélectionnés en séquences naturelles avant le clonage. Nous pourrions également avoir eu un problème d'amplification pour passer des oligonucléotides naturels sélectionnés par séquençage aux séquences modifiées, lors de la production en quantité de ces derniers avant leur évaluation par SPR. Il est donc envisagé d'optimiser à nouveau les conditions de PCR en testant par exemple d'autres polymérases. Enfin il est également possible que les séquences choisies ne soient pas actives envers la cible G4 mais que d'autres le soient, il est donc prévu de tester d'autres séquences parmi toutes celles séquencées.

Au-delà de la caractérisation et de la sélection de ligands, nos G-quadruplexes fortement stabilisés trouvent également d'autres applications. Je n'ai pas présenté dans ce manuscrit des travaux réalisés pour l'assemblage de bioconjugués G-quadruplexes antiviraux, ou encore des travaux qui vont permettre d'identifier *in-cellulo* les protéines en interactions avec les G-quadruplexes. Ces pistes de recherches constituent autant de perspectives pour les années à venir.

Partie expérimentale

I. Materiels 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. β -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). MMQ1, PDS and PhenDC3 ligands were kindly provided by M. P. Teulade-Fichou and colleagues (Institut Curie, Paris, France) L8 and L14 ligands were kindly provided by F. Thomas and colleagues (DCM, Grenoble, France). Trizma base (tris) and all the reagents for acrylamide gel preparation were from Sigma Aldrich (St Quentin Fallavier, France). The 103-nucleotide single-stranded DNA (ssDNA) library and 5'-FAM-reverse primer (24 bases) were from Eurogentec (Liege, Belgium). Heavier forward primer (20bases-spacer C18- 23 bases) was from Eurofins Genomics (Ebersberg, Germany). GeneAmp 10X PCR buffer II, AmpliTaq Gold DNA polymerase and SYBR[®] green PCR Kit were provided from Applied Biosystems (Foster city, California, USA). Vent DNA polymerase was provided from New England Biolabs (Evry, France). 5-[(3-Indolyl)propionamide-N-allyl]-2'-deoxyuridine-5'-Triphosphate was from TriLink Technologies.

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 3-hydroxy piccolinic acid (HPA, 45 mg; ammonium citrate 4 mg in 500 μ L H₂O/CH₃CN) 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. CH₃CN and TFA were HPLC quality.

II. Peptides synthesis

II.1. General procedure for 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 μ m) and detection by UV at 214 nm and 250 nm and by electron spray ionization mass spectrometry. A 1mL/min flow linear gradient from 95% solvent A (0.1% citric acid in water) and 5% solvent B (0.1% citric acid in acetonitrile/water: 9/1) to 100% B for 4 minutes was applied.

RP-HPLC purifications were performed on a Gilson system with Nucleosil C18 column (100 Å, 250 x 21 mm, 7 μ m) with UV monitoring at 214 nm and 250 nm. A 20 mL/min flow linear gradient I

was applied from 95% solvent A¹ (0.1% TFA in water) and 5% solvent B¹ (0.1% TFA in acetonitrile/water: 9/1) to 100% B¹ for 20 minutes or a 20 mL/min flow linear gradient II was applied from 95% solvent A² (0.1% formic acid in water) and 5% solvent B² (0.1% formic acid in acetonitrile/water: 9/1) to 100% B¹ for 20 minutes.

II.1.1. General procedure for linear protected peptides synthesis by SPPS

All peptide were synthesis using commercially available 2-Chlorotrityl resin (loading of 0.83 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. A DCM/MeOH/DIEA solution (17/2/1, v/v/v) was then added. 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 0.40 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. The synthesizer uses amino acid at 4 eq, DIEA and HBTU as activator. Supported peptide was cleaved from the resin using a 1% TFA solution in DCM (5 x 20 mL) or 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.

II.1.2. 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.

II.1.3. General procedure for N-ε Boc lysine deprotection

Peptides were dissolved in DCM/TFA/H₂O (45/45/10, 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.

II.1.4. 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.

II.1.5. General procedure for coupling protected aminoxy linker on peptide

Peptides were dissolved in anhydrous DMF (25 mL) and *N*-Boc-aminoxyacetic acid *N*-hydroxysuccinimide ester (1.2 eq) was added. The pH was adjusted to 8-9 with DIEA. The reaction was stirred at room temperature and monitoring by UPLC until the completion of the reaction. The

solvent was then evaporated under vacuum and the crude peptides were precipitated in ether. The crude products were purified on RP-HPLC and freeze-dried to obtain a white powder.

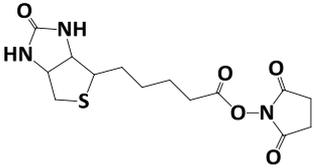
II.1.6. General procedure for coupling chloroacetamide linker on peptide

Peptides were dissolved in DMF to reach a $3 \cdot 10^{-3}$ 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. 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.

II.1.7. General procedure for oxyamine déprotection

Peptides were dissolved in TFA/TIS/H₂O (80/16/4, v/v/v) solution and the mixture was stirred for 15h at room temperature. The solvent was then evaporated under vacuum and the crude peptides were precipitated in ether as a white powder.

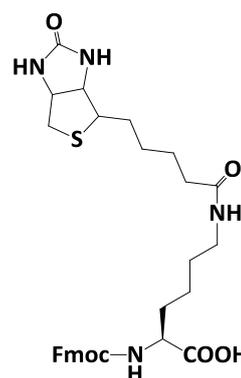
II.1.8. Experimental procedure

Biotin succinimidyl ester 2a	
<p>To a solution of biotin (1eq, 12 mmol, 3 g) in anhydrous DMF (100mL) were added <i>N</i>-hydroxysuccinimide (1.2 eq, 15 mmol, 1,7 g) and DCC (1.2 eq, 15 mmol, 3,1 g). The solution is warmed until the complete dilution of product and the mixture was stirred for 30h at room temperature. The mixture was filtrated using cellite in order to eliminate the DCU then the solvent was evaporated and the crude product is tritured in ether. After filtration the product was obtained as a white solid (11.8 mmol, 4 g, yield: 98%).</p> <p>ESI-MS (+): <i>m/z</i> calcd for C₁₄H₂₀N₃O₅S: 342.4, <i>m/z</i> found: 342.2 [M+H]⁺</p>	

Fmoc-Lys(Biotin)-OH **2**

To a solution of Fmoc-Lys-OH (1 eq, 6.8 mmol, 2.5 g) in DMF (30 mL) was added **2a** (1.2 eq, 8.1 mmol, 2.7 g). The pH was adjusted to 8-9 using DIEA and the mixture was stirred for 2h. The pH was reduced to 1 using a 2M Hydrochloric acid solution and the white precipitate is filtrated and washed with cooled water. The product was freeze-dried with ether and was obtained as a white solid (5.4 mmol, 3.2 g) yield: 80%.

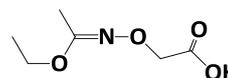
ESI-MS (+): m/z calcd for $C_{31}H_{39}N_4O_6S$: 595.7, m/z found: 595.3 $[M+H]^+$



2-(1-Ethoxyethylideneaminoxy)acetic acid **15a**

To a stirred solution of iodoacetic acid (1 eq, 49 mmol, 9 g) in water (19 mL) at 0°C was added aqueous NaOH (3.0 mL, 40% w/w). The resulting solution was allowed to heat to room temperature, after ethyl *N*-hydroxyacetimidate (1.2 eq, 59 mmol, 6.1 g) was added followed by aqueous NaOH (4.5 mL, 40% w/w) and water (19 mL) (pH of solution >12). After 5 h of stirring at 80°C and cooling to room temperature, mixture was washed with DCM (3 x 50 mL). Aqueous phase was cooled at 0°C and the pH was adjusted to 2-3 with a 1 M hydrochloride solution. The acidified water phase was then extracted with ethyl acetate (4 x 50 mL) (NB: After each extraction the pH was adjusted to 2-3 with a 1 M hydrochloride solution). The combined organic phases from this last extraction was washed with brine (50 mL), dried over Na_2SO_4 , and concentrated in vacuum, providing compound as yellow oil (26.5 mmol, 4.24 g, yield: 54%).

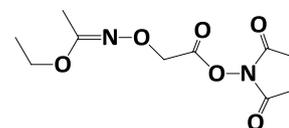
1H NMR (400 MHz, $CDCl_3$), δ (ppm): 4.48 (2H, s), 4.00 (2H, q, $J = 7.2$ Hz), 2.01 (3H, s), 1.27 (3H, t, $J = 7.2$ Hz)



N-hydroxysuccinimidyl 2-(1-Ethoxyethylideneaminoxy)acetate **15b**

To a stirred solution of 2-(1-Ethoxyethylideneaminoxy)acetic acid **15a** (1 eq, 26 mmol, 4.16 g) and *N*-hydroxysuccinimide (1.2 eq, 31.2 mmol, 3.6 g) in ethyl acetate/dioxane (120 mL, 1/1). DCC (1.2 eq, 31.2 mmol, 6.4 g) was added at 0 °C in one portion. The resulting mixture was stirred at room temperature overnight. The formed DCU was filtered off and the filtrate concentrated under vacuum. The obtained residue was dissolved in ethyl acetate (200 mL), and the solution was washed with 5% aqueous NaHCO₃ (2 x 75 mL), water (75 mL), and brine (75 mL). The organic solution was dried over Na₂SO₄ and evaporated under vacuum to obtain yellow oil (22.4 mmol, 5.8 g, yield: 86%).

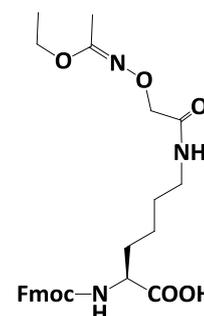
¹H NMR (400 MHz, CDCl₃), δ (ppm): 4.78 (2H, s), 4.01 (2H, q, J = 7.2 Hz), 2.84 (4H, s), 1.98 (3H, s), 1.27 (3H, t, J = 7.2 Hz)



Fmoc-Lys(Eei-Aoa)-OH **15**

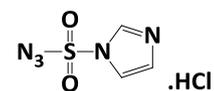
The Fmoc-Lys-OH (1.2 eq, 24 mmol, 8.8 g) was dissolved in DMF (100 mL), the pH was adjusted to 8-9 using DIEA. A solution of compound **15b** (1 eq, 20 mmol, 5.2 g) in DMF (40 mL) was added dropwise during 20 min and the mixture was stirred for 2h. The solvent was evaporated. The obtained residue was dissolved in ethyl acetate (100 mL), and the solution was washed with 10% aqueous citric acid (3 x 50 mL), and brine (100 mL). The organic solution was dried over Na₂SO₄ and evaporated under vacuum to obtain yellow powder (23.3 mmol, 12 g, yield: 97%).

ESI-MS (-): *m/z* calcd for C₂₇H₃₂N₃O₇: 510.2, *m/z* found: 510.3 [M-H]⁻



imidazole-1-sulfonyl azide **5**

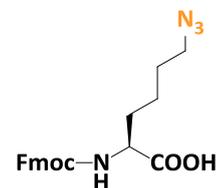
Sodium azide (1 eq, 90mmol, 5.9 g) was dissolved in MeCN (80 mL). Then sulfonyl chloride (1eq, 90 mmol, 12.1 g) were added dropwise at 0°C and the solution was stirred overnight at room temperature. Imidazole (2eq, 180 mmol, 12.4 g) was slowly added at 0°C and the reaction was stirred at room temperature. After 3h ethyl acetate (80 mL) was added and the product was washed with water (2 x 80mL) then with saturated NaHCO₃ solution (2 x 80 mL). The product was dried over Na₂SO₄ and filtrated. Hydrochloric acid in EtOH (8.5 mL d'AcCl in 30 mL of EtOH) was added dropwise, the solution was cooled at 0°C to obtain the crystallization of the product. After 2h the product was filtrated was washed with ethyl acetate (3x 40mL). The product was obtained as white crystals (70.7 mmol, 12.2 g, yield: 78%).



¹H NMR (400 MHz, D₂O), δ (ppm): 9.53 (1H, dd, J = 1.3, 1.6 Hz), 8.09 (1H, dd, J = 1.6, 2.2 Hz), 7.68 (1H, dd, J = 1.3, 2.2 Hz)

Fmoc-azidonorleucine **6**

Fmoc-Lys-OH (1 eq, 12.7 mmol, 4.7 g), CuSO₄ (0.04 eq, 0.51 mmol, 0.12 g), NaHCO₃ (3.5 eq, 44.5 mmol, 3.74 g) and ISAHC **5** (1.2 eq, 15.2 mmol, 2.6 g) were dissolved in MeOH/H₂O solution (50mL, 1/1). pH 8-9 was adjusted using a solution of saturated NaHCO₃. The mixture was stirred overnight at room temperature. pH 2 was adjusted using a 3M hydrochloric acid solution and ethyl acetate (60 mL) was added. The solution was washed with brine (2 x 60 mL), saturated NaHCO₃ solution (8 x 60 mL) and brine (2 x 60 mL). The organic solution was dried over Na₂SO₄ and evaporated under vacuum to obtain yellow solid (12.4 mmol, 4.9 g, yield: 98%).

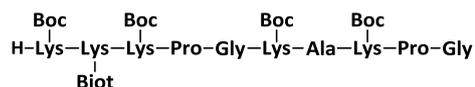


ESI-MS (+): *m/z* calcd for C₂₁H₂₂N₄O₄Na: 417.2, *m/z* found: 417.2 [M+Na]⁺

Linear peptide **1c**

The linear peptide was synthesized using general procedure **II.1.1.** and modified amino acid **2**. The crude product was used without purification.

R_t = 1.83 min

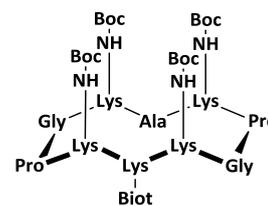


Cyclic peptide **1d**

The linear peptide **1c** (353 mg) was cyclized using general procedure **II.1.2**. The crude product was purified on RP-HPLC (gradient I) and freeze-dried to obtain a white powder (0.14 mmol, 231 mg, yield: 66% for two steps).

$R_t = 2.00$ min

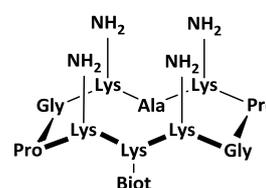
ESI-MS (+): m/z calcd for $C_{77}H_{132}N_{17}O_{20}S$: 1648.0, m/z found: 1648.4
[M+H]⁺



Peptide **1e**

The peptide **1d** (258 mg) was deprotected using general procedure **II.1.3**. The crude product was used without any further purification.

$R_t = 1.02$ min

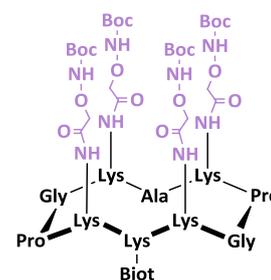


Peptide **1f**

The peptide **1f** was obtained using general procedure **II.1.5**. and peptide **1e** (200 mg). The crude product was purified on RP-HPLC (gradient I) and freeze-dried to obtain a white powder (0.10 mmol, 194 mg, yield: 63% for two steps).

$R_t = 1.80$ min

ESI-MS (+): m/z calcd for $C_{85}H_{144}N_{21}O_{28}S$: 1940.2, m/z found: 1939.9
[M+H]⁺

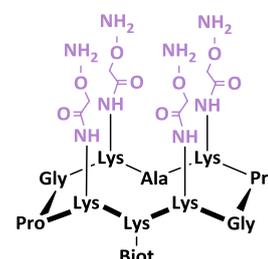


Peptide **1**

The peptide **1f** (3 mg) was deprotected using general procedure **II.1.7**. The crude product was used without any further purification. The yield was considered as quantitative

$R_t = 1.16$ min

ESI-MS (+): m/z calcd for $C_{65}H_{111}N_{21}O_{20}S$: 1539.7, m/z found: 1538.1
[M+H]⁺



Linear peptide **4b**

The linear peptide was synthesized using general procedure **II.1.1.** and modified amino acid **2** and **6**. The crude product was used without purification.

$R_t = 1.50$ min

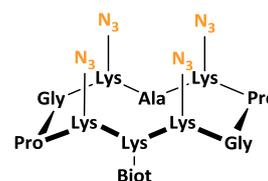


Cyclic peptide **4**

The linear peptide **4b** (271 mg) was cyclized using general procedure **II.1.2.** The crude product was purified on RP-HPLC (gradient I) and freeze-dried to obtain a white powder (0.13 mmol, 176 mg, yield: 66% for two steps).

$R_t = 1.89$ min

ESI-MS (+): m/z calcd for $C_{57}H_{92}N_{25}O_{12}S$: 1351.5, m/z found: 1350.9 $[M+H]^+$



Linear peptide **8b**

The linear peptide was synthesized using general procedure **II.1.1.** and modified amino acid **2** and **6**. The crude product was used without purification.

$R_t = 1.45$ min

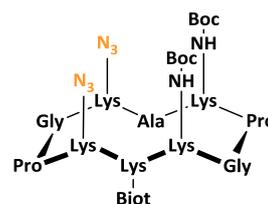


Cyclic peptide **8c**

The linear peptide **8b** (280 mg) was cyclized using general procedure **II.1.2.** The crude product was purified on RP-HPLC (gradient I) and freeze-dried to obtain a white powder (0.13 mmol, 190 mg, yield: 67% for two steps).

$R_t = 2.10$ min

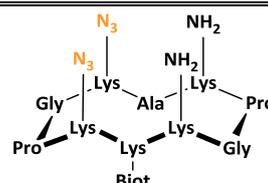
ESI-MS (+): m/z calcd for $C_{67}H_{112}N_{21}O_{16}S$: 1498.8, m/z found: 1498.3 $[M+H]^+$



Peptide **8d**

The peptide **8c** (0.12 mmol, 180 mg) was deprotected using general procedure **II.1.3.** The crude product was used without any further purification.

$R_t = 1.09$ min

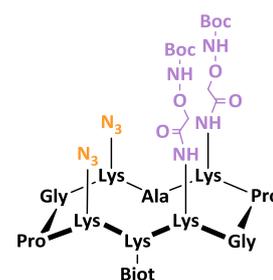


Peptide 8e

The peptide **8e** was obtained using general procedure **II.1.5.** and peptide **8d**. The crude product was purified on RP-HPLC (gradient I) and freeze-dried to obtain a white powder (0.080 mmol, 131 mg, yield: 67% for two steps).

$R_t = 1.87$ min

ESI-MS (+): m/z calcd for $C_{71}H_{118}N_{23}O_{20}S$: 1644.8, m/z found: 1644.2 $[M+H]^+$

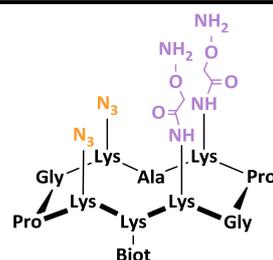


Peptide 8

The peptide **8e** (3 mg) was deprotected using general procedure **II.1.7.** The crude product was used without any further purification. The yield was considered as quantitative

$R_t = 1.43$ min

ESI-MS (+): m/z calcd for $C_{61}H_{102}N_{23}O_{16}S$: 1445.7, m/z found: 1444.4 $[M+H]^+$



Linear peptide 11b

The linear peptide was synthesized using general procedure **II.1.1.** and modified amino acid **2** and **6**. The crude product was used without purification.

$R_t = 1.35$ min

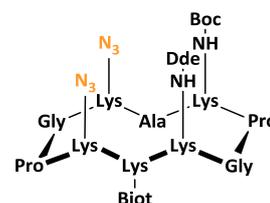


Cyclic peptide 11c

The linear peptide **11b** (325 mg) was cyclized using general procedure **II.1.2.** The crude product was purified on RP-HPLC (gradient I) and freeze-dried to obtain a white powder (0.090 mmol, 140 mg, yield: 50% for two steps).

$R_t = 1.89$ min

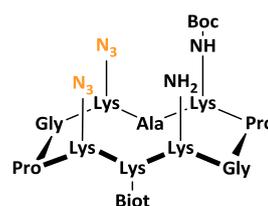
ESI-MS (+): m/z calcd for $C_{72}H_{115}N_{21}NaO_{16}S$: 1584.8, m/z found: 1584.8 $[M+Na]^+$



Peptide 11d

The peptide **11c** (0.063 mmol, 100 mg) was deprotected using general procedure **II.1.4**. The crude product was used without any further purification.

$R_t = 1.38$ min

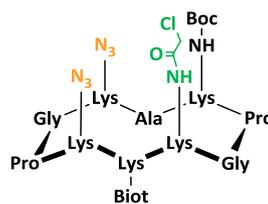


Peptide 11e

The peptide **11e** was obtained using general procedure **II.1.6** and peptide **11d**. The crude product was purified on RP-HPLC (gradient I) and freeze-dried to obtain a white powder (0.026 mmol, 39 mg, yield: 41% for two steps).

$R_t = 1.80$ min

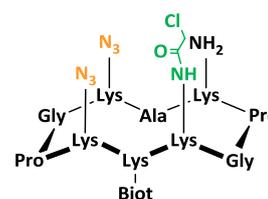
ESI-MS (+): m/z calcd for $C_{64}H_{104}ClN_{21}NaO_{15}S$: 1496.7, m/z found: 1496.7 $[M+Na]^+$



Peptide 11f

The peptide **11e** (0.067 mmol, 100 mg) was deprotected using general procedure **II.1.3**. The crude product was used without any further purification.

$R_t = 1.25$ min

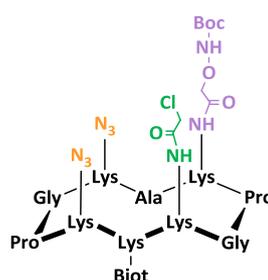


Peptide 11g

The peptide **11g** was obtained using general procedure **II.1.5** and peptide **11f**. The crude product was purified on RP-HPLC (gradient I) and freeze-dried to obtain a white powder (0.049 mmol, 77 mg, yield: 73% for two steps).

$R_t = 1.73$ min

ESI-MS (+): m/z calcd for $C_{66}H_{107}ClN_{22}NaO_{17}S$: 1569.7, m/z found: 1569.8 $[M+Na]^+$

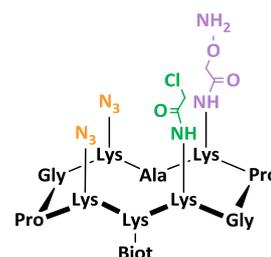


Peptide 11

The peptide **11g** (3 mg) was deprotected using general procedure **II.1.7**. The crude product was used without any further purification. The yield was considered as quantitative.

$R_t = 1.44$ min

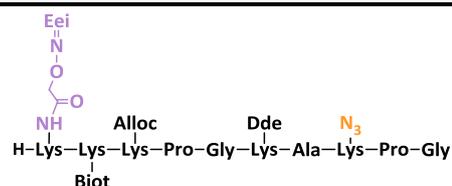
ESI-MS (+): m/z calcd for $C_{61}H_{100}ClN_{22}O_{15}S$: 1447.7, m/z found: 1447.5 $[M+H]^+$



Linear peptide 14b

The linear peptide was synthesized using general procedure **II.1.1** and modified amino acid **2**, **6** and **15**. The crude product was used without purification.

$R_t = 1.35$ min

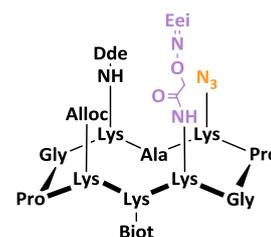


Cyclic peptide 14c

The linear peptide **14b** (365 mg) was cyclized using general procedure **II.1.2**. The crude product was purified on RP-HPLC (gradient II) and freeze-dried to obtain a white powder (0.15 mmol, 238 mg, yield: 66% for two steps).

$R_t = 1.86$ min

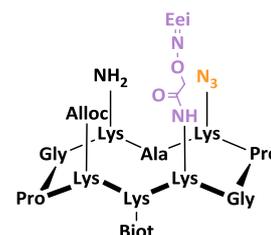
ESI-MS (+): m/z calcd for $C_{77}H_{123}N_{20}O_{19}S$: 1663.8, m/z found: 1664.5 $[M+H]^+$



Peptide 14d

The peptide **14c** (0.15 mmol, 250 mg) was deprotected using general procedure **II.1.4**. The crude product was used without any further purification.

$R_t = 1.34$ min

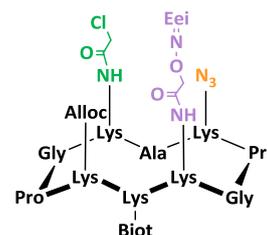


Peptide 14e

The peptide **14e** was obtained using general procedure **II.1.6.** and peptide **14d**. The crude product was purified on RP-HPLC (gradient II) and freeze-dried to obtain a white powder (0.09 mmol, 142 mg, yield: 60% for two steps).

$R_t = 1.74$ min

ESI-MS (+): m/z calcd for $C_{69}H_{112}ClN_{20}O_{18}S$: 1575.8, m/z found: 1576.2 $[M+H]^+$

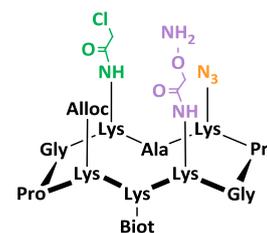


Peptide 14

The peptide **14e** (3 mg) was deprotected using general procedure **II.1.7.** The crude product was used without any further purification. The yield was considered as quantitative.

$R_t = 1.51$ min

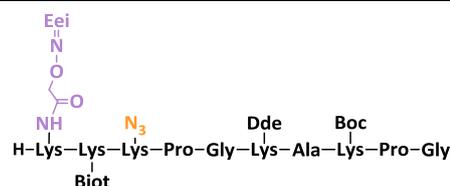
ESI-MS (+): m/z calcd for $C_{65}H_{106}ClN_{20}O_{17}S$: 1507.1, m/z found: 1506.2 $[M+H]^+$



Linear peptide 19b

The linear peptide was synthesized using general procedure **II.1.1.** and modified amino acid **2**, **6** and **15**. The crude product was used without purification.

$R_t = 1.48$ min

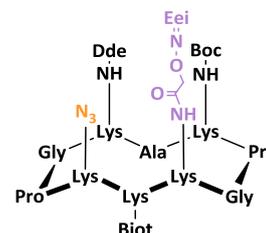


Cyclic peptide 19c

The linear peptide **19b** (331 mg) was cyclized using general procedure **II.1.2.** The crude product was purified on RP-HPLC (gradient II) and freeze-dried to obtain a white powder (0.19 mmol, 320 mg, yield: 60% for two steps).

$R_t = 1.96$ min

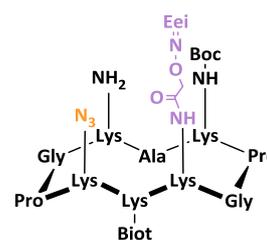
ESI-MS (+): m/z calcd for $C_{78}H_{127}N_{20}O_{19}S$: 1679.9, m/z found: 1680.5 $[M+H]^+$



Peptide 19d

The peptide **19c** (0.12 mmol, 200 mg) was deprotected using general procedure **II.1.4**. The crude product was used without any further purification.

$R_t = 1.39$ min

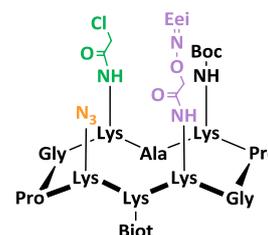


Peptide 19e

The peptide **19e** was obtained using general procedure **II.1.6**. and peptide **19d**. The crude product was purified on RP-HPLC (gradient II) and freeze-dried to obtain a white powder (0.040 mmol, 63.6 mg, yield: 33% for two steps).

$R_t = 1.84$ min

ESI-MS (+): m/z calcd for $C_{70}H_{116}ClN_{20}O_{18}S$: 1591.8, m/z found: 1593.3 $[M+H]^+$

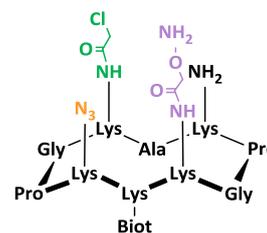


Peptide 19

The peptide **19e** (3 mg) was deprotected using general procedure **II.1.7**. The crude product was used without any further purification. The yield was considered as quantitative.

$R_t = 1.20$ min

ESI-MS (+): m/z calcd for $C_{61}H_{102}ClN_{20}O_{15}S$: 1421.7, m/z found: 1421.9 $[M+H]^+$



Linear peptide 22b

The linear peptide was synthesized using general procedure **II.1.1**. and modified amino acid **2**. The crude product was used without purification.

$R_t = 1.67$ min

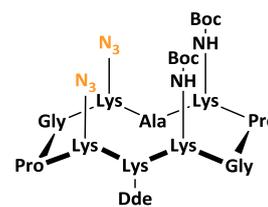


Cyclic peptide **22c**

The linear peptide **22b** (304 mg) was cyclized using general procedure **II.1.2**. The crude product was purified on RP-HPLC (gradient I) and freeze-dried to obtain a white powder (0.14 mmol, 195 mg, yield: 65% for two steps).

$R_t = 2.07$ min

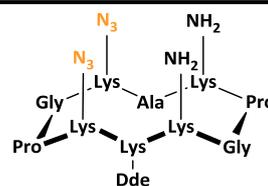
ESI-MS (+): m/z calcd for $C_{67}H_{110}N_{19}O_{16}$: 1436.8, m/z found: 1437.1
[M+H]⁺



Peptide **22d**

The peptide **22c** (0.098 mmol, 141 mg) was deprotected using general procedure **II.1.3**. The crude product was used without any further purification.

$R_t = 1.55$ min

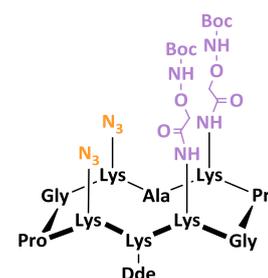


Peptide **22e**

The peptide **22e** was obtained using general procedure **II.1.5**. and peptide **22d**. The crude product was purified on RP-HPLC (gradient I) and freeze-dried to obtain a white powder (0.059 mmol, 93.3 mg, yield: 60% for two steps).

$R_t = 1.98$ min

ESI-MS (+): m/z calcd for $C_{71}H_{116}N_{21}O_{20}$: 1582.8, m/z found: 1582.4
[M+H]⁺

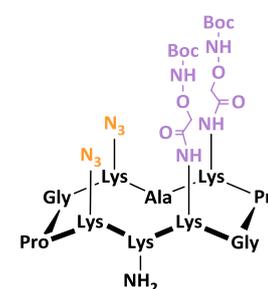


Peptide **22f**

The peptide **22e** (0.057 mmol, 90 mg) was deprotected using general procedure **II.1.4**. The crude product was used without any further purification. The crude product was purified on RP-HPLC (gradient I) and freeze-dried to obtain a white powder (0.042 mmol, 108 mg, yield: 75%).

