



**HAL**  
open science

# Photoactivatable G-quadruplex ligands: synthesis, activation, reactivity, and biological application

Charles-Henri Buffet

## ► To cite this version:

Charles-Henri Buffet. Photoactivatable G-quadruplex ligands: synthesis, activation, reactivity, and biological application. Organic chemistry. Université Paris-Saclay; Università degli studi (Pavie, Italie), 2024. English. NNT: 2024UPASF009 . tel-04523314

**HAL Id: tel-04523314**

**<https://theses.hal.science/tel-04523314>**

Submitted on 27 Mar 2024

**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.

# Photoactivatable G-quadruplex ligands: synthesis, activation, reactivity, and biological application

*Activation photochimique de ligands covalents et sélectifs de  
l'ADN G-quadruplex*

**Thèse de doctorat de l'université Paris-Saclay  
et de l'université de Pavie**

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 les unités de recherche **Chimie et modélisation pour la  
biologie du cancer** (Université Paris-Saclay, CNRS, Inserm) et **Lab G** (Université de  
Pavie), sous la direction de **Daniela VERGA**, Chargée de recherche, la co-direction de  
**Mauro FRECCERO**, Professeur, le co-encadrement de **Marie-Paule TEULADE-  
FICHO**, Directrice de recherche, et de **Filippo DORIA**, Assisant professeur

**Thèse soutenue à Paris-Saclay, le 16 février 2024, par**

**Charles-Henri BUFFET**

## **Composition du Jury**

Membres du jury avec voix délibérative

<b>Joanne XIE</b> Professeure, ENS Paris-Saclay	Présidente
<b>Thomas LAVERGNE</b> Chargé de recherche, HDR, Université Grenoble Alpes	Rapporteur & Examineur
<b>Antonio RANDAZZO</b> Professeur, Università degli Studi di Napoli Federico II	Rapporteur & Examineur
<b>Paola ARIMONDO</b> Directrice de recherche, HDR, Institut Pasteur	Examineur
<b>Stefano PROTTI</b> Professeur assistant, Università degli Studi di Pavia	Examineur



**Titre :** Activation photochimique de ligands covalents et sélectifs de l'ADN G-quadruplex

**Mots clés :** G-quadruplexes, Ligand G4 photoactivable, Études photochimiques, Biophysique, Biochimie

**Résumé :** Les G-quadruplexes sont des structures non canoniques des acides nucléiques formés au niveau de région riche en guanine et en présence d'ions potassium. Ces structures sont connues pour se former dans le génome humain (> 10 000) et semblent participer à la régulation de phénomènes biologiques.

Le but du projet est de développer et étudier des ligands trifonctionnels pour les utiliser comme outils dans un protocole d'immunoprecipitation pour mieux comprendre l'accessibilité et la dynamique des G4s via une association covalente entre le ligand et la structure G4. Les ligands sont composés de i) un coeur PDC qui est capable d'interagir spécifiquement avec les G4s, ii) un groupe photoactivable capable de piéger covalamment les G4s, et iii) un alcyne terminal pour la fonctionnalisation post-piégeage par CuAAC.

La première partie présente la synthèse et les études biophysiques et biochimiques des nouvelles familles de ligands G4 photoactivables. La conservation des propriétés du composé PDC vis à vis les G4s a été évalué par des expériences de FRET-melting et G4-FID. Leur capacité a générer une liaison covalente avec la structure G4 a été confirmée par gel d'électrophorèse en condition dénaturante, et les sites d'alkylations ont été identifiés par séquençage chimique et enzymatique.

La seconde partie présente les études photochimiques réalisées en présence de prototypes. La nature des espèces réactives a été étudiée par Laser Flash Photolysis. Pour conclure, la photoréactivité a été étudiée en présence de différent nucléosides et les adduits produits ont été caractérisés par RMN.

**Title:** Photoactivatable G-quadruplex ligands: synthesis, activation, reactivity and biological applications

**Keywords:** G-quadruplexes, Photoactivatable G4 ligands, Photochemical studies, Biophysics, Biochemistry

**Abstract:** G-quadruplexes (G4s) are non-canonical four-stranded nucleic acid structures generated by the folding of G-rich sequences in the presence of potassium ions. G4s have been found to form in the human genome (> 10,000) and seem to play a role in the regulation of many cellular processes.

The aim of this project is to develop and study trifunctional ligands that can be used in an immunoprecipitation methodology to understand more about G4 accessibility and dynamics by means of a covalent association between the G4 ligand and the G4 structure. These ligands are composed of i) the PDC core able to interact selectively with G-quadruplexes, ii) a photoactivatable moiety (benzophenone, benzyl and aliphatic diazirine, and aryl azide) able to trap covalently G-quadruplexes, and iii) a terminal alkyne for post-covalent binding functionalisation by CuAAC reaction.

The first part describes the synthesis and the biophysical and biochemical studies of the two new families of photoactivatable G4 ligands (in-line and branched compounds). The conservation of the specificity of PDC toward G4s was evaluated by FRET-melting and G4-FID assays. Their ability to generate a covalent bond with the G4 structures was assessed by denaturing gel electrophoresis and the identification of the alkylation sites determined by both chemical and enzymatic sequencing experiments.

The second part describes the photochemical studies performed in the presence of mimic prototypes. The nature of the reactive intermediates was studied by Laser Flash photolysis. Finally, photoreactivity was studied in the presence of different nucleosides and generated adducts were characterised by NMR.



Now one must get drunk and drink  
With all one's might, since Myrsilus has died

**fr.332**

Do not turn your heart to troubles,  
One gets nowhere becoming vexed,  
O Bycchis! It is the best of medicines  
To bring wine to get drunk

**fr. 335**

**Alcaeus**

Νῦν χρῆ μεθύσθην καί τινα πέρ  
βίαν

πώνην, ἐπειδὴ κάτθανε  
Μύρσιλος

**fr.332**

Οὐ χρῆ κάκοισι θυμον ἐπιτρέπην,  
προκόψομεν γὰρ οὐδὲν  
ἀσάμενοι,

ὦ Βύκχι, φαρμάκων δ' ἄριστον  
οἶνον ἐνεικαμένοις μεθύσθην

**fr. 335**



*« Il ne faut pas douter un seul moment de la suprématie qu'a la vérité sur le  
rêve »*

*"One must not doubt for a single moment the supremacy that truth has over dreams"*

**Noël Bernard**



## Acknowledgements

First, I would like to thank my thesis director **Dr. Daniela Verga** for allowing me to pursue my PhD under her supervision on her multidisciplinary project. I am grateful to have been given the opportunity to be trained in many methods and techniques across different fields, and also thank her help in becoming more critical and attentive to details.

I would like to thank my Italian (i.e. in Italy) thesis director **Pr. Mauro Freccero** to have agreed to this Joint-Cotutelle adventure without knowing me at first. I am grateful for his supervision and presence during the project, and for this opportunity to have come to Pavia to realise part of my PhD.

In addition to my thesis directors, I would like to thank **Dr. Marie-Paule Teulade-Fichou** for her interesting remarks and good advices during my PhD, and **Dr. Filippo Doria** for his daily help in Pavia and particularly with the different HPLCs used in this work.

I would like to thank **Dr. Marcel Hollenstein** and **Dr. Maria Duca** for being part of the annual thesis monitoring committee and giving me good advices about my research.

I would like to thank my PhD jury **Pr. Joanne Xie**, **Dr. Thomas Lavergne**, **Pr. Antonio Randazzo**, **Dr. Paola Arimondo**, and **Pr. Stefano Protti**, to the time they have devoted to examine the results of my research, and for our discussion during the PhD defence.

I would like to thank all the people who helped me with my experiments. In France, I would like to thank **Mrs. Corinne Landras Guetta** for her help with the synthesis of the DEG series, **Dr. Thibaut Masson** for his help and for training me in biophysics and biochemistry, **Mrs. Gaëlle Fontaine** for her help and training in cell culture. **Dr. Yu Luo**, **Dr. Rahima Chennoufi** and **Dr. Jean Chatain** for our discussions and their good advices. In Italy, I thank **Dr. Alessandra Benassi** for her help in the laboratory and for the synthesis of the *N*-methylethylamide benzophenone, **Dr. Valentina Pirota** for her assistance and good advices, **Pr. Mariella Mella** and **Teresa Recca** for the NMR assistance. I would also like to thank all members of our unit in France and Italy for their openness to discussions, helpful advices and assistance for experiments.

I am grateful to my family, particularly to my mother, **Martine Peyrieres**, for unconditional support and assistance. I would like to thank all my friends for their support and particularly grateful to **Laure de Pontual** and **Clara Testard** for all our good time since we met.



## Résumé en français (French summary) :

Les G-quadruplexes (G4s) sont des structures à quatre brins non canoniques de l'ADN qui se forment par l'interaction d'au moins deux G-quartets par  $\pi$ -stacking.<sup>56</sup> Les G-quartets sont des structures tétramériques coplanaires composées de quatre guanines qui interagissent ensemble par des liaisons hydrogènes d'Hoogsteen. Ces G4s se forment uniquement lorsque l'ADN duplexe est déroulé et séparé en deux brins dont le brin riche en guanines forme le G4 avec la stabilisation indispensable par des cations potassiums ( $K^+$ ). De très nombreuses expériences *in vitro* et *in cellulo* supportent l'hypothèse de la formation des G4s *in vivo*.<sup>48</sup> De plus, la cartographie des structures G4s génomique fixée *in cellulo* et identifiées par un anticorps anti-G4 appelé « BG4 » a permis d'identifier plus de 10 000 de ses séquences formant des G4s.<sup>91,92</sup> Et les études *in silico* et *in vitro* de cartographie des structures G4s du génome ont permis de démontrer un enrichissement de ses structures au niveau d'éléments régulateurs qui permettent de réguler la transcription et la traduction.<sup>48</sup> Ces structures sont aussi présentes au niveau télomérique pour la protection des extrémités chromatiniennes et inhiber la télomérase responsable de l'extension télomérique,<sup>97</sup> et au niveau de sites de réplication pouvant jouer un rôle sur l'initiation de la réplication.<sup>48</sup> C'est pour ses raisons que les G4s sont devenues des cibles importantes d'études dues à l'importance des processus biologiques qu'elles régulent et sont souvent associées à un développement tumoral lors de leurs dérégulations.

