

Nouveaux ligands fonctionnels des G-quadruplexes pour l'imagerie cellulaire et la capture covalente d'ADN génomique

Thibaut Masson

► To cite this version:

Thibaut Masson. Nouveaux ligands fonctionnels des G-quadruplexes pour l'imagerie cellulaire et la capture covalente d'ADN génomique. Chimie organique. Université Paris-Saclay, 2021. Français. NNT: 2021UPASF056 . tel-03552676

HAL Id: tel-03552676 https://theses.hal.science/tel-03552676

Submitted on 2 Feb 2022

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.



Nouveaux ligands fonctionnels des Gquadruplexes pour l'imagerie cellulaire et la capture covalente d'ADN génomique

Novel functional G-quadruplex ligands for cell imaging and genomic DNA covalent capture

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°571 : Sciences chimiques : molécules, matériaux, instrumentation et biosystèmes (2MIB) Spécialité de doctorat: Chimie Graduate School : Chimie. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **CMBC** (Université Paris-Saclay, CNRS, INSERM), sous la direction de **Marie-Paule TEULADE-FICHOU**, Directrice de Recherche, et le co-encadrement de **Daniela VERGA**, Chargée de Recherche

Thèse présentée et soutenue à Paris-Saclay, le 13 Décembre 2021, par

Thibaut MASSON

Composition du Jury

Boris VAUZEIL	LES	Précident	
Directeur de Re	echerche, ICSN (UPR 2301)	Fresident	
Eric DEFRANC	Q	Papportour & Evaminatour	
Professeur, Uni	versité Grenoble Alpes (UAR 2607)		
Maria DUCA			
Chargée de Re	cherche, HDR, Université Côte	Rapporteur & Examinatrice	
d'Azur (UMR 72	272)		
Anthony BUG	AUT		
Chargé de Recl	nerche, Université de Bordeaux	Examinateur	
(UMR5320-U12	212)		
Dominique GUIANVARC'H		Eveninetrice	
Professeure, Université Paris-Saclay (UMR 8182)		Examinative	
Marcel HOLLE	NSTEIN	Eveningtour	
Chargé de Recl	nerche, Institut Pasteur (UMR3523)	Examinateur	
Marie-Paule T	EULADE-FICHOU		
Directrice de Re	echerche, Institut Curie (UMR9187-	Directrice de thèse	
U1196)			

LHESE DE DOCTORAT

NNT : 2021UPASF056

« Je préfère la force du raisonnement à l'éloquence du style. »

Napoléon Bonaparte

Remerciements

Je tiens tout d'abord à remercier Marie-Paule Teulade-Fichou de m'avoir donné l'opportunité de travailler au sein de son laboratoire, tout d'abord en stage de Master et ensuite pour ces trois ans de thèse. Merci pour ses conseils, ses remarques et ses idées qui ont aiguillé mon travail de thèse.

Je remercie, évidemment, Daniela Verga, officieusement ma deuxième directrice de thèse qui m'a suivi, aidé et rassuré tout au long de ces trois ans. Peu de mots me viennent pour exprimer ma gratitude envers elle. Je suis fier d'avoir été son premier thésard, même si cela n'aura pu être officialisé. Elle m'a appris l'organisation, la critique des résultats et la rigueur de travail ; en bref, elle a fait de moi un meilleur scientifique. Pour tout ça, merci.

Je tiens à remercier tous les membres du jury d'avoir accepté d'évaluer ces travaux de doctorat. Merci à Maria Duca et Éric Defrancq d'avoir accepté d'évaluer ce manuscrit en tant que rapporteur. Merci aussi à Dominique Guianvarc'h, Anthony Bugaut, Marcel Hollenstein et Boris Vauzeilles pour avoir accepté de prendre le temps d'évaluer ce travail de thèse.

Un grand merci à toutes les personnes qui ont travaillé sur les différents projets de cette thèse. Merci à Corinne Landras Guetta, une excellente chimiste qui m'a aidé pour la synthèse des différents produits et qui m'a accompagné tout au long de ces trois ans. Merci à Anne Cucchiarini (social manager et DJ à ses heures perdues), Laurent Sevaille (partenaire de labo et maitre du piranha), Eugénie Laigre, Charles-Henri Buffet, et Rahima Chennoufi qui vient de rejoindre l'équipe. Merci évidemment aux différents stagiaires que j'ai encadrés : Nadège Heuberger, Nakid Moindjie et Simona Serra. Je n'oublie pas, bien entendu, Sophie Bombard, qui m'a particulièrement aidé pour mes expériences de biochimie et qui a pris de son temps pour discuter des résultats.

Merci à toutes les personnes avec qui j'ai partagé le grand bureau. Le support psychologique apporté par toutes ces personnes a grandement aidé à surmonter les difficultés de la thèse. Merci aux sœurs jumelles Marie Auvray et Chloé Breton-Patient avec qui j'ai passé d'excellents moments, même pendant les passages à vide. Mention spéciale pour les références culturelles de Chloé qui m'a fait réviser mes classiques pendant ces trois ans : « Tempora mori, tempora mundis recorda, voilà ça par exemple, ça ne veut absolument rien dire, mais l'effet reste le même ». Merci, évidemment, à Jaime Franco-Pinto, toujours à l'écoute et toujours présent pour aider. Je suis content que tous ses voyages aient fini par le mener en France et j'espère qu'il y restera longtemps.

4

Merci évidemment à toutes les personnes du laboratoire du dessus : Romain Najman, Claire Beauvineau, Delphine Naud, Pauline Fornarelli et Florence Mahuteau-Betzer. Merci à eux pour leur aide et pour leur bonne humeur permanente.

Merci à tous ceux qui ont participé à rendre cette thèse plus agréable : Oksana Reznichenko, Deepanjan Gosh, Julie Le Bescont, Coralie Caron, Abhijit Saha, Michela Zuffo et Pedro Guixens-Gallardo.

Je n'oublie évidemment pas Nathalie Marie, toujours disponible pour aider, je ne sais pas comment j'aurais pu faire sans elle pour toute la partie administrative.

Merci à toutes les personnes de cette unité, pour leur joie de vivre, leur écoute, leur aide.

Merci à ma famille d'avoir toujours été présente. Merci à mes parents qui m'ont toujours poussé à donner le meilleur de moi-même, je n'aurais jamais rien accompli sans leur soutien. Merci particulièrement à mon grand-père et à ma mère qui m'ont aidé à corriger ce manuscrit. Merci à mon meilleur pote, toujours présent après plus de 23 ans d'amitié.

Et finalement merci à ma compagne qui partage ma vie depuis maintenant 6 ans. Je ne sais pas comment j'aurais fait sans son aide, sa « positive attitude », sa gentillesse, son humour, sa générosité... Pour tout ça, merci.

Sommaire

Remerciements	4
Abréviations	
Chapitre I :Introduction	13
1. Les acides nucléiques	15
a. Structures principales de l'ADN et de l'ARN	15
i. L'ADN	15
ii. L'ARN	
b. Structures non canoniques	
2. Les G-quadruplexes	21
a. Structures des G-quadruplexes	21
b. Détermination de structure	25
i. Étude par Dichroïsme Circulaire (CD)	25
ii. Étude par diffraction des rayons X	28
iii. Étude par spectroscopie RMN	31
3. Localisation des G-quadruplexes au sein du génome	
a. Localisation par analyses in silico	
b. Localisation par analyses expérimentales in vitro	35
i. Localisation par inhibition de polymérases	35
ii. Localisation par réactions biochimiques	
iii. Localisation par Immunoprécipitation de la chromatine (ChIP)	
4. Rôles des G-quadruplexes	
a. Les télomères	39
b. G-quadruplexes génomiques	40
5. Ligands des G-quadruplexes	43
a. Structures et interactions	43
b. Méthodes pour l'étude des interactions ligands/G-quadruplexes	51
i. Analyses des propriétés physiques des ligands G-quadruplexes	51

	ii.	Résonance Plasmonique de Surface (SPR)	53
	iii.	Fluorescent Intercalator Displacement (FID)	54
	iv.	FRET-melting	57
6.	Ob	jectifs du travail de doctorat	60
Cha	Chapitre II : Immunodétection des G-quadruplexes guidée par des ligands65		
1.	Inti	roduction	67
а	. [Détection des G-quadruplexes par immunofluorescence	67
Ł). [Détection de ligands des G-quadruplexes	69
	i.	Détection de ligands des G-quadruplexes fluorescents	70
	ii.	Fonctionnalisation de ligands G-quadruplexes connus par des fluorophores	74
С	. F	Principe de l'immunodétection de ligands des G-quadruplexes	78
2.	Syr	nthèse des ligands pour la méthodologie G4-GIS	81
а	. 9	Synthèse des précurseurs PDC-Alk/N3	81
Ł). 9	Synthèse des dérivés 5-BrdU-Alk/N3	84
С	. F	Réaction de cycloaddition alcyne azoture catalysée au cuivre	85
3.	Étu	ude de l'affinité des ligands	
а	. 3	Séquences utilisées pour les analyses biophysiques	89
Ł). E	Étude par FRET-melting	93
С	. E	Étude par FID	99
	i.	Étude en plaques 96-puits	
	ii.	Étude en cuves	103
c	I. E	Étude par une expérience ELISA modifiée	106
4.	Étu	Ide de la réaction de CuAAC avec ou sans structures G-quadruplexes	
5.	Imi	munodétection de ligands des G-quadruplexes in cellulo	112
6.	Со	nclusion	120
Cha	pitre	e III : Photo-alkylation de structures G-quadruplexes	123
1.	Inti	roduction	125
а	. (Capture de structures G-quadruplexes par des ligands spécifiques	125

b	•	Alkylation de structures G-quadruplexes	128
	i.	Méthodologie PAL	128
	ii.	. Méthodologie PAL appliquée aux structures G-quadruplexes	130
C.		Application de la photo-alkylation pour la capture d'ADN G-quadruplex génomique	132
2.	S	ynthèse des ligands G-quadruplexes photoactivables	135
a		Synthèse des composés linéaires	135
b	-	Synthèse des composés non linéaires	139
3.	Et	tudes biophysiques de l'interaction des ligands avec les G-quadruplexes	143
a		FRET-melting	143
b	-	FID	145
c.		CD	147
4.	Ét	tude de la réaction de Huisgen catalysée au cuivre	152
5.	P	hoto-alkylation des structures G-quadruplexes	155
a		Photo-alkylation de la séquence télomérique humaine 22AG	156
b		Photo-alkylation de la séquence cMyc22	163
c.		Photo-alkylation de la séquence CEB25wt	165
d		Sélectivité de la réaction de photo-alkylation	168
6.	C	onclusion	170
Cha	pit	re IV: Étude de l'impact de la régioisomérie de dérivés bisquinolines/bisquinoliniums	173
1.	Ir	ntroduction	175
2.	S	ynthèse des composés	177
a	•	Synthèse des composés PDC et PhenDC Bispyro	177
b		Synthèse du composé PDC2	179
3.	Ét	tudes biophysiques	180
a		FRET-melting	180
b		FID	181
4.	C	onclusion	185
Cha	pit	tre V : Conclusion générale et perspectives	189

Partie expérimentale199		
Synthesis and characterization2		
Biophysical experiments		
FRET-melting experiments25	55	
FID experiments25	56	
a. Cuve experiments25	56	
b. High-throughput G4-FID25	57	
Circular Dichroism25	57	
Biochemistry experiments		
Sample preparation25	58	
Photo-alkylation experiments25	58	
Références		

Abréviations

- 3D-SIM : 3D-Structured Illumination Microscopy
- 5-BrdU : 5-bromo-2'-désoxyuridine
- ADN : Acide désoxyribonucléique
- ARN : Acide ribonucléique
- ARNr : ARN ribosomique
- ARNt : ARN de transfert
- BSA : Bovine Serum Albumin
- CD : Circluar Dichroism
- cIKP : Cyclic imidazole/lysine polyamide
- CuAAC : Copper-catalyzed Alkyne Azide Cycloaddition
- DAPI: 4',6-diamidino-2-phénylindole
- DBU : 1,8-Diazabicyclo[5.4.0]undéc-7-ène
- DCM : Dichlorométhane
- DIPEA : N,N-Diisopropyléthylamine
- DMAP: 4-diméthylaminopyridine
- DMF: N,N-Diméthylformamide
- DMSO : Diméthylsulfoxyde
- DRX : Diffraction des Rayons X
- ELISA : Enzyme-Linked Immunosorbent Assay
- FAM : Fluorescéine
- FID : Fluorescent Intercalator Displacement
- FLIM : Fluorescence Lifetime Imaging
- FRET : Förster Resonance Energy Transfer
- G4 : G-quadruplex
- G4-GIS : G4-ligand Guided Immunofluorescent Staining

HATU : hexafluorophosphate de-N, N-diméthylméthaniminium HILO : Highly Inclined Laminated Optical HOSA : Hydroxylamine-O-sulfonic acid Hp : Hairpin hPIP : Hairpin pyrrol/imidazole polyamide HPLC : High Performance Liquid Chromatography HRP : Horseradish peroxidase NaAsc : Ascorbate de sodium NHEIII : Nuclease Hypersensitive element III NMIA : N-methylisatoic anhydride NMM : N-methyl mesoporphyrine PAL : Photoaffinity labeling PCC : Pyridinium chlorochromate PDC : Pyridinedicarboxamide **PDS** : Pyridostatin PEG : Polyéthylène glycol PQS : Putative G-Quadruplex Sequence qPCR : quantitative Polymerase Chain Reaction RMN : Résonnance Magnétique Nucléaire **ROS** : Reactive Oxygene Species SHAPE : Selective 2'-Hydroxyl Acylation analyzed by Primer Extension SPAAC : Strain-Promoted Alkyne Azide Cycloaddition **SPE : Solid Phase Extraction** SPR : Surface Plasmon Resonance

- TAMRA : Tetraméthylrhodamine
- TBAF : Fluorure de tert-butylammonium
- TBDMSCI : Chlorure de tert-butyldiméthylsilyle

- TBTA : Tris(benzyltriazolylmethyl)amine
- TEA : Triéthylamine
- TEG : Triéthylène glycol
- THPTA : Tris(benzyltriazolylmethyl)amine
- TMB: 3,3',5,5'-tétraméthylbenzidine
- TO : Thiazole Orange
- TOD : Thiazole Orange Displacement
- TSS : Transcription Start Site

Chapitre I : Introduction

1. Les acides nucléiques

a. Structures principales de l'ADN et de l'ARN

Les acides nucléiques sont des biopolymères retrouvés dans les cellules de chaque organisme vivant. Ces acides nucléiques existent sous deux structures différentes : l'acide désoxyribonucléique (ADN) et l'acide ribonucléique (ARN). Ces acides nucléiques sont formés par l'enchaînement de nucléotides, eux-mêmes constitués d'une base azotée, d'un sucre et de groupements phosphates. Le sucre est un pentose de type D-ribose pour les structures d'ARN et D-2-désoxyribose pour les structures d'ADN. Ces derniers sont reliés par une liaison phosphodiester entre la position 3' du sucre et la position 5' du sucre adjacent. Une base azotée est présente sur chacune de ces structures au niveau du carbone anomérique du ribose ou désoxyribose (Figure 1A). Il existe 5 bases azotées différentes qui peuvent être des pyrimidines (un hétérocycle aromatique à 6 chaînons) ou des purines (un hétérocycle aromatique à 6 chaînons fusionné à un cycle imidazole à 5 chaînons). Le groupe des pyrimidines est composé d'une cytosine (C), d'une thymine (T) et d'un uracile (U). L'uracile est uniquement présent dans les brins d'ARN et remplace la thymine, uniquement présente dans les brins d'ADN. Le groupe des purines quant à lui est composé d'une adénine (A) et d'une guanine (G) et ces dernières sont présentes à la fois au sein des brins d'ARN et d'ADN (Figure 1B).



Figure 1: A) Structure représentative du motif nucléotide composant les acides nucléiques. B) Structures des bases hétérocycliques composant l'ADN et l'ARN.

i. L'ADN

Il a été démontré en 1953 par J. Watson et F. Crick que les bases azotées au sein des brins d'ADN étaient capables de former des liaisons hydrogène fortes, donnant à l'ADN une structure hélicoïdale en double brin¹. En effet, une adénine et une thymine sont capables d'interagir entre elles par deux

liaisons hydrogène tandis qu'une guanine et une cytosine sont capables d'interagir entre elles par trois liaisons hydrogène (Figure 2A). Dans cette structure à double hélice, les deux brins complémentaires sont antiparallèles : l'extrémité 5' du premier brin d'ADN interagit avec l'extrémité 3' du second brin d'ADN. La forme la plus commune de l'ADN à double hélice est la forme B : c'est une double hélice droite possédant entre 10,4 et 10,5 paires de bases par tour et qui peut se former dans les conditions hydratées de la cellule² (Figure 2B).



Figure 2: A) Représentation schématique des liaisons hydrogène de type Watson-Crick entre les bases A-T et C-G. B) Représentation d'une structure B adoptée par un brin d'ADN. Image tirée de Garcia-Ramos *et al.*³

Dans les cellules humaines et plus généralement dans les cellules eucaryotes, l'ADN est principalement retrouvé sous sa forme B. Entre deux divisions cellulaires, les 3,2 milliards de paires de bases du génome humain se retrouvent compactées dans 23 paires de chromosomes au sein de chaque noyau cellulaire. Cette importante compaction de l'ADN est due aux histones : des protéines chargées positivement avec lesquelles les groupements phosphates de l'ADN, chargés négativement, peuvent interagir. 8 histones vont pouvoir former un octamère et environ 147 paires de bases d'ADN vont s'enrouler autour de ce complexe, formant le nucléosome. Les nucléosomes se replient ensuite pour former une fibre de chromatine sous forme de boucles. Ces boucles sont de nouveau pliées et enroulées pour former la chromatide d'un chromosome⁴ (Figure 3). La capacité d'expression des gènes est ainsi directement affectée par l'état de la chromatine : si elle est condensée (hétérochromatine), les gènes ne pourront pas être exprimés, car l'ADN sera inaccessible à la machinerie transcriptionnelle. À l'inverse, si la chromatine est ouverte (euchromatine), l'expression génétique pourra avoir lieu.



Figure 3: Schéma représentatif du principe de formation des chromosomes. Image tirée et adaptée de : HAL Id : tel-02003415.

ii. L'ARN

L'ARN, contrairement à l'ADN, est retrouvé très majoritairement sous forme monocaténaire (simple brin), due notamment à la présence de la fonction hydroxyle en position 2' du sucre⁵. Comme cité précédemment, la base thymine est remplacée par la base uracile et l'ARN va ainsi former très majoritairement des structures secondaires intramoléculaires par des interactions A-U et C-G. Ces structures secondaires sont essentielles aux fonctions et à la stabilité des brins d'ARN et sont majoritairement constituées de différentes boucles et de bulges⁶ (ou « renflements ». Figure 4).



Figure 4: Exemple d'une structure à ARN constituée de boucles et de bulges.

Tandis que l'ADN joue un rôle de stockage de l'information génétique et reste majoritairement dans le noyau, l'ARN joue plusieurs rôles complexes de façon transitoire. Par exemple, l'ARN messager (ARNm) permet de transporter l'information génétique portée par l'ADN jusqu'aux ribosomes dans le but de synthétiser les protéines associées à ce gène⁷. Ce rôle de transport de l'information génétique est aujourd'hui le plus connu, mais l'ARN permet de réaliser d'autres processus nécessaires comme la régulation de l'expression génétique⁸ ou la catalyse de réactions biochimiques⁹ par des ARN non codants. Un autre rôle intéressant des ARN est le rôle de guide (ou template) : ces séquences d'ARN se lient à des protéines et permettent de guider l'action de ces dernières sur des séquences d'ARN ou d'ADN complémentaires¹⁰.

b. Structures non canoniques

Comme vu précédemment, la forme B de l'ADN est la plus répandue *in cellulo*. Cependant, plusieurs autres structures à double hélice ont été décrites et les deux formes les plus connues sont la forme A et la forme Z. La première est une hélice droite plus large que la forme B présentant environ 11 paires de bases par tour. C'est une structure pouvant être adoptée par l'ADN pour favoriser les interactions avec des enzymes¹¹ et qui est régulièrement observée dans des conditions de faible hydratation nécessaires pour l'analyse par DRX¹² (pour Diffraction des Rayons X, Figure 5). La forme Z quant à elle est une structure à double hélice gauche moins enroulée que la forme B avec environ 12 paires de bases par tour (Figure 5). C'est une structure pouvant notamment se former lors de

surenroulements induits par la transcription de certains brins d'ADN ainsi que dans les régions hyperméthylées des CpG-islands¹³.



Figure 5: Représentation des structures A et Z de l'ADN. Image tirée de Garcia-Ramos et al.³

Ces structures double brins sont donc des structures pouvant se former dans une grande majorité du génome. Cependant, la répartition des bases de l'ADN génomique n'est pas aléatoire et certaines parties du génome peuvent contenir des séquences particulières comme des répétitions de bases ou des répétitions de bases inversées (ou en miroir). Ces motifs répétitifs peuvent donner lieu à la formation de structures non canoniques comme l'ADN cruciforme, la jonction de Holliday, les i-motifs, l'ADN et l'ARN G-quadruplex, etc. L'ADN cruciforme et les structures en épingle à cheveux peuvent se former lorsqu'une séquence présente une symétrie des bases inversée, par exemple : 5'-TGACGT[...]ACGTCA-3'. Dans ce cas, le simple brin d'ADN peut se replier sur lui-même par appariement des bases et former un motif en épingle à cheveux (Figure 6A). L'ADN cruciforme est composé de deux motifs en épingle à cheveux de deux brins d'ADN différents et d'une jonction à 4 voies (Figure 6B). Ces structures sont connues pour avoir de nombreux rôles biologiques comme la modification du surenroulement de l'ADN¹⁴, la stabilisation des brins d'ARN et la régulation de la vitesse de transcription des ARNm¹⁵.



Figure 6: Structures possibles de l'ADN présentant une répétition des bases inversée avec A) l'ADN en épingle à cheveux et B) l'ADN cruciforme.

Dans les zones du génome répétitives en cytosines peuvent se former des structures secondaires nommées i-motifs (pour intercalated motif). Ce sont des structures à 4 brins formées par l'empilement de paires de bases C:C⁺, stabilisées à pH légèrement acide (Figure 7). La nécessité d'un pH légèrement acide pour stabiliser les i-motifs posait des questions sur la pertinence de formation de telles structures *in vivo*. Récemment, Zeraati *et al.* ont développé un anticorps spécifique de ces structures¹⁶. Leur étude a permis de montrer que les i-motifs pouvaient tout particulièrement se former lors de la phase G1 du cycle cellulaire, phase marquée par de hauts taux de transcription. Les i-motifs sont aujourd'hui suspectés de participer à la régulation d'expression de certains gènes¹⁷ ainsi qu'à la régulation de la réplication de l'ADN¹⁸.



Figure 7: A) Représentation schématique des interactions C:C+ et B) Représentation d'une structure i-motif.

L'intérêt porté aux i-motifs est arrivé de pair avec la découverte des structures G-quadruplexes. En effet, si les i-motifs peuvent se former lorsqu'un simple brin d'ADN possède une répétition de cytosines, les G-quadruplexes peuvent quant à eux se former lorsqu'un simple brin possède une répétition de guanines. Parmi toutes les structures non canoniques pouvant être adoptées par l'ADN, les structures G-quadruplexes sont, de loin, les structures les plus étudiées depuis ces 20 dernières années.

2. Les G-quadruplexes

a. Structures des G-quadruplexes

Les G-quadruplexes (G4s) sont des structures non canoniques dont l'existence *in vitro* est connue depuis 1962¹⁹. Cette structure était alors produite par autoassociation de guanines au sein d'un gel d'acide guanylique et observée par DRX. C'est 27 ans plus tard que cette structure a été observée pour la première fois sur un brin d'ADN télomérique²⁰. Globalement, les G-quadruplexes peuvent être formés lorsque l'ADN ou l'ARN est sous forme simple brin et présente une séquence de type 5'- $G_{\ge 2}N_nG_{\ge 2}N_nG_{\ge 2}N_nG_{\ge 2}-3'$. En effet, quatre répétitions d'au moins deux guanines successives peuvent s'autoassembler pour former au moins 2 G-quartets (ou G-tétrades, Figure 8A) interagissant entre eux par π -stacking et reliés par des boucles d'un à plusieurs nucléotides (Figure 8B). Ces quartets sont formés de 4 guanines agissant à la fois comme donneurs et accepteurs de liaisons hydrogène à travers des interactions de type Hoogsteen. En effet, la position N1 d'une guanine sert de donneur à l'oxygène O6 d'une seconde guanine et l'azote N7 de la première guanine sert d'accepteur de liaison hydrogène de la position N2 d'une seconde guanine.



Figure 8: A) Structure d'un G-quartet formé par l'autoassociation de 4 guanines par des liaisons hydrogène de Hoogsteen. B) Schéma d'une structure G-quadruplex formée par l'empilement de 3 G-quartets.

La forte densité électronique due aux quatre guanines est compensée par la présence de cations monovalents localisés soit au centre de chaque quartet soit entre deux quartets en fonction de la taille du cation et de la topologie du G-quadruplex. En effet, les oxygènes O6 des quatre guanines sont orientés vers le centre du G-quadruplex et peuvent se complexer avec les cations présents dans le milieu : principalement le potassium K⁺ et le sodium Na⁺ en conditions physiologiques. Le potassium peut être coordiné par 8 atomes O6 entre deux quartets²¹ (133 pm de rayon, Figure 9A) et le Na⁺ peut être coordiné soit par 8 atomes O6 entre deux quartets soit coordiné par 4 atomes O6 directement dans le plan d'un quartet²¹ (95 pm de rayon, Figure 9B). En plus de ces deux cations

monovalents, de nombreux autres ont montré la capacité de stabiliser les G-quadruplexes tels que Rb⁺, le Ba²⁺, le Ca²⁺ etc²¹.



Figure 9: Structures cristallographiques de G-quadruplexes formés par A) une séquence du promoteur *B-raf* (G₃CG₄AG₅A₂G₃) en présence d'ions K⁺ représentés ici en violet (PDB : 4H29²²) et B) le tétramère de la séquence (TG₄T) en présence d'ions Na⁺ représentés ici en rose (PDC : 352D²³). Par souci de clarté, le brin d'ADN en premier plan a été mis en transparence et les molécules d'eau ont été supprimées.

Les structures G-quadruplexes sont des structures à 4 brins très polymorphiques et plusieurs facteurs vont influencer leur structure : le type d'oligonucléotide (ADN ou ARN), la composition de la séquence, la molécularité (le nombre de brins impliqués dans la formation du G-quadruplex), le nombre de nucléotides compris dans les boucles ainsi que la direction de ces dernières, le nombre et la conformation des guanines impliquées dans la formation du G-quadruplex, etc²⁴.

Les G-quadruplexes peuvent donc être formés par association d'un (monomoléculaire), de deux (bimoléculaire), de trois (trimoléculaire) ou de quatre (tétramoléculaire) brins d'ADN ou d'ARN différents (Figure 10).



Figure 10: Représentation schématique de structures G-quadruplexes formées par association d'un ou plusieurs brins d'oligonucléotides.

Les G-quadruplexes peuvent adopter trois types de topologies en fonction de l'orientation des brins : la topologie parallèle, antiparallèle ou hybride. Une structure parallèle sera formée de brins de même orientation (Figure 11A), une structure antiparallèle sera formée de deux brins dans un sens et de deux brins dans l'autre (Figure 11B) et la topologie hybride sera formée par un brin dans une direction et des trois autres brins dans l'autre (Figure 11C). L'orientation de ces brins est directement modulée par la conformation des guanines dont la liaison N-glycosidique peut être *syn* ou *anti* (Figure 11D). Ainsi, les structures parallèles ne seront composées que de guanines *anti* tandis que les structures antiparallèles et hybrides seront composées d'un mélange de guanines *syn* et *anti*. Les conformations adoptées par les guanines vont permettre de former un total de trois types de boucles différentes : les boucles latérales, diagonales ou « propeller ». Les boucles latérales et diagonales vont relier deux brins de directions opposées entre eux tandis que les boucles « propeller » vont relier deux brins de même direction.



Figure 11: Représentation des différentes topologies adoptées par les G-quadruplexes avec A) la topologie parallèle, formée uniquement de guanines anti et possédant 3 boucles « propeller », B) la topologie antiparallèle formée d'un mélange de guanines *syn* et *anti* et possédant deux boucles latérales et une boucle diagonale, C) la topologie hybride formée d'un mélange de guanines *syn* et *anti* et possédant deux boucles latérales et une boucle « propeller », et D) les structures des guanines *syn* et *anti*.

Le nombre et le type de nucléotides présents dans les boucles influent fortement sur la topologie ainsi que sur la stabilité d'un G-quadruplex. En effet, les G-quadruplexes à ADN possédant de « petites » boucles ont tendance à former des G-quadruplexes parallèles très stables. Au contraire, les séquences avec de « longues » boucles ont tendances à former des G-quadruplexes antiparallèles ou hybrides et sont généralement moins stables²⁵. Dans le cas des G-quadruplexes à ARN, la présence de la fonction hydroxyle en position 2' augmente les contraintes stériques et a tendance à favoriser la conformation *anti* des guanines. Cette contrainte stérique favorise donc majoritairement la formation de structures G-quadruplexes parallèles, généralement plus stables que leurs homologues ADN²⁶.

La dernière caractéristique structurale des G-quadruplexes réside au niveau des sillons. Étant des structures hélicoïdales à 4 brins, elles forment quatre sillons de tailles différentes dépendant directement de l'agencement des guanines. Ces sillons peuvent être larges, moyens ou étroits. Deux guanines *syn-syn* ou *anti-anti* adjacentes d'un même quartet formeront un sillon moyen, une guanine *syn* perpendiculaire à une guanine *anti* formera un sillon étroit et une guanine *anti* perpendiculaire à une guanine syn formera un sillon large²⁷.



Figure 12: Représentation schématique des trois sillons créés par l'enchaînement de guanines de conformation *anti* et *syn*.

Enfin, certaines séquences contenant des insertions de bases au sein même d'une répétition de guanines sont capables de former des structures G-quadruplexes. Par exemple, la séquence d(TTGTGGGTGGGTGGGTGGGT) possède un résidu thymine au sein de la première répétition de guanines. Sa structure en solution K⁺ a été déterminée par RMN et a montré la formation d'un G-quadruplex possédant 3 boucles « propeller » ainsi qu'un renflement directement après la première guanine impliquée dans la formation du G-quadruplex. Dans le but de vérifier si la position de ce renflement impactait la formation de G-quadruplexes, de nombreuses autres séquences ont été étudiées par RMN en faisant varier sa position et toutes ces séquences étaient capables de former des structures G-quadruplexes bien définies²⁸. La formation de ces structures permet d'élargir la quantité de G-quadruplexes pouvant potentiellement se former par rapport à la séquence reconnue « G₂₂N_n G



Figure 13: Structure adoptée par la séquence d(TT<u>G</u>T<u>GG</u>T<u>GGG</u>T<u>GGG</u>T<u>GGG</u>T) déterminée par RMN montrant la présence d'une bulge au sein de la première répétition de guanines. Image tirée de A. T. Phan *et al.*²⁸

b. Détermination de structure

Comme vu précédemment, les G-quadruplexes sont des structures très polymorphiques et déterminer précisément la structure adoptée par une séquence riche en guanines est primordial pour la compréhension de leur importance biologique. Par conséquent, de nombreuses méthodes d'analyses ont été développées dans le but de les étudier.

i. Étude par Dichroïsme Circulaire (CD)

Le dichroïsme circulaire est un phénomène optique survenant lors de l'interaction d'un faisceau de lumière polarisée circulairement avec un milieu optiquement actif. En effet, ce phénomène survient lorsque la composante droite et la composante gauche de la lumière polarisée sont absorbées différemment par le milieu (Figure 14A). Le dichroïsme circulaire est aujourd'hui principalement utilisé pour déterminer les structures secondaires des protéines et de l'ADN. En effet, deux caractéristiques sont nécessaires à un échantillon pour obtenir un spectre CD : il doit être à la fois chiral et absorber la lumière. Dans le cas des protéines et de l'ADN, l'absorbance de ces molécules biologiques est généralement comprise entre 200 et 300nm et la chiralité est assurée par les centres asymétriques des acides aminés dans le cas de protéines et des carbones anomériques des sucres dans le cas de l'ADN.



Figure 14: Représentation schématique A) du principe de dichroïsme circulaire: la lumière polarisée droite est absorbée différemment par le milieu chiral que la lumière polarisée gauche et B) des deux moments dipolaires de transition électronique d'une guanine.

Lorsqu'une séquence d'ADN est riche en guanine et est suspectée de former une ou plusieurs structures G-quadruplexes, l'analyse par dichroïsme circulaire est une méthode rapide et efficace pour en connaitre la topologie. Dans le cas des G-quadruplexes, l'absorption de la lumière d'une longueur d'onde supérieure à 210 nm est assurée par les guanines tandis que l'effet des groupements pentose et des liaisons phosphodiesters est négligeable. La lumière polarisée induit ainsi deux transitions π - π * des guanines à 250 et 279nm dont les moments dipolaires de transition se retrouvent sur les deux axes d'une guanine²⁹ (Figure 14B). Ces deux excitons vont pouvoir interagir avec les guanines présentes au niveau des autres quartets. Dans une structure G-quadruplex, les quartets sont empilés et légèrement tournés les uns par rapport aux autres. Cette rotation est directement due à la conformation des guanines impliquées dans les quartets adjacents et induit une chiralité qui va impacter l'énergie des excitons couplés entre les guanines. Cette chiralité sera donc observée sur un spectre de dichroïsme circulaire par une différence d'intensité et de longueur d'onde des deux transitions π - π *.

Dans une structure G-quadruplex, les guanines de deux quartets adjacents peuvent interagir entres elles selon une même face « head to head » (ou tête à tête) ou « tail to tail » (ou queue à queue) ou selon deux faces différentes « head to tail » ou « tail to head »³⁰ (Figure 15). L'alternance de ces interactions va ainsi modifier la polarité du quadruplex, donner lieu à différentes énergies de transition π - π ^{*31} et donc différentes spectres CD qui peuvent être classifiés en trois groupes.



Figure 15 : Représentation de la face head et de la face tail des guanines (à gauche) et exemple d'interactions head to tail (H-to-T) et head to head (H-to-H) formées par deux quartets. Image tirée de : Quadruplex Nucleic Acids³⁰.

Pour un G-quadruplex parallèle formé de guanines uniquement *anti*, les guanines de deux quartets adjacents s'empilent de façon H – T (pour « head to tail ») et seront attribués au groupe I. Pour les G-quadruplexes dont les guanines *syn* et *anti* s'enchaînent sans s'alterner (*anti-anti-syn, syn-syn-anti*, *syn-anti-anti* ou *anti-syn-syn*), les guanines s'empilent en alternant H – H et H – T (ou T – H) et font partie du groupe II. Dans le cas des G-quadruplexes formés par l'alternance de guanines *syn-anti-syn* ou *anti-syn-anti*, les guanines vont s'empiler par des interactions de type H – H et T – T (groupe III)³².

Finalement, la conformation uniquement *anti* des guanines d'une structure parallèle donnera un intense pic positif de CD aux alentours de 260 nm. Tandis qu'une structure antiparallèle, qui contient donc des guanines de conformation à la fois *syn* et *anti*, présentera un pic positif aux alentours de 290 nm ainsi qu'un pic négatif aux alentours de 260 nm. Le mélange de conformation syn et anti des guanines peut aussi former des structures hybrides qui possèdent 3 brins dans la même orientation et 1 brin d'orientation inverse. Cette topologie implique la formation de deux bandes positives aux alentours de 290 et 265 nm (Figure 16).



Figure 16: Exemple de spectre CD de la séquence antiparallèle « basket-type » Tel22³³ en tampon Na⁺ (noire), de la séquence parallèle dérivée du proto-oncogène c-Myc³⁴ en tampon Na⁺ (rouge) et de la séquence hybride 2GKU³⁵ en tampon K⁺ (bleue). Image tirée de M. Schmid.³⁶

L'analyse par CD est une méthode qualitative puissante, rapide et non destructive qui permet donc de discriminer des topologies différentes entre plusieurs séquences d'ADN et est aujourd'hui largement utilisée dans le monde entier.³⁷ Cette méthode offre l'avantage d'être réalisée en milieu aqueux et permet donc d'obtenir des résultats pouvant représenter la réalité *in vivo*. Cette méthode permet, de plus, de pouvoir visualiser en temps réel les changements de topologie d'une séquence lors de la modification des conditions expérimentales (température, salinité, etc.) comme démontré par Gray *et al.* sur la séquence d'ADN télomérique humain Tel22.^{38,39} Elle a de plus été la première méthode à avoir démontré que la topologie des G-quadruplexes formés par la séquence d'ADN télomérique humain dépendait du nombre de répétitions du motif (TTAGGG)_n et que la séquence GGG(TTAGGG)₃ adoptait une topologie hybride 3+1 en tampon K⁺⁴⁰ contrairement à la topologie parallèle observée dans les cristaux.⁴¹

Aujourd'hui, le dichroïsme circulaire est principalement limité par l'impossibilité de discriminer plusieurs structures en solution ainsi que dans l'incapacité à connaître l'arrangement des structures G-quadruplexes à un niveau atomique. Cependant, l'équipe de Rafael del Villar-Guerra a récemment réalisé et analysé une bibliothèque de spectres CD de 23 structures G-quadruplexes dans le but de développer un algorithme capable de déterminer les enchainements de bases *anti-anti, syn-anti* et *anti-syn* au sein d'une même séquence d'ADN.⁴² Cette étude participera, par enrichissement de cette bibliothèque, à faire du dichroïsme circulaire une méthode puissante capable d'aider à la caractérisation complète des structures G-quadruplexes.

ii. Étude par diffraction des rayons X

La diffraction des rayons X est une technique d'analyse reposant sur l'interaction élastique, c'est-àdire sans perte d'énergie, d'un rayon X incident avec un échantillon ordonné (cristal). Les rayons X ont une longueur d'onde d'environ quelques Angströms correspondant à la distance interatomique d'une molécule et va permettre de modifier le nuage électronique d'une molécule. L'interaction des rayons X et du nuage électronique de la maille cristalline résulte en la diffusion des rayons sans modification de fréquence de ces derniers dans certaines directions de l'espace uniquement (Figure 17A). La répartition spatiale des rayons diffusés permet de déterminer l'organisation moléculaire de l'échantillon (Figure 17B).



Figure 17: Représentation schématique de la diffraction par rayons X. A) Principe de fonctionnement de l'analyse par diffraction des rayons X, B) Exemple de structure G-quadruplex déterminée par DRX: dimère formé par deux Gquadruplexes parallèles d'une séquence dérivée du promoteur d'oncogène KRAS⁴³ (PDB: 6N65).

Cette technique d'analyse est aujourd'hui largement utilisée dans différents domaines pour déterminer les structures fines d'échantillons cristallins (molécules organiques cristallisées, céramiques, protéines, acides nucléiques, etc.). L'analyse par diffraction des rayons X a été la première méthode capable de caractériser les interactions Hoogsteen impliquées dans la formation de G-quartets au sein d'un gel séché d'acide guanylique en 1962 par M. Gellert *et al.*¹⁹ La première structure DRX d'un G-quadruplex a été déterminée 40 ans plus tard par G. Parkinson *et al.* sur la séquence télomérique humaine⁴¹.

Depuis ce jour, cette méthode est largement utilisée pour déterminer avec précision l'arrangement atomique des structures G-quadruplexes. En effet, l'utilisation de cette méthode en tandem avec la spectrométrie RMN a permis de connaître la topologie et la géométrie détaillée d'un très grand nombre de séquences formant des structures G-quadruplexes. L'obtention de ces structures fines est primordiale pour en comprendre les caractéristiques (géométrie des quartets, solvatation, localisation des contre-ions, géométrie des boucles, etc.) ainsi que pour développer des molécules capables d'interagir spécifiquement avec certaines de ces structures. Certains complexes de petites molécules en interaction avec des structures G-quadruplexes ont de plus été observés, permettant de mieux comprendre les interactions mises en jeu et la sélectivité potentielle d'une molécule envers une structure G-quadruplexe.^{44,45}

La diffraction des rayons X a aussi permis de connaitre avec précision la structure d'hélicases responsables de la déstructuration de l'ADN double brin *in vivo*. Ces hélicases sont des protéines et certaines sont capables de « dérouler » les G-quadruplexes et participent donc à leur rôle de régulation de phénomènes biologiques. Il existe de nombreuses familles d'hélicases qui diffèrent entre elles par leurs séquences d'acides aminés⁴⁶ comme DHX36, Pif1, BLM, etc. Ces hélicases sont primordiales au sein du vivant, car de nombreuses maladies sont liées à la dysfonction de ces dernières (syndromes de Fanconi, Werner, Bloom, etc.). Pour cette raison, il est nécessaire de connaitre leurs structures précises et de comprendre le mécanisme d'action de ces enzymes,

29

aujourd'hui encore largement méconnu. Récemment, un complexe de l'hélicase DHX36 et d'un brin d'ADN contenant un motif G-quadruplex ainsi qu'un motif double-brin a été co-cristallisé, révélant l'interaction spécifique de l'hélicase avec cette séquence d'ADN. Le motif N-terminal spécifique de l'hélicase DHX36 se structure en hélice α et interagit spécifiquement avec l'ADN double brin tandis que le subdomaine « OB » interagit spécifiquement avec le G-quadruplex parallèle.⁴⁷ La même année, la structure RX d'un complexe d'une hélicase bactérienne RecQ (CsRecQ) et d'un Gquadruplex déstructuré montre qu'une guanine impliquée dans la formation de la structure G4 est séquestrée dans une poche de l'hélicase, réalisant les mêmes liaisons hydrogène que dans un quartet de guanines (Figure 18). Cette structure RX montre donc la capacité de cette hélicase à retourner une guanine impliquée dans une structure quadruplex suggérant un possible mécanisme d'action *in vivo* des hélicases spécifiques des structures quadruplexes.⁴⁸





Aujourd'hui, l'analyse par diffraction des rayons X est principalement limitée par deux facteurs : la difficulté d'obtenir des cristaux diffractants avec une résolution suffisante (généralement < 5 Å) et l'obtention de résultats qui ne reflètent pas totalement la réalité en solution. En effet, l'obtention de cristaux capables d'être soumis à l'analyse par DRX est généralement l'étape limitante dans la caractérisation de structures. La cristallisation de séquences d'ADN ou d'ARN doit suivre un protocole strict et de nombreuses conditions expérimentales doivent souvent être testées (type et concentration de sels, agents de précipitation, additifs, etc.)⁴⁹. De plus, l'obtention de cristaux implique obligatoirement l'obtention d'une topologie unique du G-quadruplex. Cela ne permet donc pas toujours de pouvoir extrapoler les résultats obtenus en milieu cristallin à un milieu physiologique qui implique parfois un mélange de structures pour une même séquence.

iii. Étude par spectroscopie RMN

La spectroscopie RMN est une technique d'analyse très largement utilisée pour la caractérisation de molécules en solution et plus rarement sous forme solide. Cette technique repose sur l'interaction d'un champ magnétique extérieur avec les noyaux possédant un spin non nul. Ces noyaux agissent comme des aimants et vont réagir à un champ magnétique externe puissant en produisant des ondes électromagnétiques dont la fréquence dépend des caractéristiques du noyau ainsi que de son environnement électronique. Cette méthode permet de connaître précisément la structure de molécules complexes en solution. Elle offre l'avantage d'être non-destructive, de ne pas nécessiter de modifications des échantillons et permet d'analyser des échantillons en milieu physiologique.

La spectroscopie RMN est très bien adaptée à la détermination de structures des acides nucléiques et des protéines, elle permet notamment d'étudier la cinétique et la dynamique de formation des Gquadruplexes. Lorsqu'une séquence d'ADN est sous forme G4, les protons présents sur les azotes N1 (protons iminos, Figure 19A) vont présenter un déplacement chimique entre 10,5 et 12,0 ppm contrairement aux protons iminos des autres structures de l'ADN qui vont présenter des déplacements chimiques plus déblindés (entre 13 et 14 ppm). Les interactions de types Hoogsteen vont permettre à ces protons d'être peu échangeables avec le solvant (généralement un tampon) et vont donc produire des signaux relativement fins⁵⁰ (Figure 19B).



Figure 19: Représentation schématique A) d'un quartet de guanines formé par les liaisons hydrogène de type Hoogsteen et B) de la structure adoptée par la séquence bcl2Mid (gauche), dérivée du promoteur d'oncogène BCL2 ainsi que le spectre RMN-¹H correspondant à la région des protons de l'azote N1 de la séquence bcl2Mid (droite). Images tirées de Yang *et al.*⁵¹

L'intégration de ces protons permet de déterminer directement le nombre de quartets mis en jeu. Par exemple, 12 protons présents entre 10,5 et 12 ppm indiquent la présence d'un G-quadruplex à 3 quartets. Il est important de noter qu'un nombre supérieur de protons au nombre attendu peut indiquer la présence de plusieurs structures G-quadruplexes en solution,⁵² rendant généralement le spectre très compliqué à interpréter. Dans ce cas-là, la méthode la plus simple est de favoriser la formation d'une seule conformation. Cela peut être réalisé en modifiant la séquence directement ou en modifiant les conditions de préparation d'échantillons ainsi que les conditions expérimentales. En effet, l'ajout et/ou la suppression de bases aux extrémités de la séquence permettent parfois de favoriser une seule conformation. C'est le cas de la séquence télomérique TAGGG(TTAGGG)₃, majoritairement sous forme 1 de l'hybride 3+1 en tampon K⁺ qui passe sous forme 2 lors de l'ajout de deux résidus thymine à l'extrémité 3' (Figure 20).⁵³



Figure 20: Représentation schématique de la structure quadruplex majoritairement adoptée par la séquence A) TAGGG(TTAGGG)₃ et B) TAGGG(TTAGGG)₃TT en tampon K⁺.

La modification des conditions expérimentales et de préparation de l'échantillon peut aussi favoriser la formation d'une seule topologie. Le folding de la séquence (annealing à haute température puis refroidissement rapide ou lent), la concentration en ADN et en sels, la température, le pH et l'ajout d'agents encombrants par exemple sont des facteurs qui vont impacter la topologie finale de la structure G-quadruplex.⁵⁰ La détermination structurale est ensuite réalisée à l'aide de différentes expériences RMN, notamment NOESY, COSY, HMBC, etc. Réaliser un enrichissement isotopique en ¹⁵N et en ¹³C à des sites spécifiques du G-quadruplex permet en général d'aider à l'attribution des signaux RMN.^{54,34}

Les avancées techniques et technologiques ont permis récemment de pouvoir étudier des structures G-quadruplexes et des complexes petites molécules/G-quadruplexes directement *in cellulo*.^{55,56} Par exemple, l'équipe de J. L. Mergny a démontré qu'une structure G-quadruplex tétramoléculaire formée par la séquence (TG₄T)₄ et injectée directement dans des ovocytes de *Xenopus laevis* pouvait être analysée par RMN *in cellulo*. Cette structure G-quadruplex possède un temps de vie relativement long de 20 h, démontrant sa résistance aux différentes réactions enzymatiques⁵⁷. Pour pouvoir être suivi par RMN, l'oligonucléotide a été enrichi en isotope ¹⁵N et ¹³C. Ils ont de plus montré que l'incubation *in cellulo* de la séquence d'ADN et du ligand 360A, connu pour interagir fortement et sélectivement avec les structures quadruplexes, entraînait des changements drastiques

du spectre RMN pouvant être attribués à différents phénomènes comme de la précipitation ou une augmentation des constantes d'affinités de certaines enzymes. Cette observation montre la difficulté de l'étude directement *in cellulo* de l'interaction des G-quadruplexes avec des petites molécules et représente un vrai défi pour la compréhension des interactions ligands/G-quadruplex dans un milieu biologique complexe.

Ainsi, de nombreuses structures d'intérêt sont désormais connues grâce à ces différentes méthodes d'analyse. On retrouve notamment des structures G-quadruplexes provenant de séquences ADN et ARN jouant un rôle important : les promoteurs de gènes c-Myc^{58,59}, KRAS^{43,60}, VEGF⁵⁴, cKit (cKit1⁶¹, cKit2⁶² et cKit*⁶³) et Bcl2⁵¹, la séquence télomérique humaine⁶⁴, les minisatellites Ceb25⁶⁵ etc. Cependant, il existe toujours un besoin d'obtenir les structures précises des G-quadruplexes biologiquement intéressantes dans le but de pouvoir développer des molécules capables d'interagir sélectivement avec ces structures.⁶⁶

3. Localisation des G-quadruplexes au sein du génome

a. Localisation par analyses in silico

Depuis la découverte des structures G-quadruplexes, de nombreuses études ont tenté de prouver leur formation au sein du génome humain. Ainsi, plusieurs études bio-informatiques ont démontré que les séquences pouvant potentiellement adopter des structures G-quadruplexes (PQS, pour Putative G-Quadruplex Sequence) n'étaient pas réparties uniformément au sein du génome.

Les études bio-informatiques vont analyser un génome donné et calculer le nombre de PQS en se basant sur un certain nombre de critères comme le nombre guanines au sein des répétitions, le nombre de quartets et la longueur des boucles. Par exemple, le logiciel Quad-Parser⁶⁷ a permis de chercher les séquences pouvant adopter des structures G-quadruplexes de motif : $G_{3-5}N_xG_{3-5}N_xG_{3-5}$ $_5N_xG_{3-5}$ (avec N = A, T, C ou G ; X = 1-7) au sein du génome humain. Cette analyse a permis de prédire la possible formation d'environ 370 000 G-quadruplexes, un nombre confirmé la même année par une autre méthode d'analyse bio-informatique utilisant la recherche de motifs similaires.⁶⁸ Le problème majoritaire des analyses bio-informatiques vient du grand polymorphisme structurel inhérent aux G-quadruplexes. En effet, certains G-quadruplexes possédant des boucles bien plus longues que 7 nucléotides peuvent se former *in vitro* et être très stables⁶⁹ et certains G-quadruplexes à deux quartets⁷⁰ peuvent également se former. Finalement, cette méthode ne peut pas non plus prendre en compte les G-quadruplexes possédant des bulges⁷¹, les G-quadruplexes formés par des séquences comprenant plus de quatre motifs de répétition de guanines ou la présence de cytosines pouvant former des liaisons hydrogène de Watson-Crick et empêcher la formation des liaisons hydrogène de Hoogsteen.

Dans le but d'être le plus précis possible pour la détection des G-quadruplexes à travers le génome, d'autres logiciels bio-informatiques ont été développés comme G4P Calculator⁷², QuadBase⁷³; G4Hunter⁷⁴ et d'autres^{75,76}. Par exemple, G4Hunter permet de vérifier la richesse en guanines d'un brin du génome et de donner un score de probabilité de formation de structures quadruplexes en faisant varier un seuil d'acceptabilité. Par exemple, un seuil élevé (supérieur à 1,7) dénombrera relativement peu de G-quadruplexes à travers le génome, mais avec peu de résultats faux-positifs tandis qu'un seuil faible (inférieur à 1) dénombrera une grande quantité de G-quadruplexes avec plus de faux-positifs. Cette méthode permet de dénombrer entre deux et dix fois plus de G-quadruplexes à travers le génome, taix fois plus de G-quadruplexes à travers le génome et dix fois plus de G-quadruplexes de faux-positifs. Cette méthode permet de dénombrer entre deux et dix fois plus de G-quadruplexes à travers le génome, taix fois plus de G-quadruplexes de travers le génome entre deux et dix fois plus de G-quadruplexes de travers le génome que les différentes méthodes citées précédemment.

Finalement, d'autres méthodes de calculs utilisant l'intelligence artificielle ont été développées dans le but d'apprendre et de faire évoluer les logiciels pour être le plus précis possible quant à la détermination de formation de PQS au sein d'ADN génomique⁷⁷ et d'ARN⁷⁸.

Ces différentes méthodes de calculs ont permis de révéler une répartition non homogène des PQS à travers le génome suggérant des implications biologiques importantes. En effet, les PQS semblent tout particulièrement surreprésentés au niveau des sites d'initiations de la transcription (TSS, pour Transcription Start Sites) correspondant au premier nucléotide de l'ADN qui est transcrit en ARN. Plus particulièrement, les PQS sont surreprésentées au niveau des promoteurs, des régions d'ADN génomiques d'environ 1000 bases présents en amont des TSS sur lesquels vont se fixer les ARN polymérases. Ces promoteurs sont essentiels à la transcription de l'ADN en ARN et environ 40% de ces derniers possèdent au moins une séquence pouvant potentiellement former un G-quadruplex avec un enrichissement d'un facteur supérieur à 6 comparé avec le reste du génome⁷⁹. Les PQS sont de plus surreprésentés dans les promoteurs de proto-oncogènes : des gènes dont les mutations sont associées à la survenue de cancers comme les gènes RAS⁸⁰ ou c-Myc⁸¹. Environ 66% de ces promoteurs de proto-oncogènes possèdent au moins une PQS, ce qui est significativement plus élevé que pour les autres promoteurs de gènes⁷⁴.

34

Parmi tous ces logiciels développés jusqu'à présent et malgré l'aide des avancées technologiques, il est toujours très difficile d'envisager la détermination exacte et précise du nombre de Gquadruplexes formé dans un contexte aussi complexe que celui de la chromatine. Si les Gquadruplexes peuvent se former lorsque l'ADN est sous forme simple brin, il est certain que le nombre de PQS calculé par les différents logiciels sera toujours surestimé par rapport au nombre réel au sein de la chromatine. Ainsi, certaines techniques expérimentales ont été développées pour déterminer la formation de structures G-quadruplexes au sein du génome.

b. Localisation par analyses expérimentales in vitro

i. Localisation par inhibition de polymérases

Les polymérases sont des enzymes capables de catalyser la synthèse d'un brin complémentaire d'ADN ou d'ARN à partir d'une matrice de référence. Lorsque la matrice de référence est un brin d'ADN, l'enzyme capable de synthétiser le brin d'ADN complémentaire est une ADN polymérase et l'enzyme capable de synthétiser le brin d'ARN complémentaire est une ARN polymérase. Lorsque la matrice de référence est un brin d'ARN, la synthèse du brin d'ADN complémentaire est assurée par une transcriptase inverse. La synthèse de l'ADN ou de l'ARN peut être inhibée par différents phénomènes biologiques et notamment par l'apparition de structures secondaires comme les G-quadruplexes qui vont agirent comme des barrages à la progression de l'enzyme sur ces brins d'ADN⁸² ou d'ARN⁸³ (Figure 21). Cette capacité des G-quadruplexes à inhiber la progression des polymérases couplée à du séquençage a été utilisée dans le but de localiser l'apparition de G-quadruplexes au sein d'un brin d'ADN⁸⁴ (G4-seq) ou d'ARN⁸⁵ (rG4-seq).




Pour ces études, un brin d'ADN génomique de lymphocyte B ou d'ARN de cellules cancéreuses HeLa a été isolé, purifié et immobilisé sur une surface. Un premier séquençage de référence a été réalisé à l'aide d'ADN polymérases pour les brins d'ADN et de transcriptases inverses pour les brins d'ARN en conditions ne stabilisant pas les G-quadruplexes (LiCl ou NaCl). Un deuxième séquençage a ensuite été réalisé après stabilisation sélective des G-quadruplexes par des ions K⁺ ou par le ligand PDS. Les deux séquençages ont ensuite été comparés et les sites de blocage de la polymérase ont été associés à la formation de motifs G-quadruplexes. Ces analyses ont pu prédire la formation d'environ 700 000 et 525 000 structures G-quadruplexes à ADN au sein du génome humain en présence de PDS et de K⁺ respectivement. L'analyse du transcriptome humain a montré la formation d'environ 13 400 et 3 800 motifs G-quadruplexes à ARN en conditions PDS+K⁺ ou K⁺ seul respectivement. Ces analyses permettent de confirmer les prédictions faites par les analyses in silico avec un grand nombre de Gquadruplexes pouvant se former au niveau des promoteurs de gènes, des régions intragéniques d'oncogènes, des régions 5' et 3' non-traduites d'ARN (UTR pour Untranslated Region) ou des régions micro-ARN. Ces études révèlent une répartition hétérogène suggérant un rôle important de régulation de l'expression génétique par les G-quadruplexes. Cependant, ces études sont réalisées sur des brins d'ADN ou d'ARN extraits et purifiés et ne prennent donc pas en compte le contexte cellulaire avec les différentes protéines pouvant s'associer aux G4s.

ii. Localisation par réactions biochimiques

Il est possible de localiser la formation d'un G-quadruplex ou d'une autre structure non canonique de l'ADN ou de l'ARN en exploitant la différence de réactivité et d'accessibilité des bases impliquées dans la formation de ces structures. Par exemple, il est possible d'oxyder sélectivement les thymines non appariées d'un simple brin d'ADN par réaction avec du KMnO₄. Cette réaction sur la thymine permet de bloquer la forme simple brin et le rendre particulièrement sensible aux nucléases S1⁸⁶, des enzymes capables de cliver sélectivement les liaisons phosphodiester de simples brins d'ADN ou d'ARN. Cette technique a ainsi pu être couplée à une analyse *in silico* permettant de prévoir la formation de structures non canoniques et plus particulièrement de structures G-quadruplexes⁸⁷. Cette étude a permis de confirmer que les G-quadruplexes étaient surreprésentés au niveau des régions régulatrices de gènes comme les TSS.

Dans le cas de l'ARN, il a été démontré que la nucléophilie des groupements hydroxyles en position 2' dépendait directement de la structure secondaire de cet ARN. En effet, une base ARN contrainte dans une certaine conformation par appariement avec une autre base verra l'électronégativité du groupement 2'-OH réduit comparé à une base libre⁸⁸. Partant de cette observation, l'acylation

36

sélective des fonctions 2'-OH des nucléotides libres a permis développer une méthode permettant de déterminer la structure secondaire de brins d'ARN⁸⁹ (SHAPE pour *Selective 2'-Hydroxyl Acylation analyzed by Primer Extension*). Cette méthode implique l'acylation des groupements 2'-OH par la N-methylisatoic anhydride (NMIA) des bases non contraintes, inhibant par la suite la réaction d'une transcriptase inverse, permettant ainsi de cartographier les sites d'acylation et d'en déduire la structure secondaire du brin d'ARN (Figure 22). Cette méthode a été appliquée à différentes structures d'ARN d'intérêts⁹⁰ et une méthode optimisée pour les structures G-quadruplexes a récemment été développée⁹¹.



Figure 22: Différence de réactivité des groupements 2'-OH d'un nucléotide contraint et d'un nucléotide libre face à un électrophile.

iii. Localisation par Immunoprécipitation de la chromatine (ChIP)

Dans le but de localiser les structures G-quadruplexes au sein du génome en gardant le milieu complexe de la chromatine, une méthode couplant la reconnaissance des structures G-quadruplexes par un anticorps sélectif (BG4) avec le séquençage à haut débit a été développée⁹² (G4-ChIP-seq). Dans cette méthode, BG4⁹³, est incubé avec de la chromatine extraite de cellules humaines. La chromatine est ensuite fragmentée et les fragments d'ADN reconnus par BG4 sont immunoprécipités avant d'être analysés par séquençage à haut débit. Cette méthode a permis de dénombrer environ 10 000 structures G-quadruplexes au sein de cellules kératinocytaires épidermiques humaines précancéreuses HaCaT, soit environ 1,4% des 700 000 structures G-quadruplexes déterminées par G4-seq précédemment cité⁸⁴. Ce résultat indique donc clairement l'importance de l'état de la chromatine pour la formation de structures G-quadruplexes. En effet, si la chromatine est décondensée par incubation d'un inhibiteur d'histone désacétylase, le nombre de structures G-quadruplexes détectées entre les cellules saines et les cellules cancéreuses.

En effet, cette méthode a montré qu'uniquement 1 000 G-quadruplexes pouvaient être détectés en cellules kératinocytaires épidermiques humaines normales (NHEK). Encore une fois, la répartition des G-quadruplexes déterminée par G4-ChIP-seq semble non homogène avec une quantité plus élevée au niveau des promoteurs de gènes à activité transcriptionnelle élevée ainsi qu'au niveau des oncogènes et proto-oncogènes. Cette méthode a permis de mettre en lumière l'importance de la structure de la chromatine pour la formation de G-quadruplexes. Cependant, il est important de noter que cette méthode requiert la fixation des cellules, ce qui impacte obligatoirement la dynamique de formation des G-quadruplexes.

Plus récemment, une protéine (G4P⁹⁴) a été développée à partir du domaine de reconnaissance des G4s⁹⁵ de l'hélicase RHAU, une hélicase à ARN connue pour interagir et déstructurer les Gquadruplexes⁹⁶. La protéine G4P est composée de 64 acides aminés pour un poids moléculaire d'environ 6,7 kDa et interagit avec différentes structures G-quadruplexes avec un K_D de l'ordre du nM tandis qu'aucune interaction n'est détectée avec des séquences d'ADN duplex. Alors qu'une étape de fixation et de perméabilisation est nécessaire pour pouvoir incuber des anticorps spécifiques des Gquadruplexes in cellulo, il a été possible d'exprimer cette protéine directement in vivo après transfection d'un plasmide au sein des cellules cancéreuses A549. L'avantage est donc de pouvoir identifier la formation des G-quadruplexes dans un contexte de cellules vivantes. L'ADN lié aux protéines G4P a ensuite pu être isolé par immunoprécipitation et une étape de séquençage a permis de détecter environ 123 000 séquences différentes. Des analyses bio-informatiques ont ensuite permis de corréler ces pics avec la formation de PQS. Environ 1/3 des 123 000 pics montrait un enrichissement attribué à la formation de structures G-quadruplexes canoniques dans 60% des promoteurs et 70% des gènes. Cette méthode a permis de détecter un nombre plus élevé de structures G-guadruplexes in vivo comparé à l'immunoprécipitation de la chromatine à l'aide de BG4 en cellules fixées. BG4 et G4P possèdent une affinité similaire envers les G-quadruplexes et cette différence du nombre de G-quadruplexes détectés peut provenir du fait que G4P entre en compétition avec les protéines in vivo tandis que BG4 ne peut déplacer les protéines bloquées par l'action du paraformaldéhyde. Il est de plus possible que la présence de G4P in vivo induise la formation de structures G-quadruplexes, augmentant ainsi le nombre de séquences isolées. Finalement, il important de noter que ces deux études ont été réalisées sur deux lignées cellulaires différentes, ce qui pourrait aussi expliquer cette différence.

Ainsi, les différentes méthodes de localisation des G-quadruplexes à travers le génome ont montré une répartition non homogène de ces séquences, suggérant de possibles rôles au sein du vivant.

4. Rôles des G-quadruplexes

a. Les télomères

Les télomères sont des brins d'ADN non codants situés aux extrémités des chromosomes et ont été les premiers à être suspectés de former des G-quadruplexes. Dans les cellules d'animaux vertébrés, les télomères sont composés de répétitions (TTAG₃) d'une longueur allant de 2kb à environ 10kb et dont l'extrémité 3' est prolongée par un simple brin d'ADN d'environ 100-150 bases possédant la même répétition (TTAG₃, Figure 23A) pouvant se replier et envahir le brin complémentaire en formant une boucle T et une boucle D (Figure 23B). Cette structure a été détectée la première fois par microscopie électronique⁹⁷ et a été confirmée par la suite^{98,99}. Les télomères sont étroitement liés à plusieurs protéines formant un complexe nucléoprotéique nommé « shelterin»¹⁰⁰. Ce dernier permet de protéger les chromosomes de la dégradation¹⁰¹, de la fusion « end to end »¹⁰² et empêche l'extrémité linéaire des chromosomes d'être reconnue comme des cassures simples ou doubles brins¹⁰³.



Figure 23: Représentation schématique de A) l'ADN télomérique présent à l'extrémité des chromosomes et B) la structure en lasso formée par l'invasion de l'extrémité 3' sur le brin complémentaire (D-loop).

Dans les années 1960, L. Hayflick démontre que les cellules somatiques ne sont capables de se diviser qu'un nombre fini de fois (entre 50 et 70) avant d'entrer en sénescence : c'est la limite de Hayflick¹⁰⁴. Plus tard, cette observation a été corrélée avec le fait que les télomères se raccourcissent à chaque division cellulaire : lorsque le brin d'ADN télomérique devient trop court, la cellule entre en sénescence. Cette sénescence programmée des cellules somatiques agit comme une barrière au développement de cellules tumorales qui, elles, se multiplient à l'infini. Cette différence de comportement est aujourd'hui connue pour être intimement liée à la télomérase, une enzyme capable d'allonger l'ADN télomérique. Cette enzyme est fortement produite dans les cellules cancéreuses et quasi absente des cellules somatiques et participe ainsi à l'immortalisation de 85 à 90% des cancers¹⁰⁵.

L'ADN télomérique peut former de nombreuses structures G-quadruplexes et ces dernières étaient initialement connues pour participer à l'inhibition de la télomérase¹⁰⁶. De nombreuses petites molécules capables de stabiliser les structures G-quadruplexes ont été spécifiquement développées dans le but d'inhiber l'activité de la télomérase dans les cellules cancéreuses. En effet, la sous-unité hTERT de l'enzyme qui permet d'ajouter des bases d'ADN à l'extrémité 3' du brin d'ADN télomérique ne peut interagir avec ce dernier que lorsqu'il est sous forme simple brin. Par conséquent, stabiliser le dernier motif G-quadruplex du brin d'ADN télomérique par des petites molécules permet d'inhiber l'activité de la télomérase. L'activité anticancéreuse des ligands G-quadruplexes est aujourd'hui plutôt attribuée à la capacité de ces derniers à déplacer certaines protéines du « shelterin complex »; induisant des dommages à l'ADN et menant à la sénescence, l'inhibition de la croissance et l'apoptose des cellules cancéreuses¹⁰⁷. Plus récemment, il a été mis en évidence que la télomérase était capable de partiellement dérouler les structures G-quadruplexes parallèles intermoléculaires de la séquence télomérique humaine et d'en allonger l'extrémité *in vitro*¹⁰⁸. Cette étude suggère que les structures G-quadruplexes de l'ADN télomérique ne permettent pas juste de réguler l'activité de la télomérase.

b. G-quadruplexes génomiques

Comme vu précédemment, les séquences pouvant potentiellement former des G-quadruplexes sont retrouvées à travers le génome avec une surreprésentation aux alentours des TSS et plus particulièrement au niveau des TSS d'oncogènes et de proto-oncogènes. Ces structures ne peuvent se former que lorsque le brin d'ADN n'est pas apparié au brin complémentaire ce qui peut impliquer un rôle lors de différents phénomènes comme la réplication, la transcription et la traduction.



Figure 24: Illustration représentant les possibles implications des G-quadruplexes lors de phénomènes biologiques comme A) la réplication, B) la transcription et C) la traduction. Image tirée et modifiée de Nakanishi et Seimiya¹⁰⁹.

Ainsi, l'apparition de structures G-quadruplexes lors du processus réplicatif va permettre de réguler la synthèse de l'ADN (Figure 24A). En effet, la formation d'une structure quadruplex lors de la réplication de l'ADN peut bloquer l'avancée de l'ADN polymérase ou au contraire aider à démarrer la réplication^{110,111}. La formation de structures G-quadruplexes peut de plus obstruer la progression de la fourche de réplication, induisant des cassures doubles brins qui peuvent mener à une instabilité génomique¹¹².