$R_t = 1.45$ min

ESI-MS (+): m/z calcd for $C_{61}H_{104}N_{21}O_{18}$: 1418.6, m/z found: 1418.6
[M+H]⁺

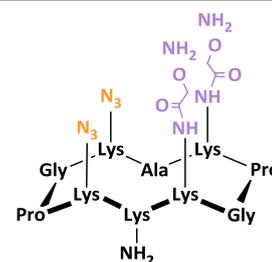


Peptide 22

The peptide **22f** (3 mg) was deprotected using general procedure **II.1.7**. The crude product was used without any further purification. The yield was considered as quantitative.

$R_t = 1.40$ min

ESI-MS (+): m/z calcd for $C_{51}H_{88}N_{21}O_{14}$: 1218.7, m/z found: 1218.3
[M+H]⁺



III. Oligonucleotides and conjugates synthesis

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

RP-HPLC analysis were performed on a Waters HPLC system using C18 Nucleosil column (Macherey-Nagel, 100 \AA , 250 x 4.6 mm, 5 μ m) with UV-monitoring at 260 nm and 280 nm. A 1 mL/min flow linear gradient was applied. Solvent A³ (50 mM triethylammonium acetate buffer with 5% acetonitrile) and solvent B³ (acetonitrile with 5% water) were used. A stepwise gradient of 0-30% B³ in 20 min followed by 30 to 100% B³ in 10 min was applied for gradient III. A stepwise gradient of 0-60% B³ in 20 min then from 60 to 100% B³ in 10 min was applied for the gradient IV. The RP-HPLC purifications of oligonucleotides were performed on a Gilson system with Nucleosil C-18 column (Macherey-Nagel, 100 \AA , 250 x 10 mm, 7 μ m) using 4 mL/min flow linear gradient from 100% solvent A³ and 0% B to 45% B³ for 20 minutes with UV-monitoring at 260 nm and 280 nm.

AELC analysis were performed on a Dionex system using Dionex column (DNAPac PA-100, 4 x 250 mm) with UV-monitoring at 260 nm and 280 nm. A 1.5 mL/min flow linear gradient from 95% solvent A⁴ (Tris buffer 0.4 M pH 7 with 5% acetonitrile) and 5% solvent B⁴ (Tris buffer 0.4 M pH 7, lithium perchlorate trihydrate buffer 0.4 M with 5% acetonitrile) to 50% B⁴ for 24 minutes was applied at 75°C. AELC purification were performed on a Dionex system using Dionex column (DNAPac PA-100, 9 x 250 mm) using 5 mL/min flow linear gradient from 95% solvent A⁴ and 5% B⁴ to 50% B⁴ for 24 minutes at 75°C with UV-monitoring at 260 nm and 280 nm.

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 $\epsilon_{260\text{nm}}$ were estimated according to the nearest neighbour model).

III.1. General procedure for oligonucleotides synthesis

III.1.1. General procedure for RNA synthesis

After synthesis, cyanoethyl protecting groups were removed using a solution of 1 M 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 then desalted and freeze-dried.

III.1.2. General procedure for 3'diol DNA synthesis

After synthesis, cyanoethyl protecting groups were removed using a solution of 1 M DBU in acetonitrile. Cleavage from the resin and deprotection were performed in ammoniac/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.

III.1.3. General procedure for DNA synthesis expected 3' diol DNA

Cleavage from the resin and deprotection were performed in ammoniac/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.

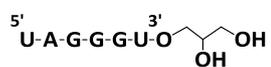
III.1.4. General procedure for 3'diol DNA and RNA oxidation

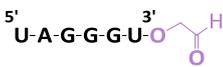
30 eq of sodium periodate were added to a solution of 3'diol oligonucleotide in water to obtain a final concentration of 10^{-3} M. The mixture was stirred for 30 min at room temperature in dark conditions. Excess of NaIO_4 was then removed by SEC. The oxidation was considered quantitative and the crude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

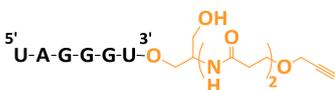
III.1.5. General procedure for 3'SSOH DNA deprotection

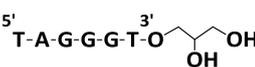
Dithiothreitol (100 eq) was added to a solution of 3'SSOH oligonucleotide in Tris.HCl buffer 1 M, pH 8.5 to obtain a final concentration of 10^{-3} M. The mixture was stirred for 1h at room temperature. The product was purified on RP-HPLC then desalted and freeze-dried.

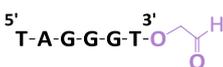
III.1.6. Experiment procedure

3'diol RNA 3b	
<p>Oligonucleotide 3' 3b diol was obtained from automated synthesis on a 3'-glyceryl CPG resin at 1 μmol scale using general procedure III.1.1. The crude product was purified on Dionex and freeze-dried (319 nmol, yield: 32%, $\epsilon_{260\text{nm}} = 62700 \text{ M}^{-1} \cdot \text{cm}^{-1}$).</p> <p>$R_t = 8.2$ min (AELC)</p> <p>ESI-MS (-): m/z calcd for $\text{C}_{61}\text{H}_{77}\text{N}_{24}\text{O}_{46}\text{P}_6$: 2067.3, m/z found: 2067.3 [M-H]⁻.</p>	 <p>The chemical structure shows the sequence 5'-U-A-G-G-G-U-O-3'. The 3' phosphate group is linked to a glycerol moiety, which has two hydroxyl groups (OH) at the 2' and 3' positions of the glycerol backbone.</p>

3'aldehyde RNA 3	
<p>Oligonucleotide 3' diol 3b (189 nmol) was oxidized using general procedure III.1.4. The crude product was used without any further purification. The yield was considered as quantitative ($\epsilon_{260\text{nm}} = 62700 \text{ M}^{-1} \cdot \text{cm}^{-1}$).</p> <p>$R_t = 8.4 \text{ min}$ (AELC)</p> <p>ESI-MS (-): m/z calcd for $\text{C}_{60}\text{H}_{73}\text{N}_{24}\text{O}_{45}\text{P}_6$: 2035.3, m/z found: 2035.2 [M-H]⁻.</p>	

3'alcyne RNA 7	
<p>Oligonucleotide 3' alcyne 7 was obtained from automated synthesis on a 3'-alkyne-modifier serinol CPG resin at 1 μmol scale using general procedure III.1.1. The crude product was purified on Dionex and freeze-dried (400 nmol, yield: 40%, $\epsilon_{260\text{nm}} = 62700 \text{ M}^{-1} \cdot \text{cm}^{-1}$).</p> <p>$R_t = 7.7 \text{ min}$ (AELC)</p> <p>ESI-MS (-): m/z calcd for $\text{C}_{70}\text{H}_{89}\text{N}_{26}\text{O}_{48}\text{P}_6$: 2248.5, m/z found: 2248.7 [M-H]⁻.</p>	

3'diol DNA 9c	
<p>Oligonucleotide 3' alcyne 9c was obtained from automated synthesis on a 3'-alkyne-modifier serinol CPG resin at 1 μmol scale using general procedure III.1.2. The crude product was purified on RP-HPLC and freeze-dried (450 nmol, yield: 45%, $\epsilon_{260\text{nm}} = 61700 \text{ M}^{-1} \cdot \text{cm}^{-1}$).</p> <p>$R_t = 9.8 \text{ min}$ (AELC)</p> <p>ESI-MS (-): m/z calcd for $\text{C}_{63}\text{H}_{81}\text{N}_{24}\text{O}_{40}\text{P}_6$: 2000.3, m/z found: 2000.5 [M-H]⁻.</p>	

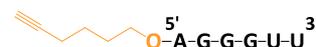
3'aldehyde DNA 9	
<p>Oligonucleotide 3' diol 9c (200 nmol) was oxidized using general procedure III.1.4. The crude product was used without any further purification. The yield was considered as quantitative ($\epsilon_{260\text{nm}} = 61700 \text{ M}^{-1} \cdot \text{cm}^{-1}$).</p> <p>$R_t = 10.2 \text{ min}$ (AELC)</p> <p>ESI-MS (-): m/z calcd for $\text{C}_{62}\text{H}_{77}\text{N}_{24}\text{O}_{39}\text{P}_6$: 1968.3, m/z found: 1968.3 [M-H]⁻.</p>	

5'alkyne RNA **10**

Oligonucleotide **10** was obtained from automated synthesis using commercially available 5'-hexynyl (β -cyanoethyl) phosphoramidite at 1 μ mol scale, using general procedure **III.1.1**. The crude product was purified on Dionex and freeze-dried (400 nmol, yield: 40%, $\epsilon_{260\text{nm}} = 63700 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 7.9 \text{ min}$ (AELC)

ESI-MS (-): m/z calcd for $\text{C}_{64}\text{H}_{79}\text{N}_{24}\text{O}_{44}\text{P}_6$: 2074.3, m/z found: 2074.3 [M-H]⁻.

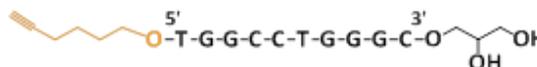


5'alkyne 3'diol DNA **12b**

Oligonucleotide **12b** was obtained from automated synthesis using commercially available 5'-hexynyl (β -cyanoethyl) phosphoramidite 3'-glyceryl CPG resin at 1 μ mol scale, using general procedure **III.1.2**. The crude product was purified on RP-HPLC and freeze-dried (500 nmol, yield: 50%, $\epsilon_{260\text{nm}} = 86800 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 14.7 \text{ min}$ (RP-HPLC, gradient III)

ESI-MS (-): m/z calcd for $\text{C}_{106}\text{H}_{138}\text{N}_{38}\text{O}_{68}\text{P}_{11}$: 3373.1, m/z found: 3372.8 [M-H]⁻.



5'alkyne 3'aldehyde DNA **12**

Oligonucleotide **12b** (330 nmol) was oxidized using general procedure **III.1.4**. The crude product was used without any further purification. The yield was considered as quantitative ($\epsilon_{260\text{nm}} = 86800 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 14.8 \text{ min}$ (RP-HPLC, gradient III)

ESI-MS (-): m/z calcd for $\text{C}_{105}\text{H}_{134}\text{N}_{38}\text{O}_{67}\text{P}_{11}$: 3341.1, m/z found: 3340.9 [M-H]⁻.



5'S-trt 3'NH₂ DNA **13b**

Oligonucleotide **13b** was obtained from automated synthesis using commercially available 3'-PT-amino-modifier C3 CPG resin and 5'-thiol-modifier C6 at 1 μmol scale, using general procedure III.1.3. The crude product was purified on RP-HPLC and freeze-dried (410 nmol, yield: 41%, $\epsilon_{260\text{nm}} = 89500 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 19.5$ min (RP-HPLC, gradient IV)

ESI-MS (-): m/z calcd for C₁₁₇H₁₄₄N₄₁O₅₈P₁₀S: 3394.4, m/z found: 3394.3 [M-H]⁻.

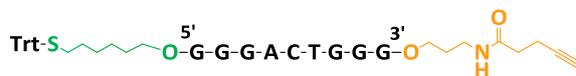


5'S-trt 3'alkyne DNA **13c**

4-pentynoic acid succinimidyl ester (30 eq, 10.8 μmol, 2.1 mg) was added to a H₂O/DMF (9/1) solution of oligonucleotide **13b** (1 eq, 360 nmol) to obtain a final concentration of 10⁻³ M. The pH was adjusted to 8-9 using DIEA. The solution was stirred for 4h at room temperature. The product was then desalted on NAP 25 and the fractions were collected to obtain the crude product (UV-monitoring at 260 nm). The coupling was considered quantitative and the crude was used in the next step without further purification ($\epsilon_{260\text{nm}} = 89500 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 19.5$ min (RP-HPLC, gradient IV)

ESI-MS (-): m/z calcd for C₁₂₂H₁₄₈N₄₁O₅₉P₁₀S: 3474.5, m/z found: 3474.2 [M-H]⁻.



5'SH 3'alkyne DNA **13**

Oligonucleotide **13c** (352 nmol) was dissolved in 0.1M triethylammonium acetate buffer (pH 6.5, concentration 10^{-3} M), 1M aqueous silver nitrate solution (0.15 vol) and acetic acid (0.5 vol) were added. The mixture was stirred for 45 min at room temperature. 1M DTT aqueous solution (0.20 vol) was next added and the mixture was stirred for 15 min at room temperature. The mixture was then centrifuged, the supernatant was conserved and the pellet was washed with 0.1M triethylammonium acetate buffer (pH 6.5). The product was desalted on NAP 25 and the fractions were collected to obtain the crude product (UV-monitoring at 260 nm). The coupling was considered quantitative and the crude oligonucleotide **13** was used in the next step without further purification ($\epsilon_{260\text{nm}} = 89500 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 12.8$ min (RP-HPLC, gradient IV)

ESI-MS (-): m/z calcd for $\text{C}_{103}\text{H}_{134}\text{N}_{41}\text{O}_{59}\text{P}_{10}\text{S}$: 3232.2, m/z found: 3231.9 [M-H]⁻.



5'alkyne 3'diol DNA **16b**

Oligonucleotide **16b** 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 **III.1.2**. The crude product was purified on RP-HPLC and freeze-dried (450 nmol, yield: 45%, $\epsilon_{260\text{nm}} = 114800 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 13.8$ min (RP-HPLC, gradient III)

ESI-MS (-): m/z calcd for $\text{C}_{119}\text{H}_{151}\text{N}_{46}\text{O}_{73}\text{P}_{12}$: 3765.4, m/z found: 3765.3 [M-H]⁻.



5'alkyne 3'aldehyde DNA **16**

Oligonucleotide **16b** (400 nmol) was oxidized using general procedure **III.1.4**. The crude product was used without any further purification. The yield was considered as quantitative ($\epsilon_{260\text{nm}} = 114800 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 14.0$ min (RP-HPLC, gradient III)

ESI-MS (-): m/z calcd for $\text{C}_{118}\text{H}_{147}\text{N}_{46}\text{O}_{72}\text{P}_{12}$: 3733.4, m/z found: 3733.3 [M-H]⁻.

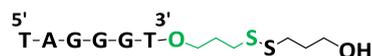


3'SSOH DNA **17c**

Oligonucleotide **17c** was obtained from automated synthesis on a 3'-thiol-modifier C3 S-S CPG resin at 1 μmol scale using general procedure **III.1.3**. The crude product was purified on RP-HPLC and freeze-dried (410 nmol, yield: 41%, $\epsilon_{260\text{nm}} = 61700 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 15.7$ min (RP-HPLC, gradient III)

ESI-MS (-): m/z calcd for $\text{C}_{66}\text{H}_{87}\text{N}_{24}\text{O}_{39}\text{P}_6\text{S}_2$: 2089.3, m/z found: 2089.5 [M-H]⁻.

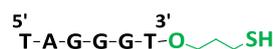


3'SH DNA **17**

Oligonucleotide **17c** (350 nmol) was deprotected using general procedure **III.1.5**. The crude product was purified on RP-HPLC and freeze-dried (259 nmol, yield: 74%, $\epsilon_{260\text{nm}} = 61700 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 17.3$ min (RP-HPLC, gradient III)

ESI-MS (-): m/z calcd for $\text{C}_{63}\text{H}_{81}\text{N}_{24}\text{O}_{38}\text{P}_6\text{S}$: 2000.4, m/z found: 1999.9 [M-H]⁻.



3'SSOH DNA **18c**

Oligonucleotide **18c** was obtained from automated synthesis on a 3'-thiol-modifier C3 S-S CPG resin at 1 μmol scale using general procedure **III.1.3**. The crude product was purified on RP-HPLC and freeze-dried (460 nmol, yield: 46%, $\epsilon_{260\text{nm}} = 114800 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 15.8 \text{ min}$ (RP-HPLC, gradient III)

ESI-MS (-): m/z calcd for $\text{C}_{116}\text{H}_{149}\text{N}_{46}\text{O}_{69}\text{P}_{11}\text{S}_2$: 3694.6, m/z found: 3695.2 [M-H]⁻.



3'SH DNA **18**

Oligonucleotide **18c** (400 nmol) was deprotected using general procedure **III.1.5**. The crude product was purified on RP-HPLC and freeze-dried (288 nmol, yield: 72%, $\epsilon_{260\text{nm}} = 114800 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 17.8 \text{ min}$ (RP-HPLC, gradient III)

ESI-MS (-): m/z calcd for $\text{C}_{113}\text{H}_{142}\text{N}_{46}\text{O}_{68}\text{P}_{11}\text{S}$: 3605.4, m/z found: 3605.4 [M-H]⁻.

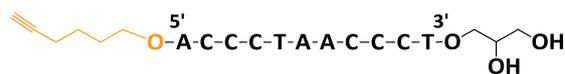


5'alkyne 3'diol DNA **20b**

Oligonucleotide **20b** 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 **III.1.2**. The crude product was purified on RP-HPLC and freeze-dried (450 nmol, yield: 45%, $\epsilon_{260\text{nm}} = 98100 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 16.9 \text{ min}$ (RP-HPLC, gradient III)

ESI-MS (-): m/z calcd for $\text{C}_{113}\text{H}_{150}\text{N}_{37}\text{O}_{71}\text{P}_{12}$: 3534.3, m/z found: 3533.9 [M-H]⁻.



5'alkyne 3'aldehyde DNA 20	
<p>Oligonucleotide 20b (400 nmol) was oxidized using general procedure III.1.4. The crude product was used without any further purification. The yield was considered as quantitative ($\epsilon_{260\text{nm}} = 98100 \text{ M}^{-1} \cdot \text{cm}^{-1}$).</p> <p>$R_t = 17.1 \text{ min}$ (RP-HPLC, gradient III)</p> <p>ESI-MS (-): m/z calcd for $\text{C}_{112}\text{H}_{146}\text{N}_{37}\text{O}_{70}\text{P}_{12}$: 3502.2, m/z found: 3502.6 [M-H]⁻.</p>	

5'alkyne 3'SSOH DNA 21b	
<p>Oligonucleotide 21b was obtained from automated synthesis using commercially available 5'-hexynyl (β-cyanoethyl) phosphoramidite and 3'-thiol-modifier C3 S-S CPG resin at 1 μmol scale using general procedure III.1.3. The crude product was purified on RP-HPLC and freeze-dried (440 nmol, yield: 44%, $\epsilon_{260\text{nm}} = 98100 \text{ M}^{-1} \cdot \text{cm}^{-1}$).</p> <p>$R_t = 16.8 \text{ min}$ (RP-HPLC, gradient III)</p> <p>ESI-MS (-): m/z calcd for $\text{C}_{116}\text{H}_{156}\text{N}_{37}\text{O}_{70}\text{P}_{12}\text{S}_2$: 3624.5, m/z found: 3625.5 [M-H]⁻.</p>	

5'alkyne 3'SH DNA 21	
<p>Oligonucleotide 21b (400 nmol) was deprotected using general procedure III.1.5. The crude product was purified on RP-HPLC and freeze-dried (222 nmol, yield: 77%, $\epsilon_{260\text{nm}} = 98100 \text{ M}^{-1} \cdot \text{cm}^{-1}$).</p> <p>$R_t = 17.9 \text{ min}$ (RP-HPLC, gradient III)</p> <p>ESI-MS (-): m/z calcd for $\text{C}_{113}\text{H}_{151}\text{N}_{37}\text{O}_{69}\text{P}_{12}\text{S}$: 3534.4, m/z found: 3534.9 [M-H]⁻.</p>	

III.2. General procedure for conjugates synthesis

III.2.1. General procedure for oxime ligations

Aldehyde oligonucleotide (1.2 eq by aminoxy function, concentration 10^{-3} M) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5) and aminoxy peptide was added. The solution was stirred at 55°C for 45 min. The product was purified on RP-HPLC or AELC then desalted and freeze-dried.

III.2.2. General procedure for CuAAC reactions in intermolecular condition

To a solution of alkyne oligonucleotide (1.2 eq, concentration 10^{-3} M) and azide peptide (1 eq) in 100 mM HEPES buffer (pH 7.4) were added CuSO_4 (3 eq), THPTA (15 eq) and sodium ascorbate (15 eq). The reaction was stirred at 37°C for 2h and quenching with 0.5 M EDTA solution (50 eq). The product was purified on RP-HPLC or AELC then desalted and freeze-dried.

III.2.3. General procedure for CuAAC reactions in intramolecular condition

To a solution of alkyne-azide conjugate (1.2 eq, concentration 10^{-4} M) in 100 mM HEPES buffer (pH 7.4) were added CuSO_4 (6 eq), THPTA (30 eq) and sodium ascorbate (30 eq). The reaction was stirred at 37°C for 2h and quenching with 0.5 M EDTA solution (50 eq). The product was purified on RP-HPLC then desalted and freeze-dried.

III.2.4. General procedure for thiol-chloroacetamide coupling

To a solution of thiol oligonucleotide (2 eq, concentration 10^{-3} M) and chloroacetamide peptide (1 eq) in a solution of $\text{H}_2\text{O}/\text{CH}_3\text{CN}$ 9/1 were added TCEP (2 eq), 500 mM KCl, DIEA (45 eq), KI (60 eq). The reaction was stirred at room temperature for 3h. The product was purified on RP-HPLC then desalted and freeze-dried.

III.2.5. General procedure for diazo-transfer reaction

To a solution of amine conjugate (1 eq, concentration 3.10^{-4} M) in a solution of NaHCO_3 50 mM/(MeOH/ H_2O , 1/1) 75/25 at pH 8.5 were added CuSO_4 (2.8 eq) and ISAHC (20 eq). The reaction was stirred at 60°C for 30 min and quenching with 0.5 M EDTA solution (50 eq). The product was purified on RP-HPLC then desalted and freeze-dried.

III.2.6. Experiment procedure

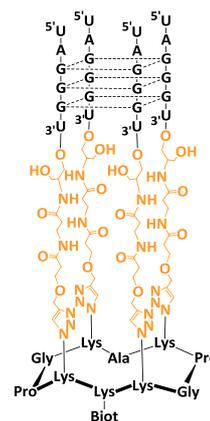
Conjugate C	
<p>The oxime ligation was carried out with aldehyde containing oligonucleotide 3 (4.8 eq, 153.6 nmol) and peptide 1 (1 eq, 32 nmol) according to the general procedure III.2.1. The crude was purified on AELC then conjugate C was desalted by SEC and freeze dried. (15.6 nmol, yield: 48%, $\epsilon_{260\text{nm}} = 250008 \text{ M}^{-1}\cdot\text{cm}^{-1}$).</p> <p>$R_t = 19.6$ min (AELC)</p> <p>MALDI-MS (-): m/z calcd for $\text{C}_{305}\text{H}_{398}\text{N}_{117}\text{O}_{196}\text{P}_{24}\text{S}$: 9614.5, m/z found: 9615.8 $[\text{M}-\text{H}]^-$.</p>	

Conjugate D

CuAAC was carried out with alkyne containing oligonucleotide **7** (4.8 eq, 250 nmol) and peptide **4** (52 nmol) according to the general procedure III.2.2. The crude was purified on AELC then conjugate **D** was desalted by SEC and freeze dried. (26 nmol, yield: 50%, $\epsilon_{260\text{nm}} = 250008 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 18.5 \text{ min}$ (AELC)

MALDI-MS (-): m/z calcd for $\text{C}_{337}\text{H}_{450}\text{N}_{129}\text{O}_{204}\text{P}_{24}\text{S}$: 10342.2, m/z found: 10342.2 $[\text{M-H}]^-$.

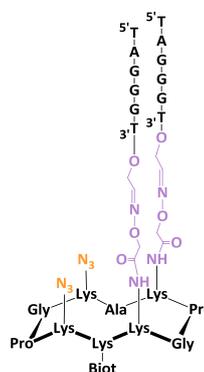


Conjugate E₁

The oxime ligation was carried out with aldehyde containing oligonucleotide **9** (2.4 eq, 367.2 nmol) and peptide **8** (1 eq, 153 nmol) according to the general procedure III.2.1. The crude was purified on RP-HPLC then conjugate **E₁** was desalted by SEC and freeze dried. (73.4 nmol, yield: 48%, $\epsilon_{260\text{nm}} = 123400 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 12.6 \text{ min}$ (AELC)

MALDI-MS (-): m/z calcd for $\text{C}_{185}\text{H}_{252}\text{N}_{71}\text{O}_{92}\text{P}_{12}\text{S}$: 5346.2, m/z found: 5348.5 $[\text{M-H}]^-$.

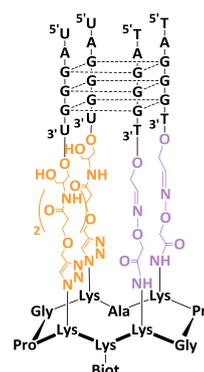


Conjugate E

CuAAC was carried out with alkyne containing oligonucleotide **7** (2.4 eq, 166.8 nmol) and conjugate **E₁** (69.5 nmol) according to the general procedure III.2.2. The crude was purified on AELC then conjugate **E** was desalted by SEC and freeze dried. (34.8 nmol, yield: 50%, $\epsilon_{260\text{nm}} = 247800 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 17.5 \text{ min}$ (AELC)

MALDI-MS (-): m/z calcd for $\text{C}_{325}\text{H}_{431}\text{N}_{123}\text{O}_{188}\text{P}_{24}\text{S}$: 9844.0, m/z found: 9846.7 $[\text{M-H}]^-$.

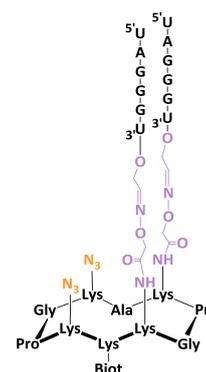


Conjugate F₁

The oxime ligation was carried out with aldehyde containing oligonucleotide **3** (2.4 eq, 498 nmol) and peptide **8** (1 eq, 207.5 nmol) according to the general procedure **III.2.1**. The crude was purified on AELC then conjugate **F₁** was desalted by SEC and freeze dried. (99.6 nmol, yield: 48%, $\epsilon_{260\text{nm}} = 125400 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 14.2 \text{ min}$ (AELC)

MALDI-MS (-): m/z calcd for $\text{C}_{181}\text{H}_{245}\text{N}_{71}\text{O}_{104}\text{P}_{12}\text{S}$: 5483.0, m/z found: 5481.8 $[\text{M-H}]^-$.

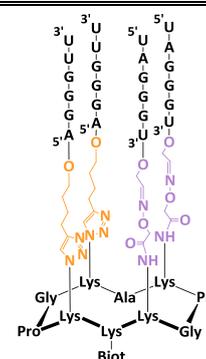


Conjugate F

CuAAC was carried out with alkyne containing oligonucleotide **10** (2.4 eq, 174 nmol) and conjugate **F₁** (72.5 nmol) according to the general procedure **III.2.2**. The crude was purified on AELC then conjugate **F** was desalted by SEC and freeze dried. (24 nmol, yield: 33%, $\epsilon_{260\text{nm}} = 251800 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 18.6 \text{ min}$ (AELC)

MALDI-MS (-): m/z calcd for $\text{C}_{309}\text{H}_{405}\text{N}_{119}\text{O}_{192}\text{P}_{24}\text{S}$: 9633.6, m/z found: 9632.5 $[\text{M-H}]^-$.

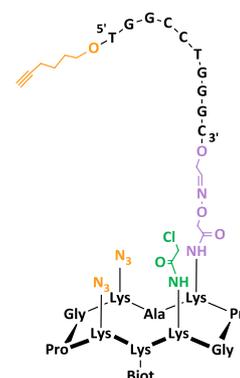


Conjugate G₁

The oxime ligation was carried out with aldehyde containing oligonucleotide **12** (1 eq, 1.3 μmol) and peptide **11** (2 eq, 1.6 μmol) according to the general procedure **III.2.1**. The crude was purified on RP-HPLC then conjugate **G₁** was desalted by SEC and freeze dried. (1.1 μmol , yield: 84%, $\epsilon_{260\text{nm}} = 86800 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 21.7 \text{ min}$ (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{166}\text{H}_{231}\text{N}_{60}\text{O}_{81}\text{P}_{11}\text{S}\text{Cl}$: 4771.2, m/z found: 4770.8 $[\text{M-H}]^-$.

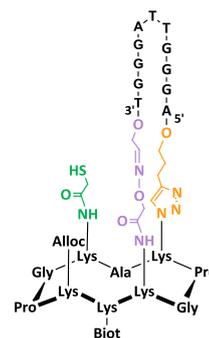


Conjugate H₂

CuAAC was carried out with conjugate H₁ (237 nmol) according to the general procedure III.2.3. The crude was purified on RP-HPLC then conjugate H₂ was desalted by SEC and freeze dried. (156 nmol, yield: 66%, $\epsilon_{260\text{nm}} = 114800 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 19.1 \text{ min}$ (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{183}\text{H}_{250}\text{N}_{66}\text{O}_{88}\text{P}_{12}\text{SCl}$: 5221.5, m/z found: 5221.4 [M-H]⁻.

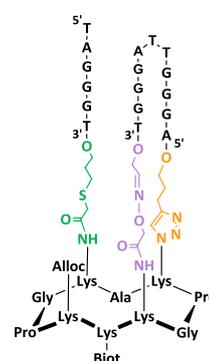


Conjugate H

The thiol chloroacetamide coupling was carried out with thiol containing oligonucleotide **17** (2 eq, 100 nmol) and conjugate H₂ (1 eq, 50 nmol) according to the general procedure III.2.4. The crude was purified on RP-HPLC then conjugate H was desalted by SEC and freeze dried. (25 nmol, yield: 50%, $\epsilon_{260\text{nm}} = 176500 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 17.9 \text{ min}$ (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{246}\text{H}_{329}\text{N}_{90}\text{O}_{126}\text{P}_{18}\text{S}_2$: 7184.4, m/z found: 7186.7 [M-H]⁻.

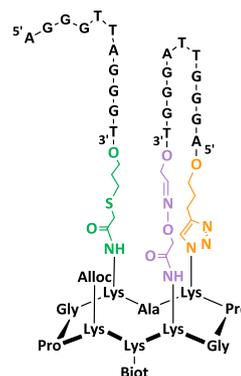


Conjugate I

The thiol chloroacetamide coupling was carried out with thiol containing oligonucleotide **18** (2 eq, 100 nmol) and conjugate H₂ (1 eq, 50 nmol) according to the general procedure III.2.4. The crude was purified on RP-HPLC then conjugate I was desalted by SEC and freeze dried. (30 nmol, yield: 60%, $\epsilon_{260\text{nm}} = 229600 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 17.1 \text{ min}$ (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{296}\text{H}_{390}\text{N}_{112}\text{O}_{156}\text{P}_{23}\text{S}_2$: 8789.4, m/z found: 8792.1 [M-H]⁻.

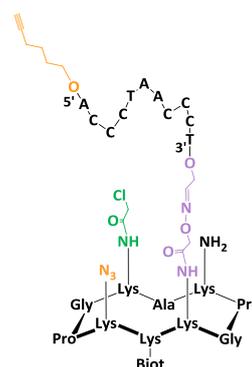


Conjugate J₁

The oxime ligation was carried out with aldehyde containing oligonucleotide **20** (1 eq, 500 nmol) and peptide **19** (2 eq, 1 μmol) according to the general procedure **III.2.1**. The crude was purified on RP-HPLC then conjugate **J₁** was desalted by SEC and freeze dried. (325 nmol, yield: 65%, $\epsilon_{260\text{nm}} = 98100 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 22.3$ min (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{173}\text{H}_{245}\text{N}_{57}\text{O}_{84}\text{P}_{12}\text{SCl}$: 4906.3, m/z found: 4905.8 $[\text{M-H}]^-$.