L'objectif du projet de thèse est de développer de nouveaux ligands G4s fonctionnalisés par un groupement photoactivable et un alcyne terminal pour réaliser une méthodologie dite de marquage par photoaffinité. Cette dernière permet de convertir une interaction non covalente en une interaction covalente par l'activation UV d'un groupement photoactivable qui génère une espèce réactive qui va réagir pour former une liaison covalente à proximité du domaine d'interaction.<sup>185,186</sup> La présence d'un alcyne terminal doit permettre l'isolation des complexes [structures G4 : ligands G4] par l'introduction d'un haptène par chimie click. L'objectif à long terme de l'équipe est de développer une nouvelle méthodologie basée sur le marquage par photoaffinité pour la cartographie des structures G4s accessibles dans le génome. L'avantage est d'éviter la fixation chimique du génome pour garder le dynamisme génomique de la formation des G4s, et la formation d'une liaison covalente avec les G4s pour éviter une dissociation et relocalisation du ligand G4 avec un site génomique différent.

Pour cela, le ligand G4 fonctionnalisé est le PDC (aussi appelé « 360A ») qui est un ligand qui possède une grande affinité et sélectivité pour les G4s, qui a été développé dans le laboratoire et qui a déjà été adapté de très nombreuses fois due à sa simplicité de synthèse. Deux familles de composés ont été développés avec les ligands « **linéaires** » et les ligands « **branchés** ». Les ligands « **linéaires** » sont obtenus par la fonctionnalisation du PDC avec la chaîne linéaire 4-diazirine-7-octynoate qui porte la diazirine aliphatique comme groupement photoactivable et l'alcyne terminal. Le PDC et la chaîne linéaire sont reliés par un lien diamine C2, C4 ou diéthylène glycol (DEG). Ces ligands ont été synthétisés par Dr. T. Masson durant sa thèse. Les ligands « **branchés** » sont obtenus par la fonctionnalisation du PDC avec 1) un des trois groupements photoactivables : benzophénone (Bph), benzyl trifluorométhyl diazirine (N2) et tétrafluorophenyl azide (N3), et 2) un des quatre espaceurs portant la fonction alcyne terminal : C1, C3, DEG et triéthylène glycol (TEG). Ces ligands ont été synthétisés par Mme. C. Guetta sauf pour la série DEG qui ont été synthétisés par moi-même (*voir chapitre II.1.2. Synthesis of Branched DEG G4 ligands*, p.94).

En premier, il a été nécessaire de vérifier que la fonctionnalisation du PDC n'a pas un effet négatif significatif sur l'interaction et surtout sur la sélectivité versus de l'ADN duplexe. Pour cela, deux méthodes biophysiques ont été utilisées, le FRET-melting (mesure du point de fusion) et le G4-FID (déplacement d'une sonde d'intercalation fluorescente), sur un panel de six G4s couvrant une large diversité de structures. Les résultats (*voir II.2.3. Compounds evaluation*, p.106) montrent que les ligands « **linéaires** » conservent les propriétés excellentes du PDC pour l'interaction et la sélectivité envers les structures G4s. En revanche, les ligands « **branchés** » montrent que la fonctionnalisation résulte en une importante diminution de l'interaction et même de la sélectivité pour certaines structures G4 versus de l'ADN duplexe. Le classement général de l'interaction est : ligands « **linéaires** » > C3/N2 et C1/N3 > N3 > Bph et N2.

Secondement, la stratégie de marquage par photoaffinité des structures G4s a été confirmée par gel d'électrophorèse en condition dénaturante, ce qui permet de séparer les séquences G4s alkylés des séquences G4s non alkylés. Pour cela, les séquences G4s 22AG, cMYC22 et CEB25*wt* ont été marqués au niveau de l'extrémité 5' par du phosphore-32, puis équilibré avec un des ligands étudiés puis irradié à 360 nm pendant 1 h avant d'être séparé sur gel. Les résultats (*III.1. G4 Alkylation efficiency*, p.117) pour les ligands « **linéaires** » montrent un haut rendement d'alkylation pour 22AG (57-63%) et CEB25*wt* (64-72%), et un peu moins important avec cMYC22 (39-48%).

Il est intéressant d'observer une importante différence de rendement d'alkylation entre CEB25 *wt* et cMYC22 alors qu'ils partagent une topologie similaire. Cela peut s'expliquer parce que CEB25 *wt* a une longue boucle centrale de 9 nucléotides qui rend la structure beaucoup plus accessible à l'alkylation que les courtes boucles de cMYC22. Différemment les ligands « **branchés** », l'alkylation a l'air importante pour cMYC22 et CEB25 *wt* ( $\geq 50\%$ ). Cependant, le rendement de l'alkylation est extrêmement faible avec 22AG (**Bph** 8-18%, **N2** 3-5%, and **N3** 6-9%), ce qui démontre une préférence structurelle de l'alkylation des ligands « **branchés** » envers certains G4s. Pour conclure, le marquage par photoaffinité des G4s *in vitro* est efficace par les ligands « **linéaires** », et efficace mais structurellement sélectif par les ligands « **branchés** ».

Pour la suite de la thèse, nous avons décidé de nous concentrer sur la caractérisation de l'alkylation, que cela soit sur la structure G4 ou en utilisant un modèle pour identifier et caractériser le produit de l'alkylation avec des nucléosides. Et nous avons décidé de nous concentrer sur la réactivité des diazirines aliphatiques (**I-N2**) présent sur les ligands « **linéaires** » et sur la benzophénone présente sur la série **Bph** des ligands « **branchés** ».

Ce concentrant sur la benzophénone, la localisation des sites d'alkylation a été déterminé par séquençage (*voir III.2. Chemical and enzymatic sequencing methods*, p.122). Au début, un séquençage chimique a été tenté utilisant du diméthyle sulfate et de la pipéridine pour fragmenter la séquence d'ADN au niveau des guanines et des sites d'alkylations, mais résulta en une désalkylation, signifiant que l'alkylation est thermiquement réversible. L'utilisation d'un séquençage enzymatique utilisant l'exonucléase 3' qui digère l'ADN du 3' au 5' et est bloqué lors de la digestion d'un nucléotide alkylé. Cela a permis d'identifier la position de l'alkylation sur 22AG. Le résultat montre que l'alkylation par les ligands de la série **Bph** est sélectif pour la thymidine et indépendant de la position des boucles. Un résultat similaire avait été obtenu par Verga *et al.* avec le ligand PDC fonctionnalisé avec une benzophénone.<sup>179</sup> Pour aller plus loin, un modèle a été développé où la benzophénone est fonctionnalisée avec une fonction *N,N*-diéthylamide pour reproduire la fonction qui relie la benzophénone au reste du ligand sur la série **Bph**. La génération de l'espèce réactive du modèle « **Bph-COEt<sub>2</sub>** » a été évalué par Laser Flash Photolyse (LFP) avec une durée de vie limité à moins d'une dizaine de nanosecondes, ce qui est un comportement différent comparé à la benzophénone avec une durée de vie d'une dizaine de microsecondes (*voir IV.3. Laser flash photolysis studies*, p.145). Vu les résultats de l'expérience de séquençage, sa réactivité avec la thymidine a été évalué et montre une cinétique de la réaction lente mais avec la formation uniquement d'oxétanes avec majoritairement les régioisomères hémiaminales éthers et minoritairement les

régioisomères éthers qui ont été caractérisés par RMN 700 MHz (*voir IV.4. Benzophenone photoadducts distributions*, p.153). De plus, le traitement thermique des deux régioisomères montre que seul l'hémiaminal éther est thermiquement réversible alors que l'éther est thermiquement stable. Pour conclure sur la benzophénone, le modèle « **Bph-COEt<sub>2</sub>** » montre clairement une particularité avec sa durée de vie extrêmement réduite (< 50 ns) comparé à la benzophénone. Cependant, le modèle garde la même sélectivité pour les thymidines et forme aussi des oxétanes par une réaction de Paternò-Büchi comme pour la benzophénone. C'est, à ma connaissance, le premier groupement photoactivable qui possède un état triplet réversible avec une durée de vie au niveau du nanoseconde, ce qui pourrait permettre de limiter les dommages par photo-oxydation souvent associés à la benzophénone mais reste à être étudié et confirmé ou infirmé.

Pour finir en se concentrant sur la diazirine aliphatique qui est présente sur les ligands « **linéaires** », le séquençage chimique et enzymatique n'a pas permis de déterminer la position de l'alkylation sur les G4s. Un modèle a été développé par Mme. C. Guetta reprenant la chaîne portant la diazirine aliphatique et l'alcyne terminal en la fonctionnalisant avec une *N*-méthylamide « **-N<sub>2</sub>CONHMe** ». La génération de l'espèce réactive a été évalué par LFP mais n'a pas permis de la caractériser due à sa trop faible durée de vie. La réactivité du modèle a été évalué avec différents nucléosides et montre la formation d'adduits uniquement avec la thymidine (**dT**) et la 2'-deoxyguanosine (**dG**). La caractérisation par RMN 700 MHz de ses adduits (*voir IV.5. Study of the photoreactivity of aliphatic diazirine*, p.161) montre la formation d'un produit majoritaire (66% HPLC) pour les deux nucléosides correspondant à une insertion N-H avec NH3 pour **dT** et NH1 pour **dG**. Et la formation d'un produit minoritaire (33% HPLC) qui correspond à une insertion O-H avec O2/O4 pour **dT** et O6 pour **dG**. La compétition de réactivité entre **dT** et **dG** avec le modèle montre une préférence marquée pour **dT**. Pour conclure, la diazirine aliphatique montre une préférence pour l'alkylation intermoléculaire de site acide comme déjà observé avec les acides aminés par West *et al.*<sup>198</sup> Le mécanisme en jeu est sûrement due à la réactivité de l'espèce réactive diazoisomère qui a été caractérisé par O'Brien *et al.*<sup>233</sup> comme l'espèce réactive réagissant avec les sites acides, et par la formation d'un intermédiaire en pair d'ions qui expliquerait la proportion aussi élevée d'insertion O-H (33%) qui ne serait pas possible d'obtenir par une insertion O-H directe à partir des formes énols. Cependant, la réactivité intramoléculaire avec les structures G4s doit être plus complexe avec une partie similaire avec la réactivité intermoléculaire due au diazoisomère mais avec en addition la réactivité du carbène qui reste à être évalué.