Dans le cas de la transcription, les G-quadruplexes peuvent soit diminuer soit augmenter la synthèse de l'ARN en fonction de leur positionnement (Figure 24B). Les G-quadruplexes étaient initialement reconnus pour leur capacité à bloquer l'avancée d'ARN polymérases, mais plusieurs études montrent que la formation de ces structures peut initier et/ou augmenter la transcription d'un gène ^{113,114}. Les G-quadruplexes jouent donc un rôle très important dans la régulation de la transcription et leurs surreprésentations au niveau des promoteurs d'oncogènes en font des cibles thérapeutiques attrayantes¹¹⁵. En effet, des motifs G-quadruplexes peuvent être retrouvés au sein des oncogènes participants aux 6 caractéristiques nécessaires au développement et l'immortalisation des cellules cancéreuses. Par exemple, les gènes c-Myc¹¹⁶ et KRAS⁴³ sont responsables de l'autosuffisance des cellules cancéreuses, le gène Bcl-2¹¹⁷ permet aux cellules cancéreuses d'éviter le processus apoptotique, le gène hTERT¹¹⁸ est responsable de la prolifération illimitée, etc. (Figure 25)



Figure 25 : Représentation schématique des 6 caractéristiques nécessaires au développement de cellules cancéreuses associées aux structures G-quadruplexes retrouvées au niveau des promoteurs des gènes. Figure adaptée de Hurley et al.¹¹⁷

Par exemple, il a été démontré que le gène c-Myc, responsable de la prolifération des cellules et qui est surexprimé dans 80% des cellules cancéreuses est transcrit à partir du promoteur NHEIII₁ (Nuclease Hypersensitive element III₁) qui possède une séquence formant plusieurs structures G4 *in vitro*¹¹⁹. La modification de ce promoteur par une séquence mutée ne formant pas de motif G-quadruplexes a montré une surexpression du gène c-Myc, démontrant la capacité des G-quadruplexes à réguler la transcription d'un gène¹²⁰.

Enfin, après transcription d'un gène riche en guanines, le brin d'ARN en résultant peut potentiellement former des G-quadruplexes (Figure 24C). Les différentes techniques précédemment citées cherchant à localiser les structures G-quadruplexes au sein du génome ont révélé une surreprésentation de ces structures au niveau des régions 5'-UTR et 3'-UTR comparé aux régions codantes¹²¹. Notamment, les 5'-UTR sont des régions de l'ARN messager situées en amont du site de début de traduction et sont reconnues par le ribosome afin de démarrer la synthèse de protéines. Ces régions 5'-UTR jouent un rôle majeur dans la régulation de la traduction et il a été démontré que la présence de G-quadruplexes au sein de ces derniers pouvait réduire drastiquement la traduction de l'ARNm. Par exemple, la région 5'-UTR du proto-oncogène NRAS est composée de 254 nucléobases et possède la séquence 5'-G₃AG₄CG₃UCUG₃-3' pouvant former une structure G-quadruplex stable¹²². Dans le but de vérifier si cette structure G-quadruplex pouvait avoir un impact sur l'efficacité de la traduction, S. Kumari *et al.* ont muté cet ARNm pour qu'il ne puisse pas former

de structures G-quadruplexes. La traduction *in vitro* par du lysat de réticulocytes de lapin a montré une efficacité de traduction 3,6 fois supérieure pour l'ARNm ne possédant pas de motif Gquadruplex, démontrant la capacité de ces derniers à réguler le phénomène de traduction. Cette étude était la première à démontrer la capacité des G4s à réguler le phénomène de traduction et a été confirmée par la suite¹²³.

5. Ligands des G-quadruplexes

a. Structures et interactions

Comme vu précédemment, les G-quadruplexes sont responsables de la régulation de nombreux phénomènes biologiques et permettent notamment l'inhibition de la télomérase et la transcription d'oncogènes et de proto-oncogènes. Dans le but de moduler leur activité et de mieux comprendre leurs implications biologiques, de nombreuses petites molécules ont été développées pour cibler les G-quadruplexes.

La structure en hélice à double brin de l'ADN étant la forme la plus courante au sein du vivant, les ligands des G-quadruplexes doivent être extrêmement sélectifs de ces derniers. Pour ce faire, plusieurs modes d'interactions peuvent assurer cette sélectivité. Les G-quadruplexes, comme vu dans l'introduction, possèdent des sillons de tailles différentes en fonction de l'orientation des guanines, des boucles de tailles et de natures différentes ainsi qu'un squelette glycophosphate. La plus grande particularité des structures G-quadruplexes réside néanmoins dans la large surface aromatique riche en électrons formée par l'interaction de quatre guanines et c'est cette particularité qui est exploitée dans la conception de molécules sélectives vis-à-vis de l'ADN duplex. Les molécules interagissant avec un quartet externe des G4 sont souvent des molécules possédant une large surface aromatique déficitaire en électrons qui interagissent avec ce dernier par π -stacking ainsi que par des interactions de van der Waals, des interactions hydrophobes et électrostatiques²⁴. Cette surface aromatique est généralement plus large que celle des intercalants classiques de l'ADN dans le but d'empêcher ces derniers d'interagir avec les sillons de l'ADN duplex. Ces ligands peuvent de plus être fonctionnalisés par des chaînes latérales protonées à pH physiologique comme des amines dans le but d'augmenter la solubilité et de réaliser des interactions électrostatiques avec les groupements phosphates des nucléotides présents au niveau des boucles. Optimiser chacune de ces interactions permet de développer des ligands qui ont à la fois une forte constante d'affinité pour les structures G4 et une grande sélectivité par rapport aux autres structures de l'ADN. Un nombre très élevé de ligands a été synthétisé à ce jour, certains seront présentés ici dans le but de décrire les différents modes d'interactions ainsi que les effets de ces ligands *in cellulo*.

Trois ligands de type bisquinolines particulièrement affins des G-quadruplexes sont présentés ciaprès (Figure 26).



Figure 26: Exemple de trois ligands bisquinolines, PDS, PDC et PhenDC3.

Ces trois ligands possèdent des structures planes en forme de V pour la pyridostatine (PDS) et la pyridodicarboxamide (PDC) et en forme de U pour la phénantrolinedicarboxamide (PhenDC3). Cette structure particulière est stabilisée par les liaisons hydrogène intramoléculaires des fonctions amides avec l'azote de la pyridine pour la PDS et le PDC et avec les deux azotes de la phénantroline pour le PhenDC3¹²⁴. Ces structures planes assurent une interaction particulièrement efficace avec les quartets externes des G-quadruplexes.

Le ligand PDS¹²⁵ est un ligand sélectif des structures G-quadruplexes et est régulièrement utilisé comme référence dans les tests cellulaires. Son large squelette aromatique ainsi que ses trois chaines amines chargées à pH physiologique permettent d'interagir avec les quartets externes des G-quadruplexes et avec les nucléotides présents dans les boucles. Le premier effet cellulaire décrit pour la PDS était sa capacité à induire des dommages à l'ADN en affectant l'intégrité du complexe « shelterin » de l'ADN télomérique¹²⁵. Plus tard, la PDS a montré la capacité d'arrêter la prolifération de cellules cancéreuses en ciblant spécifiquement les gènes à fort taux de guanines comme démontré par analyse ChIP-seq¹²⁶. Le ligand PDC¹²⁷ est un ligand de type bisquinolinium très affin des G-quadruplexes avec une très bonne sélectivité vis-à-vis de l'ADN duplex. Ce ligand est capable d'inhiber l'activité de la télomérase *in vitro* et d'induire de nombreuses dysfonctions télomériques au sein de plusieurs lignées cellulaires de tumeurs cérébrales¹²⁷. La modification de ce cycle pyridine central par un cycle phénantroline a permis de former le ligand PhenDC3, permettant un recouvrement parfait du quartet externe de la séquence Pu24T, dérivée de l'oncogène c-Myc, comme démontré par analyse RMN¹²⁸ (Figure 27). Le composé PhenDC3 est, à ce jour, l'un des meilleurs ligands G4s. Il possède une affinité nanomolaire (K_p) pour de nombreuses structures G-

quadruplexes et a montré la capacité d'inhiber l'action de l'hélicase Pif1 sur le minisatellite G4-Ceb1 *in vitro*. L'insertion du minisatellite G4-Ceb1 dans le génome de levures de boulanger induit une instabilité génomique en absence de l'hélicase Pif1 et est exacerbée lors de l'ajout de PhenDC3, démontrant la capacité de ce dernier à stabiliser les structures G-quadruplexes de la séquence G4-Ceb1 *in vivo*¹²⁹. Ces travaux pionniers ont montré la capacité du PhenDC3, et plus largement des ligands G4s, à cibler une séquence G4 dans un organisme modèle.

Ces trois ligands PDS, PDC et PhenDC3 représentent aujourd'hui les ligands des G-quadruplexes les plus populaires et sont régulièrement utilisés comme références pour des tests biophysiques et biologiques.





Certains ligands macrocycliques ont été développés dans le but d'interagir sélectivement avec les Gquadruplexes. C'est notamment le cas de la telomestatine et de certains de ses dérivés comme L2H2-6M et le ligand non cyclique TOxaPy (Figure 28).





La telomestatine est un macrocycle naturel composé de 5 cycles oxazole, deux cycles methyloxazole et un cycle thiazoline qui a été isolée en 2001 à partir d'une bactérie¹³⁰. Cette molécule, contrairement à celles présentées ci-dessus est non-chargée, mais interagit tout de même par interaction π -stacking avec les quartets externes. La telomestatine possède une certaine sélectivité pour l'ADN télomérique et est capable d'inhiber fortement la télomérase¹³¹. La telomestatine étant une molécule naturelle relativement complexe, sa synthèse est la principale limitation à son utilisation et d'autres dérivés plus simples ont été développés. Ainsi, le ligand TOxaPy est un dérivé acyclique de la telomestatine composé de quatre cycles oxazole, deux cycles pyridines et un cycle benzénique. Ce ligand possède une bonne sélectivité pour la topologie antiparallèle de l'ADN télomérique comparé à la topologie hybride en interagissant notamment avec les sillons de ce dernier¹³². De nombreux autres dérivés de la telomestatine ont été synthétisés par l'équipe de Nagasawa¹³³. Notamment, le ligand L2H2-6M(2)OTD est un dérivé cyclique de la telomestatine fonctionnalisé par 2 amines libres pouvant se protoner à pH physiologique. Le complexe formé de ce dérivé avec une structure de l'ADN télomérique humain a été déterminé par RMN¹³⁴ et a montré une interaction par π-stacking ainsi que deux interactions électrostatiques par les amines latérales (Figure 29).



Figure 29: Structure du complexe formé par le ligand L2H2-6M(2)OTD et la séquence télomérique TT(G₃TTA)₃G₃A déterminé par RMN vu de côté (gauche) et du dessus (droite). PDB : 2MB3¹³⁴.

Les dérivés de porphyrines représentent une catégorie importante de ligands des G-quadruplexes. La large surface aromatique chargée caractéristique des porphyrines permet généralement d'interagir fortement avec les quartets externes par π -stacking (Figure 30).



Figure 30: Exemple de trois ligands des G-quadruplexes de la famille des porphyrines (TmPyP4 et NMM) et Porphyrazine (3,4-TMPyPz).

Parmi ces derniers, TmPyP4¹³⁵ est un dérivé porphyrine fonctionnalisé par quatre Nmethylpyridinium et est largement utilisé comme référence dans différents tests biophysiques. Cette molécule possède de plus la capacité d'induire la formation de ROS (pour *reactive oxygen species*) après irradiation aux UVA. La génération de ROS par TmPyP4 permet l'oxydation des résidus guanines dans les cellules leucémiques humaines et d'en induire l'apoptose¹³⁶. TmPyP4 interagit avec les quartets externes des G-quadruplexes comme démontré par RMN sur la séquence Pu24I, dérivée de l'oncogène c-Myc⁵⁸ (Figure 31) ou sur le dimère de la séquence télomérique humaine à ARN TERRA¹³⁷.



Figure 31: Structure du complexe formé par le ligand TmPyP4 et la séquence Pu24I dérivée de l'oncogène c-Myc déterminé par RMN vu de côté (gauche) et du dessus (droite). PDB : 2A5R⁵⁸.

Cependant, le composé TmPyP4 n'est pas sélectif des G-quadruplexes vis-à-vis des autres structures de l'ADN comme démontré par J. Ren et J. B. Chaires¹³⁸. Cette pauvre sélectivité a mené au développement de nouvelles porphyrines comme le composé 3,4-TMPyPz¹³⁹. Ce dernier interagit fortement avec l'ADN télomérique humain avec une stœchiométrie 1:1 comme le TmPyP4 et a

montré la capacité de favoriser la formation de la topologie antiparallèle de la séquence hybride Htelo. Des titrations UV-Vis ont déterminé un K_D d'environ 200 nM pour le composé 3,4-TMPyPz contre environ 14 µM pour le composé TmPyP4 démontrant une nette optimisation des interactions avec ce G-quadruplex. Le dérivé N-methyl mesoporphyrine IX¹⁴⁰ (NMM) est une porphyrine ne possédant pas de charges positives contrairement aux deux autres composés présentés, mais montre une sélectivité particulière pour l'ADN G-quadruplex parallèle par rapport aux autres structures de l'ADN¹⁴¹. Curieusement, le K_D de NMM avec la séquence Tel22 montre une affinité relativement modeste (K_D \simeq 10 µM), mais présente une capacité à stabiliser fortement cette dernière contre la dénaturation thermique¹⁴². Une intéressante particularité de NMM est sa faible fluorescence en solution qui est fortement exacerbée lorsqu'il interagit sélectivement avec des structures Gquadruplexes, permettant de nombreuses applications de fluorescence^{143,141}. Le complexe de NMM en interaction avec la séquence télomérique Tel22 a été déterminé par DRX et a montré une interaction avec le quartet par le macrocycle ainsi qu'une interaction électrostatique par la fonction carboxylate avec le sillon. Le groupement méthyle est de plus orienté vers le centre du tétrade permettant d'optimiser les interactions hydrophobes¹⁴⁴ (Figure 32).



Figure 32: Représentation du complexe NMM-Tel22 déterminé par DRX vu de côté (gauche) et vu du dessus (droite). PDB : 4FXM.

Un problème rencontré dans la recherche de nouvelles molécules interagissant avec les structures Gquadruplexes est le manque de sélectivité pour une structure particulière. En effet, bien que les Gquadruplexes possèdent un grand polymorphisme structurel, le principal mode d'interaction des ligands se fait par reconnaissance du quartet externe qui est très similaire d'une séquence à une autre. Ainsi, dans le but de développer des ligands sélectifs d'une structure G-quadruplex, plusieurs molécules ont été imaginées et synthétisées (Figure 33).



Figure 33: Exemples de ligands G-quadruplexes développés dans le but de discriminer une structure G-quadruplex parmi d'autres avec un dimère du ligand 360A, le ligand TH3 et les molécules hybrides cIKP-hPIP.

Par exemple, certaines séquences comme les télomères peuvent former un enchainement de plusieurs structures G-quadruplexes. Par conséquent, certaines molécules ont été développées dans le but d'interagir spécifiquement avec ces G-quadruplexes multimériques. Récemment, des dimères du ligand 360A séparés par différentes longueurs de linker ont été synthétisés et leur capacité à se lier à l'ADN télomérique a été étudiée¹⁴⁵. Notamment, le dimère dont le linker est composé de deux motifs Peg (360A-n₂) a montré des constantes d'affinités 30 à 425 fois supérieures pour le dimère de l'ADN télomérique humain G2T1 comparé aux autres séquences monomériques et dimériques étudiées. Il a été ainsi montré que ce ligand était capable de se lier avec une stœchiométrie 2:1 (360A-n₂:ADN) pour une séquence formant deux G-quadruplexes consécutifs. Il a alors été supposé qu'un dimère 360A-n₂ était capable de se lier aux deux quartets des deux G-quadruplexes adjacents. De plus, ce ligand est capable d'inhiber l'action de la télomérase avec un IC₅₀ de 4,6 μM contre un IC₅₀ de 9 μM pour le monomère 360A.

Comme décrit précédemment, le gène c-Myc est un proto-oncogène connu pour participer à la croissance incontrôlée des nombreuses tumeurs et est surexprimé dans de nombreuses cellules cancéreuses comme les cancers du sein, du côlon, des poumons etc¹¹⁶. Le gène c-Myc est aujourd'hui reconnu comme étant une importante cible thérapeutique, mais aucun inhibiteur de ce gène n'a pour l'instant été capable d'atteindre les essais cliniques¹⁴⁶. Cibler directement la protéine c-Myc

semble de plus difficilement envisageable, car elle ne propose pas de site actif bien défini et possède un temps de vie court d'environ 20-30 min. La découverte de la formation d'une structure Gquadruplex intramoléculaire au sein du promoteur de c-Myc a mené à de nombreux efforts pour sélectivement interagir avec ce dernier et de nombreuses molécules ont été développées¹⁴⁷. Par exemple, le composé TH3¹⁴⁸ est un thiazole polyamide analogue structurel de la distamycin A¹⁴⁹, un ligand non sélectif des G-quadruplexes interagissant avec les quartets externes et les sillons. La molécule TH3 est capable de stabiliser une séquence dérivée de c-Myc très efficacement (22 °C déterminé par FRET-melting) comparé à d'autres séquences dérivées d'autres promoteurs comme cKit1, cKit2 et BCL-2. De plus, l'addition de 50 équivalents d'ADN duplex n'a pas modifié les températures de stabilisation, indiquant une faible affinité pour l'ADN double brin. Une titration par fluorescence a montré un K_D d'environ 0,25 µM pour la séquence Myc22 et un K_D de 1,17 à 1,34 µM pour les autres structures G-quadruplexes, démontrant une sélectivité 4 à 5 fois supérieures pour la structure Myc. L'interaction de TH3 avec la séquence Myc22 a été étudiée par RMN et a montré une interaction avec les deux quartets externes en 3' et 5' avec une stœchiométrie 2:1 (TH3:ADN). Enfin, l'incubation de TH3 dans les cellules cancéreuses HeLa et A549 a montré une activité antiproliférative qui a été attribuée au niveau plus faible d'expression du gène c-Myc.

Toujours dans le but d'interagir sélectivement avec le gène c-Myc, des ligands hybrides capables d'interagir à la fois avec une structure G-quadruplex et avec la séquence d'ADN duplex à proximité ont été développés. Notamment, des hybrides cIKP-hPIP (pour cyclic imidazole/lysine polyamide - hairpin pyrrol/imidazole polyamide) ont été synthétisés par le groupe de Sugiyama.¹⁵⁰ Ces molécules hybrides contiennent un motif cyclique cIKP capable d'interagir avec les structures G-quadruplexes¹⁵⁰ et un motif polyamide hPIP capable d'interagir sélectivement avec le petit sillon de la séquence duplex à proximité du G-quadruplex de l'oncogène c-Myc (Figure 34).



Figure 34: Illustration de l'interaction des ligands hybrides: la partie cIKP interagit avec la structure G-quadruplex et le motif hPIP interagit avec le sillon de l'ADN duplex à proximité de la structure G-quadruplex. Image tirée de Asamitsu *et al.*¹⁵⁰

Ainsi, trois séquences hPIP ont été synthétisées et testées pour leur affinité avec la séquence duplex à proximité du G-quadruplex c-Myc et deux ont été retenues pour être couplées avec la partie cIKP. Ces deux hPIP ont été couplés au motif cIKP par des linkers de tailles différentes pour donner quatre ligands finaux. Des titrations CD et des analyses par FRET-melting ont montré que les ligands étaient bien capables d'interagir à la fois avec la structure G-quadruplex et avec la structure duplex du gène c-Myc. Des analyses SPR ont de plus montré une bonne sélectivité pour la séquence G4/duplex d'intérêt (K_D = 3,4 nM) par rapport à la même séquence modifiée d'une paire de bases (K_D = 17 nM) pour une des molécules. Cette stratégie de molécules hybrides ouvre donc des possibilités pour le développement de ligands sélectifs d'autres structures G-quadruplexes présents dans les séquences d'autres oncogènes.

b. Méthodes pour l'étude des interactions ligands/G-quadruplexes

La synthèse de nouvelles molécules interagissant spécifiquement avec les structures G-quadruplexes implique donc d'étudier leurs différents modes d'interactions, leur affinité ainsi que leur sélectivité. Pour ce faire, différentes méthodes existent, bien souvent adaptées de celles utilisées pour l'étude des structures duplexes. Chacune de ces méthodes possèdent leurs avantages et leurs inconvénients, il est donc nécessaire d'en utiliser plusieurs en parallèle dans le but de comprendre au mieux les interactions entre les petites molécules et les structures G-quadruplexes. Certaines de ces méthodes sont présentées ici.

i. Analyses des propriétés physiques des ligands G-quadruplexes

1. Titration par UV-Vis

Lorsque les ligands des G-quadruplexes d'intérêt possèdent une bande d'absorption à la fois suffisamment intense et éloignée de l'absorbance de l'ADN G4, il est possible de réaliser des titrations dans le but d'obtenir des informations sur l'affinité du ligand avec la structure d'intérêt. Généralement, une titration UV-vis est réalisée par ajout d'une concentration croissante d'ADN ou d'ARN sur une concentration fixe de ligand et le spectre d'absorption est enregistré. Si le ligand interagit avec la structure G-quadruplex, son spectre d'absorption sera légèrement modifié et l'ajout croissant de G-quadruplex devrait mener à la formation d'un plateau correspondant à une stœchiométrie bien définie du complexe ligand:G4 (Figure 35). La courbe de titration résultant de la

51

variation de l'absorbance en fonction de la concentration est ensuite tracée pour calculer la constante d'affinité.



Figure 35: Exemple de titration UV-Vis d'un ligand G-quadruplex Cu(II)-Salphen (20 μM) par une séquence d'ADN télomérique humain (0 à 0,5 équivalent). La flèche noire montre la diminution du pic d'absorbance induit par l'ajout d'ADN. Image tirée et modifiée de Campbell *et al.*⁴⁴

2. Titration par fluorescence

Tout comme pour les analyses UV-Vis, la titration de fluorescence permet d'obtenir une valeur de constante d'association du ligand sur une structure G-quadruplex donnée. Dans ce cas, le ligand G-quadruplex doit avoir une fluorescence intrinsèque relativement faible lorsqu'il est libre en solution et une fluorescence exacerbée lorsqu'il interagit avec sa cible. Cela peut être le réalisé par différents mécanismes lors de l'interaction du ligand avec un G-quadruplex: une molécule dont la rotation libre autour d'une liaison clé peut être bloquée dans une conformation comme la thioflavin T¹⁵¹; la désagrégation par les G-quadruplexes d'auto-agrégats non fluorescents comme les naphtalenes diimides (NDI)¹⁵² ou comme certains squaraines¹⁵³ (Figure 36) ; ou inversement l'induction d'agrégats fluorescents par les G-quadruplexes comme les tetraphenylethenes¹⁵⁴.



Figure 36: Exemple de titration de fluorescence du composé SQgl (2 μM), dérivé de la famille des squaraines avec la séquence VEGF (0 à 4 équivalents). Image tirée et adaptée de Grande *et al.*¹⁵³

ii. Résonance Plasmonique de Surface (SPR)

La résonnance plasmonique de surface (SPR pour Surface Plasmon Resonance) a premièrement été introduite au début des années 1990 pour l'étude biochimique de nouvelles biomolécules¹⁵⁵ et a été adaptée par la suite à l'étude des G-quadruplexes¹⁵⁶. Succinctement, la SPR est une technique permettant de détecter les évènements d'interactions se produisant à la surface d'un film métallique généralement d'or ou d'argent. La lumière polarisée incidente est directement orientée vers cette surface à l'aide d'un prisme et l'angle de réflexion de la lumière réfléchie est détecté. Des structures G-quadruplexes peuvent être immobilisées sur cette surface métallique et l'interaction avec des petites molécules modifiera légèrement l'angle de réflexion de la lumière. Cette expérience est réalisée en flux ce qui permet de déterminer les caractéristiques cinétiques et thermodynamiques des ligands étudiés (Figure 37).



Figure 37: Schéma du principe d'analyse par SPR. Des structures G-quadruplexes sont immobilisées sur une surface et l'angle de réflexion est mesuré (1). Les ligands sont ensuite ajoutés au flux (2) jusqu'à atteindre un équilibre (3). L'ajout de ligands est arrêté et les ligands se dissocient (4) jusqu'à atteindre l'équilibre (5).

Cette méthode d'analyse est plus compliquée à mettre en œuvre que les titrations par UV-Vis et par fluorescence, mais possède l'avantage d'être rapide et de ne pas dépendre des propriétés spectroscopiques des ligands. C'est une technique régulièrement employée pour la détermination des constantes d'affinités de nombreux composés sur différentes structures et offre l'avantage de pouvoir déterminer une sélectivité potentielle entre différentes topologies de G-quadruplexes^{157,158}.

iii. Fluorescent Intercalator Displacement (FID)

Le FID est une méthode d'analyse biophysique permettant de déterminer indirectement l'affinité ainsi que la sélectivité d'une petite molécule vis-à-vis d'une séquence d'ADN ou d'ARN. Cette méthode a premièrement été développée en 2001 pour être appliquée à l'ADN double brin et visait alors à déterminer la fixation de petites molécules sur de nombreuses séquences d'ADN différentes¹⁵⁹. C'est une méthode facile à mettre en œuvre, peu coûteuse et qui permet de tester de nombreuses molécules sur différentes séquences. Cette méthode a ensuite été adaptée pour pouvoir être appliquée aux structures G-quadruplexes premièrement en cuve¹⁶⁰ puis par la suite en plaque de 96 à 384-puits pour l'analyse en haut débit¹⁶¹.

La méthode de FID est basée sur la différence de fluorescence d'une sonde lorsqu'elle interagit avec l'ADN (sonde on) et lorsqu'elle est libre en solution (sonde off). Cet effet repose ici sur la réduction du degré de liberté de la sonde lorsqu'elle interagit avec la structure G-quadruplex ce qui permet de restaurer l'émission de fluorescence après excitation à une longueur d'onde appropriée. Ajouter, par la suite, une concentration croissante de ligands des G-quadruplexes permet de réaliser une compétition entre la formation du complexe sonde:G4 et ligand:G4 (Figure 38). Si le ligand possède une constante d'affinité suffisamment élevée pour la structure G-quadruplex, il sera alors capable de déplacer la sonde et une diminution de fluorescence sera observée.



Figure 38: Schéma du principe de fonctionnement du FID : l'ajout d'un ligand G-quadruplex déplace la sonde « on-off » et permet de diminuer l'émission de fluorescence.

Il existe de nombreuses sondes possédant cette capacité « on-off », mais la plus utilisée est le thiazole orange (TO, Figure 39A). Cette sonde possède l'avantage d'avoir une affinité modérée pour les structures G-quadruplexes (Ka $\approx 3.10^6$ M⁻¹ pour l'ADN télomérique humain par exemple¹⁶²) ainsi qu'une excellente exaltation de fluorescence lorsqu'elle est liée aux structures G-quadruplexes (au minimum 450 fois) par rapport à la fluorescence lorsqu'elle est libre en solution (Rendement quantique $\varphi = 2.10^{-4}$). Elle possède un maximum d'absorbance à environ 490 nm et un maximum d'émission de fluorescence à environ 520 nm (Figure 39B). D'autres sondes ont été appliquées au FID dans le cas où les ligands étudiés possédaient un spectre d'absorption superposé au spectre d'absorption ou d'émission du TO, on retrouve notamment la sonde TO-PRO3 et Hoechst 33258 (Figure 39A).



Figure 39: Caractéristiques spectrales des différentes sondes avec A) la structure des trois sondes utilisées pour le FID et B) spectres d'absorbances (lignes pointillées) et d'émissions de fluorescence (lignes pleines) du Hoechst 33258 (jaune), du TO (orange) et du TO-PRO3 (bleu). Image tirée et adaptée de Largy *et al.*¹⁶¹

L'avantage de ces sondes est qu'elles ne présentent pas de sélectivités particulières pour une structure G-quadruplex par rapport à d'autres ou par rapport à l'ADN duplex. La sélectivité des ligands étudiés peut ainsi directement être mesurée en appliquant cette expérience à d'autres structures d'ADN. Il est cependant intéressant de noter que la stœchiométrie de sonde par rapport à l'ADN dépend directement du mode d'interaction entre ce dernier et la structure étudiée. Par exemple, il a été démontré par titration de fluorescence que le TO et l'ADN G4 se liaient avec un rapport 1:1 pour presque toutes les structures G-quadruplexes¹⁶². La seule exception est dans le cas des structures G-quadruplexes tétramoléculaires qui présentent deux quartets externes équivalents disponibles pour interagir avec la sonde, et implique donc l'interaction de deux TO pour un G4. Dans le cas de l'ADN duplex, il a été démontré que le rapport de TO et d'ADN duplex dépendait directement de la longueur de la séquence. Par exemple, la séquence ds26 régulièrement utilisée se lie avec le TO avec un rapport de 2:1 en faveur du TO.

Les expériences sont réalisées dans un tampon riche en ions K⁺, 0,25 μ M d'ADN ou d'ARN est généralement utilisé et 2 ou 3 équivalents de sonde dépendant de la stœchiométrie sont ajoutés. Une addition croissante de ligands est ensuite réalisée (généralement de 0 à 10 équivalents) et la fluorescence est mesurée. La diminution de fluorescence est ensuite représentée en pourcentage de sonde déplacée en fonction de l'ajout de ligand par : $\% TOD = 100 - (F/F_0)x100$. Avec *Fo* la fluorescence maximale, c'est-à-dire la fluorescence de l'ADN et du TO sans ajout de ligand, et *F* la fluorescence mesurée à chaque ajout de ligand. Cela permet ainsi de tracer la variation de fluorescence de la sonde qui reflète le pourcentage de sonde déplacée en fonction de ligand ajoutée (Figure 40).



Figure 40: Représentation d'une courbe de déplacement de sonde fluorescente en fonction de la concentration de ligand ajoutée. Les valeurs correspondantes au DC₅₀ et au TOD 1 μM sont représentées.

Deux valeurs sont généralement utilisées pour comparer des ligands entre eux : le DC₅₀ et le TOD 1 μ M. Le DC₅₀ correspond à la concentration de ligand nécessaire pour déplacer 50% de la sonde et est plutôt utilisé dans le cas des ligands de forte affinité avec les structures étudiées. Le pourcentage de déplacement à 1 μ M de ligand, noté TOD 1 μ M, est utilisé lorsque les ligands étudiés ne sont pas capables de déplacer près de 100% de la sonde. De plus, la sélectivité d'un ligand pour une structure G4 par rapport à la structure duplex peut être déterminée par le rapport du DC₅₀ obtenu sur le G4 et le DC₅₀ obtenu sur une séquence d'ADN duplex.

L'analyse par FID est donc une méthode rapide et peu coûteuse à mettre en œuvre et permet de comparer rapidement l'efficacité de plusieurs ligands différents à déplacer la sonde d'une structure d'ADN ou d'ARN. Elle est cependant limitée par deux facteurs : premièrement, il n'est pas possible de déterminer une constante d'affinité en raison de la compétition entre le ligand et la sonde, excepté si le ligand possède une affinité bien supérieure à celle de la sonde (au moins deux ordres de grandeur) ; deuxièmement, une réelle compétition entre la sonde et le ligand n'est possible que s'ils interagissent au niveau du même site du G-quadruplex. Ainsi, si un ligand ne déplace pas efficacement la sonde du G-quadruplex, cela peut provenir du mode d'interaction qui n'est pas le même. Pour ces raisons, il est toujours nécessaire de coupler cette méthode d'analyse avec une autre méthode comme le FRET-melting.

iv. FRET-melting

Le principe de l'effet FRET (pour Förster Resonance Energy Transfert) repose sur l'interaction de résonnance de deux molécules où une première molécule dite « donneur » peut transférer son

énergie d'excitation à une deuxième molécule dite « accepteur ». Ces deux molécules doivent être éloignées d'environ 10 à 100 Å avec un effet FRET qui diminue drastiquement à une distance supérieure. Cet effet a été appliqué à l'étude des G-quadruplexes à travers l'expérience de FRETmelting, initialement développée sur des séquences d'ADN télomérique¹⁶³ puis sur de nombreuses autres séquences par la suite^{164,165}. Dans cette expérience, une séquence connue pour former une structure G-quadruplex est fonctionnalisée en position 5' par le donneur d'énergie et en position 3' par l'accepteur d'énergie. Les fluorophores les plus utilisés sont la fluorescéine (FAM) comme donneur et la tetramethyl-rhodamine (TAMRA) comme accepteur, mais d'autres fluorophores peuvent être incorporés comme des cyanines ou des Bodipy¹⁶⁶. Lorsque la séquence d'ADN ou d'ARN est structurée, les deux fluorophores sont proches dans l'espace et la fluorescence de la fluorescéine est partiellement désactivée par effet FRET avec la rhodamine. Lors de la dénaturation thermique, le G-quadruplex se déstructure et les deux groupements s'éloignent, diminuant l'effet FRET et restaurant la fluorescence de la fluorescéine. De cette manière, il est possible de suivre la dénaturation de la structure en fonction de l'augmentation de la température (Figure 41). Il est théoriquement aussi possible de suivre la diminution de fluorescence de la rhodamine, mais cette méthode pose certains problèmes de reproductibilité comme mentionné par De Cian et al.¹⁶⁴



Figure 41: Schéma représentatif de l'expérience de FRET-melting appliquée aux G-quadruplexes et structures de la fluorescéine FAM et de la tetramethyl-rhodamine TAMRA.

Il est important de noter que l'ajout de fluorophores aux positions 5' et 3' peut avoir un impact plus ou moins grand sur la stabilité ainsi que sur la topologie de la structure¹⁶⁵, il est donc important de vérifier que la fonctionnalisation n'a pas d'impact délétère sur la formation de la structure Gquadruplex.

Il est possible de mesurer une différence de température de dénaturation lorsqu'un ligand capable de stabiliser la structure G-quadruplex est ajouté. La différence de température nécessaire à restaurer la moitié de la fluorescence de la fluorescéine avec et sans ligand est notée $\Delta T_{1/2}$ et permet de caractériser l'affinité du ligand pour la structure étudiée (Figure 42).



Figure 42: Représentation schématique d'une courbe de dénaturation d'une structure G-quadruplex seule (bleue) et avec un ligand (orange).

Dans le but de pouvoir observer une augmentation de la température de dénaturation, il est impératif que l'oligonucléotide seul n'ait pas un Tm trop élevé (50 < Tm < 70 °C). Pour cela, il est possible d'utiliser différentes concentrations en KCl pour moduler la stabilité du G4 et donc la température de dénaturation : les séquences très stables sont utilisées dans un tampon K⁺1 (KCl 1mM/LiCl 99mM/cacoLi 10mM, pH 7,2) et les séquences moins stables sont utilisées dans un tampon K⁺10 (KCl 10mM/LiCL 90mM/cacoLi 10mM, pH 7,2).

En plus de mesurer l'affinité d'un ligand pour une structure G-quadruplex, cette méthode permet aussi d'évaluer sa sélectivité par rapport à une autre structure. En effet, l'ajout d'un compétiteur non marqué par un fluorophore peut avoir un impact sur la température de stabilisation d'un ligand. Ce compétiteur peut être une séquence d'ADN double brin, une autre structure G-quadruplex ou d'autres structures secondaires. Il est important de noter que le compétiteur devrait toujours avoir une température de dénaturation supérieure à la séquence étudiée ainsi qu'à son complexe avec le ligand d'intérêt et ne devrait pas être complémentaire de la séquence G-quadruplex.

L'analyse par FRET-melting est donc une méthode rapide et robuste pour connaitre la stabilisation induite par un ligand de G-quadruplex. Elle peut être réalisée directement en plaque 96-puits et l'évolution de la fluorescence en fonction de la température est mesurée à l'aide d'une machine RT-PCR (pour Real Time Polymerase Chain Reaction).

6. Objectifs du travail de doctorat

Comme présenté plusieurs fois lors de l'introduction, les G-quadruplexes sont des structures pouvant se former par autoassociation de guanines. Ce sont des structures très stables *in vitro* et de nombreux travaux tendent à démontrer leur existence et leur rôle de régulateur de phénomènes biologiques *in vivo*. Des études bio-informatiques ont démontré que les PQS n'étaient pas répartis uniformément au sein du génome suggérant des implications biologiques importantes. La densité, la longueur et la quantité de ces PQS ont de plus été mis en relation avec le stade d'évolution de différentes espèces suggérant un rôle évolutif des G-quadruplexes dans le vivant¹⁶⁷. L'étude des structures G-quadruplexes est devenue, en moins de 20 ans, un sujet de recherche extrêmement important pour différents domaines scientifiques comme démontré par l'augmentation croissante du nombre de publications chaque année (Figure 43).



Figure 43: Graphique représentant le nombre de publications par année sur les G-quadruplexes recensé par le site PubMed.

Aujourd'hui et malgré un nombre grandissant de molécules interagissant avec les structures Gquadruplexes, il existe toujours un manque d'outils moléculaires pour étudier les implications précises de ces structures au sein du vivant. Tous les outils développés jusqu'ici possèdent leurs avantages et leurs inconvénients, il est donc nécessaire de les faire évoluer pour pallier les différents biais qu'ils peuvent impliquer. Dans ce contexte, de nouveaux outils moléculaires ont été développés pour l'étude *in cellulo* des structures G-quadruplexes.

Une première partie du travail doctoral a été consacrée à l'élaboration d'une nouvelle méthode pour la visualisation de ligands G-quadruplexes en cellules. À l'heure actuelle, plusieurs méthodes existent pour essayer de visualiser la formation de structures G-quadruplexes directement *in cellulo* et

chacune de ces méthodes seront introduites en détail dans le chapitre correspondant. Brièvement, la première méthode est l'imagerie par immunofluorescence d'anticorps spécifiques des Gquadruplexes, notamment développée par l'équipe de S. Balasubramanian⁹³. L'avantage de cette méthode est la haute résolution obtenue par l'amplification du signal de fluorescence, mais le principal inconvénient est la fixation nécessaire des cellules avant l'incubation des anticorps. Une deuxième méthode est le développement de sondes dont les propriétés de fluorescence sont modifiées lorsqu'elles interagissent spécifiquement avec leurs cibles. Cette méthode peut parfois être appliquée directement in vivo mais nécessite une accumulation de molécules au sein de la cellule pour pouvoir détecter le signal. Une dernière méthode consiste à développer des ligands des G-quadruplexes à la fois très sélectifs et très affins de ces structures et de les fonctionnaliser avec des fluorophores. Cette méthode possède l'avantage de pouvoir être utilisée en cellules vivantes et d'utiliser des ligands qui sont reconnus pour leur affinité avec les G-quadruplexes. Elle possède néanmoins deux inconvénients, le premier est que la fonctionnalisation par un fluorophore peut avoir un impact négatif sur la reconnaissance des G-quadruplexes par les ligands et qu'il est nécessaire, encore une fois, d'avoir une réelle accumulation de ligands à l'intérieur de la cellule pour obtenir un rapport signal/bruit satisfaisant. Dans ce contexte, nous avons développé une nouvelle méthodologie dans le but de suivre avec précision la répartition de petites molécules en couplant la technique d'imagerie par immunofluorescence et la fonctionnalisation de ligands connus des G4s par un haptène (Figure 44). Ainsi, le ligand de référence PDC a été fonctionnalisé par des bras espaceurs de tailles et de polarités différentes. Ces linkers sont fonctionnalisés à leurs extrémités soit par une fonction azoture soit par une fonction alcyne. Parallèlement, l'haptène 5-BrdU (une molécule capable d'être reconnue par un anticorps spécifique disponible commercialement) a été fonctionnalisé soit par une fonction alcyne soit par une fonction azoture. Ces deux dérivés ont été couplés entre eux par une réaction de CuAAC pour générer les ligands de type PDC-BrdU. Les ligands avant et après fonctionnalisation ont été testés pour leur affinité ainsi que pour leur sélectivité pour les structures G-quadruplexes. La détection d'un ligand PDC-BrdU lié à différentes structures G4s par un anticorps anti-BrdU a ensuite été évaluée par une méthode ELISA modifiée. Les composés avant et après fonctionnalisation ont ensuite été utilisés directement in cellulo pour l'imagerie par immunofluorescence. Dans le cas des composés non fonctionnalisés par la 5-BrdU, la réaction de CuAAC a été réalisée directement in cellulo dans le but de former les composés finaux in situ.



Figure 44: Schéma représentatif de la stratégie imaginée pour la détection de ligands des G-quadruplexes par immunofluorescence.

La deuxième partie du travail doctoral a été focalisée sur le développement de nouvelles molécules permettant l'identification exacte des sites de liaisons des ligands des G-quadruplexes au niveau du génome. Depuis plusieurs années, un certain nombre de méthodes expérimentales ont été développées dans le but de comprendre quels endroits du génome étaient susceptibles de former des G-quadruplexes¹⁶⁸. Le séquençage couplé à l'immunoprécipitation d'un anticorps sélectif des G-quadruplexes (G4 ChIP-Seq⁹²) a permis de dénombrer environ 10 000 G-quadruplexes comme décrit dans l'introduction. Le problème majeur de cette méthode repose encore une fois sur la fixation des cellules nécessaire avant l'incubation avec les anticorps, impactant la dynamique de formation des G-quadruplexes n'a été développée dans le but de cartographier les sites d'interactions de ces derniers avec l'ADN génomique. Actuellement, la seule méthode emploie l'incubation de dérivés biotinylés de ligands de référence^{169,170} au sein de lysats cellulaires et l'extraction d'ADN ou d'ARN ne repose donc que sur des interactions non covalentes (π -stacking et électrostatiques). Ces interactions, bien que pouvant être fortes avec certaines structures G-quadruplexes, ne peuvent empêcher la dissociation et la relocalisation des ligands lors de l'étape d'extraction et de séquençage.

Dans ce contexte, de nouveaux dérivés du ligand de référence PDC ont été synthétisés. Ces molécules sont fonctionnalisées par des fonctions photoactivables et une fonction alcyne. Deux types de ligands trifonctionnels ont été réalisés : des ligands possédant un bras espaceur linéaire et des ligands possédant un bras espaceur non linéaire (Figure 45). Les fonctions photoactivables peuvent être de plusieurs types, mais le principe de fonctionnement est le même : ce sont des fonctions qui peuvent générer des espèces électrophiles très réactives après photo-irradiation et qui peuvent réagir avec les fonctions à proximité. Dans ce cas, les fonctions photoactivables devraient être capables de sélectivement alkyler les G-quadruplexes lorsque les ligands interagissent avec ces

62

derniers. Ces différentes molécules sont donc testées par études biophysiques pour connaitre leur affinité et sélectivité pour les structures G-quadruplexes et leur capacité à générer des adduits covalents avec certaines séquences G-quadruplexes est évaluée par gels d'électrophorèse.

Le but final est donc de pouvoir incuber ces molécules *in vivo* et générer des liens covalents entre les ligands et les séquences formant des structures G-quadruplexes après photo-irradiation. Un haptène (5-BrdU ou biotine) pourra ensuite être ajouté au ligand G-quadruplex directement *in cellulo* à l'aide d'une réaction bioorthogonale sur la fonction alcyne ce qui permettra de réaliser l'immunoprécipitation. Finalement, le séquençage à haut débit permettra de cartographier avec précision les différents sites d'alkylation et de mettre en relation l'accessibilité des structures G-quadruplexes avec différents paramètres comme le type de cellules ou le cycle cellulaire.



Figure 45: Structures des différents composés photoactivables linéaires (gauche) et non linéaires (droite).

Finalement, une dernière partie du travail doctoral a été consacrée à la synthèse de dérivés des composés PDC et PhenDC3 dans le but de comprendre l'importance de la position du lien amide sur la partie quinoline. En effet, le ligand PDS, un ligand de référence des G-quadruplexes et un analogue structurel du PDC est constitué d'un cycle pyridine reliée aux quinolines en position 2 tandis que la pyridine du ligand PDC est reliée aux quinolines en position 3. De plus, des dérivés du ligand PDS fonctionnalisés par des bras bis-pyrrolidines ont été synthétisés par l'équipe de S. Balasubramanian, démontrant la capacité d'induire des dommages à l'ADN télomérique. Ces dérivés impliquent un lien amide en position 2 et, dans le but de comprendre l'impact de cette régioisomérie, des dérivés PhenDC et PDC –bisPyro ont été synthétisés et évalués par des études biophysiques.



Figure 46: Structures des ligands PDS et PDC avec, en rouge, le numéro de la position du lien amide sur la quinoline.

Chapitre II : Immunodétection des

G-quadruplexes guidée par des

<u>ligands</u>

1. Introduction

Comme présenté dans l'introduction générale, les motifs G-quadruplexes peuvent potentiellement se former à travers différentes régions du génome ainsi que du transcriptome et leur formation peut avoir un impact sur le fonctionnement de différents processus biologiques comme la réplication, la transcription et la traduction. Dans le but de vérifier leur formation dans un contexte cellulaire, plusieurs méthodes ont été développées. Parmi ces dernières, une méthode consiste à visualiser directement *in cellulo* la formation de G-quadruplexes par immunodétection d'anticorps spécifiques ou par imagerie de petites molécules. Ces différentes approches possèdent leurs avantages et inconvénients et sont souvent utilisées en parallèle pour visualiser efficacement les potentiels G-quadruplexes en cellules.

a. Détection des G-quadruplexes par immunofluorescence

La détection par immunofluorescence repose sur l'imagerie cellulaire après incubation d'un anticorps spécifique d'une biomolécule donnée¹⁷¹. L'anticorps peut être fonctionnalisé par des fluorophores, dans ce cas l'imagerie cellulaire est réalisée directement après incubation des anticorps (immunofluorescence directe). Dans le cas où l'anticorps n'est pas fonctionnalisé par des fluorophores, l'imagerie est réalisée en deux étapes : 1) la reconnaissance de la biomolécule ciblée par l'anticorps primaire et 2) la reconnaissance de l'anticorps primaire par des anticorps secondaires fonctionnalisés par des fluorophores (immunofluorescence indirecte). Globalement, cette deuxième technique permet d'avoir un meilleur rapport signal/bruit, car plusieurs anticorps secondaires peuvent reconnaitre un anticorps primaire, donnant ainsi lieu à une amplification du signal de fluorescence.

Historiquement, le premier anticorps décrit pour interagir préférentiellement avec les structures Gquadruplexes par rapport à d'autres structures de l'ADN a été isolé en 1998¹⁷². L'anticorps me^VIIB4 a été testé sur différentes séquences d'ADN formant différentes structures et a montré des constantes d'affinités supérieures pour les séquences formant des G-quadruplexes. Notamment, cet anticorps présentait une affinité supérieure pour des séquences parallèles dérivées de l'ADN télomérique avec, par exemple, un Ka $\approx 1,5.10^7$ M⁻¹ pour la séquence (T₂G₄)₄, que pour les autres séquences étudiées. Trois ans après, un autre anticorps, Sty49, a été développé pour sélectivement interagir avec les structures G-quadruplexes et fut le premier à être utilisé pour visualiser les G-quadruplexes en cellules. Cet anticorps possède un K_D de 3 à 5 nM pour des séquences d'ADN télomériques sans sélectivité particulière pour une topologie parallèle ou antiparallèle¹⁷³. L'immunodétection indirecte de ces anticorps a de plus permis de montrer la formation de foci au sein des macronucléus de l'espèce de cilié Stylonychia lemnae attribuée à la formation de structures G-quadruplexes. En 2013, l'équipe de Balasubramanian a développé BG4, un nouvel anticorps interagissant avec un K_D de l'ordre du nM pour différentes structures G-quadruplexes^{93,174}. Ce dernier a été utilisé pour détecter les G-quadruplexes formés par différentes lignées cellulaires à l'aide d'un protocole d'immunofluorescence indirecte en trois étapes : 1) incubation de BG4 pour reconnaitre les structures G-quadruplexes, 2) incubation d'anticorps secondaires sélectifs de BG4 et 3) incubation d'anticorps tertiaires fonctionnalisés par des fluorophores sélectifs des anticorps secondaires. Cette expérience a notamment permis d'observer la formation de foci au sein du noyau des cellules U2OS (Figure 47A) et MCF-7 attribuée à la présence d'ADN-G4, mais aussi du noyau et du cytoplasme des cellules MRC5 attribuée à la présence d'ARN-G4¹⁷⁵. Les chromosomes de la lignée cellulaire cancéreuse HeLa ont été isolés et l'immunodétection de BG4 a permis de montrer la formation de foci sur des chromosomes en métaphase et plus particulièrement aux extrémités de ces derniers, démontrant la possible formation de G-quadruplexes au niveau des télomères (Figure 47B).



Figure 47: Visualisation de l'anticorps BG4 dans A) les noyaux des cellules U2OS et B) à l'intérieur (gauche) et aux extrémités (droite) des chromosomes provenant de cellules cancéreuses HeLa⁹³.

L'immunodétection des structures G-quadruplexes a ensuite été réalisée après synchronisation des cellules dans une phase particulière du cycle cellulaire⁹³. Cette expérience a permis de montrer une faible quantité de foci lorsque les cellules sont synchronisées dans la phase G0/G1, une phase marquée par l'absence de réplication. Lorsque les cellules étaient synchronisées en phase réplicative S, la quantité de foci observée était alors 4,8 fois supérieure par rapport à la phase G0/G1, démontrant l'augmentation du nombre de structures G-quadruplexes lors de la réplication de l'ADN. Finalement, le traitement des cellules par la PDS a permis de multiplier par 2,9 le nombre de foci formé au niveau des noyaux des cellules U2OS comparé aux cellules non traitées. Cette observation a été imputée à la capacité de la PDS à stabiliser la formation transitoire des G-quadruplexes au sein

des cellules. BG4 est ainsi devenu un anticorps de référence pour étudier la localisation des Gquadruplexes et est notamment utilisé pour déterminer si une petite molécule est capable d'interagir avec les G-quadruplexes in cellulo^{176,177}. D'autres anticorps ont cependant été développés comme l'anticorps 1H6¹⁷⁸ la même année en 2013 et l'anticorps D1¹⁷⁹ développé en 2016. L'anticorps 1H6 possède des K_D de l'ordre du nM pour différentes structures G-quadruplexes à l'image de BG4. L'incubation de 1H6 avec de nombreuses lignées cellulaires différentes a très majoritairement montré la formation de foci intranucléaires. L'ajout de compétiteurs comme des séquences d'ADN ou d'ARN a montré une baisse drastique du nombre de foci lorsque le compétiteur était une séquence ADN-G4 et un nombre inchangé lorsque le compétiteur était une séquence duplex. De plus, des cellules déficientes en hélicases spécifiques des G-quadruplexes ont été incubées avec 1H6 et l'immunodétection a montré une augmentation du nombre de foci, démontrant la capacité de cet anticorps à cibler spécifiquement les G-quadruplexes. L'anticorps D1 est, quant à lui, un anticorps sélectif de la topologie parallèle de la séquence télomérique humaine avec un K_D d'environ 17 nM. La comparaison entre la quantité de signal détectée au niveau des télomères par l'anticorps D1, sélectif de la topologie parallèle, et l'anticorps BG4, non sélectif d'une topologie particulière, a montré qu'environ 77% des télomères formaient des structures parallèles in cellulo.

Finalement, l'imagerie par immunofluorescence est une méthodologie puissante pour la détection de structures G-quadruplexes *in cellulo* avec une excellente résolution et un bon rapport signal/bruit. Il est cependant important de noter que l'incubation d'anticorps nécessite la perméabilisation, et donc, la fixation des cellules en amont. Ces traitements modifient indubitablement la formation et la déstructuration des G-quadruplexes *in vivo*. L'immunodétection ne permet donc pas de rendre compte exactement de ce qu'il se passe au sein d'une cellule vivante et devrait toujours être utilisée avec des techniques complémentaires.

b. Détection de ligands des G-quadruplexes

Dans le but de traquer la formation de structures G-quadruplexes *in cellulo* à l'aide de petites molécules, deux méthodes existent actuellement. La première méthode consiste en la détection par imagerie cellulaire de ligands de G-quadruplexes intrinsèquement fluorescents ou dont les propriétés de fluorescences vont être modifiées lors de l'interaction avec la cible et la deuxième méthode est la visualisation de ligands des G-quadruplexes fonctionnalisés par des fluorophores.

i. Détection de ligands des G-quadruplexes fluorescents

La détection de ligands des G-quadruplexes intrinsèquement fluorescents a été la première méthode utilisée dans le but de suivre la répartition de ces derniers *in cellulo*. Historiquement, le premier ligand de G-quadruplexes utilisé pour l'imagerie cellulaire est le ligand métallique QP3(In)¹⁸⁰, dérivé de la porphyrine TmPyP4 (Figure 48A). Ce composé est peu fluorescent, mais l'imagerie de fluorescence après 3 jours d'incubation dans les cellules cancéreuses MCF7 a tout de même permis de montrer une localisation majoritairement nucléaire du composé. La quarfloxine¹⁸¹ (ou CX-3543) est le premier ligand des G-quadruplexes à avoir été testé par des essais cliniques (stoppés en phase II). Il a été montré que cette molécule était capable de déstructurer le complexe nucléoline/G4 au niveau des nucléoles et d'induire l'apoptose de cellules cancéreuses (Figure 48B). Le nucléole est un sous-compartiment nucléaire non délimité par une membrane et qui est principalement constitué de protéines, d'ADN et d'ARN. Cette région du noyau sert notamment à la transcription d'ARN ribosomique et est donc une cible intéressante pour les ligands des G-quadruplexes. L'imagerie cellulaire après incubation de la quarfloxine dans les cellules cancéreuses pulmonaires A549 a notamment permis de montrer une localisation de cette dernière au niveau des nucléoles.



Figure 48: Imagerie de fluorescence après incubation *in vivo* de A) QP3(In) 40 μM dans les cellules MCF7 et B) la quarfloxin 5 μM dans les cellules A549.

Ces deux molécules interagissent avec les structures G-quadruplexes, mais leurs propriétés de fluorescence ne sont pas optimisées pour la détection précise de formation des G-quadruplexes. Par conséquent, d'autres molécules ont été développées et optimisées pour détecter efficacement les petites molécules *in cellulo* comme des sondes « on-off » ou des molécules dont le temps de vie de fluorescence change lorsqu'elles interagissent avec leur cible. Par exemple, QUMA-1, un dérivé de

coumarine a été incubé en cellules HeLa et l'imagerie de fluorescence a permis de montrer la formation de nombreux foci majoritairement localisés au niveau du cytoplasme des cellules vivantes et fixées¹⁸² (Figure 49A). Dans le but de vérifier si l'ARN était la cible de la molécule au niveau du cytoplasme, différentes expériences de digestion des séquences à ARN ont été réalisées. Notamment, la destruction de l'ARN par différents traitements comme la formamide, la RNase T1 (coupe l'ARN au niveau des guanines) et la RNase A (coupe l'ARN sans sélectivité) a permis de faire disparaitre complètement la fluorescence de QUMA1 (Figure 49B, Figure 49D). La digestion partielle de l'ARN par la RNase III (sélectif des structures ARN double brin) n'a pas permis de diminuer la fluorescence, démontrant une sélectivité intéressante de QUMA1 pour les structures à simple brin (Figure 49C). Dans le but de vérifier si QUMA1 était capable de reconnaitre les structures Gquadruplexes en cellules, plusieurs expériences ont été réalisées. Notamment, l'incubation de compétiteurs interagissant avec les G-quadruplexes comme BRACO19 ou cPDS a permis de diminuer drastiquement la fluorescence de QUMA1, traduisant le déplacement de cette dernière par les ligands (Figure 49E-F). La déstructuration des G-quadruplexes à ARN par ajout d'oligonucléotides riches en cytosines ou par l'augmentation de l'expression de l'hélicase DHX36, une hélicase connue pour déstructurer les G-quadruplexes in vivo⁹⁶, a aussi permis de diminuer la quantité de fluorescence de QUMA1 (Figure 49G-H).



Figure 49: Structure de QUMA1 et imagerie cellulaire par microscopie confocale après incubation de QUMA1 A) sans traitement et après traitement par B) formamide pour détruire les structures secondaires d'ARN, C) RNase III pour détruire les ARN double brins, D) RNase T1 pour couper l'ARN au niveau des guanines, E-F) les ligands des Gquadruplexes BRACO19 et cPDS, G-H) des séquences d'oligonucléotides riches en C et augmentation de l'expression de DHX36 pour aider à la dénaturation des ARN G-quadruplexes. Image tirée de Chen *et al.*¹⁸²

Les avancées techniques et technologiques ont de plus permis de pouvoir observer en temps réel l'évolution de la fluorescence de certaines sondes directement *in vivo*. Par exemple, la sonde IMT
développée en 2018 a permis de suivre en temps réel l'évolution de la fluorescence au sein de cellules vivantes HeLa¹⁸³. Cette sonde est un dérivé de benzothiazole et possède une excellente photostabilité ainsi qu'une bonne capacité « on-off » avec une fluorescence au moins 200 fois supérieure lorsqu'elle interagit avec une structure G-quadruplex par rapport à sa fluorescence libre en solution. L'incubation en cellules vivantes et le suivi en temps réel de la fluorescence a permis d'observer la création de foci au cours du temps en atteignant un maximum de fluorescence en 7 minutes (Figure 50A). Le traitement des cellules par une concentration croissante du ligand PDS a notamment permis de montrer une diminution drastique de la fluorescence de la sonde IMT, démontrant sa capacité à suivre efficacement l'interaction de la PDS avec les G-quadruplexes *in vivo* (Figure 50B).



Figure 50: Évolution de la fluorescence d'IMT au sein de cellules vivantes Hela A) au cours du temps et B) après traitement avec une concentration croissante de ligand PDS. Image tirée de Zhan *et al.*¹⁸³

De nombreuses autres sondes ont été développées dans le but de visualiser efficacement la formation de structures G-quadruplexes *in cellulo* comme des complexes métalliques, des dérivés de cyanines etc.^{184,185} La mesure du temps de vie de fluorescence peut aussi être appliquée *in cellulo* pour déterminer la formation de structures G-quadruplexes. Le temps de vie de fluorescence (τ) est une caractéristique inhérente aux molécules fluorescentes qui correspond au temps moyen dans lequel une molécule rester à l'état excité avant d'émettre un photon. Certaines de ces molécules peuvent avoir un temps de vie de fluorescence qui va être modifié lorsqu'elles interagissent avec leurs cibles et l'imagerie par temps de vie de fluorescence (FLIM pour fluorescence lifetime imaging microscopy) permet d'obtenir des informations sur l'environnement de la molécule. Historiquement,

o-BMVC (Figure 51) a été la première sonde utilisée par FLIM dans le but de démontrer la formation de structures G-quadruplexes in cellulo¹⁸⁶. En effet, cette sonde possède un temps de vie de fluorescence d'environ 2,8 ns lorsqu'elle interagit avec l'ADN télomérique humain et un temps de vie de fluorescence d'environ 1,2 ns lorsqu'elle interagit avec l'ADN duplex. L'analyse par FLIM après incubation de o-BMVC en cellules CL1-0 a permis de montrer la formation de G-quadruplexes dans le cytoplasme (Figure 51a). Sur la Figure 51b, les points blancs représentent la molécule o-BMVC lorsqu'elle est analysée avec un temps de vie de fluorescence supérieure à 2,4 ns et la fluorescence rouge représente toutes les molécules avec un temps de vie de fluorescence inférieure à 2,4 ns. Lorsque la mesure du temps de vie de fluorescence est augmenté à 2,8 ns (Figure 51c), le nombre de points blancs (>2,8 ns) diminue et la quantité de fluorescence en rouge (<2,8 ns) augmente. Cette expérience démontre que l'analyse par FLIM de la sonde o-BMVC est capable de discriminer in cellulo les structures G-quadruplexes des autres structures formées par l'ADN et l'ARN. La possibilité d'étudier uniquement les molécules interagissant avec les structures G-quadruplexes a par la suite permis d'étudier l'interaction de ligands G-quadruplexes directement in cellulo. Par exemple, l'incubation de la sonde o-BMVC suivi de l'ajout du ligand TmPyP4 a permis de montrer une diminution drastique du nombre de foci dû à l'interaction de o-BMVC avec les structures Gquadruplexes, démontrant ainsi la capacité de TmPyP4 à déplacer la sonde des G-quadruplexes¹⁸⁷.



Figure 51 : Imagerie par temps de vie de fluorescence des cellules vivantes CL1-0 après incubation de la sonde o-BMVC et analyse par différents temps de vie de fluorescence (2,4 et 2,8 ns). Image tirée de Tseng *et al.*¹⁸⁶

Récemment, la sonde DAOTA-M2, un dérivé triangulénium, a été synthétisée par l'équipe de Ramon Villar et sa capacité à discriminer différentes structures d'oligonucléotides par temps de vie de fluorescence a été étudiée¹⁸⁸. Il a été montré que la molécule DAOTA-M2 possédait un temps de vie de fluorescence d'environ 7 ns avec les séquences d'ADN (ADN duplex et simple brin), environ 11 ns avec les séquences d'ARN (duplex et simple brin) et environ 15 ns avec les structures G-quadruplexes (ADN et ARN) *in vitro*. L'imagerie cellulaire de DAOTA-M2 en cellules vivantes U2OS a permis de montrer que cette sonde interagissait principalement avec l'ADN nucléaire et l'ADN mitochondrial comme démontré par différentes expériences de colocalisation. Comme pour la sonde o-BMVC, la molécule DAOTA-M2 a été utilisée pour visualiser la formation de structures G-quadruplexes par

FLIM en cellules vivantes. Le traitement des cellules par le ligand des G-quadruplexes PDS a de plus permis d'observer une baisse du temps de vie de fluorescence de DAOTA-M2, attribuée au déplacement de la sonde par la PDS. En 2021, la même équipe a élargi l'application de déplacement de temps de vie de fluorescence à d'autres molécules¹⁸⁹ et a de plus montré que la quantité de structures G-quadruplexes détectée par imagerie FLIM sur DAOTA-M2 était directement impactée par l'étape de fixation des cellules. La fixation des cellules est nécessaire pour l'incubation de certaines molécules ou pour l'incubation d'anticorps, mais impacte la morphologie de la cellule et peut induire des dommages¹⁸⁹. Il est donc nécessaire de prendre des précautions sur l'analyse des résultats en cellules fixées.

ii. Fonctionnalisation de ligands G-quadruplexes connus par des fluorophores

Le but de fonctionnaliser un ligand des G-quadruplexes par un fluorophore est de combiner l'affinité et la sélectivité du ligand pour les structures G-quadruplexes et la possibilité de visualiser le fluorophore par imagerie de fluorescence. Le premier ligand des G-quadruplexes couplé à un fluorophore développé est le ligand 360A (PDC) fonctionnalisé par un motif thiazole orange sur la position 4 de la pyridine (PDC-L-TO, Figure 52) ou directement sur une des unités quinolinium (PDC-*M*-TO)¹⁹⁰. La présence du fluorophore a cependant montré un impact non négligeable sur l'affinité du ligand PDC pour l'ADN G-quadruplex comparé à l'ADN duplex comme démontré par titration de fluorescence ainsi que par FRET-melting. Suite à cette étude, d'autres ligands couplés au motif TO ont été synthétisés comme le dérivé d'isaindigotone ISTO¹⁹¹ (Figure 52). Ce dernier a été notamment utilisé pour l'imagerie cellulaire, mais était alors limitée par la faible fluorescence du motif TO. D'autres fluorophores ont par la suite été couplés à des ligands G-quadruplexes comme le motif BODIPY sur le macrocyle L1H1-7OTD¹⁹² dérivé de la telomestatine (L1BOD-7OTD) ou sur le ligand PDC¹⁹³ (PDC-BODIPY) ainsi que des motifs cyanines notamment sur les dérivés d'acridine BRACO19¹⁹⁴ (BRACO19-MonoCy) et bispyrrolidine de la PDS¹⁹⁵ (PDP-Cy5, Figure 52). Ce dernier a notamment été testé par imagerie cellulaire après transfection de structures G-quadruplexes ADN (c-kit) ou d'ARN (BCL2) fonctionnalisées par une fluorescéine au sein de cellules cancéreuse HeLa. La colocalisation entre les G-quadruplexes fonctionnalisés et le PDP-Cy5 a montré que ce dernier était capable de sélectivement interagir avec ces structures.



Figure 52: Structures des ligands de G-quadruplexes connus fonctionnalisés par un fluorophore de type TO, BODIPY ou Cyanine.

Très récemment, l'équipe de Balasubramanian a utilisé un dérivé du ligand PDS-Bispyro¹⁹⁶ fonctionnalisé par une rhodamine (SiR-PyPDS) pour suivre la formation de structures G-quadruplex in vivo à l'aide de la microscopie HILO¹⁹⁷ (pour Highly Inclined Laminated Optical). Cette technique d'imagerie permet, en théorie, de pouvoir détecter des signaux de fluorescence à l'échelle d'une seule molécule. Le ligand SiR-PyPDS interagit avec plusieurs G-quadruplexes avec un K_D de l'ordre du µM et montre une exaltation de fluorescence lorsqu'il interagit avec les G-quadruplexes comparé avec l'ADN duplex. La détection de fluorescence à l'échelle d'une molécule a ensuite été réalisée in vitro par ajout de ligand SiR-PyPDS sur une structure G-quadruplex formée par la séquence Myc préalablement immobilisée sur une surface (Figure 53A). Les images ont ainsi été enregistrées pendant un temps d'exposition de 500 ms et une moyenne de 867 spots de fluorescence a été détectée. Ces spots de fluorescence ont été attribués aux évènements de structuration et de déstructuration des G-quadruplexes. Dans le but de confirmer que la formation de ces spots était due aux interactions ligands/G4, la même expérience a été réalisée avec la molécule contrôle SiRiPyPDS qui n'interagit que faiblement avec les G-quadruplexes et uniquement 66 évènements ont été observés. Une compétition a ensuite été réalisée avec l'incubation du ligand PhenDC3 dans le but de déplacer le ligand SiR-PyPDS des G-quadruplexes. Dans les conditions expérimentales, 250 pM de SiR-PyPDS sont utilisés et un excès de PhenDC3 (10 µM, soit une concentration 40 000 fois supérieure) était cependant nécessaire pour empêcher presque complètement la détection d'évènements d'interactions. La molécule a ensuite été utilisée directement *in cellulo* et l'imagerie de fluorescence HILO a pu être réalisée en cellules U2OS (Figure 53B). L'accumulation de 400 images après 100 ms d'exposition a permis de détecter la formation d'environ 79 évènements d'interactions au sein des noyaux en l'espace de 40 s. Le nombre d'évènements d'interactions a ensuite été mis en relation avec le cycle cellulaire : lorsque les cellules étaient synchronisées en phase S, environ 208 évènements étaient détectables tandis qu'aucun évènement n'était détectable en phase G0/G1. La phase S étant une phase active de réplication, la quantité de G-quadruplexes potentiellement formés est donc plus élevée que lors de la phase G0/G1.



Figure 53: Structure du ligand SiR-PyPDS et A) imagerie *in vitro* d'une surface recouverte de structures G-quadruplexes Myc après ajout de ligand SiR-PyPDS (250 pM) et B) imagerie de cellules vivantes U2OS après incubation de SiR-PyPDS (20 nM) pendant 20 min. Image tirée de Di Antonio *et al.*¹⁹⁶

La microscopie HILO est donc une technique très intéressante pour étudier avec précision le comportement de petites molécules *in cellulo* mais est limitée par son coût et sa complexité d'utilisation plus élevés que la microscopie confocale classique. De plus, il est très important de noter que la fonctionnalisation de ligands de référence par des fluorophores peut avoir un impact important sur l'affinité et la sélectivité envers les structures G-quadruplexes. En effet, les fluorophores sont souvent des molécules aromatiques encombrantes et peuvent gêner la reconnaissance des structures G-quadruplexes par le ligand¹⁹⁰. Pour pallier ce problème, la fonctionnalisation par le fluorophore peut être réalisée directement *in cellulo* à l'aide d'une réaction bioorthogonale¹⁹⁸. Les réactions bioorthogonales sont des réactions chimiques pouvant se produire en milieu cellulaire sans interférer avec les autres molécules présentes en solution. L'avantage de

cette méthode est qu'elle permet l'incubation de ligands de référence fonctionnalisés par un groupement peu encombrant ce qui permet d'éviter une modification trop importante de l'affinité et de la sélectivité de ces dernières. Par exemple, le ligand PDS a été fonctionnalisé par une fonction alcyne au niveau de la chaine de la pyridine centrale. Cette fonction a permis de réaliser une réaction de CuAAC (pour Copper-catalyzed alkyne azide cycloaddition) avec le fluorophore Alexa Fluor 594 fonctionnalisé par un azoture¹²⁶. Cette réaction a été réalisée après incubation de cellules vivantes U2OS avec la PDS-Alk et après fixation des cellules. En effet, la réaction de CuAAC nécessite l'ajout d'un excès de cuivre pour catalyser la réaction et ne peut donc pas être appliquée en cellules vivantes au niveau des noyaux des cellules. La colocalisation avec hPif1, une hélicase connue pour déstructurer les G-quadruplexes, a ainsi pu permettre de vérifier que les G4s étaient la cible de ce ligand (Figure 54).