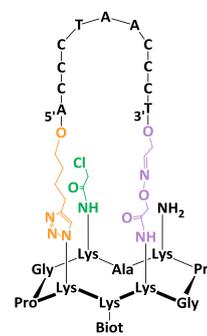


Conjugate J₂

CuAAC was carried out with conjugate **J₁** (300 nmol) according to the general procedure **III.2.3**. The crude was purified on RP-HPLC then conjugate **J₂** was desalted by SEC and freeze dried. (216 nmol, yield: 72%, $\epsilon_{260\text{nm}} = 98100 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 19.1$ min (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{173}\text{H}_{245}\text{N}_{57}\text{O}_{84}\text{P}_{12}\text{SCl}$: 4906.3, m/z found: 4907.2 $[\text{M-H}]^-$.

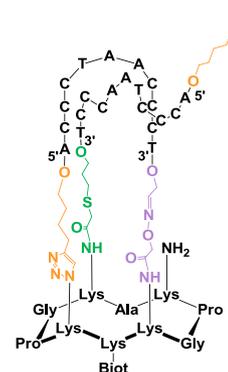


Conjugate J₃

The thiol chloroacetamide coupling was carried out with thiol containing oligonucleotide **21** (2 eq, 200 nmol) and conjugate **J₂** (1 eq, 100 nmol) according to the general procedure **III.2.4**. The crude was purified on RP-HPLC then conjugate **J₃** was desalted by SEC and freeze dried. (74 nmol, yield: 74%, $\epsilon_{260\text{nm}} = 229600 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 17.5$ min (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{286}\text{H}_{395}\text{N}_{94}\text{O}_{153}\text{P}_{24}\text{S}_2$: 8405.2, m/z found: 8403.5 $[\text{M-H}]^-$.

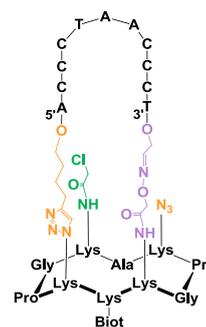


Conjugate J₄

The diazo transfer reaction was carried out conjugate J₂ (1 eq, 100 nmol) according to the general procedure III.2.5. The crude was purified on RP-HPLC then conjugate J₄ was desalted by SEC and freeze dried. (50 nmol, yield: 50%, $\epsilon_{260\text{nm}} = 98100 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 18.8 \text{ min}$ (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{173}\text{H}_{243}\text{N}_{59}\text{O}_{84}\text{P}_{12}\text{SCl}$: 4932.3, m/z found: 4933.3 [M-H]⁻.

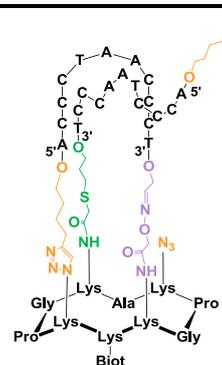


Conjugate J₅ by “voie A”

The diazo transfer reaction was carried out conjugate J₃ (1 eq, 40 nmol) according to the general procedure III.2.5. The crude was purified on RP-HPLC then conjugate J₅ was desalted by SEC and freeze dried. (31.6 nmol, yield: 79%, $\epsilon_{260\text{nm}} = 229600 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 17.7 \text{ min}$ (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{286}\text{H}_{393}\text{N}_{96}\text{O}_{153}\text{P}_{24}\text{S}_2$: 8430.8, m/z found: 8433.8 [M-H]⁻.

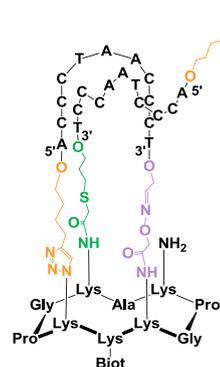


Conjugate J₅ by “voie B”

The thiol chloroacetamide coupling was carried out with thiol containing oligonucleotide **21** (2 eq, 80 nmol) and conjugate J₄ (1 eq, 40 nmol) according to the general procedure III.2.4. The crude was purified on RP-HPLC then conjugate J₅ was desalted by SEC and freeze dried. (24 nmol, yield: 60%, $\epsilon_{260\text{nm}} = 229600 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

$R_t = 17.8 \text{ min}$ (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{286}\text{H}_{393}\text{N}_{96}\text{O}_{153}\text{P}_{24}\text{S}_2$: 8430.8, m/z found: 8431.2 [M-H]⁻.

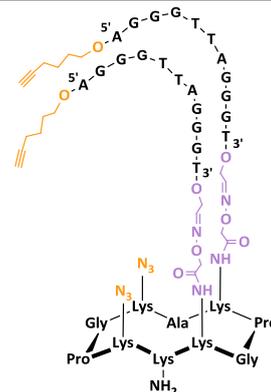


Conjugate K₁

The oxime ligation was carried out with aldehyde containing oligonucleotide **16** (2.4 eq, 800 nmol) and peptide **22** (1 eq, 1.08 μmol) according to the general procedure **III.2.1**. The crude was purified on RP-HPLC then conjugate **K₁** was desalted by SEC and freeze dried. (650 nmol, yield: 61%, $\epsilon_{260\text{nm}} = 229600 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 18.2 \text{ min}$ (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{287}\text{H}_{378}\text{N}_{113}\text{O}_{156}\text{P}_{24}$: 8650.1, m/z found: 8652.6 [M-H]⁻.

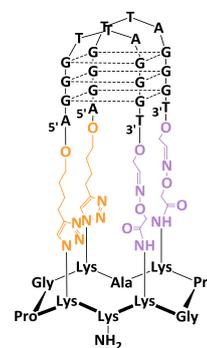


Conjugate K₂

CuAAC was carried out with conjugate **K₁** (590 nmol) according to the general procedure **III.2.3**. The crude was purified on RP-HPLC then conjugate **K₂** was desalted by SEC and freeze dried. (478 nmol, yield: 81%, $\epsilon_{260\text{nm}} = 229600 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 16.1 \text{ min}$ (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{287}\text{H}_{378}\text{N}_{113}\text{O}_{156}\text{P}_{24}$: 8650.1, m/z found: 8650.5 [M-H]⁻.

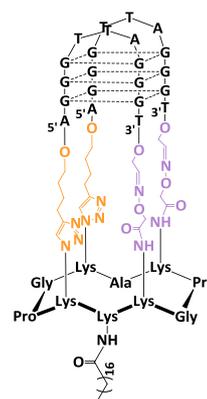


Conjugate K

To a solution of conjugate **K₂** (1 eq, 290 nmol) in H₂O (100 mL) was added a solution of N-hydroxysuccinimidyl C18 linker (60 eq, 0.017 mmol, 3.4 mg) in DMF (100 μL). The pH was adjusted to 8-9 using DIEA and the mixture was stirred for 2h. The crude was desalted by SEC and used without any further purification. The yield was considered as quantitative ($\epsilon_{260\text{nm}} = 229600 \text{ M}^{-1}\cdot\text{cm}^{-1}$).

$R_t = 12.1 \text{ min}$ (RP-HPLC, gradient III)

MALDI-MS (-): m/z calcd for $\text{C}_{305}\text{H}_{412}\text{N}_{113}\text{O}_{157}\text{P}_{24}$: 8912.1, m/z found: 8611.3 [M-H]⁻.



IV. General procedure for Thermal difference spectrum

Thermal difference spectrum resulting from the subtraction of the 5°C UV spectrum from the 80°C or 90°C UV spectrum. UV spectrums were recorded on a Varian Cary 400 UV-visible spectrophotometer using 1 cm length quartz cuvette. Spectra were recorded at 5°C and 80°C or 5°C and 90°C with wavelength range of 220 to 320 nm.

V. General procedure for circular dichroism studies

Circular dichroism studies were performed after thoroughly desalting of the products by SEC on NAP 25 cartridge. A step of annealing was applied by heating the sample at 80°C for 5 min in buffer (Tris 10 mM pH 7.4 without salt or with salt) and cooling it over 2h to room temperature. Analyses were recorded on a Jasco J-810 spectropolarimeter using 1 cm length quartz cuvette. Spectra were recorded at 20°C or every 5°C in a range of 5 to 80°C or 5 to 90°C with wavelength range of 220 to 330 nm. For each temperature, the spectrum was an average of three scans with a 0.5 s response time, a 1 nm data pitch, a 4 nm bandwidth and a 200 nm.min⁻¹ scanning speed. Melting temperatures were obtained using Boltzmann fit on Origin software. Each curve fit was only accepted with a $r_{\text{value}} > 0.99$.

VI. General procedure for NMR experiments

NMR spectra were carried out at 80 mm and at 25°C. The program uses p11gpse water suppression using 1-1 echo pulse sequence with spoil gradient pulse. The gradient file name is SMSQ1.100 that uses a 20% gradient ratio for 600 ms gradient pulse. The delay for binomial water suppression is $D19 = 60 \mu\text{sec}$ ($D19 = 1/(2*d)$ in which $d = \text{distance of next null (in Hz)}$). The experiments were recorded with 40960 data points. 7168 scans were acquired over a spectral width of 10000 Hz.

VII. General procedure for SPR and BLI experiments

VII.1. General procedure for SPR experiments

SPR data were obtained using a BIAcore T200 (GE Healthcare). Measurements were performed on gold surfaces. Sensor chips were cleaned by UV-ozone treatment for 10 min followed by rinsing with MilliQ water and ethanol. The cleaned gold surfaces were then functionalized according to the following procedure. First, mixed self-assembled monolayers (SAMs) were formed at room temperature by dipping gold sensors in the thiol mixture overnight: 90% HS-(CH₂)₁₁-EG4-OH and 10% HS-(CH₂)₁₁-EG6-Biotin (1 mM total thiol concentration in EtOH). After overnight adsorption, gold sensors were rinsed with ethanol and dried under nitrogen. The surface was then inserted into the BIAcore T200 device. All measurements were performed at 25°C by using a running buffer (RB) composed of HEPES buffered saline: 0.01M HEPES pH 7.4, 50 mM NaCl, 100 mM KCl, and 0.005% or 0.5% v/v surfactant P20. Streptavidin (100 ng.mL⁻¹) was injected (10 $\mu\text{L.min}^{-1}$) onto the biotinylated SAM until saturation of the surface (around 2500 RU). The different G-quadruplex systems of motifs **A**, **B**, **C**, **E** and **G** were injected at 2 mL.min⁻¹ onto the streptavidin-coated SAM surfaces until saturation (final response approximately 300 RU). Binding experiments were conducted at 30 mL.min⁻¹ (or 80 mL.min⁻¹ for PhenDC3 and PDS) by injection of the different ligands (dissolved in RB, injection time: 200 s, dissociation time: 300 s, stabilization time: 300 s) at different concentrations by using a multicycle method. A streptavidin surface, prepared as described above, was used as reference. Curves obtained on the reference surface were deduced from the curves recorded on the recognition one, thereby allowing elimination of refractive index changes owing to buffer effects.

VII.2. General procedure for BLI experiments

BLI data were obtained using Octet Red 96. BLI sensors with SA were purchased from Forte Bio (PALL). Prior to use, they were immersed 10 minutes in buffer before functionalization to dissolve the sucrose layer. Then the surfaces were dipped in G-quadruplex or G-triplex solution of motifs **H** and **I** (association time: 900 s, stabilization time: 600 s). During the binding experiments, the sensors containing motifs **H** or **I** were dipped in ligand solution (association time: 120 s, dissociation time: 180 s) at different concentrations by using kinetic experiment. A sensor without G-quadruplex or G-triplex was used as reference.

VII.3. Analysis of SPR and BLI experiments

The binding rate constants of G-quadruplex/ligand interactions were calculated by means of a nonlinear analysis of the association and dissociation curves by using the apparatus software (Biacore evaluation T200 for SPR and Data analysis 9.0 for BLI). The data were fitted by using a heterogeneous ligand model (two interactions per G4 target). The association rate constants, k_{on1} and k_{on2} , and the dissociation rate constants, k_{off1} and k_{off2} , as well as the theoretical maximal response, R_{max1} and R_{max2} , of the two interactions were determined. The equilibrium dissociation constants were obtained from the binding rate constants as $K_{D1}=k_{off1}/k_{on1}$ and $K_{D2}=k_{off2}/k_{on2}$. We report the thermodynamic dissociation constants that were consistent between independent experiments and for which the theoretical maximum response (R_{max}) is consistent with 1:1 interactions. We chose not to report values for the second interaction that might involve non-stoichiometric binding and/or non-specific (ligands stacking) interactions. The reported values are the mean of representative independent experiments, and the errors provided are standard deviations from the mean.

VIII. General procedure for SELEX experiments

VIII.1. SELEX procedure

Capillary electrophoresis selection was performed using a Beckman Coulter P/ACE MDQ system (Fullerton, California, USA) with exchangeable UV absorbance and laser-induced fluorescence (LIF) detectors (λ_{ex} 488 nm and λ_{em} 520 nm) equipped with a 80 cm total length (70 cm effective length) fused silica capillary (Polymicro Technologies Inc., Phoenix, Arizona, USA) and 75 μ m inner diameter. Prior to the sample injection, the capillary was conditioned through the following washing steps at 20 psi: 1 M NaOH for 5 min, water for 5 min and migration buffer (50 mM Tris, pH 7.8, 6 mM Triton X100) for 30 min. The voltage and temperature applied during separation was respectively 25 kV and 15°C for all the experiments.

Before selection, an 1 μ M modified ssDNA library was heated to 80° C for 5 min in 4X binding buffer (final concentrations: 50 mM Tris, pH 7.8, 50 mM NaCl, 1.5 mM MgCl₂; 6 mM Triton) and left at room temperature for 10 min. The G4-C18 was then added to the library to a final concentration of 150 μ M for the first round of selection, 15 μ M and 0.15 μ M for rounds 2 and 3, respectively. The ssDNA library was allowed to incubate with the target at room temperature for at least 20 min to ensure that the binding reached equilibrium. The equilibrated sample was injected by applying 0.7 psi to the capillary inlet for 40 seconds and detected using UV at 200 nm to monitor the separation.

Approximately 209 nL were injected, corresponding to almost $1.2 \cdot 10^{11}$ sequences were introduced into the capillary during the first selection round. During a selection round, the eluate was collected into 30 μL migration buffer until the unbound DNA began to elute. The injection, separation and collection process was repeated nine more times for round 1, 2 and 3. Input DNA concentration was determined by absorbance at 260 nm on a Shimadzu UV mini-1240 spectrophotometer with a Tray Cell from Hellma Analytics (*Müllheim*, Germany). Subsequent rounds of selection used collected DNA, amplified and purified from the previous round as the input DNA. The concentration of input DNA in later selection rounds was also 1 μM , approximately 10^{12} sequences were introduced into the capillary in one round of selection.

VIII.2. PCR Amplification and purification of sequences by gel electrophoresis

The initial modified library and single-stranded DNA candidates were generated using a 5'-FAM-reverse primer and a heavier forward primer during each round of the selection procedure. The forward primer is made heavier with a succession of six C_3 links extended with a 5' DNA stretch of 20 nucleotides. During PCR, this six C_3 region and the 20-nucleotide stretch cannot be amplified by the Vent DNA polymerase. A PCR product with two strands of unequal length is consequently synthesized. Each strand is then easily purified on a denaturing polyacrylamide gel. All PCRs were performed using a Biometra cycler from Labgene (Archamps, France). Master mix was made by combining 650 μL nuclease-free water, 80 μL deoxyribonucleotide triphosphate (dNTPs with Tryptamino-dUTP) (2.5 mM of each) (Invitrogen, Cergy-Pontoise, France), 10 μL each of heavier forward and reverse primers (10^{-4} M), 30 μL MgCl_2 (100 mM), and 100 μL GeneAmp 10X PCR buffer II. After mixing, 20 μL ($0.04 \text{ U} \cdot \mu\text{L}^{-1}$) of the Vent DNA polymerase were added. To finish, 100 μL of DNA collected during selection were added. This mixed solution was divided equally over ten thin-walled tubes that were subjected to PCR. The thermal cycling regime was as follows: initial denaturation for 5 min at 95 °C, and then cycling for 30 s at 95 °C, 30 s at 60 °C and 60 s at 72 °C for 29 cycles. After a Nanosep[®] 3K purification (Pall, Washington, New York, USA), the samples, which contained amplified products, were resolved on a 12% acrylamide gel at an applied voltage of 300 V. The band corresponding to the selected aptamers migrated with the same velocity of a77-FAM-library (used as a witness) and was visualized by UV-shadow method at 254 nm. After that, it was cut and eluted for 1 hour at 65 °C and 1 hour at 4 °C, in 1 mL of the extraction buffer (100 mM Tris-HCl, 500 mM NaCl, 1 mM EDTA; pH 7.4). To remove acrylamide, the extracted product was transferred into a Nanosep[®] 3K device, the retention membrane was replaced by glass wool and the system was centrifuged during 10 min at 14,000 rpm with MIKRO 220R centrifuge ($r = 87 \text{ mm}$) (Hettich centrifuges, Tuttlingen, Germany). Filtrate was removed, transferred in another native Nanosep[®] 3K device and centrifuged for 90 min at 5,000 rpm at 15 °C. The material retained on the top of the porous membrane was washed with 100 μL of water and centrifuged again. Finally ssDNA amount for each round was quantified by UV absorbance at 260 nm for binding affinity analysis of DNA pools. Three rounds of selection were performed, and the progress of selection was monitored using NECEEM to estimate the bulk affinity.

Subsequently, enriched modified DNAs from round 1, 2 and 3 were amplified by polymerase chain reaction (PCR) using AmpliTaq Gold DNA polymerase with nonlabeled forward and reverse primers and the four natural dNTPs. The amplified samples were sent for cloning and sequencing realized by Biofidal (Vaulx en Velin, France).

VIII.3. Monitoring of the SELEX progress

NECEEM-based measurement

CE separation and incubation buffer conditions were identical to those used during the SELEX process. Equilibrium mixture was composed of 100 nM 5'-FAM-ssDNA (purified modified library or purified selected pools after PCR amplification) and G4-C18 (40 μ M or 80 μ M) incubated for 30 min. Samples were prepared as previously seen and were injected by applying 0.7 psi pressure at the inlet of the capillary for 20 s. Detection of species was insured by LIF detector.

Références bibliographiques

1. Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618–619 (1986).
2. Dahm, R. Friedrich Miescher and the discovery of DNA. *Dev. Biol.* **278**, 274–288 (2005).
3. Paweletz, N. Walther Flemming: pioneer of mitosis research . *Nat Rev Mol Cell Biol* **2**, 72–75 (2001).
4. Avery, O. T., MacLeod, C. M. & McCarty, M. Studies on the Chemical Nature Of the Substance Inducing transformation of Pneumococcal types. *J. Exp. Med.* **79**, 137–158 (1944).
5. Watson, J. D. & Crick, F. H. C. Molecular Structure Of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
6. Tipson, R. S. Phoebus Aaron Theodor Levene 1869–1940. *Adv. Carbohydr. Chem.* **12**, 1–12 (1957).
7. Chargaff, E., Zamenhof, S. & Green, C. Human Desoxypentose Nucleic Acid: Isolation of Desoxypentose Nucleic Acid from Human Sperm. *Nature* **165**, 756–757 (1950).
8. WILKINS, M. H. F., GOSLING, R. G. & SEEDS, W. E. Physical Studies of Nucleic Acid: Nucleic Acid: an Extensible Molecule? *Nature* **167**, 759–760 (1951).
9. FRANKLIN, R. E. & GOSLING, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **171**, 740–741 (1953).
10. Wing, R. *et al.* Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755–758 (1980).
11. Gellert, I., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.* **48**, 2013–2018 (1962).
12. Pinnavaia, T. J., Marshall, C. L., Fisk, C. L., Miles, H. T. & Becker, E. D. Alkali Metal Ion Specificity in the Solution Ordering of a Nucleotide, 5'-Guanosine Monophosphate. *J. Am. Chem. Soc.* **100**, 3625–3627 (1978).
13. Hardin, C. C., Watson, T., Corregan, M. & Bailey, C. Cation-Dependent Transition between the Quadruplex and Watson-Crick Hairpin Forms of d(CGCG3GCG). *Biochemistry* **31**, 833–841 (1992).
14. Xu, Y., Kaminaga, K. & Komiyama, M. G-Quadruplex Formation by Human Telomeric Repeats-Containing RNA in Na Solution. *J. Am. Chem. Soc.* 798–801 (2008).
15. Martadinata, H. & Phan, A. T. Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K⁺ solution. *J. Am. Chem. Soc.* **131**, 2570–2579 (2009).
16. Collie, G. W., Haider, S. M., Neidle, S. & Parkinson, G. N. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.* **38**, 5569–5580 (2010).
17. Randall, A. & Griffith, J. D. Structure of long telomeric RNA transcripts: The G-rich RNA forms a compact repeating structure containing G-quartets. *J. Biol. Chem.* **284**, 13980–13986 (2009).
18. Tang, C. F. & Shafer, R. H. Engineering the quadruplex fold: Nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. *J. Am. Chem. Soc.* **128**, 5966–5973 (2006).

19. Huang, H. *et al.* NIH Public Access. *Nat. Chem. Biol.* **10**, 686–691 (2014).
20. Xiao, C.-D., Ishizuka, T. & Xu, Y. Antiparallel RNA G-quadruplex Formed by Human Telomere RNA Containing 8-Bromoguanosine. *Sci. Rep.* **7**, 6695–6703 (2017).
21. Marušič, M., Šket, P., Bauer, L., Viglasky, V. & Plavec, J. Solution-state structure of an intramolecular G-quadruplex with propeller, diagonal and edgewise loops. *Nucleic Acids Res.* **40**, 6946–6956 (2012).
22. Marchand, A. & Gabelica, V. Native electrospray mass spectrometry of DNA G-quadruplexes in potassium solution. *J. Am. Soc. Mass Spectrom.* **25**, 1146–1154 (2014).
23. Shirude, P. S. & Balasubramanian, S. Single molecule conformational analysis of DNA G-quadruplexes. *Biochimie* **90**, 1197–1206 (2008).
24. Mergny, J. L., Li, J., Lacroix, L., Amrane, S. & Chaires, J. B. Thermal difference spectra: A specific signature for nucleic acid structures. *Nucleic Acids Res.* **33**, 1–6 (2005).
25. Karsisiotis, A. I. *et al.* Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chemie Int. Ed.* **50**, 10645–10648 (2011).
26. Tóthová, P., Krafčíková, P. & Víglaský, V. Formation of Highly Ordered Multimers in G Quadruplexes. *Biochemistry* **53**, 7013–7027 (2014).
27. Limongelli, V. *et al.* The G-triplex DNA. *Angew. Chemie Int. Ed.* **52**, 2269–2273 (2013).
28. 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. *Angew. Chemie Int. Ed.* **53**, 4107–4112 (2014).
29. Jiang, H.-X. *et al.* Divalent cations and molecular crowding buffers stabilize G-triplex at physiologically relevant temperatures. *Sci. Rep.* **5**, 9255 (2015).
30. Gehring, K., Mergny, J.-L. & Guéron, M. A tetramic DNA structure with protonated cytosine-cytosine base pairs. *Nat. Lett.* **363**, 561–565 (1993).
31. 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 Edited by I. Tinoco. *J. Mol. Biol.* **299**, 123–144 (2000).
32. Lieblein, A. L., Buck, J., Schlepckow, K., Fürtig, B. & Schwalbe, H. Time-resolved NMR spectroscopic studies of DNA i-motif folding reveal kinetic partitioning. *Angew. Chemie Int. Ed.* **51**, 250–253 (2012).
33. Weil, J. *et al.* Stabilization of the i-motif by intramolecular adenine-adenine-thymine base triple in the structure of d(ACCCT). *Acta Crystallogr. Sect. D* **55**, 422–429 (1999).
34. Rajendran, A., Nakano, S. & Sugimoto, N. Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. *Chem. Commun.* **46**, 1299–301 (2010).
35. 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* **53**, 1586–1594 (2014).

36. Zhou, J. *et al.* Formation of i-motif structure at neutral and slightly alkaline pH. *Mol. Biosyst.* **6**, 580–586 (2010).
37. Schaffitzel, C. *et al.* In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Styloynchia lemnae* macronuclei. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8572–7 (2001).
38. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* **12**, 847–854 (2005).
39. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative Visualization of DNA G-quadruplex Structures in Human Cells. *Nat. Chem.* **5**, 182–186 (2013).
40. Biffi, G., Antonio, M. Di, Tannahill, D. & Balasubramanian, S. Visualization and selective chemical targeting of RNA G- quadruplex structures in the cytoplasm of human cells. **6**, 75–80 (2014).
41. Henderson, A. *et al.* Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* **42**, 860–869 (2014).
42. Laguerre, A. *et al.* Visualization of RNA-Quadruplexes in Live Cells. *J. Am. Chem. Soc.* **137**, 8521–8525 (2015).
43. Laguerre, A., Wong, J. M. Y. & Monchaud, D. Direct visualization of both DNA and RNA quadruplexes in human cells via an uncommon spectroscopic method. *Sci. Rep.* **6**, 32141 (2016).
44. Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **33**, 2901–2907 (2005).
45. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916 (2005).
46. 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* **31**, 8112–8119 (1992).
47. The structure of telomeric DNA. *Curr. Opin. Struct. Biol.* **13**, 275–283 (2003).
48. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402–5415 (2006).
49. Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405–413 (1985).
50. Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DNA structures. *Nature* **350**, 718–720 (1991).
51. Gomez, D., Mergny, J. L. & Riou, J. F. Detection of telomerase inhibitors based on G-quadruplex ligands by a modified telomeric repeat amplification protocol assay. *Cancer Res.* **62**, 3365–3368 (2002).
52. Cuesta, J., Read, M. A. & Neidle, S. The design of G-quadruplex ligands as telomerase inhibitors. *Mini Rev. Med. Chem.* **3**, 11–21 (2003).

53. Müller, S. & Rodriguez, R. G-quadruplex interacting small molecules and drugs: from bench toward bedside. *Expert Rev. Clin. Pharmacol.* **7**, 663–679 (2014).
54. Xu, Y. *et al.* Oligonucleotide models of telomeric DNA and RNA form a hybrid G-quadruplex structure as a potential component of telomeres. *J. Biol. Chem.* **287**, 41787–41796 (2012).
55. Martadinata, H. & Phan, A. T. Formation of a stacked dimeric G-quadruplex containing bulges by the 5-terminal region of human telomerase RNA (hTERC). *Biochemistry* **53**, 1595–1600 (2014).
56. Bhartiya, D., Chawla, V., Ghosh, S., Shankar, R. & Kumar, N. Genome-wide regulatory dynamics of G-quadruplexes in human malaria parasite *Plasmodium falciparum*. *Genomics* **108**, 224–231 (2016).
57. Holder, I. T. & Hartig, J. S. A matter of location: Influence of G-quadruplexes on *Escherichia coli* gene expression. *Chem. Biol.* **21**, 1511–1521 (2014).
58. Metifiot, M., Amrane, S., Litvak, S. & Andreola, M.-L. G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Res.* **42**, 12352–12366 (2014).
59. Perrone, R. *et al.* Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 nef Coding Region: Implications for Antiviral Activity. *PLoS One* **8**, 1–14 (2013).
60. Perrone, R. *et al.* A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *J. Med. Chem.* **56**, 6521–6530 (2013).
61. Tlučková, K. *et al.* Human papillomavirus G-quadruplexes. *Biochemistry* **52**, 7207–7216 (2013).
62. Artusi, S. *et al.* Visualization of DNA G-quadruplexes in herpes simplex virus 1-infected cells. *Nucleic Acids Res.* **44**, 10343–10353 (2016).
63. Wang, S.-R. *et al.* Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. *Cell Chem. Biol.* **23**, 1113–1122 (2017).
64. Fleming, A. M., Ding, Y., Alenko, A. & Burrows, C. J. Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. *ACS Infect. Dis.* **2**, 674–681 (2016).
65. Perrone, R. *et al.* Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19. *J. Antimicrob. Chemother.* **69**, 3248–3258 (2014).
66. Perrone, R. *et al.* Synthesis, Binding and Antiviral Properties of Potent Core-Extended Naphthalene Diimides Targeting the HIV-1 Long Terminal Repeat Promoter G-Quadruplexes. *J. Med. Chem.* **58**, 9639–9652 (2015).
67. Wei, D., Parkinson, G. N., Reszka, A. P. & Neidle, S. Crystal structure of a c-kit promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. *Nucleic Acids Res.* **40**, 4691–4700 (2012).
68. Yang, D. & Hurley, L. H. Structure of the biologically relevant G-quadruplex in the c-MYC promoter. *Nucleosides. Nucleotides Nucleic Acids* **25**, 951–968 (2006).
69. Cogoi, S. & Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **34**, 2536–2549 (2006).

70. Diveshkumar, K. V. *et al.* Specific Stabilization of c-MYC and c-KIT G-Quadruplex DNA Structures by Indolylmethyleindanone Scaffolds. *Biochemistry* **55**, 3571–3585 (2016).
71. Beaudoin, J. D. & Perreault, J. P. 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* **38**, 7022–7036 (2010).
72. Bugaut, A. & Balasubramanian, S. 5'-UTR RNA G-quadruplexes: Translation regulation and targeting. *Nucleic Acids Res.* **40**, 4727–4741 (2012).
73. Beaudoin, J. D. & Perreault, J. P. Exploring mRNA 3'-UTR G-quadruplexes: Evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.* **41**, 5898–5911 (2013).
74. Rouleau, S. G., Glouzon, J.-P. S., Brumwell, A., Bisailon, M. & Perreault, J.-P. 3' UTR G-Quadruplexes regulate miRNA binding. *RNA* **23**, 1172–1179 (2017).
75. Endoh, T., Kawasaki, Y. & Sugimoto, N. Suppression of gene expression by G-quadruplexes in open reading frames depends on G-quadruplex stability. *Angew. Chemie - Int. Ed.* **52**, 5522–5526 (2013).
76. Endoh, T. & Sugimoto, N. Unusual -1 ribosomal frameshift caused by stable RNA G-quadruplex in open reading frame. *Anal. Chem.* **85**, 11435–11439 (2013).
77. Song, J., Perreault, J.-P., Topisirovic, I. & Richard, S. RNA G-quadruplexes and their potential regulatory roles in translation. *Translation* **4**, e1244031 (2016).
78. Endoh, T. & Sugimoto, N. Mechanical insights into ribosomal progression overcoming RNA G-quadruplex from periodical translation suppression in cells. *Sci. Rep.* **6**, 22719–22727 (2016).
79. Yu, Z. *et al.* ILPR G-quadruplexes formed in seconds demonstrate high mechanical stabilities. *J. Am. Chem. Soc.* **131**, 1876–1882 (2009).
80. Kendrick, S. *et al.* 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. *J. Am. Chem. Soc.* **136**, 4161–4171 (2014).
81. Sutherland, C., Cui, Y., Mao, H. & Hurley, L. H. A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1. *J. Am. Chem. Soc.* **138**, 14138–14151 (2016).
82. Phan, A. T. & Mergny, J.-L. Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res.* **30**, 4618–4625 (2002).
83. Dhakal, S. *et al.* Coexistence of an ILPR i-Motif and a partially folded structure with comparable mechanical stability revealed at the single-molecule level. *J. Am. Chem. Soc.* **132**, 8991–8997 (2010).
84. Cao, Q. *et al.* G-quadruplex DNA targeted metal complexes acting as potential anticancer drugs. *Inorg. Chem. Front.* **4**, 10–32 (2017).
85. Xiong, Y. X., Huang, Z. S. & Tan, J. H. Targeting G-quadruplex nucleic acids with heterocyclic alkaloids and their derivatives. *Eur. J. Med. Chem.* **97**, 538–551 (2015).
86. Müller, S. & Rodriguez, R. G-quadruplex interacting small molecules and drugs: from bench toward bedside. *Expert Rev. Clin. Pharmacol.* **7**, 663–679 (2014).