# Contents

<b>ACKNOWLEDGEMENTS</b> .....	<b>9</b>
<b>RESUME EN FRANÇAIS (FRENCH SUMMARY)</b> : .....	<b>11</b>
<b>CONTENTS</b> .....	<b>17</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>21</b>
<b>CHAPTER I: INTRODUCTION</b> .....	<b>25</b>
1. NUCLEIC ACIDS .....	27
1.1. RNA specificities .....	30
1.2. Non-canonical nucleic acid structures .....	31
2. THE G-QUADRUPLEXES.....	33
2.1. G-quadruplex structures and topologies.....	34
2.2. Structural and topological characterisation .....	38
2.2.1. Circular dichroism spectroscopy .....	39
2.2.2. DMS footprinting .....	40
2.2.3. Characterisation of the molecular structure .....	41
2.3. G-quadruplex genomic localisation.....	42
2.3.1. In silico approach.....	42
2.3.2. In vitro and in vivo methodologies .....	43
2.4. G-quadruplex biological roles .....	47
2.4.1. Telomeres .....	48
2.4.2. G-quadruplex as regulatory elements.....	49
3. G-QUADRUPLEX TARGETING .....	52
3.1. G-quadruplex Ligands .....	52
3.2. Anti-G-quadruplex antibodies.....	61
3.3. Interaction studies.....	62
3.3.1. FRET-melting.....	63
3.3.2. G4-FID .....	65
3.3.3. CD.....	68
3.3.4. UV-visible and fluorescence titration .....	68
4. G-QUADRUPLEX ALKYLATION .....	70
4.1. Thermally activatable compounds and metalation agents .....	70
4.2. Photoactivatable G4 ligands.....	73
4.3. Photoaffinity labelling.....	75
<b>AIMS OF THE WORK</b> .....	<b>81</b>
<b>CHAPTER II: SYNTHESIS OF PHOTOACTIVATABLE G4 LIGANDS AND BIOPHYSICAL EVALUATION</b> .....	<b>87</b>
1. SYNTHESIS .....	92
1.1. PDC-NH <sub>2</sub> synthesis .....	92
1.2. Synthesis of <b>Branched DEG</b> G4 ligands.....	94
1.2.1. Synthesis of the common linker intermediate .....	95
1.2.2. Synthesis of the photoactivatable G4 ligands .....	96
2. BIOPHYSICAL EVALUATION OF THE PHOTOACTIVATABLE G4 LIGANDS .....	98
2.1. DNA and RNA sequences used in the biophysical studies.....	98
2.1.1. G4 DNA sequences and related topologies.....	99
2.1.2. RNA sequence and related topology .....	102
2.2. FRET-melting and G4-FID assays.....	103

2.2.1. FRET-melting protocol.....	103
2.2.2. G4-FID protocol .....	104
2.2.3. TO binding to VEGF G-quadruplex.....	105
<b>2.3. Compounds evaluation .....</b>	<b>106</b>
2.3.1. <b>22AG/F21T</b> .....	106
2.3.2. <b>22CTA/F21CTAT</b> .....	107
2.3.3. <b>cMYC22/FcMYC22T</b> .....	108
2.3.4. <b>CEB25wt/FCEB25wtT</b> .....	109
2.3.5. <b>VEGF/FVEGFT</b> .....	110
2.3.6. <b>FHIVPRO1T</b> .....	111
2.4. <i>RNA G-quadruplex in FRET-melting</i> .....	112
2.5. <i>Conclusion</i> .....	113
<b>CHAPTER III: PHOTOCHEMICAL REACTIVITY WITH G-QUADRUPLEX STRUCTURES.....</b>	<b>115</b>
1. G4 ALKYLATION EFFICIENCY .....	117
1.1. <i>Photoreactivity of in-line compounds</i> .....	119
1.2. <i>Photoreactivity of the branched compounds</i> .....	120
2. CHEMICAL AND ENZYMATIC SEQUENCING METHODS.....	122
2.1. <i>Chemical sequencing on the in-line alkylated adducts</i> .....	122
2.2. <i>Chemical sequencing on the benzophenone alkylated adducts</i> .....	124
2.3. <i>Enzymatic sequencing with 3'-exonuclease</i> .....	125
2.3.1. 3'-exonuclease treatment of <b>in-line</b> covalent adduct.....	125
2.3.2. 3'-exonuclease treatment of <b>benzophenone</b> covalent adduct .....	127
3. CONCLUSION.....	131
<b>CHAPTER IV: DETECTION OF THE REACTIVE SPECIES BY LFP AND STRUCTURAL CHARACTERISATION OF THE PHOTOADDUCTS BY NMR .....</b>	<b>133</b>
1. PRESENTATION OF THE PHOTOACTIVATABLE MOIETIES .....	135
1.1. <i>Benzophenone</i> .....	135
1.2. <i>Aryl azide</i> .....	137
1.3. <i>Diazirine</i> .....	138
2. SYNTHESIS OF THE PROTOTYPE PHOTOACTIVATABLE MOIETIES .....	141
2. PHOTOPHYSICAL STUDIES.....	142
2.1. <i>Photoactivatable moieties</i> .....	142
2.2. <i>In-line and branched compounds</i> .....	144
3. LASER FLASH PHOTOLYSIS STUDIES .....	145
3.1. <i>Photochemical investigation of benzophenone derivatives</i> .....	146
3.1.1. Detection and characterisation of the reactive species.....	147
3.1.2. Reactivity study with nucleosides .....	149
3.2. <i>Photochemical characterisation of the In-line prototype</i> .....	152
4. BENZOPHENONE PHOTOADDUCTS DISTRIBUTIONS.....	153
4.1. <i>Product distribution</i> .....	153
4.2. <i>Structural characterisation of the photoadducts</i> .....	154
5. STUDY OF THE PHOTOREACTIVITY OF ALIPHATIC DIAZIRINE .....	161
5.1. <i>Photolysis study</i> .....	161
5.2. <i>Photoproduct distribution</i> .....	164
5.3. <i>Structural characterisation of the photoadducts</i> .....	166
5.3.1. <b>dT</b> adducts .....	166
5.3.2. <b>dG</b> adducts.....	169
5.4. <i>Competitive product distribution</i> .....	173
6. CONCLUSION.....	174

6.1. Conclusion on the benzophenone derivatives.....	174
6.2. Conclusion on the <i>in-line</i> prototype.....	175
<b>CONCLUSIONS AND PERSPECTIVES.....</b>	<b>177</b>
<b>APPENDIX .....</b>	<b>187</b>
<b>EXPERIMENTAL SECTION.....</b>	<b>191</b>
<b>SYNTHESIS.....</b>	<b>193</b>
<b>BIOPHYSICAL EXPERIMENTS.....</b>	<b>212</b>
<b>BIOCHEMISTRY EXPERIMENTS.....</b>	<b>214</b>
<b>PHOTOCHEMICAL EXPERIMENTS.....</b>	<b>216</b>
<b>BIBLIOGRAPHY .....</b>	<b>219</b>



## List of abbreviations

### 5

5-BrdU: 5-bromo-2'-deoxyuridine

### A

A: Adenine

### B

Bph: Benzophenone

Bph-CONEt<sub>2</sub>: *N,N*-diethylamide benzophenone

Bph-CONHMe: Methylamide benzophenone

Bph-COOH: Carboxylic acid benzophenone

BSA: Bovine Serum Albumin

### C

C: Cytosine

CD: Circular Dichroism

Chem-IP-seq: Chemical immunoprecipitation sequencing

ChIP: Chromatin immunoprecipitation

CuAAC: Copper(I)-Catalysed Alkyne-Azide Cycloaddition

CUT&Tag: Cleavage Under Targets and Tagmentation

### D

dA: 2'-Deoxyadenosine

dC: 2'-Deoxycytidine

DC<sub>50</sub>: Concentration of ligand require to displace 50% of the Thiazole Orange