Plus récemment, des dérivés du ligand PhenDC3 fonctionnalisés par un alcyne ou un azoture ont été synthétisés¹⁹⁹. Le dérivé PhenDC3-Alk a été couplé avec le fluorophore Cy5-N3 à l'aide d'une réaction de CuAAC et le dérivé PhenDC3-N3 a été couplé avec le fluorophore Cy5-DBCO à l'aide d'une réaction SPAAC²⁰⁰ (pour Strain-Promoted Alkyne Azide Cycloaddition). Cette dernière réaction utilise le fluorophore Cy5 fonctionnalisé par un alcyne contraint dans un cycle à 8 chaînons (cyclooctyne). La tension de l'alcyne linéaire au sein d'un cycle augmente sa réactivité et permet de réaliser des réactions de cycloaddition sans ajout de cuivre pour catalyser la réaction²⁰¹. L'incubation de cellules cancéreuses HT-29 par le dérivé PhenDC3-Alk suivi de la fixation et de la réaction de CuAAC avec le fluorophore Cy5-N3 a permis de montrer la localisation du ligand au niveau des nucléoles (Figure 55A). La même expérience a été réalisée avec le ligand PhenDC3-N3 et le fluorophore Cy5-DBCO et montrait alors une localisation principalement nucléoplasmique sans marquage des nucléoles (Figure 55B). Cette étude a ainsi permis de montrer que l'ajout de cuivre pouvait participer à la relocalisation

des ligands et des fluorophores indépendamment de leur cible via la formation probable de complexes alcyne/Cu(I). Par conséquent, des contrôles rigoureux devraient être réalisés lors de l'étude de réactions de CuAAC *in cellulo*.



Figure 55 : Imagerie cellulaire après incubation de A) 5 μ M de PhenDC3-Alk et réaction de cycloaddition catalysée au cuivre avec le fluorophore Cy5-N3 et B) 5 μ M de PhenDC3-N3 et réaction de SPAAC avec le fluorophore Cy5-DBCO. Les flèches blanches montrent les nucléoles et la barre de mesure représente 5 μ m. Image tirée de Lefebvre *et al.*¹⁹⁹

Finalement, la fonctionnalisation de ligands de référence des G-quadruplexes par des fluorophores permet de connaitre la localisation de ces derniers au sein des cellules. Cependant, la fluorescence des ligands après fonctionnalisation avec un fluorophore peut ne pas être suffisante pour détecter avec précision et avec une bonne résolution la formation des complexes G4s/ligands.

c. Principe de l'immunodétection de ligands des G-quadruplexes

Dans le but d'améliorer la sensibilité de détection de ligands des G-quadruplexes par imagerie de fluorescence, une nouvelle méthode couplant les avantages de l'immunofluorescence et de la reconnaissance de ligands des G-quadruplexes a été développée. Cette méthode appelée G4-GIS (pour G4-ligand guided immunofluorescence staining) se base sur la fonctionnalisation du ligand de référence PDC par un haptène, ici la 5-bromo-2'-désoxyuridine (5-BrdU, Figure 56). Un haptène est

une molécule généralement de faible poids moléculaire pouvant être sélectivement reconnue par un anticorps²⁰².



Figure 56: Structure du PDC et de la 5-BrdU.

Dans notre cas, fonctionnaliser le ligand PDC par la 5-BrdU permettra de pouvoir visualiser *in cellulo* la localisation de ces ligands par immunofluorescence indirecte après reconnaissance de la 5-BrdU par un anticorps spécifique. Utiliser la 5-BrdU présente au moins deux avantages. Le premier vient de la taille relativement faible de la 5-BrdU qui permettra en théorie de ne pas avoir un effet trop délétère sur l'affinité du ligand PDC envers les structures G-quadruplexes. Le deuxième avantage est le fait que cet haptène synthétique soit exogène aux cellules. Cela permettra, en théorie, de limiter le bruit de fond dû à la reconnaissance non spécifique par l'anticorps primaire. Cette méthode permet d'allier l'affinité du ligand PDC envers les G-quadruplexes avec la très bonne résolution apportée par l'amplification de fluorescence du protocole d'immunofluorescence indirecte. Quatre dérivés du ligand PDC fonctionnalisés par des linkers de taille et de polarité différentes portant une fonction alcyne ou azoture ont été synthétisés (Figure 57A). Deux dérivés de la 5-BrdU fonctionnalisés en position 3' par une fonction alcyne et azoture ont aussi été synthétisés (Figure 57B).



Figure 57: Structures des différents dérivés PDC-Alk/N3 (A) et 5-BrdU-Alk/N3 (B) synthétisés.

Ces fonctions alcynes et azotures permettent ainsi de pouvoir fonctionnaliser les dérivés PDC-Alk/N3 avec le partenaire 5-BrdU-Alk/N3 adapté par une réaction de CuAAC pour former 4 ligands PDC-BrdU finaux (Figure 58). Cette réaction de click pourra de plus être réalisée directement *in situ* après incubation des dérivés PDC-Alk/N3 *in cellulo*. Cela nous permettra de pouvoir comparer l'impact de la présence de la 5-BrdU sur l'affinité et la sélectivité du ligand PDC ainsi que de pouvoir comparer la différence de localisation au sein des cellules.



Figure 58: Structure des composés finaux PDC-4,2-BrdU, PDC-4,3-BrdU, PDC-4,0-BrdU et PDC-4,PEG-BrdU.

Les différentes molécules synthétisées ont ainsi été testées pour leur affinité et leur sélectivité pour les structures G-quadruplexes par deux techniques biophysiques : le FID et le FRET-melting. La réaction de cycloaddition catalysée au cuivre a ensuite été évaluée à faible concentration en milieu aqueux en l'absence ou en présence de structures G-quadruplexes dans le but d'optimiser l'application de cette réaction *in cellulo*. La reconnaissance des composés PDC-BrdU liés à des structures G-quadruplexes par un anticorps spécifique de la 5-BrdU a ensuite été évaluée par un test ELISA modifié. Enfin, l'immunodétection des ligands PDC-BrdU a été évaluée *in cellulo*. Dans un premier cas, les composés PDC-BrdU sont incubés en cellules A549 pendant 16 h et l'incubation par l'anticorps primaire anti-BrdU puis par les anticorps secondaires fonctionnalisés par des fluorophores est réalisée (Figure 59, voie 1.). La répartition des ligands est ensuite directement évaluée par imagerie cellulaire. Dans un deuxième cas, les précurseurs PDC-Alk/N3 sont incubés en cellules vivantes pendant 16 h et la réaction avec le partenaire 5-BrdU-Alk/N3 est réalisée après fixation des cellules (Figure 59, voie II). L'immunodétection est ensuite réalisée après incubation successif des anticorps primaires et secondaires.



Figure 59: Schéma représentatif de l'imagerie par immunofluorescence. I.a) Incubation des cellules avec les dérivés PDC-BrdU, I.b) Reconnaissance de l'haptène 5-BrdU par l'anticorps primaire anti-BrdU, I.c) Incubation par des anticorps secondaires fonctionnalisés par des fluorophores spécifiques des anticorps primaires. II.a) Incubation des cellules par les dérivés du PDC fonctionnalisés par un alcyne ou un azoture, II.b) Réaction de chimie click de type CuAAC *in cellulo* entre les dérivés du PDC et les dérivés de la 5-BrdU.

2. Synthèse des ligands pour la méthodologie G4-GIS

a. Synthèse des précurseurs PDC-Alk/N3

Comme présenté précédemment, quatre dérivés du ligand **PDC** sont synthétisés. Ces dérivés sont fonctionnalisés par un linker dont la taille et la polarité sont modulées pour vérifier les potentiels effets sur l'affinité envers les G-quadruplexes ainsi que sur la pénétration cellulaire. Ces linkers sont fonctionnalisés à leur extrémité par une fonction alcyne ou azoture. Tous ces composés sont synthétisés à partir de l'intermédiaire commun **PDC-CI** (Schéma 1).



Schéma 1 : Structures des composés PDC-Alk/N3 et de l'intermédiaire commun PDC-Cl.

La synthèse du **PDC-CI** a premièrement été décrite en 3 étapes en 2016²⁰³ puis en deux étapes en 2018¹²⁴ à partir de l'acide chélidamique. Elle implique la formation du dichlorure d'acyle ainsi que la chloration du cycle pyridine sur la forme phénol en position 4 par réaction d'un excès de chlorure de thionyl en présence d'une quantité catalytique de DMF. Après 24 h, le milieu réactionnel est concentré sous vide et la formation de l'intermédiaire trichloré est vérifiée par RMN. Ce dernier est ensuite directement additionné sur un excès de 3-aminoquinoline en condition basique pour former le **PDC-CI** avec 75% de rendement en deux étapes (Schéma 2).



Schéma 2: Synthèse en deux étapes du PDC-Cl à partir de l'acide chélidamique.

Les trois composés **PDC-4,2-Alk**, **PDC-4,3-Alk** et **PDC-4,PEG-N3** sont ensuite préparés suivant la même voie synthétique (Schéma 3). La synthèse est débutée par la substitution nucléophile aromatique (S_NAr) de la position 4 de la pyridine du **PDC-Cl** par la 1,4-diaminobutane comme publié précédemment²⁰⁴ pour obtenir le composé **PDC-4-NH**₂ avec 90% de rendement. Ce dernier est ensuite dérivé par réaction de couplage peptidique en présence d'EDCI et HOBt avec trois différents acides carboxyliques commerciaux : l'acide 4-butynoique, l'acide 5-pentynoique et l'acide 11-azido-3,6,9-trioxaundecanoique avec respectivement 77, 70 et 70% de rendement. Ces trois composés sont finalement méthylés sur les azotes intracycliques des quinolines à l'aide d'un excès d'iodométhane

pour donner les composés **PDC-4,2-Alk**, **PDC-4,3-Alk** et **PDC-4,PEG-N3** avec respectivement 95, 93 et 76% de rendement.



Schéma 3: Voie de synthèse des composés PDC-4,2-Alk, PDC-4,3-Alk et PDC-4,PEG-N3.

Le dernier composé **PDC-4,0-N3** possédant une chaine courte est obtenu selon une voie synthétique différente (Schéma 4). La synthèse est débutée par une réaction de SNAr sur le composé **PDC-CI** par la 4-amino-1-butanol pour donner le composé **9** avec 66% de rendement après purification par chromatographie flash. Le composé **10** est ensuite formé par réaction de chlorure de mésyle sur la fonction hydroxyle en présence de base. La réaction doit être contrôlée attentivement pour éviter la formation du produit d'élimination. Le composé **11** est ensuite obtenu après réaction de substitution nucléophile S_N2 avec l'azoture de sodium. Finalement, le dérivé **PDC-4,0-N3** est obtenu par méthylation des deux quinolines à l'aide d'iodomoéthane. La réaction est réalisée en une nuit à 40 °C et des lavages successifs permettent d'obtenir le composé avec 87% de rendement.



Schéma 4: Voie de synthèse du composé PDC-4,0-N3.

Quatre précurseurs **PDC-Alk/N3** possédant différents bras espaceurs ont ainsi été synthétisés et seront ensuite testés pour leur affinité et leur sélectivité envers les structures G-quadruplexes. Ces précurseurs seront de plus couplés avec le partenaire **5-BrdU-Alk/N3** adapté pour former les dérivés **PDC-BrdU**.

b. Synthèse des dérivés 5-BrdU-Alk/N3

Les deux dérivés de la 5-BrdU fonctionnalisés en position 3' du cycle désoxyribose par une fonction alcyne (**5-BrdU-Alk**) ou une fonction azoture (**5-BrdU-N3**) sont synthétisés par deux routes synthétiques différentes de 3 et 4 étapes respectivement (Schéma 5). Ces deux composés sont synthétisés à partir de l'intermédiaire commun **13**. Ce dernier est formé par la protection de l'alcool primaire en position 5' par réaction de 1,1 équivalent de TBDMSCI en présence de DMAP et d'imidazole. La réaction est sélective de l'hydroxyle en position 5' dans ces conditions et permet d'obtenir le composé monoprotégé avec 79% de rendement après purification.

La synthèse du composé **5-BrdU-Alk** est ensuite réalisée par substitution nucléophile du bromure de propargyl après déprotonation de l'alcool en position 3' par réaction d'hydrure de sodium. L'alcool en position 5' est ensuite déprotégé par réaction de TBAF 1M dans le THF pour donner le composé **5-BrdU-Alk** avec un rendement de 49% sur trois étapes.

La synthèse du composé **5-BrdU-N3** est continuée à partir du composé monométhylé **13** par un réarrangement de Pummerer modifié²⁰⁵ par réaction de l'anhydride acétique en présence d'acide

acétique et de DMSO. Le composé **16** est ainsi obtenu avec un rendement de 94% après purification. Le groupement thioéther est ensuite oxydé en bromure de sulfonium en présence de dibrome puis substitué par réaction de l'azoture de sodium pour former le composé **17** avec un rendement de 53% après purification²⁰⁶. La position 5' est finalement déprotégée par réaction de TBAF pour donner le composé **5-BrdU-N3** avec un rendement de 25% sur quatre étapes.



Schéma 5: Synthèse des composés 5-BrdU-Alk et 5-BrdU-N3.

Les deux partenaires **5-Brdu-Alk** et **5-BrdU-N3** ont été synthétisés avec de bons rendements en 3 et 4 étapes respectivement. Ces molécules seront donc mises à réagir par réaction de CuAAC avec les dérivés du PDC décrits dans la partie précédente.

c. Réaction de cycloaddition alcyne azoture catalysée au cuivre

La réaction de cycloaddition alcyne azoture catalysée au cuivre est une réaction de cycloaddition 1,3dipolaire entre un alcyne (dipolarophile) et un azoture (fonction 1,3-dipolaire). La réaction de cycloaddition 1,3-dipolaire entre un alcyne et un azoture est connue depuis la fin du 19^{ème} siècle lorsque la première synthèse de triazoles a été publiée par A. Michael²⁰⁷. Elle a ensuite été largement développée et rendue populaire au milieu du 20^{ème} siècle par les travaux de R. Huisgen²⁰⁸. Cette réaction alors non catalysée nécessitait des conditions thermiques élevées et produisait un mélange des isomères 1,4 et 1,5 ce qui rendait son utilisation limitée (Schéma 6).



Schéma 6: Réaction de cycloaddition alcyne azoture formant un mélange d'isomères 1,4 et 1,5.

C'est en 2002 que sont publiés deux travaux, l'un de Sharpless²⁰⁹ et l'autre de Meldal²¹⁰, reportant tous deux l'effet catalytique du cuivre sur cette réaction. En effet, ces deux travaux démontrent que l'atome de cuivre au degré d'oxydation +I permet de catalyser la réaction en produisant exclusivement l'isomère 1,4. Le mécanisme réactionnel premièrement postulé par Sharpless impliquait alors la formation d'une espèce intermédiaire alcyne-Cu-azoture. Ce mécanisme a évolué au cours des années jusqu'en 2013 où un mécanisme réactionnel impliquant deux atomes de cuivre est proposé par l'équipe de Fokin²¹¹ (Figure 60). Ce mécanisme implique la formation d'une liaison σ et d'une liaison π par deux atomes de cuivre différents avec la fonction alcyne et a été démontré par différentes études^{212,213}.



Figure 60: Mécanisme réactionnel de la réaction de cycloaddition alcyne azoture impliquant deux atomes de cuivre. Image tirée de Fokin *et al.*²¹¹

La catalyse de cette réaction par le cuivre permet de l'insérer parfaitement dans le concept de chimie click énoncé dans les années 2000 par Sharpless²¹⁴. En effet, c'est une réaction bio-orthogonale fonctionnant souvent avec de très bons rendements dans des conditions douces, ne produisant pas

ou peu de sous-produits et qui peut s'effectuer en milieu aqueux. Cette réaction nécessite cependant la formation *in situ* de cuivre au degré d'oxydation +I, une espèce instable qui s'oxyde spontanément en cuivre au degré d'oxydation +II par l'action de dioxygène. Par conséquent, appliquer cette réaction directement *in cellulo* nécessite l'ajout d'agents réducteurs, très généralement de l'ascorbate de sodium, ainsi que l'ajout de ligands du Cu(I). Ces derniers sont des chélatants du cuivre au degré d'oxydation +I et permettent de le stabiliser même en présence d'oxygène^{215,216} (Figure 61).



Figure 61: Exemples de ligands du cuivre au degré d'oxydation +I.

Dans le but d'obtenir les composés finaux **PDC-BrdU**, plusieurs conditions de réaction ont été testées comme différentes quantités de CuSO₄, différentes quantités d'ascorbate de sodium ainsi que l'ajout ou non de ligand du cuivre TBTA. Finalement, pour un équivalent de dérivé **PDC**, 1,2 équivalent de partenaire **5-BrdU** était ajouté ainsi que 30 mol% de CuSO₄ et 1,2 équivalent d'ascorbate de sodium dans un mélange DMF/H₂O (v/v 3/1). Initialement, 0,8 équivalent de TBTA était ajouté dans le milieu pour s'affranchir du risque de l'oxydation du Cu(I). Cependant, la présence des deux charges positives des molécules **PDC-BrdU** ainsi que la grande quantité de réactifs impliquaient une purification de ces derniers par HPLC préparative (HPLCprep). Cette étape de purification a mené à la dégradation presque totale des différents composés synthétisés et il est aujourd'hui difficile de rationaliser cette dégradation. Plusieurs hypothèses ont été émises comme une dégradation induite par le cuivre ou par les conditions acides nécessaires à l'élution par HPLCprep, mais aucune d'entre elles n'ont été validées à ce jour. Finalement, il a été possible de s'affranchir de l'ajout de TBTA en dégazant tous les solvants et les réactifs ainsi qu'en faisant bien attention à garder le milieu sous atmosphère inerte. Dans ces conditions, les produits finaux étaient produits en maximum 16 h de réaction (Schéma 7).



Schéma 7: Réaction de CuAAC entre les précurseurs PDC-Alk/N3 et les partenaires 5-BrdU-Alk/N3 pour générer les composés PDC-BrdU.

Une fois la conversion totale des réactifs observée par LC/MS, le milieu réactionnel est directement ajouté à une résine Chelex pré-tamponnée à pH 7 par un tampon d'acétate d'ammonium. La résine Chelex est une résine recouverte de groupements carboxylates capables de chélater le cuivre efficacement. Le milieu réactionnel et la résine chelex sont mélangés doucement pendant une trentaine de minutes avant d'être filtrés sur un filtre en coton. La résine est rincée deux fois avec de la DMF et le milieu réactionnel est finalement évaporé à l'aide d'un concentrateur SpeedVac. Ce dernier permet de pouvoir évaporer la DMF sans avoir besoin de chauffer à plus de 30 °C. Le produit résultant est ensuite précipité par addition de DCM, filtré et lavé plusieurs fois avec du DCM et de l'eau, permettant d'obtenir les composés **PDC-4,2-BrdU**, **PDC-4,3-BrdU**, **PDC-4,0-BrdU** et **PDC-4,PEG-BrdU** avec 61, 68,78 et 48% de rendement respectivement après séchage sous vide.

Ainsi, les différents composés précurseurs **PDC-Alk/N3** et les différents composés finaux **PDC-BrdU** ont été synthétisés avec succès et sont testés pour leur affinité et leur sélectivité pour les structures G-quadruplexes par FRET-melting et FID. Une expérience ELISA modifiée a de plus été réalisée pour

un composé sur trois séquences dans le but de vérifier la bonne reconnaissance de la 5-BrdU par l'anticorps primaire lorsque le ligand PDC-BrdU est lié au G-quadruplex.

3. Étude de l'affinité des ligands

a. Séquences utilisées pour les analyses biophysiques

Dans le but de mesurer l'affinité et la sélectivité des ligands pour les structures G-quadruplexes, plusieurs séquences G4s ont été étudiées. Toutes ces séquences forment des structures bien définies en solution ce qui nous permettra de rationaliser les résultats obtenus. Un panel de 8 structures G-quadruplexes (6 séquences ADN, 2 séquences ARN) représentant les différentes topologies (parallèle, antiparallèle et hybride) a été testé par FID et FRET-melting. Dans le but de vérifier la sélectivité des ligands, deux séquences ont été utilisées en compétition : une séquence d'ARN de transfert (ARNt) pour évaluer la sélectivité pour l'ARN-G4 et une séquence d'ADN duplex (ds26) pour étudier la sélectivité de l'ADN-G4.



Parallèle 5'-AAGGGTGGGTGTAAGTGTGGGTGGGT-3'

Figure 62: Représentation de la structure G-quadruplex adoptée par la séquence CEB25wt.

La séquence CEB25wt (Figure 62) est une séquence de 26 bases d'ADN qui est localisée au sein du minisatellite CEB25. Ce minisatellite possède un ratio de bases G/C de 53%, mais la quasi-totalité (92%) des guanines se trouve sur le même brin, faisant de ce minisatellite un réel foyer à structures G4⁶⁵. Cette séquence adopte une topologie uniquement parallèle en tampon K⁺, possédant deux boucles d'un nucléotide ainsi qu'une longue boucle de 9 nucléotides au niveau de la boucle centrale. Malgré cette longue boucle de 9 nucléotides, cette séquence possède une certaine stabilité

thermique avec une température de dénaturation de 55,1 °C par UV-Vis dans un tampon KPi (phosphate de potassium) 1 mM et 10 mM posphate de lithium.²¹⁷



Figure 63: Représentation de la structure G-quadruplex adoptée par la séquence CEB25L111T.

La séquence CEB25L111T (Figure 63) est une séquence directement dérivée de la séquence CEB25wt. En effet, elle possède la même topologie parallèle, mais possède une courte boucle centrale composée d'uniquement une thymine. L'étude de la stabilité thermique de ces séquences en fonction de la taille de la boucle centrale a montré que plus la boucle était courte et plus la structure était stable. En effet, la température de dénaturation de cette séquence a été mesurée à 73,4 °C par UV-Vis dans un tampon KPi 1 mM et 10 mM posphate de lithium.^{218,217}



Parallèle 5'-TGAGGGTGGGTAGGGTAG-3'

Figure 64: Représentation de la structure G-quadruplex adoptée par la séquence cMyc22-G14T-G23T.

La séquence Myc22-G14T/G23T (abrévié cMyc22 ou Myc22, Figure 64) est une séquence complètement parallèle de 22 nucléotides, dérivée de la séquence Pu27 naturellement présente au niveau d'un promoteur de l'oncogène c-Myc (NHE III1) responsable, lors de sa surexpression, de la prolifération non contrôlée de presque tous les cancers humains.⁸¹ La séquence Pu27 (5'-TG₄AG₃TG₄AG₃TG₄AAGG-3') est très polymorphique en solution et forme plusieurs structures G4s différentes²¹⁹, la séquence Myc22-G14T/G23T a donc été développée pour en faciliter l'étude. Comme décrit dans l'introduction, de nombreuses molécules sont aujourd'hui développées dans le but de stabiliser spécifiquement cette structure, permettant ainsi de bloquer la transcription non-contrôlée de ce gène.^{157,147}



Figure 65: Représentation de la structure G-quadruplex adoptée par la séquence cKit2.

La séquence cKit2 est composée de 20 nucléotides et adopte une topologie parallèle⁶² (Figure 65). Cette séquence est présente au niveau d'un promoteur de l'oncogène c-kit qui est aujourd'hui une importante cible thérapeutique. En effet, la dérégulation du gène c-kit a été observée dans de nombreux cancers et notamment dans plus de 80% des tumeurs stromales gastro-intestinales.²²⁰



Figure 66: Représentation de la structure G-quadruplex adoptée par la séquence 22AG.

La séquence 22AG (Figure 66) est une séquence de 22 nucléotides provenant du brin d'ADN télomérique humain. Cette séquence possède 3 répétitions (GGGTTA) et peut former 2 différentes structures hybrides 3+1 en tampon K⁺ et une structure antiparallèle en tampon Na⁺. Ces structures sont monomoléculaires, cependant, une diminution du nombre de répétitions de (GGGTTA) au sein du brin permet la formation de nombreuses autres structures bimoléculaires.⁶⁴



5'-GGGCTAGGGCTAGGGCTAGGG-3'

Figure 67: Représentation de la structure G-quadruplex adoptée par la séquence 21CTA.

La séquence 21CTA (Figure 67) est une séquence de 21 bases d'ADN dérivée de la séquence télomérique humaine remplaçant le motif (TTAGGG) par (CTAGGG). La modification de la base thymine par une base cytosine permet de favoriser une structure de topologie uniquement antiparallèle comme démontré par RMN²²¹. Cette structure implique la formation d'un G-quadruplex composé de deux quartets G•G•G•G ainsi que deux interactions Watson-Crick G•C G•C.



Figure 68: Représentation de la structure G-quadruplex adoptée par la séquence TERRA.

La séquence TERRA est la séquence d'ARN résultant de la transcription de l'ADN télomérique humain. Cette structure possède une topologie parallèle²²² et est notamment nécessaire à la formation du complexe shelterin²²³. Il a été montré que des séquences courtes d'ARN télomérique (nt < 12) forment des structures multimériques et des séquences longues (nt > 40) forment des empilements de plusieurs structures G-quadruplexes⁶⁴.



5'-rGGGAGGGGCGGGUCUGGG-3'

Figure 69 : Représentation de la structure G-quadruplex adoptée par la séquence NRAS.

La séquence NRAS est une séquence d'ARN formant une structure G-quadruplex parallèle retrouvée au niveau du promoteur de l'oncogène NRAS. Ce dernier fait partie de la famille des oncogènes RAS, responsables de la prolifération incontrôlée des cellules cancéreuses. La formation de cette structure G-quadruplex au sein du promoteur de l'oncogène NRAS permet notamment de réguler le niveau de traduction de la protéine associée¹²².

b. Étude par FRET-melting

Comme décrit précédemment, l'expérience de FRET-melting permet de caractériser la capacité d'une molécule à stabiliser une structure G-quadruplex contre la dénaturation thermique^{224,165}. Les séquences utilisées sont fonctionnalisées en 5' et en 3' par la fluorescéine (FAM) et la tétraméthylrhodamine (TAMRA) respectivement. Les séquences d'ADN et d'ARN G-quadruplexes sont dénaturées dans le tampon adapté en chauffant à 95 °C pendant 10 min et sont refroidies à 4 °C pendant 3 h. Les séquences ds26 et tRNA sont chauffés à 95 °C pendant 10 min et sont laissées à refroidir lentement à température ambiante sur 3 h. Les concentrations utilisées sont de 0,2 µM pour les séquences G-quadruplexes et de 1 µM pour les ligands. La sélectivité des ligands est évaluée en ajoutant 3,0 μ M d'ARNt pour les structures ARN-G4 et 3 μ M et 10 μ M de ds26 pour les structures ADN-G4. L'ARNt est composé de 70 à 100 nucléotides et forme une structure secondaire complexe en forme de trèfle composée de trois boucles intramoléculaires ainsi que quatre régions double brin. Dans le but de vérifier la température de dénaturation des séquences G4s seules, elles sont toujours utilisées en références sans ligands avec 0, 3 ou 10 µM de compétiteurs. Pour observer une stabilisation induite par les ligands, les séquences G-quadruplexes ne doivent pas avoir une température de dénaturation trop élevée. Ainsi, il est possible de moduler cette température en modifiant la concentration d'ions K^+ : les séquences très stables sont utilisées en tampon $K^+1(1 \text{ mM})$ KCl, 99 mM LiCl, 10 mM caco.Li, pH = 7,2) et les séquences moins stables sont utilisées en tampon K^+10 (10 mM KCl, 90 mM LiCl, 10 mM caco.Li, pH = 7,2). Les températures de dénaturation des séquences G4 avec et sans compétiteurs sont reportées dans le tableau ci-après (Tableau 1).

Nom	Séquence	T _{1/2} / °C (0/3/10 μM compétiteur)				
FCEB25wtT	5'-F-(AGGGTGGGTGTAAGTGTGGGTGGG)-T	55,8 / 55,3 / 54,4 (K + 1)				
FCEB25L111T	5'-F-(AGGGTGGGTGGGTGGGT)-T	68,5 / 70,0 / 70,6 (K⁺1)				
FcMyc22T	5'-F-(TTGA <mark>GGGTGGG</mark> TA <mark>GGGTGGG</mark> TAA)-T	52,4 / 52,0 / 51,0 (K+1)				
F21CTAT	5'-F-(<mark>GGG</mark> CTA <mark>GGG</mark> CTA <mark>GGG</mark> CTA <mark>GGG</mark>)-T	59,1 / 59,6 / 59,8 (К⁺10)				
F21T	5'-F-(<mark>GGG</mark> TTAGGGTTAGGGTTAGGG)-T	54,6 / 54,0 / 53,6 (K * 10)				
FKit2T	5'-F-(<mark>GGGCGGG</mark> CGCGA <mark>GGGAGGG</mark> G)-T	58,1 / 57,7 / 57,5 (K⁺1)				
F21RT	5'-F-r(<mark>GGG</mark> UUA <mark>GGG</mark> UUA <mark>GGG</mark> UUA <mark>GGG</mark>)-T	63,9 / 63,9 (K * 1)				
FNRAST	5'-F-r(<mark>GGGAGGGGCGGGUCUGGG</mark>)-T	68,2 / 63,9 (K +1)				

Tableau 1 : Noms, séquences et températures de dénaturation des différents oligonucléotides utilisés pour les expériences de FRET-melting.

Ces expériences ont été réalisées dans des plaques 96 puits à l'aide d'une PCR 7900HT Real-Time. Après une première incubation de la plaque à 25 °C pendant 5 min, un gradient de 25 à 95 °C est entamé par augmentation de 0,5 °C/min et la mesure de la fluorescence est réalisée à chaque cycle. La fluorescence de la FAM en fonction de la température est ensuite normalisée et la différence de température nécessaire à restaurer la moitié de la fluorescence de la FAM avec et sans ligand est reportée ($\Delta T_{1/2}$, Figure 70).



Figure 70: Exemple de courbes obtenues par FRET-melting de la séquence F21T sans ligand (0 μ M ds26 courbe noire, 3 μ M ds26 courbe rouge, 10 μ M ds26 courbe bleue) et avec le ligand PDC (0 μ M ds26 courbe magenta, 3 μ M ds26 courbe verte, 10 μ M ds26 courbe bleu foncé). La variation de la température de fusion induite par le ligand est notée $\Delta T_{1/2}$.

Ainsi, la température de stabilisation induite par les ligands a été évaluée sur toutes les séquences précédemment citées et les valeurs de $\Delta T_{1/2}$ des séquences ADN-G4 sont présentées ci-après (Figure 71). Les expériences ont été réalisées en triplicat et les barres d'erreurs représentent l'écart-type des trois valeurs obtenues.





Pour toutes les séquences G-quadruplexes, le ligand **PDC** non fonctionnalisé est ajouté en tant que ligand de référence pour évaluer l'impact de la fonctionnalisation sur la stabilisation des structures G-quadruplexes. Ainsi, les données obtenues par FRET-melting indiquent une très forte stabilisation induite par les différents ligands synthétisés et plusieurs éléments peuvent en être retirés :

- Solution of the second provided the second provided the second provided to the second tot the secon
- Les différents ligands ne stabilisent pas les structures G-quadruplexes avec autant d'efficacité. Si l'on prend en compte la température de dénaturation propre à chacun des oligonucléotides, on observe que le ligand PDC et les précurseurs PDC-Alk/N3 stabilisent les séquences F21T, FcMyc22T, F21CTAT, FCEB25L111T et FKit2T très efficacement avec une température de dénaturation d'environ 85 °C tandis qu'ils stabilisent la séquence FCEB25wtT jusqu'à environ 75 °C. Cela peut notamment être expliqué par la présence de la longue boucle centrale de 9 nucléotides qui peut donner une certaine dynamique à la structure et empêcher les ligands d'interagir par π-stacking avec autant d'efficacité que pour les autres structures.
- Tous les ligands stabilisent efficacement les structures G-quadruplexes en présence de compétiteur duplex. En effet, l'ajout de 3 μM (15 équivalents) ou de 10 μM (50 équivalents) ne semble pas impacter fortement les températures de stabilisation induites par les

composés précurseurs **PDC-Alk/N3** ou les composés finaux **PDC-BrdU**. La baisse maximum de $\Delta T_{1/2}$ induite par l'ajout de compétiteur est d'environ 4 °C pour le composé **PDC-4,3-Alk** sur la séquence FCEB25L111T qui passe ainsi de 15,8 °C à 11,5 °C avec 0 et 3 µM de ds26. Il est cependant important de noter que l'ajout d'ADN compétiteur peut modifier légèrement la température de dénaturation de la séquence seule. Ici, la séquence FCEB25L111T possède une température de dénaturation de 68,5 °C sans compétiteur et de 70,0 °C en présence de 3 µM de ds26. La baisse de $\Delta T_{1/2}$ observée est donc aussi partiellement due à l'augmentation de la température de dénaturation de la séquence. L'excellente sélectivité observée pour les dérivés **PDC-Alk/N3** et **PDC-BrdU** pour l'ADN G-quadruplex par rapport à l'ADN duplex présente donc un réel avantage pour l'utilisation de ces molécules *in cellulo*.

La longueur du bras espaceur n'a que très peu d'influence sur la stabilisation des structures G-quadruplexes. Notamment, aucune différence de $\Delta T_{1/2}$ n'est observable pour les différents composés précurseurs PDC-Alk/N3. Que le bras espaceur soit long (PDC-4,PEG-N3), court (PDC-4,0-N3) ou de taille moyenne (PDC-4,2/4,3-Alk), les températures de stabilisation sont similaires d'un composé à un autre. Dans le cas des composés PDC-BrdU, quelques différences selon les séquences sont notables, notamment pour les deux dérivés PDC-4,PEG-BrdU et PDC-4,0-BrdU. Par exemple, le composé PDC-4,0-BrdU stabilise légèrement moins efficacement la structure cMyc22 que les trois autres composés ($\Delta T_{1/2} \simeq 10$ °C contre $\Delta T_{1/2} \simeq$ 15 °C). La structure G-quadruplex formée par la séquence cMyc22 est très compacte et un bras espaceur trop court peut certainement impacter le caractère π -stacking du PDC. Le composé PDC-4,PEG-BrdU induit, quant à lui, une stabilisation légèrement plus faible que les trois autres composés pour la séquence FCEB25wtT ($\Delta T_{1/2} \simeq 5$ °C contre $\Delta T_{1/2} \simeq 9$ °C) et une stabilisation légèrement plus haute pour la séquence FCEB25L111T ($\Delta T_{1/2} \simeq 11$ °C contre $\Delta T_{1/2}$ \simeq 8 °C). La séquence CEB25wt possède une longue boucle de 9 nucléotides notamment composée de deux adénines. Un bras espaceur plus long peut augmenter le caractère dynamique de la molécule et peut ainsi augmenter les probabilités que la 5-BrdU interagisse avec les nucléotides présents dans la boucle. Dans le cas de la séquence CEB25L111T, un bras espaceur plus long semble améliorer la capacité du ligand PDC à interagir avec le quartet externe du G-quadruplex en diminuant la gêne stérique.

La stabilisation induite par les ligands a de plus été évaluée sur deux séquences d'ARN Gquadruplexes, les données sont présentées ci-après (Figure 72).



Figure 72 : Variation de la température de fusion induite par le ligand de référence PDC, les dérivés PDC-Alk/N3 et les dérivés PDC-BrdU en présence de 0 μM (noir) et 3 μM (gris foncé) d'ARNt pour les séquences A) F21RT et B) FNRAST. Les barres d'erreurs représentent l'écart-type de trois expériences individuelles.

Dans le cas de la séquence transcrite de l'ADN télomérique humain F21RT, les composés précurseurs PDC-Alk/N3 stabilisent la structure G-quadruplex aussi efficacement que le PDC non fonctionnalisé avec un $\Delta T_{1/2} \simeq 22$ °C soit jusqu'à une température d'environ 86 °C. Les ligands fonctionnalisés par la 5-BrdU ont, quant à eux, une affinité légèrement affectée avec un $\Delta T_{1/2} \simeq 15$ °C pour les composés PDC-4,2-BrdU, PDC-4,3-BrdU et PDC-4,0-BrdU tandis que le composé PDC-4,PEG-BrdU induit une stabilisation plus faible avec un $\Delta T_{1/2}$ d'environ 9 °C. L'ajout de 3 μ M d'ARNt n'a pas d'influence sur stabilisation induite par les ligands précurseurs PDC-Alk/N3 indiquant une bonne sélectivité pour l'ARN G-quadruplex. Dans les cas des composés PDC-BrdU, une augmentation des $\Delta T_{1/2}$ d'environ 3 °C est observée pour les composés PDC-4,2/4,3-BrdU et le composé PDC-4,0-BrdU ainsi qu'une augmentation de $\Delta Tm_{1/2}$ d'environ 6 °C pour le composé PDC-4,PEG-BrdU. La température de dénaturation de la séquence F21RT ne variant pas lors de l'ajout de 3 µM d'ARNt, observer une augmentation de la stabilisation induite par les ligands est relativement étonnant. L'ajout des ligands PDC-BrdU sur la structure G-quadruplex pourrait induire de nouvelles interactions avec l'ARNt et participer à la formation d'un complexe plus stable que le complexe ligand/G4. Cette augmentation de la température de stabilisation est notamment confirmée avec la séquence FNRAST. Les ligands **PDC-Alk/N3** stabilisent fortement cette séquence avec un $\Delta T_{1/2} \simeq 20$ °C sans ARNt et avec un $\Delta T_{1/2} \simeq$ 24 °C avec 3 μ M d'ARNt. Cette différence de $\Delta T_{1/2}$ d'environ 4 °C pour les ligands précurseurs PDC-Alk/N3 est ici attribuée à la température de dénaturation plus faible de la séquence lorsqu'elle est en présence d'ARNt que lorsqu'elle est seule ($\Delta T_{1/2}$ = 68,2 contre 63,9 °C). Dans le cas des composés

PDC-BrdU, une augmentation encore plus importante de la température de stabilisation est observée. Par exemple, le composé **PDC-4,2-BrdU** stabilise la séquence FNRAST avec un $\Delta T_{1/2} \simeq 17$ °C sans ARNt et avec un $\Delta T_{1/2} \simeq 25$ °C.

L'expérience de FRET-melting a ainsi permis de montrer plusieurs caractéristiques intéressantes des composés précurseurs **PDC-Alk/N3** ainsi que des composés finaux **PDC-BrdU**. Globalement, tous les ligands stabilisent très efficacement les différentes séquences G-quadruplexes étudiées. Les différents composés sont ensuite testés par FID.

c. Étude par FID

Comme décrit dans l'introduction, l'expérience de FID est basée sur le déplacement d'une sonde « on-off » non sélective des G-quadruplexes par des ligands. Cette expérience a été réalisée sur les différentes séquences G-quadruplexes utilisées par FRET-melting ainsi que sur la séquence ds26 dans le but de vérifier la sélectivité des ligands par rapport à l'ADN duplex. Pour réaliser cette expérience, la sonde « on-off » utilisée est le thiazole orange (TO), car elle possède une bonne exaltation de fluorescence lors de l'interaction avec les G-quadruplexes ou avec la séquence ds26²²⁵ et possède un spectre d'absorbance et d'émission de fluorescence²²⁶ (Figure 73A) qui ne se recouvrent pas avec le spectre d'absorbance des ligands synthétisés (Figure 73B). Les expériences de FID ont tout d'abord été réalisées en haut débit dans des plaques 96-puits puis en cuves. Les résultats de ces deux méthodes sont discutés ci-après.



Figure 73: A) Spectre d'absorption (trait plein) et d'émission de fluorescence (trait pointillé) de la sonde TO. B) Spectres d'absorption des composés PDC-Alk/N3, PDC-BrdU et PDC à 5 µM dans le tampon K⁺100. Spectres du TO tirés de Rongeat *et al.*²²⁶

i. Étude en plaques 96-puits

Pour les expériences de FID en haut débit, des plaques 96-puits NBS (Non-Binding Surface) sont remplies avec un mélange de 10 μ L d'oligonucléotide 5 μ M (Concentration finale = 0,25 μ M) préalablement structurés dans un tampon K⁺100 pour s'assurer de la stabilité de la structure G4, et de 2 équivalents de TO pour les G-quadruplexes ou 3 équivalents de TO pour le ds26 (0,5 μ M et 0,75 μ M respectivement). Les puits sont ensuite remplis avec une concentration croissante de ligands ainsi qu'un volume adapté de tampon K⁺100 pour obtenir un volume final de 200 μ L comme précédemment publié¹⁶¹ (Figure 74).

	Concentration en ligands (µM)												
Ligand 1	0	0,125	0,25	0,375	0,5	0,625	0,75	1	1,25	1,5	2	2,5	
	0	0,125	0,25	0,375	0,5	0,625	0,75	1	1,25	1,5	2	2,5	
Ligand 2	0	0,125	0,25	0,375	0,5	0,625	0,75	1	1,25	1,5	2	2,5	
	0	0,125	0,25	0,375	0,5	0,625	0,75	1	1,25	1,5	2	2,5	
Ligand 3	0	0,125	0,25	0,375	0,5	0,625	0,75	1	1,25	1,5	2	2,5	
	0	0,125	0,25	0,375	0,5	0,625	0,75	1	1,25	1,5	2	2,5	
Ligand 4	0	0,125	0,25	0,375	0,5	0,625	0,75	1	1,25	1,5	2	2,5	
	0	0,125	0,25	0,375	0,5	0,625	0,75	1	1,25	1,5	2	2,5	

Figure 74: Représentation d'une plaque 96-puits et de la concentration en ligands pour 0,25 μ M d'ADN et 0,50 ou 0,75 μ M de TO.

La plaque est, par la suite, agitée lentement pendant 5 min avant que l'émission de fluorescence du TO ne soit mesurée à 544 nm après excitation à 485 nm à l'aide d'un lecteur de plaque. Le déplacement de fluorescence (% TOD) est ensuite calculé et tracé en fonction de la concentration en ligand comme décrit dans l'introduction (Chapitre I.5.b.iii). Les courbes de déplacements du TO obtenues sur la séquence 22AG sont présentées ci-après (Figure 75). Le **PhenDC3** et le **PDC** sont utilisés comme ligands de référence.



Figure 75: Courbes de déplacement de fluorescence du TO sur la séquence 22AG. Les barres d'erreurs représentent l'écart-type de trois expériences individuelles.

Tout d'abord, il est intéressant de noter que le **PhenDC3** est capable de déplacer le TO extrêmement efficacement avec 85% de déplacement à 1 μ M de ligand (soit 4 équivalents par rapport à l'ADN) et un déplacement quasi total avec 2 μ M de ligand. Le ligand **PDC** non fonctionnalisé déplace légèrement moins efficacement le TO que le ligand **PhenDC3** avec un TOD 1 à μ M de 52% et un déplacement maximal à 72% avec 2,5 μ M de ligand. Les 4 ligands **PDC-Alk/N3** déplacent le TO tout aussi efficacement que le PDC non-fonctionnalisé, confirmant ainsi les résultats obtenus par FRET-melting sur l'absence de différence selon la longueur du bras espaceur. L'affinité des composés finaux **PDC-BrdU** est relativement affectée par la présence de la base et semble, de plus, dépendre de la longueur de la chaine. Par exemple, le ligand **PDC-4,0-BrdU** déplace beaucoup moins efficacement la sonde avec un TOD 1 μ M de 17,1% tandis que le dérivé **PDC-4,PEG-BrdU** est relativement moins affecté avec un TOD 1 μ M de 36,9%. Les composés fonctionnalisés par un linker de taille moyenne **PDC-4,2/4,3-Alk** ont, quant à eux, une affinité plus haute que le composé **PDC-4,0-BrdU** et plus faible que le composé **PDC-4,PEG-BrdU**. Cette expérience montre donc que plus la base est éloignée de la partie PDC, plus le ligand est capable de déplacer efficacement la sonde.

Cette expérience de FID a été réalisée sur les séquences cMyc22, CEB25wt, CEB25L111T, 21CTA et ds26. Les TOD 1 µM des différents ligands sont présentés ci-après (Figure 76).



Figure 76: Graphique représentant les TOD 1 µM des ligands sur les différentes séquences étudiées. Les barres d'erreurs représentent l'écart-type de trois expériences individuelles.

Les observations réalisées pour la séquence 22AG se confirment aussi pour les autres séquences. Les composés PDC-Alk/N3 déplacent le TO aussi efficacement que le composé PDC non fonctionnalisé sur toutes les autres séquences. Ils déplacent notamment très efficacement la sonde de la séquence cMyc22 avec un TOD 1 µM d'environ 80%. Cette augmentation du TOD pour la séquence Myc22 comparée, par exemple, avec la séquence 22AG peut être expliquée par l'accessibilité du quartet externe pour les ligands. En effet, la structure cMyc22 présente uniquement des boucles de type « propeller » tandis que la structure 22AG présente au moins deux boucles latérales qui peuvent gêner l'interaction des ligands avec le G-quartet externe. Aucun des ligands PDC-Alk/N3 et PDC-BrdU ne déplacent le TO de la séquence ds26 ce qui confirme la sélectivité de ces derniers pour les structures G-quadruplexes. De plus, l'effet de la taille de la chaine des composés PDC-BrdU constaté sur la séquence 22AG est aussi observé sur toutes les autres séquences avec un effet encore plus marqué sur les séquences du minisatellite CEB25 et sur la séquence cMyc22. Cependant, le très faible déplacement du TO par les composés PDC-4,2/4,3-BrdU et le composé PDC-4,0-BrdU des séquences CEB25wt, CEB25L111T et Myc22 est très déconcertant au vu de la bonne stabilisation des Gquadruplexes observée par FRET-melting. En effet, dans ces conditions de FID en haut-débit, les composés PDC-4,2/4,3-BrdU ne semblent interagir que très faiblement avec la séquence Myc22 et ne semblent pas interagir du tout avec la séquence CEB25wt. Le composé PDC-4,0-BrdU quant à lui ne semble ni interagir avec la séquence CEB25wt ni avec la séquence cMyc22 et très faiblement avec la séquence CEB25L111T avec un TOD 1 µM de 6,6 µM. De plus, les valeurs de DC₅₀ des ligands de références **PhenDC3** et **PDC** sont plus faibles que celles reportées dans la littérature¹⁶¹. Plusieurs conditions ont été testées pour améliorer le déplacement comme l'utilisation d'autres plaques 96-puits (plaques d'autres fabricants, plaque en quartz), l'utilisation d'un robot automatisé pour réaliser la plaque (Pipetmax – Gilson[®]) ou la modification de la longueur d'onde de détection d'émission de fluorescence (520 ou 544 nm), mais rien n'améliorait significativement le déplacement de la sonde. Dans le but de vérifier si les ligands étaient tout de même capables d'interagir avec les G-quadruplexes par FID, ces derniers ont été testés par FID en cuve.

ii. Étude en cuves

Ces expériences ont été réalisées comme précédemment publié¹⁶² : une solution de 1 mL d'oligonucléotide structuré en solution K⁺100 (0,25 μ M) est préparée et 0,50 μ M (pour les séquences G-quadruplexes) ou 0,75 μ M (pour la séquence duplex) de TO est ajouté. Après un temps d'équilibration, la fluorescence du TO est mesurée dans des cuves en quartz à l'aide d'un spectrofluorimètre entre 510 et 750 nm après excitation du TO à 495 nm. La titration est ensuite réalisée par un ajout croissant de ligand (de 0,125 μ M à 2,5 μ M soit de 0,5 équivalent à 10 équivalents) avec un temps d'équilibration de 3 min après chaque ajout. La fluorescence de la sonde est ensuite mesurée après chaque ajout de ligand (Figure 77A) et le pourcentage de sonde déplacée est calculé comme décrit dans l'introduction (Chapitre I.5.b.iii). Le TOD est ensuite tracé en fonction de la concentration du ligand (Figure 77B).



Figure 77: Exemple d'une expérience de FID réalisée avec la séquence 22AG. A) Spectres de fluorescence du TO après chaque addition du ligand PDC-4,3-Alk et B) le déplacement de TO en fonction de la concentration des différents ligands. Les barres d'erreurs représentent l'écart-type de deux expériences individuelles.

L'expérience de déplacement de fluorescence en cuve sur la séquence 22AG avec les différents ligands montre une nette différence de TOD par rapport à l'expérience réalisée haut débit (Figure 75). Notamment, le ligand **PDC** non fonctionnalisé déplace bien plus efficacement le TO de la

structure G-quadruplex avec un TOD 1 μM de 75% contre 52% en haut débit. Les ligands précurseurs **PDC-Alk/N3** sont, comme pour l'expérience en plaque, capables de déplacer la sonde avec une efficacité similaire à celle du **PDC** avec un TOD entre 70 et 78%. Les ligands **PDC-BrdU** déplacent légèrement moins efficacement la sonde avec un TOD compris entre 57 et 45%. Comme pour l'analyse en haut débit, le composé **PDC-4,0-BrdU** est le moins efficace pour déplacer le TO de la séquence 22AG, mais l'effet de la longueur de la chaine semble moins important pour cette expérience. Il y a en effet une différence de 20% de TOD entre le ligand **PDC-4,0-BrdU** et le composé **PDC-4,PEG-BrdU** par FID en haut débit tandis qu'il n'y a qu'environ 10% de différence par FID en cuve. Les TOD 1 μM sont cependant bien plus élevés par FID en cuve avec une augmentation de près de 30% de TOD pour le ligand **PDC-4,0-BrdU**. Cette expérience semble mieux représenter la forte interaction observée par FRET-melting et a ainsi été réalisée sur les différentes séquences étudiées. Les pourcentages de TOD 1 μM sont représentés ci-après (Figure 78).



Figure 78: TOD 1 µM des ligands sur les différentes séquences d'ADN-G4 étudiées. Les barres d'erreurs représentent l'écart-type de deux expériences individuelles.

Les TOD 1 μ M des différents ligands sur les structures G4 montrent une forte capacité à déplacer le TO avec notamment peu de différences entre les composés **PDC-BrdU**. Ces derniers sont de plus capables d'induire un déplacement comparable aux composés précurseurs **PDC-Alk/N3** sur la séquence cMyc22. De plus, les différents ligands interagissent faiblement avec l'ADN duplex avec un TOD 1 μ M d'environ 20%. Ces ligands déplacent très fortement le TO sur les différentes structures Gquadruplexes avec des TOD 1 μ M allant de plus de 90% à minimum 45%. Le déplacement induit à 1 μ M est donc pour certains ligands déjà arrivé à un palier ce qui ne permet pas de rendre réellement compte des différences d'affinité de ces ligands. Pour étudier des ligands d'aussi forte affinité, il est nécessaire de tracer le DC₅₀ plutôt que de tracer les TOD 1 μ M. Le DC₅₀ représente la concentration nécessaire pour déplacer 50% de la sonde (Chapitre I.5.b.iii) et est plutôt utilisé pour les ligands de forte affinité tandis que le TOD 1 μ M est plutôt utilisé pour les ligands d'affinité moyenne. Les DC₅₀ obtenus pour les différents ligands sont représentés ci-après (Figure 79).



Figure 79: Valeurs de DC₅₀ obtenues pour chacun des ligands sur les différentes séquences d'ADN G-quadruplexes.

Le DC₅₀ représente donc la concentration nécessaire de ligand pour induire le déplacement de 50% de la sonde. Par conséquent, plus la valeur est faible et plus le ligand est capable de déplacer la sonde efficacement. Ici, tracer le DC₅₀ permet tout d'abord de confirmer que la fonctionnalisation du **PDC** par les différents bras espaceurs n'a que peu, voire pas du tout d'impact sur la capacité des ligands **PDC-Alk/N3** à déplacer la sonde. Les différents composés **PDC et PDC-Alk/N3** possèdent des DC₅₀ compris entre 0,2 et 0,45 µM en fonction de la séquence étudiée. Globalement, les composés **PDC-BrdU** possèdent des DC₅₀ supérieurs aux composés **PDC-Alk/N3** notamment pour les structures 22AG, 21CTA et CEB25wt. Cette diminution de la capacité à interagir avec le G-quadruplex pour ces structures peut notamment venir de la présence des boucles latérales des séquences 22AG et 21CTA ainsi que de la présence de la longue boucle centrale de la séquence CEB25wt. Le FID a de plus été réalisé sur les deux séquences d'ARN G4 TERRA et NRAS utilisé pour le FRET-melting (Figure 80).



Figure 80: Valeurs de DC₅₀ obtenues pour chacun des ligands sur les séquences d'ARN TERRA et NRAS.

Ces expériences montrent une forte affinité des ligands **PDC-Alk/N3** avec des DC₅₀ similaires au ligand de référence **PDC** compris entre 0,2 et 0,35 μ M selon la séquence et les dérivés **PDC-BrdU** interagissent avec la séquence NRAS avec des DC₅₀ de 0,3 à 0,4 μ M. Seul le dérivé **PDC-4,PEG-BrdU** interagit plus faiblement avec environ 0,7 μ M de ligand nécessaire pour déplacer 50% de la sonde de la séquence TERRA comparé à environ 0,35 μ M pour le ligand **PDC-4,0-BrdU**.

Globalement, les résultats obtenus par l'expérience de FID réalisée en cuve permettent de montrer une forte affinité des ligands synthétisés en permettent notamment de montrer une affinité plus grande des composés **PDC-BrdU** comparés à l'expérience de FID réalisée en plaque. L'avantage de réaliser cette expérience en plaque est de pouvoir analyser un grand panel de composés sur différentes séquences très rapidement. Cependant, la sensibilité plus faible d'un lecteur de plaque par rapport à un spectrofluorimètre peut conduire à des résultats sous-estimés pour certains ligands. Pour cette raison, il est nécessaire de toujours réaliser l'expérience de FID en parallèle à l'expérience de FRET-melting.

d. Étude par une expérience ELISA modifiée¹

Dans le but de vérifier la bonne reconnaissance de la base 5-BrdU par l'anticorps spécifique lorsqu'elle est attachée au ligand PDC en présence de structures G4s, une expérience modifiée d'ELISA indirect a été réalisée. Le principe de l'expérience est le suivant : un antigène est reconnu par

¹ Étude réalisée par le Dr. Daniela Verga

un anticorps spécifique (Abl) et ce dernier est ensuite reconnu par un anticorps secondaire (Abl). L'anticorps secondaire est conjugué avec une enzyme, généralement une peroxydase (HRP pour HorseRadish Peroxidase). L'ajout d'un substrat enzymatique permet ensuite de pouvoir quantifier les antigènes reconnus par les anticorps en mesurant l'absorbance du produit de la réaction.

Les résultats des expériences biophysiques indiquent que les ligands **PDC-BrdU** interagissent globalement tous aussi efficacement avec les G-quadruplexes, uniquement le composé **PDC-4,3-BrdU** est donc testé par cette méthode. Trois séquences sont étudiées ici : deux séquences formant des G-quadruplexes (hTelo dérivée de l'ADN télomérique humain et c-Myc) et une séquence formant une structure en épingle à cheveux (abrévié hp pour hairpin). Ces dernières sont fonctionnalisées par un groupement biotine et sont immobilisées à différentes concentrations (de 0,1 à 100 nM) sur une plaque recouverte de streptavidine après avoir été foldées en tampon 100 mM KCl / 10 mM Tris HCl, pH = 7,2. Le ligand **PDC-4,3-BrdU** (250 nM) est ensuite ajouté et incubé 1 h avant que les plaques ne soient bloquées à l'aide d'albumine de sérum bovin (BSA pour Bovine Serum Albumin). L'anticorps anti-BrdU de souris (AbI) et l'anticorps anti-souris couplé avec l'enzyme HRP (AbII) sont finalement incubés 1 h l'un après l'autre. Le substrat (TMB) de la peroxydase et le peroxyde d'hydrogène sont ensuite ajoutés et la réaction est arrêtée par ajout d'HCl 2 M (Figure 81A). L'absorbance du substrat est ensuite mesurée à 450 nm à l'aide d'un lecteur de plaque Fluostar Omega et est tracée en fonction de la concentration d'ADN (Figure 81B).




Une augmentation nette de l'absorbance à 450 nm est observée avec l'augmentation de la concentration d'ADN G-quadruplex démontrant la capacité de l'anticorps Abl à détecter les composés **PDC-4,3-BrdU** liés aux structures G-quadruplexes formées par hTelo et c-Myc. L'augmentation de l'absorbance en fonction de la concentration d'ADN permet ainsi de démontrer que l'anticorps Abl ne peut reconnaître qu'uniquement les ligands liés à l'ADN G-quadruplex. Cela permet de plus de pouvoir calculer une constante de dissociation pour les deux séquences G-quadruplexes après fitting des courbes suivant un modèle « one binding site ». Cette expérience a permis de montrer que le composé **PDC-4,3-BrdU** possédait une constante de dissociation de l'ordre du nM avec des K_D de 3,5 nM et 9,8 nM pour la séquence cMyc22 et la séquence hTelo respectivement, démontrant la très forte affinité de ce ligand pour les structures G-quadruplexes notamment comparé à des ligands de référence²²⁷. Finalement, aucune variation d'absorbance à 450 nm n'est observée lorsque la séquence utilisée est une séquence hp, démontrant la sélectivité du ligand **PDC-4,3-BrdU** pour les structures G-quadruplexes.

Les expériences biophysiques de FRET-melting et de FID ainsi que l'expérience ELISA avec le composé **PDC-4,3-BrdU** montrent généralement une forte affinité des différents composés synthétisés pour les structures G-quadruplexes. Les ligands non fonctionnalisés par la 5-BrdU possèdent une affinité similaire à celle du ligand de référence PDC tandis que les ligands **PDC-BrdU** possèdent une affinité relativement plus faible. Cette différence, bien que modeste, permet de confirmer la nécessité de réaliser la fonctionnalisation *in situ* à l'aide d'une réaction bioorthogonale. La réaction de cycloaddition alcyne/azoture entre les dérivés **PDC-Alk/N3** et les partenaires **5-BrdU-Alk/N3** a ainsi été étudiée dans différentes conditions pour s'assurer notamment de son fonctionnement *in cellulo*.

4. Étude de la réaction de CuAAC avec ou sans structures Gquadruplexes²

Dans le but de pouvoir appliquer la réaction de Huisgen catalysée au cuivre *in cellulo*, plusieurs conditions de réaction ont été étudiées en milieu aqueux. Notamment, la cinétique de réaction du composé **PDC-4,2-Alk** avec le composé **5-BrdU-N3** a été réalisée. Pour cette expérience, une

² Étude réalisée avec l'aide d'Anne Cucchiarini et Eugénie Laigre

première solution de CuSO₄.5H₂O (1,5 μ L, 10 mM), de THPTA (1,5 μ L, 50 mM) et de NaAsc (1,5 μ L, 40 mM) dans 11,6 μ L d'eau est préparée et est ajoutée à une solution de **PDC-4,2-Alk** (0,75 μ L, 1 mM) dans 11,6 μ L d'eau. Le partenaire **5-BrdU-N3** (1,5 μ L 1 mM) est finalement ajouté et le milieu réactionnel est analysé par HPLC après différents temps de réaction. Les concentrations finales sont :

- 25 μM **PDC-4,2-Alk**
- 50 μM 5-BrdU-N3
- 500 μM CuSO₄.5H₂O
- 2500 μM THPTA
- 2000 μM NaAsc

Un large excès d'agent réducteur (NaAsc) ainsi que de ligand du Cu(I) (THPTA) est ajouté dans le but de limiter l'oxydation de cuivre +I en cuivre +II par le dioxygène. Les cellules utilisées pour l'imagerie ne pourront en effet pas être mises sous atmosphère inerte et un excès de réactifs limitant l'oxydation du cuivre sera donc nécessaire. Le pourcentage de produit formé est obtenu par le rapport de l'aire du pic du composé PDC-4,2-BrdU formé et de l'aire du pic du composé de départ PDC-4,2-Alk. Le pourcentage de produit formé est reporté ci-après (Figure 82).



Figure 82: Pourcentage de produit PDC-4,2-BrdU formé après différents temps de réaction. Les barres d'erreurs représentent l'écart-type de deux expériences individuelles.

La cinétique de réaction montre une conversion rapide des réactifs avec environ 45% de produit formé en 30 min. Après 30 min de plus, la quantité de produit formé n'augmente que de 13% pour atteindre 58% de produit formé. Après 2 h, la réaction arrête d'évoluer avec un maximum de conversion d'environ 75%, soit une augmentation de 17% en 1 h de réaction supplémentaire. Cette étude montre que le cuivre +I s'oxyde rapidement en cuivre +II malgré un large excès d'ascorbate de sodium et de ligand THPTA. L'impact de la présence de structures G-quadruplexes sur la réaction de CuAAC a ensuite été étudié. Pour cela, trois séquences formant des G-quadruplexes ont été choisies : la séquence 22AG, la séquence cMyc22 et la séquence TERRA. Ces séquences sont structurées en tampon K⁺10 (25 µM) et ajoutées à une solution de ligand **PDC-Alk/N3** (25 µM). Cette solution est ensuite ajoutée à un mélange de CuSO₄.5H₂O (500 µM), THPTA (2500 µM) et d'ascorbate de sodium (2000 µM). Le partenaire **5-BrdU-Alk/N3** (50 µM) adapté est finalement ajouté et le mélange est laissé réagir 1 h. Le mélange réactionnel est ensuite dilué avec de l'eau et passé à travers une colonne SPE (Solid Phase Extraction) préalablement conditionnée. Cette étape permet de retirer les séquences d'ADN ou d'ARN dans le but de pouvoir analyser le milieu réactionnel par HPLC. Le pourcentage de produit formé est ensuite déterminé par HPLC. Le même protocole est appliqué aux différents ligands **PDC-Alk/N3** sans ajouter de séquences G-quadruplexes dans le but de vérifier l'impact de la présence d'ADN ou d'ARN sur la réaction. Un exemple de chromatogramme de la réaction de CuAAC entre le composé **PDC-4,3-Alk** et la base **5-Brdu-N3** sans oligonucléotide ou en présence de la séquence cMyc22 est présenté ci-après (Figure 83).



Figure 83: Exemple de chromatogramme obtenu après réaction du ligand PDC-4,3-Alk et 5-BrdU avec (trait bleu) et sans (trait noir) oligonucléotide.

Lorsque la réaction entre le composé **PDC-4,3-Alk** et le composé **5-BrdU-N3** est réalisée sans oligonucléotide dans le milieu, une conversion de 57% est observée. Ce rendement est notamment similaire à celui du composé **PDC-4,2-Alk** observé lors de l'étude de la cinétique de réaction. Après ajout d'un équivalent de la séquence cMyc22 structurée, la formation du produit **PDC-4,3-BrdU** augmente jusqu'à environ 97%, démontrant une accélération de la cinétique de réaction en présence de cette structure G-quadruplex. Cette expérience a ainsi été réalisée avec tous les dérivés sur les trois séquences formant des G-quadruplexes (Figure 84).



Figure 84: Pourcentage des différents composés PDC-BrdU formés sans oligonucléotide (barres noires) ou en présence de 25 µM de cMyc22 (barres rouges), de 22AG (barres bleues) et de TERRA (barres magenta). Les barres d'erreurs représentent l'écart-type de trois expériences individuelles.

Il est intéressant de noter que, non seulement la présence de structures G-quadruplexes n'empêche pas la formation des composés **PDC-BrdU**, mais surtout qu'elle améliore les taux de conversions. Cette augmentation significative peut notamment être attribuée au fait que l'ADN ou l'ARN peut agir comme une plateforme qui rapproche les deux réactifs l'un de l'autre comme précédemment publié sur l'ADN duplex²²⁸ et plus tard sur l'ADN et l'ARN G-quadruplex²²⁹. Cette constatation est d'autant plus significative pour les structures G-quadruplexes parallèles cMyc22 et TERRA qui présentent des quartets externes peu encombrés contrairement à la structure 22AG qui présente une boucle latérale au niveau du quartet externe. De plus, la réaction semble mieux fonctionner, après 1h, pour les composés **PDC-4,0-N3** et **PDC-4,PEG-N3** avec une conversion de presque 100% en présence des structures G-quadruplexes. Dans ce cas, la complexation du Cu(I) avec la fonction alcyne est peut-être facilitée lorsque cette dernière est présente sur la base 5-BrdU plutôt que sur la partie PDC qui interagit avec un G-quadruplex.

Ces expériences ont ainsi permis de démontrer que la réaction de cycloaddition alcyne-azoture catalysée au cuivre entre les partenaires **PDC-Alk/N3** et **5-BrdU-Alk/N3** pouvait avoir lieu en conditions aqueuse et en présence d'oligonucléotides. Notamment, les structures G-quadruplexes étudiées permettent d'améliorer le rendement de la réaction après 1 h. Finalement, les différents ligands synthétisés ont été testés *in cellulo* dans le but de déterminer leur localisation par immunofluorescence.

5. Immunodétection de ligands des G-quadruplexes in cellulo³

Avant de pouvoir étudier la répartition des ligands **PDC-BrdU** *in cellulo*, il est nécessaire de connaître la cytotoxicité de ces derniers. Pour cela, 20 000 cellules de la lignée A549 sont incubées dans 24 puits avec des concentrations croissantes de ligand et les cellules sont comptées après 72 h d'incubation (Tableau 2).

	IC₅₀ / μM	
PDC-4,2-Alk	27,8 ± 1,56	
PDC-4,3-Alk	45,8 ± 0,56	
PDC-4,0-N3	16,1 ± 0,21	
PDC-4,PEG-N3	> 100	
PDC-4,2-BrdU	53,0 ± 7,97	
PDC-4,3-BrdU	63,5 ± 19,31	
PDC-4,0-BrdU	21,2 ± 1,10	
PDC-4,PEG-BrdU	> 100	

Tableau 2 : Valeurs d'IC₅₀ obtenus pour les différents composés synthétisés déterminés par comptage de cellules vivantes A549 après 72 h d'incubation.

Les IC_{50} obtenus sur les composés avant et après fonctionnalisation par la 5-BrdU montrent une toxicité relativement modérée, notamment pour les composés à linker longs **PDC-4,PEG-N3** et **PDC-4,PEG-BrdU**. Les composés à chaine courte **PDC-4,0-N3** et **PDC-4,0-BrdU** sont plus toxiques avec des IC_{50} de 16,1 et 21,2 μ M respectivement.

Pour vérifier la possibilité de détecter les dérivés **PDC-BrdU** *in cellulo* par immunofluorescence, les cellules A549 sont premièrement incubées 16 h avec les 4 composés **PDC-4,2-BrdU**, **PDC-4,3-BrdU**, **PDC-4,0-BrdU** ou **PDC-4,PEG-BrdU** (5 μ M, 10 μ M ou 15 μ M). Les cellules sont ensuite fixées et perméabilisées avant que les différents anticorps ne puissent être incubés. L'anticorps primaire de souris (AbI) anti-BrdU est incubé pendant 1 h avant que ne soit incubé 40 min l'anticorps secondaire (AbII) anti-souris fonctionnalisé par des fluorophores Alexa Fluor 488. Un exemple d'images obtenues par microscopie à champ large (3D-SIM) après incubation des composés **PDC-BrdU** à 5 μ M est présenté ci-après (Figure 85).

³ Étude réalisée par le Dr. Daniela Verga



Figure 85: Exemple d'images de cellules A549 obtenues par microscopie à champ large sans ajout de ligand (contrôle) ou après 16 h d'incubation des différents composés PDC-BrdU (5 μM). Les noyaux sont marqués au DAPI. Les images sont acquises séparément pour le DAPI (bleu) et l'Alexa Fluor (vert) et sont ensuite fusionnées (merge). La barre d'échelle représente 20 μm.