87. Drygin, D. *et al.* Anticancer activity of CX-3543: A direct inhibitor of rRNA biogenesis. *Cancer Res.* **69**, 7653–7661 (2009).
88. Campbell, N. H. *et al.* Molecular Basis of Structure-Activity Relationships between Salphen Metal Complexes and Human Telomeric DNA Quadruplexes. *J. Med. Chem.* **55**, 209–222 (2012).
89. Georgiades, S. N., Abd Karim, N. H., Suntharalingam, K. & Vilar, R. Interaction of metal complexes with G-quadruplex DNA. *Angew. Chemie Int. Ed.* **49**, 4020–4034 (2010).
90. Lecarme, L. *et al.* Interaction of polycationic Ni(II)-salophen complexes with G-quadruplex DNA. *Inorg. Chem.* **53**, 12519–12531 (2014).
91. Lecarme, L. *et al.* Efficient Inhibition of Telomerase by Nickel–Salophen Complexes. *ChemMedChem* **4**, 1133–1136 (2016).
92. Sun, D., Guo, K., Rusche, J. J. & Hurley, L. H. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* **33**, 6070–6080 (2005).
93. De Cian, A., Guittat, L., Shin-ya, K., Riou, J.-F. & Mergny, J.-L. Affinity and selectivity of G4 ligands measured by FRET. *Nucleic Acids Symp. Ser.* **49**, 235–236 (2005).
94. Incles, C. M. *et al.* A G-quadruplex telomere targeting agent produces p16-associated senescence and chromosomal fusions in human prostate cancer cells. *Mol. Cancer Ther.* **3**, 1201–1206 (2004).
95. Hounsou, C. *et al.* G-quadruplex recognition by quinacridines: A SAR, NMR, and biological study. *ChemMedChem* **2**, 655–666 (2007).
96. Rodriguez, R. *et al.* A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.* **130**, 15758–15759 (2008).
97. De Cian, A., DeLemos, E., Mergny, J. L., Teulade-Fichou, M. P. & Monchard, D. Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J. Am. Chem. Soc.* **129**, 1856–1857 (2007).
98. Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M. P. & Phan, A. T. Solution structure of a G-quadruplex bound to the bisquinolinium compound phen-DC3. *Angew. Chemie - Int. Ed.* **53**, 999–1002 (2014).
99. Fedoroff, O. Y., 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* **39**, 15083–15090 (2000).
100. Li, X., Peng, Y., Ren, J. & Qu, X. Carboxyl-modified single-walled carbon nanotubes selectively induce human telomeric i-motif formation. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19658–19663 (2006).
101. Zhao, C., Ren, J. & Qu, X. Single-walled carbon nanotubes binding to human telomeric i-motif DNA under molecular-crowding conditions: More water molecules released. *Chem. - A Eur. J.* **14**, 5435–5439 (2008).

102. Chen, Y. *et al.* Insights into the biomedical effects of carboxylated single-wall carbon nanotubes on telomerase and telomeres. *Nat. Commun.* **3**, 1074–1087 (2012).
103. Cui, Y. *et al.* Molecular population dynamics of DNA structures in a bcl-2 promoter sequence is regulated by small molecules and the transcription factor hnRNP LL. *Nucleic Acids Res.* **42**, 5755–5764 (2014).
104. Wright, E. P. *et al.* Mitoxantrone and Analogues Bind and Stabilize i-Motif Forming DNA Sequences. *Sci. Rep.* **6**, 39456–39463 (2016).
105. Noer, S. L. *et al.* Folding dynamics and conformational heterogeneity of human telomeric G-quadruplex structures in Na⁺ solutions by single molecule FRET microscopy. *Nucleic Acids Res.* **44**, 464–471 (2016).
106. 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* **39**, 297–308 (1996).
107. Galibert, M. *et al.* Application of click-click chemistry to the synthesis of new multivalent RGD conjugates. *Org. Biomol. Chem.* **8**, 5133–5138 (2010).
108. Pifferi, C., Berthet, N. & Renaudet, O. Cyclopeptide scaffolds in carbohydrate-based synthetic vaccines. *Biomater. Sci.* **5**, 953–965 (2017).
109. Jencks, W. P. Studies on the Mechanism of Oxime and Semicarbazone Formation¹. *J. Am. Chem. Soc.* **81**, 475–481 (1959).
110. Huisgen, R. 1,3-Dipolar Cycloadditions. Past and Future. *Angew. Chem. Int. Ed.* **2**, 565–598 (1963).
111. Singh, Y., Murat, P. & Defrancq, E. Recent developments in oligonucleotide conjugation. *Chem. Soc. Rev.* **39**, 2054 (2010).
112. Schoch, J., Wiessler, M. & Ja, A. Post-Synthetic Modification of DNA by Inverse-Electron-Demand Diels - Alder Reaction. 8846–8847 (2010)
113. Schoch, J., Ameta, S. & Jäschke, A. Inverse electron-demand Diels–Alder reactions for the selective and efficient labeling of RNA. *Chem. Commun.* **47**, 12536 (2011).
114. Defrancq, E. & Messaoudi, S. Palladium-Mediated Labeling of Nucleic Acids. *ChemBioChem* **18**, 426–431 (2017).
115. Viault, G. *et al.* The first ‘ready-to-use’ benzene-based heterotrifunctional cross-linker for multiple bioconjugation. *Org. Biomol. Chem.* **11**, 2693 (2013).
116. Joshi, T., Patra, M., Spiccia, L. & Gasser, G. Di-heterometalation of thiol-functionalized peptide nucleic acids. *Artif. DNA. PNA XNA* **4**, 11–18 (2013).
117. Ligeour, C. *et al.* Synthesis of galactoclusters by metal-free thiol ‘click chemistry’ and their binding affinities for pseudomonas aeruginosa lectin Ica. *European J. Org. Chem.* **2014**, 7621–7630 (2014).
118. Murat, P. *et al.* A novel conformationally constrained parallel G quadruplex. *Chembiochem* **9**, 2588–2591 (2008).

119. Bonnet, R., Lavergne, T., Gennaro, B., Spinelli, N. & Defrancq, E. Construction of anti-parallel G-quadruplexes through sequential templated click. *Chem. Commun.* **51**, 4850–4853 (2015).
120. Murat, P. *et al.* Template-Assembled Synthetic G-Quadruplex (TASQ): A Useful System for Investigating the Interactions of Ligands with Constrained Quadruplex Topologies. *Chem. - A Eur. J.* **16**, 6106–6114 (2010).
121. Merrifield, R. B. Solid Phase Peptide Synthesis. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).
122. Matteucci, M. D. & Caruthers, M. H. Synthesis of Deoxyoligonucleotides on a Polymer Support. *J. Am. Chem. Soc.* **103**, 3185–3191 (1981).
123. 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. *Chem. - A Eur. J.* **14**, 9135–9138 (2008).
124. Azzalin, C. M., Reichenbach, P., Khoriant, L., Giulotto, E. & Lingner, J. Telomeric Repeat Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science (80)*. **318**, 798–801 (2007).
125. Schoeftner, S. & Blasco, M. A. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell Biol.* **10**, 228–236 (2008).
126. Redon, S., Reichenbach, P. & Lingner, J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* **38**, 5797–5806 (2010).
127. Redon, S., Zemp, I. & Lingner, J. A three-state model for the regulation of telomerase by TERRA and hnRNPA1. *Nucleic Acids Res.* **41**, 9117–9128 (2013).
128. Cusanelli, E., Romero, C. & Chartrand, P. Telomeric Noncoding RNA TERRA Is Induced by Telomere Shortening to Nucleate Telomerase Molecules at Short Telomeres. *Mol. Cell* **51**, 780–791 (2013).
129. Hirashima, K. & Seimiya, H. Telomeric repeat-containing RNA/G-quadruplex-forming sequences cause genome-wide alteration of gene expression in human cancer cells in vivo. *Nucleic Acids Res.* **43**, 2022–2032 (2015).
130. Rondón, A. G., Jimeno, S. & Aguilera, A. The interface between transcription and mRNP export: From THO to THSC/TREX-2. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1799**, 533–538 (2010).
131. Pfeiffer, V., Crittin, J., Grolimund, L. & Lingner, J. The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *EMBO J.* **32**, 2861–2871 (2013).
132. Xu, Y. *et al.* Finding a human telomere DNA–RNA hybrid G-quadruplex formed by human telomeric 6-mer RNA and 16-mer DNA using click chemistry: A protective structure for telomere end. *Bioorg. Med. Chem.* **22**, 4419–4421 (2014).
133. Zhang, J., Zheng, K., Xiao, S., Hao, Y. & Tan, Z. Mechanism and manipulation of DNA : RNA hybrid G- quadruplex formation in transcription of G-rich DNA. *J. Am. Chem. Soc.* **136**, 1381–1390 (2014).

134. Zhao, Y. *et al.* Real-Time Detection Reveals Responsive Cotranscriptional Formation of Persistent Intramolecular DNA and Intermolecular DNA:RNA Hybrid G-Quadruplexes Stabilized by R-Loop. *Anal. Chem.* **89**, 6036–6042 (2017).
135. Wu, R. Y., Zheng, K. W., Zhang, J. Y., Hao, Y. H. & Tan, Z. Formation of DNA: RNA hybrid G-quadruplex in bacterial cells and its dominance over the intramolecular DNA G-quadruplex in mediating transcription termination. *Angew. Chemie - Int. Ed.* **54**, 2447–2451 (2015).
136. Edupuganti, O. P., Singh, Y., Defrancq, E. & Dumy, P. New strategy for the synthesis of 3',5'-bifunctionalized oligonucleotide conjugates through sequential formation of chemoselective oxime bonds. *Chem. - A Eur. J.* **10**, 5988–5995 (2004).
137. Laurent, A., de Lambert, B., Charreyre, M.-T., Mandrand, B. & Chaix, C. A one step derivatization of controlled pore glass for oligonucleotide solid-phase synthesis. *Tetrahedron Lett.* **45**, 8883–8887 (2004).
138. Hardouin, J., Cremer, A.-G. & Delmas, A.-F. Investigation of in-source decay of oxime-linked peptide by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **25**, 2106–2112 (2011).
139. Goddard-Borger, E. D. & Stick, R. V. An efficient, inexpensive, and shelf-stable diazotransfer reagent: Imidazole-1-sulfonyl azide hydrochloride. *Org. Lett.* **9**, 3797–3800 (2007).
140. Sigman, D. S., Graham, D. R., Aurora, V. D. & Stern, A. M. Oxygen-dependent Cleavage of DNA by the 1,10-Phenanthroline.Cuprous Complex. *J. Biol. Chem.* **254**, 12269–12272 (1979).
141. Bonnat, L. *et al.* Templated Formation of Discrete RNA and DNA:RNA Hybrid G-Quadruplexes and Their Interactions with Targeting Ligands. *Chem. - A Eur. J.* **22**, 3139–3147 (2016).
142. Piekna-przybylska, D., Sullivan, M. A., Sharma, G. & Bambara, R. A. U3 Region in the HIV 1 Genome Adopts a G Quadruplex Structure in Its RNA and DNA Sequence. (2015).
143. Amrane, S. *et al.* Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: A potential target for anti-HIV drug development. *J. Am. Chem. Soc.* **136**, 5249–5252 (2014).
144. Špringer, T., Šípová, H., Vaisocherová, H., Štěpánek, J. & Homola, J. Shielding effect of monovalent and divalent cations on solid-phase DNA hybridization: Surface plasmon resonance biosensor study. *Nucleic Acids Res.* **38**, 7343–7351 (2010).
145. Bonnat, L. *et al.* Template-Mediated Stabilization of a DNA G-Quadruplex formed in the HIV-1 Promoter and Comparative Binding Studies. *Chem. - A Eur. J.* **23**, 5602–5613 (2017).
146. Cerofolini, L. *et al.* G-triplex structure and formation propensity. *Nucleic Acids Res.* **42**, 13393–13404 (2014).
147. Rajendran, A. *et al.* Small molecule binding to a G-hairpin and a G-triplex: a new insight into anticancer drug design targeting G-rich regions. *Chem. Commun. (Camb)*. **51**, 9181–4 (2015).
148. Amato, J. *et al.* Discovery of the first dual G-triplex/G-quadruplex stabilizing compound: a new opportunity in the targeting of G-rich DNA structures? *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 1271–1280 (2017).
149. Foulard, S., Rasmussen, M. O., Razkin, J., Boturyn, D. & Dumy, P. 1-Ethoxyethylidene, a new group for the stepwise SPPS of aminooxyacetic acid containing peptides. *J. Org. Chem.* **73**, 983–991 (2008).

150. Dickerhoff, J. & Weisz, K. Flipping a G-tetrad in a unimolecular quadruplex without affecting its global fold. *Angew. Chemie - Int. Ed.* **54**, 5588–5591 (2015).
151. Gajarský, M. *et al.* Structure of a Stable G-Hairpin. *J. Am. Chem. Soc.* **139**, 3591–3594 (2017).
152. Bonnet, R., Murat, P., Spinelli, N. & Defrancq, E. Click–click chemistry on a peptidic scaffold for easy access to tetrameric DNA structures. *Chem. Commun.* **48**, 5992 (2012).
153. Lartia, R., Bonnat, L., Murat, P. & Defrancq, E. Preparation of azido containing oligonucleotides through diazo transfer reaction. *Curr. Protoc. Nucleic Acid Chem.* 1–7 (2012).
154. Pockrand, I. & Raether, H. Surface plasma oscillations at sinusoidal silver surfaces. *Appl. Opt.* **16**, 1784–1786 (1977).
155. Collie, G. *et al.* Selectivity in small molecule binding to human telomeric RNA and DNA quadruplexes. *Chem. Commun.* 7482–7484 (2009).
156. Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotech* **33**, 877–881 (2015).
157. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Meth* **13**, 841–844 (2016).
158. Marton, S., Cleto, F., Krieger, M. A. & Cardoso, J. Isolation of an aptamer that binds specifically to *E. coli*. *PLoS One* **11**, 1–17 (2016).
159. González, V. M., Elena Martín, M., Fernández, G. & García-Sacristán, A. Use of aptamers as diagnostics tools and antiviral agents for human viruses. *Pharmaceuticals* **9**, 1–34 (2016).
160. Tran, D. T. *et al.* Selection of aptamers against Ara h 1 protein for FO-SPR biosensing of peanut allergens in food matrices. *Biosens. Bioelectron.* **43**, 245–251 (2013).
161. Pfeiffer, F. & Mayer, G. Selection and Biosensor Application of Aptamers for Small Molecules. *Front. Chem.* **4**, 1–21 (2016).
162. Watrin, M., Von Pelchrzim, F., Dausse, E., Schroeder, R. & Toulmé, J. J. In vitro selection of RNA aptamers derived from a genomic human library against the TAR RNA element of HIV-1. *Biochemistry* **48**, 6278–6284 (2009).
163. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
164. Mendonsa, S. D. & Bowser, M. T. In Vitro Evolution of Functional DNA Using Capillary Electrophoresis. *J. Am. Chem. Soc.* **126**, 20–21 (2004).
165. Mosing, R. K., Mendonsa, S. D. & Bowser, M. T. Capillary electrophoresis-SELEX selection of aptamers with affinity for HIV-1 reverse transcriptase. *Anal. Chem.* **77**, 6107–6112 (2005).
166. Mendonsa, S. D. & Bowser, M. T. In vitro selection of aptamers with affinity for neuropeptide Y using capillary electrophoresis. *J. Am. Chem. Soc.* **127**, 9382–9383 (2005).
167. Yang, J. & Bowser, M. T. Capillary electrophoresis-SELEX selection of catalytic DNA aptamers for a small-molecule porphyrin target. *Anal. Chem.* **85**, 1525–1530 (2013).

168. Ochsner, U. A., Green, L. S., Gold, L. & Janjic, N. Systematic selection of modified aptamer pairs for diagnostic sandwich assays. *Biotechniques* **56**, 125–133 (2014).
169. Gawande, B. N. *et al.* Selection of DNA aptamers with two modified bases. *Proc. Natl. Acad. Sci.* **114**, 2898–2903 (2017).
170. Rohloff, J. C. *et al.* Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Mol. Ther. - Nucleic Acids* **3**, e201 (2014).
171. Vaught, J. D. *et al.* Expanding the chemistry of DNA for in vitro selection. *J. Am. Chem. Soc.* **132**, 4141–4151 (2010).
172. Bailey, T. L. *et al.* MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, 202–208 (2009).
173. Park, Y. M., Squizzato, S., Buso, N., Gur, T. & Lopez, R. The EBI search engine: EBI search as a service—making biological data accessible for all. *Nucleic Acids Res.* **45**, W545–W549 (2017).

Annexes

G-Quadruplexes

Templated Formation of Discrete RNA and DNA:RNA Hybrid G-Quadruplexes and Their Interactions with Targeting Ligands

Laureen Bonnat,^[a, b] Jérôme Dejeu,^[a, b] Hugues Bonnet,^[a, b] Béatrice Génaro,^[a, b]
Olivier Jarjays,^[a, b] Fabrice Thomas,^[a, b] Thomas Lavergne,^{*[a, b]} and Eric Defrancq^{*[a, b]}

Abstract: G-rich RNA and DNA oligonucleotides derived from the human telomeric sequence were assembled onto addressable cyclopeptide platforms through oxime ligations and copper-catalyzed azide-alkyne cycloaddition (CuAAC) reactions. The resulting conjugates were able to fold into highly stable RNA and DNA:RNA hybrid G-quadruplex (G4) architectures as demonstrated by UV, circular dichroism (CD), and NMR spectroscopic analysis. Whereas rationally designed parallel RNA and DNA:RNA hybrid G4 topologies could be obtained, we could not force the formation of an antiparallel

RNA G4 structure, thus supporting the idea that this topology is strongly disfavored. The binding affinities of four representative G4 ligands toward the discrete RNA and DNA:RNA hybrid G4 topologies were compared to the one obtained with the corresponding DNA G4 structure. Surface plasmon resonance (SPR) binding analysis suggests that the accessibility to G4 recognition elements is different among the three structures and supports the idea that G4 ligands might be shaped to achieve structure selectivity in a biological context.

Introduction

The double-helical structure of DNA within which two antiparallel strands are held together through canonical A/T and G/C base pairing was established by the seminal contributions of Watson, Crick, Wilkins, and Franklin over half a century ago. Beyond double-helical-based structures such as B-DNA, A-DNA, and Z-DNA, the past decades have brought accumulating evidence of the existence and biological relevance of four-stranded nucleic acid motifs called quadruplexes. In particular, DNA G-quadruplexes (G4-DNA) formed by G-rich DNA sequences have been the subject of intensive investigations during the past two decades.

G4 formation does not involve canonical Watson–Crick base pairing but relies on the stacking of Hoogsteen hydrogen-bonded guanine tetrads.^[1] Sequencing and bioinformatics analysis of the human genome have identified hundreds of thousands of sequences with the potential to form stable G4 structures.^[2] Interestingly, a significant statistical enrichment of G4-forming sequences within key regulatory regions has been uncovered. DNA G4-forming sequences can be found at telo-

meres, which are responsible for eukaryote chromosome stability and replication capability, and are over-represented in the promoter region of numerous genes, including proto-oncogenes c-MYC, c-KIT, and KRAS. The propensity of the G4 structure to be stabilized by physiologically abundant monovalent cations, such as K⁺ and Na⁺, along with reports suggesting that DNA G4-forming sequences are evolutionary conserved among species, contributed significantly to increased attention toward this particular nucleic acid architecture.

Besides DNA-G4, G-rich RNA sequences are also prone to fold into stable G4 architectures (RNA-G4). Among other locations, RNA-G4-forming sequences can be found in numbers within the 5'- and 3'-untranslated regions and in the open reading frame of mRNA.^[3] To date, RNA G4 has been suspected to be involved in several phenomena such as translation regulation, pre-mRNA processing, and mRNA targeting.^[4] Owing to the single-stranded nature of transcribed RNA, in vivo formation of G4-RNA is expected to occur more easily than G4-DNA.

Strong arguments have been provided for the formation of both DNA and RNA G4 structures within cells with the use of G4-specific antibodies^[5] and binding-activated fluorescent G4-targeting ligands.^[6] Several G4-binding regulatory proteins have been identified, and G4 formation is now suspected to be involved in numerous pathogenic processes including degenerative disorders, oncogene regulation, and viral infections.^[7]

More recently, intermolecular hybrid DNA:RNA quadruplex structures (HQ) have attracted increasing interest owing to the fact that HQ motifs, which only require two stretches of guanine repeats from both DNA and RNA strands to fold, could be more abundant than either DNA and RNA-G4.^[8] It was reported that HQ formation could modulate transcription in vitro^[9] and

[a] L. Bonnat, Dr. J. Dejeu, H. Bonnet, B. Génaro, Dr. O. Jarjays, Prof. F. Thomas, Dr. T. Lavergne, Prof. E. Defrancq
Université Grenoble Alpes, Département de Chimie Moléculaire, UMR-5250, 38000 Grenoble (France)
E-mail: thomas.lavergne@ujf-grenoble.fr
eric.defrancq@ujf-grenoble.fr

[b] L. Bonnat, Dr. J. Dejeu, H. Bonnet, B. Génaro, Dr. O. Jarjays, Prof. F. Thomas, Dr. T. Lavergne, Prof. E. Defrancq
CNRS, 38000 Grenoble (France)

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/chem.201504351>.

in vivo,^[10] and their prevalence was recently evidenced.^[11] The spontaneous formation, within a model cell system, of HQ motifs between biologically relevant G-rich mRNA sequences and a rationally designed extraneous synthetic G-rich oligonucleotide was exploited as an elegant approach to repress gene translation and decrease the proliferation of cancer cells.^[12]

The biological function associated with G4 nucleic acids has spurred a large number of studies directed toward the identification of G4-specific binding probes and drugs.^[13] Hundreds of distinct small-molecule ligands with high affinities toward DNA and RNA G4 motifs have been reported, and many have demonstrated the ability to affect cellular processes.^[14] For example, the association of exogenous ligands with the DNA G4 formed within the human telomere was found to increase its thermodynamic stability, thus affecting the activity of telomerase, an enzyme of clinical interest for aging and cancer.^[15]

However, most of the compounds assembled to date exhibit insufficient specificity for a particular G4 topology, which precludes the meticulous examination of distinct G4 biological functions and restricts their therapeutic potential. Encouragingly, binding selectivity might eventually be achieved owing to the fact that G4 displays remarkable structural heterogeneity. Indeed, a G4 folded structure can vary in the number of stacked G-tetrads; the composition, length, and polarity of G-tract connecting sequences (e.g., loops); and the width of the resulting four grooves.^[1]

Yet the identification of structure-selective ligands is complicated by another essential feature of G4 nucleic structure, namely, the intrinsic polymorphic nature of synthetic G4 targets isolated in solution.^[16] Indeed, numerous in vitro studies have shown that small variations in their experimental conditions—such as the nature and concentration of metal cations, the solvent viscosity, the molecular concentration, and the folding procedure—could trigger major structural reorganizations and lead to intricate structural mixtures in solution.^[17] As

we begin to grasp the rationales for the structural variation observed in vitro, those observations cannot fully be transposed to the complex intracellular milieu, and the topology adopted by G4 in vivo remains largely unknown.

In this comprehensive context, the elaboration of molecular systems able to mimic a broad range of well-defined G4 topology would be of significant interest. Such topologically controlled and stabilized G4 architectures would find application in vitro with the development of structure-specific G4 binding ligands and in vivo with the identification of structure-specific G4 binding partners (e.g., proteins).

Over the past few years, we and others^[18] have demonstrated that the thermodynamic stability of distinct G4 folds could be increased by site-specifically fixing DNA-G4-forming sequences onto rigid molecular scaffolds. By using regioselectively addressable cyclopeptide platforms, a full parallel tetrameric (Figure 1B)^[19] along with an antiparallel dimeric DNA-G4 topology were assembled.^[20] As expected, the resulting folded topologies displayed high stability and reduced polymorphism and could find applications in the exhaustive evaluation of ligand binding properties.^[21] The G-rich sequences that were connected to molecular templates were derived from the human telomeric repeat sequence $d(5'AG_3T_2AG_3^3)$ owing to the prevalent role of this particular DNA sequence in chromosome regulation and cancer-cell processes. Truly selective ligands for G4 folds present in the telomeric region would be of particular interest, in line with the complexity and abundance of G4-forming sequences in this region of the genome. No fewer than six distinct G4 architectures have been proposed on the basis of high-resolution structural investigations of the telomeric DNA G4 primary sequence.^[22] Telomeres are also transcribed into non-coding G-rich RNA repeats (TERRA) that have been shown to fold into stable RNA-G4. The TERRA transcript has been suggested to play a role in the interactions with proteins at chromosome ends^[23] and was found to alter gene expression in te-

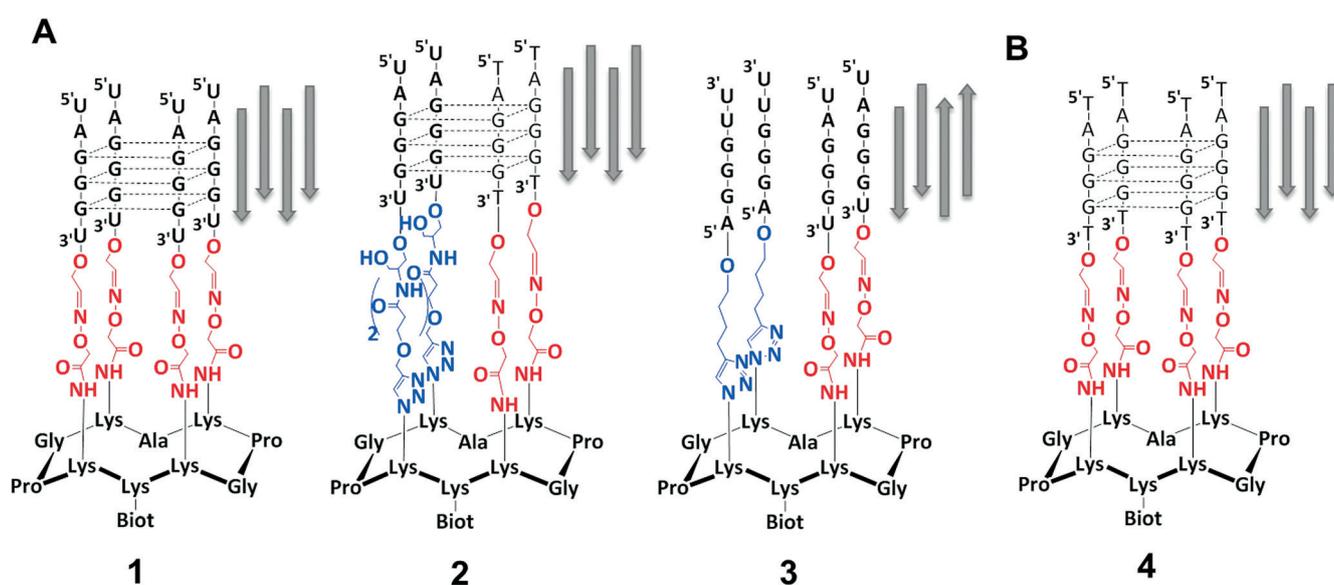


Figure 1. A) RNA and DNA:RNA conjugates 1–3 assembled in this study. B) Previously reported DNA conjugate 4. Ribonucleotides are bolded.

lomere-elongated cancer cells.^[24] Finally, it was suggested that hybrid DNA:RNA G4 structures could also arise from the intermolecular association of telomeric DNA and RNA G-rich sequences adding to the significance of TERRA transcript for telomere end protection against degradation and chromosome end fusions.^[25]

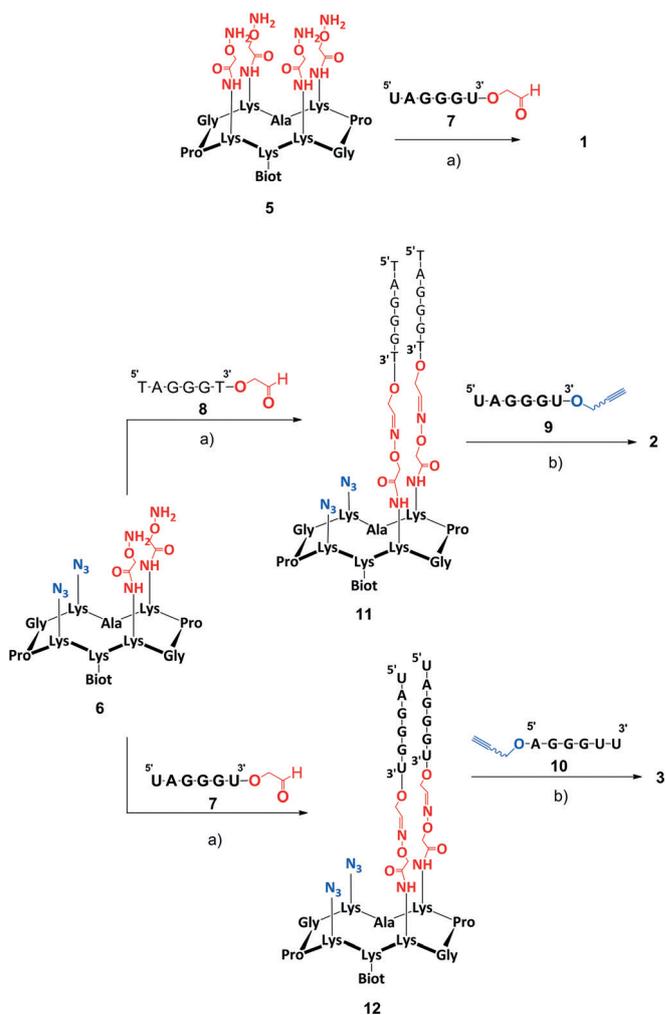
In this paper, we report the synthesis, characterization, and targeting of discrete RNA and DNA:RNA hybrid G4 structures derived from the corresponding biologically relevant telomeric G4 folds (Figure 1A). Three molecular systems were devised. RNA conjugate system **1** and DNA:RNA conjugate system **2** were assembled with the expectation that they would fold into stable parallel G4 topologies analogous to those reported for the parental RNA and DNA:RNA telomeric G4. On the other hand, RNA strands of opposite polarities were fixed within the RNA conjugate system **3** to probe for the propensity of G-rich RNA sequences to fold into an antiparallel G4 architecture when subjected to physical constraints. Unlike with DNA G4, the structural repertoire of RNA G4 has been shown to be rather limited, as in all (but one)^[26] reports the G4-forming RNA strands exclusively appears to run in parallel orientations. To assemble conjugate **1**, four oxime ligation reactions were simultaneously performed between an aminoxy-containing cyclopeptide platform and RNA strands that bear aldehyde functions. To achieve the controlled association of DNA and RNA strands within conjugates **2** and **3**, sequential oxime ligation and copper-catalyzed azide-alkyne cycloaddition (CuAAC) were performed between an aminoxy- and azide-containing cyclopeptide and DNA and/or RNA strands that bore either aldehyde or alkyne functions. Both oxime and CuAAC reactions were achieved with excellent yields, and the resulting conjugates **1–3** were obtained with high purity. As expected, both conjugates **1** and **2** were able to fold into highly stable parallel G4 topologies as validated by UV thermal difference spectra (TDS), circular dichroism (CD), and NMR spectroscopy, thereby expanding our collection of telomeric G4 mimics to RNA-based structures. Strikingly, but in coherence with previous reports,^[27] the G-rich RNA strands within conjugate **3** refused to adopt the forced antiparallel orientation as our experimental data instead support the formation of a peculiar mixture of parallel G4 architectures. The ability of four ligands, representative of well-known classes of G4 binders, to associate with RNA and DNA:RNA G4 **1** and **2** was assessed using surface plasmon resonance (SPR). When compared with the previously described and analogous DNA G4 **4**, the reported binding affinities and specificities support the idea that selective recognition of related but distinct DNA, RNA, and DNA:RNA G4 can indeed be achieved.