DEG: Diethylene glycol

dG: 2'-Deoxyguanosine

DMS: Dimethyl sulfate

DNA: Deoxyribonucleic acid

DOSY: Diffusion-Ordered Spectroscopy

dsDNA: Double-Stranded DNA

dT: Thymidine

dU: 2'-Deoxyuridine

### F

FAM: Fluorescein

FID: Fluorescence Intercalator Displacement

FRET: Förster Resonance Energy Transfer

## G

G: Guanine

G4: G-quadruplex

G4P: G4 potential

## H

HMBC: Heteronuclear Multiple Bond Correlation

HSQC: Heteronuclear Single Quantum Correlation

## I

I-motif: Intercalated-motif

ISC: Intersystem crossing

## K

$K_a$ : Association constant

$k_D$ : Dissociation constant

$k_{obs}$ : Observed rate constant

## L

LFP: Laser Flash Photolysis

*l*-N<sub>2</sub>: Aliphatic diazirine

*l*-N<sub>2</sub>COOH: Aliphatic diazirine carboxylic acid prototype

*l*-N<sub>2</sub>NHMe: Aliphatic diazirine methylamide prototype

## M

miRNA: MicroRNA

mRNA: Messenger RNA

## N

N<sub>2</sub>: Benzyl trifluoromethyl diazirine

N<sub>3</sub>: Tetrafluorophenyl azide

ncRNA: Non-Coding RNA

NDI: Naphthalene diimide

NGS: Next-Generation Sequencing

NHS: Nuclease Hypersensitive Site

NMR: Nuclear Magnetic Resonance

NOESY: Nuclear Overhauser Effect Spectroscopy

nt: Nucleotide

## O

OQ: Observed G-quadruplex

## P

PAL: Photoaffinity Labelling

PDC: Pyridodicarboxamide

PhenDC: Phenanthroline derivatives

PQS: Putative G-quadruplex Sequence

Pu: Purine-rich

## Q

QM: Quinone-Methide

QMP: Quinone-Methide Precursor

## R

RNA: Ribonucleic acid

## S

scFv: Single-Chain Variable Fragment

Seq: Sequencing

siRNA: Small Interfering RNA

## T

T: Thymine

TAMRA: Tetramethylrhodamine

TASQ: Template-Assembled Synthetic G-Quadruplex

TEG: Triethylene glycol

T<sub>m</sub>: Melting Temperature

TO: Thiazole Orange

TOD: Thiazole Orange Displacement

tRNA: Transfer RNA

TSS: Transcription Starting Site

## U

U: Uracil ; Uridine

UTR: Untranslated region

## Δ

ΔT<sub>½</sub>: Stabilisation temperature



Chapter I:  
Introduction



# 1. Nucleic acids

Nucleic acids are macromolecules essential to life and are composed of two families: the deoxyribonucleic acid (DNA) and the ribonucleic acid (RNA). DNA was first isolated as a DNA/protein mixture by Miescher who called it Nuclein.<sup>1,2</sup> Later, the mixture was purified to give DNA and was mainly characterised by Miescher, Kossel and Altmann, the latter named the pure substance nucleic acid.<sup>3</sup> The work of Kossel shows that the nucleic acid contained pyrimidine and purine structures, a sugar and a phosphate group.<sup>4-6</sup> It was finally demonstrated that the sugar is a 2'-deoxyribose, which gave its name deoxyribonucleic acid, and the four bases were identified.

Deoxyribonucleic acid is a biopolymer constituted by monomeric units called nucleotides represented in Figure 1-A. The latter are divided into three subunits: a 2'-deoxyribose sugar, one to three phosphate groups (mono-, di- or triphosphate nucleotide) connected to the C5' of the sugar, and one of the four nucleobases tethered to the anomeric carbon (C1') through an N-glycosidic bond. The monophosphate nucleotides are joined together in a linear chain through a 3',5'-phosphodiester linkage between the 5' phosphate group of a nucleotide and the 3'-hydroxyl group of the subsequent nucleotide. It forms an alternating 2'-deoxyribose-phosphate backbone represented in Figure 1-B.

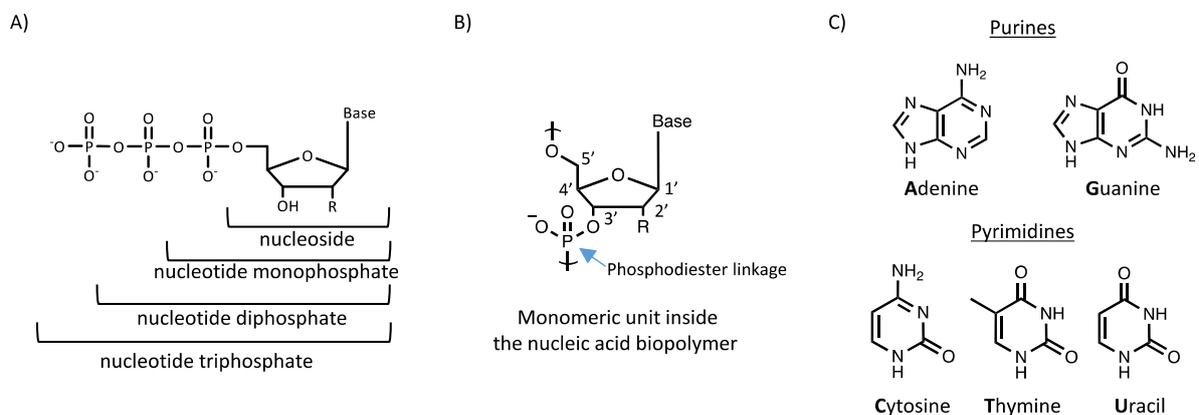


Figure 1: A) General molecular structure of the nucleoside and nucleotides. B) Structure of the monomeric unit inside the linear chain of the nucleic acid biopolymer. R = -H for DNA and -OH for RNA. C) Molecular structure of the different purine and pyrimidine bases.

The nucleobases can be separated into two structural groups shown in Figure 1-C: the purines with Adenine (**A**) and Guanine (**G**), and the pyrimidines with Cytosine (**C**) and Thymine (**T**) (the latter is replaced by Uracil (**U**) in RNA sequences). Chargaff discovered that A/T and G/C had the same proportion inside a pair.<sup>7,8</sup> It was later explained as A is complementary to T by forming two hydrogen bonds, and C complements G by forming three hydrogen bonds as shown in Figure 2-A. These hydrogen bonds are called Watson-Crick-Franklin hydrogen bonds. Their formation drives the two complementary polynucleotide strands to interact together by running in opposite directions from each other, i.e. the two strands are antiparallel, forming the antiparallel right-handed double helix DNA represented in Figure 2-B.<sup>9,10</sup>

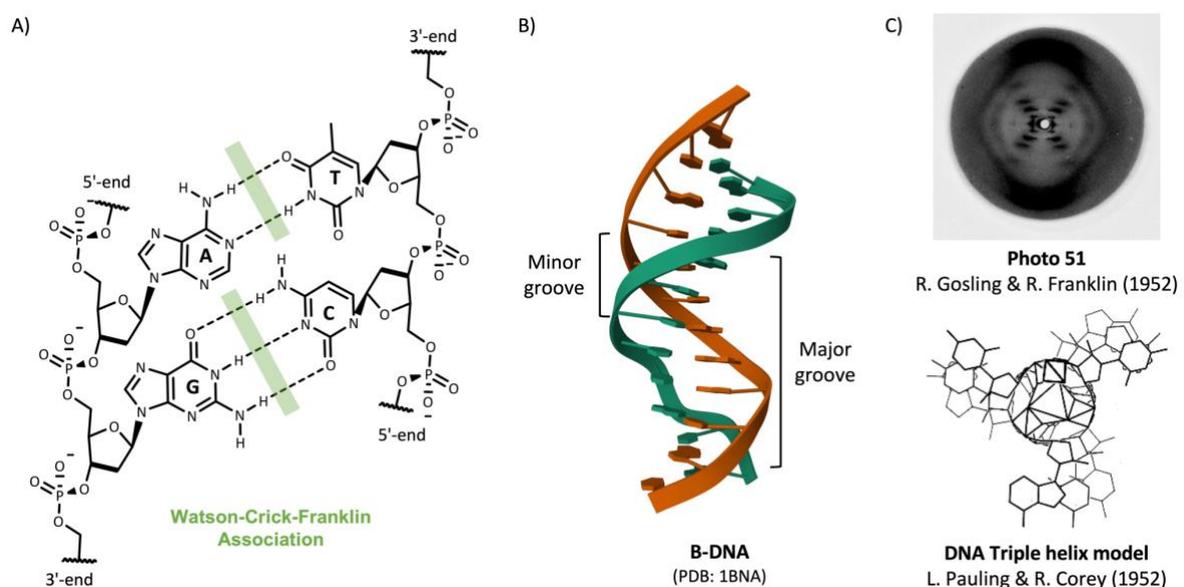


Figure 2: A) Double-strand structure with the base pair association. B) Example of B-DNA structure from a B-DNA dodecamer resolved by X-ray crystallography (PDB: 1BNA).<sup>10</sup> C) Photo 51 (top) obtained by Gosling and Franklin in 1952 of a B-DNA fibre crystal, and the DNA triple helix representation of Pauling and Corey (Pauling and Corey, 1952).<sup>12</sup>

An intensive focus was given to the characterisation of the three-dimensional structure of the DNA during the mid-twenty. It was only in 1953 that Watson and Crick, and later acknowledged Franklin, published the double helix structure of the DNA<sup>11</sup> inspired by the triple helix DNA model of Pauling and Corey<sup>12</sup> and by the Photo 51 from Franklin and Gosling<sup>13</sup> both represented in Figure 2-C. Photo 51 is a bold X-ray diffraction picture of the hydrate DNA crystal showing a clear helicoidal structure. The double helix described by Watson and Crick corresponds to the most common form of DNA found in nature, called B-DNA. The double helix is right-handed with 10.5 base pairs per turn and contains a 22 Å wide major groove and 12 Å wide minor groove.<sup>14</sup>

At first, DNA was considered to be too simple to store hereditary information as it contains only four bases, and proteins were considered to be better candidates with their twenty-two amino acids. Griffith experiments<sup>15,16</sup> and Avery-MacLeod-McCarty<sup>17</sup> experiments demonstrated the storage of the hereditary information on the DNA by transforming non-virulent bacteria into virulent bacteria uniquely via the transfer of DNA originating from a virulent bacteria strain. In addition, this was confirmed by Hershey and Chase<sup>18</sup> observing radiolabelled viral DNA transfer from bacteriophages to bacteria.