Ces images montrent tout d'abord que l'immunodétection des ligands **PDC** couplés à la base **5-BrdU** est possible. Après 16 h d'incubation à 5 μ M, les différents ligands montrent une localisation très majoritairement cytoplasmique. Quelques foci de faible intensité sont néanmoins observables au sein des noyaux. Il est intéressant de noter que la longueur du bras espaceur des composés **PDC-BrdU** n'impacte pas la localisation du signal de fluorescence. Le contrôle réalisé sans ajout de ligand montre une absence complète de fluorescence, démontrant la spécificité de l'anticorps primaire utilisé.

Les images obtenues après incubation des cellules par 10 et 15 μ M de ligand pendant 16 h montrent une augmentation du signal de fluorescence sans modifier la répartition des ligands. L'augmentation de la concentration des ligands augmente cependant le signal du bruit de fond qui peut être attribué à un marquage non spécifique des ligands (Figure 86).



Figure 86: Exemple d'images de cellules A549 obtenues par microscopie à champ large avec une concentration croissante de ligands PDC-4,PEG-BrdU (5, 10 et 15 μM). Les noyaux sont marqués en bleu à l'aide de DAPI. La barre d'échelle représente 20 μm.

Comme démontré par les analyses biophysiques, la présence de la 5-BrdU sur la partie PDC peut impacter plus ou moins significativement l'affinité pour les structures G-quadruplexes. Dans le but de pallier cette perte, la réaction de CuAAC a été réalisée directement *in cellulo*. Un premier contrôle visant à évaluer l'impact de la présence de la base **5-BrdU-Alk/N3** sur la détection d'immunofluorescence a été réalisé. Pour cela, les cellules sont incubées avec 20 µM de composé **5-BrdU-Alk** ou **5-BrdU-N3** pendant 16 h avant de réaliser le protocole d'immunofluorescence précédemment décrit (Figure 87).



Figure 87: Images de cellules A549 obtenues par microscopie à champ large après incubation de 20 µM de 5-BrdU-Alk ou de 5-BrdU-N3. Les noyaux sont marqués en bleu à l'aide de DAPI. La barre d'échelle représente 20 µm.

L'absence de fluorescence après 16 h d'incubation des composés **5-BrdU-Alk/N3** montre ainsi que ces derniers ne sont plus présents dans le milieu après les différentes étapes de lavage. Enfin, les cellules ont été incubées avec les dérivés **PDC-4,2-Alk**, **PDC-4,3-Alk**, **PDC-4,0-N3** et **PDC-4,PEG-N3** pendant 16 h à différentes concentrations avant d'être fixées et perméabilisées. La réaction de Huisgen catalysée au cuivre est ensuite réalisée en 1 h dans le noir après ajout d'un mélange de **5-BrdU-Alk/N3** (50 µM), CuSO₄.5H₂O (100 µM), THPTA (500 µM) et d'ascorbate de sodium (5000 µM). Le protocole d'immunodétection est ensuite réalisé comme décrit précédemment (Figure 88).



Figure 88: Images de cellules A549 obtenues par microscopie à champ large après incubation de 5 µM des différents ligands précurseurs avec ou sans cuivre dans le milieu réactionnel. Les noyaux sont marqués à l'aide de DAPI.

Il est tout d'abord intéressant de noter l'absence de fluorescence lorsque le cuivre n'est pas ajouté dans le cocktail réactionnel, démontrant encore une fois la détection spécifique de la 5-BrdU lorsqu'elle est liée au ligand. Lorsque le cuivre est présent dans le milieu, le fort signal de fluorescence montre que la réaction de click fonctionne parfaitement. Les composés se localisent encore une fois majoritairement dans le cytoplasme des cellules dans ces conditions en formant quelques foci au niveau des noyaux. Les images semblent de plus indiquer un léger effet de la longueur du bras espaceur : les ligands PDC-4,0-N3 et PDC-4,PEG-N3 donnent des signaux de fluorescence légèrement plus faibles que pour les composés PDC-4,2-Alk et PDC-4,3-Alk. Notamment, le composé PDC-4,3-Alk possède une localisation particulière avec la formation de foci majoritairement perinucléaires alors que le composé PDC-4,3-BrdU fonctionnalisé *ex-situ* se localisait globalement à travers tout le cytoplasme (Figure 89). Généralement, la fonctionnalisation *in situ* permet d'obtenir un signal de fluorescence plus intense comparé à la fonctionnalisation *ex-situ*.



Figure 89 : Comparaison de la localisation des ligands PDC-4,3-BrdU et PDC-4,3-Alk après 16 h d'incubation. Les flèches blanches indiquent la présence de foci intranucléaires. La barre d'échelle représente 20 µm.

Dans le but de vérifier l'avantage de la détection de ligands des G-quadruplexes par immunofluorescence comparée à la détection de ligands fonctionnalisés par des fluorophores, le ligand **PDC-4,3-Alk** a été fonctionnalisé *in -situ* par le fluorophore **Cy5-N3** (Figure 90).



Figure 90 : Images de cellules A549 obtenues par microscopie à champ large après 16 h d'incubation du composé PDC-4,3-Alk (5 μM) avec ou sans cuivre dans le milieu réactionnel. Les noyaux sont marqués en bleu à l'aide de DAPI. La barre d'échelle représente 20 μm.

L'imagerie de fluorescence après fonctionnalisation *in situ* du ligand **PDC-4,3-Alk** par le fluorophore **Cy5** aboutit à un marquage moins bien défini du cytoplasme avec un signal de fluorescence diffus comparé à l'immunodétection de la 5-BrdU (Figure 89). Dans ces conditions, le ligand **PDC-4,3-Alk** se localise au niveau du cytoplasme et s'accumule dans les régions périnucléaires mais ne génère pas de foci comme avec le protocole d'immunodétection. Cette dernière permet ainsi de pouvoir déterminer avec une meilleure résolution la localisation précise des ligands au sein des cellules.

Ensuite, dans le but de déterminer la cible principale des ligands, la digestion de l'ARN simple brin a été réalisée par traitement des cellules à la RNase A. Au vu de la localisation principalement cytoplasmique et des données biophysiques, il est en effet probable que la cible principale des ligands soit l'ARN présent dans le cytoplasme des cellules. Pour cela, les cellules ont été incubées pendant 16 h avec 5 μM des différents ligands et la dégradation de l'ARN est réalisée après fixation et perméabilisation des cellules par traitement de 0,1 mg/mL de RNase A pendant 1 h à 37 °C. Dans le cas des composés **PDC-Alk/N3**, la réaction de Huisgen catalysée au cuivre est réalisée avant la dégradation de l'ARN. L'immunodétection est ensuite réalisée comme précédemment décrit (Figure 91).



Figure 91: Images de cellules A549 obtenues par microscopie à champ large après 16 h d'incubation des différents ligands (5 μM) avec ou sans traitement par la RNase A. Les noyaux sont marqués en bleu à l'aide de DAPI. Les barres d'échelles représentent 20 μm.

La dégradation de l'ARN par le traitement à la RNase A montre une diminution drastique de la fluorescence pour tous les composés, démontrant la préférence des ligands pour les structures à ARN *in cellulo*. De plus, le traitement à la RNase A permet de révéler la présence de foci intranucléaires, auparavant masqués par la forte intensité du signal de fluorescence cytoplasmique (Figure 92).



Figure 92: Agrandissement d'une image de cellules A549 obtenues par microscopie à champ large après 16 h d'incubation du ligand PDC-4,2-BrdU (5 μM) et après traitement par la RNase A. Les noyaux sont marqués en bleu à l'aide de DAPI. Les barres d'échelles représentent 20 μm.

Cette observation permet de montrer que les différents ligands interagissent aussi avec l'ADN cellulaire et prouve la très bonne sensibilité de l'immunodétection de ligands des G-quadruplexes.

Les différentes expériences réalisées ont permis de souligner l'efficacité de la méthode G4-GIS pour la détection de ligands des G-quadruplexes *in cellulo* notamment comparé à la détection de ligands fonctionnalisés par des fluorophores. Les différents composés se localisent globalement uniformément dans le cytoplasme des cellules excepté le composé **PDC-4,3-Alk** qui se localise préférentiellement dans une région périnucléaire de la cellule. La diminution de la fluorescence après traitement à la RNase A ainsi que cette localisation particulière peut notamment indiquer une localisation au niveau du réticulum endoplasmique. Ce dernier est un organite riche en ARN ribosomique (ARNr) qui peut notamment former des structures G-quadruplexes *in cellulo*²³⁰.

Finalement, bien que les données *in vitro* montrent une excellente sélectivité des ligands pour les structures G-quadruplexes plutôt que pour l'ADN duplex ou l'ARNt, il est nécessaire de vérifier si les ligands sont capables d'interagir avec les G-quadruplexes *in cellulo*. Pour ce faire, une des expériences les plus répandues est de vérifier la colocalisation du signal de fluorescence attribué au ligand avec le signal de fluorescence d'un anticorps spécifique des G-quadruplexes^{231,108}. Ainsi, après l'incubation du ligand **PDC-4,3-Alk** (5 et 15 μM) et la réaction de click *in situ*, l'anticorps BG4 et l'anticorps anti-BrdU ont été co-incubés. L'immunodétection de BG4 (Alexa Fluor 555) et de l'anticorps anti-BrdU (Alexa Fluor 488) est finalement réalisée (Figure 93).



Figure 93: Image d'une cellule A549 obtenue par microscopie à champ large après immunodétection des anticorps anti-BrdU (Alexa Fluor 488) et des anticorps BG4 (Alexa Fluor 555). Le noyau est marqué en bleu à l'aide de DAPI. La barre d'échelle représente 10 μm.

L'immunofluorescence des anticorps BG4 révèle la formation de foci au niveau du noyau ainsi qu'au niveau du cytoplasme des cellules. Les foci formés par les composés **PDC-4,3-Alk** sont localisés, comme précédemment observé, majoritairement au niveau du cytoplasme et dans une moindre mesure au niveau du noyau. Cependant, la superposition des deux signaux montre une colocalisation non significative des anticorps anti-BrdU et des anticorps anti-BG4. Le composé **PDC-4,3-Alk** et l'anticorps BG4 possédant une constante de dissociation sensiblement identique (de l'ordre du nM⁹³), l'absence de colocalisation peut probablement être expliquée par l'encombrement stérique des anticorps. Dans ce cas, deux anticorps primaires sont dirigés vers une même cible : la structure G-quadruplex pour BG4 et le complexe G-quadruplex/ligand **PDC-4,3-Alk** pour l'anticorps anti-BrdU. La détection de BG4 se fait ensuite par incubation successive d'anticorps secondaires et tertiaires et la détection de l'anticorps anti-BrdU se fait par incubation d'un anticorps secondaire (Figure 94). Au total, 5 anticorps différents doivent se retrouver spatialement proches pour observer une colocalisation.



Figure 94: Schéma représentatif des différents anticorps nécessaires à l'observation d'une colocalisation entre les foci dus de l'immunodétection de la 5-BrdU et de l'immunodétection de BG4.

D'autres méthodes permettant de prouver la reconnaissance des structures G-quadruplexes *in cellulo* pourront être étudiées par la suite. Par exemple, réaliser l'immunodétection des différents composés au sein de lignées cellulaires dont l'expression d'hélicases sélectives des G-quadruplexes a été modulée pourrait influer sur le nombre de foci observé. En effet, si les G-quadruplexes ne sont pas déstructurés par l'action d'hélicases, les ligands pourraient reconnaitre un plus grand nombre de G-quadruplexes et une fluorescence plus importante serait alors observée¹⁷⁸. A contrario, si la déstructuration des G-quadruplexes est augmentée par une expression plus importante d'hélicases, une diminution de la fluorescence due au ligand devrait être observée¹⁸².

6. Conclusion

Une nouvelle méthodologie nommée G4-GIS a été développée dans le but d'améliorer la résolution d'imagerie de fluorescence nécessaire à la localisation des ligands des G-quadruplexes en cellules. Cette méthode repose sur la détection et l'amplification de fluorescence par des anticorps reconnaissant sélectivement ces ligands. Pour le développement de cette méthodologie, deux catégories de ligands ont été synthétisées: les ligands PDC fonctionnalisés par l'haptène 5-BrdU nécessaire à l'immunodétection et les ligands précurseurs non fonctionnalisés par la 5-BrdU. Dans le premier cas, l'immunodétection est réalisée directement après un certain temps d'incubation et, dans le deuxième cas, une réaction bioorthogonale entre les précurseurs et la 5-BrdU est réalisée avant l'immunodétection. La fonctionnalisation *in situ* permet de limiter l'impact négatif de la fonctionnalisation sur l'affinité du ligand pour les structures G-quadruplexes.

La molécule PDC est un ligand de référence des structures G-quadruplexes. Elle possède un très fort K_a pour de nombreuses structures G-quadruplexes, induit une stabilisation contre la dénaturation thermique élevée pour la plupart des G-quadruplexes et possède une bonne sélectivité vis-à-vis de l'ADN double brin. Elle a premièrement été décrite pour sa capacité à induire l'apoptose des cellules cancéreuses en interagissant avec les structures G-quadruplexes des télomères¹²⁷. C'est une molécule possédant une solubilité relativement bonne comparée à d'autres ligands bis-cationiques et la fonctionnalisation au niveau du cycle pyridine permet de limiter la perte d'affinité pour les structures G-quadruplexes. La méthodologie G4-GIS a donc été logiquement appliquée à des dérivés de ce ligand.

Quatre dérivés du ligand PDC fonctionnalisés par des bras espaceurs de nature et de polarité différentes ont été synthétisés avec succès. Ces dérivés sont fonctionnalisés à l'extrémité des bras espaceurs par une fonction alcyne ou une fonction azoture dans le but de pouvoir les fonctionnaliser par la suite à l'aide d'une réaction de CuAAC. Deux partenaires de chimie click dérivés de l'haptène 5-BrdU fonctionnalisés par une fonction alcyne ou une fonction azoture ont ensuite été synthétisés avec succès. Ces deux précurseurs ont ensuite été liés par une réaction optimisée de Huisgen catalysée au cuivre. Les ligands avant (PDC-Alk/N3) et après (PDC-BrdU) fonctionnalisation par l'haptène ont été étudiés pour leur capacité à interagir avec l'ADN et l'ARN.

L'interaction des ligands avec l'ADN et l'ARN G-quadruplex a été étudiée par deux méthodes biophysiques. La première, le FRET-melting, permet d'étudier la stabilisation induite par les ligands contre la dénaturation thermique. La deuxième, le FID, permet d'étudier l'interaction entre une structure ADN ou ARN par déplacement d'une sonde de type « on-off » non sélective pour les structures G-quadruplexes. Ces études ont permis de montrer que les ligands avant la fonctionnalisation par la 5-BrdU (**PDC-Alk/N3**) interagissent avec les structures G-quadruplexes aussi efficacement que le ligand de référence **PDC** et aucune perte de sélectivité n'est détectée pour ces dérivés. Les ligands fonctionnalisés par l'haptène (**PDC-BrdU**) montrent, quant à eux, une interaction légèrement plus faible que le ligand de référence PDC. La sélectivité pour les structures G-quadruplexes n'est cependant par altérée. Ces observations démontrent notamment la nécessité de réaliser la fonctionnalisation directement en cellules. Un des composés **PDC-BrdU** a ensuite été testé par une expérience ELISA modifiée. Cette expérience a permis de démontrer la capacité d'un anticorps anti-BrdU à reconnaitre spécifiquement la base 5-BrdU lorsque le composé **PDC-BrdU** interagit avec les structures G-quadruplexes. Cette étude a notamment permis de montrer un K_D de l'ordre du nM, démontrant la très forte affinité de cette série de ligands.

Dans le but de pouvoir fonctionnaliser *in cellulo* les dérivés **PDC-Alk/N3**, la réaction bioorthogonale de Huisgen catalysée au cuivre a été étudiée dans différentes conditions. Notamment, l'étude de la cinétique de réaction en milieu aqueux a permis de montrer un taux de conversion de 60% après une heure de réaction et une désactivation du cuivre(I) après environ 2 h de réaction. Cette réaction a ensuite été étudiée en présence de deux structures G4-ADN et une structure G4-ARN. Après une heure de réaction, une augmentation des taux de conversion est globalement observée pour les différents ligands avec les structures G-quadruplexes, démontrant la capacité des structures G-quadruplexes à rapprocher spatialement les différents réactifs.

Finalement, la détection des dérivés PDC-BrdU en cellules par la stratégie G4-GIS a été réalisée. L'incubation directe des composés PDC-BrdU fonctionnalisés ex-situ par la 5-BrdU suivi de l'immunodétection par l'incubation d'anticorps primaires et secondaires a permis de dévoiler avec précision la répartition des différents composés au sein des cellules. Après 16 h d'incubation, l'immunodétection a permis de montrer une répartition majoritairement cytoplasmique des composés avec la formation de quelques foci au sein des noyaux. Les composés précurseurs de chimie click PDC-Alk/N3 ont ensuite été incubés pendant 16 h avant d'être fonctionnalisés par la 5-BrdU directement in cellulo à l'aide de la réaction de Huisgen catalysée au cuivre. La détection par immunofluorescence a permis de montrer la formation de nombreux foci intenses dans la région périnucléaire pour le composé PDC-4,3-Alk et une localisation diffuse dans le cytoplasme pour les autres composés. L'immunodétection des ligands PDC-Alk/N3 après fonctionnalisation in situ permet d'obtenir des signaux de fluorescence plus intenses que l'immunodétection directe des ligands PDC-BrdU fonctionnalisés ex-situ. Dans le but de comparer la détection par la méthode G4-GIS et la détection de ligands couplés à des fluorophores, l'imagerie de fluorescence a été réalisée après fonctionnalisation in situ du ligand PDC-4,3-Alk par le fluorophore Cy5. La fonctionnalisation par ce fluorophore montre un signal de fluorescence relativement diffus qui souligne donc l'avantage de l'amplification de fluorescence pour améliorer la résolution des images. Un traitement par la RNase A a de plus permis de montrer que la cible majoritaire des dérivés du PDC est l'ARN cytoplasmique et probablement l'ARN G-quadruplex tout en dévoilant la formation de foci au sein du noyau.

Cette nouvelle méthodologie permet ainsi d'enrichir le nombre d'outils disponibles pour l'étude des structures G-quadruplexes *in cellulo*. Cette méthodologie permet d'améliorer sensiblement la résolution d'imagerie de fluorescence pour les ligands des G-quadruplexes. Elle pourra notamment être appliquée à d'autres ligands que les dérivés du PDC après une simple fonctionnalisation d'un ligand de référence par une fonction azoture ou alcyne.

Chapitre III : Photo-alkylation de

structures G-quadruplexes

1. Introduction

Comme décrit dans l'introduction, plusieurs méthodes visant à localiser les séquences du génome pouvant former des structures G-quadruplexes ont été développées. Ces méthodes reposent soit sur des études bio-informatiques soit sur des études expérimentales *in vitro* ou *in cellulo* basées sur la reconnaissance des G-quadruplexes par un anticorps (BG4⁹²) ou d'une protéine (G4P⁹⁴). Malheureusement, ces méthodes requièrent soit la fixation des cellules avant l'incubation de l'anticorps, soit l'utilisation d'une protéine difficile d'accès. Dans le but de pallier ces problèmes et d'enrichir le nombre d'outils moléculaires pour l'étude de la localisation des séquences formant des G-quadruplexes *in cellulo*, plusieurs méthodes employant des petites molécules ont été développées.

a. Capture de structures G-quadruplexes par des ligands spécifiques

L'identification des séquences adoptant des structures G-quadruplexes au sein du génome par l'utilisation de ligands repose sur la fonctionnalisation de ces derniers par un groupement biotine. Une fois la petite molécule en présence de G-quadruplexes, le motif biotine permet l'extraction des complexes ADN ou ARN / ligand du milieu biologique par ajout de billes magnétiques recouvertes de streptavidine. Une fois les ligands extraits du milieu, les brins d'ADN ainsi capturés à l'aide de ces ligands sont analysés. Cette méthode dite de pull-down est couramment utilisée en biologie : historiquement, la première expérience de capture d'ADN-G4 par un ligand a été publiée par Muller *et al.* en 2010²³². Dans cette étude, un dérivé de la PDS fonctionnalisé par un groupement biotine a été développé (PDB-Biot, Figure 95A). Une première expérience de pull-down *in vitro* a permis de montrer que ce ligand était capable de sélectivement capturer les structures G-quadruplexes au sein d'un mélange de différentes séquences d'ADN (Figure 95B). L'extraction supportée d'ADN Gquadruplex par le dérivé PDB-Biot a ensuite été réalisée sur des fragments d'ADN génomique de cellules HT1080. Le pull-down suivi d'analyses qPCR a notamment permis de montrer que le ligand PBP-Biot était capable d'extraire efficacement l'ADN télomérique de ce mélange d'ADN.



Figure 95: Expérience d'extraction supportée avec A) la structure du ligand PDB-Biot dérivé de la PDS et B) analyse par gel d'électrophorèse après l'extraction supportée d'un mélange de différentes séquences d'ADN. Image tirée et adaptée de Muller *et al.*²³²

En 2019, le ligand PDB-PEG₂₄-biotin (Figure 96A) dérivé de la PDS fonctionnalisée par un long linker PEG et une biotine a été développé et utilisé pour l'analyse des G-quadruplexes au sein du génome humain²³³. Pour cela, l'ADN génomique de cellules HEK 293T a été extrait, fragmenté et restructuré en présence d'une forte concentration d'ions K⁺. En parallèle, des billes recouvertes de streptavidine ont été préincubées avec le ligand PDB-PEG₂₄-biotin. Les fragments d'ADN génomique ont ensuite été ajoutés aux ligands liés aux billes magnétiques et les brins d'ADN G-quadruplexes capturés ont finalement été analysés par NGS (Figure 96B). Au total, environ 51 000 PQS ont été identifiées à l'aide de cette méthode d'extraction supportée avec un enrichissement de séquences pouvant former des G-quadruplexes au niveau des régions intergéniques et au niveau des introns. Cette observation est notamment en accord avec les différentes méthodes développées pour l'étude de la localisation des structures G-quadruplexes. Cette étude a permis de montrer que les séquences (ATGG)_n étaient distribuées à 44,7% dans les régions intergéniques et à 55,3% au niveau des introns, suggérant de possibles rôles pour les G-quadruplexes à deux quartets sur l'épissage des séquences pré-ARNm.



Figure 96: A) Structure du composé PDB-PEG₂₄-biotin. B) Schéma général du protocole d'extraction supportée de fragments d'ADN génomiques suivi du séquençage à haut débit. Image tirée et adaptée de Zeng *et al.*²³³

Récemment, un dérivé du ligand DOTASQ²³⁴ fonctionnalisé par un groupement biotine (BioTASQ, Figure 97A) a été développé. Le ligand DOTASQ est une molécule biomimétique composée de quatre guanines pouvant s'autoassocier et interagir avec les quartets externes des structures Gquadruplexes par π -stacking. Le dérivé biotinylé BioTASQ a ainsi été utilisé pour réaliser l'extraction et l'analyse d'ARN G-quadruplexes in vitro et in cellulo^{170,235} par G4RP-seq (pour G4-RNA precipitation-sequencing). Une première expérience in vitro de capture de séquences Gquadruplexes et duplexes fonctionnalisées par une FAM a permis de montrer une bonne sélectivité envers les séquences G-quadruplexes et plus particulièrement pour les séquences d'ARN G4. Le ligand BioTASQ a ensuite été utilisé sur des extraits de cellules humaines MCF7 (Figure 97B). Pour cela, les cellules ont premièrement été traitées par du formaldéhyde dans le but de stabiliser et bloquer la formation de structures secondaires. Les lysats cellulaires ont ensuite été obtenus avant d'être incubés avec le ligand BioTASQ. Les séquences liées aux ligands ont par la suite été capturées par ajout des billes magnétiques recouvertes de streptavidine et les fragments d'ARN récupérés ont été analysés par RT-PCR. L'analyse spécifique de deux séquences ARN G4 : VEGF⁵⁴ et NRAS¹²² a permis de montrer un enrichissement lors de la capture par le ligand BioTASQ. De plus, le traitement préalable des cellules par les ligands BRACO-19²³⁶ et RHPS4²³⁷ a permis d'augmenter l'enrichissement de ces deux séquences. Par exemple, le traitement par ces deux ligands a permis de multiplier par 6,5 et 10,2 la quantité d'ARN G4 attribuée à la séquence VEGF, démontrant la capacité de ces ligands à stabiliser ces structures G-quadruplexes in cellulo. Finalement, cette méthodologie a été réalisée sur une lignée de cellules cancéreuses humaines dans le but de séquencer la totalité des brins d'ARN extraits (G4RP-seq). Cette étude a permis de montrer la formation de nombreuses structures Gquadruplexes à travers le transcriptome. Le traitement des cellules par les ligands BRACO-19 et RHPS4 a de plus permis d'accroître le nombre de séquences d'ARN G4 extraites par le ligand BioTASQ.



Figure 97 : Protocole d'extraction supportée avec A) la structure du ligand BioTASQ utilisé et B) le schéma de l'expérience de capture d'ARN G4 d'extraits cellulaires.

Comme présenté dans cette introduction les méthodologies de capture d'ADN ou d'ARN Gquadruplex par des ligands des G-quadruplexes reposent uniquement sur des interactions non covalentes de type π -stacking, électrostatiques, etc. Si ces interactions peuvent être très fortes pour certaines structures G-quadruplexes, la vaste diversité de ces structures *in cellulo* impactera indubitablement le nombre de séquences pouvant être extraites. De plus, les interactions non covalentes entre une structure G-quadruplex et un ligand ne peuvent empêcher la relocalisation de ces derniers lors des différentes étapes de lavage. Dans le but d'améliorer l'extraction d'une cible spécifique d'un milieu biologique, une méthode permettant la formation d'un lien covalent entre une petite molécule et sa cible a été développée.

b. Alkylation de structures G-quadruplexes

i. Méthodologie PAL

La méthodologie PAL (pour Photoaffinity Labelling) est une technique robuste développée initialement dans le but d'élucider les cibles principales de petites molécules²³⁸. Cette méthode

s'appuie sur la fonctionnalisation d'une petite molécule d'intérêt par une fonction photoactivable ainsi que par un motif chimique permettant généralement l'extraction supportée²³⁹.

Les fonctions photoactivables ainsi ajoutées aux molécules d'intérêt permettent, après irradiation UV-Vis, de pouvoir générer des liens covalents avec leur cible. L'irradiation permet de générer des espèces très réactives pouvant réagir avec les molécules à proximité. Idéalement, la fonction photoactivable doit avoir le moins d'impact sur l'affinité de la molécule pour sa cible, être stable sans irradiation et posséder une longueur d'onde d'activation supérieure à 300 nm²⁴⁰. Les réactions générées par ces espèces très réactives doivent être rapides et non-réversibles pour minimiser le marquage non spécifique. Les fonctions photoactivables les plus utilisées sont les dérivés de tétrafluorophenyl azotures, les benzophénones ainsi que les trifluoromethylphenyl diazirines (Figure 98)^{241,240}. La fonctionnalisation par quatre atomes de fluor du cycle phényle des composés arylazotures permet notamment d'éviter des réactions parasites de cyclisation et d'expansion de cycle²⁴². Les différentes fonctions présentées ici sont généralement stables, disponibles commercialement et très réactives, mais sont relativement encombrées stériquement. Dans le but de ne pas trop impacter l'affinité de la petite molécule pour sa cible, des dérivés de diazirines aliphatiques ont été développés²⁴³ (Figure 98).



Figure 98: Structures principales des fonctions photoactivables utilisées pour la méthodologie PAL.

L'irradiation UV de chacune de ces fonctions permet la formation d'espèces très réactives nécessaires au marquage d'une cible biologique. Par exemple, l'irradiation à 300 nm des dérivés arylazotures permet de générer des nitrènes, l'irradiation aux alentours de 350 nm des dérivés de la benzophénone permet de générer des espèces diradicalaires et l'irradiation à environ 360 nm des dérivés diazirines permet de générer des carbènes²⁴⁴. Toutes ces espèces peuvent ensuite réagir par de nombreuses réactions différentes avec les fonctions disponibles à proximité^{240,245,246} (Figure 99).



Figure 99: Exemples de réactions possibles après irradiation de la fonction A) diazirine, B) azoture et C) benzophénone. Cette méthodologie a ainsi été appliquée à de nombreuses petites molécules dans le but d'identifier leur cible *in cellulo*^{247,248,249}.

ii. Méthodologie PAL appliquée aux structures G-quadruplexes

Très récemment en 2021, deux études indépendantes l'une de l'autre ont reporté l'utilisation de dérivés de la PDS fonctionnalisée par une fonction diazirine aliphatique ainsi que d'une fonction alcyne dans le but d'identifier les protéines interagissant avec les G-quadruplexes *in cellulo*. La première étude publiée par l'équipe de Zhou²⁵⁰ utilise un dérivé du ligand PDB (PDB-DA-A, Figure 100) et la deuxième étude publiée par l'équipe de Balasubramanian²⁵¹ utilise deux dérivés de la PDS (Photo-PDS, Figure 100).



Figure 100: Structures des molécules PDB-DA-A²⁵⁰ et Photo-PDS²⁵¹ utilisées pour la méthodologie PAL.

Dans ces deux études, différentes lignées cellulaires sont incubées in vivo avec les dérivés photoactivables avant d'être irradiées à 365 nm. Les molécules ainsi irradiées sont ensuite soumis à une réaction de Huisgen catalysée au cuivre avec une biotine-N3 après avoir récupéré les lysats cellulaires. Les ligands fonctionnalisés par la biotine sont capturés par extraction supportée à l'aide de billes magnétiques recouvertes de streptavidine. Finalement, les protéines liées aux ligands sont isolées et analysées par LC-MS/MS. Plusieurs réplicats biologiques sont réalisés et uniquement les protéines isolées à chaque réplicat sont considérées comme des résultats positifs. Dans le cas du ligand PDB-DA-A, la capture de protéines après photo-crosslinking a permis d'isoler 114 et 281 protéines différentes au sein des cellules SV589 et MM231 respectivement. Parmi ces protéines, 8 nouvelles protéines ont été isolées et caractérisées pour leur capacité à interagir avec différentes séquences G-quadruplexes, élargissant ainsi le nombre de potentielles cibles thérapeutiques. Dans le cas des molécules Photo-PDS, l'incubation de cellules HEK293T et l'extraction a permis d'isoler 248 et 209 protéines différentes à l'aide des composés Photo-PDS-1 et Photo-PDS-2 respectivement. Parmi ces dernières, 201 protéines identiques ont été isolées indépendamment de la taille du linker. Il est de plus intéressant de noter que ces deux différentes études ont identifié 61 protéines communes sur l'ensemble des protéines isolées, démontrant l'importance du ligand ainsi que de la lignée cellulaire étudiée. Ces deux études sont ainsi les premiers exemples de l'application de la méthodologie PAL pour l'extraction et l'identification de protéines associées aux G-quadruplexes.

En plus d'utiliser la méthodologie PAL pour la capture de protéines interagissant avec les Gquadruplexes, il est possible de l'appliquer directement pour l'alkylation d'ADN G-quadruplex. En effet, D. Verga *et al.* ont synthétisé, en 2013, des dérivés du ligand PDC fonctionnalisé par différents bras espaceurs ainsi qu'une fonction photoactivable de type benzophénone ou tétrafluorophenyl azoture. Des expériences d'irradiation à 365 nm en présence de différentes structures Gquadruplexes ont permis de démontrer la capacité de ces ligands à réagir sélectivement avec les Gquadruplexes. Par exemple, les deux ligands possédant un bras espaceur de 4 atomes de carbone fonctionnalisés par une fonction benzophénone et azoture aromatique permettent d'alkyler, tous les deux, 26% de la structure G-quadruplex formée par la séquence télomérique humaine 22AG (Figure 101). L'augmentation de la taille du bras espaceur par 8 atomes de carbone réduit considérablement la quantité d'alkylation de la séquence 22AG (Figure 101). Des études de séquençages des adduits covalents ainsi formés ont permis de déterminer la localisation des liaisons covalentes sur la séquence 22AG. Dans le cas du ligand PDC-C4-Bp, une alkylation sélective des thymines présentes au niveau des boucles du G-quadruplex, certainement par réaction de Paterno-Büchi, est observée. Le composé PDC-C4-N3, quant à lui, réagit exclusivement avec les guanines impliquées dans la formation du G-quartet externe.



Figure 101: Structure des composés photoactivables PDC-PEG-N3 et PDC-PEG-Bp.

Cette étude a ainsi été la première à démontrer la capacité de ligands des G-quadruplexes fonctionnalisés par des groupements photoactivables à générer des adduits covalents sélectivement avec l'ADN G-quadruplex. Utiliser cette capacité pour la capture d'ADN génomique pourrait notamment être réalisé dans le but de déterminer la localisation des séquences pouvant adopter des structures G-quadruplexes.

c. Application de la photo-alkylation pour la capture d'ADN G-quadruplex génomique

Comme décrit précédemment, utiliser des ligands sélectifs des G-quadruplexes pour réaliser la capture d'ADN ou d'ARN G4 présent dans les cellules offre plusieurs avantages : utiliser des petites molécules permet de pouvoir incuber les cellules *in vivo* ; les ligands développés peuvent posséder une excellente affinité et sélectivité pour les G-quadruplexes ; il est possible de fonctionnaliser les ligands par un marqueur après interaction avec la cible à l'aide d'une réaction bioorthogonale dans le but d'éviter la perte d'affinité et/ou de sélectivité du ligand. Ces différents avantages sont cependant largement contrebalancés par un désavantage majeur : la possible perte d'information lors de l'étape

d'extraction supportée. En effet, si un ligand n'est pas lié covalemment à la structure G-quadruplex d'origine, il pourrait en théorie se dissocier et/ou se relocaliser sur d'autres structures.

Dans le but de pallier ce problème, la méthodologie de séquençage d'ADN après extraction supportée de petites molécules est couplée à la méthodologie PAL. Pour cela, des dérivés du ligand PDC fonctionnalisés par un groupement photoactivable ainsi que par un alcyne vrai ont été imaginés et synthétisés. Deux types de ligands ont été développés : des ligands linéaires et des ligands non linéaires (Figure 102). D'un point de vue structurel, les ligands linéaires sont moins encombrés stériquement mais offrent peu de possibilités de fonctionnalisation comparé aux dérivés non linéaires.



Figure 102: Structures des différents ligands linéaires et non linéaires synthétisés.

Une fois les différents dérivés synthétisés, leur affinité et leur sélectivité pour les G-quadruplexes ont été déterminées par des expériences de FRET-melting et de FID et l'effet de ces ligands sur la topologie des G-quadruplexes a ensuite été évalué par des analyses de dichroïsme circulaire. Finalement, la capacité de ces ligands à générer des adduits covalents avec différentes structures Gquadruplexes après irradiation a été évaluée par gel d'électrophorèse.

Ces molécules permettront, par la suite, de pouvoir étudier efficacement la formation et l'accessibilité de structures G-quadruplexes au niveau du génome. Les cellules seront incubées directement avec les ligands possédant les meilleures propriétés (affinité, sélectivité, réactivité) avant d'être irradiées dans le but de générer sélectivement un lien covalent entre l'ADN G-quadruplex et les ligands. Une fois cette photo-activation réalisée, les ligands seront soumis à une réaction de CuAAC avec l'hapten 5-BrdU-N3 utilisé précédemment. La capacité de cet haptène à être reconnu par l'anticorps primaire lorsqu'il est lié à un dérivé du PDC a déjà été évaluée lors du projet

précédent et sera de nouveau utilisée. Une fois la photo-alkylation réalisée, les brins d'ADN liés aux ligands seront capturés par ajout de billes magnétiques recouvertes d'anticorps anti-BrdU. Les brins d'ADN ainsi récupérés pourront finalement être isolés et analysés par séquençage à haut débit (Figure 103A). Cette méthodologie permettra de vérifier la localisation précise du lien covalent entre le ligand et l'ADN. Des analyses bio-informatiques seront menées en parallèle dans le but de vérifier la formation de structures G-quadruplexes au niveau des sites de liaison des ligands. Dans le but de vérifier que la génération d'un lien covalent entre le ligand et l'ADN G-quadruplex offre un réel avantage, ce protocole sera aussi réalisé avec les dérivés non covalents PDC-BrdU précédemment synthétisés (Figure 103B).



Figure 103: Schéma représentatif du protocole de séquençage après extraction supportée de l'ADN. A) Protocole de séquençage pour les ligands fonctionnalisés par un groupement photoactivable. B) Protocole de séquençage pour les ligands non fonctionnalisés par un groupement photoactivable.

2. Synthèse des ligands G-quadruplexes photoactivables

a. Synthèse des composés linéaires

Les composés PDC-N2 linéaires sont synthétisés à partir du linker **4-N2-octynoic (26)**²⁵², préparé en 4 étapes à partir du 4-pentyn-1-ol commercial (Schéma 8).



Schéma 8: Schéma réactionnel pour la synthèse du linker 4-N₂-octynoic acid et structure du catalyseur thiazolium Cat1 et de l'acide hydroxylamine O-sulfonique (HOSA).

La synthèse du linker est débutée par oxydation du produit commercial 4-pentyn-1-ol par réaction du chlorochromate de pyridinium (PCC) en présence d'acétate de sodium pendant 1 h. Le brut réactionnel est ensuite simplement dilué avec de l'éther diéthylique et filtré sur colonne chromatographique de silice. L'aldéhyde **23** étant très volatil, le mélange est concentré précautionneusement et est directement engagé dans la deuxième réaction. Le composé **23** en solution dans le dioxane est ajouté goutte à goutte, sur 4 h, à un mélange d'acrylate d'éthyle et de sel de thiazolium **Cat1** dans le dioxane à 80 °C. Cette réaction est une réaction de Stetter²⁵³ : le sel de thiazolium va permettre de modifier la nature électrophile du carbone de l'aldéhyde pour en faire un carbone nucléophile. Cette réaction fait donc partie de la chimie de type Umpolung, initialement décrite par D. Seebach²⁵⁴. Le carbone nucléophile ainsi généré par le sel de thiazolium peut réaliser une addition de Michael sur l'acrylate d'éthyle pour générer le composé **24**. L'addition très lente de l'aldéhyde, la température relativement haute et le temps de réaction très long de 54 h permet de

limiter au maximum la formation du produit de condensation 1,2 de l'aldéhyde sur lui-même et de favoriser l'addition 1,4 sur l'acrylate d'éthyle (Schéma 9).



Schéma 9: Mécanisme de la réaction de Stetter (bleu) et de la condensation benzoïque 1,2 (rouge).

Le produit **24** est ainsi obtenu avec 19% de rendement sur deux étapes après purification par chromatographie flash. Dans le but d'améliorer le rendement, deux autres bases régulièrement utilisées pour ce type de réaction ont été testées : la DIPEA et la DBU, mais aucune de ces bases n'ont permis d'améliorer le rendement. Ce faible rendement peut notamment être expliqué par le manque de réactivité de la fonction ester 1,4-insaturée. En effet, la réaction de Stetter implique une addition de Michael et est donc directement impactée par l'effet électroattracteur relativement faible de la fonction ester conjuguée à la double liaison, notamment comparé à des fonctions aldéhydes ou cétones. La deuxième explication de ce faible rendement vient de la grande volatilité de l'intermédiaire aldéhyde **23** qui le rend très difficile à manipuler. La température élevée (80 °C) nécessaire à la réaction de Stetter peut de plus faire volatiliser le produit lors de l'ajout goutte à goutte. Dans le but de réduire la volatilité de l'aldéhyde **23** et de limiter de potentielles réactions parasites avec la fonction alcyne, une réaction de protection à l'aide de chlorure de triméthylsilyle en

présence de *n*-BuLi a été réalisée²⁵⁵. La fonction hydroxyle a ensuite été oxydée en aldéhyde par réaction de PCC comme précédemment décrit et l'aldéhyde a été utilisé pour la réaction de Stetter. Hélas, aucun produit correspondant au produit attendu n'a pu être isolé (Schéma 10).



Schéma 10: Schéma réactionnel envisagé pour la synthèse de la molécule 2'.

Face à cet échec et aux difficultés rencontrées lors de cette synthèse, d'autres essais d'optimisation n'ont pas été réalisés. Ainsi, l'ester **24** est ensuite saponifié par réaction de LiOH pour donner l'acide **25** avec un rendement quasi quantitatif (Schéma 8). Dans le but d'obtenir le composé diazirine **4-N₂-octynoic acid (26**), plusieurs conditions de réaction ont été testées (Tableau 3).

Source d'ammoniac	Oxydants	Rendement max
NH₃.MeOH 7M	DIPEA/I2	28%
NH _{3(I)}	KOH/O ₂	-
	DIPEA/I ₂	58%

Tableau 3: Conditions de réaction testées pour l'obtention du linker diazirine.

Les conditions de réaction publiées pour obtenir ce linker implique l'utilisation d'ammoniac en solution dans le MeOH à une concentration de 7 M ainsi qu'une oxydation de l'intermédiaire diaziridine par le couple DIPEA/I₂²⁵². Dans ces conditions, le composé attendu était obtenu avec 28% de rendement au maximum. Dans le but de maximiser la formation de l'intermédiaire imine nécessaire à la réaction, de l'ammoniac liquide a été utilisé et une oxydation par ajout de DIPEA/I₂ a permis d'obtenir le composé attendu avec 58% de rendement. L'oxydation par une base en présence d'air comme précédemment publié²⁵⁶ n'a pas permis d'obtenir le composé **4-N₂-octynoic acid (26)**. Les meilleures conditions de réaction impliquent donc la formation préalable d'ammoniac liquide. Pour ce faire, l'ammoniac est produit par réaction de soude et de chlorure d'ammonium en présence d'eau. Cette réaction est très exothermique et produit l'ammoniac sous forme gazeux. Les vapeurs sont passées à travers une colonne d'oxyde de calcium pour capter les traces d'eau potentiellement produites et sont recondensées dans un tube scellé refroidi à -78 °C. L'acide **25** est ensuite dissous dans l'ammoniac liquide et le mélange est agité 5 h à température ambiante. Le mélange est ensuite

refroidi à -70 °C et l'acide hydroxylamine O-sulfonique (HOSA) est ajouté lentement. Le tube est scellé et le mélange réactionnel est agité 1h30 à température ambiante. Finalement, le mélange est refroidi une dernière fois à -70 °C, le tube scellé est ouvert et le mélange est laissé revenir à température ambiante très lentement pendant la nuit dans le but d'évaporer l'ammoniac. Une fois l'ammoniac évaporé, le mélange est filtré, dissous dans du méthanol et la DIPEA est ajoutée. Enfin, du diiode est additionné jusqu'à obtenir une coloration marron persistante. Le mélange réactionnel est ensuite traité et purifié par colonne chromatographique sur silice pour donner le produit **4-N₂-octynoic acid (26)** avec 58% de rendement. Le mécanisme de formation de la diazirine est présenté ci-après (Schéma 11).



Schéma 11: Mécanisme de la réaction de formation de la diazirine.

Dans le but de synthétiser différents ligands PDC fonctionnalisés par cette diazirine, trois bras espaceurs ont été introduits sur le cycle pyridine (Schéma 12).



Schéma 12: Synthèse des trois dérivés du PDC fonctionnalisés par un linker court (PDC-2NH₂), un linker moyen (PDC-4NH₂) et un linker long (PDC-PEG₂-NH₂).

Ainsi, trois dérivés fonctionnalisés par trois linkers de taille et de polarité différentes ont été synthétisés à partir du **PDC-CI**. Ces composés sont obtenus par réaction de S_NAr sur le **PDC-CI** par la 1,2-diaminoéthane, la 1,4-diaminobutane et la 1,8-diamino-3,6-dioxaoctane. Une simple précipitation dans l'eau ainsi que plusieurs lavages suffisent à obtenir les trois composés **PDC-2NH₂** (27), **PDC-4NH₂** (2) et **PDC-PEG2-NH₂** (28) avec 86, 88 et 78% de rendement respectivement. Ces

différentes molécules sont ensuite couplées avec la diazirine précédemment synthétisée (Schéma 13).



Schéma 13 : Synthèse des dérivés PDC-2-N2-Alk, PDC-4N2-Alk et PDC-PEG₂-Alk.

Ainsi, la diazirine **4-N₂-octynoic acid** (**26**) est couplée aux différents dérivés **PDC-NH2** précédemment synthétisés par une réaction de couplage peptidique en présence d'HATU et de DIPEA. Une purification par chromatographie flash permet d'obtenir les composés **29**, **30 et 31** avec 76, 90 et 100% de rendement respectivement. Finalement, la méthylation des quinolines par réaction d'iodométhane permet d'obtenir les composés finaux **PDC-2-N2-Alk**, **PDC-4N2-Alk** et **PDC-PEG2-Alk**. Ces molécules seront testées pour leur affinité et leur sélectivité pour les structures G-quadruplexes ainsi que pour leur capacité à réaliser des liaisons covalentes avec ces dernières.

b. Synthèse des composés non linéaires⁴

Trois types de ligands non linéaires ont été synthétisés avec succès. Ces trois groupes possèdent une fonction photo-activable aromatique (diazirine, azoture, benzophénone) ainsi qu'une fonction alcyne à l'extrémité d'un bras espaceur plus ou moins long. La synthèse de ces différents composés commence par la formation des acides carboxyliques fonctionnalisés par un alcyne vrai et une

⁴ Synthèse réalisée par Corinne Landras Guetta et Matthieu Seze

fonction photoactivable. La synthèse des composés possédant un bras espaceur de type propyne et pentyne est présentée ci-après (Schéma 14).



Schéma 14: Synthèse des acides carboxyliques fonctionnalisés par une fonction photoactivable ainsi qu'une fonction alcyne de type propyne et pentyne.

La synthèse est débutée par la réaction de S_N2 par la propargyl amine ou la 4-pentyn amine sur le methyl 4-bromobutanoate pour donner les composés **35** et **36** avec 54 et 61% de rendement respectivement. Les conditions expérimentales ont été adaptées de Pícha et al.²⁵⁷ dans le but de limiter la réaction de lactamisation intramoléculaire entre l'amine secondaire et la fonction ester. Les groupements photoactivables aromatiques sont ensuite introduits après formation des chlorures d'acyles pour donner les différents ester C1 (**37-39**) et C3 (**40-42**) avec des rendements de faibles à quantitatifs. Les acides carboxyliques correspondants (**43-48**) sont ensuite simplement obtenus par réaction de saponification avec de bons rendements.

Les trois dérivés photoactivables fonctionnalisés par un bras espaceur de type TEG sont synthétisés suivant une voie de synthèse différente de celle présentée pour les composés C1 et C3 (Schéma 15).



Schéma 15: Synthèse des acides carboxyliques fonctionnalisés par une fonction photoactivable ainsi qu'une fonction alcyne éloignée de l'amine tertiaire par un linker TEG.

Dans le but de synthétiser les acides carboxyliques fonctionnalisés par une fonction photoactivable ainsi qu'une fonction alcyne espacée par un linker TEG, le composé **52** est premièrement préparé par synthèse convergente de deux intermédiaires de synthèse. Le composé **50** est un triéthylène glycol fonctionnalisé aux extrémités par un atome de brome et un alcyne vrai. Il est obtenu en deux étapes à partir du triéthylène glycol²⁵⁸. La première étape est une réaction de S_N2 sur le bromure de propargyl après déprotonation de l'hydroxyle primaire par du tBuOK. La fonctionnalisation de l'alcool primaire par un atome de brome est ensuite réalisée par une réaction d'Appel²⁵⁹ en présence de triphénylphosphine et de tétrabromure de carbone pour donner le composé **50** avec 88% de rendement. En parallèle, le composé **51** est produit par estérification de l'acide 4-aminobutanoïque par réaction de chlorure de thionyle en présence d'éthanol. Ces deux intermédiaires de synthèse sont ensuite mis à réagir en présence de base. L'addition lente du dérivé bromé sur un excès d'amine permet notamment d'éviter la formation de l'amine tertiaire. Le composé **52** est ainsi obtenu avec 66% de rendement après purification. Les trois fonctions photoactivables diazirine, azoture et benzophénones sont ensuite couplés avec ce dernier par réaction de substitution nucléophile d'acyle après formation des chlorures d'acyles correspondants. Les différents esters ainsi produits sont saponifiés en présence de soude pour donner les composés **56**, **57**, et **58** avec 97, 95 et 72% de rendement respectivement.

Au total, neuf acides carboxyliques fonctionnalisés par un groupement photoactivable et une fonction alcyne ont été synthétisés. La fonction alcyne est éloignée du cœur PDC et de la fonction photoactivable par trois linkers différents. Ces modulations pourront avoir un effet sur les rendements de photo-alkylation et de la réaction du Huisgen catalysée au cuivre avec le dérivé de la 5-BrdU. Ces acides carboxyliques sont finalement couplés au ligand de référence PDC (Schéma 16).



Schéma 16: Synthèse des dérivés photoactivables non linéaires du PDC à partir du composé PDC-4-NH2.

La synthèse des dérivés non linéaires du PDC est ainsi continuée par le couplage peptidique entre les différents acides carboxyliques préalablement synthétisés avec le composé **PDC-4-NH2** (2). Cette réaction est réalisée soit en présence d'HATU soit en présence d'EDCI.HOBt pour donner les dérivés **59-67** avec des rendements de moyens à bons (de 42 à 85%). La méthylation des quinolines est finalement réalisée par réaction d'iodométhane pendant 16 h à 40 °C. De simples lavages à l'eau et à l'éthanol permettent d'obtenir huit dérivés trifonctionnels non linéaires avec de bons rendements. Malheureusement, le dérivé **PDC-C3-N3 (72)** n'a pas pu être obtenu, car une dégradation de la fonction azoture (N3) en amine (NH2) était observée par LC/MS. Des essais de méthylation des quinolines dans d'autres conditions sont actuellement menés dans le but d'obtenir ce composé sans induire la dégradation de la fonction photoactivable.

Au total, onze composés photoactivables dérivés du ligand PDC ont été synthétisés. Parmi ces derniers, trois possèdent une fonction diazirine aliphatique éloignée du cœur PDC par trois linkers

différents. Les dérivés non linéaires sont fonctionnalisés par trois fonctions photoactivables aromatiques différentes : une fonction diazirine, azoture ou benzophénone. La fonction alycne est, quant à elle, éloignée du reste de la molécule par une linker court (propyne), un linker moyen (pentyne) ou un linker long (TEG). En théorie, ce linker ne devrait avoir qu'un impact modéré sur l'affinité des ligands pour les structures G-quadruplexes ainsi que sur leur capacité à générer des adduits covalents avec ces derniers. Cependant, la présence de la fonction alcyne pourrait générer des réactions secondaires après l'activation des groupements photoactivables par irradiation. Moduler la proximité spatiale de la fonction alcyne permettra de pouvoir choisir uniquement les ligands capables de réaliser efficacement la réaction de Huisgen catalysée au cuivre après photoirradiation. Les dérivés ainsi synthétisés ont été testés pour leur capacité à interagir avec les structures G-quadruplexes par différentes études biophysiques. Ces expériences sont actuellement en cours de réalisation, mais des résultats préliminaires sont présentés ici.

3. Etudes biophysiques de l'interaction des ligands avec les Gquadruplexes

a. FRET-melting⁵

Dans le but de caractériser la capacité des ligands synthétisés à stabiliser les structures Gquadruplexes contre la dénaturation thermique, des expériences de FRET-melting ont été réalisées. Les résultats de cette étude sont présentés ici pour trois séquences : la séquence télomérique humaine 22AG (F21T), la séquence FcMyc22T dérivée de la séquence Pu27 naturellement présente au sein du promoteur du gène c-Myc et la séquence FCEB25wtT dérivée du minisatellite CEB25. La sélectivité des composés est vérifiée en ajoutant une séquence de compétiteur duplex ds26 à une concentration de 3 et 10 µM comme précédemment décrit (Figure 104).

⁵ Étude réalisée avec l'aide de Charles-Henri Buffet


Figure 104: Résultats des expériences de FRET-melting des composés photoactivables dérivés du PDC sur les séquences F21T, FCEB25wtT et FcMyc22T. Les pointillés rouges délimitent les différents ligands en fonction de la nature du linker.

Ces résultats de FRET-melting permettent de montrer de nombreuses différences en fonction des ligands et des séquences étudiées. Il est difficile d'en ressortir une tendance générale, mais plusieurs points peuvent tout de même être remarqués :

Les différents groupes de ligands synthétisés ne stabilisent pas aussi efficacement les structures G-quadruplexes. Notamment, les ligands linéaires PDC-4-N2-Alk, PDC-2-N2-Alk et PDC-PEG-N2-Alk stabilisent généralement avec une meilleure efficacité les différentes structures que les autres composés. Par exemple, les trois composés stabilisent la séquence F21T avec un ΔT_{1/2} d'environ 25 °C tandis que les autres composés stabilisent cette séquence de 10 à 20 °C. Parmi les ligands linéaires, le composé à linker long PDC-PEG-N2-Alk stabilise plus efficacement que les deux autres composés pour les séquences FCEB25wtT et FcMyc22T. Les stabilisations induites par les trois autres catégories (C1, C3 et PEG) ne semblent ni dépendre de la taille du linker espaçant la fonction alcyne ni de la fonction photoactivable. Il est cependant intéressant de noter que les dérivés de la benzophénone stabilisent globalement aussi efficacement que les autres composés. En effet, une affinité

plus faible pouvait être attendue pour les dérivés de la benzophénone au vu de l'encombrement stérique de cette fonction.

- ► Les différents ligands ont une affinité plus ou moins affectée par rapport au ligand de référence **PDC** en fonction de la séquence étudiée. Par exemple, le **PDC** stabilise la séquence F21T avec un $\Delta T_{1/2}$ d'environ 32 °C et le ligand **PDC-C3-N2** stabilise cette même séquence avec un $\Delta T_{1/2}$ d'environ 22 °C, une baisse d'environ 30% est donc observée. Ces ligands stabilisent la séquence FCEB25wtT avec des $\Delta T_{1/2}$ de 16 et 6 °C respectivement, une baisse d'environ 65% de la stabilisation est observée dans ce cas.
- Les températures de stabilisation ne sont pas ou peu affectées par la présence d'ADN duplex. Il est intéressant de noter que tous les composés synthétisés possèdent une excellente sélectivité pour l'ADN G-quadruplex par rapport à l'ADN duplex. En effet, l'ajout de 3 ou 10 μM d'ADN compétiteur duplex (soit 15 équivalents et 50 équivalents) n'induit pas de diminution des températures de stabilisation des différents composés sur ces trois séquences.

Il est globalement difficile d'identifier une tendance réelle à l'aide de ces expériences de FRETmelting. D'autres expériences doivent être réalisées sur d'autres séquences dans le but d'étoffer la quantité de données biophysiques sur la capacité des ligands à stabiliser les structures Gquadruplexes. Pour compléter cette étude, des expériences de FID en cuve ont été réalisées.

b. FID

Comme pour les composés PDC-BrdU, la sonde utilisée ici est le thiazole orange TO, car elle possède de bonnes propriétés de fluorescence ainsi qu'un spectre d'absorbance et de fluorescence qui ne se recouvrent pas avec les spectres d'absorbances des composés photoactivables (Figure 105).



Figure 105 : Spectres d'absoprtion des différents composés photoactivables à une concentration de 5 μ M en tampon K⁺100.

Les résultats du FID sur la séquence télomérique humaine 22AG sont présentés ici, mais d'autres séquences seront testées par la suite (Figure 106).



Figure 106: Résultats obtenus par expérience de FID. A) Courbes de déplacement de la sonde TO. B) Graphique représentant les différents DC₅₀ obtenus Les pointillés rouges délimitent les différents ligands en fonction de la nature du linker. Les barres d'erreurs représentent l'écart-type de deux expériences individuelles.

Cette expérience de FID sur la séquence F21T avec les différents composés synthétisés indique globalement une bonne capacité des ligands à déplacer le TO de la structure G-quadruplex. Les deux composés linéaires **PDC-4-N2-Alk** et **PDC-PEG-N2-Alk** déplacent notamment le TO avec un DC₅₀ d'environ 0,42 μ M tandis que le PDC possède un DC₅₀ de 0,3 μ M. Cette expérience permet notamment de montrer que la fonctionnalisation par un bras espaceur de taille moyenne et longue impacte relativement peu la capacité du PDC à déplacer la sonde. Le composé **PDC-2-N2-Alk** possède quant à lui un DC₅₀ légèrement plus élevé de 0,56 μ M. Trois autres composés possèdent un DC₅₀ similaire : le **PDC-C1-N3**, **PDC-C1-Bp** et le **PDC-TEG-N3**. Des DC₅₀ supérieurs à 0,7 μ M sont observés

pour les ligands restant avec notamment le composé **PDC-C1-N2** qui possède le DC₅₀ le plus élevé d'environ 1,0 μ M.

Globalement, les deux expériences de FRET-melting et de FID sont en accord pour les différents ligands : les ligands qui stabilisent efficacement la séquence F21T montrent une forte capacité à déplacer le TO du G-quadruplex et les ligands qui stabilisent moins efficacement cette séquence montrent des DC₅₀ plus élevés. La fonctionnalisation du **PDC** par des groupements photoactivables aromatiques induit généralement une diminution de l'interaction comparée aux ligands fonctionnalisés avec une chaine aliphatique. D'autres séquences seront testées par ces deux méthodes par la suite.

c. CD

Avant de pouvoir étudier la capacité des différents ligands à générer des adduits covalents avec l'ADN, des expériences de titration CD ont été réalisées. Ces analyses permettront de vérifier si les ligands induisent une modification de la topologie des structures G-quadruplexes et permettront de pouvoir rationaliser les résultats des expériences de photo-alkylation.

Pour ces expériences, les séquences d'ADN G-quadruplexes sont structurées à 3 μ M dans un tampon K⁺100. Les spectres CD sont ensuite enregistrés entre 200 et 500 nm avant l'ajout de ligand puis après l'ajout de 3, 6 et 15 μ M de ligands (soit 1, 2 et 5 équivalents). L'ellipticité θ mesurée pour chaque longueur d'onde est convertie en ellipticité molaire ($\Delta \epsilon$) exprimé en M⁻¹.cm⁻¹. Le $\Delta \epsilon$ représente la différence de coefficient d'extinction molaire entre la composante droite et gauche de la lumière polarisée circulairement et qui permet d'affranchir les résultats de la concentration et de la longueur du chemin optique³⁷. L'étude par CD a été réalisée sur les trois séquences utilisées par FRET-melting : la séquence hybride 22AG et les séquences parallèles cMyc22 et CEB25wt. Les expériences de CD ont été réalisées dans ces conditions pour le composé de référence **PDC**, le dérivé linéaire **PDC-4-N2-Alk** et les dérivés **PDC-C1** et **PDC-C3**.

Les analyses CD réalisées sur la séquences télomérique humaine 22AG sans ajout de ligand permettent de montrer que la séquence forme une structure hybride en solution avec un maximum d'intensité à environ 290 nm et un épaulement à environ 270 nm comme précédemment décrit⁶⁴ (Figure 107). L'ajout de 1 équivalent de ligand induit un changement de la signature CD de la topologie hybride pour une topologie antiparallèle avec un maximum d'intensité à environ 290 nm et

un pic négatif à environ 265 nm. Cette observation a précédemment été décrite²⁶⁰ et est attribuée à la capacité des ligands bisquinolinium **PDC** et **PhenDC3** à induire un changement de conformation de la structure. L'ajout de 5 équivalents de ligand induit, de plus, une baisse de l'intensité du signal CD dans certains cas.



Figure 107: Spectres CD de la séquence 22AG (3 µM) en tampon K+100 sans ligand (noir), ou après addition de 1 eq (rouge), 2 eq (bleu) et 5 eq (magenta) de ligand.

Les spectres CD de la séquence cMyc22 montrent une topologie parallèle avec un maximum d'intensité à environ 260 nm comme préalablement décrit³⁴ (Figure 108). L'ajout de 1 et 2 équivalents de ligand ne modifie pas la topologie de la structure G-quadruplex, mais l'ajout de 5 équivalents induit, dans certains cas, la diminution de l'ellipticité de la séquence cMyc22. En effet, cette diminution est observée pour tous les ligands excepté le ligand **PDC** et le ligand **PDC-C3-Bp** (Figure 110).



Figure 108: Spectres CD de la séquence cMyc22 (3 μM) en tampon K+100 sans ligand (noir), ou après addition de 1 eq (rouge), 2 eq (bleu) et 5 eq (magenta) de ligand.

Les spectres CD de la séquence CEB25wt (Figure 109) permettent de montrer la topologie parallèle de cette séquence avec le maximum d'intensité à environ 260 nm⁶⁵ et l'ajout d'une concentration croissante de ligand ne semble pas impacter la topologie de la séquence. Cependant, l'ajout de ligands induit généralement une diminution de l'intensité de l'ellipticité à 265 nm de la séquence CEB25wt.



Figure 109: Spectres CD de la séquence CEB25wt (3 μM) en tampon K+100 sans ligand (noir), ou après addition de 1 eq (rouge), 2 eq (bleu) et 5 eq (magenta) de ligand.

Il est intéressant de noter que la baisse d'intensité du signal CD dépend directement du ligand pour les deux séquences cMyc22 et CEB25wt. Par exemple, les ligands **PDC-C3-N2** et **PDC-C1-N3** induisent une forte diminution de l'intensité du signal CD à 265 nm tandis que le ligand **PDC** induit une diminution légère de l'intensité du spectre CD (Figure 110). La diminution de l'intensité d'un spectre CD est difficile à rationaliser, mais une possible explication pourrait être la précipitation du complexe ligand/G-quadruplex à 15 μM de ligand.



Figure 110: Évolution de l'intensité du signal CD de la séquence CEB25wt et de la séquence cMyc22 (3 μM) à 265 nm sans ligand et après ajout de 3 μM, 6 μM et 15 μM de ligand.

L'observation d'une diminution de l'intensité du spectre CD de structures G-quadruplexes lors d'une titration de petites molécules a déjà été reportée dans la littérature^{261,262}. Notamment, la titration CD du composé TmPyP4 et d'un dérivé de triarylpyridine avec la séquence ARN M3Q et la séquence ADN c-kit-1 respectivement a montré une forte diminution d'ellipticité à 260 nm de ces structures G-quadruplexes (Figure 111). Supportée par d'autres expériences (RMN, gels d'électrophorèse, UV-vis, etc.), cette observation a ainsi été attribuée à la capacité de ces ligands à déstructurer les G-quadruplexes formés pas ces séquences.



Figure 111: Effet de l'addition d'une concentration croissante de ligand sur l'intensité de spectres CD. A) Titration de TmPyP4 avec la séquence ARN M3Q (4 μM). B) Titration de triarylpyridine (1) avec la séquence ADN c-Kit1 (10 μM). Images tirées de Waller *et al.*²⁶¹ et Morris *et al.*²⁶²

Dans le but de mieux comprendre la diminution d'intensité du spectre CD observée, une titration a été réalisée sur la séquence CEB25wt avec les composés de référence **PDC** et **PhenDC3** et le composé photoactivable **PDC-C1-N3** avec une concentration croissante de 0 à 24 μ M (0 à 8 équivalents, Figure 112).

Cette expérience montre tout d'abord que l'intensité de l'ellipticité de la séquence CEB25wt ne diminue plus à 265 nm après ajout de 1 µM de ligand PDC, indiquant probablement la formation du complexe PDC/G-quadruplex avec une stœchiométrie 1 :1. Ensuite, les titrations avec le ligand **PhenDC3** et **PDC-C1-N3** montrent que l'intensité du spectre CD est modifiée après chaque ajout de ligand. Le ligand **PhenDC3** induit notamment une diminution drastique de l'ellipticité à 265 nm et 290 nm. Cette différence de comportement entre le composé **PDC** et le composé **PhenDC3** et **PDC-C1-N3** est très déconcertante au vu de leur structure moléculaire très proche. Le composé **PhenDC3** étant un ligand très affin des structures G-quadruplexes et qui est capable de les stabiliser

fortement²⁶³, il est peu probable que cet effet observé par analyse CD résulte de la déstructuration des G-quadruplexes. Un tel effet pourrait notamment être attribué à de l'agrégation, de la précipitation ou à l'induction d'un signal CD et des précautions devraient toujours être prises lors de l'interprétation de spectres de dichroïsme circulaire²⁶⁴.



Figure 112: Spectres CD de la séquence CEB25wt (3 µM) sans ligand ou après ajout d'une concentration croissante de PDC, PhenDC3 et PDC-C1-N3.

D'autres expériences ont été réalisées (UV-Vis/CD) dans le but de comprendre la diminution d'ellipticité observée, mais aucune conclusion ne peut être tirée pour le moment.

4. Étude de la réaction de Huisgen catalysée au cuivre

Dans le but de pouvoir appliquer la méthodologie de capture d'ADN G-quadruplex présenté dans l'introduction, il est nécessaire de vérifier que les ligands sont capables d'être couplés à la 5-BrdU à l'aide de la réaction de Huisgen catalysée au cuivre. Pour cela, tous les ligands synthétisés ont été soumis à une réaction de CuAAC avec la **5-BrdU-N3** précédemment synthétisée. Les conditions de réaction ont été optimisées et ont été réalisées de la sorte : le dérivé du **PDC photoactivable** (2,5 µl, 2 mM) est ajouté à la **5-BrdU-N3** (5 µL, 2 mM) en solution dans 108,5 µL d'H₂O. En parallèle, une deuxième solution contenant le THPTA (40 µL, 10 mM) et l'ascorbate de sodium (40 µM, 10 mM) est

préparée. Finalement, le sulfate de cuivre (4 μ L, 10 mM) est ajouté à la deuxième solution et celle-ci est directement additionnée sur le mélange de ligand **PDC photoactivable** et de **5-BrdU-N3**. Le volume final est de 200 μ L et les concentrations sont :

- 25 µM PDC photoactivable
- 50 μM **5-BrdU-N3**
- 200 μM CuSO₄.5H₂O
- 2000 μM THPTA
- 2000 μM NaAsc

Comme pour la réaction de CuAAC présenté dans le projet précédent (Chapitre II.4), le mélange est agité 1 h avant d'être injecté et analysé par HPLC analytique. En parallèle, chaque échantillon est analysé par LC/MS dans le but de vérifier la masse du produit formé. Un exemple d'analyse de la réaction du composé **PDC-4-N2-Alk** avec le composé **5-BrdU-N3** est présenté ci-après (Figure 113).



Figure 113: Réaction de CuAAC entre le composé PDC-4-N2-Alk et le composé 5-BrdU-N3. A) Exemple de chromatogramme HPLC obtenu sans ajout de cuivre ou après 1 h de réaction. B) Analyse LC/MS après 1 h de réaction et C) un zoom sur le spectre de masse du pic attribué à la formation du produit PDC-4-N2-BrdU.

Cette expérience permet tout d'abord de montrer que la réaction du ligand **PDC-4-N2-Alk** avec le dérivé **5-BrdU-N3** est très rapide dans ces conditions avec un taux de conversion de presque 90% en 1 h (Figure 113A). L'analyse LC/MS du milieu réactionnel (Figure 113B) permet, de plus, de pouvoir vérifier la formation du produit attendu par analyse de masse (Figure 113C). Ces analyses ont ensuite été réalisées pour tous les différents ligands synthétisés (Figure 114).



Figure 114: Taux de conversion de la réaction de CuAAC des différents composés photoactivables après 1 h de réaction avec le composé 5-BrdU-N3 déterminés par analyse HPLC. Les barres d'erreurs représentent l'écart-type de deux expériences individuelles.

La réaction de CuAAC réalisée sur les différents ligands permet généralement de former environ 90% de produit attendu après 1 h et plus de 95% dans le cas des composés PDC-C1-N2, PDC-C1-N3 et PDC-C1-Bp. Curieusement, les ligands PDC-C3-N2 et PDC-C3- Bp montrent des taux de conversion légèrement plus faibles avec environ 60 et 70% de conversion respectivement. Cette observation permet notamment de montrer que la taille du linker séparant l'amine tertiaire de la fonction alcyne peut avoir un impact sur la réaction de click.