Results and Discussion

Synthesis of conjugates

The molecular scaffolds used to site-specifically connect the nucleic acid strands consist of a cyclic decapeptide with two independent and addressable faces. The “upper” domain displays diversely functionalized lysine side chains to anchor the

oligonucleotide strands while the “lower” domain is functionalized by a biotin for surface immobilization (SPR studies). The unique structural features of the template contribute to prevent undesired steric hindrances between the presented elements.^[28] To assemble RNA system **1**, four oligoribonucleotides derived from the TERRA telomeric sequence $r(5'UAGGGUX^3')$ **7** (in which X represents a 3'-aldehyde linker) were attached through oxime bond formation onto cyclodecapeptide **5** (Scheme 1). To assemble HQ system **2**, two oligodeoxyribonu-



Scheme 1. Synthesis of conjugates **1–3**. a) NH_4OAc 0.4 M, pH 4.5, 55 °C, 45 min; b) HEPES 100 mM, CuSO_4 (3 equiv), THPTA (15 equiv), sodium ascorbate (15 equiv), 37 °C, 2 h. Ribonucleotides are bold.

cleotides derived from the telomeric DNA sequence $(5'TAGGGTX^3')$ **8** and two oligoribonucleotides of sequence $r(5'UAGGGU^3')$ **9** (in which Y represents a 3'-alkynyl linker) were successively connected to cyclodecapeptide **6** by means of stepwise oxime ligations and CuAAC reactions. To prepare RNA system **3**, two oligoribonucleotides **7** and two oligoribonucleotides of sequence $r(5'ZAGGGU^3')$ **10** (in which Z represents a 5'-hexynyl linker) were successively connected to cyclodeptide **6**. To assemble scaffolds **5** and **6**, peptide precursors were assembled using standard 9-fluorenylmethoxycarbonyl (Fmoc)/

*t*Bu solid-phase peptide synthesis (SPPS) on an acid-labile 2-chlorotrityl resin and cyclized in solution according to reported procedures (see the Supporting Information).^[29] The aminoxy groups were introduced on lysine side chains by the addition of *N*-Boc-aminoxyacetic acid (Boc = *tert*-butyloxycarbonyl) *N*-hydroxysuccinimide (NHS) ester and subsequent acidic deprotection. Azido moieties on scaffold **6** were introduced during SPPS using the *L*-azidonorleucine building block. To assemble oligonucleotides *r*(⁵UAGGGUX³) **7** and *d*(⁵TAGGGTX³) **8**, solid-phase synthesis using phosphoramidite chemistry was performed on a glyceryl solid support. Oligoribonucleotides were prepared using pivaloyloxymethyl (PivOM) groups as 2'-OH protection.^[30] Oligonucleotides **7** and **8** were obtained after oxidative cleavage of the 3'-diol linker generated upon release from the solid support. No side reaction was observed during the NaIO₄ oxidative cleavage of RNA-containing 3'-diol. Alkyne-containing oligoribonucleotides *r*(⁵UAGGGUY³) **9** and *r*(⁵ZUAGGGU³) **10** were prepared using a 3'-alkyne modifier serinol resin and 5'-hexynyl phosphoramidite, respectively. The synthesis of conjugate **1** was carried out in ammonium acetate buffer (pH 4.5) at 55 °C by using a slight excess amount of aldehyde-containing oligoribonucleotide **7**. Reaction monitoring by reverse-phase HPLC (RP-HPLC) revealed a complex mixture with broadening signals that we attributed to the coexistence of folded and unfolded RNA G4 structures within **1** along with higher aggregates that involve **7** (Figure S11 in the Supporting Information). To obtain a clearer picture of the actual molecular components of the crude reaction mixture, and to enable the efficient isolation of conjugate **1**, we investigated the use of anion-exchange liquid chromatography (AELC) under denaturing conditions. We found that the use of a chaotropic salt gradient (lithium perchlorate) at high temperature (75 °C) significantly improved the resolution between peaks. Under those stringent conditions, a major peak that corresponded to conjugate **1** was observed along with a minor peak that corresponded to an intermediate conjugate that bore three RNA strands (Figure S12 in the Supporting Information). Purification using a preparative AELC system followed by thorough desalting using size-exclusion chromatography (SEC) and a C18 cartridge afforded pure conjugate **1** with an isolated yield of 48%. MALDI-TOF MS analysis showed excellent agreement between the experimentally determined molecular weight and the calculated value (Supporting Information).

Toward conjugates **2** and **3**, oxime ligation reactions were first performed between peptide scaffold **6** and aldehyde-functionalized oligonucleotides **8** and **7**, respectively. The formation of the resulting intermediates was confirmed by mass spectrometry (ESI), and AELC-purified compounds **11** and **12** were subjected to CuAAC reactions with alkyne-functionalized oligonucleotides **9** and **10**, respectively. Under standard CuAAC conditions, which include CuSO₄, tris(benzyltriazolylmethyl)amine (THPTA) ligand, and sodium ascorbate in 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer, the conversion of **11** and **12** to **2** and **3**, respectively, proceeded to near completion within two hours at 37 °C (see Figures S18 and S24 in the Supporting Information). Again, major peaks that corresponded to the desired conjugates were observed along with minor peaks that corresponded to intermediate conjugates that bore only three RNA and/or RNA and DNA strands. Purified and desalted conjugates **2** and **3** were isolated and characterized by means of mass spectrometry (MALDI).

Characterization of G-quadruplex topologies

To gain insight into the topology adopted by the RNA and DNA strands within conjugates **1–3**, we first performed the analysis of their thermal absorbance difference spectra (TDS).^[31] Thermal difference spectra were collected in the presence of Na⁺ or K⁺ or in the absence of added cations. In all cases, TDS profiles were similar and characteristic of G4 architectures with two positive peaks at around 243 and 273 nm and a negative peak at around 295 nm (see the Supporting Information).

To further confirm the formation of G4 structures and to determine the resulting topologies, CD analyses were performed.^[32] The CD spectra of **1**, **2**, and **3** were recorded in the presence of Na⁺ or K⁺ or in the absence of added cations (Figure 2). In all cases, the CD spectra display peaks that are characteristic of a parallel G4 topology with a positive peak at 263 nm and a negative peak around 240 nm. Although this particular CD signature was expected for conjugates **1** and **2**, it supports the idea that the RNA strands within conjugate **3** refused to adopt the forced antiparallel orientation and managed to reorganize themselves to fold into a parallel G4 topology. We also observed a diminished CD signal for conjugate **3**. A possible interpretation would be that the flexibility required

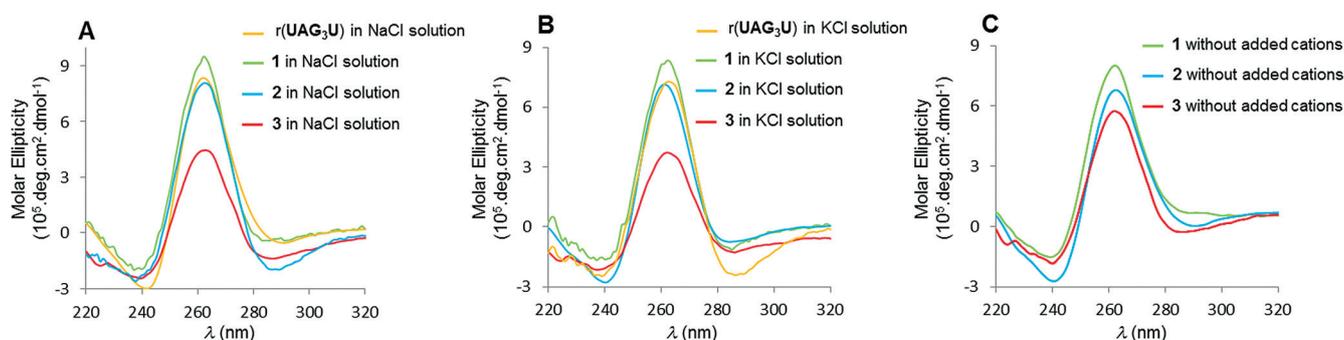


Figure 2. Normalized CD spectra of G-quadruplex-forming sequence *r*(UAG₃U) (6.4 μM) and conjugates **1–3** (1.6 μM) in 10 mM Tris-HCl, pH 7.4, 20 °C in the presence of either A) 10 mM NaCl, B) 100 μM KCl, or C) without added cations.

to reorient the strand polarities might come at the cost of one G-tetrad as some guanines might be involved in the formation of hairpin-like structures. Regardless of the exact structure, and in line with previous unsuccessful attempts to govern the formation of antiparallel RNA G4,^[27] this unexpected major structural reorganization in the context of constrained template-assembled nucleic acids further emphasizes that the formation of antiparallel RNA G4 is strongly disfavored.

Preliminary ¹H NMR spectroscopic studies were performed to further confirm the formation of G4 structures in the RNA and RNA:DNA systems 1–3. ¹H NMR spectroscopic analysis of 1–3 in sodium-based aqueous solutions showed the presence of guanine imino protons within the range of $\delta = 10$ –12 ppm that are characteristic of G4 folds (Figure 3). As expected, only three distinct proton signals were obtained for the symmetrical RNA G4 structure 1. As anticipated from the CD analysis, the NMR spectra of conjugate 3 show broad signals that suggest the formation of intricate topologies, the identification of which is beyond the scope of this paper.

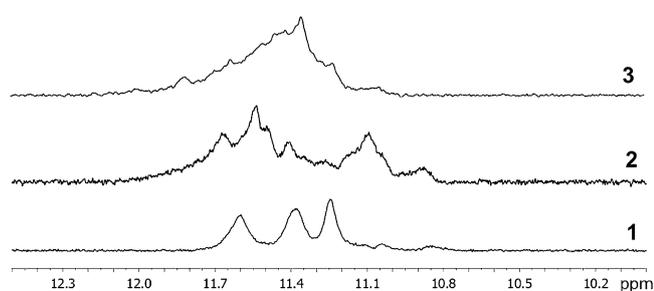


Figure 3. The imino proton region of 1D ¹H NMR spectra of 1–3 in 10 mM sodium phosphate buffer, pH 7.4 in the presence of 100 mM NaCl.

The thermal stabilities of constrained systems 1–3 were obtained from CD denaturation studies. The melting temperature (T_m) of the G4 structures was collected for each conjugate in the presence of either 10 mM NaCl, 100 μ M KCl, or without added cations. The sodium and potassium salt concentrations were selected to allow for an accurate comparison of the thermal stability of the architectures of 1–3 with the one obtained by the intermolecular association of four $r^5(\text{UAG}_3\text{U}^3)$ strands and the one obtained for previously reported DNA conjugate 4.^[19] As shown in Table 1, conjugate 1 was found to fold into a G4 structure that is more stable than those that involve DNA strands in either sodium- or potassium-based solutions. This observation was expected as more favorable enthalpy contributions are associated with the formation of G4 structures from RNA strands than from DNA strands. The significant difference between the thermal stability of RNA G4 1 and its unconstrained counterpart [$4 \times r^5(\text{UAG}_3\text{U}^3)$] (ΔT_m of 20 and 12 °C in NaCl and KCl solution, respectively) are in agreement with previously reported observations that show the significant gain in stability provided by fixation on the molecular template.^[19] Strikingly, we observed a melting transition well above room temperature (45 °C) for the RNA G4 1 folded in the absence of stabilizing cations.

Table 1. Melting temperatures values from CD denaturation studies.					
Conditions ^[a]	r(UAG ₃ U)	RNA system 1	DNA:RNA system 2	RNA system 3	DNA system 4
NaCl 10 mM	45	66	51	44	51
KCl 100 μ M	46	58	47	43	48
Without added cations	n.d.	45	34	27	32

[a] G4 structures were first annealed by heating at 80 °C for 5 min in Tris-HCl buffer (10 mM, pH 7.4) with 10 mM NaCl or 100 μ M KCl or without added cations and cooled over 2 h to room temperature. Samples 1–4 were prepared at 1.6 μ M (see the Supporting Information); r(UAG₃U) is at 6.4 μ M.

We found out that, under any conditions investigated, the hybrid DNA:RNA G4 structure within system 2 was as stable as the DNA G4 structure 4, which supports the idea that this particular molecular architecture is likely to arise when DNA and RNA G-rich sequences co-localize within cellular compartments. To rule out the formation of intermolecular RNA G4 structures that would involve the RNA strands of two distinct conjugate 2, we confirmed that the melting temperature was independent of the concentration (see Figures S34 and S35 in the Supporting information). The stabilized intramolecular DNA:RNA G4 structure presented within system 2 is, to the best of our knowledge, the first relevant model system for the investigation of the binding capacities of ligands toward this peculiar nucleic acid structure of biological interest.

In agreement with the folding interpretation provided above, the parallel G4 structure formed within 3 was found to be significantly less stable than the one formed within conjugate 1, with ΔT_m values ranging from 15 to 20 °C. Although we confirmed that the folding and/or unfolding processes were independent of the molecular concentration, we observed hysteresis between melting and annealing curves for this particular conjugate (Figures S35 and S36 in the Supporting Information). Whereas the hysteresis was moderate in the presence of Na⁺ or in the absence of added cations (observed ΔT_m of 4 and 3 °C, respectively), it was remarkable in the presence of K⁺ (observed ΔT_m of 16 °C), which suggests a slower association rate under this particular condition. Collectively those data support the idea that the RNA strands within conjugate 3 engage in an intricate folding pathway governed by unfavorable kinetics to adopt a parallel G4 architecture of substantial thermodynamic stability.

SPR binding analysis

Extensive research has been carried out to identify small-molecule ligands with the ability to distinguish G4 structures from other abundant nucleic acid structures such as duplex DNA. By building on numerous successes, the research field is now shifting toward the development of ligands with the ability to distinguish distinct G4 folds. In this context, conjugates 1, 2, and 4 constitute the first benchmark platform to evaluate, in parallel, the propensity of ligands to associate with DNA, RNA,

or DNA:RNA hybrid G4 folds. The template-assembled DNA G4 conjugate **4** has been used in previous studies to investigate the interaction modes of families of G4 binding ligands.^[21b,33] Good agreements were found between those experiments and structural and affinity studies that involve the parental telomeric sequence from which the conjugate **4** derives.^[21b] To begin to compare the interactions between G4 ligands and the distinct RNA, DNA:RNA, and DNA G4 conjugates **1**, **2**, and **4**, we collected four reported G4 ligands—TMPyP4, MMQ1, BRACO-19, and NiSal-3im (Figure 4)—and carried out SPR binding analysis.

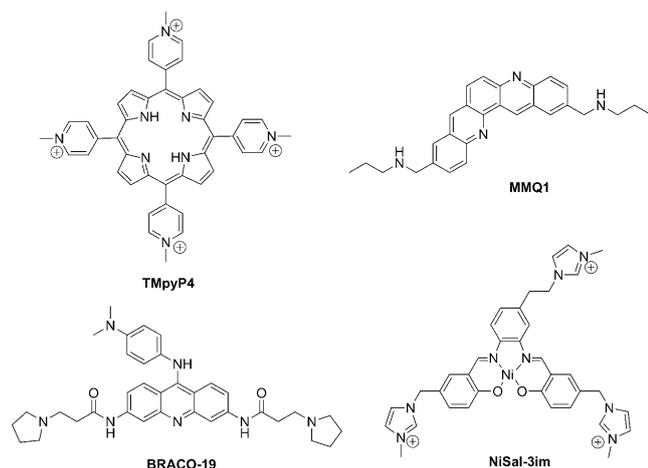


Figure 4. Structure of the G4 ligands subjected to SPR binding analysis.

TMPyP4 is a tetracationic porphyrin. Its association with G4 has been extensively studied *in vitro* and in cellular contexts.^[34] There is an overall agreement that the major contribution to G4 binding by the TMPyP4 porphyrin comes from π - π stacking with the external G tetrad. MMQ1 is a quinacridine derivative that has been shown to stabilize G4 structures and can operate as a telomerase inhibitor in cellular contexts.^[35] Again, the binding mode of MMQ1 ligands is dominated by stacking interactions. BRACO-19 is a trisubstituted acridine derivative that has demonstrated anticancer^[36] and antiviral activity^[37] through the stabilization of G4 at the telomere or in HIV-infected cells. In addition to stacking interactions, two flexible side chains provide BRACO-19 with the opportunity to engage contacts within the grooves of G4 structures.^[38] Lastly, NiSal-3im is a recently reported nickel(II) salophen complex that consists of an electro-deficient polyaromatic core functionalized with alkylimidazolium arms of different lengths (compound **8** in the literature).^[21a] NiSal-3im demonstrated a substantial preference for the G4 structure over DNA duplex, and the interaction has been shown to rely on both π stacking over the external G tetrad and electrostatic interactions within opposite grooves.

To perform the SPR binding analysis, conjugates **1**, **2**, and **4** were first anchored onto streptavidin-coated gold surfaces, and similar immobilization levels were obtained. Each ligand was injected simultaneously on all channels to allow for the direct comparison of recorded responses. Association/dissocia-

Table 2. Dissociation constants (K_D) [nM] for the interaction of ligands with G4 structures.

Ligands	RNA system 1	DNA:RNA system 2	DNA system 4
TmPyP4	197 ± 21	348 ± 30	317 ± 34
MMQ1	1449 ± 120	882 ± 70	722 ± 63
BRACO-19	261 ± 31	79 ± 11	8 ± 4
NiSal-3im	543 ± 58	306 ± 33	290 ± 31

tion curves were recorded, and the dissociation constant values were calculated from the kinetic parameters of the interactions (Table 2; see also the Supporting Information).

We found out that TMPyP4 displays very little selectivity for a given G4 architecture. This behavior was expected as the external G-tetrad configuration is likely to be very similar in all three target systems. On the other hand, MMQ1, NiSal-3im, and BRACO-19 were able to distinguish, to various extents, between the distinct G4 structures. Indeed, those three ligands showed higher affinity for DNA-based G4 than for RNA G4, which suggests the existence of specific interactions provided by the various side chains. Although the selectivity is modest for MMQ1 and NiSal-3im (up to twofold), BRACO-19 affinity appears to significantly increase with the numbers of DNA strands within a given G4 topology (up to 30-fold). It was reported that the 2'-hydroxyl groups within the RNA-based G4 architecture were capable of engaging in extensive interactions within the G4 structure and with the surrounding environment, thereby altering the groove distances.^[39] The lower affinity of BRACO-19 ligands for the RNA G4 structure might be connected to the steric constraints imposed by the 2'-hydroxyl groups because its side chains might not fit the particular RNA G4 grooves. In fact, the higher affinity of BRACO-19 for DNA versus the RNA telomeric G4 structure has already been reported in the past.^[40] In this previous work, molecular dynamics experiments suggested that unfavorable electronic repulsion might occur between the amide carbonyl oxygen atom of the BRACO-19 side chain and a 2'-hydroxyl group present within the RNA G4 groove. It is also interesting to note that the affinity of BRACO-19 toward the RNA:DNA hybrid G4 structure **2** is distinct from those obtained for RNA and DNA G4 motifs **1** and **4**. This observation demonstrates that the accessibility of G4 recognition elements is indeed different among sequence-related DNA, RNA, and DNA:RNA hybrid G4 model structures and supports the idea that structure selectivity might be achieved through the appropriate selection of aromatic chromophores and side chains.

Conclusion

A molecular engineering approach was developed to assemble and stabilize discrete RNA and DNA:RNA hybrid G4 topologies related to biologically relevant telomeric G4 structures. Importantly we report the first methodology to isolate and stabilize a DNA:RNA hybrid G4 motif, a structure that has been reported to be a putative new target for drug design. When compared with the corresponding DNA and RNA G4 motif, we un-

veiled that the hybrid structure is in fact as stable as the former. Even though the DNA, RNA, and DNA:RNA conjugates investigated in this study are lacking loop sequences, an important feature for ligand recognition, they constitute an extensive benchmark platform to begin assessing the specificity of G4-targeting ligands. By using our platform, we provide some evidence that supports the idea that selective targeting of distinct but related DNA, RNA, and DNA:RNA G4 folds is achievable. Structure-selective ligands in the context of telomeric DNA and RNA G4-forming sequences would significantly contribute to elucidate the existence and differentiate the functions of those G4 motifs from one another selectively in a cellular context. In a more general context, our study contributes to enable the identification of truly specific biological tracers and therapeutic modules for the *in vivo* targeting and visualization of G4.

In the course of this work, we also further emphasize that the formation of an antiparallel topology is strongly disfavored for RNA G4 structures. Even under forced orientations, we could not overcome the thermodynamic barrier that prevents the formation of this topology.

Beyond their uses as model target systems, the stable G4 conjugates reported in this study, along with future conjugates assembled by following the same methodologies, might find numerous applications that range from biology to biotechnology. We are currently investigating the use of stabilized G4 systems to highlight and identify G4 specific protein binders in cells. The unique features of template-assembled G4 motifs are also being exploited to assemble potent G4 decoys and aptamers, and for the design of G4-based sensors and catalysts.^[41] Efforts in those directions and progress in the development of loop-containing well-defined G4 architectures of biological significance will be reported in due course.

Experimental Section

General

ESI mass spectra were performed using an Esquire 3000 spectrometer from Bruker or an Acquity UPLC/MS system from Waters equipped with an SQ Detector 2. MALDI-TOF mass spectra were performed using an Autoflex Bruker with hydropiccolinic acid (HPA) (45 mg, ammonium citrate (4 mg) in 500 μ L H₂O/CH₃CN) as matrix. Peptides were analyzed in the positive mode and oligonucleotides and conjugates in the negative mode. NMR spectra were obtained using a Bruker Avance III spectrometer. All solvents and reagents used were of the highest purity commercially available.

Synthesis

Peptide scaffolds **5** and **6** were prepared according to previously reported procedures.^[29] RNA oligonucleotides were prepared with PivOM chemistry^[30] using an Expedite synthesizer at 1 μ mol scale using 3'-glyceryl and 3'-alkyne-modifier Serinol CPG for 3'-modification and 5'-hexynyl (β -cyanoethyl) phosphoramidite for 5'-modification (see the Supporting Information). RP-HPLC analysis was performed using a Waters HPLC system with a C18 Nucleosil column (Macherey-Nagel, 100 Å , 250 \times 4.6 mm, 5 μ m) with UV monitoring at 260 and 280 nm. A 1 mL min⁻¹-flow linear gradient was applied.

Solvent A¹ (50 mM triethylammonium acetate buffer with 5% acetonitrile) and solvent B¹ (acetonitrile with 5% water) were used. A stepwise gradient of 0–30% B¹ in 20 min followed by 30% to 100% B¹ within 5 min was applied. RP-HPLC purifications of oligonucleotides were performed using a Gilson system with a Nucleosil C-18 column (Macherey-Nagel, 100 Å , 250 \times 10 mm, 7 μ m) and a 4 mL min⁻¹-flow linear gradient from 100% A¹ and 0% B¹ to 45% B¹ for 20 min with UV monitoring at 260 and 280 nm. Anion-exchange HPLC (AE-HPLC) analysis was performed on a Dionex system using Dionex column (DNAPac PA-100, 4 \times 250 mm) with UV monitoring at 260 and 280 nm. A 1.5 mL min⁻¹-flow linear gradient from 95% A² (Tris buffer 0.4 M pH 7 with 5% acetonitrile) and 5% solvent B² (Tris buffer 0.4 M pH 7, lithium perchlorate trihydrate buffer 0.4 M with 5% acetonitrile) to 50% B² for 24 min was applied at 75 °C. Anion-exchange HPLC (AELC) purification was performed using a Dionex system with a Dionex column (DNAPac PA-100, 9 \times 250 mm) and a 5 mL min⁻¹ flow linear gradient with solvent A² and B² at 75 °C with UV monitoring at 260 and 280 nm. Desalting of oligonucleotides was carried out by size-exclusion chromatography (SEC) using a NAP 25 cartridge according to the manufacturer's protocol.

General procedure for oxime ligations

Aldehyde oligonucleotide (1.2 equiv by oxyamine function) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5, concentration 10⁻³ M), and free aminoxy peptide was added. The solution was stirred at 55 °C for 45 min.

General procedure for CuAAC reactions

CuSO₄ (3 equiv by azido function), THPTA (15 equiv by azido function), and sodium ascorbate (15 equiv by azido function) were added to a solution of alkyne oligonucleotide (1.2 equiv by azido function, concentration 10⁻³ M) in 100 mM HEPES buffer (pH 7.4). The reaction was stirred at 37 °C for 2 h and quenched with a solution of 0.5 M ethylenediaminetetraacetic acid (EDTA; 50 equiv by azido function).

Circular dichroism studies

Prior to CD analysis, compounds **1–4** were thoroughly desalted by SEC, then an annealing step was applied by heating the sample at 80 °C for 5 min in buffer (Tris 10 mM pH 7.4 without salt or with 10 mM NaCl or with 100 μ M KCl) and cooling it over 2 h to room temperature. Analyses were recorded using a Jasco J-810 spectropolarimeter with a 1 cm-length quartz cuvette. Spectra were recorded at 20 °C or every 5 °C over a range of 5 to 80 °C with a wavelength range of 220 to 330 nm. For each temperature, the spectrum was an average of three scans with a 0.5 s response time, a 1 nm data pitch, a 4 nm bandwidth, and a 200 nm min⁻¹ scanning speed. Melting temperatures were obtained using the Boltzmann fit with Origin software. Each curve fit was only accepted with an $r_{\text{value}} > 0.99$.

NMR spectroscopic experiments

¹H NMR spectroscopic (500 MHz) experiments were carried out with 80 μ M conjugates at 25 °C in sodium phosphate buffer (10 mM, pH 7.4) and in the presence of 10% D₂O and 100 mM NaCl.

SPR experiments

SPR data were obtained using a BIAcore T200 (GE Healthcare). Measurements were performed on gold surfaces. Sensor chips were cleaned by UV-ozone treatment for 10 min followed by rinsing with MilliQ water and ethanol. The cleaned gold surfaces were then functionalized according to the following procedure. First, mixed self-assembled monolayers (SAMs) were formed at room temperature by dipping gold sensors in the thiol mixture overnight: 90% HS-(CH₂)₁₁-EG4-OH and 10% HS-(CH₂)₁₁-EG6-Biotin (1 mM total thiol concentration in EtOH). After overnight adsorption, gold sensors were rinsed with ethanol and dried under nitrogen. The surface was then inserted into the BIAcore T200 device. All measurements were performed at 25 °C by using a running buffer (RB) composed of HEPES buffered saline: 0.01 M HEPES pH 7.4, 50 mM NaCl, 100 mM KCl, and 0.005% v/v surfactant P20. Streptavidin (100 ng mL⁻¹) was injected (10 μL min⁻¹) onto the biotinylated SAM until saturation of the surface (around 2500 RU). The different G quadruplex systems of **1**, **2**, and **4** were injected at 2 μL min⁻¹ onto the streptavidin-coated SAM surfaces until saturation (final response approximately 300 RU). Binding experiments were conducted at 30 μL min⁻¹ by injection of the different ligands (dissolved in RB, injection time: 200 s, dissociation time: 300 s, stabilization time: 300 s) at different concentrations by using a multi-cycle method. A streptavidin surface, prepared as described above, was used as reference. Curves obtained on the reference surface were deduced from the curves recorded on the recognition one, thereby allowing elimination of refractive index changes owing to buffer effects. The binding rate constants of G-quadruplex DNA/ligand interactions were calculated by means of a nonlinear analysis of the association and dissociation curves by using the BIAcore T200 (v2.0.1) SPR kinetic evaluation software. The data were fitted by using a heterogeneous ligand model (two interactions per G4 target; see the Supporting Information). The association rate constants, k_{on1} and k_{on2} , and the dissociation rate constants, k_{off1} and k_{off2} , as well as the theoretical maximal response, R_{max1} and R_{max2} of the two interactions were determined. The equilibrium dissociation constants were obtained from the binding rate constants as $K_{D1} = k_{off1}/k_{on1}$ and $K_{D2} = k_{off2}/k_{on2}$. We report the thermodynamic dissociation constants that were consistent between independent experiments and for which the theoretical maximum response (R_{max}) is consistent with 1:1 interactions. We chose not to report values for the second interaction that might involve non-stoichiometric binding and/or non-specific (ligands stacking) interactions (see the Supporting Information). The reported values are the mean of representative independent experiments, and the errors provided are standard deviations from the mean.

Acknowledgements

This work was supported by ANR Quarndiems (ANR-12-BSV8-0008-04), the Labex ARCAN (ANR-11-LABX-0003-01), and the AGIR program (UGA). The NanoBio-ICMG platforms (FR 2607) are acknowledged for their support.

Keywords: circular dichroism · click chemistry · DNA structures · G-quadruplexes · RNA structures

- [1] A. I. Karsisiotis, C. O'Kane, M. Webba da Silva, *Methods* **2013**, *64*, 28–35.
 [2] a) V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, S. Balasubramanian, *Nat. Biotechnol.* **2015**, *33*, 877–881; b) J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2005**, *33*, 2908–2916.