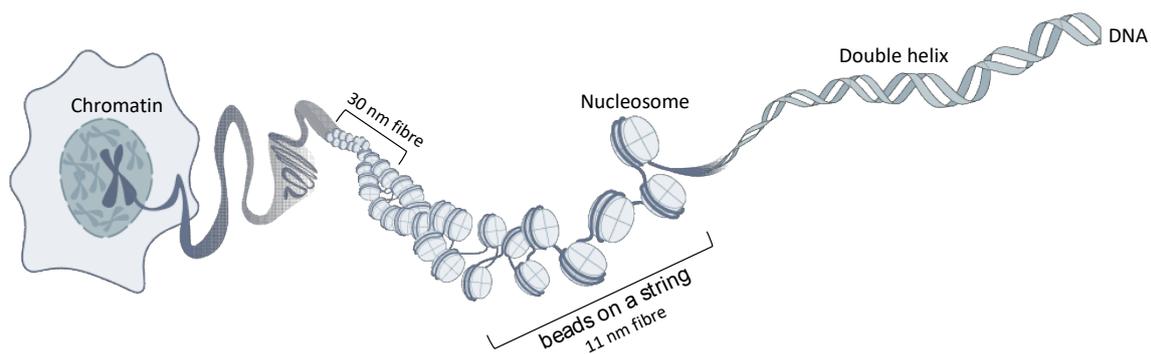


Figure 3: Chromatin organisation inside the nucleus. The DNA wrap around the nucleosome to form the "beads on a string" structure. A more compact structure can be achieved by internucleosomal interaction to form 30 nm fibre structure.

In the eukaryote, the genetic information is divided into long linear DNA molecules called chromosomes. Each chromosome can be present in more than one copy, e.g. mammalian genomes are composed of homologous copies of each chromosome, i.e. mammals are diploid organisms.<sup>19</sup> The DNA is stored with different compaction levels in the nucleus as a substance called chromatin shown in Figure 3. The 2 nm double helix DNA is wrapped around histone octamer to form the nucleosome as the basic structural unit. Nucleosome repeats form the "beads on a string" 11 nm fibre, the first order packaging structure. Adding H1 histone to the nucleosomes compacted them in a 30 nm fibre through internucleosomal interaction.<sup>20</sup> These two levels of DNA compaction form the active euchromatin (i.e. open chromatin). Further compaction can be achieved with scaffold proteins forming the heterochromatin (i.e. closed chromatin). Furthermore, the chromosomes can achieve ultimate compaction by forming X-shape chromosomes during the metaphase. The DNA packaging impact directly the activity of the chromatin: the euchromatin is characterised by high transcriptional activity, whereas the heterochromatin is silenced.

## 1.1. RNA specificities

Ribonucleic acid is similar to the DNA structure except that ribose replaces the 2'-deoxyribose, and uracil replaces the thymine. The rest of the structure represented in Figure 4-A remains identical, in which each nucleotide is bridged by a 3',5'-phosphodiester linkage and with the base pairs A/U and G/C. However, RNA molecules often exist in a single-stranded form capable of forming intramolecular secondary structures or as DNA to form a double-strand by interacting with a complementary strand. Another difference with DNA is the short lifetime of the RNA molecules due to a high regulation *in vivo* to regulate the translation of protein.<sup>21-23</sup> In addition, the RNA molecules are also less stable *in vitro* due to the presence of the 2'-hydroxyl that is susceptible to hydrolyse the phosphodiester linkage as shown in Figure 4-B.<sup>24</sup>

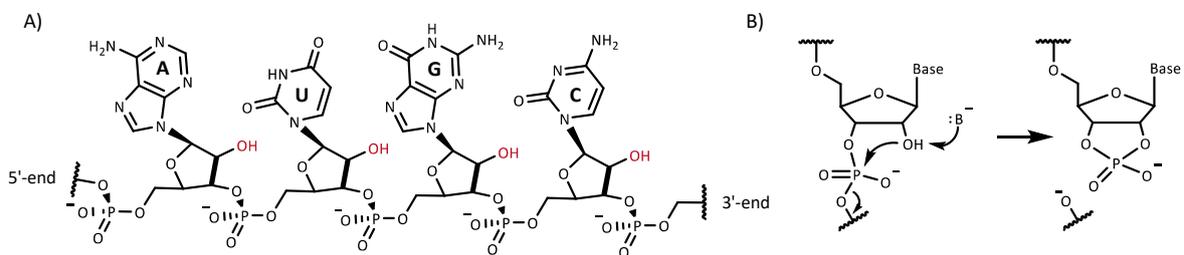


Figure 4: A) Structure representation of the single stranded RNA with the 2'-hydroxyl in red. B) Schematic representation of the phosphodiester hydrolyses by the attack on the 2'-hydroxyl group.

RNA is well-known for the coding messenger RNAs (mRNAs) that are translated into protein by the ribosome. However, RNAs are not limited to mRNAs and many non-coding RNAs (ncRNAs) exist for many biological purposes, and they can fold into various structures.<sup>25</sup> For example, microRNAs (miRNAs) or small interfering RNAs (siRNAs) can interact with the 5'UTR of an mRNA to form a double strand and prevent its translation. This duplex interaction forms an antiparallel right-handed double helix similar to the B-DNA called A-RNA shown in Figure 5-A.<sup>26,27</sup> Not only RNA can form a RNA duplex but RNA:DNA hybrids are known to exist such as represented in Figure 5-B and to have important biological functions.<sup>28</sup> For example, initiation of the DNA replication requires complementary RNA primers as starting site for the DNA polymerase.<sup>29</sup> Another interesting and important structure is the transfer RNA (tRNA) that transports the amino acids during translation of mRNAs into proteins by the ribosome. The tRNA structure comprises four hairpin loops forming a three-dimensional L structure shown in Figure 5-C.<sup>30,31</sup>

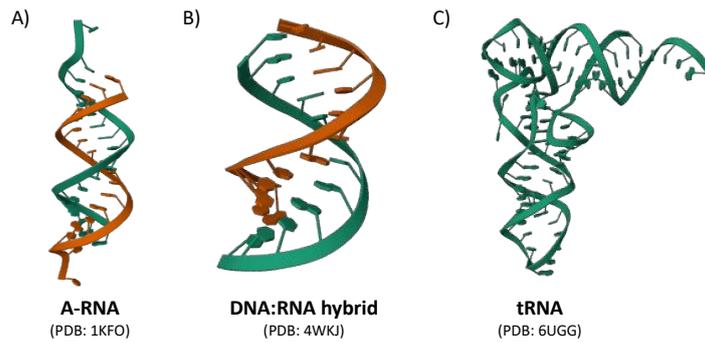


Figure 5: Three-dimensional structure of diverse RNA structures such as A) A-RNA double helix (PDB: 1KFO),<sup>27</sup> B) DNA:RNA hybrid (PDB: 4WKJ),<sup>28</sup> and C) tRNA (PDB: 6UGG).<sup>31</sup>

## 1.2. Non-canonical nucleic acid structures

Nucleic acid sequences are very similar, due to a low diversity of the nucleobases, yet forming a large diversity of three-dimensional structures. For example, the duplex DNA presented previously as forming a B-DNA structure can also switch transiently to A-DNA or Z-DNA. A-DNA exhibited in Figure 6-B is an antiparallel right-handed double helix,<sup>32</sup> more compact than B-DNA, observed in dehydrated conditions and during protein binding if water is stripped off from the DNA.<sup>33,34</sup> Z-DNA displayed in Figure 6-D is an antiparallel left-handed double helix,<sup>35</sup> considered to play a significant role in transcription.<sup>36</sup> The typical Watson-Crick-Franklin interaction (Figure 6-A) can also be observed intramolecularly with the formation of hairpin (also called stem-loop) represented in Figure 6-E,<sup>37</sup> when two regions of the same strand, usually complementary or palindromic sequences, base-pair to form an intramolecular double helix. Hairpins are essential secondary structures in RNA as implicated in the mRNA protection and transcription termination.<sup>38</sup>

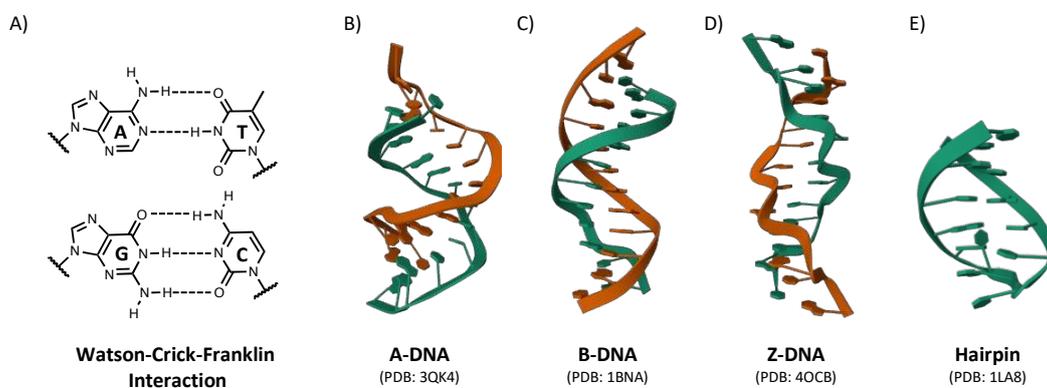


Figure 6: A) Watson-Crick-Franklin interaction of the DNA base pairs, and three-dimensional structure of the B) A-DNA (PDB: 3QK4),<sup>32</sup> C) B-DNA (PDB: 1BNA),<sup>10</sup> D) Z-DNA (PDB: 4OCB),<sup>35</sup> and E) hairpin (PDB: 1LA8).<sup>37</sup>

However, the Watson-Crick-Franklin interaction is not the only interaction possible between the nucleobases and other types of interaction can be implicated in the formation of intermolecular non-canonical DNA structures. The association of a third strand can generate the formation of triplex DNAs (H-DNA) as shown in Figure 7-A,<sup>39</sup> and is associated to the regulation of gene expression.<sup>40</sup> The association of three DNA strands among each other can generate the formation of a three-way junction as displayed in Figure 7-B that can form transiently during DNA replication.<sup>41,42</sup>