Finalement, la réaction de CuAAC entre les différents ligands photoactivables et la **5-BrdU-N3** devra être étudiée en présence d'ADN G-quadruplex après avoir réalisé la réaction de photo-alkylation. En effet, il est possible que la génération d'espèces très réactives comme des carbènes, des nitrènes ou des espèces diradicalaires puisse réagir intramoléculairement avec la fonction alcyne. Pour cela, l'expérience envisagée sera de réaliser la réaction de CuAAC avec un fluorophore commercial (Cy5-N3 par exemple) après avoir préalablement irradié les différents ligands en présence d'ADN G-quadruplex. Les échantillons pourront ensuite être séparés sur gel d'électrophorèse et analysés²⁵¹.

5. Photo-alkylation des structures G-quadruplexes

La capacité des différents ligands synthétisés à sélectivement générer des liaisons covalentes avec les G-quadruplexes après irradiation a été étudiée par gel d'électrophorèse en conditions dénaturantes. Trois séquences ont été étudiées: la séquence télomérique humaine 22AG, la séquence cMyc22 et la séquence CEB25wt. Tous les ligands ont pu être utilisés lors de ces expériences hormis le composé **PDC-PEG-N2-Alk** qui a été synthétisé trop tardivement.

Pour ces expériences de photo-alkylation, les différentes séquences G4s sont marquées par un atome de phosphore radioactif (³²P) à l'aide d'une réaction enzymatique. La réaction de phosphorylation est réalisée en position 5' de l'oligonucléotide par l'enzyme polynucléotide kinase T4 incubée pendant 1 h à 37 °C en présence d'ATP-³²P. Le milieu réactionnel est ensuite purifié par gel d'électrophorèse dénaturant ou directement filtré. L'utilisation d'ADN radiomarqué offre plusieurs avantages. Le premier avantage est la grande sensibilité de détection qui permet de pouvoir travailler à de faibles concentrations d'ADN radiomarqué. Le deuxième avantage est de pouvoir éviter la fonctionnalisation de l'ADN par un fluorophore qui peut avoir un effet négatif sur l'interaction avec le ligand et qui pourrait surtout générer des réactions parasites après photo-irradiation en présence des ligands.

Ainsi, les différentes réactions de photo-alkylation ont été réalisées en présence de 10 μM d'ADN non marqué et d'une quantité suffisante d'ADN marqué (entre 20 000 et 50 000 c/s). L'ADN Gquadruplex est structuré par chauffage à 95 °C pendant 10 min et laissé refroidir lentement à température ambiante en tampon K⁺100. Les différents ligands sont ensuite ajoutés et sont laissés reposer à l'abri de la lumière pendant environ 1 h avant d'être soumis à l'irradiation. Différents temps d'irradiation sont réalisés à une longueur d'onde de 365 nm et les échantillons sont analysés par gel d'électrophorèse d'acrylamide en conditions dénaturantes (Figure 115A). Ainsi, lorsque l'alkylation est réalisée sur un brin d'ADN, sa masse molaire sera plus élevée que le brin d'ADN nonalkylé et une bande retardée devrait être observée. Les rendements d'alkylation sont ensuite calculés par le rapport de l'intensité de la radioactivité émise par les produits d'alkylation et de l'intensité totale de radioactivité. L'irradiation des échantillons est réalisée avec un nombre de lampes de 2 à 10 (15W par lampe, 365 nm) au sein d'un photoréacteur multiray (Figure 115B).



Figure 115: Photo-alkylation des G-quadruplexes. A) Schéma représentatif du principe de l'étude de la photo-alkylation des structures G-quadruplexes. B) Photographie du photoréacteur utilisé pour cette étude.

a. Photo-alkylation de la séquence télomérique humaine 22AG

Les expériences de photo-alkylations ont tout d'abord été réalisées et optimisées sur la séquence télomérique 22AG. Plusieurs paramètres peuvent ainsi être modifiés pour optimiser le rendement d'alkylation des composés comme le nombre d'équivalents de ligands par rapport à l'ADN G-quadruplex, le nombre de lampes utilisées pour irradier les échantillons ainsi que le temps d'irradiation. Ainsi, des premières expériences d'irradiations pendant 5, 15, 30 et 60 min avec 4 lampes en présence de la séquence 22AG (10 µM) et des ligands des séries **PDC-linéaires**, **PDC-C1** et **PDC-C3** (20 µM) ont été réalisées (Figure 116).



Figure 116 : Résultats des réactions de photo-alkylation analysées par gels d'électrophorèse d'acrylamide 15% en conditions dénaturantes sur la séquence 22AG (10 µM) avec 4 lampes (365 nm) pour les composés PDC-4-N2-Alk, PDC-2-N2-Alk, PDC-C1-N2, PDC-C1-N3, PDC-C1-Bp, PDC-C3-N2 et PDC-C3-Bp (20 µM) non irradiés (Ø) ou irradiés 5, 15, 30 et 60 min. Les deux bandes « 22AG » correspondent aux contrôles sans ligands non irradiés (Ø) ou irradiés 60 min. Le trait noir délimite un gel qui a été rogné. Les structures des différentes fonctions photoactivables sont représentées au-dessus des noms des ligands.

Ces premières expériences permettent de montrer plusieurs résultats intéressants. Tout d'abord, il est important de noter que la séquence 22AG sans ligand ne présente aucun produit de dégradation due à l'irradiation pendant 1 h à 365 nm. Cette observation est importante, car ces expériences de photo-alkylation seront, par la suite, réalisées au sein de cellules vivantes et l'irradiation doit avoir le moins d'effet possible sur l'ADN. Ensuite, les différents composés irradiés en présence de la séquence 22AG sont tous capables de former au moins un nouveau produit de migration retardée, attribué à la photo-alkylation de la structure G-quadruplex. Les rendements de photo-alkylation sont très différents d'une fonction photoactivable à une autre. Par exemple, les deux dérivés linéaires **PDC-4-N2-Alk** et **PDC-2-N2-Alk** sont capables de réagir très efficacement avec la séquence 22AG avec plus de 30% de rendement d'alkylation en 5 min d'irradiation et plus de 65% de rendement après 1 h d'irradiation. Les autres composés azotures, diazirines et benzophénones montrent des rendements d'alkylation plus faibles. Par exemple, l'irradiation du composé azoture **PDC-C1-N3** et des composés

benzophénones PDC-C1-Bp et PDC-C3-Bp pendant 1 h permet d'alkyler la séquence 22AG avec des rendements modérés compris entre 8 et 14% dans ces conditions. De plus, les deux composés PDC-C1/C3-Bp sont capables de former deux bandes pouvant notamment être attribuées à la formation de deux produits d'alkylation différents. Finalement, l'irradiation des composés PDC-C1-N2 et PDC-C3-N2 permet d'obtenir un rendement maximum d'alkylation d'environ 3% après 1 h. Le faible rendement d'alkylation pour ces deux fonctions photoactivables N2-aromatiques est relativement étonnant au vu de la grande réactivité des dérivés N2-linéaires. Cette observation peut notamment être expliquée par une mauvaise proximité spatiale entre la fonction photoactivable et la structure G-quadruplex. Dans le but de vérifier si les rendements d'alkylations pouvaient être améliorés en augmentant la quantité d'énergie apportée au milieu, les expériences ont été répétées dans les mêmes conditions en utilisant la totalité des lampes du photo-réacteur (Figure 117).

0

			$\mathcal{A}_{H} \xrightarrow{H} \mathcal{A}_{H} \xrightarrow{H} \mathcal{A}_{H} \xrightarrow{N=N} \mathcal{A}_{H}$					P ^{A^AN N F F F}					
	22	AG		PDC-4-N2-Alk					PDC-C1-N2				
t / min	Ø	60	ø	5 34%	15 49%	30 56%	60 57%	ø	5 3%	15 3%	30 3%	60 3%	
	-	-	-										
				Ø 1 ↓ 13 PDC-C1-N3					PDC-C1-Bp				
			Ø	5 3%	15 6%	30 8%	60 9%	ø	5 4%	15 11%	30 15%	60 1 7%	
			_	_	_	_	_			-	-	-	
			0	P P N N N N N N N F F F									
				PDC-C3-N2					PDC-C3-Bp				
			ø	5 2%	15 3%	30 3%	60 3%	ø	5 :	15 3 7% 3	0 60 3% 10))%	

Figure 117 : Résultats des réactions de photo-alkylation analysées par gels d'électrophorèse d'acrylamide 15% en conditions dénaturantes sur la séquence 22AG (10 µM) avec 10 lampes (365 nm) pour les composés PDC-4-N2-Alk, PDC-C1-N2, PDC-C1-N3, PDC-C1-Bp, PDC-C3-N2 et PDC-C3-Bp (20 µM) non irradiés (Ø) ou irradiés 5, 15, 30 et 60 min. Les deux bandes « 22AG » correspondent aux contrôles sans ligands non irradiés (Ø) ou irradiés 60 min. Le trait noir délimite un gel qui a été rogné. Les structures des différentes fonctions photoactivables sont représentées au-dessus des noms des ligands.

L'irradiation des différents composés par les 10 lampes du photo-réacteur permet d'augmenter légèrement la quantité d'alkylation pour les dérivés fonctionnalisés par un groupement benzophénone PDC-C1-Bp et PDC-C3-Bp. Par exemple, l'irradiation du composé PDC-C1-Bp par 4 ou 10 lampes en présence de la séquence 22AG permet de générer des adduits avec un rendement de 14 et 17% respectivement. Globalement, l'augmentation du nombre de lampes ne permet pas d'améliorer les rendements d'alkylation pour les autres composés. Dans ces conditions, le dérivé PDC-4-N2-Alk reste le ligand permettant d'alkyler la séquence 22AG le plus efficacement avec environ 50% d'alkylation après 15 min d'irradiation. Finalement, la réaction de photo-alkylation a été réalisée pour les composés PDC-TEG en présence de la séquence 22AG avec 10 lampes (Figure 118).



Figure 118: Résultats des réactions de photo-alkylation analysées gels d'électrophorèse d'acrylamide 15% en conditions dénaturantes sur la séquence 22AG (10 μM) avec 10 lampes (365 nm) pour les composés PDC-TEG-N2, PDC-TEG-N3 et PDC-TEG-Bp (20 μM) non irradiés (Ø) ou irradiés 5, 15, 30 et 60 min. Les deux bandes « 22AG » correspondent aux contrôles sans ligands non irradiés (Ø) ou irradiés 60 min. Les structures des différentes fonctions photoactivables sont représentées au-dessus des noms des ligands.

Cette expérience permet notamment de confirmer que la quantité d'alkylation dépend surtout de la fonction photoactivable et très peu de la taille ou de la polarité du linker. Par exemple, le composé **PDC-TEG-N3** et le composé **PDC-C1-N3** permet d'alkyler 8 et 9% respectivement la séquence 22AG. Cependant, de légères différences sont observables entre les composés **PDC-C1-Bp**, **PDC-C3-Bp** et **PDC-TEG-Bp** dont la quantité de photo-alkylation est respectivement de 17, 10 et 13%.

Dans le cas des composés non linéaires, il est intéressant de noter que la quantité d'alkylation continue d'augmenter entre 30 et 60 min après irradiation par 4 ou 10 lampes alors que les composés linéaires arrivent à un palier après 30 min. Dans le but de vérifier s'il est possible d'améliorer les rendements d'alkylation, tous les composés non linéaires sont irradiés pendant 2 h en présence de la séquence 22AG, les résultats sont résumés ci-après (Figure 119).



Figure 119: Rendements d'alkylation obtenus pour les différents ligands non linéaires avec la séquence 22AG après irradiation avec 10 lampes pendant 5, 15, 30, 60 et 120 min.

Ce graphique permet de montrer que la quantité d'alkylation n'augmente pas ou peu entre 1 h ou 2 h d'irradiation en présence des différents composés. Il permet de plus de montrer que les ligands non linéaires les plus efficaces pour réaliser un lien covalent avec l'ADN sont les dérivés de la benzophénone **PDC-C1-Bp** et **PDC-TEG-Bp** suivi des ligands **PDC-C1-N3**, **PDC-C3-Bp** et **PDC-TEG-N3**. Il est important de noter que les différentes valeurs obtenues lors de ces expériences sont difficiles à rationaliser avec les données de biophysique. Par exemple, le meilleur composé trifonctionnel testé par photochimie, le ligand **PDC-C1-Bp**, stabilise la séquence télomérique F21T avec un $\Delta T_{1/2}$ d'environ 15 °C par FRET-melting tandis que l'un des moins bon photo-alkylants, le ligand **PDC-C3-N2**, stabilise la séquence F21T avec un $\Delta T_{1/2}$ d'environ 23 °C. Cette observation montre que les rendements d'alkylation dépendent principalement de la nature du groupement photoactivable. Globalement, les différents ligands possédant une fonction diazirine aromatique ne sont que très peu capables d'alkyler la séquence 22AG.

Les spectres d'absorption des différents composés réalisés précédemment (Figure 105) montrent que ces derniers absorbent plus efficacement à des longueurs d'onde inférieures à 365 nm. Dans le but de vérifier si la longueur d'onde des lampes avait un impact sur l'activation des différents groupements photoactivables, les ligands ont été irradiés en présence de 22AG avec 4 lampes de 310 nm. Uniquement les gels des composés **PDC-C1-N2**, **PDC-C1-N3** et **PDC-C1-Bp** sont présentés ici (Figure 120).



Figure 120: Résultats des réactions de photo-alkylation analysées par gels d'électrophorèse d'acrylamide 15% en conditions dénaturantes sur la séquence 22AG (10 μM) avec 4 lampes (310 nm) pour les composés PDC-C1-N2, PDC-C1-N3 et PDC-C1-Bp (20 μM) non irradiés (Ø) ou irradiés 5, 15, 30 et 60 min. Les deux bandes « 22AG » correspondent aux contrôles sans ligands non irradiés (Ø) ou irradiés 60 min. Le trait noir délimite un gel qui a été rogné. Les structures des différentes fonctions photoactivables sont représentées au-dessus des noms des ligands.

Il est tout d'abord intéressant de noter que la séquence 22AG sans ligand irradié pendant 60 min à 310 nm présente deux produits de migration plus rapide indiquant une réaction de l'ADN avec la lumière. Ces produits peuvent notamment résulter de la réaction de cycloaddition [2+2] de deux thymines adjacentes²⁶⁵ au niveau des boucles TTA de la séquence 22AG. Globalement, la diminution de la longueur d'onde, et donc l'augmentation de l'énergie apportée aux échantillons, permet d'augmenter les quantités d'alkylation. Après seulement 5 min, les rendements d'alkylation sont de 7, 20 et 31% pour les composés PDC-C1-N2, PDC-C1-N3 et PDC-C1-Bp respectivement. Les rendements sont donc multipliés par 2, 10 et 15 comparés à l'irradiation par 4 lampes à 365 nm. La forte augmentation de l'efficacité de la réaction par la diminution de la longueur d'onde des lampes pour les composés PDC-C1-N3 et PDC-C1-Bp semble indiquer une bonne proximité spatiale entre la structure G-quadruplex et ces fonctions photoactivables. La faible augmentation d'alkylation induite par le composé PDC-C1-N2 après irradiation à 310 nm semble, contrairement aux autres composés, indiquer une mauvaise proximité spatiale entre la fonction diazirine et le G-quadruplex. Pour confirmer cette hypothèse, une solution du ligand PDC-C1-N2 (200 µM) dans l'eau a été irradiée 5 min avec 10 lampes (365 nm) sans présence de structures G-quadruplexes en solution avant d'être analysée par HPLC et LC/MS (Figure 121).

161



Figure 121: Chromatogrammes HPLC obtenu après injection du ligand PDC-C1-N2 (200 μM) dans l'eau sans irradiation (Ø) et après 5 min d'irradiation par 10 lampes à 365 nm et zoom sur le pic principal obtenu après analyse LC/MS. La structure de la fonction photoactivable avant irradiation est représentée en noir et la structure supposée de la fonction photoactivable après irradiation est représentée en bleu.

L'analyse HPLC du composé **PDC-C1-N2** après irradiation pendant 5 min par 10 lampes à 365 nm dans l'eau permet de montrer que ce dernier forme un produit majoritaire avec plus de 90% de conversion. L'analyse LC/MS du milieu réactionnel a de plus permis de montrer une masse de 860,6, pouvant notamment correspondre au produit de l'insertion du carbène dans la liaison O-H de l'eau comme précédemment décrit²⁶⁶. D'autres études de photolyse seront réalisées dans le but de caractériser les espèces réactives générées par les différents composés. Cette expérience permet cependant de démontrer que les dérivés aromatiques de la diazirine sont activés et quenchés très rapidement dans l'eau. Ainsi, en présence de structures G-quadruplexes, si le carbène généré n'est pas directement en contact d'un substrat avec lequel réagir, il sera désactivé presque instantanément. Cela permet de prouver que la faible alkylation observée avec ces composés vient d'une mauvaise proximité spatiale entre la diazirine et les nucléotides de la structure G-quadruplex.

L'étude des réactions de photo-alkylation a ensuite été réalisée sur la séquence cMyc22 avec les différents composés synthétisés.

b. Photo-alkylation de la séquence cMyc22

Au vu des résultats des réactions de photo-alkylation sur la séquence 22AG, des premières expériences ont été réalisées avec les deux diazirines linéaires sur la séquence cMyc22 (Figure 122).



Figure 122: Résultats des réactions de photo-alkylation analysées par gels d'électrophorèse d'acrylamide 15% en conditions dénaturantes sur la séquence cMyc22 (10 μM) avec les composés PDC-4-N2-Alk et PDC-2-N2-Alk (20 μM) non irradiés (Ø) ou irradiés 5, 15, 30 et 60 min par A) 4 lampes et B) 10 lampes (365 nm). Les deux bandes « cMyc22 » correspondent aux contrôles sans ligands non irradiés (Ø) ou irradiés 60 min. Les structures des différentes fonctions photoactivables sont représentées au-dessus des noms des ligands. Les flèches rouges montrent la présence de bandes de faible intensité.

Les réactions des composés **PDC-4-N2-Alk** et **PDC-2-N2-Alk** sur la séquence cMyc22 après irradiation par 4 lampes (Figure 122A) ou 10 lampes (Figure 122B) permettent de montrer que les composés sont capables d'alkyler efficacement cette séquence. Par exemple, 30 min d'irradiation par 4 lampes permettent de générer deux produits d'alkylation avec 5 et 27% de rendement respectivement pour les deux ligands. L'augmentation du nombre de lampes permet, contrairement aux expériences avec la séquence 22AG, d'augmenter les rendements d'alkylation pour ces deux composés en formant de nouvelles bandes. Il est intéressant de noter que les produits d'alkylation majoritairement formés migrent plus rapidement que l'ADN de référence. Cette observation est difficile à rationaliser, mais peut provenir de la formation d'un produit très stable et plus compact que la séquence cMyc22 dénaturée et qui pourrait passer plus facilement à travers le réseau formé par le gel de polyacrylamide. Dans le but d'essayer de déstabiliser ces potentielles structures, les échantillons ont été chauffés à 95 °C pendant 10 min en présence de formamide avant d'être séparés par gel d'électrophorèse, mais aucune différence n'a pu être remarquée.

Ensuite, les réactions de photo-alkylation ont été réalisées avec les composés non linéaires sur la séquence cMyc22. Au vu des rendements modérés obtenus sur la séquence 22AG, 10 lampes ont directement été utilisées. Les résultats obtenus pour les ligands après 60 min d'irradiation sont présentés ci-après (Figure 123).



Figure 123: Résultats des réactions de photo-alkylation analysées par gels d'électrophorèse d'acrylamide 15% en conditions dénaturantes sur la séquence cMyc22 (10 μM) avec 10 lampes (365 nm) pour les composés PDC-C1-N2, PDC-C1-N3, PDC-C1-Bp, PDC-C3-N2, PDC-C3-Bp, PDC-TEG-N2, PDC-TEG-N3 et PDC-TEG-Bp (20 μM) non irradiés (Ø) ou irradiés 60 min. Les deux bandes « cMyc22 » correspondent aux contrôles sans ligands non irradiés (Ø) ou irradiés 60 min. Le trait noir délimite un gel qui a été rogné. Les structures des différentes fonctions photoactivables sont représentées au-dessus des noms des ligands.

Ces expériences de photo-alkylation de la séquence cMyc22 en présence des ligands non linéaires montrent que tous les ligands, excepté le ligand PDC-C1-N2, sont capables d'alkyler plus ou moins efficacement cette séquence. Tout comme pour les dérivés PDC-4-N2-Alk et PDC-2-N2-Alk, les produits formés après irradiation montrent une migration accélérée par rapport à l'ADN de référence. Cependant, dans le cas des dérivés non linéaires, les produits d'alkylation forment des bandes très proches de l'ADN non-alkylé ce qui rend difficile la quantification des rendements d'alkylation. Il est cependant intéressant de remarquer que les ligands PDC-TEG-N3 et PDC-TEG-Bp sont capables d'induire la formation presque totale d'un nouveau produit de migration accélérée, suggérant une alkylation totale ou presque de la séquence cMyc22. Dans le but de suivre la formation de ces produits au cours du temps, une expérience de suivi cinétique a été réalisée dans les mêmes conditions avec les composés PDC-TEG-N2, PDC-TEG-N3 et PDC-TEG-Bp (Figure 124).



Figure 124: Résultats des réactions de photo-alkylation analysées par gels d'électrophorèse d'acrylamide 15% en conditions dénaturantes sur la séquence cMyc22 (10 μM) avec 10 lampes (365 nm) pour les composés PDC-TEG-N2, PDC-TEG-N3 et PDC-TEG-Bp (20 μM) non irradiés (Ø) ou irradiés 5, 15, 30 et 60 min. Les deux bandes « cMyc22 » correspondent aux contrôles sans ligands non irradiés (Ø) ou irradiés 60 min. Les structures des différentes fonctions photoactivables sont représentées au-dessus des noms des ligands.

Le suivi cinétique de la réaction de photo-alkylation par les ligands TEG montre que la formation du nouveau produit est très rapide. Après seulement 5 min d'irradiation, tous les composés sont capables de générer un nouveau produit de migration légèrement plus rapide que celle de l'ADN cMyc22 non-alkylé. L'augmentation du temps d'irradiation permet généralement de former un ou plusieurs produits d'alkylation de migration plus rapide que l'ADN non-alkylé.

c. Photo-alkylation de la séquence CEB25wt

Finalement, l'étude de la réaction de photo-alkylation par les différents ligands a été réalisée sur la séquence CEB25wt. La cinétique de réaction des deux composés diazirines linéaires **PDC-4-N2-Alk** et **PDC-2-N2-Alk** est présentée ci-après (Figure 125).



Figure 125: Résultats des réactions de photo-alkylation analysées par gels d'électrophorèse d'acrylamide 15% en conditions dénaturantes sur la séquence CEB25wt (10 μM) avec 4 lampes (365 nm) pour les composés PDC-4-N2-Alk et PDC-2-N2-Alk (20 μM) non irradiés (Ø) ou irradiés 5, 15, 30 et 60 min. Les deux bandes « CEB25wt » correspondent aux contrôles sans ligands non irradiés (Ø) ou irradiés 60 min. Le trait noir délimite un gel qui a été rogné. Les structures des différentes fonctions photoactivables sont représentées au-dessus des noms des ligands.

L'irradiation par 4 lampes des composés **PDC-4-N2-Alk** et **PDC-2-N2-Alk** en présence de la séquence CEB25wt montre une forte capacité de ces derniers à alkyler la structure G-quadruplex formée par cette séquence. Par exemple, les ligands **PDC-4-N2-Alk** et **PDC-2-N2-Alk** sont capables de générer un nouveau produit de migration plus rapide avec 56 et 58% de rendement respectivement après 30 min d'irradiation. Il est aussi intéressant de noter que la migration de la bande CEB25wt avec chacun des ligands sans irradiation (Ø) est légèrement plus rapide que la bande de la séquence CEB25wt sans ligand avec ou sans irradiation. Cette observation est très reproductible entre différentes expériences et pourrait provenir de la formation d'un complexe très stable entre la structure Gquadruplex et le ligand.

La réaction de photo-alkylation a ensuite été étudiée sur cette séquence avec les différents composés non linéaires (Figure 126).



Figure 126: Résultats des réactions de photo-alkylation analysées par gels d'électrophorèse d'acrylamide 15% en conditions dénaturantes sur la séquence CEB25wt (10 μM) avec 4 lampes (365 nm) pour les composés PDC-C1-N2, PDC-C1-N3, PDC-C1-Bp, PDC-C3-N2 et PDC-C3-Bp (20 μM) et 10 lampes pour les composés PDC-TEG-N2, PDC-TEG-N3 et PDC-TEG-Bp (20 μM) non irradiés (Ø) ou irradiés 5, 15, 30 et 60 min. Les deux bandes « CEB25wt » correspondent aux contrôles sans ligands non irradiés (Ø) ou irradiés 60 min. Les structures des différentes fonctions photoactivables sont représentées au-dessus des noms des ligands.

Un exemple de suivi cinétique de la réaction est présenté ici avec les différents ligands non linéaires. Les dérivés **PDC-C1** et **PDC-C3** ont été irradiés avec 4 lampes tandis que les composés **PDC-TEG** ont été irradiés directement avec 10 lampes. Comme avec la séquence cMyc22, l'irradiation des composés en présence de la séquence CEB25wt permet de former un nouveau produit de migration légèrement accélérée par rapport à la séquence CEB25wt, mais très difficile à séparer. De plus, augmenter le pourcentage d'acrylamide à 20% n'a pas permis d'améliorer la séparation des différents produits formés (résultats non montrés ici).

Finalement, seule une étude par séquençage permettra de démontrer la nature des produits générés par la photo-alkylation.

d. Sélectivité de la réaction de photo-alkylation

La formation d'un lien covalent entre une structure G-quadruplex et les ligands photoactivables devra, *in fine*, être réalisée directement en cellule. Il est donc impératif que la réaction de photoalkylation des structures G-quadruplexes soit la plus sélective possible. Pour le vérifier, des expériences de compétition ont été réalisées en présence d'ADN G-quadruplex et d'un excès d'ADN duplex. Pour cela, une seule condition d'irradiation a été choisie pour chaque ligand capable de réagir efficacement avec l'ADN G-quadruplex : 30 min et 4 lampes pour les ligands linéaires, 1 h et 10 lampes pour les ligands non linéaires. Les réactions ont été réalisées en présence de 10, 20, 50 et 100 μ M de compétiteur duplex (ds26), soit 1, 2, 5 et 10 équivalents par rapport à l'ADN G-quadruplex. Un exemple de réaction sur la séquence 22AG pour les différents ligands est présenté ci-après (Figure 127).



Figure 127: Résultats des réactions de photo-alkylation analysées par gels d'électrophorèse d'acrylamide 15% en conditions dénaturantes sur la séquence 22AG (10 μM) pour les composés PDC-4-N2-Alk et PDC-2-N2-Alk (20 μM, 4 lampes, 30 min) et pour les composés PDC-C1-Bp, PDC-C3-Bp et PDC-TEG-Bp (20 μM, 10 lampes, 1 h) en présence de 0, 10, 20, 50 et 100 μM de ds26. Les deux bandes « 22AG » correspondent aux contrôles sans ligands non irradiés (Ø) ou irradiés 30 min. Les structures des différentes fonctions photoactivables sont représentées au-dessus des noms des ligands.

Cette expérience permet de montrer que la réaction de photo-alkylation n'est pas ou peu affectée par la présence d'ADN duplex dans le milieu. Par exemple, dans le cas du composé **PDC-4-N2-Alk**, la présence de 10 équivalents de ds26 ne diminue que de 3% le rendement d'alkylation de la séquence

22AG. Pour le cas du composé **PDC-2-N2-Alk**, la présence de 10 équivalents diminue de 14% le rendement de photo-alkylation. Finalement, pour les composés non linéaires fonctionnalisés par une benzophénone, la présence d'ADN duplex dans le milieu ne diminue pas les rendements d'alkylation. Cette expérience prouve, tout comme l'expérience de FRET-melting, que les ligands synthétisés sont très sélectifs pour la structure G-quadruplex formée par la séquence 22AG comparé à l'ADN duplex.

Enfin, cette réaction a été réalisée dans les mêmes conditions pour les séquences cMyc22 et CEB25wt avec les ligands linéaires **PDC-4-N2-Alk** et **PDC-2-N2-Alk**. Les rendements d'alkylation sont normalisés par rapport aux rendements d'alkylation obtenus sans présence d'ADN compétiteur et les résultats sont présentés ci-après pour les trois séquences 22AG, cMyc22 et CEB25wt (Figure 128).



Figure 128 : Quantité d'alkylation normalisée sans compétiteur ou avec 10, 20, 50 et 100 μM de compétiteur. Les ligands PDC-4-N2-Alk et PDC-2-N2-Alk sont irradiés 30 min avec 4 lampes et les composés PDC-C1-Bp, PDC-C3-Bp et PDC-TEG-Bp sont irradiés 1 h avec 10 lampes.

Ces graphiques permettent de montrer que, même avec un large d'excès d'ADN duplex, les quantités d'alkylation des G-quadruplexes par irradiation des différents ligands sont relativement peu affectées. En effet, la diminution maximale du rendement de réaction est de 15% avec le ligand **PDC-2-N2-Alk** sur la séquence 22AG pour une concentration de 50 μ M de compétiteur. De plus, l'augmentation à 100 μ M de ds26 ne semble pas avoir un impact supérieur pour ce ligand. Globalement, tous les ligands possèdent une forte sélectivité envers les structures G-quadruplexes, ce qui représente un réel atout pour l'utilisation de ces molécules *in cellulo*.

6. Conclusion

Une nouvelle méthodologie de séquençage permettant l'identification de la localisation et de l'accessibilité des structures G-quadruplexes au niveau du génome est actuellement en cours de développement. Cette méthodologie s'appuie sur la capacité de dérivés du ligand PDC à interagir sélectivement avec l'ADN G-quadruplex. Ces dérivés sont fonctionnalisés par un alcyne et par des groupements photoactivables pouvant générer des espèces électrophiles très réactives après irradiation. Ces espèces pourront réagir sélectivement avec les structures G-quadruplexes formées à dans le génome en générant un lien covalent entre le ligand et l'ADN. Une fois la liaison réalisée, les molécules seront soumises à une réaction de Huisgen catalysée au cuivre avec la 5-BrdU-N3 dans le but de réaliser la capture des fragments d'ADN. Cette méthodologie permettra à la fois de rendre compte de la formation de G-quadruplexes *in vivo* et de minimiser la relocalisation des ligands.

Pour mener à bien ce projet, plusieurs ligands ont été synthétisés. Ces ligands diffèrent par la nature de la fonction photoactivable ainsi que par la taille et la polarité des différents bras espaceurs. Ainsi, trois ligands linéaires et huit ligands non linéaires ont été synthétisés. Les ligands linéaires sont fonctionnalisés par une fonction diazirine aliphatique plus ou moins éloignée du ligand PDC. Les ligands non linéaires sont fonctionnalisés par trois différentes fonctions photoactivables : une fonction diazirine aromatique, une fonction azoture aromatique et une benzophénone. De plus, la fonction alcyne est plus ou moins éloignée du reste de la molécule par différents bras espaceurs. Cette grande variété de ligands permettra de vérifier l'impact du positionnement de la fonction alcyne et des fonctions photoactivables sur l'affinité des ligands ainsi que sur leur capacité à générer des adduits covalents avec l'ADN.

Les différents ligands synthétisés sont ensuite testés pour leur capacité à interagir avec différentes séquences G-quadruplexes par des études de FRET-melting et de FID. Trois séquences ont été testées par FRET-melting et une séquence par FID, ces expériences sont actuellement en cours de réalisation et d'autres séquences seront utilisées par la suite. Ces expériences biophysiques montrent globalement des différences entre les ligands et il est pour l'instant difficile de rationaliser les résultats en fonction de la structure des molécules. Généralement, les ligands linéaires montrent des interactions plus fortes avec les G-quadruplexes que les ligands non linéaires par ces deux méthodes mais cette observation n'est pas toujours vérifiée en fonction de la séquence. Il est cependant

important de remarquer que tous les ligands synthétisés sont très sélectifs de l'ADN G-quadruplex par rapport à une séquence d'ADN duplex compétiteur.

Ensuite, la réaction de Huisgen catalysée au cuivre entre les ligands synthétisés et la 5-BrdU-N3 a été étudiée par HPLC. Globalement, cette étude montre que la réaction est très rapide pour la majeure partie des ligands lorsqu'il n'y a pas d'ADN G-quadruplex dans le milieu. Dans le but de s'assurer que la réaction est toujours possible après avoir généré un lien covalent avec l'ADN, des expériences supplémentaires seront nécessaires. Une expérience notamment envisagée sera de réaliser la réaction de CuAAC entre les ligands photoactivables et un fluorophore après avoir irradié ces derniers en présence d'ADN. La séparation par gel d'électrophorèse permettra ensuite de pouvoir visualiser la formation du produit ligand/G4/fluorophore.

Enfin, la capacité des ligands à générer liaisons covalentes avec différentes structures Gquadruplexes a été évaluée par gel d'électrophorèse. Pour ces expériences, différentes conditions ont été évaluées comme le temps d'irradiation, la concentration des composés et la quantité d'énergie apportée au milieu. Globalement, les ligands linéaires permettent de réaliser des liens covalents très efficacement avec les séquences d'ADN G-quadruplexes 22AG, cMyc22 et CEB25wt. Ces expériences ont de plus permis de montrer que le type de fonction photoactivable des ligands non linéaires avait un impact très marqué sur les rendements d'alkylation. Dans notre cas, les fonctions photoactivables pouvaient être classées en fonction des rendements obtenus : Bp > N3 > N2. Enfin, l'irradiation en présence d'un excès de compétiteur d'ADN duplex a permis de montrer que la réaction de photo-alkylation était très sélective pour l'ADN G-quadruplex.

Finalement, de nombreuses expériences restent à être réalisées. Comme expliqué précédemment, la réaction de CuAAC doit être réalisée avec un fluorophore après irradiation des composés en présence d'ADN. Ensuite, le protocole d'extraction supportée après génération d'adduits covalents et le protocole de séquençage doivent être optimisés sur des extraits cellulaires. Dans une dernière partie, les ligands seront directement utilisés *in cellulo* dans le but de déterminer avec précision la localisation et l'accessibilité des G-quadruplexes au sein du génome.

Chapitre IV: Étude de l'impact de la

régioisomérie de dérivés

bisquinolines/bisquinoliniums

1. Introduction

Comme décrit dans l'introduction, de nombreuses petites molécules ont été développées pour interagir sélectivement avec les G-quadruplexes. Parmi ces molécules, les dérivés bisquinoliniums comme le PDC¹²⁷ et le PhenDC3²⁶³ ou les dérivés bisquinolines comme la PDS¹²⁵ sont des ligands interagissant avec une forte affinité avec les structures G-quadruplexes (Figure 129). D'un point de vue structurel, la PDS diffère des dérivés bisquinoliniums PDC et PhenDC3 en deux points : le premier est le nombre de charges positives ainsi que leur localisation et le deuxième est la position de la liaison reliant la fonction amide à la fonction quinoline. Dans le cas de la PDS, trois charges sont présentes à pH physiologique au niveau des fonctions amines des trois chaines latérales tandis que pour les composés bisquinoliniums, les deux charges sont situées sur les azotes des quinolines et sont obtenues par méthylation de ces derniers. De plus, la régioisomérie entre la fonction amide et la quinoline n'est pas la même pour la PDS que pour le PDC et le PhenDC3. En effet, la quinoline est liée en position 2 à la fonction amide pour la PDS tandis que le groupement quinolinium est lié en position 3 à la fonction amide pour les dérivés PDC et PhenDC3.



Figure 129: Structures des ligands PDS, PDC et PhenDC3 avec, en rouge, la numérotation des atomes des groupements quinoline ou quinolinium.

En 2012, Müller *et al.* ont synthétisé un large panel de composés dérivés de la PDS²⁶⁷. Certains de ces dérivés sont fonctionnalisés par différentes chaines latérales possédant des amines primaires ou tertiaires. Parmi ces dérivés, le composé PDS-Pyro (Figure 130) stabilise aussi efficacement par FRET-melting la séquence télomérique humaine que le ligand de référence PDS ($\Delta T_{1/2} = 35,5$ et 34,8 °C respectivement). Ce même composé non fonctionnalisé par la troisième chaine sur le cycle pyridine (PDC2-Pyro, Figure 130) a montré, quant à lui, une température de stabilisation légèrement plus faible, mais toujours très importante ($\Delta T_{1/2} = 22,0$ °C).



Figure 130: Structures des ligands PDS-Pyro et PDC2-Pyro.

Il est intéressant de noter que tous ces ligands possèdent cette même fonctionnalisation en position 2 de la quinoline. Dans le but de comprendre l'importance de la régioisomérie de la fonction amide sur l'interaction avec les G-quadruplexes ainsi que l'importance de la position des charges, plusieurs dérivés du ligand PDC et PhenDC3 ont été synthétisés (Figure 131).



Figure 131: Structures des différents isomères dérivés du PDC et du PhenDC3 synthétisés.

Ces dérivés sont fonctionnalisés par deux bras pyrrolidines et la jonction entre la quinoline et la liaison amide est réalisée soit en position 2 (PDC2/PhenDC2-Pyro) soit en position 3 (PDC/PhenDC3-Pyro) de la quinoline. Le dérivé PDC2-Pyro a de plus été fonctionnalisé en position 4 de la pyridine par la 1,4-diaminobutane dans le but de vérifier l'importance de la présence d'une amine libre sur ces dérivés. Finalement, le PDC2, un isomère du PDC dont les quinolines sont fonctionnalisées en position 2 a été synthétisé. Ce composé permettra de vérifier, par comparaison avec le PDC, l'impact de la régioisomérie de la fonction amide sur l'interaction avec les G-quadruplexes. Les différents dérivés ainsi synthétisés ont été évalués par FRET-melting et FID pour leur affinité et leur sélectivité avec les G-quadruplexes.

2. Synthèse des composés⁶

a. Synthèse des composés PDC et PhenDC Bispyro

La synthèse des dérivés PDC et PhenDC fonctionnalisés par deux groupements pyrrolidines avec une jonction en position 2 ou 3 de la quinoline est débutée par la synthèse de la **2-** et **3-amino-4pyroquinoline (80** et **82** respectivement, Schéma 17).



Schéma 17 : Synthèse des composés 2- et 3-amino-4-pyroquinoline.

Le dérivé 2-amino-4-pyrroquinoline est premièrement synthétisé en deux étapes à partir de la 4aminoquinolin-4-ol (**77**). La première étape est la chloration de la fonction hydroxyle en position 4 par réaction de trichlorure de phosphoryle pendant 16 h. Le composé **78** ainsi formé est ensuite soumis à une réaction de S_NAr avec la N-(2-hydroxyethyl)pyrrolidine (**79**) préalablement déprotonée par réaction de tBuOK pour former le composé **2-amino-4-pyroquinoline** (**80**) avec 50% de rendement après purification. Le régioisomère **3-amino-4-pyroquinoline** (**82**) est, quant à lui, obtenu à partir de la 4-chloroquinoline-3-amine (**81**) commercialement disponible par réaction de S_NAr avec la N-(2-hydroxyethyl)pyrrolidine (**79**) dans les mêmes conditions.

Les produits **PDC/PDC2-Pyro** et **PhenDC3/2-Pyro** sont ensuite obtenus par réaction de couplage peptidique (Schéma 18) entre les dérivés 2/3-amino-4-pyroquinoline et l'acide 2,6-pyridinedicarboxylique (**83**) ou l'acide 1,10-phenanthroline-2,9-dicarboxylique (**84**).

⁶ Synthèse réalisée avec l'aide de Laurent Sevaille



Schéma 18: Synthèse des composés PDC/PDC2-Pyro et PhenDC3/PhenDC2-Pyro.

Les deux dérivés **82** et **80** précédemment synthétisés sont mis à réagir avec l'acide 2,6pyridinedicarboxylique (**83**) en présence d'EDCI.HOBt et de DIPEA pour donner les composés **PDC-Pyro (85)** et **PDC2-Pyro (86)** avec 48 et 34% de rendement respectivement après purification par HPLCprep.

Les dérivés **PhenDC3-Pyro (87)** et **PhenDC2-Pyro (88)** sont obtenus par réaction des composés **82** et **80** avec l'acide 1,10-phénantroline-2,9-dicarboxylique (84) en présence d'EDCI.HOBt ou d'HATU avec respectivement 47 et 97% de rendement. Ces derniers sont purifiés par de simples lavages après précipitation dans l'eau.

Un autre composé nommé **PDC2-Pyro-4NH2** avait préalablement été synthétisé au laboratoire (Figure 132). Ce dernier est un dérivé du **PDC2-Pyro** fonctionnalisé en position 4 de la pyridine par la 1,4-diaminobutane. Cette molécule diffère de la **PDS-Pyro** par la chaine présente sur la pyridine : la chaine est composée de deux carbones supplémentaires et est connectée par un atome d'azote plutôt qu'un atome d'oxygène. Cette molécule est testée par FRET-melting et FID dans le but de rationaliser l'impact d'une troisième chaine protonable sur la capacité des molécules à interagir avec les G-quadruplexes.



Figure 132: Structure de la molécule PDC2-Pyro-4NH2 dérivée du ligand PDS-Pyro.

b. Synthèse du composé PDC2

Dans le but de vérifier l'impact de la régioisomérie de la liaison amide connectant le groupement quinolinium au cycle pyridine sur la reconnaissance des G-quadruplexes, un dérivé du PDC fonctionnalisé en position 2 de la quinoline a été synthétisé (Schéma 19). La synthèse est tout d'abord débutée par réaction de couplage peptidique entre la 2-aminoquinoline commercialement disponible et l'acide 2,6-pyridinedicarboxylique (**83**) pour donner le composé **89** avec 70% de rendement. Pour obtenir le composé **PDC2**, les conditions classiques de méthylation des quinolines (MeI, DMF, 40 °C) n'ont pu être appliquées, car aucune réaction ne se produisait. Plusieurs conditions ont été testées dans le but d'obtenir le composé **PDC2** et les meilleures conditions trouvées impliquent la méthylation des quinolines par le triflate de méthyle à 110 °C par irradiations micro-ondes. Le composé **PDC2** a ainsi été obtenu avec 25% de rendement après purification HPLCprep.



Schéma 19: Synthèse du composé PDC2 en deux étapes à partir de l'acide 2,6-pyridinedicarboxylique.
3. Études biophysiques

a. FRET-melting

Dans le but de caractériser la capacité des différents ligands à interagir avec l'ADN G-quadruplex, quatre séquences ont été testées : la séquence télomérique humaine F21T, la séquence dérivée du proto-oncogène FcMyc22T et deux séquences du minisatellite CEB25 : la séquence FCEB25wtT et la séquence modifiée FCEB25L111TT. Les expériences de FRET-melting ont été réalisées dans les mêmes conditions que précédemment décrit : 0,2 μ M d'ADN G4 fonctionnalisé par une FAM et une TAMRA, 1,0 μ M de ligand et 0, 3 ou 10 μ M d'ADN compétiteur duplex. Les ligands PDS, PDC et PhenDC3 sont utilisés comme ligands de référence et les résultats obtenus sont présentés ci-après (Figure 133).



Figure 133: Résultats des expériences de FRET-melting sur les séquences F21T, FcMyc22T, FCEB25wtT et FCEB25L111TT avec les différents composés synthétisés ainsi que les ligands PDS, PDC et PhenDC3 comme ligands de référence. Les barres d'erreur représentent l'écart-type de deux expériences individuelles.

Les différents résultats obtenus par FRET-melting indiquent généralement une faible capacité des ligands synthétisés à stabiliser les structures G-quadruplexes étudiées. Par exemple, les composés **PDC/PDC2-Pyro**, **PhenDC3/PhenDC2-Pyro** et **PDC2** stabilisent la séquence F21T avec un $\Delta T_{1/2}$ de

moins de 4 °C sans présence de compétiteur. Cette faible capacité à stabiliser les structures Gquadruplexes est, de plus, observée pour toutes les séquences étudiées ici. Le composé PDC2-Pyro-**4NH2** stabilise de façon légèrement plus efficace les différentes séquences étudiées avec un $\Delta T_{1/2}$ d'environ 15 °C pour la séquence F21T. Les effets sur la température de fusion des différents Gquadruplexes sont tout de même très éloignés de ceux observés pour le ligand PDS qui stabilise la séquence F21T avec un $\Delta T_{1/2}$ d'environ 30 °C. Cette différence peut notamment indiquer l'importance de la présence d'amines primaires au sein de cette famille de molécules. L'encombrement stérique généré par la présence des groupements pyrrolidines peut, de plus, avoir un effet négatif sur la reconnaissance des G-quadruplexes. Ensuite, il est très déconcertant de ne pas retrouver l'effet de stabilisation rapporté dans la littérature sur la séquence F21T pour le composé **PDC2-Pyro** : $\Delta T_{1/2}$ = 2,5 °C dans nos conditions contre $\Delta T_{1/2}$ = 20 °C dans les conditions de Müller *et al* ²⁶⁷ Étrangement, la seule différence notable entre les conditions testées par Müller et al. et les nôtres est le tampon de l'expérience. En effet, Müller et al. ont réalisé les expériences de FRET-melting dans un tampon K.caco 60 mM. Cependant, la modification du tampon n'a pas permis d'améliorer sensiblement les températures de stabilisation observées (données non montrées ici). Une deuxième observation déconcertante est la très faible capacité du ligand PDC2 à stabiliser les différentes structures Gquadruplexes. En effet, seule la position des fonctions amides sur les groupements quinoliniums est modifiée par rapport au ligand PDC et une différence de plus de 30 °C est observée pour la séquence F21T ou FcMyc22T.

Les différents ligands synthétisés ont ensuite été testés par FID sur les mêmes séquences Gquadruplexes.

b. FID

Les différentes séquences précédemment utilisées par FRET-melting sont étudiées par tests FID en plaques. Les pourcentages de TOD à 1 μ M de ligand sont présentés ci-après (Figure 134).



Figure 134: TOD 1 µM des ligands synthétisés sur les différentes séquences d'ADN G4 étudiées. Les barres d'erreurs représentent l'écart-type de deux expériences individuelles.

Les expériences de FID réalisées en plaques permettent de montrer des différences notables au sein des différents ligands. Par exemple, les ligands **PDC2-Pyro** et **PhenDC2-Pyro** permettent de déplacer environ 30% de la sonde de la séquence 22AG tandis que leurs isomères de position **PDC-Pyro** et **PhenDC3-Pyro** ne sont pas capables de déplacer la sonde. Cette observation peut notamment indiquer que les composés dérivés de la PDS protonables en solution physiologique n'adoptent pas une conformation adaptée à l'interaction avec les G-quadruplexes. L'encombrement stérique des deux bras pyrrolidines pourrait, en effet, empêcher les dérivés d'adapter la structure en « V » nécessaire à l'interaction avec le G-quartet externe et favoriser ainsi une forme ouverte comme décrit par Xie *et al*¹²⁴ (Figure 135).



Figure 135: Conformations pouvant être adoptées par le dérivé PDC-Pyro.

Les TOD 1 μ M des composés **PDC2/PhenDC2-Pyro** observés pour les autres séquences sont cependant légèrement plus faibles avec des TOD entre 20 et 10%. De plus, il intéressant de noter que le composé **PDC2-Pyro-4NH2** possède des TOD 1 μ M très proches de ceux du ligand **PDS** sur la séquence 22AG et cMyc22. Tout comme pour le ligand **PDS**, le dérivé **PDC2-Pyro-4NH2** n'est cependant pas capable de déplacer efficacement le TO des structures G-quadruplexes formées par la séquence CEB25wt et CEB25L111T.

Encore une fois, le dérivé **PDC2** ne montre aucune interaction avec les structures G-quadruplexes étudiées, ce qui est très étonnant. Les analyses par RMN ¹H de ce composé montrent que les protons des deux fonctions amides ne sont pas visibles sur les spectres, ce qui peut indiquer un échange très rapide avec le solvant. Cependant, il est aussi possible que l'étape de méthylation n'ait pas été réalisée sur les azotes des quinolines mais sur les azotes des amides. Cela pourrait expliquer l'absence d'affinité avec les G-quadruplexes et l'absence des protons amides sur les spectres RMN ¹H. Dans le but de s'assurer de la méthylation sur les quinolines, une autre voie de synthèse a été réalisée pour obtenir le composé **PDC2** (Schéma 20). Cette synthèse implique la méthylation préalable de la 2-aminoquinoline²⁶⁸ suivie de la réaction de substitution d'acyle avec le dérivé dichlorure d'acyle de l'acide pyridine-2,6-dicarboxylique. Le composé **PDC2** est finalement obtenu avec uniquement 14% de rendement après purification par HPLCprep.



Schéma 20: Synthèse du composé PDC2 par une deuxième voie de synthèse.

Les analyses RMN, LC/MS et biophysiques du composé **PDC2** nouvellement obtenu sont cependant identiques à celles du **PDC2** synthétisé précédemment. Cela indique que le composé obtenu par la première voie de synthèse possédait bien les deux groupements méthyles au niveau des azotes des quinolines. L'absence d'interaction avec les G-quadruplexes ainsi que l'absence des deux protons amides sur les spectres RMN ¹H peut aussi indiquer que les protons amides sont particulièrement acides. En effet, contrairement au ligand PDC, les doublets non liants des azotes des fonctions amides sont conjugués avec les azotes méthylés des quinolines. Cette forme de résonnance peut ainsi augmenter l'acidité de ces protons.

De simples calculs de pKa et de distribution des espèces en fonction du pH permettent de montrer que l'espèce principale en solution à pH = 7,2 n'est pas le $PDC2^{2+}$ mais certainement un mélange d'espèces chargées et non chargées (Figure 136).



Figure 136: Distribution des espèces du PDC2 en solution à différents pH calculée à l'aide du logiciel Marvin Sketch 5.11.4 (gauche). Structures et exemples des formes de résonnance du PDC2²⁺, du PDC2¹⁺ et du PDC2 neutre ainsi que les valeurs des pKa des deux fonctions amides (droite).

Dans le but de confirmer cette supposition, des mesures d'absorbance UV-Vis ont été réalisées à différents pH dans un tampon K⁺100 (Figure 137).



Figure 137: Spectres d'absorption des composés PDC et PDC2 dans le tampon K⁺100 à pH = 1,6; 4,5; 7,2 et 9,2.

Les mesures d'absorbance ont été réalisées à pH = 1,6 ; 4,5 ; 7,2 et 9,2. Tout d'abord, il est important de montrer que l'espèce majoritaire en solution ne dépend pas du pH pour le ligand **PDC**. Dans le cas du composé **PDC2**, le spectre d'absorbance dépend directement du pH. À pH = 1,6, le spectre d'absorbance montre un maximum à 333 nm qui se déplace vers le visible avec l'augmentation du pH. Ces expériences montrent que la conjugaison des amides avec les groupements quinoliniums augmente l'acidité des protons et que l'espèce principale en solution à pH = 7,2 n'est pas le PDC2²⁺. Cette déduction permet ainsi de comprendre le manque d'affinité de cette molécule pour les différents G-quadruplexes étudiés.

4. Conclusion

Dans le but d'étudier l'impact de la fonctionnalisation et de la régioisomérie de la quinoline sur des dérivés des ligands PDC et PhenDC3, plusieurs dérivés ont été synthétisés. Deux dérivés du ligand PDC et du ligand PhenDC3 non méthylés ont notamment été fonctionnalisés par deux bras pyrrolidines au niveau des positions 4 des quinolines. Ces deux ligands diffèrent entre eux par la position du lien amide sur la quinoline. Un troisième ligand dérivé du ligand PDC2-Pyro fonctionnalisé en position 4 de la pyridine centrale par la 1,4-diaminobutane a précédemment été synthétisé au laboratoire et a été évalué ici. Finalement, le dérivé PDC2 a été synthétisé. Ce dérivé est un analogue structurel du PDC dont les quinolines ne sont pas substituées en position 3 mais en position 2.

Les dérivés ainsi synthétisés ont été testés par des expériences de FRET-melting et FID sur quatre séquences différentes. Ces études biophysiques ont permis de montrer que les dérivés PDC/PhenDC-Pyro ne stabilisent que très faiblement les structures G-quadruplexes étudiées. Le composé PDC2-Pyro-4NH2 a, quant à lui, montré des températures de stabilisation plus élevées. Ce dérivé possède une amine primaire supplémentaire par rapport à la molécule PDC2-Pyro, ce qui démontre la nécessité de cette fonction pour l'interaction avec les G-quadruplexes. Il est cependant déconcertant d'observer une si faible température de stabilisation pour le ligand PDC2-Pyro qui a précédemment été synthétisé et évalué par FRET-melting sur la séquence F21T et FcMyc22T²⁶⁷. Différentes expériences ont été réalisées dans le but de comprendre cette différence de stabilisation induite, mais aucune conclusion ne peut être tirée pour le moment. Dans le cas des expériences de FID, le ligand PDC2-Pyro-4NH2 a montré une capacité à déplacer le TO des séquences 22AG et cMyc22 comparable à celle de la PDS. Au sein de la famille PDC/PhenDC-Pyro, seuls les dérivés fonctionnalisés en position 2 de la quinoline (PDC2-Pyro et PhenDC2-Pyro) ont démontré la capacité de déplacer une partie de la sonde du G-quadruplex.

Finalement, l'isomère de position du ligand PDC, le PDC2, n'a montré aucune affinité pour les différents G-quadruplexes. Cette observation était très étonnante au vu de sa structure proche de celle du PDC. Des calculs de pKa ont permis de démontrer que la première valeur de pKa des fonctions amides du PDC2 était d'environ 7,2, ce qui indique une déprotonation de la molécule à pH physiologique. Des analyses UV-Vis à différents pH ont permis de montrer que le spectre d'absorption variait entre pH = 1,6 et pH = 7,2. Ces différentes observations semblent donc indiquer que la molécule PDC2 ne possède pas réellement deux charges positives en solution mais qu'elle est, au moins partiellement, déprotonée.

Dans le but d'approfondir la compréhension de l'importance de la position de fonctionnalisation de la quinoline pour les dérivés non méthylés, il serait intéressant de synthétiser la PDS fonctionnalisée en position 3 de la quinoline (PDS3). Une synthèse envisagée est présentée ci-après (Schéma 21).



Schéma 21: Voie synthétique envisagée pour la synthèse du ligand PDS3.

Chapitre V : Conclusion générale et

perspectives

Différentes analyses du génome humain ont montré que les séquences d'ADN pouvant adopter des structures G-quadruplexes n'étaient pas distribuées de façon aléatoire et ces structures sont aujourd'hui connues pour réguler plusieurs processus biologiques. Cependant, les G-quadruplexes sont des structures très dynamiques et leur formation dépend de nombreux facteurs : il est donc nécessaire de développer des outils adaptés pour l'étude de ces structures dans un contexte cellulaire.

Dans ce contexte, une partie du travail de doctorat a été consacrée au développement d'une nouvelle méthode pour la localisation précise de ligands des G-quadruplexes in cellulo. Cette méthode combine les avantages de deux méthodes existantes : l'immunodétection et l'imagerie de ligands fluorescents. Cette méthodologie appelée G4-GIS. pour G4-ligand Guided Immunofluorescence Staining, est basée sur le développement de ligands fonctionnalisés par un haptène. Pour ce travail de thèse, le ligand PDC a été choisi pour sa forte affinité ainsi que sa bonne sélectivité envers les G-quadruplexes. L'haptène choisi ici est la 5-BrdU : cette molécule possède l'avantage d'avoir un faible poids moléculaire et d'être exogène aux cellules. Plusieurs dérivés du ligand PDC fonctionnalisés par différents bras espaceurs ainsi que par une fonction alcyne ou azoture ont été synthétisés avec succès. En parallèle, la 5-BrdU a aussi été fonctionnalisée par un alcyne ou un azoture dans le but de réaliser une réaction de CuAAC entre les dérivés PDC et les dérivés 5-BrdU (Figure 138).



Figure 138: Structures des dérivés PDC-Alk/N3 et des dérivés 5-BrdU-Alk/N3.

De cette manière, plusieurs dérivés PDC-BrdU ont été synthétisés à l'aide de cette réaction. La réaction de Huisgen catalysée au cuivre offre la possibilité d'être réalisée directement *in cellulo*. De ce fait, les différents ligands avant (PDC-Alk/N3) et après fonctionnalisation par la 5-BrdU (PDC-BrdU)

ont été testés pour leur capacité à interagir avec les G-quadruplexes. Deux méthodes biophysiques ont été utilisées pour cela : le FRET-melting et le G4-FID. Ces deux méthodes ont permis de montrer que la modification du PDC par différents bras espaceurs fonctionnalisés par un alcyne ou un azoture ne modifiait pas la capacité de ce dernier à interagir avec les G-quadruplexes. L'ajout de la 5-BrdU par réaction de click catalysée au cuivre sur les dérivés du PDC a montré un certain impact sur la capacité des ligands à interagir avec les G-quadruplexes. Cet impact est relativement modéré, mais démontre la nécessité de réaliser la fonctionnalisation du PDC par la 5-BrdU directement in situ. De plus, ces études biophysiques ont permis de montrer une bonne sélectivité pour les ADN et ARN G4s comparée à une séquence d'ADN duplex et à l'ARNt. La réaction de Huisgen catalysée au cuivre a ensuite été étudiée en l'absence ou en présence d'ADN et d'ARN G-quadruplex. Ces études ont permis de montrer que la présence de structures G4s permettait d'améliorer la cinétique de la réaction. Finalement, la localisation des ligands in cellulo a été étudiée par un protocole d'immunofluorescence indirecte (Figure 139). L'immunodétection des composés PDC-BrdU après 16 h d'incubation a permis de montrer une accumulation majoritairement cytoplasmique de ces derniers. L'immunodétection des composés PDC-Alk/N3 a ensuite été réalisée. Pour cela, les différents ligands ont été incubés in cellulo pendant 16 h avant d'être soumis à une réaction de click avec le partenaire 5-BrdU après fixation des cellules. Le protocole d'immunodétection a ensuite été réalisé dans les mêmes conditions. Cette étude a permis de montrer une accumulation cytoplasmique des ligands tout en augmentant généralement le signal de fluorescence comparé à l'immunodétection des ligands PDC-BrdU. Le composé PDC-4,3-Alk se localise notamment dans une région périnucléaire en formant des foci intenses de fluorescence. La digestion de l'ARN par traitement avec la RNase A a permis de diminuer drastiquement le signal de fluorescence dans le cytoplasme. Cette observation suggère que les différents dérivés du PDC ciblent majoritairement l'ARN in cellulo. Finalement, le ligand PDC-4,3-Alk a été couplé in situ avec le fluorophore Cy5-N3 dans le but de comparer la méthode d'immunodétection avec la méthode d'imagerie de fluorescence directe. Cette expérience a démontré que la localisation du ligand ne dépendait pas du groupement attaché à ce dernier et a permis de montrer une meilleure résolution d'image avec l'immunodétection, due notamment à l'amplification de fluorescence propre à cette technique. Finalement, des expériences de colocalisation entre BG4 et les anticorps anti-BrdU ont été réalisées dans le but de démontrer que les ligands interagissaient avec les G-quadruplexes in cellulo.



Figure 139: Schéma du principe de la méthodologie G4-GIS.

Afin de vérifier que les ligands sont capables d'interagir avec les G-quadruplexes *in cellulo*, plusieurs méthodes peuvent être envisagées. Une méthode pertinente serait de diminuer ou, au contraire, d'exacerber la production d'hélicases interagissant avec les G-quadruplexes. Si la quantité de fluorescence est modulée en fonction de la production d'hélicase, cela constituerait une preuve supplémentaire de l'interaction des ligands avec les G-quadruplexes *in cellulo*. Finalement, cette méthodologie permet d'augmenter le nombre d'outils moléculaires disponibles pour l'étude des G-quadruplexes et pourrait notamment être utilisée en combinant d'autres ligands que le PDC avec les dérivés de la 5-BrdU. Plus généralement, cette méthodologie pourrait être utilisée dans le but de localiser avec précision la répartition de n'importe quelle petite molécule *in cellulo*.

La deuxième partie du travail doctoral a été focalisée sur le développement de nouvelles molécules permettant la localisation exacte des sites de liaison des ligands des G-quadruplexes au niveau du génome. Seules quelques méthodes employant des petites molécules sélectives des G-quadruplexes ont été développées dans le but de déterminer avec précision la localisation des séquences adoptant des G-quadruplexes au sein du génome. Ces méthodes impliquent l'utilisation de ligands modifiés par un groupement biotine et reposent sur l'extraction supportée de fragments d'ADN génomiques par ajout de billes magnétiques recouvertes de streptavidine. C'est une méthodologie performante qui permet d'identifier les séquences pouvant former des structures G-quadruplexes à l'aide de petites molécules. Cependant, cette méthodologie repose sur l'interaction non covalente entre le ligand et l'ADN et ne peut empêcher la relocalisation de ces derniers pendant les différentes étapes de l'extraction supportée. Pour pallier ce problème, le marquage par photoaffinité (PAL) représente une approche élégante pour l'identification des cibles de petites molécules. Dans ce contexte, des dérivés du ligand PDC fonctionnalisés par un alcyne ainsi qu'un groupement photoactivable ont été imaginés et synthétisés. La présence de la fonction alcyne permettra, par la suite, de pouvoir réaliser une réaction de Huisgen catalysée au cuivre avec un haptène dans le but de pouvoir réaliser le

protocole d'extraction supportée. Les ligands synthétisés sont ainsi fonctionnalisés par des bras espaceurs de tailles et de polarités différentes ainsi que par une fonction diazirine, azoture ou benzophénone (Figure 140).



Figure 140: Structures des dérivés PDC-PhotoAlk linéaires et non linéaires.

Ces différents ligands ont ensuite été testés pour leur capacité à interagir avec les structures Gquadruplexes. Deux techniques ont été employées pour cela : le FRET-melting et le FID. Ces deux analyses montrent que les différentes fonctionnalisations réalisées sur le PDC ont un impact négatif plus ou moins élevé sur l'interaction avec les G-quadruplexes. Généralement, les ligands linéaires possèdent des affinités légèrement supérieures avec les différents G-quadruplexes par rapport aux ligands non linéaires. Tous les ligands testés par FRET-melting sont cependant très sélectifs pour les séquences G-quadruplexes comparées à une séquence d'ADN duplex. Les analyses par FRET-melting et FID doivent cependant être réalisées sur d'autres séquences dans le but de confirmer les différentes observations. Les ligands ont ensuite été testés pour leur capacité à générer des adduits covalents avec l'ADN G-quadruplex. Pour cela, trois séquences formant des structures Gquadruplexes ont été irradiées en présence des différents ligands. Les analyses par gels d'électrophorèse ont montré que la capacité d'alkylation des ligands était directement liée à la nature et à la position du groupement photoactivable.

Actuellement, trois autres dérivés non linéaires fonctionnalisés par un bras espaceur diéthylène glycol sont en cours de synthèse. Cette modulation permettra de vérifier l'impact de la longueur du bras espaceur sur la réaction de Huisgen catalysée au cuivre ainsi que sur la capacité des ligands à générer des adduits covalents avec les structures G-quadruplexes. Par la suite, la réaction de Huisgen catalysée au cuivre sera étudiée en présence de structures G-quadruplexes après avoir généré un lien

194

covalent par photo-irradiation entre l'ADN et le ligand. Finalement, les molécules présentant les meilleurs rendements d'alkylation seront utilisées dans le but d'étudier la formation et l'accessibilité de structures G-quadruplexes *in cellulo* au sein du génome. Après une période d'incubation, les cellules seront irradiées et les ligands interagissant avec des structures G-quadruplexes pourront générer un lien covalent avec ces derniers (Figure 141A). Les ligands seront ensuite soumis à une réaction de CuAAC avec la 5-BrdU-N3 précédemment synthétisée. Enfin, des billes magnétiques recouvertes d'anticorps anti-BrdU seront utilisées dans le but de capturer les fragments d'ADN liés aux ligands et ces derniers seront soumis à un séquençage haut débit.

Dans le cas des ligands ne présentant qu'une faible capacité d'alkylation des G4s, il serait intéressant d'évaluer leur capacité à générer des liens covalents avec des protéines interagissant avec les G-quadruplexes. En effet, si la position du groupement photoactivable sur la molécule n'est pas optimale pour l'alkylation des G4s, elle pourrait être adaptée à l'alkylation de protéines interagissant avec les G4s. Dans ce cas, après une période d'incubation *in cellulo*, les ligands seraient irradiés et pourraient générer un lien covalent avec les protéines situées à proximité des G-quadruplexes. Comme pour l'extraction d'ADN, les ligands seraient soumis à une réaction de CuAAC avec la 5-BrdU-N3 et extraits du milieu par ajout de billes magnétiques. Les protéines ainsi isolées seraient finalement soumises à des analyses protéomiques dans le but de les caractériser (Figure 141B).



Figure 141: Schéma général d'extraction d'ADN G-quadruplex (A) ou de protéines liées aux G-quadruplexes (B) après formation d'un lien covalent avec le ligand adapté.

Une dernière partie du travail de doctorat a été consacrée à l'étude de l'impact de la régioisomérie du lien amide avec la quinoline sur des dérivés bisquinolines de la pyridine et de la phénantroline. Pour cela, deux quinolines fonctionnalisées en position 4 par un bras pyrrolidine ont été synthétisées avec succès. Ces dernières ont ensuite été couplées avec l'acide 2,6-dicarboxylique pour former les composés PDC/PDC-Pyro et avec l'acide 1,10-phenanthroline-2,9-dicarboxylique pour former les composés PhenDC3/PhenDC2-Pyro. En parallèle, l'isomère de position du PDC, le PDC2, a été synthétisé (Figure 142A). Les différents ligands ont été évalués par FRET-melting et FID, révélant une très faible capacité de ces ligands à interagir avec les structures G-quadruplexes. Un composé dérivé du PDC2-Pyro, le PDC2-Pyro-4NH2 précédemment synthétisé au laboratoire, a montré une meilleure capacité à stabiliser les G-quadruplexes que le ligand PDC2-Pyro. La très faible capacité d'interaction entre les différents ligands synthétisés et les G-quadruplexes est très déconcertante, car l'un des dérivés, le PDC2-Pyro, a déjà été décrit et induisait alors un effet de stabilisation très important sur les séquences F21T et FcMyc22T. Plusieurs expériences ont été réalisées dans le but de comprendre cette différence, mais aucune conclusion ne semble pouvoir être tirée pour le moment. Dans le but de compléter cette étude sur l'impact de la régioisomérie de fonctionnalisation des quinolines sur l'affinité des ligands avec les G-quadruplexes, la synthèse du ligand PDS3 pourrait être envisagée (Figure 142B).



Figure 142 : A) Structures des ligands synthétisés pour l'étude de l'impact de la régioisomérie des dérivés bisquinolines et bisquinoliniums sur l'affinité avec les G-quadruplexes. B) Structure de la PDS3.