- [3] P. Agarwala, S. Pandey, S. Maiti, *Org. Biomol. Chem.* **2015**, *13*, 5570–5585.
 [4] S. Millevoi, H. Moine, S. Vagner, *Wires RNA* **2012**, *3*, 495–507.
 [5] a) G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, *Nat. Chem.* **2013**, *6*, 75–80; b) G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nat. Chem.* **2013**, *5*, 182–186.
 [6] A. Laguerre, K. Hukezalie, P. Winckler, F. Katranji, G. Chanteloup, M. Pirrotta, J. M. Perrier-Cornet, J. M. Wong, D. Monchaud, *J. Am. Chem. Soc.* **2015**, *137*, 8521–8525.
 [7] N. Maizels, *EMBO Rep.* **2015**, *16*, 910–922.
 [8] S. Xiao, J. Y. Zhang, K. W. Zheng, Y. H. Hao, Z. Tan, *Nucleic Acids Res.* **2013**, *41*, 10379–10390.
 [9] J. Y. Zhang, K. W. Zheng, S. Xiao, Y. H. Hao, Z. Tan, *J. Am. Chem. Soc.* **2014**, *136*, 1381–1390.
 [10] R. Y. Wu, K. W. Zheng, J. Y. Zhang, Y. H. Hao, Z. Tan, *Angew. Chem. Int. Ed.* **2015**, *54*, 2447–2451; *Angew. Chem.* **2015**, *127*, 2477–2481.
 [11] S. Xiao, J. Y. Zhang, J. Wu, R. Y. Wu, Y. Xia, K. W. Zheng, Y. H. Hao, X. Zhou, Z. Tan, *Angew. Chem. Int. Ed.* **2014**, *53*, 13110–13114; *Angew. Chem.* **2014**, *126*, 13326–13330.
 [12] D. Bhattacharyya, K. Nguyen, S. Basu, *Biochemistry* **2014**, *53*, 5461–5470.
 [13] a) S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discovery* **2011**, *10*, 261–275; b) S. A. Ohnmacht, S. Neidle, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2602–2612.
 [14] S. Müller, R. Rodriguez, *Expert rev. clin. pharmacol.* **2014**, *7*, 663–679.
 [15] S. Müller, D. A. Sanders, M. Di Antonio, S. Matsis, J. F. Riou, R. Rodriguez, S. Balasubramanian, *Org. Biomol. Chem.* **2012**, *10*, 6537–6546.
 [16] C. Musetti, A. P. Krapcho, M. Palumbo, C. Sissi, *Plos One* **2013**, *8*, e58529.
 [17] J. Palacky, M. Vorlickova, I. Kejnovska, P. Mojzes, *Nucleic Acids Res.* **2013**, *41*, 1005–1016.
 [18] a) M. Nikan, B. O. Patrick, J. C. Sherman, *Chembiochem* **2012**, *13*, 1413–1415; b) R. Ferreira, M. Alvira, A. Avino, I. Gomez-Pinto, C. Gonzalez, V. Gabelica, R. Ertija, *ChemistryOpen* **2012**, *1*, 106–114; c) G. Oliviero, N. Borbone, J. Amato, S. D'Errico, A. Galeone, G. Piccialli, M. Varra, L. Mayol, *Biopolymers* **2009**, *91*, 466–477.
 [19] P. Murat, D. Cressend, N. Spinelli, A. Van der Heyden, P. Labbe, P. Dumy, E. Defrancq, *Chembiochem* **2008**, *9*, 2588–2591.
 [20] R. Bonnet, T. Lavergne, B. Gennaro, N. Spinelli, E. Defrancq, *Chem. Commun.* **2015**, *51*, 4850–4853.
 [21] a) L. Lecarme, E. Prado, A. De Rache, M. L. Nicolau-Travers, R. Bonnet, A. van Der Heyden, C. Philouze, D. Gomez, J. L. Mergny, H. Jamet, E. Defrancq, O. Jarjayes, F. Thomas, *Inorg. Chem.* **2014**, *53*, 12519–12531; b) P. Murat, R. Bonnet, A. Van der Heyden, N. Spinelli, P. Labbe, D. Monchaud, M. P. Teulade-Fichou, P. Dumy, E. Defrancq, *Chem. Eur. J.* **2010**, *16*, 6106–6114.
 [22] V. Gabelica, *Biochimie* **2014**, *105*, 1–3.
 [23] C. Wang, L. Zhao, S. Lu, *Int. j. biol. sci.* **2015**, *11*, 316–323.
 [24] K. Hirashima, H. Seimiya, *Nucleic Acids Res.* **2015**, *43*, 2022–2032.
 [25] Y. Xu, T. Ishizuka, J. Yang, K. Ito, H. Katada, M. Komiyama, T. Hayashi, *J. Biol. Chem.* **2012**, *287*, 41787–41796.
 [26] K. D. Warner, M. C. Chen, W. Song, R. L. Strack, A. Thorn, S. R. Jaffrey, A. R. Ferre-D'Amare, *Nat. Struct. Mol. Biol.* **2014**, *21*, 658–663.
 [27] O. Mendoza, M. Porrini, G. F. Salgado, V. Gabelica, J. L. Mergny, *Chem. Eur. J.* **2015**, *21*, 6732–6739.
 [28] P. Dumy, I. M. Eggleston, G. Esposito, S. Nicula, M. Mutter, *Biopolymers* **1998**, *39*, 297–308.
 [29] R. Bonnet, P. Murat, N. Spinelli, E. Defrancq, *Chem. Commun.* **2012**, *48*, 5992–5994.
 [30] T. Lavergne, J. R. Bertrand, J. J. Vasseur, F. Debart, *Chem. Eur. J.* **2008**, *14*, 9135–9138.
 [31] J. L. Mergny, J. Li, L. Lacroix, S. Amrane, J. B. Chaires, *Nucleic Acids Res.* **2005**, *33*, e138.
 [32] A. Randazzo, G. P. Spada, M. W. da Silva, *Topics in Current Chem.* **2012**, *330*, 67–86.
 [33] a) L. Sabater, M. L. Nicolau-Travers, A. De Rache, E. Prado, J. Dejeu, O. Bombarde, J. Lacroix, P. Calsou, E. Defrancq, J. L. Mergny, D. Gomez, G. Pratviel, *J. Biol. Inorg. Chem.* **2015**, *20*, 729–738; b) L. Sabater, P. J. Fang, C. F. Chang, A. De Rache, E. Prado, J. Dejeu, A. Garofalo, J. H. Lin, J. L. Mergny, E. Defrancq, G. Pratviel, *Dalton Trans.* **2015**, *44*, 3701–3707.
 [34] L. Martino, B. Pagano, I. Fotticchia, S. Neidle, C. Giancola, *J. Phys. Chem. B* **2009**, *113*, 14779–14786.

- [35] C. Hounsou, L. Guittat, D. Monchaud, M. Jourdan, N. Saettel, J. L. Mergny, M. P. Teulade-Fichou, *ChemMedChem* **2007**, *2*, 655–666.
- [36] C. M. Incles, C. M. Schultes, H. Kempfski, H. Koehler, L. R. Kelland, S. Neidle, *Mol. Cancer Ther.* **2004**, *3*, 1201–1206.
- [37] R. Perrone, E. Butovskaya, D. Daelemans, G. Palu, C. Pannecouque, S. N. Richter, *The Journal of Antimicrobial Chemotherapy* **2014**, *69*, 3248–3258.
- [38] N. H. Campbell, G. N. Parkinson, A. P. Reszka, S. Neidle, *J. Am. Chem. Soc.* **2008**, *130*, 6722.
- [39] G. W. Collie, S. M. Haider, S. Neidle, G. N. Parkinson, *Nucleic Acids Res.* **2010**, *38*, 5569–5580.
- [40] G. Collie, A. P. Reszka, S. M. Haider, V. Gabelica, G. N. Parkinson, S. Neidle, *Chem. Commun.* **2009**, 7482–7484.
- [41] L. Stefan, T. Lavergne, N. Spinelli, E. Defrancq, D. Monchaud, *Nanoscale* **2014**, *6*, 2693–2701.

Received: October 29, 2015

Published online on January 25, 2016

Click Chemistry

Template-Mediated Stabilization of a DNA G-Quadruplex formed in the HIV-1 Promoter and Comparative Binding Studies

Laureen Bonnat,^[a, b] Laure Bar,^[a] Béatrice Génaro,^[a] Hugues Bonnet,^[a] Olivier Jarjayes,^[a] Fabrice Thomas,^[a] Jérôme Dejeu,^{*[a]} Eric Defrancq,^[a] and Thomas Lavergne^{*[a]}

Abstract: G-rich DNA oligonucleotides derived from the promoter region of the HIV-1 long terminal repeat (LTR) were assembled onto an addressable cyclopeptide platform through sequential oxime ligation, a thiol-iodoacetamide SN2 reaction, and copper-catalyzed azide-alkyne cycloaddition reactions. The resulting conjugate was shown to fold into a highly stable antiparallel G4 architecture as demonstrated by UV, circular dichroism (CD), and NMR spectroscop-

ic analysis. The binding affinities of six state-of-the-art G4-binding ligands toward the HIV-G4 structure were compared to those obtained with a telomeric G4 structure and a hairpin structure. Surface plasmon resonance binding analysis provides new insights into the binding mode of broadly exploited G4 chemical probes and further suggests that potent and selective recognition of viral G4 structures of functional significance might be achieved.

Introduction

Over the past decade significant efforts have been invested to clarify the occurrence and function of unusual nucleic acid structures that differ from canonical Watson–Crick-based duplexes. Collectively those data suggest that both DNA and RNA noncanonical structural architectures may hold a pivotal role in numerous cellular processes. Amongst high-order nucleic acid structures, the so-called G-quadruplex (G4) conformations have unarguably attracted most of the attention. G4 are nucleic acid structures within which four G-rich tracts hold together through the formation and stacking of two or more G-tetrad units (Figure 1A).^[1] The G-tetrads, formed by a network of Hoogsteen hydrogen bonds, are themselves strongly stabilized by the coordination of monovalent cations, such as K⁺ and Na⁺. The four G-tracts can run in the same or opposite directions in respect to their neighbors thus forming distinct types of G4 folds termed parallel, antiparallel (as in Figure 1), and hybrid. Depending on the strand polarity, a G4 structure comprises of four negatively charged grooves of different sizes classified as narrow, medium, or wide (Figure 1A). Finally, G-tracts are connected by nucleotide sequences of different length that may adopt distinct conformations, depending on

the polarity of consecutive strands, and which are termed lateral (as in Figure 1), diagonal, or propeller loops.

The formation of G4 structures within G-rich regions of the human genome is now firmly established. In cellulo visualization of both DNA and RNA G4 structures, through the use of G4-specific small-molecule and antibody binding agents, have been independently reported.^[2] The propensity of G-rich sequences to form G4-structures has been recently re-assessed using whole genome and transcriptome sequencing technologies^[3] and refined with in silico predictions,^[4] further supporting the prevalence of G4 structures within key regulatory regions of human genes. In particular, G4 formation at telomeres and within the promoter region of oncogenes has been found worthy of acute attention as those structures are now considered as novel anticancer drug targets. Indeed, several studies have shown that inhibition of telomerase activity and transcriptional repression of oncogenes could be achieved using G4-stabilizing binding ligands.^[5] In parallel, it has also been shown over the past years that G4-based regulation was not limited to human or higher vertebrates. Evidence of G4 formation within the genomes of protozoa,^[6] bacteria,^[7] and viruses^[8] have been reported suggesting functional significance. As to the latest family of pathogens, G4-forming sequences have been reported in the genomes of the human immunodeficiency virus (HIV-1),^[9] Papillomavirus,^[10] Herpes virus,^[11] and, recently, within the genome of viruses involved in the latest epidemic outbreaks such as in the Ebola virus^[12] and Zika virus.^[13] Analogous to anticancer strategies aiming at telomeric and oncogenic G4s, studies suggest that viral DNA and/or RNA G4 might be targeted by G4-selective ligands to impart key virus functions, paving the way for the development of new strategies against viral diseases.^[14] Such novel antiviral strategies are currently being explored in the context of HIV-1 infection. Within the HIV-1 genome, G4-forming DNA sequences can be

[a] L. Bonnat, L. Bar, B. Génaro, H. Bonnet, Dr. O. Jarjayes, Prof. F. Thomas, Dr. J. Dejeu, Prof. E. Defrancq, Dr. T. Lavergne
Univ. Grenoble Alpes, CNRS, DCM UMR-5250, 38000 Grenoble (France)
E-mail: jerome.dejeu@univ-grenoble-alpes.fr
thomas.lavergne@univ-grenoble-alpes.fr

[b] L. Bonnat
Univ. Grenoble Alpes, CNRS, DPM UMR-5063, 38000 Grenoble (France)

Supporting information and the ORCID identification number for the author of this article can be found under:
<https://doi.org/10.1002/chem.201700417>

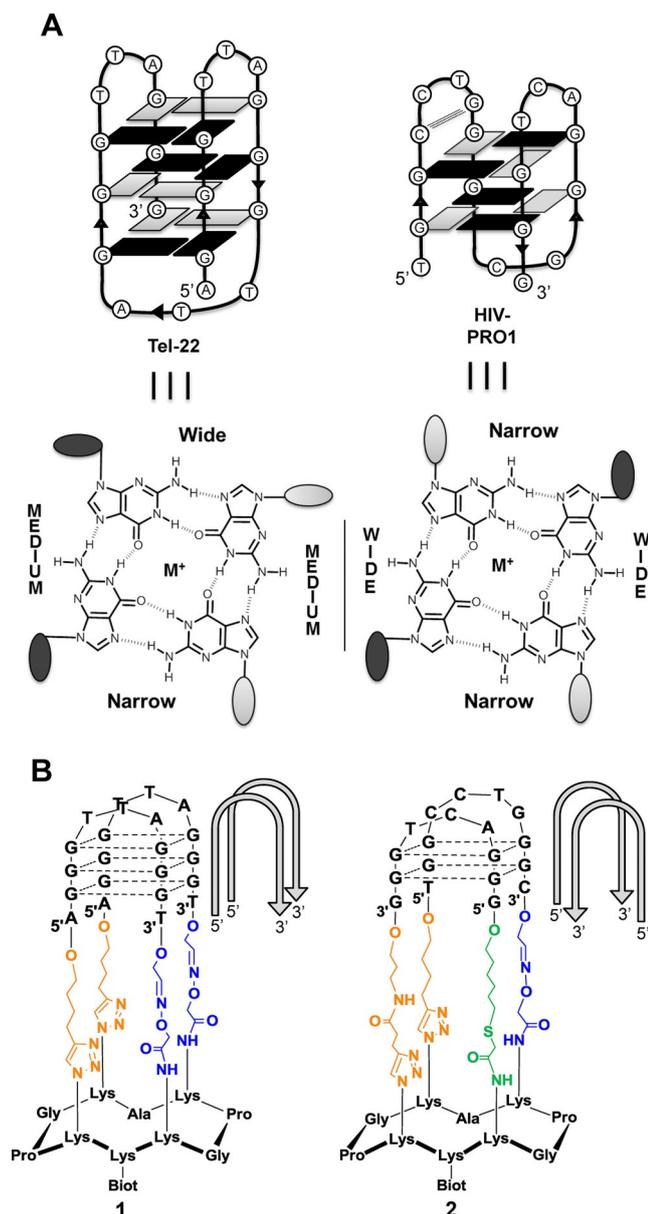


Figure 1. A) Schematic drawing of G4 structures formed by the human telomeric DNA sequence (Tel-22) and the HIV-1 promoter sequence (HIV-PRO1) and corresponding G-tetrad arrangements (black oval and box = (*anti*) guanine, grey oval and box = (*syn*) guanine. B) Previously reported DNA conjugate 1 and DNA conjugate 2 assembled in this study.

found in the coding region of the Nef gene and within the U3 promoter region of the Long Terminal Repeat (LTR).^[8] A G4-mediated transcriptional regulation of the Nef gene has been evidenced, suggesting that the G4 structure might play a central role in viral pathogenicity.^[9a] The LTR promoter contains five transcription factor binding sites and is consequently crucial for viral transcription and latency.^[9b] There is also evidence that suggests the U3 region of LTR might be dynamically forming G4 structures to bind those cellular transcription factors.^[15] Besides, G4 structures can also form in regions of the HIV-1 RNA genome where they appear to be involved in regulating the reverse transcription of the genome.^[16]

To probe the effect of G4-inducing/stabilizing ligands on HIV progression, reporter genes assays and virological assays have been implemented and acridine (BRACO-19),^[14b] porphyrin (TMPyP4)^[9a] and naphthalene derivatives (NDI)^[14a] have been shown to reduce HIV-1 infectivity with activities in the range of micromolar (BRACO-19) to nanomolar (NDI) concentrations. In the course of those studies, it has also been further emphasized that distinct G4 tracts and different G4 conformations could be associated with distinct viral functions. This topological and functional diversity suggest that a high level of drug selectivity might be achieved and that selective targeting could lead to efficient and specific antiviral treatments. Yet, most of the G4-binding ligands tested so far have been shown to interact with genomic and/or telomeric DNA and RNA G4. The identification of selective ligands for distinct G4-folds would thereby allow for further examination of G4 biological functions and the emergence of novel molecules with genuine therapeutic potential.

A first step toward the identification of a truly selective binding agent might consist in the side-by-side comparison of the association and dissociation properties of various ligands toward distinct viral, genomic, and telomeric G4-folds. However, this approach is complicated by the structural instability and/or polymorphism of G4 nucleic acids formed by synthetic G-rich sequences isolated in solution. Indeed, it has been shown that small variation in the experimental conditions, such as the nature and concentration of metal cations or the folding procedure, could trigger major G4 structural reorganizations thereby preventing the meticulous characterization of ligands/G4 interactions.^[17]

In this context, and over the past few years, we,^[18] and others,^[19] have elaborated molecular systems capable of mimicking a range of well-defined DNA and RNA G4 topologies. By site-specifically fixing DNA and/or RNA G4-forming sequences onto addressable molecular scaffolds, we have been able to trigger the formation of topologically controlled and thermally stabilized G4 architectures termed TASQs (i.e. template assembled synthetic G-quadruplexes).^[18] Those stable structural mimics have been extensively used to probe the binding mode of various families of G4 binding ligands.^[18a,20] The scaffolds used to site-specifically connect the DNA and RNA strands consist of cyclic decapeptides with two independent and addressable faces (as in Figure 1 B).^[21] The “upper” domain displays diversely functionalized side chains to attach the oligonucleotide strands, whereas the “lower” domain can be functionalized for non-perturbative surface immobilization.

An antiparallel dimeric DNA-G4, the primary sequence of which derives from the human telomeric repeat sequence Tel-22 d(5'AG₃T₂AG³) and the folded structure of which matches one of the reported native conformation, was assembled (1, Figure 1).^[18b] To achieve the assembly of this bioconjugate, sequential oxime ligations (OL) and copper-catalyzed azide-alkyne cycloaddition reactions (CuAAC) were performed between the peptide scaffold presenting aminooxy and azido moieties and two identical oligonucleotide bearing complementary aldehyde and alkyne functions, respectively. The resulting G4 topology displayed high stability, with a melting

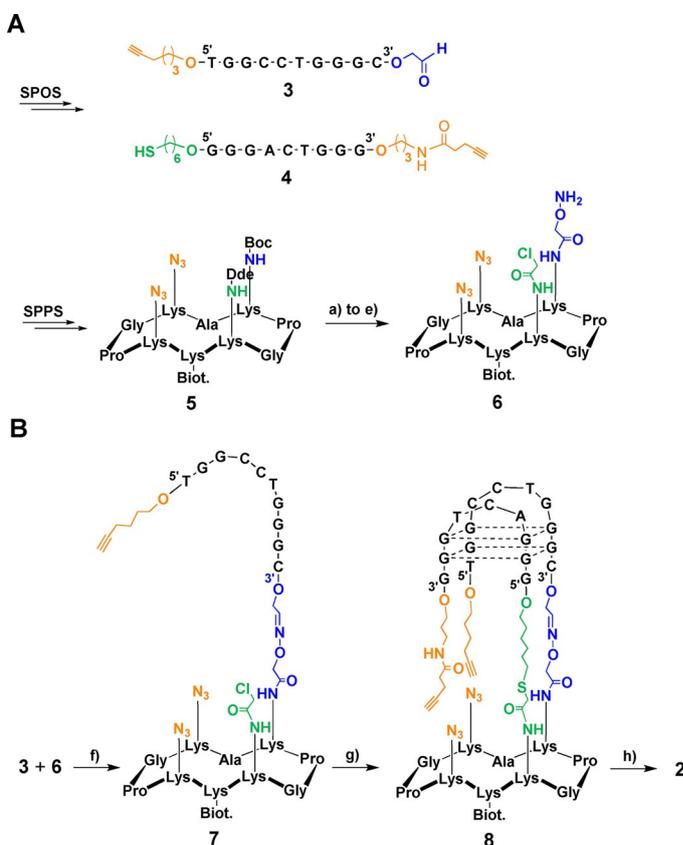
temperature differential of about 30 °C when compared with the native G4 fold, thus paving the way to the exhaustive evaluation of ligand binding properties.

An expansion of this approach to isolate and stabilize biologically relevant viral G4 structures and the subsequent characterization of their recognition by targeting ligands would be of particular interest.

In this paper, we report on the assembly, characterization, and targeting of a G4 structure derived from the LTR region of HIV-1. The U3 region of the LTR possesses nine G-tracts and, according to a predictive algorithm, twelve distinct sequences have the propensity to fold into a G4 structure. Some of those sequences have been already been shown to fold into a stable G4 structure.^[9b] For instance, the sequence d(5'TGGCCTGGGCGGGACTGGG3'), termed HIV-PRO1, has been identified as the most represented motifs after analysis of the alignment of position -57 to -78 of 1684 distinct HIV-1 subspecies.^[22] According to NMR studies, this sequence forms an antiparallel chair-type G4 topology with two stacked G-tetrads and an additional CG base pair within a connecting loop (Figure 1A). The G4 DNA forming strands run in opposite directions with respect to their two neighbors which results in the formation of two wide and two narrow grooves (Figure 1A). G4 structures folded around two G-tetrad layers are commonly less stable than three G-tetrad layers ones. Nonetheless, it has been suggested that, either thermodynamically stable or transiently folded, such structures may abundantly arise in vivo thereby contributing to the regulation of numerous cellular processes. Amrane et al., reported that, whereas the HIV-PRO1 sequence would not fold into a G4 structure in the absence of added stabilizing metal cations, the thermal stability of the antiparallel G4 structure, in potassium rich solution, was as high as 56 °C.^[22] Surprisingly, we found out that the HIV-PRO1 sequence did not fold into a G4 structure even in presence of high concentration of sodium cations (vide infra).

To mimic and further stabilize this reported folded topology, we hypothesized that G-rich DNA oligonucleotides of sequence d(5'TGGCCTGGG3') and d(5'GGACTGGG3') could be site-specifically fixed onto a cyclic peptide scaffold (**2**, Figure 1B). However, and unlike with telomeric G4-forming conjugate **1**, the site-specific attachment of the two distinct oligonucleotides onto a cyclopeptide scaffold implies the use of an additional orthogonal chemical ligation step along with OL and CuAAC (as shown in Figure 1B). Beyond the imperative of chemical orthogonality associated with the synthesis of any bioconjugate, the assembly of well-structured nucleic acid foldamers such as **2** requires the use of chemical ligations that respect additional criteria. First, the steric parameters of the newly formed chemical bonds have to be well-thought-out so that it does not affect the structural folding. For analogous reasons, the complementary functionalities have to be installed onto linkers of comparable lengths. Finally, to achieve the intramolecular macrocyclization, at least one of the chemical ligations has to proceed with high yield under stoichiometric conditions. The SN2 reaction between thiol containing biopolymers and reporters bearing iodoacetamide functions has been extensively reported in the past.^[23] We hypothesized that such

an SN2-thiol coupling reaction (TC) could be used along with OL and CuAAC for the stepwise assembly of nucleic acid foldamers **2** from bifunctionalized oligonucleotides and an adequately heterofunctionalized peptide scaffold (Scheme 1). In



Scheme 1. A) Synthesis of peptide and oligonucleotide building blocks. a) hydrazine 2% in DMF, 5 min at RT; b) chloroacetic anhydride (2 equiv) in DMF, pH 8 with DIEA, 1 h 30 min at RT; c) TFA/CH₂Cl₂/H₂O 45:45:10, 20 min at RT; d) *N*-Boc-aminoxyacetic acid *N*-hydroxysuccinimide ester (1.2 equiv) in anhydrous DMF, pH 8–9 with DIEA, 1 h at RT; e) TFA/H₂O/TIS (80:16:2), 10 min at RT. B) Assembly of conjugate **2**. f) oligonucleotide **3** (1 equiv), peptide **6** (1.2 equiv), 0.4 M ammonium acetate buffer, pH 4.5, 45 min at 55 °C; g) oligonucleotide **4** (2 equiv), TCEP (2 equiv), DIEA (45 equiv), KI (60 equiv), KCl 500 mM, pH 8.5, under argon, 5 h at RT; h) HEPES 100 mM, 10% DMF, CuSO₄ (6 equiv), THPTA (30 equiv), sodium ascorbate (30 equiv), KCl 100 mM, pH 7.4, 2 h at 37 °C.

the course of this work, we have found that the stepwise conjugation of oligonucleotide **3**, bearing an aldehyde function at its 3' end, and oligonucleotide **4**, bearing a thiol function at its 5' end, onto cyclopeptide **6**, functionalized with both an oxyamine and a chloroacetamide moieties, allowed for the assembly of an intermediate conjugate **8** that may be subjected to intramolecular CuAAC reactions to form **2**. All functionalities were found to be compatible and all ligations steps were performed with good to excellent yields. Final conjugate **2** was obtained with high purity according to reverse-phase HPLC (RP-HPLC), ESI, and MALDI-TOF mass spectrometry. As anticipated, conjugate **2** was able to fold into a highly stable antiparallel G4 topology as validated by UV thermal difference studies (TDS), circular dichroism (CD) and NMR spectroscopy.

The ability of six small-molecule ligands, representative of prominent classes of G4 binders, to associate with the HIV-G4 structure formed within **2** was assessed using surface plasmon resonance (SPR). We also investigated the interactions of those derivatives of acridine (BRACO-19, MMQ1), porphyrin (TMPyP4), quinoline (PDS, PhenDC3), and salophen (NiSal-3im) with the telomeric G4 structure formed within **1** and with a DNA-hairpin forming conjugate previously assembled (**9**, Figure S42, Supporting Information). A significant disparity in affinity and specificity was observed amongst the eighteen ligands/targets pairs that were studied. Whereas most G4-targeting ligands displayed promiscuous affinities toward the distinct G4 structures, one ligand was found to discriminate, to some extent, between **1** and **2** supporting the idea that selective recognition might be achieved. Further examination of the kinetics of SPR binding studies also revealed remarkable differences in the binding mechanism of broadly employed G4-binding chemical probes.

Results and Discussion

Synthesis of conjugates

To assemble conjugate **2**, oligonucleotides **3** $d(5'X\text{-TGGCTGGGC-}Y^3)$ and **4** $d(5'Z\text{-GGGACTGGG-}X^3)$, in which *X* represents 3'- or 5'-alkynyl linkers, *Y* a 3'-aldehyde linker, and *Z* a 5'-thiol linker, were attached through stepwise OL, TC, and CuAAC reactions onto cyclopeptide **6** (Scheme 1). To synthesize cyclopeptide scaffold **6**, peptide precursor **5** was first assembled using standard fluorenylmethoxycarbonyl (Fmoc)/*t*Bu solid-phase peptide synthesis (SPPS) on a 2-chlorotrityl resin and cyclized in solution according to reported procedures.^[24] A biotin was fixed onto the lower domain of the peptide scaffold by using a biotin-functionalized lysine building block during SPPS. Azido moieties were also introduced during SPPS using the α -azidonorleucine building block. To equip cyclopeptide scaffold **6** with an oxyamine and a chloroacetamide function, two orthogonally protected lysines were incorporated during SPPS. Starting from **5**, the chloroacetamide group was anchored on a lysine side chain after removal of bis-*N*-[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl] (DDE) using chloroacetic anhydride. The aminoxy group was then introduced on the second lysine side chain, after *tert*-butoxycarbonyl (Boc) removal, by the addition of *N*-Boc-aminoxyacetic acid *N*-hydroxysuccinimide (NHS) ester and subsequent acidic deprotection. To synthesize oligonucleotides **3** and **4**, solid-phase oligonucleotide syntheses (SPOS), using phosphoramidite chemistry, were performed on commercially available glyceryl solid support and phtalimidyl solid support, respectively. Alkynyl and thiol functional groups were introduced at the 5'-terminus of oligonucleotides **3** and **4** by using commercially available hexynylphosphoramidite and trityl-protected C6-thiol phosphoramidite, respectively (Supporting Information). Following cleavage from the solid support and deprotection steps, oligonucleotide **3** was obtained after oxidative cleavage of the 3'-diol group (see the Supporting Information). Oligonucleotide **4** was obtained by reacting pentynoic acid NHS ester

onto its 3'-amino precursor generated after release and deprotection (Supporting Information).

Toward **2**, oligonucleotide **3** and peptide **6** were first covalently grafted by oxime bond formation to form intermediate **7**. The assembly was carried out in ammonium acetate buffer (pH 4.5) at 55 °C using an excess of aminoxy-containing peptide **6**. The reaction was monitored by RP-HPLC and proceeded to completion within 45 min. Purification using RP-HPLC followed by desalting using size-exclusion chromatography (SEC) afforded pure conjugate **7** with an isolated yield of 84%. Conjugate **7** was characterized by MALDI-TOF and ESI MS analysis, which showed excellent agreement between the experimentally determined molecular weight and the calculated value (Supporting Information).

The reaction between thiol-containing biomolecules and chloroacetamide-containing reporters (and vice versa) has been investigated in a number of independent studies.^[23] The reaction is usually performed in aqueous/organic mixtures with an excess of 2 to 10 equivalents a one reactant over the other. The reaction is also performed under reducing conditions (tris(2-carboxyethyl)phosphine (TCEP)) to prevent the formation of disulfide species. Potassium iodide (KI) is frequently used for in situ activation of the reaction through the transient formation of a reactive iodoacetamide group. Lastly, the reaction is performed around pH 8.5 by the addition of tertiary amines or in the presence of an organic buffer agent.

Nonetheless, under standard reaction conditions, that is an excess of two equivalents of oligonucleotide **4** over peptide **7** in a water/acetonitrile solution (9:1 ratio) containing 20 molar equivalents of KI, 30 molar equivalents of *N,N*-diisopropylethylamine (DIEA) and two molar equivalents of TCEP, we could only observe the sluggish conversion of about 20% of conjugate **7** into conjugate **8** after 24 h at room temperature according to RP-HPLC monitoring. Whereas heating of the reaction mixture to 50 °C positively affected the conversion rate, it also resulted in the reduction of the azido moieties by TCEP. Heating of the reaction to 50 °C in the absence of TCEP resulted in the formation of dimeric disulfide-linked species from **4**, further confirming that the presence of a disulfide cleaving agent is required.