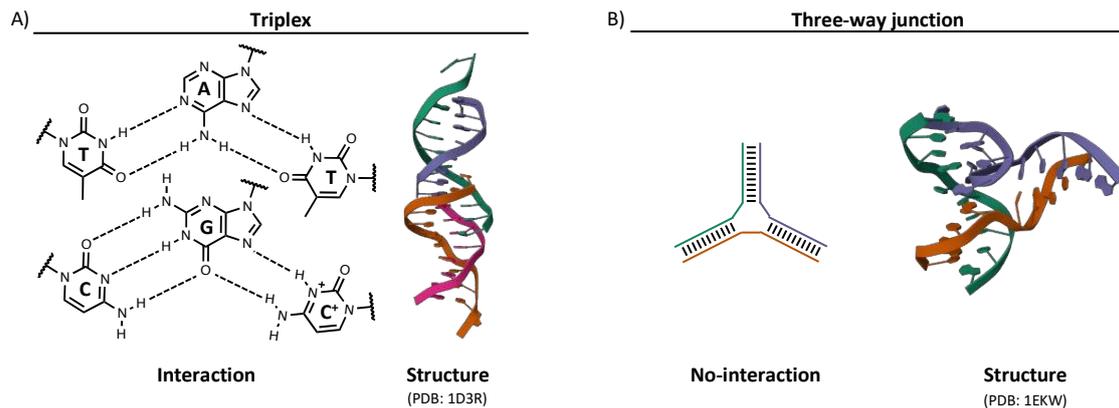


Figure 7: Representation of the interaction (left) and three-dimensional structure (right) of the A) triplex DNA (PDB: 1D3R)<sup>39</sup> and B) three-way junction (PDB: 1EKW).<sup>42</sup>

In addition, nucleic acid molecules can adopt non-canonical single- and multi-stranded structures such as G-quadruplexes and i-motifs. G-quadruplex structures represented in Figure 8-A are going to be detailed in the next section.<sup>43</sup> I-motif stands for intercalated-motif DNA and forms in repetitive C-rich regions by the association of cytosine-protonated cytosine (C-C<sup>+</sup>) base pairs stacking perpendicularly upon each other as shown in Figure 8-B.<sup>44,45</sup> As one of the cytosines must be protonated, i-motif forms at slightly acidic conditions, but evidence has shown *in vivo* formation in human cells.<sup>46</sup> The biological role played by i-motifs is yet to be totally defined but evidence suggests that they play a role in gene regulation and transcription.<sup>47</sup>

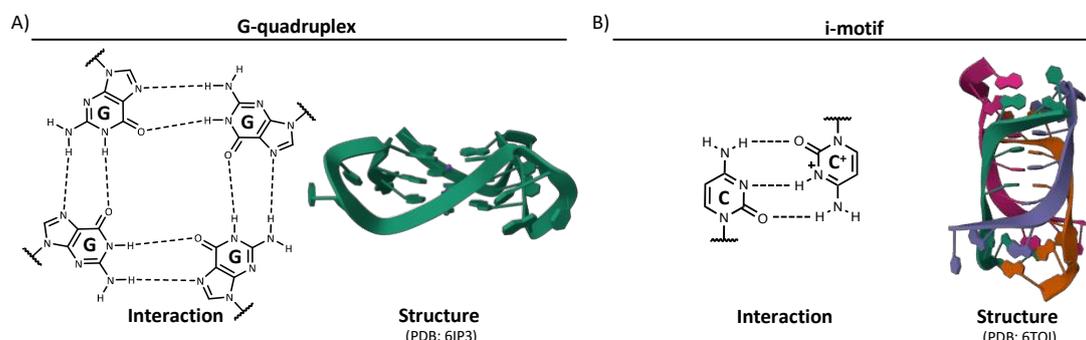


Figure 8: Representation of the interaction (left) and three-dimensional structure (right) of the A) G-quadruplex (PDB: 6IP3)<sup>43</sup> and B) i-motif (PDB: 6TQI).<sup>45</sup>

## 2. The G-quadruplexes

G-quadruplexes (G4s) are non-canonical four-stranded structures represented in Figure 9-A, generated by the folding of G-rich sequences and are formed by a minimum of two G-quartets interacting by  $\pi$ -stacking with each other. G-quartet is a square co-planar tetrameric structure composed of four guanines that interact among each other as shown in Figure 9-B, through Hoogsteen hydrogen bonding involving for each of them: the N1H and N2H as hydrogen donors, and N7 and O6 as hydrogen acceptors.<sup>48</sup> These structures are stabilised by potassium inner cations localised midway two G-quartets and forming coordination bonds with the eight O6 atoms.<sup>49</sup>

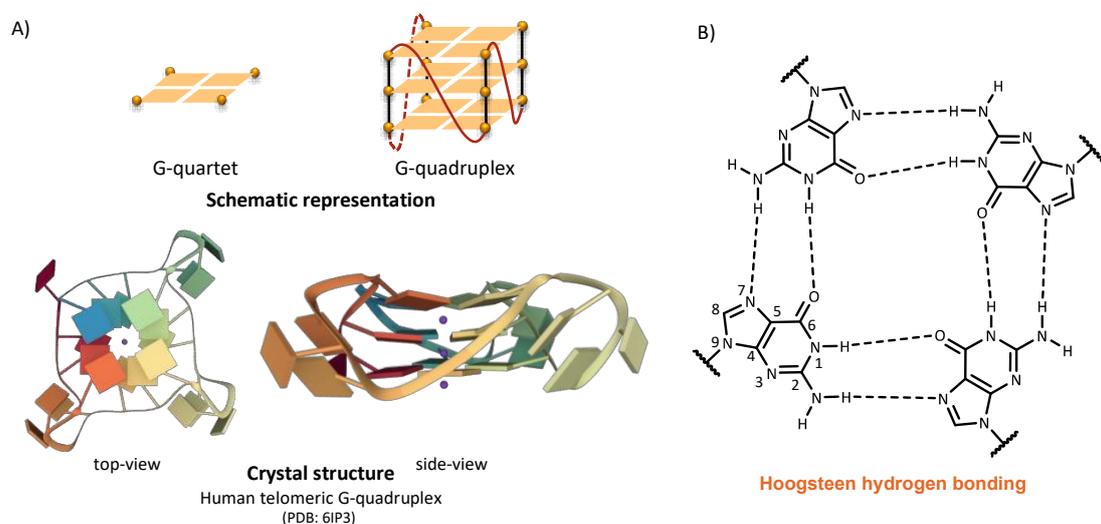


Figure 9: A) Schematic (top) and three-dimensional structure (bottom) of the parallel G-quadruplex of the human telomeric sequence (PDB: 6IP3).<sup>43</sup> B) Molecular representation of the G-quartet structure composed of four guanines interacting together by Hoogsteen hydrogen bonding.

It has been first reported by Bang *et al.* that highly concentrated solutions of guanosine in water and at low temperatures were forming viscous gels.<sup>50</sup> Later, X-ray diffraction of guanosine fibres reported by Gellert *et al.* (1962) showed the formation of a G-quartet organisation.<sup>51</sup> Furthermore, the stacking of these G-quartets among each other was observed in polyguanylic acid by Zimmerman *et al.*<sup>52</sup> and Arnott *et al.*<sup>53</sup> (1974). The first trace of a biological G-quadruplex structure was published by Henderson *et al.* (1987) as a "novel intramolecular structure containing guanine-guanine base pairs" from G-rich telomeric sequences.<sup>208</sup> The authors presented a non-canonical self-associated intramolecular structure formed by guanine-guanine base pairing through non-Watson-Crick association as a hairpin. It is now known that the structure described was a G-quadruplex and not a hairpin as proposed originally in the article.

Furthermore, the G-quadruplex structure was first described by Sen *et al.*<sup>209</sup> (1988) and Sundquist *et al.*<sup>54</sup> (1989). They demonstrated the formation of guanine tetrads through Hoogsteen hydrogen bonding and the need for potassium cations to stabilise the structure.

## 2.1. G-quadruplex structures and topologies

G-quadruplexes are highly polymorphic structures, meaning they can adopt a variety of G4 topologies, depending not only on their sequences but also on environmental conditions, such as the nature and the concentration of the cations, and the presence of molecular crowding agents.<sup>55</sup> The first parameter to take into account is the molecularity: G-quadruplexes can be intramolecular structures, where one strand folds over into a monomolecular G4, or intermolecular structures, generated by the association of two (bimolecular), three (trimolecular) or four (tetramolecular) strands as displayed in Figure 10.<sup>56</sup>

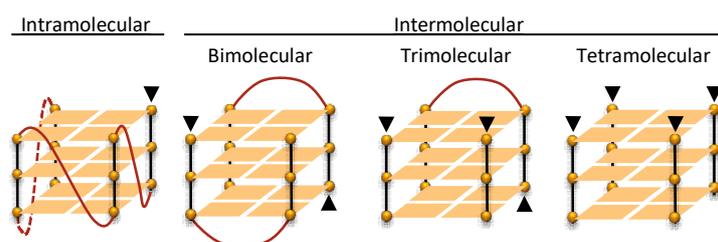


Figure 10: Structure representation of the intramolecular and intermolecular G4 structures. Black arrows indicate the starting of the 5'-end.