Partie expérimentale

Synthesis and characterization

All commercially available chemicals were bought from Sigma-Aldrich, Acros, Fluorochem, Fisher Scientific, TCI and Alfa Aesar as reagent grade and were used without further purification. Anhydrous solvents (DMF, THF and DCM) were obtained from MBRAUN Solvent Purification System. Chelex® resin 100 sodium form (50 – 100 mesh particle) was purchased from Sigma-Aldrich. Melting points were determined on a Kofler system type WME from Wagner & Munz. Thin-layer chromatography and flash chromatography were performed using pre-coated silica gel plates (Merck 60 F₂₅₄) with visualization at 254 and 366 nm. Flash chromatography was performed with silica gel 60 (40-63 µm, Merk). Deuterated solvents DMSO- d_6 and MeOH- d_4 were purchased from Eurisotop. All ¹H NMR and APT ¹³C NMR spectra were recorded on a Bruker Advance 300 MHz spectrometer except ¹H NMR of PDC-4,2-BrdU which was recorded on a 500 MHz Bruker Avance III spectrometer, using deuterated solvents and TMS as internal standard. The spectra are reported in ppm and referenced to deuterated DMSO (2.50 ppm for ¹H, 39.52 ppm for ¹³C) or deuterated MeOH (4.87 ppm for ¹H). Data for ¹H NMR are reported as follows: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, quint = quintet, dd = doublet of doublets, ddd = doublet of doublet of doublets, dt = doublet of triplets, m = multiplet, bs = broad singlet), coupling constant (Hz), integration. All compounds showed > 95% purity, assessed by HPLC analysis using either MS or UV absorbance or diode array detectors. Synthetic compounds were analyzed by LC-MS experiments carried out on a HPLC Alliance 2695 unit (Waters, USA) equipped with a Phenomenex Luna Omega Polar C18 column (50x3 mm i.d, 3 µm) and coupled to a low resolution mass spectrometry (ESI-MS) micromass ZQ 2000 (Waters).

Gradient HPLC injection:

Method A = flow 0.6 mL/min; gradient elution: eluent A% H_2O 0.1% FA (formic acid), eluent B% MeCN 0.1% FA: 0 min (95/5), 7 min (5/95), 9 min (5/95), 9 min 10 s (95/5), 15 min (95/5).

Method B = flow 0.6 mL/min; gradient elution: eluent A% H_2O 0.1% FA (formic acid), eluent B% MeCN 0.1% FA: 0 min (95/5), 7 min (0/100), 10 min (95/5).

Analytical method to control final compound purity (Column HPLC: XBridge C18 Column (150x3.0 i.d., 3 μm)):

Method C = flow 0.4 mL/min; gradient elution: eluent A% H_2O 0.1% FA (formic acid), eluent B% MeCN 0.1% FA: 0 min (95/5), 2 min (95/5), 8 min (0/100), 16 min (0/100), 16 min 30 s (95/5), 20 min (95/5) When needed, DMSO was evaporated using a SpeedVac Vacuum Concentrator Plus from Eppendorf. This work has benefited from the facilities and expertise of the Small Molecule Mass Spectrometry platform and of the High-field NMR platform of ICSN (Centre de Recherche de Gif - www.icsn.cnrs-gif.fr).

Compounds 3, 4, and 5 were synthesized with the same synthetic protocol:

Compound **2** (1.0 equiv) was mixed with the corresponding carboxylic acid (1.5 equiv), EDCI (1.5 equiv), and HOBt (0.15 equiv) in dry DMF. The resulting orange solution was stirred at room temperature under argon atmosphere for 16 h. The mixture was then concentrated under reduced pressure and the obtained solid was washed 3x with water and dried under vacuum. The crude product was purified by SiO2 chromatography.

Compound 3

4-((4-(pent-4-ynamido)butyl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Compound **2** (400 mg, 0.791 mmol), 4-pentynoic acid (116.4 mg, 1.187 mmol), EDCI (184.2 mg, 1.187 mmol), HOBt (16.0 mg, 0.119 mmol), DMF (8 mL).

Purification: DCM/MeOH from 100/0 to 94/6, leading to compound **3** as a white solid (355 mg, 77%). **Mp** = 250 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.28 (s, 2H), 9.36 (d, *J* = 2.4 Hz, 2H), 8.96 (d, *J* = 2.3 Hz, 2H), 8.04 (d, *J* = 8.6 Hz, 4H), 7.92 (t, *J* = 5.6 Hz, 1H), 7.72 (dt, *J* = 7.5, 1.4 Hz, 2H), 7.64 (dt, *J* = 7.5, 1.0 Hz, 2H), 7.59 – 7.48 (m, 3H), 3.25 (dt, *J* = 6.3, 5.5 Hz, 2H), 3.10 (dt, *J* = 6.4, 5.9 Hz, 2H), 2.75 (t, *J* = 2.5 Hz, 1H), 2.39 – 2.32 (m, 2H), 2.27 (t, *J* = 6.8 Hz, 2H), 1.72 – 1.47 (m, 4H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 170.14, 163.16, 156.29, 146.09, 144.66, 131.96, 128.69, 128.32, 127.90, 127.75, 127.18, 124.13, 83.81, 71.30, 41.74, 38.14, 34.29, 26.74, 25.57, 14.34; **LC/MS** (Method A): Retention time 6.20 min, m/z (ESI+) = 586.5 [M-H]⁺.

Compound 4

4-((4-(hex-5-ynamido)butyl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Compound **2** (400 mg, 0.791 mmol), 5-hexynoic acid (133.0 mg, 1.187 mmol), EDCI (184.2 mg, 1.187 mmol), HOBt (16.0 mg, 0.119 mmol), DMF (8 mL).

Purification: DCM/MeOH from 100/0 to 94/6, leading to compound **4** as a white solid (330 mg, 70%). **Mp** = 198 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.28 (s, 2H), 9.36 (d, *J* = 2.4 Hz, 2H), 8.97 (d, *J* = 2.2 Hz, 2H), 8.05 (d, *J* = 8.6 Hz, 4H), 7.87 (t, *J* = 5.3 Hz, 1H), 7.72 (dt, *J* = 7.6, 1.2 Hz, 2H), 7.64 (dt, *J* = 7.6, 1.0 Hz, 2H), 7.60 – 7.48 (m, 3H), 3.26 (dt, *J* = 6.0, 5.5 Hz, 2H), 3.11 (dt, *J* = 6.2, 5.8 Hz, 2H), 2.78 (t, *J* = 2.6 Hz, 1H), 2.22 – 2.11 (m, 4H), 1.66 (quint, *J* = 7.3 Hz, 2H), 1.64 – 1.49 (m, 4H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 171.35, 163.16, 156.28, 146.09, 144.66, 131.98, 128.70, 128.31, 127.90, 127.76, 127.17, 124.11, 84.12, 71.50, 41.76, 38.09, 34.22, 26.78, 25.61, 24.34, 17.44; **LC/MS** (Method A): Retention time 6.09 min, m/z (ESI+) = 600.5 [M-H]⁺.

Compound 5

4-((1-azido-11-oxo-3,6,9-trioxa-12-azahexadecan-16-yl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6dicarboxamide



Compound **2** (663.3 mg, 1.312 mmol), 11-azido-3,6,9-trioxaundecanoic acid (459 mg, 1.968 mmol), EDCI (305.5 mg, 1.968 mmol), HOBt (26.9 mg, 0.199 mmol), DMF (13mL). The reaction needs to be carried out by protecting the round bottom flask from the light.

Purification: DCM/EtOH from 100/0 to 90/10, leading to compound **5** as a white solid (577 mg, 61%). **Mp** = dec. > 48 °C; ¹H **NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.28 (s, 2H), 9.36 (d, *J* = 2.4 Hz, 2H), 8.96 (d, *J* = 2.2 Hz, 2H), 8.04 (d, *J* = 8.4 Hz, 4H), 7.78 – 7.68 (m, 3H), 7.63 (dt, *J* = 7.5, 1.0 Hz, 2H), 7.59 – 7.48 (m, 3H), 3.88 (s, 2H), 3.63 – 3.53 (m, 10H), 3.37 (t, *J* = 5.1 Hz, 2H), 3.29 – 3.14 (m, 4H), 1.66 – 1.53 (m, 4H); ¹³C **NMR** (75 MHz, DMSO-*d6*) δ (ppm): 169.14, 163.16, 156.27, 146.08, 144.66, 131.98, 128.70, 128.29, 127.88, 127.76, 127.16, 124.09, 70.25, 70.01, 69.79, 69.68, 69.60, 69.25, 49.98, 41.74, 37.74, 26.84, 25.56; **LC/MS** (Method A): Retention time 6.29 min, m/z (ESI+) = 721.5 [M-H]⁺.

Final copper-catalyzed alkyne-azide cycloaddition (CuAAC) precursors **PDC-4,2-Alk**, **PDC-4,3-Alk**, and **PDC-4,PEG-N3** were obtained following the same synthetic protocol

Compound **3**, **4**, or **5** (1.0 equiv) was dissolved in dry DMF warmed up at 40 °C. Iodomethane (235 equiv) was added and the resulting mixture was stirred at 40 °C for 16 h under argon atmosphere, until completion of the reaction. The reaction mixture was concentrated under reduced pressure and the resulting orange solid was suspended in EtOH, filtered, washed 2x with EtOH, and then 2x with Et_2O .

PDC-4,2-Alk (6)

3,3'-((4-((4-(pent-4-ynamido)butyl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



Compound **3** (300 mg, 0.512 mmol), DMF (3.8 mL), MeI (7.5 mL, 120.4 mmol), leading to **PDC-4,2-Alk** as a yellow solid (425 mg, 95%).

Mp = dec. > 218 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.71 (s, 2H), 10.11 (d, *J* = 1.8 Hz, 2H), 9.64 (d, *J* = 1.5 Hz, 2H), 8.58 – 8.52 (m, 4H), 8.24 (dt, *J* = 8.0, 1.0 Hz, 2H), 8.07 (t, *J* = 7.7 Hz, 2H), 7.94 (t, *J* = 5.5 Hz, 1H), 7.70 (t, *J* = 5.1 Hz, 1H), 7.59 (bs, 2H), 4.78 (s, 6H), 3.28 (dt, *J* = 6.0, 5.7 Hz, 2H), 3.13 (dt, *J* = 6.2, 5.9 Hz, 2H), 2.75 (t, *J* = 2.5 Hz, 1H), 2.41 – 2.31 (m, 2H), 2.31 – 2.21 (m, 2H), 1.70 – 1.48 (m, 4H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 170.15, 163.21, 156.55, 144.73, 135.68, 134.13, 133.92, 132.34, 130.34, 129.86, 129.22, 119.23, 99.51, 83.79, 71.29, 46.17, 41.75, 38.06, 34.27, 26.74, 25.33, 14.32; **LC/MS** (Method A): Retention time 5.07 min, m/z (ESI+) = 614.5; **HRMS** (ESI, *m/z*) calcd. for formula $C_{36}H_{37}N_7O_{31}$ [M-I]⁺ = 742.2003; found = 742.1992.

PDC-4,3-Alk (7)

3,3'-((4-((4-(hex-5-ynamido)butyl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



Compound **4** (300 mg, 0.500 mmol), DMF (3.8 mL), MeI (7.3 mL, 117.6 mmol), leading to **PDC-4,3-Alk** as a yellow solid (411 mg, 93%).

Mp = dec. > 218 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.71 (s, 2H), 10.11 (d, *J* = 1.8 Hz, 2H), 9.64 (d, *J* = 1.5 Hz, 2H), 8.55 (d, *J* = 9.3 Hz, 4H), 8.24 (dt, *J* = 8.0, 1.0 Hz, 2H), 8.08 (t, *J* = 7.0 Hz, 2H), 7.88 (t, *J* = 5.6 Hz, 1H), 7.71 (t, *J* = 5.1 Hz, 1H), 7.60 (bs, 2H), 4.78 (s, 6H), 3.26 (dt, *J* = 6.0, 5.6 Hz, 2H), 3.11 (dt, *J* = 6.3, 5.9 Hz, 2H), 2.78 (t, *J* = 2.6 Hz, 1H), 2.22 – 2.09 (m, 4H), 1.67 (quint, *J* = 7.3 Hz, 2H), 1.63 – 1.50 (m, 4H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 171.34, 163.16, 156.50, 144.69, 135.62, 134.02, 133.88, 132.33, 130.29, 129.83, 129.18, 119.18, 84.07, 71.50, 46.16, 41.74, 38.01, 34.18, 26.77, 25.47, 24.30, 17.39; **LC/MS** (Method A): Retention time 4.57 min, m/z (ESI+) = 628.5; **HRMS** (ESI, *m/z*) calcd. for formula $C_{37}H_{39}N_7O_3I$ [M-I]⁺ = 756.2159; found = 756.2161.

PDC-4, PEG-N3 (8)

3,3'-((4-((1-azido-11-oxo-3,6,9-trioxa-12-azahexadecan-16-yl)amino)pyridine-2,6-dicarbonyl)bis-(azanediyl))bis(1-methylquinolin-1-ium) iodide



Compound **5** (450 mg, 0.624 mmol), DMF (4.5 mL), MeI (9.1 mL, 146.7 mmol), leading to **PDC-4,PEG-** N_3 as a yellow solid (478 mg, 76%). The reaction needs to be carried out by protecting the round bottom flask from the light.

Mp = dec. > 185 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.71 (s, 2H), 10.13 (d, *J* = 0.9 Hz, 2H), 9.65 (d, *J* = 0.9 Hz, 2H), 8.55 (d, *J* = 8.7 Hz, 4H), 8.23 (t, *J* = 7.9 Hz, 2H), 8.08 (t, *J* = 7.5 Hz, 2H), 7.79 – 7.66 (m, 2H), 7.59 (bs, 2H), 4.78 (s, 6H), 3.89 (s, 2H), 3.63 – 3.53 (m, 10H), 3.39 (t, *J* = 4.9 Hz, 2H), 3.30

- 3.25 (m, 2H), 3.23 - 3.15 (m, 2H), 1.60 (m, 4H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 169.11, 163.19, 156.53, 144.72, 135.67, 134.11, 133.91, 132.37, 130.33, 129.84, 129.21, 119.21, 99.49, 70.22, 69.97, 69.76, 69.64, 69.55, 69.21, 49.95, 46.14, 41.72, 37.66, 26.80, 25.43; **LC/MS** (Method A): Retention time 4.79 min, m/z (ESI+) = 749.5; **HRMS** (ESI, *m/z*) calcd. for formula $C_{39}H_{46}N_{10}O_{6}I$ [M-I]⁺ = 877.2647; found = 877.2668.

Compound 9

4-((4-hydroxybutyl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Compound **1** (1.0 equiv, 600 mg, 1.322 mmol) mixed with 4-amino-1-butanol (18.7 equiv, 2.3 mL, 24.77 mmol) and TEA (2.7 equiv, 0.5 mL, 3.597 mmol) was heated at 90 °C under argon atmosphere for 16 h. After completion of the reaction, the crude product was directly purified by SiO_2 gel chromatography and eluted with a gradient of DCM/MeOH from 100/0 to 90/10, affording compound **9** as a white solid (441 mg, 66%).

Mp = 150 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.23 (s, 2H), 9.36 (d, *J* = 2.3 Hz, 2H), 8.94 (d, *J* = 1.9 Hz, 2H), 8.08 – 7.99 (m, 4H), 7.71 (t, *J* = 7.6 Hz, 2H), 7.63 (t, *J* = 7.4 Hz, 2H), 7.57 (bs, 2H), 7.44 (t, *J* = 5.0 Hz, 1H), 4.38 (t, *J* = 4.8 Hz, 1H), 3.48 (dt, *J* = 5.8, 5.2 Hz, 2H), 3.30 – 3.25 (m, 2H), 1.73 – 1.52 (m, 4H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 163.20, 156.33, 146.13, 144.68, 131.99, 128.72, 128.36, 127.93, 127.78, 127.22, 124.17, 60.44, 42.00, 29.95, 24.96; **LC/MS** (Method A): Retention time 5.48 min, m/z (ESI+) = 507.5 [M-H]⁺.

Compound 10

4-((2,6-bis(quinolin-3-ylcarbamoyl)pyridin-4-yl)amino)butyl methanesulfonate



Compound **9** (1.0 equiv, 40 mg, 0.079 mmol) was suspended in dry DCM (2.4 mL) at 0 °C under argon atmosphere and freshly distilled TEA (9.0 equiv, 0.099 mL, 0.712 mmol) was added under stirring. Reaction mixture was cooled down to 0°C and MsCl (4.5 equiv, 0.027 mL, 0.355 mmol) was added dropwise. The mixture was stirred at 0°C for 0.5h and then left to warm up to room temperature. The reaction was quenched by addition of ice water and a precipitate formed. The latter was filtered and the resulting white solid washed 3x with small portions of water and then once with Et₂O. The solid was dried under vacuum, yielding compound **10** as a white solid (31 mg, 67%). Once the ¹H NMR was recorded, the product was directly submitted to the next reaction.

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.36 (s, 2H), 9.42 (d, *J* = 2.3 Hz, 2H), 9.03 (d, *J* = 1.8 Hz, 2H), 8.10 - 8.03 (m, 4H), 7.75 (dt, *J* = 7.1, 1.0 Hz, 2H), 7.66 (t, *J* = 7.2 Hz, 2H), 7.62 - 7.52 (m, 3H), 4.28 (t, *J* = 6.1 Hz, 2H), 3.34 - 3.26 (m, 2H), 3.20 (s, 3H), 1.89 - 1.65 (m, 4H); **LC/MS** (Method A): Retention time 6.14 min, m/z (ESI+) = 585.5 [M-H]⁺.

Compound 11

4-((4-azidobutyl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Compound **10** (1 equiv, 1350 mg, 0.599 mmol) was directly submitted to nucleophilic substitution with NaN₃ (10 equiv, 389.2 mg, 5.99 mmol) in dry DMF (23 mL) at 100 °C for 1 h 30 min. The reaction must be protected from light. DMF was removed under reduced pressure and the resulting yellow solid was washed 3x with water and then once with Et_2O . The crude product was purified by SiO_2 gel chromatography using a gradient of DCM/EtOH from 100/0 to 90/10, yielding compound **11** as a pale yellow solid (332 mg, 77%).

Mp = 132 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.29 (s, 2H), 9.35 (d, *J* = 2.5 Hz, 2H), 8.96 (d, *J* = 2.3 Hz, 2H), 8.04 (d, *J* = 9.3 Hz, 4H), 7.72 (dt, *J* = 7.6, 1.4 Hz, 2H), 7.63 (dt, *J* = 7.8, 1.0 Hz, 2H), 7.60 – 7.50 (m, 3H), 3.43 (t, *J* = 6.3 Hz, 2H), 3.32 – 3.24 (m, 2H), 1.75 – 1.62 (m, 4H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 163.14, 156.30, 146.09, 144.65, 131.95, 128.68, 128.32, 127.89, 127.74, 127.18, 124.13, 50.39, 41.51, 25.91, 25.41; **LC/MS** (Method A): Retention time 6.88 min, m/z (ESI+) = 532.5 [M-H]⁺.

PDC-4,0-N3 (12)

3,3'-((4-((4-azidobutyl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



Compound **11** (1 equiv, 220 mg, 0.414 mmol) was dissolved in dry DMF (3.0 mL) at 40 °C under argon atmosphere and protected from light. Iodomethane (250 equiv, 6.4 mL, 103.5 mmol) was then added. The mixture was stirred at 40 °C for 16 h till completion of the reaction. The mixture was concentrated under reduced pressure and the resulting orange solid was suspended in EtOH, filtered, washed 2x with EtOH, and then 2x with Et₂O, leading to **PDC-4,0-N3** as a yellow solid (292 mg, 87%). **Mp** = dec. > 215 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.66 (s, 2H), 10.12 (d, *J* = 1.9 Hz, 2H), 9.64 (s, 2H), 8.60 – 8.48 (m, 4H), 8.24 (dt, *J* = 8.0, 1.0 Hz, 2H), 8.08 (t, *J* = 7.6 Hz, 2H), 7.65 (t, *J* = 5.4 Hz, 1H), 7.64 – 7.58 (m, 2H), 4.79 (s, 6H), 3.43 (t, *J* = 5.9 Hz, 2H), 3.35 – 3.28 (m, 2H), 1.75 – 1.65 (m, 4H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm) 163.21, 156.56, 144.78, 135.66, 134.12, 133.92, 132.38, 130.34, 129.87, 129.21, 119.22, 50.39, 46.15, 41.55, 25.90, 25.33; **LC/MS** (Method A): Retention time 5.17 min, m/z (ESI+) = 560.5; **HRMS** (ESI, *m/z*) calcd. for formula C₃₁H₃₁N₉O₂I [M-I]⁺ = 688.1645; found = 688.1622.

Compound 13

5-bromo-1-((2R,4S,5R)-5-(((tert-butyldimethylsilyl)oxy)methyl)-4-hydroxytetrahydrofuran-2-yl)pyrimidine-2,4(1H,3H)-dione



Commercially available 5-bromo-2'-deoxyuridine (**5-BrdU**, 1 equiv, 1.00 g, 3.256 mmol) was mixed with DMAP (0.13 equiv, 51.7 mg, 0.423 mmol) and imidazole (3.3 equiv, 731.5 mg, 10.75 mmol) in dry DMF (10 mL) under argon atmosphere. The mixture was cooled down to 0 °C and TBDMSCI (1.1 equiv, 539.9 mg, 3.582 mmol) dissolved in dry DMF (5 mL) was added dropwise. The reaction mixture was stirred at 0 °C for 30 min and then slowly warmed up to room temperature and stirred for 16 h. Afterwards, a saturated solution of NaHCO₃ was added and the aqueous phase was extracted 3x with

AcOEt. The combined organic layers were washed once with water, dried over MgSO₄, filtered, and concentrated under reduced pressure. The resulting crude solid was purified by SiO_2 gel chromatography using a gradient of DCM/EtOH from 100/0 to 95/5, yielding compound **13** as a white solid (1.081 g, 79%).

Mp = dec. > 72 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.85 (s, 1H), 8.01 (s, 1H), 6.10 (dd, *J* = 7.3, 6.3 Hz, 1H), 5.29 (d, *J* = 4.1 Hz, 1H), 4.22 – 4.15 (m, 1H), 3.87 (dd, *J* = 5.3, 2.7 Hz, 1H), 3.82 (dd, *J* = 11.5, 2.7 Hz, 1H), 3.73 (dd, *J* = 11.5, 3.3 Hz, 1H), 2.21 – 2.02 (m, 2H), 0.89 (s, 9H), 0.10 (d, *J* = 2.5 Hz, 6H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 159.13, 149.67, 139.39, 95.93, 87.31, 85.10, 70.50, 63.19, 25.94, 18.13, -5.35; **LC/MS** (Method A): Retention time 6.58 min, m/z (ESI+) = 423.2 [M-H]⁺. 159.13, 149.67, 139.39, 95.93, 87.31, 85.10, 70.50, 63.19, 25.94, 18.13, -5.35

Compound 14

5-bromo-1-((2R,4S,5R)-5-(((tert-butyldimethylsilyl)oxy)methyl)-4-(prop-2-yn-1-yloxy)tetrahydrofuran-2-yl)pyrimidine-2,4(1H,3H)-dione



Compound **13** (1 equiv, 500 mg, 1.187 mmol) was dissolved in anhydrous THF (22.5 mL) at room temperature and under argon atmosphere. NaH (2.5 equiv, 71.2 mg, 2.967 mmol) was then added portion wise. The mixture was stirred 30 min and propargyl bromide (2.5 equiv, 0.264 mL, 2.967 mmol, 80 wt % in toluene) was added dropwise. The reaction mixture was stirred for additional 16 h. EtOH was added to quench the reaction and the solvent was removed under reduced pressure. The resulting crude product was purified by SiO₂ gel chromatography using a gradient of DCM/EtOH from 100/0 to 95/5, affording compound **14** as a white solid (450 mg, 83%).

Mp = dec. > 45 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.87 (s, 1H), 8.02 (s, 1H), 6.04 (dd, *J* = 8.2, 5.8 Hz, 1H), 4.26 – 4.21 (m, 3H), 4.06 (dd, *J* = 5.1, 3.1 Hz, 1H), 3.83 (dd, *J* = 11.4, 3.4 Hz, 1H), 3.75 (dd, *J* = 11.4, 3.3 Hz, 1H), 3.48 (t, *J* = 2.3 Hz, 1H), 2.35 (ddd, *J* = 13.6, 5.7, 1.6 Hz, 1H), 2.12 (ddd, *J* = 13.9, 8.2, 5.9 Hz, 1H), 0.90 (s, 9H), 0.12 (d, *J* = 2.0 Hz, 6H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 159.07, 149.69, 139.27, 96.14, 85.01, 84.40, 80.01, 78.22, 77.48, 63.23, 55.84, 36.88, 25.90, 18.07, -5.40; **LC/MS** (Method B): Retention time 7.10 min, m/z (ESI+) = 461.2 [M-H]⁺.

5-BrdU-Alk (15)

5-bromo-1-((2R,4S,5R)-5-(hydroxymethyl)-4-(prop-2-yn-1-yloxy)tetrahydrofuran-2-yl)pyrimidine-2,4(1H,3H)-dione



Compound **14** (1 equiv, 400 mg, 0.871 mmol) was solubilized in THF anhydrous (8 mL) and TBAF (1.2 equiv, 1.05 mL, 1.045 mmol, 1M in THF) was slowly added at room temperature. The resulting mixture was stirred under argon atmosphere for 16 h. The solvent was then removed under vacuum and the crude was purified by SiO₂ gel chromatography, eluted with a gradient of DCM/EtOH from 100/0 to 90/10, affording **5-BrdU-Alk** as a white solid (224.5 mg, 75%).

Mp = 157 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.84 (s, 1H), 8.37 (s, 1H), 6.06 (dd, *J* = 7.6, 6.1 Hz, 1H), 5.26 (t, *J* = 4.8 Hz, 1H), 4.28 – 4.22 (m, 1H), 4.21 (d, *J* = 2.3 Hz, 2H), 3.99 (dd, *J* = 2.9, 2.6 Hz, 1H), 3.70 – 3.54 (m, 2H), 3.49 (t, *J* = 2.4 Hz, 1H), 2.32 (ddd, *J* = 13.9, 6.1, 2.6 Hz, 1H), 2.19 (ddd, *J* = 13.7, 7.6, 6.0 Hz, 1H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 159.13, 149.75, 140.13, 95.94, 84.85, 84.80, 80.23, 78.35, 77.35, 61.15, 55.96, 36.85; **LC/MS** (Method A): Retention time 4.72 min, m/z (ESI+) = 347.1 [M-H]⁺; **HRMS** (ESI, *m/z*) calcd. for formula $C_{12}H_{14}N_2O_5Br$ [M-H]⁺ = 345.0086; found = 345.0096.

Compound 16

5-bromo-1-((2R,4S,5R)-5-(((tert-butyldimethylsilyl)oxy)methyl)-4-((methylthio)methoxy)tetrahydrofuran-2-yl)pyrimidine-2,4(1H,3H)-dione



To compound **13** (1 equiv, 1.00 g, 2.373 mmol) was added a mixture of DMSO (44.6 equiv, 7.52 mL, 105.8 mmol), acetic acid (11.6 equiv, 1.58 mL, 27.6 mmol), and acetic anhydride (23.1 equiv, 5.14 mL, 54.75 mmol). The resulting mixture was stirred at room temperature for 48 h. To quench the reaction, the mixture was poured into a cold saturated solution of NaHCO₃. The pH was adjusted to 10 by adding portions of solid NaHCO₃ and the resulting aqueous solution was extracted 3x with

DCM. The combined organic layers were then dried over MgSO₄, filtered, and the solvent removed under reduced pressure. The resulting crude product was purified by SiO_2 gel chromatography eluted with a gradient of DCM/EtOH from 100/0 to 99/1, yielding compound **16** as a white solid (987 mg, 86%).

Mp = dec. > 45 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.87 (s, 1H), 8.02 (s, 1H), 6.06 (dd, *J* = 8.1, 5.9 Hz, 1H), 4.71 (s, 2H), 4.41 – 4.35 (m, 1H), 4.04 (dd, *J* = 5.3, 3.1 Hz, 1H), 3.83 (dd, *J* = 11.3, 3.7 Hz, 1H), 3.76 (dd, *J* = 11.4, 3.5 Hz, 1H), 2.32 (ddd, *J* = 13.8, 5.9, 2.1 Hz, 1H), 2.21 – 2.12 (m, 1H), 2.09 (s, 3H), 0.90 (s, 9H), 0.11 (d, *J* = 1.8 Hz, 6H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 159.07, 149.67, 139.31, 96.16, 85.07, 84.54, 75.93, 72.62, 63.15, 36.95, 25.90, 18.06, 13.30, -5.40² **LC/MS** (Method A): Retention time 7.97 min, m/z (ESI+) = 483.2 [M-H]⁺.

Compound 17

1-((2R,4S,5R)-4-(azidomethoxy)-5-(((tert-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-5bromopyrimidine-2,4(1H,3H)-dione



To a stirred solution of compound **16** (1 equiv, 380 mg, 0.789 mmol) in dry DCM (2.4 mL) cooled at - 10 °C was added dropwise a solution of Br_2 (1.2 equiv, 48.7 µL, 0.947 mmol) in dry DCM (1.2 mL). The mixture was stirred at -10 °C for 30 min sheltered from light. A suspension of NaN₃ (6 equiv, 308 mg, 4.735 mmol) in dry DMF (15 mL) was then added. The reaction mixture was slowly warmed up to room temperature and kept under stirring for 16 h. The reaction was quenched by addition of a saturated solution of NaHCO₃ and the resulting solution was extracted 3x with DCM. The combined organic layers were then dried over MgSO₄, filtered, and the solvent was removed under reduced pressure. The resulting yellow oil was then purified by SiO₂ gel chromatography eluted with a gradient of DCM/EtOH from 100/0 to 99/1, yielding compound **17** as a colorless oil (206 mg, 55%).

¹H NMR (300 MHz, DMSO-*d6*) δ (ppm): 11.89 (s, 1H), 8.01 (s, 1H), 6.07 (dd, *J* = 7.4, 6.2 Hz, 1H), 4.88 (d, *J* = 9.0 Hz, 1H), 4.85 (d, *J* = 9.0 Hz, 1H), 4.34 – 4.29 (m, 1H), 4.05 (dd, *J* = 5.7, 3.0 Hz, 1H), 3.84 (dd, *J* = 11.4, 3.5 Hz, 1H), 3.77 (dd, *J* = 11.4, 3.5 Hz, 1H), 2.37 (ddd, *J* = 13.7, 5.9, 2.5 Hz, 1H), 2.30 – 2.18 (m, 1H), 0.90 (s, 9H), 0.11 (d, *J* = 1.4 Hz, 6H); ¹³C NMR (75 MHz, DMSO-*d6*) δ (ppm): 159.14, 149.70, 139.39, 96.17, 84.97, 84.68, 80.99, 78.07, 62.96, 37.30, 25.91, 18.09, -5.39; LC/MS (Method B): Retention time 7.25 min, m/z (ESI+) = 478.3 [M-H]⁺.

5-BrdU-N3 (18)

1-((2R,4S,5R)-4-(azidomethoxy)-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-bromopyrimidine-2,4(1H,3H)-dione



Compound **17** (1 equiv, 160 mg, 0.336 mmol) was solubilized in anhydrous THF (3.2 mL) and TBAF (1.2 equiv, 0.403 mL, 0.403 mmol, 1M in THF) was slowly added at room temperature. The resulting mixture was stirred under argon atmosphere protected from the light for 16 h. The solvent was then removed under vacuum and the crude product was purified by SiO₂ gel chromatography, eluted with a gradient of DCM/EtOH from 100/0 to 90/10, affording **5-BrdU-N3** as a white solid (75 mg, 62%). **Mp** = 156 °C; ¹H **NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.75 (bs, 1H), 8.36 (s, 1H), 6.08 (t, *J* = 6.6 Hz,

1H), 5.28 (t, J = 4.6 Hz, 1H), 4.85 (s, 2H), 4.39 – 4.33 (m, 1H), 3.99 (dd, J = 6.3, 3.0 Hz, 1H), 3.70 – 3.57 (m, 2H), 2.35 – 2.26 (m, 2H); ¹³**C** NMR (75 MHz, DMSO-*d6*) δ (ppm) 159.20, 149.78, 140.18, 95.93, 85.06, 84.77, 81.11, 77.93, 60.84, 37.46; **LC/MS** (Method A): Retention time 4.78 min, m/z (ESI+) = 364.0 [M-H]⁺; **HRMS** (ESI, *m/z*) calcd. for formula C₁₀H₁₃N₅O₅Br [M-H]⁺ = 362.0100; found = 362.0098.

All **PDC-BrdU** compounds were obtained using the following general protocol:

PDC CuAAC precursor (1.0 equiv) was mixed with the appropriate 5-BrdU derivative (1.2 equiv) in DMF/distilled water 6/2 mL (3/1 (v/v) after addition of every reagents). The mixture was protected from light, stirred and purged with argon for 15 min. A solution of sodium ascorbate in water (1.2 equiv, 0.1 M) was then added followed by a solution of $CuSO_4.5H_2O$ in water (0.3 equiv, 0.1 M). The solution was purged for additional 15 min with argon and then stirred under argon atmosphere until completion of the reaction. Copper(II) ions were removed from the reaction mixture by adding Chelex[®] resin 100 Na-form, previously conditioned to pH = 7 with triethylammonium acetate (50 mM). The suspension was gently shaken for 30 min and then filtered through a cotton pad. The Chelex resin was washed twice with DMF and the resulting mixture was concentrated under reduced pressure using a SpeedVac Vacuum Concentrator to give the crude product as a brown sticky oil. Crude product was precipitated in DCM, filtered, washed several times with DCM and then with water. The brown solid was then dried under vacuum yielding the expected compound.

PDC-4,2-BrdU (19)

3,3'-((4-((4-(3-(1-((((2R,3S,5R)-5-(5-bromo-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)-2-(hydroxymethyl)tetrahydrofuran-3-yl)oxy)methyl)-1H-1,2,3-triazol-4-yl)propanamido)butyl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



PDC-4,2-Alk (1 equiv, 70 mg, 0.081 mmol), **5-BrdU-N3** (1.2 equiv, 35.0 mg, 0.086 mmol), DMF (6 mL), water (0.75 mL), sodium ascorbate (1.2 equiv, 0.97 mL, 0.97 mmol, 0.1 M), $CuSO_{4}.5H_{2}O$ (0.3 equiv, 241 µL, 0.024 mmol, 0.1 M). The reaction was complete within 16 h. Work up steps led to final compound **PDC-4,2-BrdU** as a brown solid (60 mg, 61%).

Mp = dec. > 190 °C; ¹**H** NMR (500 MHz, DMSO-*d6*) δ (ppm): 11.69 (s, 2H), 10.17 (bs, 2H), 9.70 (bs, 2H), 8.43 (bs, 4H), 8.27 (s, 1H), 8.15 (bs, 2H), 8.00 (bs, 2H), 7.97 (s, 1H), 7.81 (bs, 1H), 7.73 (bs, 1H), 7.65 – 7.40 (m, 2H), 6.03 (t, J = 6.6 Hz, 1H), 5.75 (s, 2H), 5.13 (bs, 1H), 4.70 (bs, 6H), 4.32 (bs, 1H), 3.87 (bs, 1H), 3.58 (dd, J = 11.6, 2,8 Hz, 1H), 3.51 (dd, J = 11.6, 2.3 Hz, 1H), 2.91 (t, J = 7.2 Hz, 2H), 2.16 – 2.11 (m, 2H), 1.67 – 1.51 (m, 4H); **LC/MS** (Method B): Retention time 4.56 min, m/z (ESI+) = 977.6; **HRMS** (ESI, m/z) calcd. for formula C₄₆H₄₉N₁₂O₈BrI [M-I]⁺ = 1103,2024; found = 1103.2054.

PDC-4,3-BrdU (20)

3,3'-((4-((4-(1-((((2R,3S,5R)-5-(5-bromo-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)-2-(hydroxymethyl)tetrahydrofuran-3-yl)oxy)methyl)-1H-1,2,3-triazol-4-yl)butanamido)butyl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



PDC-4,3-Alk (1 equiv, 80 mg, 0.091 mmol), **5-BrdU-N3** (1.2 equiv, 39.4 mg, 0.108 mmol, DMF (6 mL), water (0.65 mL), sodium ascorbate (1.2 equiv, 1.08 mL, 0.108 mmol, 0.1 M), CuSO₄.5H₂O (0.3 equiv,

271 μL, 0.027 mmol, 0.1 M). The reaction was complete within 16 h. Work up steps led to final compound **PDC-4,3-BrdU** as a brown solid (77 mg, 68%).

Mp = dec. > 182 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.70 (s, 2H), 10.14 (bs, 2H), 9.67 (bs, 2H), 8.47 (bs, 4H), 8.28 (s, 1H), 8.24 – 8.04 (m, 4H), 8.02 (s, 1H), 7.74 (t, J = 5.3 Hz, 1H), 7.70 – 7.40 (m, 3H), 6.03 (t, J = 6.7 Hz, 1H), 5.75 (s, 2H), 5.14 (bs, 1H), 4.73 (bs, 6H), 4.32 (bs, 1H), 3.87 (bs, 1H), 3.60 – 3.50 (m, 2H), 2.65 (t, J = 7.4 Hz, 2H), 2.19 – 2.10 (m, 4H), 1.86 (quint, J = 7.3 Hz, 2H), 1.70 – 1.50 (m, 4H); **LC/MS** (Method B): Retention time 4.66 min, m/z (ESI+) = 991.7; **HRMS** (ESI, *m/z*) calcd. for formula C₄₇H₅₁N₁₂O₈BrI [M-I]⁺ = 1117.2181; found = 1117.2233.

PDC-4,0-BrdU (21)

3,3'-((4-((4-(((((2R,3S,5R)-5-(5-bromo-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)-2-(hydroxymethyl)tetrahydrofuran-3-yl)oxy)methyl)-1H-1,2,3-triazol-1-yl)butyl)amino)pyridine-2,6-dicarbonyl)bis-(azanediyl))bis(1-methylquinolin-1-ium) iodide



PDC-4,0-N3 (1 equiv, 70 mg, 0.085 mmol), **5-BrdU-Alk** (1.2 equiv, 35.6 mg, 0.102 mmol), DMF (6 mL), water (0.75 mL), sodium ascorbate (1.2 equiv, 1.02 mL, 0.10 mmol, 0.1 M), CuSO₄.5H₂O (0.3 equiv, 257.5 μ L, 0.026 mmol, 0.1 M). The reaction was complete within 4 h. Work up steps led to final compound **PDC-4,0-BrdU** as a brown solid (78 mg, 78%).

Mp = dec. > 200 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.70 (s, 2H), 10.21 (bs, 2H), 9.69 (bs, 2H), 8.40 (bs, 4H), 8.32 (s, 1H), 8.24 – 8.10 (m, 3H), 8.00 (bs, 2H), 7.70 – 7.30 (m, 3H), 6.06 (t, *J* = 6.7 Hz, 1H), 5.14 (bs, 1H), 4.69 (bs, 6H), 4.60 (s, 2H), 4.47 (t, *J* = 6.7 Hz, 2H), 4.24 (bs, 1H), 4.01 (bs, 1H), 3.68 – 3.56 (m, 2H), 2.16 – 2.25 (m, 2H), 2.08 – 1.95 (2H), 1.72 – 1.54 (m, 2H); **LC/MS** (Method B): Retention time 4.61 min, m/z (ESI+) = 906.4; **HRMS** (ESI, *m/z*) calcd. for formula C₄₃H₄₄N₁₁O₇BrI [M-I]⁺ = 1032.1653; found = 1032.1667.

PDC-4, PEG-BrdU (22)

3,3'-((4-((1-(4-((((2R,3S,5R)-5-(5-bromo-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)-2-(hydroxymethyl)tetrahydrofuran-3-yl)oxy)methyl)-1H-1,2,3-triazol-1-yl)-11-oxo-3,6,9-trioxa-12-azahexadecan-16yl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



PDC-4,PEG-N3 (1 equiv, 70 mg, 0.070 mmol), **5-BrdU-Alk** (1.2 equiv, 28.9 mg, 0.084 mmol), DMF (6 mL), water (1.0 mL), sodium ascorbate (1.2 equiv, 0.836 mL, 0.836 mmol, 0.1 M), CuSO₄.5H₂O (0.3 equiv, 209 μ L, 0.021 mmol, 0.1 M). The reaction was complete within 6 h. Work up steps led to final compound **PDC-4,PEG-BrdU** as a brown solid (45 mg, 48%).

Mp = dec. > 148 °C; ¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.70 (s, 2H), 10.11 (bs, 2H), 9.64 (bs, 2H), 8.51 (bs, 4H), 8.32 (s, 1H), 8.22 (bs, 2H), 8.10 – 8.00 (m, 3H), 7.70 – 7.50 (m, 3H), 6.06 (t, *J* = 6.6 Hz, 1H), 5.13 (bs, 1H), 4.76 (bs, 6H), 4.58 (bs, 2H), 4.52 (t, *J* = 5.1 Hz, 2H), 4.22 (bs, 1H), 4.00 (bs, 1H), 3.89 (bs, 2H), 3.83 (t, *J* = 4.90 Hz, 2H), 3.65 – 3.50 (m, 10H), 2.25 – 2.15 (m, 2H), 1.70 – 1.52 (m, 4H); **LC/MS** (Method B): Retention time 4.64 min, m/z (ESI+) = 1095.6; **HRMS** (ESI, *m/z*) calcd. for formula $C_{51}H_{59}N_{12}O_{11}BrI [M-I]^+ = 1221.2654$; found = 1221.2644.

Compound 23

pent-4-ynal

To the mixture of pyridinium chlorochromate (1.5 equiv, 6.995 g, 32.5 mmol) and sodium acetate (0.15 equiv, 266mg, 3.25 mmol) in dried DCM was added 4-pentyn-1-ol (1 equiv, 2 mL, 21.6 mmol) dropwise at 0°C under argon atmosphere. The reaction is stirred for 2 h at room temperature and the formation of the aldehyde compound is checked by NMR (δ H aldehyde = 9.75 ppm in CDCl3). 15 mL of Et2O is directly added into the reaction mixture and crude mixture is loaded onto a silica gel chromatography and eluted with Et2O. Gathered fractions were then concentrated carefully under vacuum until \approx 70% Et2O was removed. This product was directly used in the next step.
¹**H NMR** (300 MHz, CDCl3) δ (ppm): 9.75 (s, 1H), 2.65 (t, *J* = 7.0 Hz, 2H), 2.53 – 2.41 (m, 4H), 2.25 (bs, 1H)

Compound 24

ethyl 4-oxooct-7-ynoate

The freshly synthesized **compound 23** (1 equiv, 21.6 mmol) was mixed with ethyl acrylate (2 equiv, 4.7 mL, 43.2 mmol) in 25 mL of 1,4-Dioxane. The mixture was added dropwise over 4 h to a suspension of 3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride (0.2 equiv, 1.106 g, 4.33 mmol), TEA (1 equiv, 3.0 mL, 21.6 mmol) and ethyl acrylate (2 equiv, 4.7 mL, 43.2 mmol) in 35 mL of 1,4-Dioxane at 80 °C under argon atmosphere for 54 h. Solvents were then removed under reduced pressure and the crude product was solubilized in DCM, washed with aqueous 10% sulfuric acid, a saturated solution of NaHCO3 and finally with Brine. Organic layer was dried over Na2SO4, filtered and evaporated under vacuum resulting in a red oil. The crude oil was loaded onto a silica gel chromatography and eluted by Cy/AcOEt (100/0 to 80/20) to give expected product **24** as an orange oil (766 mg, 19%).¹H NMR (300 MHz, CDCl₃) δ (ppm) 4.12 (q, *J* = 7.1 Hz, 2H), 2.78 – 2.68 (m, 4H), 2.60 (t, *J* = 6.2 Hz, 2H), 2.46 (dt, *J* = 7.2 / 2.6 Hz, 2H), 1.94 (t, *J* = 2.6 Hz, 1H), 1.25 (t, *J* = 7.1 Hz, 3H); ¹³C NMR (75 MHz, CDCl₃) δ (ppm): δ 206.75, 172.81, 68.88, 60.84, 41.48, 37.22, 28.06, 14.31, 13.05

Compound 25

4-oxooct-7-ynoic acid

To a stirred solution of compound **24** (1 equiv, 658 mg, 3.61 mmol) in 25 mL of MeOH and water (5 equiv, 0.33 mL, 18.4 mmol) was added LiOH.H2O (5 equiv, 773 mg, 18.4 mmol) at room temperature. The mixture was stirred at room temperature for 15 h. Volatiles were removed under vacuum and crude solid was solubilized in some water. pH was acidified using 6M HCl and aqueous layer was extracted multiple times with DCM. Gathered organic layers were dried with Na2SO4, filtered and evaporated under vacuum affording expected product as a golden solid (523 mg, 94%).

¹H NMR (300 MHz, CDCl₃) δ (ppm): 2.78 – 2.61 (m, 6H), 2.47 (dt, *J* = 7.2 / 2.6 Hz, 2H), 1.95 (t, J = 2.6 Hz, 1H);
 ¹³C NMR (75 MHz, CDCl₃) δ (ppm): 206.50, 178.68, 82.97, 68.98, 41.39, 36.93, 27.80, 13.05.

4-N₂-octynoic acid (26)

3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)propanoic acid

Anhydrous ammonia was produced by reaction of potassium hydroxide and ammonium chloride. Vapors were condensed in a sealed tube previously cooled down at -70 °C using an acetone-nitrogen bath. 3 mL were produced this way and compound **25** (1 equiv, 250 mg, 1.62 mmol) was slowly added. The tube was sealed, allowed to warm to room temperature and stirred for 5 h. After this time, the mixture was cooled down below -50 °C and hydroxylamine-O-sulfonic acid (1.2 equiv, 220 mg, 1.94 mmol) was added over a period of 15 min. The tube was then sealed again and the mixture was stirred at room temperature for 1h 30. The mixture was next cooled down below -50 °C and allowed to slowly warm to room temperature in order to remove ammonia overnight. The slurry mixture was then filtrated and washed several times with small portions of MeOH. Volatiles were removed under vacuum and the resulting solid was dissolved in 3 mL of MeOH. The mixture was then treated with DIPEA (1.3 equiv, 0.34 mL, 2.03 mmol), stirred for 10 min and lodine was added portion wise until a brown color persist more than 20 min, indicating total oxidation of the diaziridine function.

HCl 1N was next added followed by DCM. Layers were separated and aqueous layer was extracted once with DCM. Gathered organic layers were washed with a saturated solution of Na₂S₂O₃ twice or more if a brown color persist and finally with brine. Organic layer was dried over Na₂SO₄, filtered and evaporated under vacuum. Crude product was then purified by SiO₂ gel chromatography eluted with DCM-MeOH from 0 to 6%, resulting in expected product as a brown oil (156 mg, 58%).

¹H NMR (300 MHz, CDCl₃) δ (ppm): 2.18 (t, J = 7.6 Hz, 2H), 2.08 – 1.96 (m, 3H), 1.82 (t, J = 7.6 Hz, 2H),
1.66 (t, J = 7.1 Hz, 2H); ¹³C NMR (75 MHz, CDCl₃): δ (ppm) 177.87, 82.69, 69.46, 32.34, 28.25, 27.91,
27.61, 13.38

PDC-2-NH2 (27)

4-((2-aminoethyl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



PDC-CI (1 equiv, 353 mg, 0.78 mmol) was suspended in 3 mL of DMSO. Ethylenediamine (15 equiv, 0.78 mL, 11.7 mmol) and TEA (5 equiv, 0.54 mL, 3.89 mmol) were next added and resulting mixture was stirred at 70°C for 16 h. After that time, reaction was allowed to cool down to rt and 30 mL of water was added. Reaction mixture was centrifuged for 20 min and surnatent was carefully removed. This step was repeated once and resulting precipitate was filtrated and washed with water until neutrality was reached. The beige paste was finally dried under vacuum to give **PDC-2-NH2** as a beige solid (319 mg, 86%)

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.29 (s, 2H), 9.36 (d, *J* = 2.3 Hz, 2H), 8.96 (d, *J* = 2.0 Hz, 2H), 8.04 (d, *J* = 8.6 Hz, 4H), 7.72 (t, *J* = 7.6 Hz, 2H), 7.64 (t, *J* = 7.7 Hz, 2H), 7.58 (bs, 2H), 7.51 (t, *J* = 5.3 Hz, 1H), 3.29 – 3.21 (m, 2H), 2.80 (t, *J* = 6.3 Hz, 2H).; ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 163.18, 156.44, 146.11, 144.67, 131.97, 128.70, 128.34, 127.91, 127.76, 127.20, 124.15, 45.63, 39.52; **LC/MS** (Method B): Retention time 5.63 min, m/z (ESI+) = 478.3 [M-H]⁺.

PDC-PEG₂-NH2 (28)

4-((2-(2-(2-aminoethoxy)ethoxy)ethyl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



PDC-CI (1 equiv, 604mg, 1.33 mmol) was suspended in 6 mL of DMSO. 1,8-diamino-3,6-dioxaoctane (15 equiv, 2.9 mL, 20.0 mmol) and TEA (5 equiv, 0.92 mL, 6.65 mmol) were next added and resulting mixture was stirred at 70°C for 16 h. After that time, reaction was allowed to cool down to rt and 120 mL of water was added. Mixture was centrifuged for 20 min and surnatent was carefully removed. This step was repeated once and resulting precipitate was filtrated and washed with water until

neutrality was reached. The beige paste was finally dried under vacuum to give PDC-PEG₂-NH2 as a beige solid (589 mg, 78%)

¹H NMR (300 MHz, DMSO-*d6*) δ (ppm): 11.29 (s, 2H), 9.36 (d, J = 2.3 Hz, 2H), 8.96 (d, J = 2.1 Hz, 2H), 8.04 (d, J = 8.6 Hz, 4H), 7.72 (dt, J = 7.0, 1.1 Hz, 2H), 7.67 – 7.56 (m, 5H), 3.66 (t, J = 5.2 Hz, 2H), 3.62 – 3.51 (m, 4H), 3.50 – 3.41 (m, 2H), 3.36 (t, J = 4.8 Hz, 2H), 2.64 (t, J = 5.8 Hz, 2H).; ¹³C NMR (75 MHz, DMSO-*d6*) δ (ppm): 163.18, 156.45, 146.10, 144.67, 131.99, 128.70, 128.34, 127.91, 127.76, 127.20, 124.14, 72.62, 69.87, 69.64, 68.72, 42.13, 41.10.; **LC/MS** (Method B): Retention time 5.93 min , m/z (ESI+) = 566.6 [M-H]⁺.

Compound 29

4-((2-(3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)propanamido)ethyl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Previously synthesized diazirine acid **26** (1 equiv, 10 mg, 0.06 mmol) and **PDC-2NH2** (1.1 equiv, 31.6 mg, 0.06 mmol) were solubilized in 1.0 mL of DMF. DIPEA (3 equiv, 0.03 mL, 0.18 mmol) was added followed by HATU (1.2 equiv, 27.5 mg, 0.07 mmol). Reaction mixture was stirred at room temperature under argon atmosphere for 1h 30. Reaction mixture was concentrated under vacuum and crude product was purified by SiO_2 gel chromatography eluted with DCM-MeOH from 0 to 8% resulting in expected product as a white solid (34 mg, 90%).

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.30 (s, 2H), 9.37 (d, *J* = 2.3 Hz, 2H), 8.98 (d, *J* = 2.0 Hz, 2H), 8.10 (t, *J* = 5.3 Hz, 1H), 8.04 (d, *J* = 8.5 Hz, 4H), 7.73 (t, *J* = 7.4 Hz, 2H), 7.64 (t, *J* = 7.5 Hz, 2H), 7.60 – 7.52 (m, 3H), 3.40 – 3.35 (m, 2H), 3.31 – 3.26 (m, 2H), 2.82 (t, *J* = 2.5 Hz, 1H), 1.98 (td, *J* = 7.5, 2.7 Hz, 2H), 1.91 (t, *J* = 8.5 Hz, 2H), 1.67 (t, *J* = 7.7 Hz, 2H), 1.57 (t, *J* = 7.3 Hz, 2H).; ¹³**C NMR** (75 MHz, DMSO*d6*) δ (ppm): 171.05, 163.08, 156.40, 146.06, 144.64, 131.93, 128.67, 128.31, 127.88, 127.73, 127.17, 124.11, 99.51, 83.13, 31.40, 29.48, 28.23, 28.01, 12.64; **LC/MS** (Method B): Retention time 6.83 min, m/z (ESI+) = 626.4 [M-H]⁺.

Compound 30

4-((4-(3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)propanamido)butyl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Previously synthesized diazirine acid **26** (1 equiv, 10 mg, 0.06 mmol) was solubilized in 1.0 mL of DMF. DIPEA (3 equiv, 0.03 mL, 0.18 mmol) and HATU (1.5 equiv, 34.3 mg, 0.09 mmol) were next added followed by **PDC-4NH2** (1.1 equiv, 33.5 mg, 0.07 mmol). Reaction mixture was stirred at room temperature under argon atmosphere for 1h 30. Reaction mixture was concentrated under vacuum and crude product was purified by SiO₂ gel chromatography eluted with DCM-EtOH from 0 to 8% resulting in expected product as a white solid (30 mg, 76%).

¹H NMR (300 MHz, DMSO-*d6*) δ (ppm): δ 11.28 (s, 2H), 9.36 (d, J = 2.4 Hz, 2H), 8.96 (d, J = 2.2 Hz, 2H), 8.04 (d, J = 8.5 Hz, 4H), 7.89 (t, J = 5.5 Hz, 1H), 7.72 (dt, J = 7.7, 1.3 Hz, 2H), 7.63 (dt, J = 7.6, 1.0 Hz, 2H), 7.60 – 7.46 (m, 3H), 3.25 (dt, J = 5.9, 5.5 Hz, 2H), 3.10 (dd, J = 6.2, 5.8 Hz, 2H), 2.81 (t, J = 2.6 Hz, 1H), 1.98 (td, J = 7.4, 2.6 Hz, 2H), 1.89 (t, J = 7.6 Hz, 2H), 1.72 – 1.48 (m, 8H); ¹³C NMR (75 MHz, DMSO-*d6*) δ (ppm): 170.98, 163.63, 156.75, 146.56, 145.12, 132.43, 129.16, 128.79, 128.36, 128.22, 127.65, 124.60, 100.00, 83.64, 72.21, 42.18, 38.61, 31.94, 30.00, 28.75, 28.62, 27.14, 26.04, 13.13; LC/MS (Method B): Retention time 6.61 min, m/z (ESI+) = 654.6 [M-H]⁺.

Compound 31

4-((2-(2-(2-(3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)propanamido)ethoxy)ethoxy)ethyl)amino)-N2,N6di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Previously synthesized diazirine acid **26** (1 equiv, 29 mg, 0.17 mmol) and **PDC-PEG₂-NH2** (1.1 equiv, 109 mg, 0.19 mmol) were solubilized in 2.9 mL of DMF. DIPEA (3 equiv, 0.09 mL, 0.52 mmol) was added followed by HATU (1.2 equiv, 80 mg, 0.20 mmol). Reaction mixture was stirred at room temperature under argon atmosphere for 1h 30. Reaction mixture was concentrated under vacuum and crude product was purified by SiO₂ gel chromatography eluted with DCM-MeOH from 0 to 8% resulting in expected product as a white solid (128 mg, quant.).

¹H NMR (300 MHz, DMSO-*d6*) δ (ppm): 11.28 (s, 2H), 9.36 (d, *J* = 1.6 Hz, 2H), 8.97 (d, *J* = 1.8 Hz, 2H), 8.04 (d, *J* = 8.3 Hz, 4H), 7.92 (t, *J* = 5.4 Hz, 1H), 7.72 (dt, *J* = 7.8, 1.2 Hz, 2H), 7.67 – 7.54 (m, 5H), 3.65 (t, *J* = 5.0 Hz, 2H), 3.62 – 3.51 (m, 4H), 3.48 – 3.40 (m, 4H), 3.24 – 3.15 (m, 2H), 2.80 (t, *J* = 2.5 Hz, 1H), 1.96 (dt, *J* = 7.4, 2.6 Hz, 2H), 1.90 (t, *J* = 7.7 Hz, 2H), 1.61 (t, *J* = 7.6 Hz, 2H), 1.54 (t, *J* = 7.4 Hz, 2H). ; ¹³C NMR (75 MHz, DMSO-*d6*) δ (ppm): 170.71, 163.10, 163.03, 156.39, 145.97, 145.89, 144.52, 131.94, 131.84, 128.57, 128.30, 127.85, 127.72, 127.15, 124.16, 124.05, 83.13, 71.64, 69.76, 69.56, 69.08, 68.66, 42.02, 31.38, 29.36, 28.19, 28.09, 12.61. ; .; LC/MS (Method B): Retention time 7.15 min , m/z (ESI+) = 714.6 [M-H]⁺.

PDC-2-N2-Alk (32)

3,3'-((4-((2-(3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)propanamido)ethyl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



Previously synthesized compound **29** (1 equiv, 20 mg, 0.03 mmol) was dissolved at 40 °C in 0.3 mL of dried DMF. Mel (100 equiv, 0.2 mL, 3.2 mmol) was added slowly and mixture was stirred at 40 °C under argon atmosphere. Reaction was over after 16 h monitored by LCMS. Reaction mixture was evaporated to dryness under vacuum and crude solid was suspended in MeCN. Resulting suspension was filtered and yellow solid was washed several times with MeCN, water and then with Et2O. Resulting solid was dried under vacuum yielding expected product as a yellow solid (24 mg, 83%).

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.72 (s, 2H), 10.12 (d, *J* = 1.3 Hz, 2H), 9.65 (d, *J* = 1.1 Hz, 2H), 8.65 - 8.50 (m, 4H), 8.23 (t, *J* = 7.9 Hz, 2H), 8.14 - 8.00 (m, 3H), 7.73 (t, *J* = 5.5 Hz, 1H), 7.61 (s, 2H),

4.78 (s, 6H), 3.31 - 3.24 (m, 2H), 2.81 (t, J = 2.6 Hz, 1H), 1.98 (td, J = 7.2, 2.4 Hz, 2H), 1.96 - 1.88 (m, 2H), 1.67 (t, J = 7.7 Hz, 2H), 1.56 (t, J = 7.4 Hz, 2H).; ¹³**C** NMR (75 MHz, DMSO-*d6*): δ (ppm) 171.14, 163.17, 156.75, 144.74, 135.70, 134.10, 133.94, 132.41, 130.37, 129.89, 129.25, 119.25, 83.16, 71.78, 46.18, 37.83, 31.43, 29.51, 28.27, 28.01, 12.69 ; **LC/MS** (Method B): Retention time 5.99 min, m/z (ESI+) = 654.3; **HRMS** (ESI, *m/z*) calcd. for formula $C_{37}H_{37}IN_9O_3^+$ [M-I]⁺ = 782.2064; found = 782.2084.

PDC-4-N2-Alk (33)

3,3'-((4-((4-(3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)propanamido)butyl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



Previously synthesized compound **30** (1 equiv, 152 mg, 0.23 mmol) was dissolved at 40 °C in 2.0 mL of dried DMF. Mel (1.5 mL, 23.3 mmol, 100 equiv) was added slowly and mixture was stirred at 40 °C under argon atmosphere. Reaction was over after 16 h monitored by LCMS. Reaction mixture was evaporated to dryness under vacuum and crude solid was suspended in MeCN. Resulting suspension was filtered and yellow solid was washed several times with MeCN and then with Et2O. Resulting solid was dried under vacuum yielding expected product as a yellow solid (170 mg, 78%).

¹H NMR (300 MHz, DMSO-*d6*) δ (ppm): 11.69 (s, 2H), 10.16 (s, 2H), 9.65 (s, 2H), 8.56 (d, J = 8.9 Hz, 2H), 8.53 (d, J = 9.6 Hz, 2H), 8.21 (t, J = 7.8 Hz, 2H), 8.05 (t, J = 7.5 Hz, 2H), 7.90 (t, J = 5.1 Hz, 1H), 7.69 (t, J = 4.4 Hz, 1H), 7.65 – 7.41 (m, 3H), 4.78 (s, 6H), 3.31 - 3.23 (m, 2H), 3.11 (dt, J = 5.9, 5.6 Hz, 2H), 2.82 (t, J = 2.4 Hz, 1H), 1.98 (dt, J = 7.3, 2.4 Hz, 2H), 1.90 (t, J = 7.6 Hz, 2H), 1.72 – 1.47 (m, 8H); ¹³C NMR (75 MHz, DMSO-*d6*) δ (ppm): 170.54, 163.16, 156.50, 144.69, 135.61, 133.99, 133.88, 132.34, 130.29, 129.84, 129.19, 119.17, 83.16, 71.75, 46.18, 41.75, 38.10, 31.47, 29.53, 28.28, 28.13, 26.70, 25.47, 12.68; LC/MS (Method B): Retention time 4.84 min, m/z (ESI+) = 682.6; HRMS (ESI, *m/z*) calcd. for formula $C_{39}H_{41}IN_9O_3^+$ [M-I]⁺ = 810.2377; found = 810.2339.

PDC-PEG₂-N2-Alk (34)

3,3'-((4-((2-(2-(2-(3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)propanamido)ethoxy)ethoxy)ethyl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



Previously synthesized compound **31** (1 equiv, 24 mg, 0.03 mmol) was dissolved at 40 °C in 0.3 mL of dried DMF. Mel (100 equiv, 0.2 mL, 3.4 mmol) was added slowly and mixture was stirred at 40 °C under argon atmosphere. Reaction was over after 16 h monitored by LCMS. Reaction mixture was evaporated to dryness under vacuum and crude solid was suspended in MeCN. Resulting suspension was filtered and yellow solid was washed several times with MeCN, water and then with Et2O. Resulting solid was dried under vacuum yielding expected product as a yellow solid (27 mg, 79%).

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.71 (s, 2H), 10.13 (s, 2H), 9.65 (s, 2H), 8.55 (dd, *J* = 8.0, 3.8 Hz, 4H), 8.24 (d, *J* = 7.3 Hz, 2H), 8.07 (t, *J* = 7.5 Hz, 2H), 7.92 (t, *J* = 5.0 Hz, 1H), 7.77 (t, *J* = 5.0 Hz, 1H), 7.64 (s, 2H), 4.78 (s, 6H), 3.65 (t, *J* = 4.8 Hz, 2H), 3.62 – 3.53 (m, 4H), 3.49 – 3.38 (m, 4H), 3.18 (dt, *J* = 5.3, 5.7 Hz, 2H), 2.81 (t, *J* = 2.2 Hz, 1H), 1.96 (td, *J* = 7.3, 2.4 Hz, 2H), 1.89 (t, *J* = 7.7 Hz, 2H), 1.60 (t, *J* = 7.6 Hz, 2H), 1.53 (t, *J* = 7.4 Hz, 2H).; ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 170.81, 163.23, 156.78, 144.75, 135.70, 134.13, 133.97, 132.43, 130.38, 129.90, 129.26, 119.25, 71.79, 69.87, 69.63, 69.13, 68.78, 46.21, 42.16, 38.57, 31.41, 29.42, 28.28, 28.12.; **LC/MS** (Method B): Retention time 5.93 min , m/z (ESI+) = 742.7. **HRMS** (ESI, *m/z*) calcd. for formula $C_{41}H_{45}IN_9O_5^+$ [M-I]⁺ = 870.2588; found = 870.2582.

4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoyl chloride



4-[3-(trifluoromethyl)-3H-diazirin-3-yl]benzoic acid (1 equiv) was dissolved in SOCl₂ (27.5 equiv) and the mixture was stirred for 20 h under argon atmosphere. Volatiles were removed under vacuum and resulting solid was dried under vacuum before being used in next reaction.





The commercially available carboxylic acid (1 equiv) was dissolved in Et2O and was reacted with PCl₅ (1.1 equiv) under argon atmosphere for 2 h. Volatiles were removed under vacuum and used without purification

4-benzoylbenzoyl chloride



A mixture of commercially available benzophenone (1 equiv) in thionyl chloride (70 equiv) was heated under reflux for 2h under argon atmosphere. Volatiles were removed under vacuum and resulting colorless solid was used without purification.

Compound 35

methyl 4-(prop-2-yn-1-ylamino)butanoate

Propargylamine (2 equiv, 0.7 mL, 11 mmol) was solubilized in MeCN (9 mL) at 0 °C and methyl 4bromobutanoate (1 equiv, 0.7 mL, 5.5 mmol) in MeCN (9 mL) was added at 0 °C over 20 min. The reaction mixture was stirred at 0 °C for 1h and was then allowed to warm to rt. Reaction mixture was stirred at rt for 16 h and then 24 h at 40 °C. After coming back to rt, volatiles were removed under vacuum and directly purified by SiO_2 gel chromatography eluted with DCM-EtOH from 0 to 2%, affording compound **35** as a colorless oil (467 mg, 54%).

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 3.67 (s, 3H), 3.41 (d, *J* = 2.4 Hz, 2H), 2.72 (t, *J* = 7.0 Hz, 2H), 2.39 (t, *J* = 7.4 Hz, 2H), 2.20 (t, *J* = 2.4 Hz, 1H), 1.81 (quint, *J* = 7.2 Hz, 2H).; ¹³**C NMR** (75 MHz, CDCl₃) δ (ppm): 174.10, 82.26, 71.42, 51.70, 47.85, 38.11, 31.93, 25.06.

Compound 36

methyl 4-(pent-4-yn-1-ylamino)butanoate



pent-4-yn-1-amine, HCl (2 equiv, 294 mg, 2.46 mmol) and TEA (3 equiv, 0.5 mL, 3.69 mmol) were solubilized in MeCN (12 mL) at 0 °C and methyl 4-bromobutanoate (1 equiv, 0.15 mL, 1.23 mmol) in MeCN (12 mL) was added at 0 °C over 20 min. The reaction mixture was stirred at 0 °C for 1h and was next heated at 40 °C for 24 h. After coming back to rt, volatiles were removed under vacuum and directly purified by SiO₂ gel chromatography eluted with DCM-EtOH from 0 to 2%, affording **36** as a yellow oil (137 mg, 61%).

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 3.69 (s, 3H), 3.15 (t, *J* = 7.9 Hz, 2H), 3.12 (t, *J* = 7.9 Hz, 2H), 2.54 (t, *J* = 6.9 Hz, 2H), 2.41 (td, *J* = 6.7, 2.5 Hz, 2H), 2.31 – 2.11 (m, 4H), 2.03 (t, *J* = 2.6 Hz, 1H).; ¹³**C NMR** (75 MHz, CDCl₃) δ (ppm): 172.88, 81.79, 70.45, 52.07, 47.29, 46.88, 31.16, 24.52, 21.09, 16.23.

Compound 37

methyl 4-(N-(prop-2-yn-1-yl)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamido)butanoate



To a solution of previously prepared **35** (1 equiv, 83 mg, 0.54 mmol) and TEA (2 equiv, 0.15 mL, 1.07 mmol) in DCM (1 mL) was added dropwise a solution **4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoyl chloride** (1.5 equiv, 200 mg, 0.8 mmol) in DCM (1 mL) at 0 °C under argon atmosphere. The reaction mixture was allowed to warm to rt and was stirred 16 h at rt under argon atmosphere. The reaction mixture was cooled down to 0 °C and some water was added. Layers were separated and aqueous

layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated under vacuum. The mixture was purified by SiO₂ gel chromatography eluted with Cy-AcOEt from 0 to 20%, yielding expected compound as a pale yellow oil (197 mg, 100%).

¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.62 – 7.41 (m, 2H), 7.25 (bs, 2H), 4.37 (bs, 1H), 3.96 (bs, 1H), 3.77 – 3.32 (m, 5H), 2.57 – 2.16 (m, 3H), 2.16 – 1.83 (m, 2H).

Compound 38

methyl 4-(4-azido-2,3,5,6-tetrafluoro-N-(prop-2-yn-1-yl)benzamido)butanoate



To a solution of previously prepared **35** (1 equiv, 290 mg, 1.87 mmol) and TEA (2 equiv, 0.52 mL, 3.74 mmol) in DCM (3.2 mL) was added dropwise a solution of **4-azido-2,3,5,6-tetrafluorobenzoyl chloride** (1.5 equiv, 710 mg, 2.8 mmol) in DCM (3.2 mL) at 0 °C under argon atmosphere. The reaction mixture was allowed to warm to rt and was stirred 16 h at rt under argon atmosphere. The reaction mixture was cooled down to 0 °C and some water was added. Layers were separated and aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated under vacuum. The mixture was purified by SiO₂ gel chromatography eluted with Cy-AcOEt from 0 to 20%, yielding expected compound as a pale yellow oil (488 mg, 70%).