To account for the poor reactivity observed when the reaction was performed at room temperature, we hypothesized that unfavorable charge repulsions between **4** and the non-complementary DNA strand within conjugate **7** might contribute to prevent the formation of **8**. Repulsion of densely immobilized DNA probes have been studied in the past and cation shielding was identified as an efficient approach to stabilize such closely spaced oligonucleotides.^[25] With the aim of improving the rate of the TC reaction, we thereby explored the addition of large concentrations of metallic or organic cations. The TC reaction was performed under the conditions depicted above and in the presence of either 500 mM KCl, 500 mM NaCl, 500 mM LiCl, or 500 mM NH₄Cl. The conversion ratios were determined after 3 h using RP-HPLC analysis and were found to be 55% (KCl), 40% (NaCl), 20% (LiCl), and 5% (NH₄Cl), endorsing the beneficial outcome of the addition of large concentrations of neutralizing cations. Differences be-

tween cationic salts might be rationalized by the disparate charge screening capacities of cationic species of different nature and size.

Starting from the most effective reaction conditions (i.e. with 500 mM KCl) we further explored the experimental conditions. We observed that the addition of four molar equivalents of oligonucleotide **4**, instead of two, did not positively affect the conversion rate. On the other hand, we found out that increasing the molar equivalent ratio of KI to 60 while increasing the one of *N,N*-diisopropylethylamine (DIEA) to 45 (in order to maintain a pH of 8.5) improved the conversion ratio to 76% after 3 h at room temperature. Under this set of conditions, the reaction was found to be nearly quantitative after 5 h reaction at room temperature. Using this last set of conditions, we then explored whether we could further optimize the concentration of KCl. While the use of 1 M KCl concentration resulted in a conversion ratio similar to the one observed with 500 mM, we observed a dramatic decrease to 20% conversion after 3 h when the reaction was performed with a KCl concentration of 100 mM. Finally, we also explored the consequence of swapping the organic co-solvent of the reaction. Briefly, we found out that dimethylsulfoxide could be used in place of acetonitrile without any positive or detrimental outcome on the conversion rate, whereas the reaction was not as efficient when conducted in the presence of MeOH.

The TC reaction between oligonucleotide **4** and conjugate **7** was proficiently scaled-up under the optimized conditions (Scheme 1, g). Purification using RP-HPLC followed by desalting using SEC afforded pure conjugate **8**, with an isolated yield of 70%, according to RP-HPLC, MALDI-TOF, and ESI MS analysis (Supporting Information).

The conversion of conjugate **8** to conjugate **2** was then performed under nonstandard CuAAC conditions previously developed for the synthesis of conjugate **1**.^[18b] The intramolecular reactions were performed in 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid (HEPES) buffer (pH 7.4) and in the presence of copper sulfate (CuSO₄), tris(benzyltriazolylmethyl)amine (THPTA) ligand, sodium ascorbate, and 100 mM potassium chloride. Indeed, the addition of a large concentration of G4-stabilizing cations (i.e. KCl), before initiating the CuAAC macrocyclization reactions, was expected to drive the DNA strands of conjugate **8** to adopt an antiparallel G4 topology (as drawn in Scheme 1) thereby promoting the formation of the desired product **2**. Such conformation-assisted enhancement of the reactivity was evidenced for the assembly of **1** during which stabilizing cations were successfully used to hold the azide and alkyne groups close to each other thereby lowering the entropic cost of the reactions and preventing the formation of branched products.^[18b] The stabilization of the antiparallel G4 structure of HIV-PRO1 by K⁺ has been reported,^[22] and we could confirm the formation, in the presence of K⁺, of a stable antiparallel G4 structure within **8** (vide infra, CD analysis Figure 2 and Table 1). Thus, under the amended CuAAC reaction conditions (Scheme 1, h), and, as anticipated, the conversion of **8** to **2** proceeded to completion in 2 h at room temperature as evidenced by RP-HPLC monitoring with the appearance of a sharp peak of lower retention time (Supporting Infor-

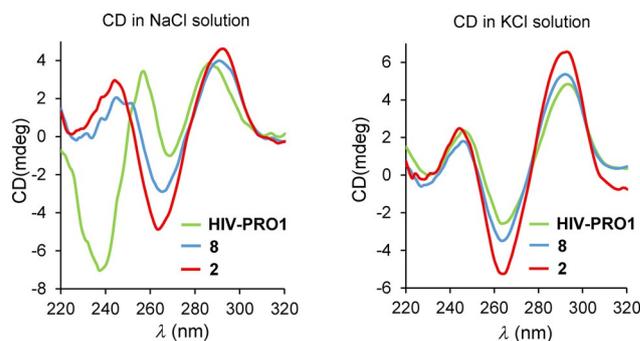


Figure 2. CD spectra of G-quadruplex-forming sequence HIV-PRO1 (green), conjugate **8** (blue) and **2** (red) at 1.6 μm in 10 mM Tris-HCl, pH 7.4, 20 °C in the presence of either 100 mM NaCl or 100 mM KCl.

Table 1. Melting temperature values from CD denaturation studies.

conditions ^[a]	HIV-PRO1		HIV-G4 system 8		HIV-G4 system 2	
	NaCl	KCl	NaCl	KCl	NaCl	KCl
<i>T_m</i> [°C]	n/a	56	51	53	67	76

[a] G4 structures were first annealed by heating at 80 °C for 5 min in Tris-HCl buffer (10 mM, pH 7.4) with 100 mM NaCl or 100 mM KCl and cooled down over 2 h to room temperature. Samples were prepared at 1.6 μm (see the Supporting Information).

mation, Figure S25). Conjugate **2** was purified using RP-HPLC and the identity and purity of the product were confirmed by RP-HPLC, ESI and MALDI-TOF MS analyses.

Characterization of G-quadruplex topologies

To confirm the formation of a stable G4 structure within both conjugates **8** and **2**, we performed the analysis of their thermal absorbance difference spectra (TDS) along with that of the parental HIV-PRO1 DNA sequence d⁽⁵⁾TGGCCTGGGCGGGACTGGG⁽³⁾.^[26] TDS were collected according to reported procedure (Supporting Information and ref. [27]) and in presence of either 100 mM Na⁺ or K⁺ (Supporting Information, Figures S29–S31). Remarkably, we observed that whereas the TDS of the parental HIV-PRO1 DNA sequence shows the characteristic peaks of G4 architectures, with a positive peak at around 273 nm and a negative peak at around 295 nm, when the data were collected in presence of K⁺, the TDS profile does not display any of those characteristic peaks in presence of Na⁺, which suggests the absence of a folded G4 structure (Supporting Information, Figure S29). On the other hand, TDS profiles of **2** as well as **8** exhibit the characteristics of G4 architectures in both Na⁺- and K⁺-rich solutions suggesting the formation of G4 structures in a broader range of experimental conditions (Supporting Information, Figures S30 and S31).

To further confirm the formation of G4 structures within **2** and **8** and to get insights into the adopted topology, CD analyses were performed in the presence of either 100 mM NaCl or 100 mM KCl.^[27] The CD spectra of **2**, **8**, and HIV-PRO1 recorded in the presence of K⁺ display peaks that are characteristic of

antiparallel G4 topologies with two positive peaks at 242 and 294 nm and a negative peak at 262 nm (Figure 2). On the other hand, CD spectra recorded in the presence of Na⁺ reveal peaks that are characteristic of antiparallel G4 topologies for both conjugates **2** and **8** but not for the HIV-PRO1 sequence. CD data are in agreement with the TDS data and suggest that the native sequence do not fold into a stable G4 structure in the presence of Na⁺. While not unprecedented, the exclusive reliance in K⁺ for folding remains a rare feature of G4-forming sequences as the formation of G4 structures is usually promoted by either Na or K cations. Those experimental data further indicate that the covalent anchoring of the DNA strands onto the peptide scaffold promote the formation of the antiparallel G4 structure in a broader range of conditions.

Preliminary ¹H NMR spectroscopic studies were also performed to further confirm the formation of G4 architectures within conjugate **2**. ¹H NMR analysis of **2** in sodium- and potassium-based aqueous solutions showed the presence of guanine imino proton resonances within the range of δ = 10–12 ppm that are characteristics of G4 folds (Supporting Information, Figures S39 and S40). In the presence of potassium, a broad lower-field peak at δ = 13.4 ppm suggests the presence of an additional Watson–Crick base pair. This is consistent with the formation of the CG base pair within the longer G-tract connecting loop (⁵CCTG³ loop) of HIV-PRO 1 as described by Amrane et al.^[22] Remarkably, ¹H NMR analysis of the native HIV-PRO1 sequence in the presence of sodium cations further confirmed the absence of a folded G4-structure as we could not observe the typical resonances of the guanine imino protons (Supporting Information, Figure S41).

Thermal stabilities of the “open” conjugate **8** and the “constrained” conjugate **2** were then evaluated by CD denaturation studies and compared with those obtained for the native HIV-PRO1 G4 structure (Table 1). The melting temperature (*T_m*) was collected for each G4 structure in the presence of either 100 mM KCl or 100 mM NaCl. As shown in Table 1, the “open” conjugate **8** was found to fold into a G4 structure that is as stable as the G4 structure formed by HIV-PRO1 in the presence of K⁺. Interestingly, whereas HIV-PRO1 does not adopt a G4 structure in the presence of NaCl, the G4 structure within **8** appears to be comparably stabilized by both sodium and potassium cations. On the other hand, a dramatic stabilization was observed for the G4 structure formed within constrained conjugate **2** in either Na⁺ or K⁺ solutions with a Δ*T_m* value of

+20 °C for the folded topology adopted in the presence of K⁺ cations. This strong stabilization is in agreement with previously reported values for the gain in stability provided by fixation on the molecular template.^[18] We also confirmed that the folding and unfolding processes were independent of the molecular concentration, thereby endorsing the prevalent formation of intramolecular G4 species (Supporting Information, Figures S35–S38). The formation of a resilient antiparallel G4 structure within conjugate **2** makes it an ideal model system to study the recognition of the HIV-PRO1 G4 conformation by natural ligands (i.e. G4-interacting proteins) and synthetic ligands.

SPR binding analysis

The identification of small-molecule ligands that selectively interact with G4 structures is essential for reliable *in vitro* biochemical and *in vivo* functional studies and might lead to new therapeutic approaches to treat human pathologies. In particular, the access to synthetic ligands that would be selective for viral G4 structures over cellular G4 structures would not merely contribute to provide a better understanding of the role of G4 in regulating pathogen processes but may also contribute to establishing novel antiviral strategies. Recent studies suggest that G4 structures formed within the HIV-1 genome may constitute promising novel targets for antiviral drug development.^[8] Indeed, the G4-forming sequences appear to be highly conserved among subspecies thereby suggesting functional significance. For instance, independent studies have shown that derivatives of porphyrin,^[15] acridine,^[14b] or naphthalene^[14a] could interact with HIV-G4 structures leading to the inhibition of key steps in the HIV-1 viral cell cycle.

To explore the propensity of G4-binding ligands to interact with the viral G4 structure formed by HIV-PRO1 and to assess the selectivity of recognition, we collected six low molecular weight compounds, archetypes of important families of G4 ligands, and we studied, by SPR, their interactions with HIV conjugate **2**, telomeric conjugate **1**, and a control hairpin structure **9** (Supporting Information, Figure S42) (Table 2). Template assembled DNA and RNA G4 model structures have been successfully exploited in the past, along with the hairpin structure **9**, to probe the interaction modes of several G4-binding ligands.^[18a,20a–d]

In the present study, we investigated the interactions of TMPyP4, MMQ1, BRACO-19, NiSal-3im, PDS, and PhenDC3 with

Table 2. Association kinetic constants (*k_{on}*), dissociation kinetic constants (*k_{off}*), and thermodynamic dissociation constants (*K_D*) for the interaction of ligands with G4 structures **2**, **1**, and hairpin structure **9**.

Ligands	HIV-G4 structure 2			Telomeric-G4 structure 1			Hairpin structure 9		
	<i>k_{on}</i> [M ⁻¹ s ⁻¹]	<i>k_{off}</i> [s ⁻¹]	<i>K_D</i> (M)	<i>k_{on}</i> [M ⁻¹ s ⁻¹]	<i>k_{off}</i> [s ⁻¹]	<i>K_D</i> [M]	<i>k_{on}</i> [M ⁻¹ s ⁻¹]	<i>k_{off}</i> [s ⁻¹]	<i>K_D</i> (M)
TMPyP4	4.5 ± 1.6 · 10 ⁵	8.7 ± 1.7 · 10 ⁻¹	2.1 ± 0.4 · 10 ⁻⁶	7.9 ± 1.0 · 10 ⁵	5.3 ± 0.7 · 10 ⁻¹	6.8 ± 0.5 · 10 ⁻⁷	7.4 ± 1.0 · 10 ⁵	5.3 ± 1.6 · 10 ⁻¹	7.0 ± 1.2 · 10 ⁻⁷
MMQ1	1.8 ± 0.4 · 10 ⁵	3.3 ± 1.1 · 10 ⁻¹	1.9 ± 0.8 · 10 ⁻⁶	1.4 ± 0.4 · 10 ⁵	5.4 ± 1.4 · 10 ⁻¹	4.5 ± 1.9 · 10 ⁻⁶	9.2 ± 3.3 · 10 ⁴	3.8 ± 0.6 · 10 ⁻¹	4.9 ± 1.3 · 10 ⁻⁶
NiSal-3im	4.4 ± 0.8 · 10 ⁵	2.9 ± 0.9 · 10 ⁻¹	6.3 ± 0.9 · 10 ⁻⁷	4.5 ± 0.8 · 10 ⁵	3.5 ± 1.0 · 10 ⁻¹	7.6 ± 0.8 · 10 ⁻⁷	2.7 ± 0.2 · 10 ⁵	7.6 ± 1.9 · 10 ⁻¹	2.9 ± 1.0 · 10 ⁻⁶
Braco	7.7 ± 0.2 · 10 ⁵	4.3 ± 0.9 · 10 ⁻²	5.5 ± 0.9 · 10 ⁻⁸	5.3 ± 0.2 · 10 ⁵	1.6 ± 0.3 · 10 ⁻¹	3.1 ± 0.5 · 10 ⁻⁷	3.6 ± 0.9 · 10 ⁵	4.2 ± 1.3 · 10 ⁻¹	1.2 ± 0.1 · 10 ⁻⁶
PDS	6.0 ± 0.0 · 10 ⁴	1.5 ± 0.4 · 10 ⁻³	2.4 ± 0.7 · 10 ⁻⁸	2.1 ± 1.2 · 10 ⁵	1.9 ± 0.3 · 10 ⁻³	1.2 ± 0.6 · 10 ⁻⁸	n.d. ^[a]	n.d. ^[a]	n.d. ^[a]
PhenDC3	2.6 ± 1.5 · 10 ⁷	5.3 ± 2.9 · 10 ⁻²	2.1 ± 0.1 · 10 ⁻⁹	4.3 ± 1.5 · 10 ⁷	9.6 ± 2.4 · 10 ⁻²	2.4 ± 0.3 · 10 ⁻⁹	n.d. ^[a]	n.d. ^[a]	n.d. ^[a]

[a] Not determinable as no signals were observed in the range of concentrations.

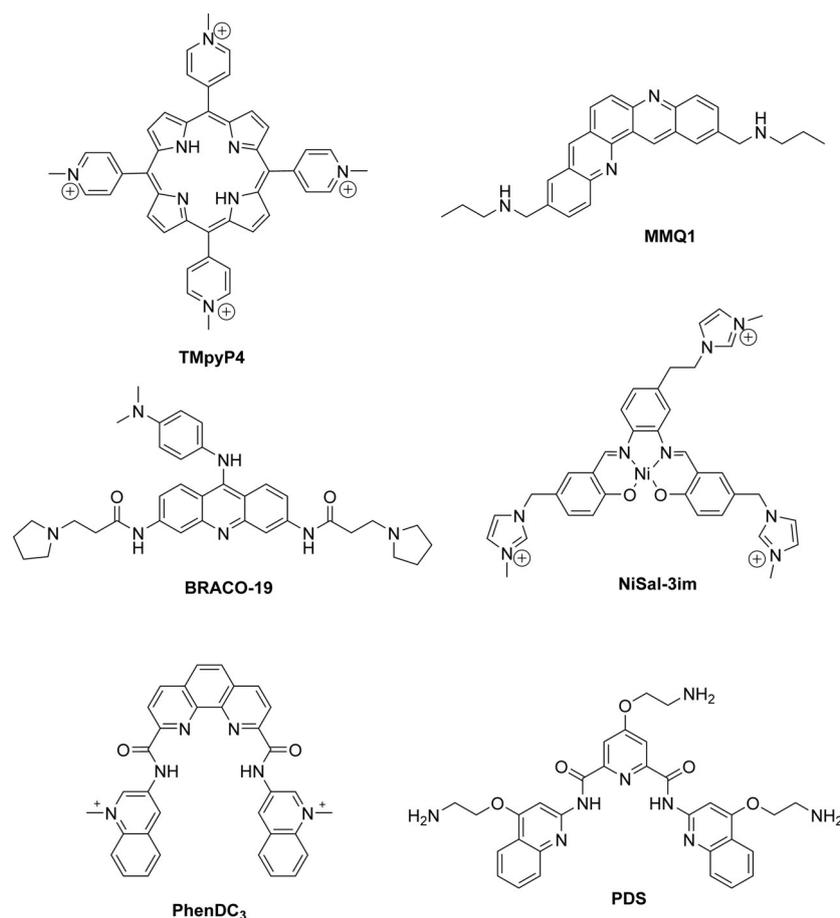


Figure 3. Structure of the G4 ligands subjected to SPR binding analysis.

1, **2**, and **9** (Figure 3). TMPyP4 is a tetracationic porphyrin that has been shown to interact with cellular and viral G4 structures.^[9a,28] MMQ1 is a quinacridine derivative that can stabilize the telomeric G4 structure thereby acting as telomerase inhibitor during cell assays.^[29] Braco-19 is a trisubstituted acridine derivative that has been shown to downregulate cancer cell proliferation through telomeric G4 binding.^[30] Braco-19 has also been used to affect HIV infected cells acting at both the reverse transcription and the post integration steps of the viral cycle.^[14b] NiSal-3im is a nickel(II) salophen complex that has been shown to associate preferably with DNA G4 structures over duplex DNA.^[20d] PDS is a strong stabilizer of G4 structures that can trigger the formation of DNA double strand breaks and to induce polymerase stalling at G4 sites.^[31] PhenDC3 is also a potent G4 stabilizer that has been used to probe the formation of G4 structures in mammalian cells.^[32] The interaction of PhenDC3 with a G4 structure formed within the C-Myc oncogene promoter has been solved by high-resolution NMR studies revealing an extended stacking interaction between the ligands and the external G-tetrad.^[33]

To perform the SPR binding analysis, conjugates **1**, **2**, and **9** were incubated onto streptavidin-coated gold surfaces until similar immobilization levels were obtained. Each ligand was injected simultaneously on all three channels for a direct comparison of recorded responses. The association/dissociation sig-

nals were recorded and the thermodynamic dissociation constant values (K_D) were calculated from the kinetic parameters of the interactions. The association/dissociation rates (k_{on}/k_{off}) and the dissociation constants are reported in Table 2. To get a clearer view of the kinetic nature of the interactions, the association/dissociation rates are plotted on a logarithmic scale on Figure 4.

In agreement with previous reports, we found that TMPyP4 and MMQ1 display little or no selectivity for G4 structures over the hairpin DNA. The thermodynamic dissociation constants of TMPyP4 and MMQ1 for conjugate **1** are in the low micromolar range, which is in good accordance with the reported value for the native telomeric G4 structure.^[20e] The affinity of TMPyP4 for the HIV-G4 structure **2** is somewhat lower (≈ 3 fold) than for the telomeric G4 structure **1**. Given that TMPyP4 possess a broad surface and can interact through π - π stacking with external G-tetrads, this difference might be rationalized by the increased loop rigidity of the HIV-G4 structure **2** within which a stably formed CG base pair may reduce the accessibility of the external tetrad. The similar K_D collected with MMQ1 for the recognition of **1** and **2** are consistent with the binding mode of MMQ1 which is also dominated by stacking interactions. All four other ligands were found to discriminate, to various extents, between G4 and hairpin structures with favorable thermodynamic dissociation constants for the first conformation.

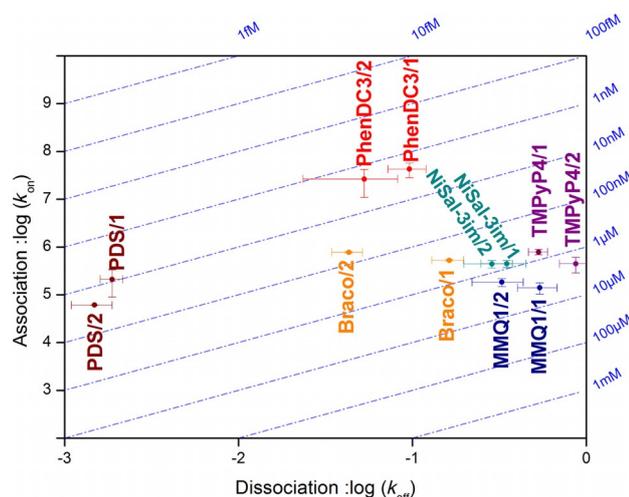


Figure 4. Isoaffinity plot and kinetic characterization as established by SPR. K_D (parallel diagonal lines) k_{on} (association kinetic constant, y axis), k_{off} (dissociation kinetic constant, x axis).

NiSal-3im comprises of a polyaromatic core functionalized with alkyimidazolium side chains that may engage in electrostatic interactions within G4 grooves and/or loops. Despite differences in tetrad accessibility and groove sizes, NiSal-3im displays identical dissociation constants of ≈ 700 nM for telomeric G4 structure **1** and HIV-G4 structure **2**. On the other hand, BRACO-19 was found to be the most selective compound. Its higher affinity for the HIV-G4 structure **2** (≈ 55 nM) versus the telomeric G4 structure (≈ 300 nM) reflects a slower dissociation rate when interacting with the **2**. The approximately sixfold increase in affinity might be rationalized by the occurrence of favorable interactions between the two flexible side chains of BRACO-19 and the grooves of the HIV-G4 structure **2**. The HIV-G4 structure **2** comprises two wide and two narrow grooves opposite to one another, whereas the telomeric G4 structure **1** features two medium grooves opposite to one another and a wide and narrow groove opposite to one another (Figure 1B). Even though only structural analysis would provide undisputable insights into the binding mechanism, we can hypothesize that the flexible side chains of BRACO-19 either better fit in the opposite wide grooves of the HIV-G4 structure or that they engage in more intricate interactions within the opposite narrow grooves of the HIV-G4 structure. Regardless of the binding mode, the good affinity and selectivity of BRACO-19 for a G4 structure derived from the LTR region of HIV-1, alongside with reported promising antiviral activities,^[14b] further consolidates the idea that this ligand and related analogues are worthy of further investigation.

Both PDS and PhenDC3 ligands are well-known for being potent G4-stabilizing agents and for being selective for G4 structures over duplex DNA. Both ligands comprise diversely substituted quinoline units connected through a pyridine (PDS) or a phenantrolin (PhenDC3) core and form crescent-shaped conformations that can interact with the G4 external tetrad. Using SPR binding analysis, we found that both PDS and Phen-DC3 display a high affinity for the HIV-G4 structures **2** with K_D in the low nanomolar range and no measurable affin-

ity for the hairpin structure **9**. These data suggest that both PDS and PhenDC3 might be exploited to probe the functional consequence of HIV-G4 targeting. Nanomolar K_D values were also obtained for the binding of PDS and PhenDC3 to the telomeric G4 structure **1**. PhenDC3, which exhibits a wider aromatic surface, was found to display the best affinities with dissociation constants of about 2 nM for both **1** and **2**. Even though PDS and PhenDC3 are not selective for a given G4 conformation, SPR binding analysis revealed remarkable aspects of their interactions with G4 structures. Indeed, the low K_D values for both PDS and PhenDC3 toward G4 conceal considerable disparities in the kinetics of association/dissociation, which suggests stark differences in their mode of interaction. As illustrated in Figure 4, PhenDC3 displays association rates that are faster ($\approx 10^7$ $\text{m}^{-1}\cdot\text{s}^{-1}$) than those of the other ligands by about two orders of magnitude. On the other hand, PDS displays very slow dissociation rates ($\approx 10^{-3}$ s^{-1}) that are ≈ 50 -fold lower than those of PhenDC3. PDS slow off-rates might be explained by the occurrence of stabilizing interactions between its protonable side chains and the negatively charged grooves of G4. Establishing such structure–kinetics relationships is of significance as molecular recognition (best described by the association rates) and complex stability (best described by the dissociation rates) are relevant predictors of the effectiveness of chemical probes and drugs.^[34] A fast association, such as that of PhenDC3, is an essential factor of efficacy as it may lead to significant therapeutic (drug) or functional (probe) concentrations in very short times. Even though there is not always a direct correlation between slow-off rates and long-lasting consequences of ligand binding, it has also been repeatedly advocated that slow-off rates were the most sought-after characteristic of probes and drugs. The ligand PDS has been successfully used for genome-^[3b] and transcriptome-wide^[3a] profiling of DNA and RNA G4 structures. PDS ligand and a close analogue have also been used to promote the visualization of the DNA G4 structure in human cells^[2d] and to mediate the isolation of G4-forming sequences from human cells.^[35] One may suggest that these experimental successes, leading to those notable insights, have been achieved thanks to the remarkable capacity of PDS to establish long-lasting interactions with G4 structures.

Collectively those binding data suggest that readily available synthetic ligands, that have been shown to interact with cellular G4 structures, might be exploited to investigate HIV infection pathways. Furthermore, even though we observed a high degree of promiscuity for the interactions between the ligands and the two distinct G4 structures, we could identify one ligand, BRACO-19, with the ability to distinguish between the distinct, yet topologically related, antiparallel G4 structures **1** and **2**.

Conclusions

A synthetic biomolecular engineering approach has been developed to stabilize a nucleic acid foldamer that mimics a G4 structure formed by the G-rich LTR region of HIV-1. We herein demonstrate that, under optimized conditions, OL, TC, and

CuAAC reactions can be sequentially exploited to assemble a sophisticated G4-forming bioconjugate. The requisite to optimize and amend well-documented chemical reactions (TC and CuAAC) in the context of the synthesis of **8** and **2** further emphasizes the higher requirements associated to the assembly of structural foldamers over the synthesis of more conventional bioconjugates. TDS, CD, and NMR biophysical studies support the idea that the resulting conjugate **2** folds into the expected antiparallel chair-type G4 topology. As evidenced by CD thermal denaturation studies, the G4 conformation adopted by **2** is highly stable. Furthermore, **2** remains folded in a broader range of experimental conditions thereby enabling the use of this conjugate for reliable interaction studies. The identification of HIV-G4-binding ligands is of particular interest as selective targeting of G4 structures derived from highly conserved viral sequences might lead to novel, efficient and specific antiviral treatments. Using SPR binding analysis we confirmed that synthetic ligands, such as PDS, PhenDC3, and BRACO-19, display high affinities for the HIV-G4 structure thereby supporting their use to probe HIV pathways. Whereas most ligands were not able to discriminate between the HIV-G4 structure and an antiparallel chair-type G4 derived from the human telomeric DNA sequence, we provide evidence that the acridine derivative BRACO-19 exhibits some selectivity toward the HIV-G4 structure. This observation demonstrates that a structure-specific ligand may indeed arise from the design or selection of molecules that would engage in discrete interactions with differential G4 recognition elements. In the course of this study, we also unveiled remarkable features of the interactions between G4 structures and broadly used binding probes such as PDS ligand. As those observations further endorse the use of PDS as a high quality probe for G4 biological research, they also stress-out the importance of meticulous assessment of binding parameters.

Beyond its use as model target for the identification of a structure-specific synthetic ligand, we anticipate that the stable HIV-G4 conjugate assembled in this study might be exploited to isolate and identify G4-specific protein binders thereby providing new insights into viral regulation pathways. We also anticipate that the methodological developments depicted in this study could be readily exploited to assemble well-ordered bioconjugates that would fold into other relevant viral G4 structures. This would be of particular interest since the formation of distinct viral G4 structures has been associated to distinct functional consequences in latency, virulence, or mutation rates. The newly developed synthetic methods may also find applications in the development of a broader range of bioconjugates and foldamers including stable mimics of other biologically relevant nucleic acid structures such as triplexes and i-motifs. Efforts in those directions are on-going and will be reported in due course.

Experimental Section

General

ESI mass spectra were performed using an Esquire 3000 spectrometer from Bruker or an Acquity UPLC/MS system from Waters equipped with an SQ Detector 2. MALDI-TOF mass spectra were performed using an Autoflex Bruker with hydropiccolinic acid (HPA) (45 mg, ammonium citrate (4 mg) in 500 mL H₂O/CH₃CN) as the matrix. Peptides were analyzed in the positive mode and oligonucleotides and conjugates in the negative mode. ¹H NMR experiments were carried out with a Bruker AVANCE III 500 MHz spectrometer equipped with a cryo-probe Prodigy. Data were processed with TopSpin software (Bruker). All solvents and reagents used were of the highest purity commercially available. Phen-DC3 was kindly provided by M.P. Teulade-Fichou and colleagues.

Peptide synthesis

General details: The reactions were monitored using an UPLC-MS system (waters) with reverse-phase chromatography using a Nucleosil C18 column (130 Å, 2.1 × 50 mm, 1.7 μm) and combined electron spray ionization mass spectrometry. A 1 mL min⁻¹ flow linear gradient from 95% solvent A (0.1% formic acid in water) and 5% solvent B (0.1% formic acid in acetonitrile/water: 9:1) to 100% B for 4 min was applied. RP-HPLC purifications were performed on a Gilson system with a Nucleosil C18 column (100 Å, 250 × 21 mm, 7 μm) with UV monitoring at 214 and 250 nm. A 20 mL min⁻¹ flow linear gradient was applied from 95% solvent A¹ (0.1% trifluoroacetic acid in water) and 5% solvent B¹ (0.1% trifluoroacetic acid in acetonitrile/water: 9:1) to 100% B¹ for 20 min.