The stacking of the G-quartets defines the final topology adopted by a G-quadruplex, and each G-quartet conformation is defined by the glycosidic bond angle (*anti* or *syn* represented in Figure 11-A) adopted by its guanines. All the different glycosidic bond angle permutations define eight possible groove-width combinations for a G-quartet.<sup>57</sup> A G-quadruplex comprises four grooves characterised by three different widths: narrow, medium, and wide represented in Figure 11-B.<sup>56</sup> The groove width is defined by the adjacent guanines glycosidic bond angles, and not by their actual size. G-quartets can only stack with another G-quartet having a similar groove width combination, which finally defines the G4 topology. In theory, an intramolecular G-quadruplex can adopt 26 different topologies but so far, only seven topologies for now have been experimentally observed.<sup>58</sup>

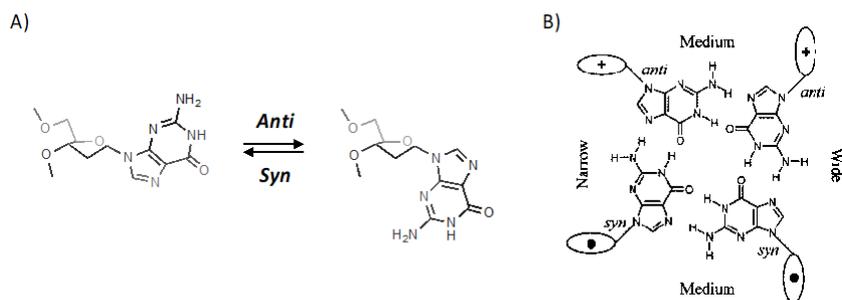


Figure 11: A) Guanine syn (left) and anti (right) conformations that can be interconvert by rotation around the anomeric bond. B) Example of G-quartet comprising one narrow, two medium, and one wide grooves (Figure from Simonsson, 2001).<sup>56</sup>

The second important component of the G-quadruplexes is the connecting loops. They are single-stranded sequences interconnecting the guanine tracts together and not interacting with the G-quartets. There are three basic types of loops: **propeller**, **lateral**, and **diagonal**.<sup>56,58</sup> A **propeller** loop (Figure 12-A) connects two adjacent guanines from two opposite quartets so that the guanine tracts are parallel. A **lateral** loop (Figure 12-B) connects two adjacent guanines from the same quartet so that the guanine tracts are antiparallel. A **diagonal** loop (Figure 12-C) connects two diagonally opposed guanines from the same quartet so that the guanine tracts are antiparallel. The length of the loop can directly impact its type. A short loop with a unique base will favour a **propeller** loop, one to two bases length favours a **lateral** loop, and a longer length is required for **diagonal** loop.

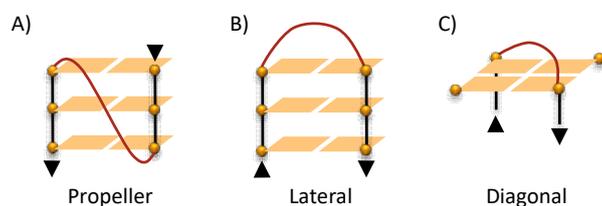


Figure 12: Representation of the connecting loops. The propeller and lateral are represented with one G-quadruplex face and the diagonal with only the G-quartet face. The black arrows indicate the orientation of guanine tracts.

In addition, the length and composition of the loop can impact the G-quadruplex stability. For example, the human minisatellite CEB25*wt* (L191) is a parallel G-quadruplex with a central 9-nt long propeller loop and two 1-nt propeller loops. Piazza *et al.*<sup>59</sup> demonstrated that the CEB25 stability increased as a function of the central loop length shortening as displayed in Figure 13 ( $T_m = 73.4$  °C for L111 and 55.1 °C for L191). Besides, variation in L111 loop composition shows that adenine was destabilising CEB25 ( $T_m = 56.5$  °C) as compared to cytosine and thymine equivalent ( $T_m = 74.7 - 73.4$  °C respectively).

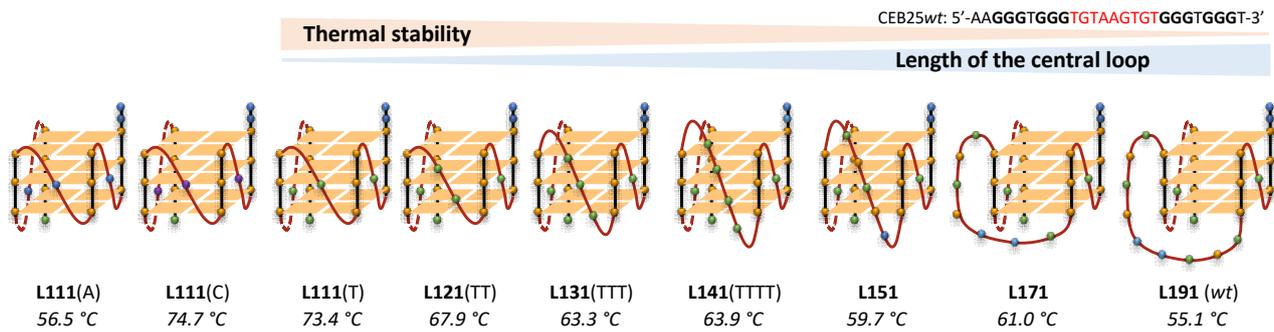


Figure 13: The stability of CEB25 is directly related to the length of the central loop.<sup>59</sup>

G-quadruplex structures are classified into three main topologies represented in Figure 14: **parallel**, **antiparallel**, and **hybrid**.<sup>56,60</sup> A **parallel** G4 has the four tracts with the same orientation, implicating that all the loops are propeller. An **antiparallel** G4 has two tracts with the same orientation and two tracts in the opposite orientation. It can form a “**chair type**” characterised by three lateral loops or a “**basket type**” with a diagonal loop and two lateral loops. A **hybrid** G4 is also called “3+1” as three tracts have the same orientation and one the opposite orientation, resulting in a propeller and two lateral loops. In the **hybrid-1** and **hybrid-2** forms, the G-quartets rearrangements are equivalent upon a 90° rotation.

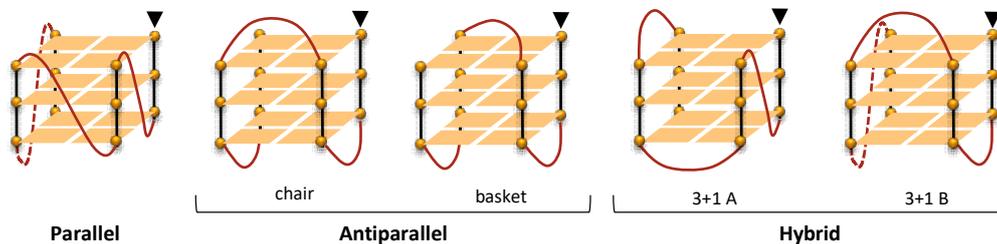


Figure 14: Representation of the different topologies that can adopt G-quadruplexes.

G-quadruplex topology can also be influenced by the nature of the inner cation. *In vitro* experiments are often performed in the presence of K-rich buffer as potassium is physiologically predominant in the cells.<sup>61</sup> However, G4s can chelate other alkali metal cations, alkaline earth metal cations and transition metals cations.<sup>62</sup> Together with potassium, the most employed cations in experimental conditions are sodium and lithium. Sodium cation has a smaller radius making the coordination less defined as the sodium can interact from the center of one G-quartet to midway two G-quartets or with a less symmetric intermediate between the two first.<sup>49</sup> This generates a lower stabilisation compared to potassium and can change the topology to a new one.

For example, the human telomeric G-quadruplex (22AG) adopts a hybrid topology in potassium<sup>63</sup> but an antiparallel topology in sodium<sup>64</sup> as displayed in Figure 15. Lithium cation does not stabilise G4 due to its small radius and is often used in experiments to balance ionic forces and as a negative control mimicking non-folding conditions.

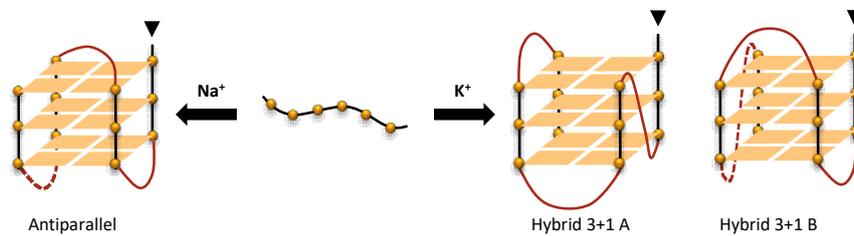


Figure 15: The human telomeric sequence 22AG can adopt multiple G4 conformation depending on the cationic conditions: 22AG adopts an antiparallel conformation in sodium<sup>64</sup> and a hybrid conformation in potassium.<sup>63</sup>

So far, we have been only describing the more classical G-quadruplex structures. However, several studies highlighted the formation of non-canonical G4 structures. G-quartets are composed of four guanines (G-G-G-G) but it is actually possible to observe mix-quartets when at least one guanine is replaced by another nucleobase. For example, 22CTA is a G-quadruplex comprising two G-quartets and one G-C-G-C mix-quartet where two pairs of C/G interact together as Watson-Crick-Franklin base pair (Figure 16-A).<sup>65-67</sup> The two pairs are coplanar and do not directly interact between each other, however, the guanines composing the two pairs interact through coordination with a potassium ion as displayed in Figure 16-B.<sup>67</sup>

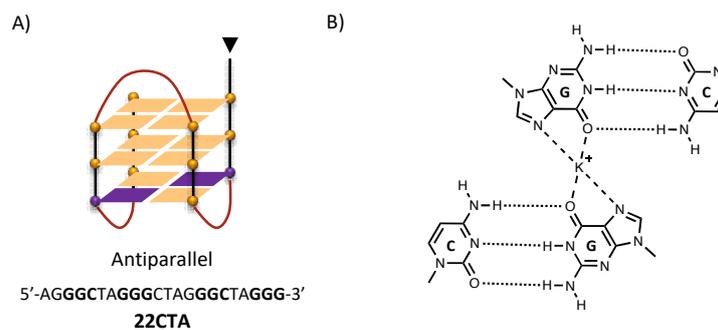


Figure 16: A) Representation of the 22CTA structure comprising a G-C-G-C mix-quartet.<sup>67</sup> B) the G-Cs interact as Watson-Crick-Franklin base pair. The two G-C pairs interact together uniquely through the potassium coordination by the guanines.