Presence of two conformers: ¹H NMR (300 MHz, CDCl₃) δ (ppm): 4.39 (d, *J* = 2.5 Hz, 1H), 3.99 (d, *J* = 2.4 Hz, 1H), 3.73 (t, *J* = 7.1 Hz, 1H), 3.70 (s, 1.5H), 3.62 (s, 1.5H), 3.44 (t, *J* = 7.3 Hz, 1H), 2.43 (t, *J* = 7.4 Hz, 1H), 2.34 – 2.24 (m, 2H), 2.03 (quint, *J* = 7.3 Hz, 1H), 1.89 (quint, *J* = 7.1 Hz, 1H).

Compound 39

methyl 4-(4-benzoyl-N-(prop-2-yn-1-yl)benzamido)butanoate



To a solution of previously prepared **35** (1 equiv, 90 mg, 0.58 mmol) and TEA (2 equiv, 0.16 mL, 1.16 mmol) in DCM (3 mL) was added dropwise a solution **4-benzoylbenzoyl chloride** (1.5 equiv, 212 mg, 0.87 mmol) in DCM (3 mL) at 0 °C under argon atmosphere. The reaction mixture was allowed to

warm to rt and was stirred 16 h at rt under argon atmosphere. The reaction mixture was cooled down to 0 °C and some water was added. Layers were separated and aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated under vacuum. The mixture was purified by SiO₂ gel chromatography eluted with DCM-AcOEt from 0 to 5%, yielding expected compound as a colorless oil (162 mg, 77%).

Presence of two conformers: ¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 7.83 (d, *J* = 8.2 Hz, 2H), 7.81 – 7.76 (m, 2H), 7.64 – 7.56 (m, 2H), 7.48 (t, *J* = 7.5 Hz, 3H), 4.39 (bs, 1H), 4.00 (bs, 1H), 3.76 – 3.53 (m, 4H), 3.50 – 3.41 (m, 1H), 2.52 – 2.15 (m, 3H), 2.14 – 1.83 (m, 2H).; ¹³**C NMR** (75 MHz, CDCl₃) δ (ppm): 196.09, 170.61, 139.39, 138.82, 137.16, 132.95, 130.29, 130.21, 128.56, 126.84, 126.34, 78.52, 51.85, 47.83, 46.41, 44.92, 39.46, 33.98, 31.90, 31.48, 30.77, 23.29, 22.41.

Compound 40

methyl 4-(N-(pent-4-yn-1-yl)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamido)butanoate



To a solution of previously prepared **36** (1 equiv, 106 mg, 0.58 mmol) and TEA (2 equiv, 0.16 mL, 1.16 mmol) in DCM (5 mL) was added dropwise a solution of **4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoyl chloride** (1.5 equiv, 216 mg, 0.87 mmol) in DCM (5 mL) at 0 °C under argon atmosphere. The reaction mixture was allowed to warm to rt and was stirred 16 h at rt under argon atmosphere. The reaction mixture was cooled down to 0 °C and some water was added. Layers were separated and aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated under vacuum. The mixture was purified by SiO₂ gel chromatography eluted with Cy-AcOEt from 0 to 20%, yielding expected compound as a pale yellow oil (72 mg, 32%).

Presence of two conformers: ¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.40 (d, *J* = 8.3 Hz, 2H), 7.23 (d, *J* = 8.2 Hz, 2H), 3.69 (bs, 2H), 3.59 (s, 3H), 3.30 (bs, 2H), 2.51 – 2.13 (m, 3H), 2.09 – 1.62 (m, 6H).; ¹³C NMR (75 MHz, CDCl₃) δ (ppm): 170.73, 138.14, 130.29, 127.20, 126.81, 123.91, 120.27, 69.37, 51.84, 48.39, 44.39, 31.19, 27.51, 22.85, 15.94.

227

Compound 41

methyl 4-(4-azido-2,3,5,6-tetrafluoro-N-(pent-4-yn-1-yl)benzamido)butanoate



To a solution of previously prepared **36** (1 equiv, 52 mg, 0.28 mmol) and TEA (2 equiv, 0.08 mL, 0.57 mmol) in DCM (1 mL) was added dropwise a solution of **4-azido-2,3,5,6-tetrafluorobenzoyl chloride** (1.5 equiv, 108 mg, 0.43 mmol) in DCM (1 mL) at 0 °C under argon atmosphere. The reaction mixture was allowed to warm to rt and was stirred 16 h at rt under argon atmosphere. The reaction mixture was cooled down to 0 °C and some water was added. Layers were separated and aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated under vacuum. The mixture was purified by SiO₂ gel chromatography eluted with DCM-AcOEt from 0 to 2%, yielding expected compound as a pale colorless oil (75 mg, 66%).

Presence of two conformers: ¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 3.69 (s, 2H), 3.64 – 3.52 (m, 3H), 3.38 – 3.22 (m, 2H), 2.41 (t, *J* = 7.3 Hz, 1H), 2.33 – 2.20 (m, 2H), 2.14 (td, *J* = 6.6, 2.6 Hz, 1H), 2.04 – 1.78 (m, 4H), 1.72 (quint, *J* = 7.0 Hz, 1H).; ¹³**C NMR** (75 MHz, CDCl₃) δ (ppm): 173.49, 172.82, 159.17, 144.47, 142.48, 142.21, 141.16, 139.06, 138.90, 83.18, 82.20, 69.47, 69.40, 51.87, 51.83, 48.26, 47.48, 44.55, 44.36, 31.00, 30.03, 27.09, 26.07, 23.40, 22.58, 16.05, 15.55.

Compound 42

methyl 4-(4-benzoyl-N-(pent-4-yn-1-yl)benzamido)butanoate



To a solution of previously prepared **36** (1 equiv, 80 mg, 0.44 mmol) and TEA (2 equiv, 0.12 mL, 0.87 mmol) in DCM (2 mL) was added dropwise a solution **4-benzoylbenzoyl chloride** (1.5 equiv, 160 mg, 0.65 mmol) in DCM (2 mL) at 0 °C under argon atmosphere. The reaction mixture was allowed to warm to rt and was stirred 16 h at rt under argon atmosphere. The reaction mixture was cooled down to 0 °C and some water was added. Layers were separated and aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated under

vacuum. The mixture was purified by SiO_2 gel chromatography eluted with DCM-AcOEt from 0 to 5%, yielding expected compound as a colorless oil (92 mg, 54%).

Presence of two conformers: ¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 7.87 – 7.76 (m, 4H), 7.61 (t, *J* = 7.4 Hz, 1H), 7.51 (d, *J* = 7.9 Hz, 2H), 7.48 (d, *J* = 8.2 Hz, 2H), 3.70 (bs, 2H), 3.60 (s, 3H), 3.43 – 3.22 (m, 2H), 2.45 (t, *J* = 6.8 Hz, 1H), 2.33 (s, 1H), 2.18 (t, *J* = 7.1 Hz, 1H), 2.13 – 1.69 (m, 6H). ¹³**C NMR** (75 MHz, CDCl₃) δ (ppm): 196.13, 171.03, 140.52, 138.39, 137.22, 132.90, 130.30, 130.19, 128.53, 126.55, 69.59, 53.56, 51.84, 48.76, 48.17, 44.35, 44.25, 31.39, 30.80, 27.50, 26.35, 23.95, 22.85, 16.30, 15.84.

Compound 43

4-(N-(prop-2-yn-1-yl)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamido)butanoic acid



To the previously synthesized **37** (1 equiv, 200 mg, 0.54 mmol) in EtOH (2.8 mL) was added dropwise NaOH (2 equiv, 0.7 mL 1.6 M, 1.09 mmol) at 0 °C. Reaction mixture was allowed to warm to rt and was stirred 16 h. Volatiles were removed under vacuum and the mixture was solubilized in water, washed 3 times with DCM and aqueous layer was then acidified to pH = 1 by addition of 10% HCl. Aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated to give **43** as a colorless oil (192 mg, 100%)

Presence of two conformers: ¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 7.67 – 7.36 (m, 2H), 7.23 (bs, 1H), 4.36 (s, 1H), 4.02 – 3.88 (m, 1H), 3.76 – 3.61 (m, 1H), 3.57 – 3.33 (m, 1H), 2.62 – 2.18 (m, 3H), 2.16 – 1.72 (m, 2H).

Compound 44

4-(4-azido-2,3,5,6-tetrafluoro-N-(prop-2-yn-1-yl)benzamido)butanoic acid



To the previously synthesized **38** (1 equiv, 50 mg, 0.13 mmol) in EtOH (0.7 mL) / DCM (1.3 mL) was added dropwise NaOH (2 equiv, 0.17 mL 1.6 M, 0.27 mmol) at 0 °C. Reaction mixture was allowed to warm to rt and was stirred 16 h. Volatiles were removed under vacuum and the mixture was

solubilized in water, washed 3 times with DCM and aqueous layer was then acidified to pH = 1 by addition of 10% HCl. Aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated to give **44** as a colorless oil (48 mg, 100%)

Presence of two conformers: ¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 4.39 (d, *J* = 2.3 Hz, 1H), 3.98 (d, *J* = 2.0 Hz, 1H), 3.74 (t, *J* = 7.1 Hz, 1H), 3.43 (t, *J* = 7.6 Hz, 1H), 2.48 (t, *J* = 7.3 Hz, 1H), 2.40 – 2.26 (m, 2H), 2.04 (quint, *J* = 7.3 Hz, 1H), 1.89 (quint, *J* = 6.8 Hz, 1H).

Compound 45

4-(4-benzoyl-N-(prop-2-yn-1-yl)benzamido)butanoic acid



To the previously synthesized **39** (1 equiv, 162 mg, 0.45 mmol) in EtOH (2mL) was added dropwise NaOH (2 equiv, 0.56 mL 1.6 M, 0.90 mmol) at 0 °C. Reaction mixture was allowed to warm to rt and was stirred 16 h. Volatiles were removed under vacuum and the mixture was solubilized in water, washed 3 times with DCM and aqueous layer was then acidified to pH = 1 by addition of 10% HCl. Aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated to give **45** as a colorless oil (120 mg, 77%)

Presence of two conformers: ¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 9.00 (bs, 1H), 7.82 (d, *J* = 8.1 Hz, 2H), 7.79 (d, *J* = 7.5 Hz, 2H), 7.60 (t, *J* = 7.3 Hz, 2H), 7.52 – 7.43 (m, 3H), 4.39 (bs, 1H), 3.99 (bs, 1H), 3.73 – 3.39 (m, 2H), 2.55 – 2.18 (m, 3H), 2.10 – 1.85 (m, 2H).

Compound 46

4-(N-(pent-4-yn-1-yl)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamido)butanoic acid



To the previously synthesized **40** (1 equiv, 138 mg, 0.35 mmol) in EtOH (2 mL) was added dropwise NaOH (2 equiv, 0.45 mL 1.6 M, 0.7 mmol) at 0 °C. Reaction mixture was allowed to warm to rt and was stirred 16 h. Volatiles were removed under vacuum and the mixture was solubilized in water,

washed 3 times with DCM and aqueous layer was then acidified to pH = 1 by addition of 10% HCl. Aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated to give **46** as a colorless oil (112 mg, 84%)

Presence of two conformers: ¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 7.41 (d, *J* = 8.3 Hz, 2H), 7.22 (d, *J* = 7.7 Hz, 2H), 3.56 (bs, 2H), 3.41 – 3.15 (m, 2H), 2.54 – 2.15 (m, 3H), 2.14 – 1.62 (m, 6H). ; ¹³**C NMR** (75 MHz, CDCl₃) δ (ppm): 177.90, 171.03, 137.83, 130.43, 127.21, 126.79, 123.91, 120.27, 69.48, 48.79, 48.25, 45.37, 44.46, 31.40, 30.74, 28.71, 27.59, 26.27, 23.07, 22.83, 16.35, 15.80.

Compound 47

4-(4-azido-2,3,5,6-tetrafluoro-N-(pent-4-yn-1-yl)benzamido)butanoic acid



To the previously synthesized **41** (1 equiv, 250 mg, 0.62 mmol) in EtOH (3 mL) was added dropwise NaOH (2 equiv, 0.8 mL 1.6 M, 1.25 mmol) at 0 °C. Reaction mixture was allowed to warm to rt and was stirred 16 h. Volatiles were removed under vacuum and the mixture was solubilized in water, washed 3 times with DCM and aqueous layer was then acidified to pH = 1 by addition of 10% HCl. Aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated to give **47** as a pale yellow oil (234 mg, 97%)

Presence of two conformers: ¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 3.63 (dd, *J* = 14.8, 7.8 Hz, 2H), 3.41 – 3.21 (m, 2H), 2.48 (t, *J* = 7.2 Hz, 1H), 2.38 – 2.25 (m, 2H), 2.15 (td, *J* = 6.6, 2.5 Hz, 1H), 2.05 – 1.80 (m, 4H), 1.73 (quint, *J* = 7.0 Hz, 1H).

Compound 48

4-(4-benzoyl-N-(pent-4-yn-1-yl)benzamido)butanoic acid



To the previously synthesized **42** (1 equiv, 150 mg, 0.38 mmol) in EtOH (2 mL) was added dropwise NaOH (2 equiv, 0.5 mL 1.6 M, 0.77 mmol) at 0 °C. Reaction mixture was allowed to warm to rt and

was stirred 16 h. Volatiles were removed under vacuum and the mixture was solubilized in water, washed 3 times with DCM and aqueous layer was then acidified to pH = 1 by addition of 10% HCl. Aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated to give **48** as a pale yellow oil (133 mg, 92%)

Presence of two conformers: ¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 7.87 – 7.74 (m, 4H), 7.61 (t, *J* = 7.4 Hz, 1H), 7.53 – 7.42 (m, 4H), 3.68 – 3.51 (m, 2H), 3.41 – 3.23 (m, 2H), 2.46 (t, *J* = 7.3 Hz, 1H), 2.38 – 2.28 (m, 1H), 2.26 – 2.14 (m, 1H), 2.12 – 1.66 (m, 6H).; ¹³**C NMR** (75 MHz, CDCl₃) δ (ppm): 196.21, 171.37, 140.22, 138.44, 137.18, 132.97, 130.34, 130.23, 128.57, 126.59, 69.70, 48.71, 48.23, 44.37, 27.48, 26.34, 22.75, 16.31, 15.86.

Compound 49

2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)ethan-1-ol

To a suspension of tBuOK (0.5 equiv, 370 mg, 3.33 mmol) in dry THF (16mL) was added TEG (1 equiv, 1.0 g, 6.66 mmol) at 0°C under argon atmosphere. The reaction mixture was allowed to stir at rt for 30 min then Propargyl Bromide (0.5 equiv, 400 mg, 3.33 mmol) in dry THF (2 mL) was added dropwise. The resulting beige-brown heterogeneous reaction mixture was let stirring for 14h at rt. The reaction mixture was diluted with THF and filtered through a celite pad. The crude mixture was concentrated under vacuum and purified by SiO2 gel chromatography eluted by DCM-AcOEt from 0 to 50% yielding expected product as a colorless oil (461 mg, 74%).

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 4.21 (d, *J* = 2.4 Hz, 2H), 3.78 – 3.65 (m, 10H), 3.64 – 3.58 (m, 2H), 2.48 (bs, 1H), 2.43 (t, *J* = 2.3 Hz, 1H).

Compound 50

3-(2-(2-(2-bromoethoxy)ethoxy)ethoxy)prop-1-yne

Br~0~0~0

Previously prepared **49** (1 equiv; 424 mg, 2.25 mmol) was dissolved in dry DCM and the solution was cooled to 0°C under argon atmosphere. PPh₃ (1.5 equiv, 886 mg, 3.38 mmol) was added and the solution was stirred for 15 min at 0°C. CBr_4 (1.5 equiv, 1218 mg, 3.38 mmol) was added and the solution was stirred for another 30min at 0°C before being allowed to warm up to rt. The reaction mixture was stirred overnight and volatiles were removed under vacuum. Crude mixture was purified

by SiO₂ gel chromatography eluted by DCM, yielding expected product as a colorless oil (540 mg, 88%)

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 4.19 (d, *J* = 2.4 Hz, 2H), 3.80 (t, *J* = 6.3 Hz, 2H), 3.71 – 3.63 (m, 8H), 3.46 (t, *J* = 6.3 Hz, 2H), 2.42 (t, *J* = 2.4 Hz, 1H).

Compound 51

4-ethoxy-4-oxobutan-1-aminium

Thionyl chloride (2.5 equiv, 5.3 mL, 72.7 mmol) was added dropwise at 0°C to a stirred suspension of commercially available 4-aminobutyric acid (1 equiv, 3.0 g, 29.0 mmol) in EtOH (54 mL). The pale yellow homogeneous reaction mixture was heated for 5h at reflux. Volatiles were removed under vacuum and resulting solid was washed with Et_2O , yielding expecting product as a colorless solid (4.6 g, 94%).

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 8.12 (s, 3H), 4.06 (q, *J* = 7.1 Hz, 2H), 2.85 – 2.69 (m, 2H), 2.41 (t, *J* = 7.4 Hz, 2H), 1.88 – 1.73 (m, 2H), 1.18 (t, *J* = 7.1 Hz, 3H).

Compound 52

ethyl 4,7,10-trioxa-13-azaheptadec-1-yn-17-oate



To a mixture of previously prepared **51** (3 equiv, 1262 mg, 7.53 mmol) and TEA (4 equiv, 1.39 mL, 10.0 mmol) in MeCN (9mL) was added dropwise at 0 °C, a solution of **50** (1 equiv, 630 mg, 2.50 mmol) in dry MeCN (9mL) under argon atmosphere. The mixture was stirred under argon at 0 °C for 30 min and then 48 h at 40 °C. Volatiles were removed under vacuum and crude mixture was purified by SiO_2 gel chromatography eluted by DCM-EtOH from 0 to 8%, yielding expected product as a yellow oil (500 mg, 66%)

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 4.17 (d, *J* = 2.4 Hz, 2H), 4.07 (q, *J* = 7.1 Hz, 2H), 3.88 (t, *J* = 5.0 Hz, 2H), 3.71 – 3.57 (m, 8H), 3.19 (t, *J* = 5.0 Hz, 2H), 3.10 (t, *J* = 7.6 Hz, 2H), 2.48 – 2.37 (m, 3H), 2.21 – 2.06 (m, 2H), 1.20 (t, *J* = 7.1 Hz, 3H).; ¹³**C NMR** (75 MHz, CDCl₃) δ (ppm): 172.64, 79.50, 74.96, 70.20, 70.16, 70.13, 68.89, 65.78, 60.85, 58.42, 47.32, 47.22, 31.33, 21.15, 14.17.

Compound 53

ethyl 13-(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoyl)-4,7,10-trioxa-13-azaheptadec-1-yn-17-oate



To a mixture of **52** (1 equiv, 100 mg, 0.33 mmol) and TEA (2 equiv, 0.09 mL, 0.66 mmol) in DCM (2.2 mL) was added dropwise a solution of previously prepared **4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoyl chloride** (1.5 equiv, 123.7 mg, 0.5 mmol) in DCM (2.2 mL) at 0 °C under argon. The mixture was stirred at 0 °C for 1 h before being allowed to warm to rt. The mixture was stirred 16 h under argon atmosphere before being cooled down to 0 °C. Water was added dropwise followed by solid NaHCO₃. Layers were separated and aqueous layer was extracted several times with DCM. Organic layers were dried over MgSO₄, filtrated and concentrated under vacuum. The mixture was purified by SiO₂ gel chromatography eluted with DCM-EtOH from 0 to 2%, yielding expected product as a pale yellow oil (100 mg, 59%).

Presence of two conformers: ¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.45 (d, *J* = 8.0 Hz, 1H), 7.41 (d, *J* = 7.4 Hz, 1H), 7.20 (d, *J* = 7.4 Hz, 2H), 4.18 (d, *J* = 2.3 Hz, 2H), 4.16 – 3.99 (m, 2H), 3.78 – 3.46 (m, 12H), 3.44 – 3.26 (m, 2H), 2.42 (t, *J* = 2.3 Hz, 1H), 2.38 (t, *J* = 6.9 Hz, 1H), 2.20 – 2.08 (m, 1H), 2.05 – 1.91 (m, 1H), 1.88 – 1.74 (m, 1H), 1.32 – 1.12 (m, 3H).): ¹³C NMR (75 MHz, CDCl₃) δ (ppm): 173.32, 171.17, 138.24, 130.18, 127.71, 127.24, 126.70, 126.63, 123.91, 120.27, 79.68, 74.72, 70.72, 70.56, 69.21, 68.63, 60.64, 58.52, 49.58, 48.91, 44.93, 44.67, 31.68, 31.10, 28.70, 28.17, 23.83, 22.74, 14.32.

Compound 54

ethyl 13-(4-azido-2,3,5,6-tetrafluorobenzoyl)-4,7,10-trioxa-13-azaheptadec-1-yn-17-oate



To a mixture of **52** (1 equiv, 40 mg, 0.16 mmol) and TEA (2 equiv, 0.04 mL, 0.31 mmol) in DCM (1 mL) was added dropwise a solution of previously prepared **4-azido-2,3,5,6-tetrafluorobenzoyl chloride** (1.5 equiv, 59 mg, 0.23 mmol) in DCM (1 mL) at 0 °C under argon. The mixture was stirred at 0 °C for 1 h before being allowed to warm to rt. The mixture was stirred 16 h under argon atmosphere before being cooled down to 0 °C. Water was added dropwise followed by solid NaHCO₃. Layers were

separated and aqueous layer was extracted several times with DCM. Organic layers were dried over MgSO₄, filtrated and concentrated under vacuum. The mixture was purified by SiO₂ gel chromatography eluted with DCM-EtOH from 0 to 1%, yielding expected product as a yellow oil (30 mg, 41%).

Presence of two conformers: ¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 4.24 – 4.09 (m, 3H), 4.05 (q, *J* = 7.1 Hz, 1H), 3.78 – 3.46 (m, 12H), 3.41 – 3.30 (m, 2H), 2.43 (t, *J* = 2.2 Hz, 1H), 2.38 (t, *J* = 7.7 Hz, 1H), 2.20 (t, *J* = 6.8 Hz, 1H), 2.02 (quint, *J* = 7.3 Hz, 1H), 1.85 (quint, *J* = 7.0 Hz, 1H), 1.31 – 1.17 (m, 3H).

Compound 55

ethyl 13-(4-benzoylbenzoyl)-4,7,10-trioxa-13-azaheptadec-1-yn-17-oate



To a mixture of **52** (1 equiv, 100 mg, 0.33 mmol) and TEA (2 equiv, 0.09 mL, 0.66 mmol) in DCM (2.2 mL) was added dropwise a solution of previously prepared **4-benzoylbenzoyl chloride** (1.5 equiv, 121 mg, 0.5 mmol) in DCM (2.2 mL) at 0 °C under argon. The reaction mixture was stirred at 0 °C for 1 h and then 8 h at room temperature. The reaction mixture was cooled down to 0 °C and some water was added followed by NaHCO₃ (sat.). Aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated under vacuum. The mixture was purified by SiO₂ gel chromatography eluted with DCM-EtOH from 0 to 3%, yielding expected compound as a pale yellow oil (86 mg, 51%).

Presence of two conformers: ¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.85 – 7.75 (m, 4H), 7.59 (t, J = 7.3 Hz, 1H), 7.51 (d, J = 7.1 Hz, 2H), 7.47 (d, J = 7.3 Hz, 2H), 4.24 – 3.97 (m, 4H), 3.83 – 3.30 (m, 14H), 2.40 (t, J = 2.3 Hz, 1H), 2.13 (t, J = 6.8 Hz, 1H), 2.05 – 1.94 (m, 2H), 1.92 – 1.77 (m, 1H), 1.30 – 1.11 (m, 3H).; ¹³C NMR (75 MHz, CDCl₃) δ (ppm): 196.17, 173.22, 172.52, 171.39, 171.08, 140.65, 140.55, 138.28, 138.20, 137.27, 132.83, 130.26, 130.17, 128.50, 127.05, 126.57, 79.68, 74.71, 70.71, 70.54, 69.18, 60.62, 58.49, 53.55, 49.57, 48.87, 44.90, 44.69, 31.67, 31.18, 23.87, 22.77, 14.33.

Compound 56

13-(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoyl)-4,7,10-trioxa-13-azaheptadec-1-yn-17-oic acid



To the previously synthesized **53** (1 equiv, 100 mg, 0.2 mmol) in EtOH (1 mL) / DCM (1 mL) was added dropwise NaOH (2 equiv, 0.25 mL 1.6 M, 0.40 mmol) at 0 °C. Reaction mixture was allowed to warm to rt and was stirred 16 h. Volatiles were removed under vacuum and the mixture was solubilized in water, washed 3 times with DCM and aqueous layer was then acidified to pH = 1 by addition of 10% HCl. Aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated to give **56** as a colorless oil (92 mg, 97%)

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 7.47 (d, *J* = 7.6 Hz, 1H), 7.43 (d, *J* = 8.2 Hz, 1H), 7.21 (d, *J* = 8.0 Hz, 2H), 4.20 (d, *J* = 2.3 Hz, 2H), 3.80 – 3.51 (m, 13H), 3.47 – 3.27 (m, 2H), 2.51 – 2.40 (m, 2H), 2.19 (t, *J* = 7.0 Hz, 1H), 1.99 (t, *J* = 6.8 Hz, 1H), 1.91 – 1.77 (m, 1H).

Compound 57

13-(4-azido-2,3,5,6-tetrafluorobenzoyl)-4,7,10-trioxa-13-azaheptadec-1-yn-17-oic acid



To the previously synthesized **54** (1 equiv, 100 mg, 0.2 mmol) in EtOH (1mL) / DCM (1 mL) was added dropwise NaOH (2 equiv, 0.25 mL 1.6 M, 0.40 mmol) at 0 °C. Reaction mixture was allowed to warm to rt and was stirred 16 h. Volatiles were removed under vacuum and the mixture was solubilized in water, washed 3 times with DCM and aqueous layer was then acidified to pH = 1 by addition of 10% HCl. Aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated to give **57** as a colorless oil (90 mg, 95%)

Presence of two conformers: ¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 4.18 (t, *J* = 2.1 Hz, 2H), 3.77 – 3.45 (m, 12H), 3.38 (t, *J* = 5.3 Hz, 2H), 2.49 – 2.39 (m, 2H), 2.29 (t, *J* = 6.4 Hz, 1H), 2.02 (quint, *J* = 7.1 Hz, 1H), 1.86 (quint, *J* = 7.1 Hz, 1H).; ¹³**C NMR** (75 MHz, CDCl₃) δ (ppm): 177.14, 176.98, 165.97, 159.28,

144.62, 141.32, 139.06, 138.83, 79.58, 74.81, 70.82, 70.67, 70.48, 70.43, 69.47, 69.22, 69.16, 68.80, 58.50, 49.22, 48.99, 45.61, 45.13, 30.98, 30.11, 23.18, 22.47.

Compound 58

13-(4-benzoylbenzoyl)-4,7,10-trioxa-13-azaheptadec-1-yn-17-oic acid



To the previously synthesized **55** (1 equiv, 300 mg, 0.6 mmol) in EtOH (3 mL) / DCM (3 3 mL) was added dropwise NaOH (2 equiv, 0.74 mL 1.6 M, 1.18 mmol) at 0 °C. Reaction mixture was allowed to warm to rt and was stirred 16 h. Volatiles were removed under vacuum and the mixture was solubilized in water, washed 3 times with DCM and aqueous layer was then acidified to pH = 1 by addition of 10% HCl. Aqueous layer was extracted several times with DCM and organic layers were dried over MgSO₄, filtered and concentrated to give **58** as a light brown oil (205 mg, 72%)

Presence of two conformers: ¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.87 – 7.76 (m, 4H), 7.61 (t, *J* = 7.4 Hz, 1H), 7.53 (d, *J* = 6.5 Hz, 2H), 7.48 (d, *J* = 7.2 Hz, 2H), 4.19 (bs, 2H), 3.86 – 3.50 (m, 12H), 3.50 – 3.27 (m, 2H), 2.48 (t, *J* = 6.6 Hz, 1H), 2.42 (t, *J* = 2.4 Hz, 1H), 2.20 (t, *J* = 7.8 Hz, 1H), 2.07 – 1.99 (m, 1H), 1.94 – 1.82 (m, 1H).; ¹³C NMR (75 MHz, CDCl₃) δ (ppm): 196.21, 177.06, 176.66, 171.64, 171.28, 140.35, 138.28, 137.19, 132.86, 130.17, 128.50, 127.02, 126.57, 79.54, 74.83, 70.68, 70.45, 69.39, 69.13, 68.92, 58.44, 49.43, 48.99, 44.92, 44.78, 31.45, 30.81, 23.51, 22.67.

Compound 59

4-((4-(4-(N-(prop-2-yn-1-yl)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamido)butanamido)butyl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Compound 43 (1.2 equiv, 20 mg, 0.06 mmol) was sheltered from light and dissolved in DMF (0.6mL) and DIEA (3 equiv, 0.02 mL, 0.14 mmol) was added under argon. The mixture was stirred 5 min under argon and HATU (1.2 equiv, 23 mg, 0.06 mmol) was slowly added. The mixture was stirred 5 min under argon and **PDC-4-NH2** (1.0 equiv, 24 mg, 0.05 mmol) was added. The reaction mixture was stirred under argon for 16 h and volatiles were removed under vacuum. The crude product was purified by SiO₂ gel chromatography eluted with DCM-EtOH from 0 to 5%, yielding **59** as a colorless solid (20 mg, 50%).

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.28 (s, 2H), 9.35 (d, *J* = 2.2 Hz, 2H), 8.96 (d, *J* = 1.9 Hz, 2H), 8.04 (d, *J* = 8.4 Hz, 4H), 7.76 – 7.59 (m, 6H), 7.59 – 7.46 (m, 4H), 7.39 – 7.29 (m, 2H), 4.27 (bs, 1H), 4.13 (dd, *J* = 5.8, 1.6 Hz, 1H), 3.97 (bs, 1H), 3.67 – 3.41 (m, 2H), 3.15 – 2.94 (m, 4H), 2.12 (bs, 1H), 2.03 – 1.70 (m, 3H), 1.69 – 1.45 (m, 4H).; **LC/MS** (Method B): Retention time 6.95 min, m/z (ESI+) = 841.4 [M-H]⁺.

Compound 60

4-((4-(4-azido-2,3,5,6-tetrafluoro-N-(prop-2-yn-1-yl)benzamido)butanamido)butyl)amino)-N2,N6di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Compound 44 (1 equiv, 175 mg, 0.49 mmol) was sheltered from light and dissolved in DMF (3.6 mL) and TEA (1 equiv, 0.07 mL, 0.49 mmol), EDCI (1 equiv, 76 mg, 0.49 mmol) and HOBt (0.4 equiv, 26 mg, 0.2 mmol) were added under argon. The mixture was stirred for 5 min and **PDC-4-NH2** (1 equiv, 247 mg, 0.49 mmol) was added. The reaction mixture was stirred under argon for 16 h and volatiles were removed under vacuum. The crude product was purified by SiO₂ gel chromatography eluted with DCM-EtOH from 0 to 5%, yielding **60** as a colorless solid (209 mg, 51%).

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.28 (s, 2H), 9.35 (d, *J* = 2.2 Hz, 2H), 8.96 (d, *J* = 1.9 Hz, 2H), 8.04 (d, *J* = 8.4 Hz, 4H), 7.93 – 7.79 (m, 1H), 7.72 (t, *J* = 7.7 Hz, 2H), 7.65 (t, *J* = 7.4 Hz, 2H), 7.60 – 7.47 (m, 3H), 4.33 (d, *J* = 2.0 Hz, 1H), 4.13 (s, 1H), 3.54 (t, *J* = 7.4 Hz, 1H), 3.30 – 3.19 (m, 4H), 3.18 – 2.98 (m, 2H), 2.13 (t, *J* = 7.3 Hz, 1H), 1.99 (t, *J* = 6.8 Hz, 1H), 1.92 – 1.81 (m, 1H), 1.79 – 1.68 (m, 1H), 1.68 – 1.39 (m, 5H).; **LC/MS** (Method B): Retention time 6.23 min, m/z (ESI+) = 846.6 [M-H]⁺.

Compound 61

4-((4-(4-(4-benzoyl-N-(prop-2-yn-1-yl)benzamido)butanamido)butyl)amino)-N2,N6-di(quinolin-3yl)pyridine-2,6-dicarboxamide



Compound 45 (1 equiv, 120 mg, 0.34 mmol) was sheltered from light and dissolved in DMF (2.6 mL) and TEA (1 equiv, 0.05 mL, 0.34 mmol), EDCI (1 equiv, 53 mg, 0.34 mmol) and HOBt (0.4 equiv, 19 mg, 0.14 mmol) were added under argon. The mixture was stirred for 5 min and **PDC-4-NH2** (1 equiv, 174 mg, 0.34 mmol) was added. The reaction mixture was stirred under argon for 16 h and volatiles were removed under vacuum. The crude product was purified by SiO₂ gel chromatography eluted with DCM-EtOH from 0 to 7%, yielding **61** as a colorless solid (177 mg, 62%).

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 10.20 (s, 2H), 9.14 (s, 2H), 8.82 (s, 2H), 8.04 (d, J = 8.1 Hz, 2H), 7.82 – 7.59 (m, 8H), 7.57 – 7.36 (m, 9H), 3.68 (bs, 2H), 3.40 – 3.10 (m, 4H), 2.12 – 1.77 (m, 11H).; **LC/MS** (Method B): Retention time 6.87 min, m/z (ESI+) = 837.6 [M-H]⁺.

Compound 62

4-((4-(A-(N-(pent-4-yn-1-yl)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamido)butanamido)butyl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Compound 46 (1.2 equiv, 20 mg, 0.05 mmol) was sheltered from light and dissolved in DMF (0.6mL) and DIEA (3 equiv, 0.02 mL, 0.14 mmol) was added under argon. The mixture was stirred 5 min under argon and HBTU (1.5 equiv, 24 mg, 0.06 mmol) was slowly added. The mixture was stirred 5 min under argon and **PDC-4-NH2** (1.0 equiv, 22 mg, 0.04 mmol) was added. The reaction mixture was

stirred under argon for 16 h and volatiles were removed under vacuum. The crude product was purified by SiO_2 gel chromatography eluted with DCM-EtOH from 0 to 5%, yielding **62** as a white solid (32 mg, 85%).

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.29 (s, 2H), 9.36 (d, *J* = 2.5 Hz, 2H), 8.97 (d, *J* = 2.5 Hz, 2H), 8.04 (d, *J* = 8.3 Hz, 4H), 7.73 (t, *J* = 7.1 Hz, 3H), 7.65 (d, *J* = 7.5 Hz, 2H), 7.59 – 7.43 (m, 4H), 7.37 – 7.23 (m, 2H), 3.31 – 2.96 (m, 8H), 2.81 (s, 1H), 2.24 – 2.09 (m, 2H), 2.02 – 1.87 (m, 2H), 1.82 – 1.41 (m, 8H).; **LC/MS** (Method B): Retention time 7.01 min, m/z (ESI+) = 869.7 [M-H]⁺.

Compound 63

4-((4-(4-azido-2,3,5,6-tetrafluoro-N-(pent-4-yn-1-yl)benzamido)butanamido)butyl)amino)-N2,N6di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Compound 47 (1.2 equiv, 28 mg, 0.07 mmol) was sheltered from light and dissolved in DMF (0.8 mL) and DIEA (3 equiv, 0.03 mL, 0.18 mmol) was added under argon. The mixture was stirred 5 min under argon and HBTU (1.5 equiv, 34 mg, 0.09 mmol) and HOBt (0.5 equiv, 4.1 mg, 0.03 mmol) were added. The mixture was stirred 5 min under argon and **PDC-4-NH2** (1.0 equiv, 31 mg, 0.06 mmol) was added. The reaction mixture was stirred under argon for 16 h and volatiles were removed under vacuum. The crude product was purified by SiO₂ gel chromatography eluted with DCM-EtOH from 0 to 5%, yielding **63** as a white solid (23 mg, 44%).

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 10.34 (s, 2H), 9.07 (s, 2H), 8.68 (s, 2H), 7.92 (t, *J* = 7.3 Hz, 2H), 7.62 (d, *J* = 8.2 Hz, 2H), 7.58 – 7.51 (m, 2H), 7.43 (t, *J* = 7.2 Hz, 2H), 7.30 (s, 2H), 7.10 (t, *J* = 5.4 Hz, 1H), 6.09 (s, 1H), 3.50 – 3.38 (m, 2H), 3.34 – 3.03 (m, 6H), 2.18 – 2.05 (m, 2H), 2.02 – 1.81 (m, 4H), 1.76 (d, *J* = 6.6 Hz, 1H), 1.71 (t, *J* = 2.4 Hz, 1H), 1.68 – 1.46 (m, 6H).; **LC/MS** (Method B): Retention time 6.96 min, m/z (ESI+) = 874.5 [M-H]⁺.

Compound 64

4-((4-(4-(4-benzoyl-N-(pent-4-yn-1-yl)benzamido)butanamido)butyl)amino)-N2,N6-di(quinolin-3yl)pyridine-2,6-dicarboxamide



Compound 48 (1 equiv, 133 mg, 0.35 mmol) was sheltered from light and dissolved in DMF (2.7 mL) and TEA (1 equiv, 0.05 mL, 0.35 mmol) was added under argon. The mixture was stirred 5 min under argon and EDCI (1 equiv, 68 mg, 0.35 mmol) and HOBt (0.4 equiv, 19 mg, 0.14 mmol) were added. The mixture was stirred 5 min under argon and **PDC-4-NH2** (1.0 equiv, 178 mg, 0.35 mmol) was added. The reaction mixture was stirred under argon for 16 h and volatiles were removed under vacuum. The crude product was purified by SiO₂ gel chromatography eluted with DCM-MeOH from 0 to 5%, yielding **64** as a white solid (184 mg, 60%).

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.28 (s, 2H), 9.35 (d, *J* = 2.4 Hz, 2H), 8.96 (d, *J* = 2.2 Hz, 2H), 8.04 (d, *J* = 8.3 Hz, 4H), 7.83 – 7.69 (m, 6H), 7.70 – 7.61 (m, 4H), 7.59 – 7.48 (m, 6H), 3.55 – 3.39 (m, 3H), 3.28 – 3.09 (m, 5H), 3.02 (bs, 1H), 2.19 – 2.11 (m, 1H), 2.05 – 1.89 (m, 2H), 1.88 – 1.40 (m, 8H).; **LC/MS** (Method B): Retention time 6.72 min, m/z (ESI+) = 865.6 [M-H]⁺.

Compound 65

4-((17-oxo-13-(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoyl)-4,7,10-trioxa-13,18-diazadocos-1-yn-22-yl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Compound 56 (1.2 equiv, 74 mg, 0.15 mmol) was sheltered from light and dissolved in DMF (2.0 mL) and DIEA (4.5 equiv, 0.09 mL, 0.57 mmol) was added under argon. The mixture was stirred 5 min under argon and HATU (1.8 equiv, 87 mg, 0.23 mmol) was slowly added. The mixture was stirred 5

min under argon and **PDC-4-NH2** (1.0 equiv, 64 mg, 0.13 mmol) was added. The reaction mixture was stirred under argon for 16 h and volatiles were removed under vacuum. The crude product was purified by SiO₂ gel chromatography eluted with DCM-EtOH from 0 to 5%, yielding **65** as a colorless solid (70 mg, 57%).

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 10.27 (s, 2H), 9.11 (s, 2H), 8.80 (s, 2H), 8.03 (d, *J* = 8.4 Hz, 2H), 7.73 (d, *J* = 7.9 Hz, 2H), 7.62 (t, *J* = 7.4 Hz, 2H), 7.51 (t, *J* = 7.4 Hz, 2H), 7.38 – 7.27 (m, 4H), 7.03 (d, *J* = 8.0 Hz, 2H), 5.92 (s, 1H), 4.14 (d, *J* = 2.2 Hz, 2H), 3.72 – 3.09 (m, 20H), 2.42 (t, *J* = 2.2 Hz, 1H), 2.26 (t, *J* = 6.3 Hz, 1H), 2.02 – 1.88 (m, 2H), 1.77 – 1.56 (m, 4H).; **LC/MS** (Method B): Retention time 7.15 min, m/z (ESI+) = 973.5 [M-H]⁺.

Compound 66

4-((13-(4-azido-2,3,5,6-tetrafluorobenzoyl)-17-oxo-4,7,10-trioxa-13,18-diazadocos-1-yn-22-yl)amino)-N2,N6-di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Compound 57 (1.2 equiv, 50 mg, 0.10 mmol) was sheltered from light and dissolved in DMF (1.22 mL) and DIEA (4.5 equiv, 0.06 mL, 0.38 mmol) was added under argon. The mixture was stirred 5 min under argon and HATU (1.8 equiv, 58 mg, 0.15 mmol) was slowly added. The mixture was stirred 5 min under argon and **PDC-4-NH2** (1.0 equiv, 43 mg, 0.09 mmol) was added. The reaction mixture was stirred under argon for 16 h and volatiles were removed under vacuum. The crude product was purified by SiO₂ gel chromatography eluted with DCM-MeOH from 0 to 5%, yielding **66** as a colorless solid (35 mg, 42%).

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 10.39 (s, 2H), 9.06 (s, 2H), 8.63 (s, 2H), 7.87 (t, *J* = 7.0 Hz, 2H), 7.64 – 7.48 (m, 4H), 7.38 (t, *J* = 7.1 Hz, 2H), 7.22 (bs, 1H), 6.13 (bs, 1H), 4.08 (d, *J* = 2.3 Hz, 1H), 3.64 – 3.30 (m, 14H), 3.29 – 3.17 (m, 4H), 3.15 – 2.99 (m, 2H), 2.40 (t, *J* = 2.3 Hz, 1H), 2.14 (t, *J* = 5.9 Hz, 1H), 2.07 – 1.91 (m, 1H), 1.90 – 1.78 (m, 1H), 1.79 – 1.67 (m, 1H), 1.45 – 1.20 (m, 4H).; ¹³C NMR (75 MHz, CDCl₃) δ (ppm): 172.67, 172.08, 163.33, 163.27, 160.34, 159.56, 156.12, 148.84, 144.93, 144.88, 144.12, 142.06, 140.81, 138.82, 131.50, 128.64, 128.14, 127.83, 127.35, 124.65, 124.49, 121.13, 111.50, 100.19, 79.63, 74.73, 70.62, 70.47, 70.35, 70.23, 69.12, 69.05, 68.17, 58.42, 53.56, 48.73, 45.22,

45.11, 42.64, 39.17, 33.26, 27.16, 26.01, 23.38.; **LC/MS** (Method B): Retention time 7.16 min, m/z (ESI+) = 978.5 [M-H]⁺.;

Compound 67

4-((13-(4-benzoylbenzoyl)-17-oxo-4,7,10-trioxa-13,18-diazadocos-1-yn-22-yl)amino)-N2,N6di(quinolin-3-yl)pyridine-2,6-dicarboxamide



Compound 58 (1.2 equiv, 110 mg, 0.23 mmol) was sheltered from light and dissolved in DMF (2.8 mL) and DIEA (4.5 equiv, 0.14 mL, 0.86 mmol) was added under argon. The mixture was stirred 5 min under argon and HATU (1.8 equiv, 130 mg, 0.34 mmol) was slowly added. The mixture was stirred 5 min under argon and **PDC-4-NH2** (1.0 equiv, 96 mg, 0.19 mmol) was added. The reaction mixture was stirred under argon for 16 h and volatiles were removed under vacuum. The crude product was purified by SiO₂ gel chromatography eluted with DCM-EtOH from 0 to 5%, yielding **67** as a colorless solid (116 mg, 63%).

¹H NMR (300 MHz, CDCl₃) δ (ppm): 10.30 (s, 2H), 9.11 (s, 2H), 8.75 (s, 2H), 8.00 (d, J = 8.3 Hz, 2H), 7.70 (d, J = 7.4 Hz, 4H), 7.67 – 7.29 (m, 14H), 5.93 (t, J = 4.5 Hz, 1H), 4.10 (d, J = 2.1 Hz, 2H), 3.71 – 3.29 (m, 16H), 3.19 (s, 3H), 2.40 (t, J = 2.3 Hz, 1H), 2.28 (t, J = 6.3 Hz, 1H), 2.05 – 1.86 (m, 6H).; ¹³C NMR (75 MHz, CDCl₃) δ (ppm): 195.95, 172.82, 172.09, 163.27, 156.02, 148.80, 145.03, 145.00, 139.90, 138.08, 136.94, 132.75, 131.44, 130.03, 128.70, 128.38, 128.04, 127.73, 127.18, 126.66, 124.49, 100.00, 79.54, 74.76, 70.50, 70.43, 70.33, 70.21, 69.00, 68.44, 58.33, 48.81, 44.75, 42.56, 39.07, 33.61, 27.08, 25.99, 23.48.; LC/MS (Method B): Retention time 6.77 min, m/z (ESI+) = 969.7 [M-H]⁺.

PDC-C1-N2 (68)

*3,3'-((4-((4-(N-(prop-2-yn-1-yl)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamido)butanamido)*butyl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



MeI (100 equiv, 0.15 mL, 2.4 mmol) was added slowly to a solution of **59** (1.0 equiv, 20 mg, 0.02 mmol) in DMF (0.2 mL) under argon atmosphere at 40°C sheltered from the light. The reaction mixture was stirred for 16h at 40°C. Volatiles were removed under vacuum and resulting solid was suspended in EtOH, washed several times with EtOH and then with Et_2O , affording **PDC-C1-N2** as a yellow solid (26 mg, 97%)

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.72 (s, 2H), 10.11 (s, 2H), 9.64 (s, 2H), 8.56 (d, *J* = 8.8 Hz, 4H), 8.25 (d, *J* = 7.3 Hz, 2H), 8.19 – 8.12 (m, 2H), 8.07 (d, *J* = 7.5 Hz, 3H), 7.71 (t, *J* = 4.7 Hz, 1H), 7.53 (s, 2H), 7.34 (s, 2H), 4.78, (s, 6H), 4.27 (bs, 1H), 3.97 (bs, 1H), 3.48 (bs, 1H), 3.31 – 2.96 (m, 6H), 2.14 (bs, 1H), 2.01 – 1.73 (m, 3H), 1.69 – 1.41 (m, 4H).; **LC/MS** (Method B): Retention time 5.09 min, m/z (ESI+) = 869.5.

PDC-C1-N3 (69)

3,3'-((4-((4-(4-(4-azido-2,3,5,6-tetrafluoro-N-(prop-2-yn-1-yl)benzamido)butanamido)butyl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



MeI (100 equiv, 1.25 mL, 20.1 mmol) was added slowly to a solution of **60** (1.0 equiv, 170 mg, 0.2 mmol) in DMF (1.5 mL) under argon atmosphere at 40°C sheltered from the light. The reaction mixture was stirred for 16h at 40°C. Volatiles were removed under vacuum and resulting solid was suspended in EtOH, washed several times with EtOH and then with Et₂O, affording **PDC-C1-N3** as a yellow solid (214 mg, 94%)

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): δ 11.73 (s, 2H), 10.12 (s, 2H), 9.65 (s, 2H), 8.56 (d, *J* = 8.7 Hz, 4H), 8.25 (t, *J* = 7.9 Hz, 2H), 8.09 (t, *J* = 7.6 Hz, 2H), 7.94 – 7.80 (m, 1H), 7.71 (t, *J* = 4.7 Hz, 1H), 7.61 (s, 2H), 4.79 (s, 6H), 4.34 (d, *J* = 1.9 Hz, 1H), 4.14 (d, *J* = 1.9 Hz, 1H), 3.57 (t, *J* = 7.4 Hz, 2H), 3.20 – 3.00 (m, 3H), 2.19 – 2.09 (m, 1H), 2.00 (t, *J* = 6.4 Hz, 1H), 1.86 (pent, *J* = 7.2 Hz, 1H), 1.81 – 1.39 (m, 6H).; **LC/MS** (Method B): Retention time 4.55 min, m/z (ESI+) = 874.6.

PDC-C1-Bp (70)

3,3'-((4-((4-(4-(4-benzoyl-N-(prop-2-yn-1-yl)benzamido)butanamido)butyl)amino)pyridine-2,6dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



MeI (100 equiv, 1.25 mL, 20.1 mmol) was added slowly to a solution of **61** (1.0 equiv, 170 mg, 0.2 mmol) in DMF (1.5 mL) under argon atmosphere at 40°C sheltered from the light. The reaction mixture was stirred for 16h at 40°C. Volatiles were removed under vacuum and resulting solid was suspended in EtOH, washed several times with EtOH and then with Et₂O, affording **PDC-C1-Bp** as a yellow solid (170 mg, 100%)

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.72 (s, 2H), 10.11 (s, 2H), 9.64 (s, 2H), 8.56 (d, J = 8.7 Hz, 4H), 8.26 (d, J = 7.8 Hz, 2H), 8.09 (t, J = 7.6 Hz, 2H), 7.83 – 7.75 (m, 4H), 7.74 – 7.65 (m, 2H), 7.61 – 7.48 (m, 5H), 4.79 (s, 6H), 4.32 (bs, 1H), 4.05 (bs, 1H), 3.37 – 2.96 (m, 6H), 2.18 (s, 1H), 2.06 – 1.74 (m, 4H), 1.70 – 1.39 (m, 4H).; **LC/MS** (Method B): Retention time 5.25 min, m/z (ESI+) = 865.5; **HRMS** (ESI, m/z) calcd. for formula C₅₂H₅₀IN₈O₅⁺ [M-I]⁺ = 993.2949; found = 993.2976.

PDC-C3-N2 (71)

3,3'-((4-((4-(N-(pent-4-yn-1-yl)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamido)butanamido)butyl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



MeI (140 equiv, 0.32 mL, 5.0 mmol) was added slowly to a solution of **62** (1.0 equiv, 31 mg, 0.04 mmol) in DMF (0.3 mL) under argon atmosphere at 40°C sheltered from the light. The reaction mixture was stirred for 16h at 40°C. Volatiles were removed under vacuum and resulting solid was suspended in EtOH, washed several times with EtOH and then with Et₂O, affording **PDC-C3-N2** as a yellow solid (24 mg, 58%)

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.73 (s, 2H), 10.09 (d, *J* = 1.3 Hz, 1H), 9.62 (d, *J* = 1.7 Hz, 2H), 8.55 (dd, *J* = 8.3, 4.9 Hz, 4H), 8.24 (dt, *J* = 7.7, 1.2 Hz, 2H), 8.08 (t, *J* = 7.7 Hz, 2H), 7.99 – 7.58 (m, 3H), 7.48 (d, *J* = 8.3 Hz, 2H), 7.39 – 7.23 (m, 2H), 4.78 (s, 6H), 3.28 – 3.06 (m, 6H), 3.01 (bs, 1H), 2.82 (bs, 1H), 2.23 – 2.09 (m, 2H), 2.03 – 1.95 (m, 3H), 1.84 – 1.52 (m, 7H), 1.52 – 1.40 (m, 1H).; **LC/MS** (Method B): Retention time 5.54 min, m/z (ESI+) = 897.7.

PDC-C3-Bp (73)

3,3'-((4-((4-(4-benzoyl-N-(pent-4-yn-1-yl)benzamido)butanamido)butyl)amino)pyridine-2,6dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



MeI (140 equiv, 1.0 mL, 16.4 mmol) was added slowly to a solution of **64** (1.0 equiv, 79 mg, 0.09 mmol) in DMF (0.9 mL) under argon atmosphere at 40°C sheltered from the light. The reaction mixture was stirred for 16h at 40°C. Volatiles were removed under vacuum and resulting solid was

suspended in EtOH, washed several times with EtOH and then with Et₂O, affording **PDC-C3-Bp** as a yellow solid (127 mg, 95%)

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm) : 11.72 (s, 2H), 10.09 (s, 2H), 9.62 (s, 2H), 8.55 (d, *J* = 9.2 Hz, 4H), 8.25 (t, *J* = 8.1 Hz, 2H), 8.09 (d, *J* = 7.6 Hz, 2H), 7.95 (s, 1H), 7.81 – 7.65 (m, 6H), 7.62 – 7.47 (m, 5H), 4.78 (s, 6H), 3.30 – 3.11 (m, 6H), 3.07 – 2.97 (m, 2H), 2.63 (s, 1H), 2.24 – 2.11 (m, 2H), 2.09 – 1.89 (m, 2H), 1.87 – 1.41 (m, 8H).; **LC/MS** (Method B): Retention time 4.94 min, m/z (ESI+) = 893.6.

PDC-TEG-N2 (74)

3,3'-((4-((17-oxo-13-(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoyl)-4,7,10-trioxa-13,18-diazadocos-1-yn-22-yl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



MeI (100 equiv, 0.32 mL, 5.14 mmol) was added slowly to a solution of **65** (1.0 equiv, 50 mg, 0.05 mmol) in DMF (0.4 mL) under argon atmosphere at 40°C sheltered from the light. The reaction mixture was stirred for 16h at 40°C. Volatiles were removed under vacuum and resulting solid was suspended in EtOH, washed several times with EtOH and then with Et₂O, affording **PDC-TEG-N2** as a yellow solid (48 mg, 74%)

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 11.53 (s, 2H), 11.33 (s, 2H), 9.91 (s, 2H), 8.18 (d, J = 8.6 Hz, 2H), 8.11 (d, J = 9.2 Hz, 2H), 8.00 (t, J = 7.4 Hz, 2H), 7.86 (t, J = 8.0 Hz, 2H), 7.56 – 7.36 (m, 4H), 6.91 (s, 1H), 5.99 (s, 1H), 4.83 (s, 6H), 4.19 (bs, 2H), 3.84 – 3.50 (m, 12H), 3.46 – 3.20 (m, 6H), 2.46 (bs, 1H), 2.37 – 2.27 (m, 2H), 2.08 – 1.95 (m, 2H), 1.80 – 1.69 (m, 4H).; **LC/MS** (Method B): Retention time 5.81 min, m/z (ESI+) = 1001.6; **HRMS** (ESI, *m/z*) calcd. for formula C₅₃H₅₇IN₁₀O₇F₃⁺ [M-I]⁺ = 1129.3409; found = 1129.3411.

PDC-TEG-N3 (75)

3,3'-((4-((13-(4-azido-2,3,5,6-tetrafluorobenzoyl)-17-oxo-4,7,10-trioxa-13,18-diazadocos-1-yn-22yl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



MeI (100 equiv, 0.25 mL, 4.1 mmol) was added slowly to a solution of **66** (1.0 equiv, 40 mg, 0.04 mmol) in DMF (0.3 mL) under argon atmosphere at 40°C sheltered from the light. The reaction mixture was stirred for 16h at 40°C. Volatiles were removed under vacuum and resulting solid was suspended in EtOH, washed several times with EtOH and then with Et₂O, affording **PDC-TEG-N3** as a yellow solid (41 mg, 79%)

¹**H NMR** (300 MHz, CDCl₃) δ (ppm): 11.49 (s, 2H), 11.32 (s, 2H), 9.91 (s, 2H), 8.20 (d, J = 8.2 Hz, 2H), 8.12 (d, J = 8.8 Hz, 2H), 8.02 (d, J = 8.0 Hz, 2H), 7.87 (t, J = 7.4 Hz, 2H), 7.42 (s, 2H), 6.54 (s, 1H), 5.93 (s, 1H), 4.83 (s, 6H), 4.19 (t, J = 2.3 Hz, 2H), 3.84 – 3.50 (m, 14H), 3.46 – 3.21 (m, 4H), 2.46 (t, J = 2.9 Hz, 1H), 2.34 – 2.26 (m, 2H), 2.08 – 1.93 (m, 2H), 1.80 - 1.66 (m, 4H).; **LC/MS** (Method B): Retention time 6.34 min, m/z (ESI+) = 1006.7.

PDC-TEG-Bp (76)

3,3'-((4-((13-(4-benzoylbenzoyl)-17-oxo-4,7,10-trioxa-13,18-diazadocos-1-yn-22-yl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide



MeI (100 equiv, 0.2 mL, 3.2 mmol) was added slowly to a solution of **67** (1.0 equiv, 31 mg, 0.03 mmol) in DMF (0.25 mL) under argon atmosphere at 40°C sheltered from the light. The reaction mixture was stirred for 16h at 40°C. Volatiles were removed under vacuum and resulting solid was suspended in EtOH, washed several times with EtOH and then with Et₂O, affording **PDC-TEG-Bp** as a yellow solid (25 mg, 62%)

¹H NMR (300 MHz, CDCl₃) δ (ppm): 11.42 (s, 2H), 11.25 (s, 2H), 9.80 (s, 2H), 8.16 – 8.05 (m, 4H), 7.96 (t, *J* = 7.9 Hz, 2H), 7.89 – 7.73 (m, 7H), 7.66 – 7.40 (m, 6H), 7.31 (s, 2H), 6.18 (s, 1H), 4.79 (s, 6H), 4.15 (d, *J* = 1.6 Hz, 2H), 3.84 – 3.45 (m, 16H), 3.35 (bs, 2H), 3.27 – 3.11 (m, 3H), 2.44 (t, *J* = 2.2 Hz, 1H), 2.38 – 2.30 (m, 1H), 2.03 (s, 2H), 1.82 – 1.70 (m, 3H).; ¹³C NMR (75 MHz, CDCl₃) δ (ppm): 196.36, 172.99, 172.13, 163.04, 156.20, 146.31, 140.66, 138.39, 137.35, 134.87, 133.71, 133.08, 132.86, 132.59, 130.54, 130.35, 130.17, 129.67, 128.75, 127.21, 117.82, 79.84, 70.87, 70.69, 69.35, 68.85, 58.67, 45.91, 39.25.; LC/MS (Method B): Retention time 5.33 min, m/z (ESI+) = 997.6; HRMS (ESI, *m/z*) calcd. for formula C₅₈H₆₂IN₈O₈⁺ [M-I]⁺ = 1125.3735; found = 1125.3689.

Compound 78

4-chloroquinolin-2-amine

H₂N N C

2-aminoquinolin-4-ol (1 equiv, 1002 mg, 6.26 mmol) and POCl₃ (10.3 equiv., 9870 mg, 6 mL, 64.4 mmol) were introduced in a flask and agitated at 120°C for 22h. After completion of the reaction, excess of POCl₃ was carefully removed under vacuum. 20 mL of boiling water was added slowly and the mixture was agitated at 120°C for 1h before NaOH (6N) was added until reaching pH = 14. The mixture was cooled down to room temperature and chloroform was added. Layers were separated and aqueous layer was extracted with chloroform. Combined organic layers were dried over MgSO₄ and solvent was evaporated to obtain compound **78** as a white solid (742 mg, 66%).

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 7.91 – 7.82 (m, 1H), 7.60 - 7.54 (m, 1H), 7.54 – 7.48 (m, 1H), 7.27 (ddd, *J* = 8.1, 6.7, 1.4 Hz, 1H), 6.96 (s, 1H), 6.65 (s, 2H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 158.0, 148.8, 140.9, 130.4, 125.7, 123.4, 122.2, 120.3, 111.8; **LC/MS** (Method B): Retention time 4.35 min, m/z (ESI+) = 179.2 [M-H]⁺.

2-amino-4-pyroquinoline (80)

4-(2-(pyrrolidin-1-yl)ethoxy)quinolin-2-amine

Potassium tert-butoxide (6.6 equiv, 6220 mg, 55.4 mmol) and THF (20 mL) were agitated at room temperature for 5 min. N-(2-hydroxyethyl)pyrrolidine (10.1 equiv, 10.0 mL, 85.1 mmol) was added to the heterogeneous solution and agitated at 95 °C. Previously synthesized compound **78** (1 equiv, 1500 mg, 8.4 mmol) was solubilised in THF (5 mL) and was added dropwise on reaction mixture at 95 °C. After 24 h, the reaction mixture was cooled down to room temperature and was diluted with DCM. The apparition of a thin solid was observed and separated by centrifugation. Organic layer was recovered and treated once with sat. NaHCO3 and sat. NaCl. Combined organic layers were dried over MgSO4 and solvent was evaporated. Resulting product was purified by Al₂O₃ flash chromatography eluted by AcOEt-EtOH from 0 to 10% yielding expected compound as a yellow solid (1090 mg, 50%)

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 7.85 – 7.76 (m, 1H), 7.48 – 7.40 (m, 1H), 7.40 – 7.35 (m, 1H), 7.15 – 7.06 (m, 1H), 6.27 (s, 2H), 6.19 (s, 1H), 4.19 (t, *J* = 5.6 Hz, 2H), 2.92 (t, *J* = 5.6 Hz, 2H), 2.62 – 2.53 (m, 4H), 1.74 – 1.65 (m, 4H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 160.8, 159.5, 148.9, 129.4, 124.9, 121.2, 120.4, 116.8, 90.6, 67.3, 54.2, 53.9, 23.2; **LC/MS** (Method B): Retention time 1.58 min, m/z (ESI+) = 258.3 [M-H]⁺.

3-amino-4-pyroquinoline (82)

4-(2-(pyrrolidin-1-yl)ethoxy)quinolin-3-amine

 NH_2

Potassium tert-butoxide (6.4 equiv, 4900 mg, 53.5 mmol) and THF (20 mL) were agitated at r.t. for 5 min. N-(2-hydroxyethyl)pyrrolidine (5.1 equiv, 5.0 mL, 42.5 mmol) was added to the heterogeneous solution and agitated at 95 °C. 4-chloroquinolin-3-amine (1 equiv, 1500 mg, 8.4 mmol) was solubilised in THF (7 mL) and was added dropwise on reaction mixture at 95 °C. After 2 days, the reaction mixture was cooled down to r.t. and was diluted with DCM. The apparition of a thin solid was observed and separated by centrifugation. Organic layer was recovered and treated once with sat. NaHCO3 and sat. NaCl. Combined organic layers were dried over MgSO4 and solvent was evaporated. Resulting product was purified by SiO₂ flash chromatography eluted by AcOEt-EtOH-TEA (90/9/1) yielding expected compound as a light brown solid (733 mg, 34%)

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 8.53 (s, 1H), 7.94 – 7.85 (m, 1H), 7.85 – 7.76 (m, 1H), 7.49 – 7.30 (m, 2H), 5.76 (s, 2H), 4.09 (t, *J* = 5.4 Hz, 2H), 2.79 (t, *J* = 5.3 Hz, 2H), 2.61 – 2.53 (m, 4H), 1.79 –

1.67 (m, 4H); ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 144.4, 142.4, 141.6, 134.7, 128.9, 126.2, 124.6, 124.2, 119.8, 71.3, 55.2, 53.7, 23.2; **LC/MS** (Method B): Retention time 1.26 min, m/z (ESI+) = 258.3 [M-H]⁺.

PDC-Pyro (85)

N2,N6-bis(4-(2-(pyrrolidin-1-yl)ethoxy)quinolin-3-yl)pyridine-2,6-dicarboxamide



2,6-pyridinedicarboxylic acid (1 equiv, 20 mg, 0.12 mmol) was mixed with DMF 3 mL, HOBt (2 equiv, 32 mg, 0.24 mmol), DIPEA (5 equiv, 0.1 mL, 0.6 mmol) and with previously synthesized 3-amino-4-pyroquinoline (3 equiv, 92 mg, 0.36 mmol). HATU (5 equiv, 227 mg, 0.6 mmol) was next added slowly and reaction mixture was stirred under argon atmosphere for 16 h. Volatiles were next removed under vacuum and crude mixture was directly purified by prep-HPLC to afford **PDC-Pyro** as a white powder (42 mg, 48%).

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.24 (s, 2H), 8.89 (s, 2H), 8.51 (d, *J* = 6.9 Hz, 2H), 8.46 – 8.34 (m, 3H), 8.09 (d, *J* = 8.4 Hz, 2H), 7.86 (t, *J* = 7.2 Hz, 2H), 7.73 (t, *J* = 7.3 Hz, 2H), 4.55 (t, *J* = 4.1 Hz, 4H), 3.67 (t, *J* = 4.3 Hz, 4H), 2.01 – 1.78 (m, 8H).; ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 163.14, 155.30, 152.14, 148.20, 147.70, 140.43, 130.10, 128.95, 126.99, 125.80, 122.78, 122.49, 119.33, 68.32, 53.81, 53.68, 22.42.; **LC/MS** (Method B): Retention time 4.31 min, m/z (ESI+) = 646.5 [M-H]⁺; **HRMS** (ESI, *m/z*) calcd. for formula $C_{37}H_{40}N_7O_4^+$ [M-H]⁺ = 646.3142; found = 646.3149.

PDC2-Pyro (86)

N2,N6-bis(4-(2-(pyrrolidin-1-yl)ethoxy)quinolin-2-yl)pyridine-2,6-dicarboxamide


2,6-pyridinedicarboxylic acid (1 equiv, 10 mg, 0.06 mmol) was mixed with DMF (1 mL), HOBt (2 equiv, 16 mg, 0.12 mmol), DIPEA (5 equiv, 0,05 mL, 0.03 mmol) and with pyro-2-aminoquinoline (2,5 equiv, 39 mg, 0.15 mmol). Mixture was stirred and HATU (5 equiv, 114 mg, 0.3 mmol) was slowly added. Reaction mixture was stirred at room temperature under argon atmosphere for 16 h. Volatiles were next removed under vacuum and crude mixture was directly purified by prep-HPLC to afford **PDC2-Pyro** as a white powder (15 mg, 34%).

¹**H NMR** (300 MHz, MeOH-*d4*) δ (ppm): 8.54 (d, *J* = 7.8 Hz, 2H), 8.41 (bs, 2H), 8.33 (d, *J* = 7.7 Hz, 1H), 8.26 (d, *J* = 7.9 Hz, 2H), 8.10 (s, 2H), 7.89 (d, *J* = 8.3 Hz, 2H), 7.73 (t, *J* = 7.3 Hz, 2H), 7.52 (t, *J* = 7.5 Hz, 2H), 4.72 (t, *J* = 4.6 Hz, 4H), 3.74 (t, *J* = 4.3 Hz, 4H), 3.47 – 3.38 (m, 8H), 2.20 – 2.04 (m, 8H).; ¹³**C NMR** (75 MHz, MeOH-*d4*) δ (ppm): 154.73, 154.35, 143.67, 140.47, 138.87, 131.62, 122.25, 118.54, 117.81, 116.50, 113.63, 111.33, 86.28, 58.77, 46.41, 45.93, 14.80.; **LC/MS** (Method B): Retention time 4.93 min, m/z (ESI+) = 646.6 [M-H]⁺; **HRMS** (ESI, *m/z*) calcd. for formula $C_{37}H_{40}N_7O_4^+$ [M-H]⁺ = 646.3142; found = 646.3168.

PhenDC3-Pyro (87)

N2,N9-bis(4-(2-(pyrrolidin-1-yl)ethoxy)quinolin-3-yl)-1,10-phenanthroline-2,9-dicarboxamide



1,10-phenanthroline-2,9-dicarboxylic acid (1 equiv, 64.8 mg, 0.24 mmol), previously synthesized 3amino-4-pyroquinoline (3.0 equiv, 184 mg, 0.72 mmol), EDC (3.1 equiv, 143 mg, 0.749 mmol) and HOBt (0.5 equiv, 16.9 mg, 0.125 mmol) were solubilized under argon atmosphere in DMF (5 mL) and stirred at room temperature for 24h. The reaction mixture was poured in water (25mL) and agitated for 5 min. The heterogeneous mixture was centrifuged, the majority of the liquid phase was removed and the solid was filtered, washed with water and dried under vacuum to affort **PhenDC3-Pyro** as a white solid (84 mg, 47%)

¹H NMR (300 MHz, DMSO-*d6*) δ (ppm): 11.51 (s, 2H), 9.79 (s, 2H), 8.92 (d, *J* = 8.3 Hz, 2H), 8.62 (d, *J* = 8.3 Hz, 2H), 8.33 (s, 2H), 8.06 (d, *J* = 8.4 Hz, 2H), 7.67 (t, *J* = 7.1 Hz, 2H), 7.30 – 7.23 (m, 2H), 7.22 – 7.14 (m, 2H), 4.02 (t, *J* = 5.2 Hz, 4H), 2.58 – 2.52 (m, 4H), 1.96 (bs, 8H), 0.94 (bs, 8H); ¹³C NMR (75)

MHz, DMSO-*d6*) δ (ppm): 162.5, 151.5, 149.1, 146.6, 146.3, 143.4, 139.0, 130.5, 129.1, 128.7, 128.4, 126.3, 123.7, 122.9, 121.6, 121.3, 73.3, 54.9, 53.2, 22.6.; **LC/MS** (Method B): Retention time 4.44 min, m/z (ESI+) = 747.6 [M-H]⁺.; **HRMS** (ESI, *m/z*) calcd. for formula C₄₄H₄₃N₈O₄⁺ [M-H]⁺ = 747.3407 ; found = 747.3438.