Synthesis of peptide 6: The linear peptide was synthesized using 2-chlorotrityl (loading of 0.83 mmol g⁻¹) commercial resin. Fmoc-Gly-OH (3 equiv) was coupled on the resin in anhydrous DCM in a glass reaction vessel fitted with a sintered glass. pH was adjusted to 8 by using DIEA. The mixture was stirred for 1 h at room temperature. A DCM/MeOH/DIEA solution (17:2:1) was then added. A Fmoc protecting group was removed using three washes with 20% piperidine solution in DMF (40 mL). The resin loading was assessed by quantification of free dibenzofulvene using UV absorbance at 299 nm (loading of 0.40 mmol g⁻¹, yield: 48%). The elongation was performed on a Syro II using Fmoc/tBu strategy on Fmoc-Gly-2-chlorotrityl resin (462 mg). Fmoc-Ala-OH, Fmoc-Gly-OH, Fmoc-Lys(Boc)-OH, Fmoc-Pro-OH, Fmoc-Lys(Dde)-OH, and Fmoc-Lys(biotin)-OH were commercially available. Fmoc-aminonorleucine was synthesized according to the reported protocol.^[36] The elongated peptide was then cleaved from the resin by using 1% TFA solution in DCM (5 × 20 mL). The resulting solution was evaporated under vacuum and the peptide was precipitated in ether to give a white powder. The crude product was used without any further purification in the next step. The cyclization of the linear peptide (325 mg) was performed in DMF at 10⁻³ M concentration in the presence of PyBOP (1.2 equiv, 0.250 mmol, 130 mg). The pH was adjusted to 8–9 by using DIEA and the solution was stirred at room temperature until complete peptide cyclization occurred (UPLC monitoring). The solvent was evaporated under vacuum then the crude peptide was precipitated in ether and purified on RP-HPLC and freeze-dried to give a white powder (0.090 mmol, 140 mg, yield: 50%). HRMS (+): *m/z*: calcd for C₇₂H₁₁₅N₂₁NaO₁₆S: 1584.8; found: 1584.8 [M+Na]⁺. The Dde protecting group was then removed using a 2% hydrazine solution in DMF. The mixture was stirred at room temperature until complete deprotection (UPLC monitoring). The solvent was evaporated under vacuum then the crude peptide was precipitated in ether to give a white

powder that was used without any further purification. The yield was quantitative. The peptide (90 mg) was next functionalized using 2 equiv of chloroacetic anhydride (0.13 mmol, 22 mg) in DMF (3.10^{-3} M) at pH 8–9. The reaction was stirred for 1 h 30 min at room temperature. The solvent was evaporated under vacuum and then the crude peptide was precipitated in ether. The crude product was purified on RP-HPLC and freeze-dried to give a white powder (0.026 mmol, 39 mg, yield: 41% for two steps). HRMS (+): m/z : calcd for $C_{64}H_{104}ClN_{21}NaO_{15}S$: 1496.7; found: 1496.7 [$M+Na$]⁺. The Boc protection on the lysine side chain was then cleaved through treatment with a DCM/TFA/H₂O (45:45:10) solution (20 mL) over 1 h at room temperature. After evaporation of the solvent under vacuum and precipitation in ether, the crude product was obtained as a white powder in quantitative yield and used in the following step without any further purification. The resulting peptide was next functionalized with an aminoxy moiety by using 1.2 equiv of *N*-Boc-aminoxyacetic acid *N*-hydroxysuccinimide ester in anhydrous DMF (25 mL). The pH was adjusted to 8–9 with DIEA and the reaction was stirred at room temperature and monitored by UPLC-MS until the completion of the reaction. Solvent evaporation and ether precipitation afforded the crude product which was purified on RP-HPLC and freeze-dried to give a white powder (0.019 mmol, 29 mg, yield: 73% for two steps). HRMS (+): m/z : calcd for $C_{66}H_{107}ClN_{22}NaO_{17}S$: 1569.7; found: 1569.8 [$M+Na$]⁺. Peptide **6** was obtained after deprotection by using a TFA/H₂O/TIS (80:16:2) solution (4 mL) for 30 min at room temperature. The solvent was evaporated under vacuum then the crude peptide was precipitated in ether to give a white powder. The crude product was used without further purification. The yield was considered as quantitative.

Oligonucleotide synthesis

General details: The oligonucleotides were prepared using β -cyanoethylphosphoramidite chemistry on a 3400 DNA synthesizer at 1 μ mol scale. RP-HPLC analyses were performed on a Waters HPLC system using C18 Nucleosil column (Macherey–Nagel, 100 Å, 250 \times 4.6 mm, 5 μ m) with UV-monitoring at 260 and 280 nm. A 1 mL min⁻¹ flow linear gradient was applied. Solvent A² (50 mM triethylammonium acetate buffer with 5% acetonitrile) and solvent B² (acetonitrile with 5% water) were used. A stepwise gradient of 0–30% B² in 20 min followed by 30 to 100% B² in 10 min was applied for the gradient I. A stepwise gradient of 0–60% B² in 20 min followed by 60 to 100% B² in 10 min was applied for the gradient II. The RP-HPLC purifications of oligonucleotides were performed on a Gilson system with Nucleosil C-18 column (Macherey–Nagel, 100 Å, 250 \times 10 mm, 7 μ m) with UV-monitoring at 260 and 280 nm using a 4 mL min⁻¹ flow linear gradient. A stepwise gradient of 0–40% B² in 20 min was applied for the gradient III. A stepwise gradient of 0–80% B² in 20 min was applied for the gradient IV

Synthesis of oligonucleotide 3: Oligonucleotide **3** was prepared according to a previously reported procedure.^[18b] ESI-MS(–) m/z : calcd for $C_{105}H_{134}N_{38}O_{67}P_{11}$: 3341.1; found: 3340.9 [$M-H$][–].

Synthesis of oligonucleotide 4: The oligonucleotide 5' *S*-trityl, 3' amino, precursor of oligonucleotide **4**, was obtained by automated DNA synthesis using commercially available 3'-PT-amino-modifier C3 CPG resin and 5'-thiol-modifier C6 at 1 μ mol scale. Cleavage from the resin and deprotection were performed in ammonia/methylamine (1:1) solution for 30 min at 65 °C. The released product was purified on RP-HPLC (gradient IV) and then desalted on a NAP-25 column and freeze-dried (410 nmol, yield: 41%). ESI-MS(–): m/z : calcd for $C_{117}H_{144}N_{41}O_{58}P_{10}S$: 3394.4; found: 3394.3 [$M-H$][–]. Then 4-pentynoic acid succinimidyl ester (30 equiv,

10.8 μ mol, 2.1 mg) was added to a solution of intermediate 5' *S*-trityl,3' amino oligonucleotide (1 equiv, 360 nmol) in H₂O/DMF (9:1) solution to reach a final concentration of 10^{-3} M. The pH was adjusted to 8–9 using DIEA. The solution was stirred for 4 h at room temperature and the crude reaction mixture was desalted on a NAP-25 column. The coupling was considered quantitative and the desalted crude was analyzed on RP-HPLC and used in the next step without further purification. ESI-MS(–): m/z : calcd for $C_{122}H_{148}N_{41}O_{59}P_{10}S$: 3474.5; found: 3474.2 [$M-H$][–]. Final deprotection of the 5' *S*-trityl group was achieved using 1 M aqueous silver nitrate solution (0.15 vol) and acetic acid (0.5 vol) in 0.1 M triethylammonium acetate buffer (pH 6.5, concentration 10^{-3} M). The mixture was stirred for 45 min at room temperature. 1 M aqueous dithiothreitol (DTT) solution (0.20 vol) was then added and the mixture was stirred for 15 min at room temperature. After centrifugation, the supernatant was conserved and the pellet was washed with 0.1 M triethylammonium acetate buffer solution (pH 6.5). The solution was then desalted on NAP 25. The coupling was considered quantitative and the crude oligonucleotide **4** was analyzed on RP-HPLC and used in the next step without further purification. ESI-MS(–): m/z : calcd for $C_{103}H_{134}N_{41}O_{59}P_{10}S$: 3232.2; found: 3231.9 [$M-H$][–].

Conjugate synthesis

Synthesis of conjugate 7 (oxime ligation): Oligonucleotide **3** (1 equiv, 1.3 μ mol) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5, concentration 10^{-3} M) and aminoxy peptide **6** (1.2 equiv, 1.56 μ mol) was added. The solution was stirred at 55 °C for 45 min. The crude mixture was purified on RP-HPLC (gradient III). Conjugate **7** was desalted on a NAP-25 column and freeze dried (1.1 μ mol, yield 84%). ESI-MS(–): m/z : calcd for $C_{166}H_{231}N_{60}O_{81}P_{11}SCI$: 4771.2; found: 4770.8 [$M-H$][–].

Synthesis of conjugate 8 (thioether formation): Conjugate **7** (1 equiv 50 nmol) and oligonucleotide **4** (2 equiv, 100 nmol) were dissolved in a solution of H₂O/CH₃CN (9:1) at 5.10^{-4} M. TCEP (2 equiv), 500 mM KCl, DIEA (45 equiv), and KI (60 equiv) were added to this solution. The reaction was stirred at room temperature for 5 h. The crude was purified on RP-HPLC (gradient III) and then conjugate **8** was desalted on a NAP-25 column and freeze dried (35 nmol, yield 70%). ESI-MS(–): m/z : calcd for $C_{269}H_{365}N_{101}O_{140}P_{21}S_2$: 7967.9; found: 7968.4 [$M-H$][–].

Synthesis of conjugate 2 (CuAAC reactions): Conjugate **8** (100 nmol) was dissolved in 100 mM HEPES buffer (pH 7.4) at 10^{-4} M. KCl (100 mM), CuSO₄ (6 equiv), THPTA (30 equiv), and sodium ascorbate (30 equiv) were added to this solution. The reaction was stirred at 37 °C for 2 h and stopped with the addition of 0.5 M ethylenediaminetetraacetic acid (EDTA) solution (50 equiv). The resulting reaction was desalted using a NAP-25 column. The crude was purified on RP-HPLC (gradient III) then conjugate **2** was desalted on a NAP-25 column and freeze dried (79.4 nmol, yield 79%). ESI-MS(–): m/z : calcd for $C_{269}H_{365}N_{101}O_{140}P_{21}S_2$: 7967.9; found: 7968.5 [$M-H$][–].

Circular dichroism studies

Prior to CD analysis, compounds **2** and **8** were thoroughly desalted by SEC then an annealing step was applied by heating the sample at 80 °C for 5 min in buffer (Tris 10 mM pH 7.4 with 100 mM NaCl or with 100 mM KCl) and cooling it over 2 h to room temperature. Analyses were recorded on a Jasco J-810 spectropolarimeter using 1 cm length quartz cuvette. Spectra were recorded at 20 °C or every 5 °C in a range of 5 to 90 °C with a wavelength range of 220 to 320 nm. For each temperature, the spectrum was an average of

three scans with a 0.5 s response time, a 1 nm data pitch, a 4 nm bandwidth and a 200 nm min⁻¹ scanning speed. Melting temperatures were obtained using Boltzmann fit on Origin software. Each curve fit was only accepted with a $r_{\text{value}} > 0.99$.

NMR experiments

NMR spectra of conjugate **2** and HIV-PRO1 were carried out at 80 μM and at 25 °C in sodium phosphate buffer or potassium phosphate buffer (20 mM, pH 7.4) and in the presence of 10% D₂O and 70 mM NaCl or 70 mM KCl. The program uses p11gpse water suppression using 1-1 echo pulse sequence with spoil gradient pulse. The gradient file name is SMSQ1.100 that uses a 20% gradient ratio for 600 μs gradient pulse. The delay for binomial water suppression is $D19 = 60 \mu\text{s}$ ($D19 = 1/(2*d)$ in which d = distance of next null (in Hz)). The experiments were recorded with 40960 data points. 7168 scans were acquired over a spectral width of 10000 Hz.

SPR experiments

SPR data were obtained on a BIAcore T200 (GE Healthcare). Measurements were performed on gold surfaces. Sensor chips were cleaned by UV-ozone treatment during 10 min after being rinsed with MilliQ water and ethanol. The cleaned gold surfaces were then functionalized according to the following procedure. Firstly, mixed self-assembled monolayers (SAMs) were formed at room temperature by dipping overnight gold sensors in the thiol mixture: 80% HS-(CH₂)₁₁-EG4-OH and 20% HS-(CH₂)₁₁-EG6-Biotin (1 mM total thiol concentration in EtOH). After overnight adsorption, gold sensors were rinsed with ethanol, dried under nitrogen, and then inserted in the BIAcore T200 apparatus. All measurements were performed at 25 °C, using a running buffer (R.B.) composed of HEPES buffered saline: 0.01 M HEPES pH 7.4, 50 mM NaCl, 100 mM KCl, and 0.5% v/v surfactant P20. Streptavidin (100 ng mL⁻¹) was injected (10 $\mu\text{L min}^{-1}$) on the biotinylated SAM until saturation of the surface (around 2500 R.U.). The conjugates **1**, **2**, and **9** were injected at 2 $\mu\text{L min}^{-1}$ on the streptavidin-coated SAM surfaces until saturation (a final response ca 300 RU). Binding experiments were conducted at 30 $\mu\text{L min}^{-1}$ (or 80 $\mu\text{L min}^{-1}$ for PhenDC3 and PDS) by injection of the different ligands (dissolved in R.B., injection time: 240 s, stabilization time: 600 s) at different concentrations using a single-cycle kinetic method. These concentrations are: 0.1, 0.25, 1, 3, 10 μM for MMQ1; 1, 5, 10, 25, 50 nM for PhenDC3; 10, 25, 50, 100, 250 nM for PDS; 15, 50, 200, 600, 3000 nM for BRACO-19; and 75, 200, 400, 1000, 3000 nM for TMPyP4, NiSal-3im. A streptavidin surface, prepared as described above, was used as a reference. Curves obtained on the reference surface were deduced from the curves recorded on the recognition one, allowing elimination of refractive index changes due to buffer effects. The binding rate constants of G-quadruplex DNA/ligands interactions were calculated by a nonlinear analysis of the association and dissociation curves using the SPR kinetic evaluation software of BIAcore T200 (evaluation Software 2.0.1). The data were fitted using a heterogeneous ligand model (two interactions per G4 target, see the Supporting Information). The association rate constants, $k_{\text{on}1}$ and $k_{\text{on}2}$, and the dissociation rate constants, $k_{\text{off}1}$ and $k_{\text{off}2}$, as well as the experimental maximal response, $R_{\text{max}1}$ and $R_{\text{max}2}$, of the two interactions were determined. The equilibrium dissociation constants were obtained from the binding rate constants as $K_{D1} = k_{\text{off}1}/k_{\text{on}1}$ and $K_{D2} = k_{\text{off}2}/k_{\text{on}2}$. We chose not to report values for the second interaction which may involve nonstoichiometric binding and/or nonspecific interactions (see the Supporting Information). The reported values are the mean of representative independent experiments and the errors

provided are standard deviations from the mean. Each SPR experiment was repeated at least three times. Mean square errors χ^2 and experimental maximum responses (R_{max}) available in the Supporting Information are taken from the best fits for each ligand.

Acknowledgements

This work was supported by "Agence National de la Recherche" (ANR-12-BSV8-0008-04 and ANR-16-CE11-0006-01), Labex ARCAN (ANR-11-LABX-0003-01), and AGIR program from Univ. Grenoble Alpes. The NanoBio-ICMG platforms (FR 2607) are acknowledged for their support. The authors thank Dr. S. Amrane for helpful suggestions regarding NMR data collection.

Conflict of interest

The authors declare no conflict of interest.

Keywords: click chemistry · G-quadruplexes · G-quadruplex ligands · HIV · nucleic acid conjugates

- [1] A. I. Karsiotis, C. O'Kane, M. Webba da Silva, *Methods* **2013**, *64*, 28–35.
- [2] a) A. Laguerre, J. M. Wong, D. Monchaud, *Sci. Rep.* **2016**, *6*, 32141; b) A. Laguerre, K. Hukezalie, P. Winckler, F. Katranji, G. Chanteloup, M. Pirrotta, J. M. Perrier-Cornet, J. M. Wong, D. Monchaud, *J. Am. Chem. Soc.* **2015**, *137*, 8521–8525; c) G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, *Nat. Chem.* **2013**, *6*, 75–80; d) G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nat. Chem.* **2013**, *5*, 182–186.
- [3] a) C. K. Kwok, G. Marsico, A. B. Sahakyan, V. S. Chambers, S. Balasubramanian, *Nat. Methods* **2016**, *13*, 841–844; b) V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, S. Balasubramanian, *Nat. Biotechnol.* **2015**, *33*, 877–881.
- [4] A. Bedrat, L. Lacroix, J. L. Mergny, *Nucleic Acids Res.* **2016**, *44*, 1746–1759.
- [5] S. Müller, R. Rodriguez, *Expert Rev. Clin. Pharm.* **2014**, *7*, 663–679.
- [6] D. Bhartiya, V. Chawla, S. Ghosh, R. Shankar, N. Kumar, *Genomics* **2016**, *108*, 224–231.
- [7] I. T. Holder, J. S. Hartig, *Chem. Biol.* **2014**, *21*, 1511–1521.
- [8] M. Metifiot, S. Amrane, S. Litvak, M. L. Andreola, *Nucleic Acids Res.* **2014**, *42*, 12352–12366.
- [9] a) R. Perrone, M. Nadai, J. A. Poe, I. Frasson, M. Palumbo, G. Palu, T. E. Smithgall, S. N. Richter, *PLoS ONE* **2013**, *8*, e73121; b) R. Perrone, M. Nadai, I. Frasson, J. A. Poe, E. Butovskaya, T. E. Smithgall, M. Palumbo, G. Palu, S. N. Richter, *J. Med. Chem.* **2013**, *56*, 6521–6530.
- [10] K. Tluckova, M. Marusic, P. Tothova, L. Bauer, P. Sket, J. Plavec, V. Viglasky, *Biochemistry* **2013**, *52*, 7207–7216.
- [11] a) B. Biswas, M. Kandpal, U. K. Jauhari, P. Vivekanandan, *BMC Genomics* **2016**, *17*, 949; b) S. Artusi, R. Perrone, S. Lago, P. Raffa, E. Di Iorio, G. Palu, S. N. Richter, *Nucleic Acids Res.* **2016**, *44*, 10343–10353; c) S. Artusi, M. Nadai, R. Perrone, M. A. Biasolo, G. Palu, L. Flamand, A. Calistri, S. N. Richter, *Antiviral Res.* **2015**, *118*, 123–131; d) J. T. Tellam, J. Zhong, L. Le-kieffre, P. Bhat, M. Martinez, N. P. Croft, W. Kaplan, R. L. Tellam, R. Khanna, *PLoS Pathog.* **2014**, *10*, e1004423; e) P. Murat, J. Zhong, L. Le-kieffre, N. P. Cowieson, J. L. Clancy, T. Preiss, S. Balasubramanian, R. Khanna, J. Tellam, *Nat. Chem. Biol.* **2014**, *10*, 358–364; f) E. K. Lee, S. Y. Kim, K. W. Noh, E. H. Joo, B. Zhao, E. Kieff, M. S. Kang, *Antiviral Res.* **2014**, *104*, 73–83.
- [12] S. R. Wang, Q. Y. Zhang, J. Q. Wang, X. Y. Ge, Y. Y. Song, Y. F. Wang, X. D. Li, B. S. Fu, G. H. Xu, B. Shu, P. Gong, B. Zhang, T. Tian, X. Zhou, *Cell. Chem. Biol.* **2016**, *23*, 1113–1122.
- [13] A. M. Fleming, Y. Ding, A. Alenko, C. J. Burrows, *ACS Infect. Dis.* **2016**, *2*, 674–681.
- [14] a) R. Perrone, F. Doria, E. Butovskaya, I. Frasson, S. Botti, M. Scalabrini, S. Lago, V. Grande, M. Nadai, M. Freccero, S. N. Richter, *J. Med. Chem.*

- 2015, 58, 9639–9652; b) R. Perrone, E. Butovskaya, D. Daelemans, G. Palu, C. Pannecouque, S. N. Richter, *J. Antimicrob. Chemother.* **2014**, 69, 3248–3258.
- [15] L. M. Harris, C. J. Merrick, *PLoS Pathog.* **2015**, 11, e1004562.
- [16] D. Piekna-Przybylska, M. A. Sullivan, G. Sharma, R. A. Bambara, *Biochemistry* **2014**, 53, 2581–2593.
- [17] a) E. Largy, A. Marchand, S. Amrane, V. Gabelica, J. L. Mergny, *J. Am. Chem. Soc.* **2016**, 138, 2780–2792; b) M. Boncina, G. Vesnaver, J. B. Chaires, J. Lah, *Angew. Chem. Int. Ed.* **2016**, 55, 10340–10344; *Angew. Chem.* **2016**, 128, 10496–10500.
- [18] a) L. Bonnat, J. Dejeu, H. Bonnet, B. Gennaro, O. Jarjayes, F. Thomas, T. Lavergne, E. Defrancq, *Chem. Eur. J.* **2016**, 22, 3139–3147; b) R. Bonnet, T. Lavergne, B. Gennaro, N. Spinelli, E. Defrancq, *Chem. Commun.* **2015**, 51, 4850–4853; c) P. Murat, D. Cressend, N. Spinelli, A. Van der Heyden, P. Labbe, P. Dumy, E. Defrancq, *ChemBioChem* **2008**, 9, 2588–2591.
- [19] a) V. Romanucci, M. Gaglione, A. Messere, N. Potenza, A. Zarelli, S. Noppen, S. Liekens, J. Balzarini, G. Di Fabio, *Eur. J. Med. Chem.* **2015**, 89, 51–58; b) M. Nikan, B. O. Patrick, J. C. Sherman, *ChemBioChem* **2012**, 13, 1413–1415; c) R. Ferreira, M. Alvira, A. Avino, I. Gomez-Pinto, C. Gonzalez, V. Gabelica, R. Eritja, *ChemistryOpen* **2012**, 1, 106–114; d) G. Oliviero, N. Borbone, J. Amato, S. D'Errico, A. Galeone, G. Piccialli, M. Varra, L. Mayol, *Biopolymers* **2009**, 91, 466–477.
- [20] a) L. Lecarme, E. Prado, A. De Rache, M. L. Nicolau-Travers, G. Gellon, J. Dejeu, T. Lavergne, H. Jamet, D. Gomez, J. L. Mergny, E. Defrancq, O. Jarjayes, F. Thomas, *ChemMedChem* **2016**, 11, 1133–1136; b) J. Dejeu, T. Lavergne, J. D. Nora, E. Defrancq, G. Pratviel, *Inorg. Chim. Acta* **2016**, 452, 98–103; c) L. Sabater, M. L. Nicolau-Travers, A. De Rache, E. Prado, J. Dejeu, O. Bombarde, J. Lacroix, P. Calsou, E. Defrancq, J. L. Mergny, D. Gomez, G. Pratviel, *J. Biol. Inorg. Chem.* **2015**, 20, 729–738; d) L. Lecarme, E. Prado, A. De Rache, M. L. Nicolau-Travers, R. Bonnet, A. van Der Heyden, C. Philouze, D. Gomez, J. L. Mergny, H. Jamet, E. Defrancq, O. Jarjayes, F. Thomas, *Inorg. Chem.* **2014**, 53, 12519–12531; e) P. Murat, R. Bonnet, A. Van der Heyden, N. Spinelli, P. Labbe, D. Monchaud, M. P. Teulade-Fichou, P. Dumy, E. Defrancq, *Chem. Eur. J.* **2010**, 16, 6106–6114.
- [21] E. Garanger, D. Boturyn, O. Renaudet, E. Defrancq, P. Dumy, *J. Org. Chem.* **2006**, 71, 2402–2410.
- [22] S. Amrane, A. Kerkour, A. Bedrat, B. Vialet, M. L. Andreola, J. L. Mergny, *J. Am. Chem. Soc.* **2014**, 136, 5249–5252.
- [23] a) B. Thomas, M. Fiore, G. C. Daskhan, N. Spinelli, O. Renaudet, *Chem. Commun.* **2015**, 51, 5436–5439; b) C. Ligeour, L. Dupin, A. Marra, G. Vergoten, A. Meyer, A. Dondoni, E. Souteyrand, J.-J. Vasseur, Y. Chevolut, F. Morvan, *Eur. J. Org. Chem.* **2014**, 7621–7630; c) G. Viault, S. Dautrey, N. Maindron, J. Hardouin, P. Y. Renard, A. Romieu, *Org. Biomol. Chem.* **2013**, 11, 2693–2705; d) T. Joshi, M. Patra, L. Spiccia, G. Gasser, *Artificial DNA, PNA & XNA* **2013**, 4, 11–18; *XNA* **2013**, 4, 11–18.
- [24] M. Galibert, P. Dumy, D. Boturyn, *Angew. Chem. Int. Ed.* **2009**, 48, 2576–2579; *Angew. Chem.* **2009**, 121, 2614–2617.
- [25] T. Springer, H. Sipova, H. Vaisocherova, J. Stepanek, J. Homola, *Nucleic Acids Res.* **2010**, 38, 7343–7351.
- [26] J. L. Mergny, J. Li, L. Lacroix, S. Amrane, J. B. Chaires, *Nucleic Acids Res.* **2005**, 33, e138.
- [27] A. Randazzo, G. P. Spada, M. W. da Silva, *Top. Curr. Chem.* **2012**, 330, 67–86.
- [28] L. Martino, B. Pagano, I. Fotticchia, S. Neidle, C. Giancola, *J. Phys. Chem. B* **2009**, 113, 14779–14786.
- [29] C. Hounsou, L. Guittat, D. Monchaud, M. Jourdan, N. Saettel, J. L. Mergny, M. P. Teulade-Fichou, *ChemMedChem* **2007**, 2, 655–666.
- [30] C. M. Incles, C. M. Schultes, H. Kempfski, H. Koehler, L. R. Kelland, S. Neidle, *Mol. Cancer Ther.* **2004**, 3, 1201–1206.
- [31] a) S. Müller, D. A. Sanders, M. Di Antonio, S. Matsis, J. F. Riou, R. Rodriguez, S. Balasubramanian, *Org. Biomol. Chem.* **2012**, 10, 6537–6546; b) R. Rodriguez, K. M. Miller, J. V. Forment, C. R. Bradshaw, M. Nikan, S. Britton, T. Oelschlaegel, B. Xhemalce, S. Balasubramanian, S. P. Jackson, *Nat. Chem. Biol.* **2012**, 8, 301–310; c) R. Rodriguez, S. Muller, J. A. Yeoman, C. Trentesaux, J. F. Riou, S. Balasubramanian, *J. Am. Chem. Soc.* **2008**, 130, 15758–15759.
- [32] A. De Cian, E. Delemos, J. L. Mergny, M. P. Teulade-Fichou, D. Monchaud, *J. Am. Chem. Soc.* **2007**, 129, 1856–1857.
- [33] W. J. Chung, B. Heddi, F. Hamon, M. P. Teulade-Fichou, A. T. Phan, *Angew. Chem. Int. Ed.* **2014**, 53, 999–1002; *Angew. Chem.* **2014**, 126, 1017–1020.
- [34] D. Guo, L. H. Heitman, I. J. AP, *ChemMedChem* **2015**, 10, 1793–1796.
- [35] S. Müller, S. Kumari, R. Rodriguez, S. Balasubramanian, *Nat. Chem.* **2010**, 2, 1095–1098.
- [36] E. D. Goddard-Borger, R. V. Stick, *Org. Lett.* **2007**, 9, 3797–3800.

 Manuscript received: January 27, 2017

Accepted Article published: March 6, 2017

Final Article published: April 5, 2017

**Synthèse et étude d'ADN et d'ARN G-quadruplexes à topologies contrôlées.
Applications pour la caractérisation et la sélection de ligands.**

Les acides nucléiques riches en guanines ou en cytosines peuvent se replier sur eux-mêmes et former des systèmes tétramériques tels que les G-quadruplexes (G4) ou les i-motifs. Ces motifs, abondamment représentés dans certaines régions du génome humain semblent contribuer à la régulation cellulaire et suscitent depuis plusieurs années un intérêt grandissant. Ils sont notamment présents dans la région télomérique, mais aussi dans les promoteurs d'oncogènes ou au sein des génomes viraux et sont impliqués dans certaines pathologies humaines. Ils représentent ainsi des cibles thérapeutiques et diagnostiques potentielles. Cependant, les G4 adoptent *in-vitro* des topologies variées qui compliquent le développement de ligands spécifiques et affins. Dans ce contexte, le laboratoire a développé le concept du TASQ pour "Template Assembled Synthetic G-Quadruplex" dans le but d'accéder à des G4 se structurant en une topologie définie.

Le premier chapitre décrit l'assemblage de mimes de motifs G4 contraints en une topologie unique. En utilisant un gabarit cyclodécapeptide rigide et différentes méthodes de conjugaison, nous avons assemblé des motifs G4 ARN parallèle et hybride ADN/ARN dérivant de la séquence télomérique ainsi qu'un motif G4 d'ADN présent dans la séquence promotrice du VIH-1. L'utilisation du concept TASQ nous a également permis de préparer un motif G-triplexe (G3), intermédiaire à la formation des motifs G4. Nous avons montré une forte stabilisation de tous les édifices G4 contraints ainsi préparés.

Le second chapitre concerne les études de caractérisation et de sélection de ligands vis-à-vis des motifs G4 et G3 contraints. La caractérisation repose sur l'évaluation de l'affinité et de la sélectivité de différentes familles de ligands pour ces édifices, par résonance plasmonique de surface ou par interférométrie bio-couche. La sélection de ligands a été réalisée par la méthode SELEX dans le but d'obtenir des aptamères affins et spécifiques d'un motif G4 contraint.

Mots clés : G-quadruplexe, G-triplexe, i-motif, ligations chimiosélective, gabarit cyclodécapeptidique, SPR, SELEX

**Synthesis and study of topologically controlled DNA and RNA G-quadruplexes.
Applications for the characterization and the selection of ligands.**

Guanines or cytosines rich nucleic acids can fold into tetrameric G-quadruplexes (G4) or i-motifs structures. G4 motifs are found within the human genome and should contribute to cellular regulation. In particular G4 are found at telomeric region and also in promoters of oncogenes or within viral genomes. They are suspected of participating in the regulation of human pathologies and have therefore been envisioned as potential therapeutic and diagnostic targets. However, the intrinsic conformational polymorphism of G4 motifs complicates the development of specific and affine ligands. In this context, the laboratory has developed the TASQ concept for "Template Assembled Synthetic G-Quadruplex" with the aim to obtain a defined G4 topology.

The first chapter reports on the assembly on the peptide template of RNA and DNA:RNA hybrid G4 structures that derive from the human telomeric sequence as well as of DNA G4 structure found within the HIV virus promoter. G-triplex (G3) motif which is supposed to be an intermediate during the formation of the G4 motifs has also been prepared. By using appropriate ligations of the oligonucleotide strands on the peptide template we were able to control the folding of G-quadruplex motifs and stabilize them.

The second chapter reports the studies for the characterization and the selection of ligands against G4 and G3 motifs. The evaluation of the affinity and selectivity of different families of ligands for these constrain motifs was performed by using surface plasmon resonance or by bio-layer interferometry. The selection of ligands was carried out by the SELEX method in order to obtain affine and specific aptamers of a constrained G4 motif.

Keywords: G-quadruplex, G-triplex, i-motif, chemoselective ligation, cyclodecapeptidic scaffold, SPR, SELEX

Lieu : Département de Chimie Moléculaire, équipe I2BM, UMR 5250, Université Grenoble Alpes, 570 rue de la chimie, 38041 Grenoble Cedex 9 (France)