PhenDC2-Pyro (88)

N2,N9-bis(4-(2-(pyrrolidin-1-yl)ethoxy)quinolin-2-yl)-1,10-phenanthroline-2,9-dicarboxamide



1,10-phenanthroline-2,9-dicarboxylic acid (1 equiv, 54.2 mg, 0.20 mmol), 2-amino-4-pyroquinoline (2.6 equiv, 133 mg, 0.52 mmol), HATU (3.13 equiv, 240 mg, 0.63 mmol) and HOBt (0.25 equiv, 6.8 mg, 0.05 mmol) were solubilised under argon atmosphere in DMF (2 mL) and stirred at room temperature. After 3h, the reaction mixture was poored in water (25mL) and agitated for 5 min. The heterogeneous mixture was centrifuged, the majority of the liquid phase was removed and the solid was filtered, washed with water and dried under vacuum to affort **PhenDC2-Pyro** as a white solid (142 mg, 97%).

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 11.20 (s, 2H), 8.95 (d, *J* = 8.3 Hz, 2H), 8.68 (d, *J* = 8.1 Hz, 2H), 8.35 (s, 2H), 8.28 (d, *J* = 8.0 Hz, 2H), 8.22 (s, 2H), 7.49 (t, *J* = 6.6 Hz, 2H), 7.31 (s, 4H), 4.73 (bs, 4H), 3.85 (bs, 4H), 3.67 – 3.39 (m, 8H), 2.14 – 1.92 (m, 8H).; ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 163.0, 161.6, 151.3, 151.1, 148.7, 146.9, 143.5, 139.1, 131.1, 130.0, 128.8, 128.6, 127.4, 124.5, 121.9, 121.5, 120.7, 118.8, 94.1, 64.2, 54.3, 52.9, 22.6.; **LC/MS** (Method B): Retention time 4.32 min, m/z (ESI+) = 747.6 [M-H]⁺; **HRMS** (ESI, *m/z*) calcd. for formula $C_{44}H_{43}N_8O_4^+$ [M-H]⁺ = 747.3407 ; found = 747.3436.

2-amino-1-methylquinolinium (91)

2-aminoquinoline (1 equiv, 1.0 g, 7.0 mmol) was dissolved in a mixture of CH₃CN (40 mL) and MeI (4.6 equiv, 2 mL, 32.1 mmol). The solution was stirred at 45 °C for 16 h. After completion, volatiles

were evaporated, the resulting solid was solubilized in MeCN and Et₂O was slowly added. The solid was filtered, washed with Et₂O and dried under vacuum, affording 2-amino-1-methylquinolinium as a white solid (1.3 g, 65%)

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm): 9.12 (bs, 2H), 8.34 (d, *J* = 9.3 Hz, 1H), 8.05 – 7.94 (m, 2H), 7.93 – 7.81 (m, 1H), 7.57 (t, *J* = 7.5 Hz, 1H), 7.15 (d, J = 9.3 Hz, 1H), 3.87 (s, 3H) ; ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 155.0, 141.6, 137.5, 132.8, 129.7, 125.2, 122.1, 116.5, 114.8, 34.3.; **LC/MS** (Method B): Retention time 3.79 min, m/z (ESI+) = 159.2.

Pyridine-2,6-dicarbonyl dichloride



2,6-pyridinedicarboxylic acid (1 equiv, 41 mg, 0.25 mmol) was reacted with SOCl₂ (30 equiv, 0.53 mL, 7.36 mmol) and DMF (1 equiv, 0.02 mL, 0.25 mmol) under reflux and under argon atmosphere for 7h. The mixture was evaporated to dryness and dried under vacuum for 16 h yielding pyridine-2,6-dicarbonyl dichloride as a white solid and was used without purification.

PDC2 (90)

2,2'-((pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium)



Previously synthesized **2-amino-1-methylquinolinium** (3 equiv, 210.3 mg, 0.74 mmol) was mixed with DIPEA (5 equiv, 0.2 mL, 1.23 mmol) and DMF (1 mL). A solution of previously prepared pyridine-2,6-dicarbonyl dichloride (1 equiv, 50.0 mg, 0.25 mmol) in DMF (1 mL) was then added dropwise under argon. Reaction mixture was stirred for 4 h and volatiles were evaporated under vacuum. Crude mixture was purified by preparative HPLC to yield **PDC2** as a yellow solid (17.0 mg, 12%)

¹**H NMR** (300 MHz, DMSO-*d6*) δ (ppm) 8.60 (d, *J* = 7.6 Hz, 2H), 8.31 (t, *J* = 7.7 Hz, 1H), 8.19 (d, *J* = 8.9 Hz, 2H), 7.98 (d, *J* = 8.7 Hz, 2H), 7.92 – 7.78 (m, 6H), 7.52 (t, *J* = 7.3 Hz, 2H), 4.20 (s, 6H).; ¹³**C NMR** (75 MHz, DMSO-*d6*) δ (ppm): 171.21, 159.48, 153.05, 140.02, 138.52, 132.51, 129.10, 126.38, 125.13, 122.73, 118.51, 117.25, 33.77.; **LC/MS** (Method B): Retention time 4.88 min, m/z (ESI+) = 448.3; **HRMS** (ESI, *m/z*) calcd. for formula $C_{27}H_{22}N_5O_2^+$ [M-H]⁺ = 448.1774; found = 448.1778.

Biophysical experiments

Oligonucleotide sequences were purchased from Eurogentec as dried samples purified by HPLC-RP. tRNA sequence from bovine liver was purchased from Sigma as dried sample. Oligonucleotides were dissolved in MilliQ water, stored at -20 °C and concentrations were determined by absorption measurements at 260 nm at 95 °C using the molar extinction coefficient provided by Eurogentec. tRNA was dissolved at $\simeq 200$ mM in a 10 mM Caco.Li buffer previously sterilized. The final concentration was calculated using 321.47 g/mol as the average MW of a *ss*RNA nucleobase and 85 the average number of nucleobases of the sequence.

FRET-melting experiments

The experiments were carried out on pre-folded G-quadruplex structures: the sequences were heated at 95 °C for 10 min in opportune buffer and left to fold at 4 °C overnight. Duplex DNA and transfer RNA (tRNA) sequences were heated at 95 °C for 10 min and left to cool down to room temperature over a period of 3 h. Buffer conditions are specified below for each G4 sequence.

As previously reported, stabilization of ligands on G4-structures was monitored via FRET-melting assay performed in 96-well plates on a real-time PCR apparatus, 7900HT Fast Real-Time PCR system. This assay was carried out with oligonucleotides mimicking biological relevant G4-forming sequences, labeled with FRET partners at each end (5'-FAM (donor) and 3'-TAMRA (acceptor)) in the presence of G4-ligands. To determine ligand G4 vs duplex selectivity, an unlabeled DNA competitor ds26 was used. In order to assess RNA-G4 vs RNA single strand selectivity, an unlabeled tRNA sequence was used. This oligonucleotide consists of an average of 85 nucleobases and was used at $\approx 3\mu$ M (≈ 15 mol equiv, ≈ 60 nucleobases equiv). Each well was filled with: 0.20 μ M (1 equiv) labeled G4 DNA or RNA, 1 μ M (5 equiv) of ligand, and dsDNA competitor at 0, 3, and 10 μ M (0, 15, and 50 mol equiv) respectively, or RNA competitor at 0 and 3 μ M, in a total volume of 25 μ L. Solutions were prepared in buffer containing lithium cacodylate 10 mM, pH 7.2, and KCl 10 mM, completed by 90 mM LiCl for F21T, F21CTAT and 1 mM KCl, completed by 99 mM LiCl for all the others G-quadruplex sequences. The FAM channel was used to collect the fluorescence signal. Stabilization of compounds with G4-structures was monitored as followed: 5 min at 25 °C and then increase of 0.5 °C every minute until reaching 95 °C. The stabilization was evaluated by determining the temperature at half denaturation

of the G4 (obtained at the half height of the sigmoidal curve, $\Delta T1/2$) in the absence and in the presence of ligand, duplex DNA, and tRNA competitor, by using Origin 2021.

A series of oligonucleotides covering a range of possible G4 conformations was used.

FID experiments

The experiments were carried out on pre-folded G-quadruplex structures: the sequences were heated at 95 °C for 10 min (10 mM lithium cacodylate (pH = 7.3) and 100 mM KCl) and left to cool down to room temperature over a period of 3 h. Duplex DNA was heated at 95 °C for 10 min (10 mM lithium cacodylate (pH = 7.3) and 100 mM KCl) and left to cool down to room temperature over a period of 3 h.

a. Cuve experiments

FID-assay was performed on a Cary Eclipse spectrofluorometer, using 1 mL quartz cell as previously reported. 0.25 μ M of folded DNA prepared in a solution of 10 mM lithium cacodylate (pH = 7.3) and 100 mM KCl was mixed with 0.50 μ M Thiazole Orange (TO) for G4-DNA and 0.75 μ M for duplex DNA to obtain a total volume of 1 mL. The solution was left to stand in the dark for 30 min and then the titration was started. Each addition of ligand (from 0.5 to 10 equiv, stock solution 0.125 μ M in DMSO) was followed by a 3 min equilibration time and fluorescence emission spectra were recorded from 510 to 750 nm (λ exc = 495 nm, 1.0 nm increment, 0.1 s integration time, and slit 5 nm). The percentage of TO displacement was calculated as follows: TOD (%) = 100 – (F / F0) x 100, where F0 is the maximum fluorescence emission area of bound TO before addition of ligand and F is the fluorescence emission area after each ligand addition. The percentage of TO displacement was then plotted as a function of the concentration of added ligand.

b. High-throughput G4-FID

High-throughput FID experiments were carried in 96-well Black Flat bottom Polystyrene NBS microplates from Corning and measurements were performed on a FLUOstar Omega microplate reader from BMG Labtech. Microplates were filled as follows: from 190 to 90 μ L of K⁺100 buffer (to reach a final volume of 200 μ L), 10 μ L of a prefolded DNA solution (5 μ M) and TO (10 μ M for G4-DNA and 15 μ M for ds26), from 0 to 100 μ L of ligand (5 μ M in K⁺100 buffer). The final concentration of ligands from column A to H was then: 0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 1.0, 1.25, 1.5, 2.0 and 2.5 μ M. Finally, the microplate was gently shook for 5 min at 500 rpm and emission of TO fluorescence in each well was measurement at 520 nm after excitation at 485 nm. TO displacement was calculated and plotted as previously mentioned.

Circular Dichroism

CD analysis were carried out using a Jasco J-815 Spectropolarimeter and 1 mL quartz cells with a path length of 0.5 nm. Oligonucleotides were prefolded at 3 μ M in K⁺100 buffer for a total volume of 1 mL. CD spectra were recorded from 500 to 210 nm with 4 spectra accumulations, a scanning speed of 100 nm/min, a data pitch of 0.5 nm, a Digital Integration Time (D.I.T) of 1 s and a bandwidth of 2.0 nm. When needed, a solution of 2 mM G4 ligand in DMSO was added from 1 to 5 equiv (3 to 15 μ M) followed by 5 min equilibration time.

Data were treated using Origin 2021 and the ellipticity θ (mdeg) was converted in $\Delta \varepsilon$, which is the difference of molar absorption coefficients of the two light components left and right. $\Delta \varepsilon$ is expressed in M⁻¹.cm⁻¹ and is calculated using the formula $\frac{\theta}{32980*l*C}$ where I is the path length in cm and C the concentration in mol/L.

Biochemistry experiments

Sample preparation

Oligonucleotide sequences were purchased from Eurogentec as dried samples purified by PAGE and were dissolved in MilliQ water, stored at -20 °C. [γ -³²P] ATP was purchased from Perkin Elmer as a 25µL solution for a total radioactivity of 9.25 MBq. Oligonucleotide sequences were 5'-radiolabeled as followed: 1 µL of 20 µM PAGE-purified oligonucleotide was mixed with 2 µL MilliQ water, 1 µL T4-PNK buffer 10x, 3 µL [γ -³²P] ATP and 3 µL of T4-PNK enzyme. The reaction mixture was incubated at 37 °C for 1 h. The sequences 22AG and cMyc22 were purified by 20% acrylamide gel electrophoresis. The sequence Ceb25WT was diluted with MilliQ water, passed through a G25 Sephadex column and directly used without further purification.

Photo-alkylation experiments

Photochemical experiments were carried out in 1.5 mL Eppendorf tubes containing a prefolded solution of 10 μ M DNA and 1.0 to 2.5 μ L of 5'-end radiolabeled sequence depending on the amount of radioactivity in K⁺100 buffer for a total volume of 18 μ L. For competition experiments, previous solution was prepared in a total volume of 16 μ L and 2 μ L of prefolded duplex DNA (ds26) at appropriate concentration was added. Finally, 2 μ L of ligand in DMSO at various concentration was added and allowed to stand in dark for 30 min. After equilibration time, samples were irradiated using a Multirays photoreactor from Helios Quartz. After irradiation, 5 μ L of loading solution (bromophenol blue and xylene cyanol in formamide/water 8/2) was added. Samples were concentrated 15 min at 30 °C using a SpeedVac, loaded on a 15 or 20% denaturing gel electrophoresis and gels were runned for 5 to 6 h at 20 W. After that time, a europium screen was exposed to the gel for 16 to 24 h depending on the amount of radioactivity. The screen was finally analyzed using a Typhoon apparatus and images were treated using ImageQuant TL v8.0.

Références

- 1. Crick, F. & Watson, J. Molecular structure of nucleic acis. *Nature* **171**, (1953).
- 2. Ghosh, A. & Bansal, M. A glossary of DNA structures from A to Z. Acta Crystallogr. Sect. D Biol. Crystallogr. 59, 620–626 (2003).
- 3. García-Ramos, J. C., Galindo-Murillo, R., Cortés-Guzmán, F. & Ruiz-Azuara, L. Metal-based drug-DNA interactions. *J. Mex. Chem. Soc.* **57**, 245–259 (2013).
- 4. Polo, S. E. & Almouzni, G. Chromatin assembly: A basic recipe with various flavours. *Curr. Opin. Genet. Dev.* **16**, 104–111 (2006).
- 5. Rich, A. The Era of RNA Awakening: Structural biology of RNA in the early years. *Q. Rev. Biophys.* **42**, 117–137 (2009).
- 6. Schroeder, S. J. Challenges and approaches to predicting RNA with multiple functional structures. *RNA* **24**, 1615–1624 (2018).
- 7. Jacob, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190**, 576–581 (1961).
- Liu, J. Control of protein synthesis and mRNA degradation by microRNAs. *Curr. Opin. Cell Biol.* 20, 214–221 (2008).
- 9. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
- 10. Nguyen, T. H. D. *et al.* Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557**, 190–195 (2018).
- 11. Whelan, D. R. *et al.* Detection of an en masse and reversible B- to A-DNA conformational transition in prokaryotes in response to desiccation. *J. R. Soc. Interface* **11**, 3–8 (2014).
- 12. Dickerson, R. DNA Structure from A to Z. *Methods in Enzomology* **211**, (1992).
- 13. Temiz, N. A., Donohue, D. E., Bacolla, A., Luke, B. T. & Collins, J. R. The role of methylation in the intrinsic dynamics of B- and Z-DNA. *PLoS One* **7**, 1–9 (2012).
- 14. Bikard, D., Loot, C., Baharoglu, Z. & Mazel, D. Folded DNA in Action: Hairpin Formation and Biological Functions in Prokaryotes. *Microbiol. Mol. Biol. Rev.* **74**, 570–588 (2010).
- 15. Choi, J. & Majima, T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* **40**, 5893–5909 (2011).
- Zeraati, M. *et al.* I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* 10, 631–637 (2018).
- 17. Wright, E. P., Huppert, J. L. & Waller, Z. A. E. Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res.* **45**, 2951–2959 (2017).
- 18. Takahashi, S., Brazier, J. A. & Sugimoto, N. Topological impact of noncanonical DNA structures on Klenow fragment of DNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9605–9610 (2017).
- 19. M. Gellert, M. N. Lipsett, D. R. D. Helix formation by guanylic acid. 191–192 (1962). doi:10.1145/3084381.3084430

- 20. Sundquist, W. I. & Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342**, 825–829 (1989).
- 21. Bhattacharyya, D., Arachchilage, G. M. & Basu, S. Metal cations in G-quadruplex folding and stability. *Front. Chem.* **4**, 1–14 (2016).
- 22. Wei, D. *et al.* Crystal structure of a promoter sequence in the B-raf gene reveals an intertwined dimer quadruplex. *J. Am. Chem. Soc.* **135**, 19319–19329 (2013).
- 23. Phillips, K., Dauter, Z., Murchie, A. I. H., Lilley, D. M. J. & Luisi, B. The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution. *J. Mol. Biol.* **273**, 171–182 (1997).
- 24. Zhang, S., Wu, Y. & Zhang, W. G-quadruplex structures and their interaction diversity with ligands. *ChemMedChem* **9**, 899–911 (2014).
- 25. Tippana, R., Xiao, W. & Myong, S. G-quadruplex conformation and dynamics are determined by loop length and sequence. **42**, 8106–8114 (2014).
- 26. Joachimi, A., Benz, A. & Hartig, J. S. A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg. Med. Chem.* **17**, 6811–6815 (2009).
- 27. Tran, P. L. T., De Cian, A., Gros, J., Moriyama, R. & Mergny, J.-L. Tetramolecular Quadruplex Stability and Assembly. in *Quadruplex Nucleic Acids* 243–273 (2012). doi:10.1007/128_2012_334
- 28. Mukundan, V. T. & Phan, A. T. Bulges in G Quadruplexes: Broadening the Definition of G Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **135**, 5017–5028 (2013).
- 29. Fülscher, M. P., Serrano-Andrés, L. & Roos, B. O. A theoretical study of the electronic spectra of adenine and guanine. *J. Am. Chem. Soc.* **119**, 6168–6176 (1997).
- 30. Randazzo, A., Spada, G. P. & Webba da Silva, M. Circular Dichroism of Quadruplex Structures. in *Quadruplex Nucleic Acids* 15–31 (2012).
- 31. Gray, D. M. *et al.* Measured and Calculated CD Spectra of G-Quartets Stacked with the Same or Opposite Polarities. *Chirality* 431–440 (2008).
- 32. Karsisiotis, A. I., O'Kane, C. & Webba da Silva, M. DNA quadruplex folding formalism A tutorial on quadruplex topologies. *Methods* 64, 28–35 (2013).
- 33. Wang, Y. & Patel, D. J. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* **1**, 263–282 (1993).
- 34. Ambrus, A., Chen, D., Dai, J., Jones, R. A. & Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **44**, 2048–2058 (2005).
- 35. Luu, K. N., Phan, A. T., Kuryavyi, V., Lacroix, L. & Patel, D. J. Structure of the human telomere in K+ solution: An intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.* **128**, 9963–9970 (2006).
- 36. Schmid, M. Conformational dynamics of G-quadruplex DNA probed by time-resolved circular dichroism. *Univ. Paris-Saclay* **Optics**, (2017).
- 37. Kejnovská, I., Renčiuk, D., Palacký, J. & Vorlíčková, M. CD Study of the G-Quadruplex Conformation. *Methods Mol. Biol.* **2035**, 25–44 (2019).

- 38. Gray, R. D., Petraccone, L., Trent, J. O. & Chaires, J. B. Characterization of a K+-induced conformational switch in a human telomeric DNA oligonucleotide using 2-aminopurine fluorescence. *Biochemistry* **49**, 179–194 (2010).
- 39. Gray, R. D., Li, J. & Chaires, J. B. Energetics and kinetics of a conformational switch in Gquadruplex DNA. J. Phys. Chem. B **113**, 2676–2683 (2009).
- Vorlíčková, M., Chládková, J., Kejnovská, I., Fialová, M. & Kypr, J. Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Res.* 33, 5851–5860 (2005).
- 41. Parkinson, G. N., Lee, M. P. H. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876–880 (2002).
- 42. del Villar-Guerra, R., Trent, J. O. & Chaires, J. B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chemie Int. Ed.* **57**, 7171–7175 (2018).
- 43. Ou, A. *et al.* High resolution crystal structure of a KRASpromoter G-quadruplex reveals a dimer with extensive poly-A ?-stacking interactions for small-molecule recognition. *Nucleic Acids Res.* **1**, 1–11 (2020).
- 44. 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).
- 45. Papi, F. *et al.* Pyridine Derivative of the Natural Alkaloid Berberine as Human Telomeric G 4 DNA Binder: A Solution and Solid-State Study . *ACS Med. Chem. Lett.* **11**, 645–650 (2020).
- 46. Mendoza, O., Bourdoncle, A., Boulé, J. B., Brosh, R. M. & Mergny, J. L. G-quadruplexes and helicases. *Nucleic Acids Res.* **44**, 1989–2006 (2016).
- 47. Chen, M. C. *et al.* Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **558**, 465–483 (2018).
- 48. Voter, A. F., Qiu, Y., Tippana, R., Myong, S. & Keck, J. L. A guanine-flipping and sequestration mechanism for G-quadruplex unwinding by RecQ helicases. *Nat. Commun.* **9**, 1–8 (2018).
- 49. Parkinson, G. N. & Collie, G. W. X-Ray Crystallographic Studies of G-Quadruplex Structures. *Methods Mol. Biol.* **2035**, 131–155 (2019).
- 50. Adrian, M., Heddi, B. & Phan, A. T. NMR spectroscopy of G-quadruplexes. *Methods* **57**, 11–24 (2012).
- 51. Dai, J., Chen, D., Jones, R. A., Hurley, L. H. & Yang, D. NMR solution structure of the major Gquadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* **34**, 5133–5144 (2006).
- 52. Lin, C., Dickerhoff, J. & Yang, D. NMR Studies of G-Quadruplex Structures and G-Quadruplex-Interactive Compounds. *Methods Mol. Biol.* **2035**, 157–176 (2019).
- Phan, A. T., Kuryavyi, V., Luu, K. N. & Patel, D. J. Structure of two intramolecular Gquadruplexes formed by natural human telomere sequences in K+ solution. *Nucleic Acids Res.* 35, 6517–6525 (2007).
- 54. Agrawal, P., Hatzakis, E., Guo, K., Carver, M. & Yang, D. Solution structure of the major Gquadruplex formed in the human VEGF promoter in K+: Insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Res.* **41**, 10584–10592 (2013).

- 55. Krafcikova, M., Robert, H., Trantirek, L. & Foldynova-trantirkova, S. In Cell NMR Spectroscopy: Investigation of G-Quadruplex Structures. *Methods Mol. Biol.* **2035**, 397–405 (2019).
- 56. Krafčík, D. *et al.* Towards profiling of the g-quadruplex targeting drugs in the living human cells using nmr spectroscopy. *Int. J. Mol. Sci.* **22**, (2021).
- 57. Salgado, G. F., Cazenave, C., Kerkour, A. & Mergny, J.-L. G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy. *Chem. Sci.* **6**, 3314–3320 (2015).
- 58. Phan, A. T., Kuryavyi, V., Gaw, H. Y. & Patel, D. J. Small-molecule interaction with a fiveguanine-tract g-quadruplex structure from the human myc promoter. *Nat. Chem. Biol.* **1**, 167– 173 (2005).
- 59. Mathad, R. I., Hatzakis, E., Dai, J. & Yang, D. C-MYC promoter G-quadruplex formed at the 5'end of NHE III 1 element: Insights into biological relevance and parallel-stranded Gquadruplex stability. *Nucleic Acids Res.* **39**, 9023–9033 (2011).
- 60. Marquevielle, J. *et al.* Structure of two G-quadruplexes in equilibrium in the KRAS promoter. *Nucleic Acids Res.* 1–10 (2020).
- 61. Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S. & Patel, D. J. Structure of an unprecedented Gquadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.* **129**, 4386–4392 (2007).
- 62. Kuryavyi, V., Phan, A. T. & Patel, D. J. Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Res.* **38**, 6757–6773 (2010).
- 63. Kotar, A., Rigo, R., Sissi, C. & Plavec, J. Two-quartet kit* G-quadruplex is formed via doublestranded pre-folded structure. *Nucleic Acids Res.* **47**, 2641–2653 (2019).
- 64. Phan, A. T. Human telomeric G-quadruplex: Structures of DNA and RNA sequences. *FEBS J.* **277**, 1107–1117 (2010).
- 65. Amrane, S. *et al.* Formation of Pearl-Necklace Monomorphic G-Quadruplexes in the Human CEB25 Minisatellite. *J. Am. Chem. Soc.* **134**, 5807–5816 (2012).
- 66. Carvalho, J., Mergny, J. L., Salgado, G. F., Queiroz, J. A. & Cruz, C. G-quadruplex, Friend or Foe: The Role of the G-quartet in Anticancer Strategies. *Trends Mol. Med.* 1–14 (2020).
- 67. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916 (2005).
- 68. Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **33**, 2901–2907 (2005).
- 69. Guédin, A., Gros, J., Alberti, P. & Mergny, J. L. How long is too long? Effects of loop size on Gquadruplex stability. *Nucleic Acids Res.* **38**, 7858–7868 (2010).
- 70. Qin, M. *et al.* Two-quartet G-quadruplexes formed by DNA sequences containing four contiguous GG runs. *J. Phys. Chem. B* **119**, 3706–3713 (2015).
- 71. Ngoc Nguyen, T. Q., Lim, K. W. & Phan, A. T. Duplex formation in a G-quadruplex bulge. *Nucleic Acids Res.* **48**, 10567–10575 (2020).
- 72. Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. **34**, 3887–3896 (2006).

- 73. Yadav, V. K., Abraham, J. K., Mani, P., Kulshrestha, R. & Chowdhury, S. QuadBase : genomewide database of G4 DNA — occurrence and conservation in human , chimpanzee , mouse and rat promoters and 146 microbes. **36**, 381–385 (2008).
- 74. Bedrat, A., Lacroix, L. & Mergny, J. Re-evaluation of G-quadruplex propensity with G4Hunter. **44**, 1746–1759 (2016).
- 75. Varizhuk, A. *et al.* The expanding repertoire of G4 DNA structures. *Biochimie* **135**, 54–62 (2017).
- 76. Beaudoin, J. D., Jodoin, R. & Perreault, J. P. New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Res.* **42**, 1209–1223 (2014).
- 77. Sahakyan, A. B. *et al.* Machine learning model for sequence-driven DNA G-quadruplex formation. *Sci. Rep.* **7**, 1–11 (2017).
- 78. Garant, J. M., Perreault, J. P. & Scott, M. S. Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. *Bioinformatics* **33**, 3532–3537 (2017).
- 79. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35**, 406–413 (2007).
- 80. Cox, A. D. & Der, C. J. Ras history: The saga continues. *Small GTPases* **1**, 2–27 (2010).
- 81. Whitfield, J. R., Beaulieu, M.-E. & Soucek, L. Strategies to Inhibit Myc and Their Clinical Applicability. *Front. Cell Dev. Biol.* **5**, 1–13 (2017).
- 82. Woodford, K. J., Howell, R. M. & Usdin, K. A novel K+-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *J. Biol. Chem.* **269**, 27029–27035 (1994).
- 83. Kwok, C. K. & Balasubramanian, S. Targeted detection of g-quadruplexes in cellular RNAs. *Angew. Chemie Int. Ed.* **54**, 6751–6754 (2015).
- 84. Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33**, 877–881 (2015).
- 85. 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. Methods* **13**, 841–844 (2016).
- 86. Kouzine, F. *et al.* Global regulation of promoter melting in naive lymphocytes. *Cell* **153**, 988 (2013).
- 87. Kouzine, F. *et al.* Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. *Cell Syst.* **4**, 344–356 (2017).
- 88. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).
- 89. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610–1616 (2006).
- 90. Weeks, K. M. & Mauger, D. M. Exploring RNA Structural Codes with SHAPE Chemistry. *Acc. Chem. Res.* 44, 1280–1291 (2011).

- 91. Kwok, C. K., Sahakyan, A. B. & Balasubramanian, S. Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA. *Angew. Chemie Int. Ed.* **55**, 8958–8961 (2016).
- 92. Hänsel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267–1272 (2016).
- 93. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–186 (2013).
- 94. Zheng, K. *et al.* Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Res.* **48**, 11706–11720 (2020).
- 95. Heddi, B., Cheong, V. V., Martadinata, H. & Phan, A. T. Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9608–9613 (2015).
- 96. Chen, M. C., Murat, P., Abecassis, K., Ferre-D'Amare, A. R. & Balasubramanian, S. Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. *Nucleic Acids Res.* **43**, 2223–2231 (2015).
- 97. Griffith, J. D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514 (1999).
- 98. Doksani, Y., Wu, J. Y., De Lange, T. & Zhuang, X. Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent t-loop formation. *Cell* **155**, 345–356 (2013).
- 99. Tomaska, L., Nosek, J., Kar, A., Willcox, S. & Griffith, J. D. A New View of the T-Loop Junction: Implications for Self-Primed Telomere Extension, Expansion of Disease-Related Nucleotide Repeat Blocks, and Telomere Evolution. *Front. Genet.* **10**, 1–9 (2019).
- 100. De Lange, T. Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
- 101. De Lange, T. T-loops and the origin of telomeres. *Nat. Rev. Mol. Cell Biol.* **5**, 323–329 (2004).
- 102. Steensel, V., B., A. S. & Lange, T. De. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413 (1998).
- 103. Zakian, V. A. Telomeres: The beginnings and ends of eukaryotic chromosomes. *Exp. Cell Res.* **318**, 1456–1460 (2012).
- 104. Shay, J. W. & Wright, W. E. Hayflick, his limit, and cellular ageing. *Nat. Rev. Mol. Cell Biol.* **1**, 72–76 (2000).
- Shay, J. W. & Wright, W. E. Role of telomeres and telomerase in cancer. *Semin. Cancer Biol.* 21, 349–353 (2011).
- 106. Zahler, A., Williamson, J., Cech, T. & Prescott, D. Inhibition of telomerase by G-quartet DNA structures. *Nature* **353**, 412–414 (1991).
- 107. Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. J. Med. Chem. 59, 5987–6011 (2016).
- 108. Moye, A. L. *et al.* Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **6**, 7643 (2015).

- 109. Nakanishi, C. & Seimiya, H. G-quadruplex in cancer biology and drug discovery. *Biochem. Biophys. Res. Commun.* **531**, 45–50 (2020).
- 110. Lerner, L. K. & Sale, J. E. Replication of G quadruplex DNA. Genes (Basel). 10, 1–25 (2019).
- 111. Prorok, P. *et al.* Involvement of G-quadruplex regions in mammalian replication origin activity. *Nat. Commun.* **10**, 1–16 (2019).
- 112. Dhaval Varshney, Jochen Spiegel, Katherine Zyner, D. T. and S. B. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* (2020).
- 113. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: Stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770–780 (2012).
- 114. Robinson, J., Raguseo, F., Nuccio, S. P., Liano, D. & Di Antonio, M. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res.* **4**, 1–13 (2021).
- 115. Rigo, R., Palumbo, M. & Sissi, C. G-quadruplexes in human promoters: A challenge for therapeutic applications. *Biochim. Biophys. Acta Gen. Subj.* **1861**, 1399–1413 (2017).
- 116. Gabay, M., Li, Y. & Felsher, D. W. MYC Activation Is a Hallmark of Cancer Initiation and Maintenance. *Cold Spring Harb. Lab. Press Med.* 1–14 (2014).
- 117. Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **277**, 3459–3469 (2010).
- 118. Monsen, R. C. *et al.* The hTERT core promoter forms three parallel G-quadruplexes. *Nucleic Acids Res.* **48**, 5720–5734 (2020).
- 119. Simonsson, T., Pecinka, P. & Kubista, M. DNA tetraplex formation in the control region of cmyc. *Nucleic Acids Res.* **26**, 1167–1172 (1998).
- 120. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci.* **99**, 11593–11598 (2002).
- 121. Qi, T., Xu, Y., Zhou, T. & Gu, W. The Evolution of G-quadruplex Structure in mRNA Untranslated Region. *Evol. Bioinforma.* **17**, 1–7 (2021).
- 122. Kumari, S., Bugaut, A., Huppert, J. L. & Balasubramanian, S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **3**, 218–221 (2007).
- 123. Kamura, T., Katsuda, Y., Kitamura, Y. & Ihara, T. G-quadruplexes in mRNA: A key structure for biological function. *Biochem. Biophys. Res. Commun.* **526**, 261–266 (2020).
- 124. Xie, X. *et al.* Topology-Selective, Fluorescent "Light-Up" Probes for G-Quadruplex DNA Based on Photoinduced Electron Transfer. *Chem. A Eur. J.* **24**, 12638–12651 (2018).
- 125. Rodriguez, R. *et al.* A novel small molecule that alters shelterin integrity and triggers a DNAdamage response at telomeres. *J. Am. Chem. Soc.* **130**, 15758–15759 (2008).
- 126. Rodriguez, R. *et al.* Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301–310 (2012).
- 127. Pennarun, G. *et al.* Apoptosis related to telomere instability and cell cycle alterations in human glioma cells treated by new highly selective G-quadruplex ligands. *Oncogene* **24**, 2917–

2928 (2005).

- 128. 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).
- 129. Piazza, A. *et al.* Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. *Nucleic Acids Res.* **38**, 4337–4348 (2010).
- 130. Shin-ya, K. *et al.* Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus. *J. Am. Chem. Soc.* **123**, 1262–1263 (2001).
- 131. Kim, M. Y., Vankayalapati, H., Shin-Ya, K., Wierzba, K. & Hurley, L. H. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex. *J. Am. Chem. Soc.* **124**, 2098–2099 (2002).
- 132. Hamon, F. *et al.* An acyclic oligoheteroaryle that discriminates strongly between diverse G-Quadruplex topologies. *Angew. Chemie Int. Ed.* **50**, 8745–8749 (2011).
- 133. Iida, K. & Nagasawa, K. Macrocyclic polyoxazoles as G-quadruplex ligands. *Chem. Rec.* **13**, 539–548 (2013).
- 134. Chung, W. J. *et al.* Solution structure of an intramolecular (3 + 1) human telomeric Gquadruplex bound to a telomestatin derivative. *J. Am. Chem. Soc.* **135**, 13495–13501 (2013).
- 135. Wheelhouse, R. T., Sun, D., Han, H., Han, F. X. & Hurley, L. H. Cationic porphyrins as telomerase inhibitors: The interaction of tetra- (N-methyl-4-pyridyl)porphine with quadruplex DNA. *J. Am. Chem. Soc.* **120**, 3261–3262 (1998).
- 136. Tada-Oikawa, S., Oikawa, S., Hirayama, J., Hirakawa, K. & Kawanishi, S. Dna damage and apoptosis induced by photosensitization of 5,10,15,20-tetrakis (n-methyl-4-pyridyl)-21h,23h-porphyrin via singlet oxygen generation. *Photochem. Photobiol.* **85**, 1391–1399 (2009).
- 137. Qi, Q. *et al.* Preferential Binding of π-Ligand Porphyrin Targeting 5'-5' Stacking Interface of Human Telomeric RNA G-Quadruplex Dimer. *J. Phys. Chem. Lett.* **10**, 2143–2150 (2019).
- 138. Ren, J. & Chaires, J. B. Sequence and Structural Selectivity of Nucleic Acid Binding Ligands. *Biochemistry* **38**, 16067–16075 (1999).
- 139. Gonçalves, D. P. N., Rodriguez, R., Balasubramanian, S. & Jeremy, K. M. Tetramethylpyridiniumporphyrazines a new class of G- quadruplex inducing and stabilising ligands. *Chem. Commun.* 44, 4685–4687 (2006).
- 140. Li, Y., Geyer, C. R. & Sen, D. Recognition of Anionic Porphyrins by DNA Aptamers. *Biochemistry* **35**, 6911–6922 (1996).
- 141. Yett, A., Lin, L. Y., Beseiso, D. & Yatsunyk, L. A. N-methyl mesoporphyrin IX as a highly selective light-up probe for G-quadruplex DNA. *J. Porphyr. Phthalocyanines* **23**, 1195–1215 (2019).
- 142. Nicoludis, J. M., Barrett, S. P., Mergny, J. & Yatsunyk, L. A. Interaction of human telomeric DNA with N- methyl mesoporphyrin IX. *Nucleic Acids Res.* **40**, 5432–5447 (2012).
- 143. Xu, S. *et al.* Analytical Methods Fluorescent G-quadruplex NMM DNA probe for the detection of silver nanoparticles in aqueous. *Anal. Methods* **7**, 1672–1675 (2015).
- 144. Nicoludis, J. M. et al. Optimized End-Stacking Provides Speci fi city of N Methyl

Mesoporphyrin IX for Human Telomeric G - Quadruplex DNA. J. Am. Chem. Soc. **134**, 20446–20456 (2012).

- 145. Liao, T. C. *et al.* A Comparative Study on High Selectivities of Human Telomeric Dimeric G-Quadruplexes by Dimeric G-Quadruplex Binders. *Chem. A Eur. J.* **24**, 15840–15851 (2018).
- 146. Chaudhuri, R., Bhattacharya, S., Dash, J. & Bhattacharya, S. Recent Update on Targeting c -MYC G - Quadruplexes by Small Molecules for Anticancer Therapeutics. *J. Med. Chem.* **64**, 42– 70 (2020).
- 147. Chen, B. J., Wu, Y. L., Tanaka, Y. & Zhang, W. Small molecules targeting c-Myc oncogene: Promising anti-cancer therapeutics. *Int. J. Biol. Sci.* **10**, 1084–1096 (2014).
- 148. Dutta, D. *et al.* Cell penetrating thiazole peptides inhibit c-MYC expression via site-specific targeting of c-MYC. *Nucleic Acids Res.* **46**, 5355–5365 (2018).
- 149. Cocco, M. J., Hanakahi, L. A., Huber, M. D. & Maizels, N. Specific interactions of distamycin with G-quadruplex DNA. *Nucleic Acids Res.* **31**, 2944–2951 (2003).
- 150. Asamitsu, S. *et al.* Simultaneous Binding of Hybrid Molecules Constructed with Dual DNA-Binding Components to a G-Quadruplex and Its Proximal Duplex. *Chem. - A Eur. J.* **24**, 4428– 4435 (2018).
- 151. Gabelica, V. *et al.* Multiple and cooperative binding of fluorescence light-up probe thioflavin t with human telomere DNA G-quadruplex. *Biochemistry* **52**, 5620–5628 (2013).
- 152. Zuffo, M. *et al.* More is not always better: finding the right trade-off between affinity and selectivity of a G-quadruplex ligand. *Nucleic Acids Res.* **46**, e115 (2018).
- 153. Grande, V., Doria, F., Freccero, M. & Würthner, F. An Aggregating Amphiphilic Squaraine: A Light-up Probe That Discriminates Parallel G-Quadruplexes. *Angew. Chemie Int. Ed.* **56**, 7520–7524 (2017).
- 154. Hong, Y. *et al.* Label-free fluorescent probing of G-quadruplex formation and real-time monitoring of DNA folding by a quaternized tetraphenylethene salt with aggregation-induced emission characteristics. *Chem. A Eur. J.* **14**, 6428–6437 (2008).
- 155. Englebienne, P., Van Hoonacker, A. & Verhas, M. Surface plasmon resonance: Principles, methods and applications in biomedical sciences. *Spectroscopy* **17**, 255–273 (2003).
- 156. Redman, J. E. Surface plasmon resonance for probing quadruplex folding and interactions with proteins and small molecules. *Methods* **43**, 302–312 (2007).
- 157. Calabrese, D. R. *et al.* Chemical and structural studies provide a mechanistic basis for recognition of the MYC G-quadruplex. *Nat. Commun.* **9**, 1–15 (2018).
- 158. Hu, M., Wu, T., Huang, Q. & Jin, G. New substituted quinoxalines inhibit triple-negative breast cancer by specifically downregulating the c-MYC transcription. *Nucleic Acids Res.* **47**, 10529–10542 (2019).
- 159. Boger, D. L., Fink, B. E., Brunette, S. R., Tse, W. C. & Hedrick, M. P. A simple, high-resolution method for establishing DNA binding affinity and sequence selectivity. *J. Am. Chem. Soc.* **123**, 5878–5891 (2001).
- 160. Monchaud, D., Allain, C. & Teulade-Fichou, M. P. Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of

putative ligands. Bioorganic Med. Chem. Lett. 16, 4842-4845 (2006).

- 161. Largy, E., Hamon, F. & Teulade-Fichou, M. P. Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures. *Anal. Bioanal. Chem.* **400**, 3419–3427 (2011).
- 162. Monchaud, D. *et al.* Ligands playing musical chairs with G-quadruplex DNA: A rapid and simple displacement assay for identifying selective G-quadruplex binders. *Biochimie* **90**, 1207–1223 (2008).
- 163. Mergny, J. L. & Maurizot, J. C. Fluorescence resonance energy transfer as a probe for Gquartet formation by a telomeric repeat. *ChemBioChem* **2**, 124–132 (2001).
- 164. Cian, A. De *et al.* Fluorescence-based melting assays for studying quadruplex ligands. *Methods* **42**, 183–195 (2007).
- 165. De Rache, A. & Mergny, J. L. Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay. *Biochimie* **115**, 194–202 (2015).
- 166. Sapsford, K. E., Berti, L. & Medintz, I. L. Materials for fluorescence resonance energy transfer analysis: Beyond traditional donor-acceptor combinations. *Angew. Chemie Int. Ed.* **45**, 4562–4589 (2006).
- 167. Wu, F. *et al.* Genome-wide analysis of DNA G-quadruplex motifs across 37 species provides insights into G4 evolution. *Commun. Biol.* **4**, 1–11 (2021).
- 168. Raguseo, F., Chowdhury, S., Minard, A. & Di Antonio, M. Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome. *Chem. Commun.* **56**, 1317–1324 (2020).
- 169. Zeng, W. *et al.* Small-molecule-based human genome G4 profiling reveals potential gene regulation activity. *Chem. Commun.* **55**, 2269–2272 (2019).
- 170. Yang, S. Y. *et al.* Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.* **9**, (2018).
- 171. Im, K., Mareninov, S., Palma Diaz, M. F. & Yong, W. H. An introduction to Performing Immunofluorescence Staining Kyuseok. *Methods Mol. Biol.* **1897**, 299–311 (2019).
- 172. Brown, B. A. *et al.* Isolation and characterization of a monoclonal anti-quadruplex DNA antibody from autoimmune 'viable motheaten' mice. *Biochemistry* **37**, 16325–16337 (1998).
- 173. Schaffitzel, C. *et al.* In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8572–8577 (2001).
- 174. Javadekar, S. M., Nilavar, N. M., Paranjape, A., Das, K. & Raghavan, S. C. Characterization of Gquadruplex antibody reveals differential specificity for G4 DNA forms. *DNA Res.* **27**, 1–17 (2020).
- 175. Biffi, G., Di Antonio, M., Tannahill, D. & Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* **6**, 75–80 (2014).
- 176. Domarco, O. *et al.* Subcellular Duplex DNA and G-Quadruplex Interaction Profiling of a Hexagonal PtII Metallacycle. *Angew. Chemie Int. Ed.* **58**, 8007–8012 (2019).

- 177. Laguerre, A. *et al.* Visualization of RNA-Quadruplexes in Live Cells. *J. Am. Chem. Soc.* **137**, 8521–8525 (2015).
- 178. Henderson, A. *et al.* Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* **42**, 860–869 (2014).
- 179. Liu, H. Y. *et al.* Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chem. Biol.* **23**, 1261–1270 (2016).
- 180. Izbicka, E. *et al.* Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res.* **59**, 639–644 (1999).
- 181. Drygin, D. *et al.* Anticancer activity of CX-3543: A direct inhibitor of rRNA biogenesis. *Cancer Res.* **69**, 7653–7661 (2009).
- 182. Chen, X. C. *et al.* Tracking the Dynamic Folding and Unfolding of RNA G-Quadruplexes in Live Cells. *Angew. Chemie Int. Ed.* **57**, 4702–4706 (2018).
- 183. Zhang, S. *et al.* Real-time monitoring of DNA G-quadruplexes in living cells with a small-molecule fluorescent probe. *Nucleic Acids Res.* **46**, 7522–7532 (2018).
- 184. Largy, E., Granzhan, A., Hamon, F., Verga, D. & Teulade-Fichou, M.-P. Visualizing the quadruplex: from fluorescent ligands to light-up probes. *Pept. Mater.* **310**, 1–26 (2011).
- 185. Ma, D. L. *et al.* Recent Developments in G-Quadruplex Probes. *Chem. Biol.* **22**, 812–828 (2015).
- 186. Tseng, T.-Y. *et al.* Fluorescent probe for visualizing guanine-quadruplex DNA by fluorescence lifetime imaging microscopy. *J. Biomed. Opt.* **18**, 101309 (2013).
- 187. Tseng, T. Y., Chu, I. Te, Lin, S. J., Li, J. & Chang, T. C. Binding of small molecules to Gquadruplex DNA in cells revealed by fluorescence lifetime imaging microscopy of o-BMVC foci. *Molecules* 24, 1–13 (2019).
- 188. Shivalingam, A. *et al.* The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat. Commun.* **6**, (2015).
- 189. Summers, P. A. *et al.* Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat. Commun.* **12**, 1–11 (2021).
- 190. Yang, P., De Cian, A., Teulade-Fichou, M. P., Mergny, J. L. & Monchaud, D. Engineering bisquinolinium/thiazole orange conjugates for fluorescent sensing of G-quadruplex DNA. *Angew. Chemie Int. Ed.* **48**, 2188–2191 (2009).
- 191. Yan, J. W. *et al.* Development of a new colorimetric and red-emitting fluorescent dual probe for G-quadruplex nucleic acids. *Chem. Commun.* **50**, 6927–6930 (2014).
- 192. Tera, M. *et al.* Visualization of G-quadruplexes by using a BODIPY-labeled macrocyclic heptaoxazole. *Org. Biomol. Chem.* **8**, 2749–2755 (2010).
- 193. Largy, E., Hamon, F. & Teulade-Fichou, M. P. A streptavidin paramagnetic-particle based competition assay for the evaluation of the optical selectivity of quadruplex nucleic acid fluorescent probes. *Methods* **57**, 129–137 (2012).
- 194. Percivalle, C., Mahmood, T. & Ladame, S. Two-in-one: A pH-sensitive, acridine-based,

fluorescent probe binds G-quadruplexes in oncogene promoters. *Medchemcomm* **4**, 211–215 (2013).

- 195. Wu, F. *et al.* Visualization of G-quadruplexes in gel and in live cells by a near-infrared fluorescent probe. *Sensors Actuators, B Chem.* **236**, 268–275 (2016).
- 196. Di Antonio, M. *et al.* Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.* **12**, 832–837 (2020).
- 197. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 159–161 (2008).
- 198. Nguyen, S. S. & Prescher, J. A. Developing bioorthogonal probes to span a spectrum of reactivities. *Nature Reviews Chemistry* **4**, 476–489 (2020).
- 199. Lefebvre, J., Guetta, C., Poyer, F., Mahuteau-Betzer, F. & Teulade-Fichou, M. P. Copper–Alkyne Complexation Responsible for the Nucleolar Localization of Quadruplex Nucleic Acid Drugs Labeled by Click Reactions. *Angew. Chemie - Int. Ed.* **56**, 11365–11369 (2017).
- 200. Agard, N. J., Baskin, J. M., Prescher, J. A., Lo, A. & Bertozzi, C. R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **1**, 644–648 (2006).
- 201. Sletten, E. M., Nakamura, H., Jewett, J. C. & Bertozzi, C. R. Difluorobenzocyclooctyne: Synthesis, reactivity, and stabilization by β-cyclodextrin. J. Am. Chem. Soc. 132, 11799–11805 (2010).
- 202. Leif, R. C., Stein, J. H. & Zucker, R. M. A short history of the initial application of anti-5-BrdU to the detection and measurement of S phase. *Cytometry* **58A**, 45–52 (2004).
- 203. Betzer, J. *et al.* Linking of Antitumor trans NHC-Pt (II) Complexes to G-Quadruplex DNA Ligand for Telomeric Targeting. *Bioconjug. Chem.* **27**, 1456–1470 (2016).
- 204. Verga, D., Hamon, F., Poyer, F., Bombard, S. & Teulade-Fichou, M. P. Photo-cross-linking probes for trapping G-quadruplex DNA. *Angew. Chemie Int. Ed.* **53**, 994–998 (2014).
- 205. Zavgorodny, S. *et al.* 1-Alkylthioalkylation of Nucleoside Hydroxyl Functions and Its Synthetic Applications: a New Versatile Method in Nucleoside Chemistry. *Tetrahedron Lett.* **32**, 7593–7596 (1991).
- 206. Varizhuk, A. M. *et al.* Synthesis of triazole-linked oligonucleotides with high affinity to DNA complements and an analysis of their compatibility with biosystems. *J. Org. Chem.* **78**, 5964–5969 (2013).
- 207. Liang, L. & Astruc, D. The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) 'click' reaction and its applications. An overview. *Coord. Chem. Rev.* **255**, 2933–2945 (2011).
- 208. Huisgen, R. & Huisgen, R. 1,3-Dipolar Cycloadditions. Past and Future. *Angew. Chemie Int. Ed. English* **2**, 565–598 (1963).
- 209. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise huisgen cycloaddition process: Copper(I)-catalyzed regioselective 'ligation' of azides and terminal alkynes. *Angew. Chemie Int. Ed.* **41**, 2596–2599 (2002).
- 210. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-Triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057–3064 (2002).

- 211. Fokin, V. V., Malik, J. A. & Worrel, B. T. Direct Evidence of a Dinuclear Copper Intermediate in Cu(I)- Catalyzed Azide–Alkyne Cycloadditions. *Science (80-.).* **340**, 457–460 (2013).
- 212. Ben El Ayouchia, H., Bahsis, L., Anane, H., Domingo, L. R. & Stiriba, S. E. Understanding the mechanism and regioselectivity of the copper(i) catalyzed [3 + 2] cycloaddition reaction between azide and alkyne: A systematic DFT study. *RSC Adv.* **8**, 7670–7678 (2018).
- 213. Jin, L., Tolentino, D. R., Melaimi, M. & Bertrand, G. Isolation of bis(copper) key intermediates in Cu-catalyzed azide-alkyne "click reaction. *Sci. Adv.* **1**, (2015).
- 214. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie Int. Ed.* **40**, 2004–2021 (2001).
- 215. Rodionov, V. O., Presolski, S. I., Díaz, D. D., Fokin, V. V. & Finn, M. G. Ligand-accelerated Cucatalyzed azide-alkyne cycloaddition: A mechanistic report. *J. Am. Chem. Soc.* **129**, 12705– 12712 (2007).
- 216. Donnelly, P. S., Zanatta, S. D., Zammit, S. C., White, J. M. & Williams, S. J. 'Click' cycloaddition catalysts: Copper(I) and copper(II) tris(triazolylmethyl)amine complexes. *Chem. Commun.* **233**, 2459–2461 (2008).
- 217. Piazza, A. *et al.* Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J.* **34**, 1718–1734 (2015).
- 218. Rachwal, P. A., Brown, T. & Fox, K. R. Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. *Biochemistry* **46**, 3036–3044 (2007).
- 219. Le, H. T. *et al.* Not All G-Quadruplexes are Created Equally: An Investigation of the Structural Polymorphism of the c-Myc G-Quadruplex- Forming Sequence and its Interaction with the Porphyrin meso- Tetra(N-methyl-4-pyridyl)porphine Huy. *Org. Biomol. Chem.* **10**, 9393–9404 (2012).
- 220. Zaman Huri, H., Saleem, M., Abbaspour Babaei, M., Ahmadipour, F. & Kamalidehghan, B. Receptor tyrosine kinase (c-Kit) inhibitors: a potential therapeutic target in cancer cells. *Drug Des. Devel. Ther.* **10**, 2443–2459 (2016).
- 221. Lim, K. W. *et al.* Sequence variant (CTAGGG)n in the human telomere favors a G-quadruplex structure containing a G·C·G·C tetrad. *Nucleic Acids Res.* **37**, 6239–6248 (2009).
- 222. 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).
- 223. Biffi, G., Tannahill, D. & Balasubramanian, S. An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J. Am. Chem. Soc.* **134**, 11974–11976 (2012).
- 224. Renčiuk, D. *et al.* A FRET-based screening assay for nucleic acid ligands. *Methods* **57**, 122–128 (2012).
- 225. Tran, P. L. T., Largy, E., Hamon, F., Teulade-Fichou, M. P. & Mergny, J. L. Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures. *Biochimie* **93**, 1288–1296 (2011).
- 226. Rongeat, N. et al. Flow cytometer based on triggered supercontinuum laser illumination.

Cytom. Part A **81**, 611–617 (2012).

- 227. 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).
- 228. Poulin-Kerstein, A. T. & Dervan, P. B. DNA-Templated Dimerization of Hairpin Polyamides. J. Am. Chem. Soc. **125**, 15811–15821 (2003).
- 229. Di Antonio, M. *et al.* Selective RNA versus DNA G-quadruplex targeting by situ click chemistry. *Angew. Chemie Int. Ed.* **51**, 11073–11078 (2012).
- 230. Mestre-Fos, S., Ito, C., Moore, C. M., Reddi, A. R. & Williams, L. D. Human ribosomal Gquadruplexes regulate heme bioavailability. *J. Biol. Chem.* **295**, 14855–14865 (2020).
- 231. Das, T., Panda, D., Saha, P. & Dash, J. Small Molecule Driven Stabilization of Promoter G-Quadruplexes and Transcriptional Regulation of c-MYC. *Bioconjug. Chem.* **29**, 2636–2645 (2018).
- 232. Müller, S., Kumari, S., Rodriguez, R. & Balasubramanian, S. Small-molecule-mediated Gquadruplex isolation from human cells. *Nat. Chem.* **2**, 1095–1098 (2010).
- 233. Zeng, W. *et al.* Small-molecule-based human genome G4 profiling reveals potential gene regulation activity. *Chem. Commun.* **55**, 2269–2272 (2019).
- 234. Haudecoeur, R., Stefan, L., Denat, F. & Monchaud, D. A model of smart g-quadruplex ligand. *J. Am. Chem. Soc.* **135**, 550–553 (2013).
- 235. Renard, I. *et al.* Small-molecule affinity capture of DNA/RNA quadruplexes and their identification in vitro and in vivo through the G4RP protocol. *Nucleic Acids Res.* **47**, 502–510 (2019).
- Double, J. A. *et al.* The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Res.* 65, 1489–1496 (2005).
- 237. Salvati, E. *et al.* Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J. Clin. Invest.* **117**, 3236–3247 (2007).
- 238. Lapinsky, D. J. Tandem photoaffinity labeling-bioorthogonal conjugation in medicinal chemistry. *Bioorganic Med. Chem.* **20**, 6237–6247 (2012).
- 239. Smith, E. & Collins, I. Photoaffinity labeling in target-and binding-site identification. *Future Med. Chem.* **7**, 159–183 (2015).
- 240. Hill, J. R. & Robertson, A. A. B. Fishing for Drug Targets: A Focus on Diazirine Photoaffinity Probe Synthesis. *J. Med. Chem.* **61**, 6945–6963 (2018).
- 241. A. Fleming, S. Chemical reagents in photoaffinity labeling. *Tetrahedron* **51**, 12479–12520 (1995).
- 242. Cline, M. R., Mandel, S. M. & Platz, M. S. Identification of the reactive intermediates produced upon photolysis of p-azidoacetophenone and its tetrafluoro analogue in aqueous and organic solvents: Implications for photoaffinity labeling. *Biochemistry* **46**, 1981–1987 (2007).
- 243. Das, J. Aliphatic diazirines as photoaffinity probes for proteins: Recent developments. *Chem. Rev.* **111**, 4405–4417 (2011).

- 244. Sakurai, K., Ozawa, S., Yamada, R., Yasui, T. & Mizuno, S. Comparison of the reactivity of carbohydrate photoaffinity probes with different photoreactive groups. *ChemBioChem* **15**, 1399–1403 (2014).
- 245. Vodovozova, E. L. Photoaffinity labeling and its application in structural biology. *Biochem.* **72**, 1–20 (2007).
- 246. Wang, Y. C. *et al.* Unravelling nitrene chemistry from acyclic precursors: Recent advances and challenges. *Org. Chem. Front.* **8**, 1677–1693 (2021).
- 247. Jessen, K. A. *et al.* The discovery and mechanism of action of novel tumor-selective and apoptosis-inducing 3,5-diaryl-1,2,4-oxadiazole series using a chemical genetics approach. *Mol. Cancer Ther.* **4**, 761–771 (2005).
- 248. Li, Z. *et al.* Design and synthesis of minimalist terminal alkyne-containing diazirine photocrosslinkers and their incorporation into kinase inhibitors for cell- and tissue-based proteome profiling. *Angew. Chemie - Int. Ed.* **52**, 8551–8556 (2013).
- 249. Kotoku, N. *et al.* Synthesis and evaluation of effective photoaffinity probe molecule of furospinosulin-1, a hypoxia-selective growth inhibitor. *Bioorganic Med. Chem.* **22**, 2102–2112 (2014).
- 250. Su, H. *et al.* Photoactive G-Quadruplex Ligand Identifies Multiple G-Quadruplex-Related Proteins with Extensive Sequence Tolerance in the Cellular Environment. *J. Am. Chem. Soc.* **143**, 1917–1923 (2021).
- 251. Zhang, X., Spiegel, J., Martínez Cuesta, S., Adhikari, S. & Balasubramanian, S. Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. *Nat. Chem.* **13**, (2021).
- 252. Parker, C. G. *et al.* Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **168**, 527-541.e29 (2017).
- 253. Stetter, H. Catalyzed Addition of Aldehydes to Activated Double Bonds. *Angew. Chem. Int. Ed. Engl.* **15**, 639–647 (1976).
- 254. Seebach, D. Methods of Reactivity Umpolung. *Angew. Chemie Int. Ed. English* **18**, 239–258 (1979).
- 255. Pabbaraja, S. *et al.* Synthesis and biological evaluation of 12-, 13-, 14-membered macrolides and open chain 2,6-trans-disubstituted dihydropyran analogues for aspergillides. *Tetrahedron Lett.* **59**, 2570–2576 (2018).
- 256. Wang, L. *et al.* Base-mediated one-pot synthesis of aliphatic diazirines for photoaffinity labeling. *Molecules* **22**, 1389 (2017).
- 257. Pícha, J. *et al.* Tri-Orthogonal Scaffolds for the Solid-Phase Synthesis of Peptides. *European J. Org. Chem.* **2018**, 5180–5192 (2018).
- 258. Cagnoni, A. J., Varela, O., Gouin, S. G., Kovensky, J. & Uhrig, M. L. Synthesis of multivalent glycoclusters from 1-thio-β-D-galactose and their inhibitory activity against the β-galactosidase from E. coli. *J. Org. Chem.* **76**, 3064–3077 (2011).
- 259. Appel, R. Tertiary Phosphane/Tetrachloromethane, a Versatile Reagent for Chlorination, Dehydration, and P-N Linkage. *Angew. Chemie Int. Ed.* **14**, 801–811 (1975).
- 260. Bončina, M. et al. Thermodynamic fingerprints of ligand binding to human telomeric G-

quadruplexes. Nucleic Acids Res. 43, 10376–10386 (2015).

- 261. Waller, Z. A. E., Sewitz, S. A., Hsu, S. T. D. & Balasubramanian, S. A small molecule that disrupts G-quadruplex DNA structure and enhances gene expression. *J. Am. Chem. Soc.* **131**, 12628–12633 (2009).
- 262. Morris, M. J., Wingate, K. L., Silwal, J., Leeper, T. C. & Basu, S. The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res.* **40**, 4137–4145 (2012).
- 263. De Cian, A., DeLemos, E., Mergny, J. L., Teulade-Fichou, M. P. & Monchaud, D. Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J. Am. Chem. Soc.* **129**, 1856–1857 (2007).
- 264. Mitteaux, J. *et al.* Identifying G-Quadruplex-DNA-Disrupting Small Molecules. *J. Am. Chem. Soc.* **143**, 12567–12577 (2021).
- 265. Goodsell, D. S. The Molecular Perspective: Ultraviolet Light and Pyrimidine Dimers. *Oncologist* **6**, 298–299 (2001).
- 266. Dubinsky, L., Krom, B. P. & Meijler, M. M. Diazirine based photoaffinity labeling. *Bioorganic Med. Chem.* **20**, 554–570 (2012).
- 267. Müller, S. *et al.* Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. *Org. Biomol. Chem.* **10**, 6537–6546 (2012).
- 268. Meguellati, K., Spichty, M. & Ladame, S. Synthesis, spectroscopic and DNA alkylating properties of malondialdehyde (MDA) bis-imine fluorescent adducts-Synthèse Nmethylquinolinium. *Mol. Biosyst.* **6**, 1694–1699 (2010).



ÉCOLE DOCTORALE

Sciences Chimiques: Molécules, Matériaux, Instrumentation et Biosystèmes (2MIB)

Titre : Nouveaux ligands fonctionnels des G-quadruplexes pour l'imagerie cellulaire et la capture covalente d'ADN génomique

Mots clés : Ligands de G-quadruplex, Immunodétection, Synthèse organique, Chimie click, Photo-alkylation, Evaluations biophysiques et Biochimiques.

Résumé : Les G-quadruplexes (G4s) sont des structures secondaires non canoniques, pouvant être adoptées par des brins d'ADN ou d'ARN riches en guanines possédant au moins quatre répétitions de deux guanines. Ces structures sont aujourd'hui connues pour participer à la régulation de phénomènes biologiques comme la réplication, la transcription et la traduction. Depuis la découverte de ces structures, de nombreuses petites molécules (appelées ligands G4) ont été développées dans le but d'interagir sélectivement avec les Gquadruplexes. Cependant, caractère très le dynamique de ces structures en fait des cibles difficiles à étudier en conditions cellulaires. Il est donc nécessaire de développer des outils moléculaires permettant leur étude en cellules vivantes.

Dans ce contexte, une nouvelle méthodologie, G4-GIS développée. appelée été Cette а méthodologie permet de suivre avec précision la localisation de petites molécules in cellulo et combine les avantages de l'utilisation de ligands avec l'immunofluorescence. Pour le développement de ce projet, plusieurs dérivés du ligand des Gquadruplexes PDC ont été fonctionnalisés avec un haptène (5-BrdU), pouvant être reconnu par un anticorps spécifique. La synthèse de ces dérivés implique tout d'abord la fonctionnalisation du PDC par un groupement azoture ou alcyne (PDC-Alk/N3). En parallèle, l'haptène commercial 5-BrdU a été fonctionnalisé par les mêmes groupements azoture ou alcyne (5-BrdU-Alk/N3). Les différents ligands PDC ont ensuite été couplés aux partenaires 5-BrdU-Alk/N3 adaptés par une réaction de CuAAC pour former les composés PDC-BrdU. Cette réaction de être réalisée in cellulo peu, plus, (fonctionnalisation in situ) dans le but d'éviter une potentielle perte d'affinité ou de sélectivité des ligands par la présence de la 5-BrdU. Les différents ligands avant et après fonctionnalisation ont ensuite

été testés pour leur capacité à interagir avec les structures G-quadruplexes par deux méthodes biophysiques. L'imagerie cellulaire des ligands ensuite PDC-BrdU а été réalisée par immunofluorescence. De plus, l'immunodétection a pu être réalisée après fonctionnalisation in situ des ligands PDC-Alk/N3 avec la 5-BrdU. Cette méthodologie d'immunodétection de ligands des G-quadruplexes permet ainsi de détecter avec une excellente résolution la localisation de ligands des G-quadruplexes in cellulo.

Déterminer la localisation exacte de formation de structures G-quadruplexes à travers le génome représente un réel défi et un enjeu majeur pour la compréhension des rôles biologiques des Gquadruplexes. Pour cela, une nouvelle méthode combinant le marguage par photoaffinité de G4s (PAL) avec l'immunoprécipitation а été dérivés PDC développée. du ligand Des fonctionnalisés à la fois par un groupement photoactivable et une fonction alcyne ont été synthétisés. L'affinité et la sélectivité de ces ligands pour les structures G-quadruplexes ont été évaluées par deux méthodes biophysiques sur différentes séquences G4s. Par ailleurs, la capacité des ligands à générer de façon sélective des liens covalents avec l'ADN G-quadruplex après photoirradiation a été étudiée. Par la suite, ces molécules seront utilisées en cellules dans le but de réaliser l'immunoprécipitation de fragments d'ADN liés aux ligands après fonctionnalisation in situ avec la 5-BrdU par une réaction de CuAAC. Cette approche permettra de comprendre quelles séquences du génome forment des G4s ainsi que leur accessibilité par des ligands. Cette méthodologie permettra ainsi de clarifier les rôles biologiques associés à la formation de ces structures secondaires.

Title : Novel functional G-quadruplex ligands for cell imaging and genomic DNA covalent capture

Keywords : G-quadruplex ligands, Organic synthesis, Biophysical and biochemical assays Immunodetection, Click chemistry, Photo-alkylation

Abstract : G-quadruplexes (G4s) are non-canonical secondary structures that can be formed in G-rich DNA or RNA strands containing at least four repeats of two guanines. Today, these structures are known to regulate important biological processes such as replication, transcription and translation. Since their discovery, many small molecules (called G4 ligands) have been developed in the aim to selectively interact with these structures. However, G4 dynamic nature makes them difficult to study in cellular conditions. Therefore, it is still necessary to develop new molecular tools to study them in live cells.

In this context, a new methodology called G4-GIS has been developed. This methodology allows to follow with higher resolution the localization of small molecules in cells combines the and advantages of using G4 ligands and immunofluorescence. For the development of this project, several derivatives of the known Gquadruplex ligand PDC were functionalized with a hapten (5-BrdU), which can be recognized by a specific antibody. At first, the synthesis focused on the functionalization of the PDC core with a linker bearing either an azide or an alkyne (PDC-Alk/N3). In parallel, the commercially available 5-BrdU was functionalized with an azide or alkyne group (5-BrdU-Alk/N3). The different PDC ligands were then tethered to the adapted 5-BrdU-Alk/N3 partners by a CuAAC reaction to form the 5BrdU functionalized PDC derivatives. This reaction can be performed in situ the presence of the G4 target (in functionalization), avoiding potential loss of affinity or selectivity due to the presence of the hapten. The different PDC derivatives, before and after functionalization, were then evaluated for their ability to bind G-quadruplex structures by two biophysical assays.

Cell imaging of PDC-BrdU ligands was then performed by immunofluorescence staining. In addition, this method was performed after in situ functionalization of PDC-Alk/N3 ligands with 5-BrdU moiety. This G-quadruplex ligand immunodetection methodology allows to visualize excellent resolution with an the precise localization of G-quadruplex ligands in cells.

Determining the exact localization of Gquadruplex structures throughout the genome is a major challenge for unravelling the biological roles of G-quadruplexes. For this purpose, a new method combining G4 photoaffinity labeling (PAL) and immunoprecipitation capture has been developed. PDC derivatives functionalized with both a photoactivatable moiety and a terminal alkyne were synthesized. The affinity and selectivity of these ligands towards G-quadruplex structures have been evaluated by two biophysical methods and the ability to selectively generate covalent bonds with G-quadruplex DNA after photo-irradiation was studied on different G4forming sequences. In the next future, these G4 ligands will be employed in cells and, after immuno-tag (5BrdU) functionalization by CuAAC reaction, will be used to develop a chemical immunoprecipitation methodology. By using this approach, it will be possible to understand which G4s are really forming in a dynamic system and their accessibility by the ligands in order to clarify their roles associated with biological functions such as transcription.