In addition, it is possible to observe Watson-Crick-Franklin base pair interacting through  $\pi$ -stacking with the external G-quartets such as for HIV-1 G4 represented in Figure 17. HIV-Pro1 G-quadruplex presents a G-C Watson-Crick pair stacking with the upper G-quartet.<sup>68</sup> Another related example is represented by the longer sequence of HIV-1 LTR-III, which was reported to form a G-C hairpin with a GTG apical loop.<sup>69</sup>

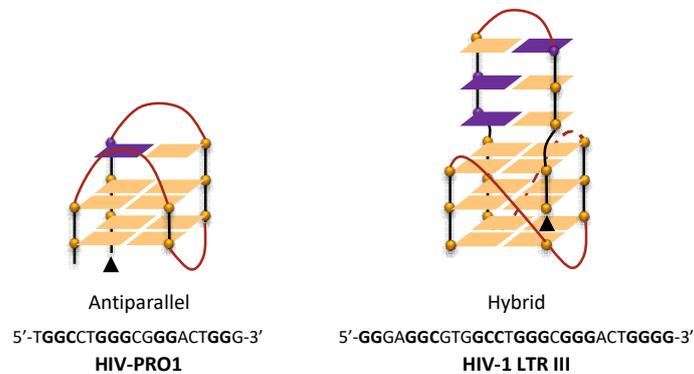


Figure 17: Representation of the HIV-PRO1 (left)<sup>68</sup> and HIV-1 LTR III (right)<sup>69</sup> comprising a GC Watson-Crick base pair that form a hairpin in HIV-1 LTR III.

## 2.2. Structural and topological characterisation

Due to all this structural variability, G-quadruplex structures require tools for their characterisation since each G4 sequence can have structural particularities that are hard to predict using *in silico* approaches. Thus, it is essential to have powerful methods to characterise the G4 structures formed by a given sequence, from its general topology using Circular Dichroism (CD) spectroscopy to its fine structure using NMR spectroscopy and X-ray crystallography.

### 2.2.1. Circular dichroism spectroscopy

A routine method to determine or confirm G-quadruplex folding topology is the circular dichroism (CD) spectroscopy. This technique allows to determine the chirality of molecules through their optical activity by measuring the difference in absorption of the left and right circularly polarised light. This method has been widely applied to determine the secondary structure of biomolecules such as proteins and nucleic acids because they present a specific CD spectral signature, e.g. B-form duplex DNA is characterised by a maximum at 270-280 nm and a minimum at 245 nm.<sup>70</sup>

G-quadruplexes are chiral molecules and each topology has a characteristic signature represented in Figure 18:<sup>71,72</sup> Parallel structures have a maximum at 260 nm and a minimum at 240 nm. Antiparallel structures form two maxima at 240 nm and 295 nm, and a minimum at 260 nm. Two maxima characterise hybrid topology at 270 nm and 290 nm, and a minimum at 240 nm.

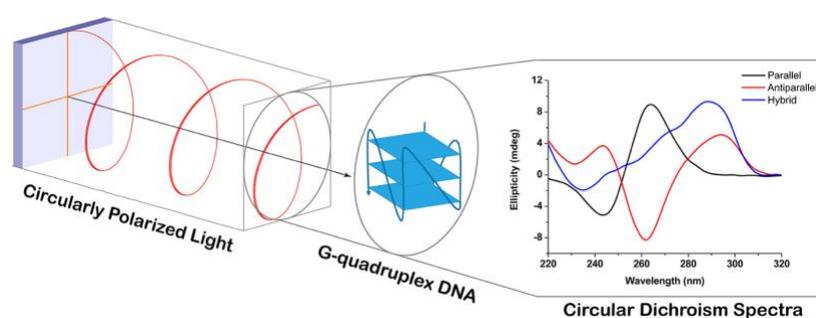
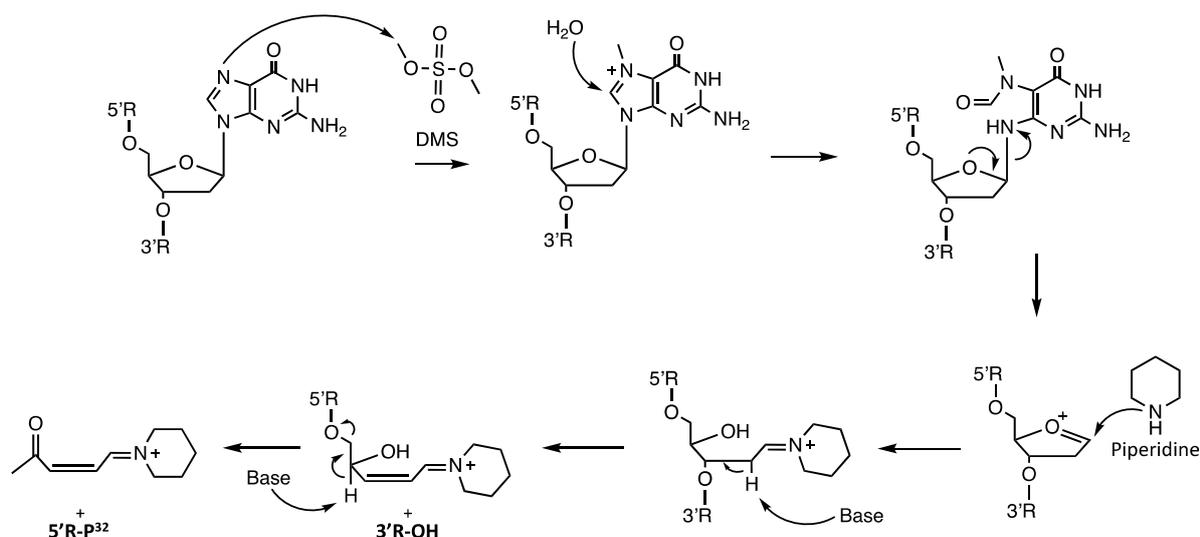


Figure 18: Representation of the CD experiment and of the characteristic spectral signatures for each topology (figure from Carvalho et al., 2017).<sup>71</sup>

The advantages of this method are: i) the easy preparation of the sample, and ii) its non-destructive character. The disadvantage is the qualitative nature of the information obtained by CD: The topology of the more populated structure generated by the sequence is observed, and it is not impossible to have a full overview of all the possible conformations present in solution. In addition, the spectra must be carefully analysed since the CD signal is an additive signal and is generated by all the species present in the solution, and therefore, it could lead to misinterpretation.

### 2.2.2. DMS footprinting

G-quadruplex sequences may have more guanines in their sequences than the ones involved in the formation of the G-quartets. This implies that not necessarily all the guanines present in the sequence are involved in G4 formation. Guanines involved in G-quartet formation can be identified using Maxam and Gilbert sequencing. It is a method developed by Maxam and Gilbert in 1977 using a chemical DNA-cleavage agent, specific for a nucleobase, to sequence a radiolabelled DNA.<sup>73,74</sup> This method has regained popularity for guanine sequencing using dimethyl sulfate (DMS) and is better known now as DMS footprinting.<sup>75</sup> Guanine cleavage mechanism is presented in Scheme 1: it starts with the methylation of the N7-position of the guanine by the DMS to form the unstable 7-methyl guanine that undergoes a base-mediated rearrangement ending with its elimination from the DNA backbone and DNA cleavage. The backbone cleavage end with a  $\beta$ -elimination on the C2' position by the piperidine to produce the  $^{32}\text{P}$ -5' fragment and the 3' fragment. The  $^{32}\text{P}$ -5' fragment can be revealed by autoradiography and each band corresponds to the -1mer of each cleaved guanine in the sequence.



*Scheme 1: Mechanism of the sequencing using DMS/piperidine: after methylation the basic treatment induces the  $\beta$ -elimination reaction which release the radiolabelled  $^{32}\text{P}$ -end 5'-fragment and the 3'-fragment.*

When the guanine is involved in a G-quartet, the N7 position is inaccessible to DMS methylation due to its involvement in the formation of Hoogsteen hydrogen bonds. Therefore, the absence of cleavage bands compared to the unfolded DMS-treated reference allows the identification of the guanines involved in the G-quartet formation.

### 2.2.3. Characterisation of the molecular structure

The previously described methods are indirect and give limited information on the G-quadruplexes structures. This limitation has been overcome by using X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy to achieve the resolution of the G4 molecular structures. The first provides precise structure and localisation of the coordinated cations within a single crystal. However, there is no dynamics and the obtained structure could be distorted by the crystal packing forces or not represent the dominant species in solution.<sup>76</sup> NMR spectroscopy allows an accurate identification of the structure in solution. Hence, these two methods give complementary information on the structural characteristics (quartets and loops geometries, cation localisation, and solvation).

NMR is a standard spectroscopic method used to identify and characterise the structure of both small organic molecules and macromolecules. For the identification of G-quadruplexes, it is important to have a single kinetic stable structure in solution for a complete characterisation avoiding peak overlapping. Thus, it is often required to induce the formation of a single conformer from a wild-type sequence by changing the experimental conditions (pH, temperature, salt type and concentration, folding) or by modifying the sequence.

Nucleic acid structures are characterised by NMR peaks with specific chemical shifts. The latter are determined by different hydrogen bonding interactions between the nucleobases. Remarkably, imino proton peaks from guanines involved in a G-quartet generate a signal in the 10.5 – 12.0 ppm region.<sup>77</sup> These signals are generated by the slower exchange of the imino protons involved in the Hoogsteen hydrogen bonds. Structural determination still requires performing different homonuclear, heteronuclear, and nuclear Overhauser effect experiments. The final assignment can present some ambiguities that can be resolved by isotopic enrichment by <sup>13</sup>C and <sup>15</sup>N labelling of specific sites, or by substituting a nucleobase by a derivative lacking or modifying a related peak. Furthermore, NMR spectroscopy has been successfully used to follow the <sup>13</sup>C-/<sup>15</sup>N-labelled d(TG<sub>4</sub>T)<sub>4</sub> G-quadruplex distribution inside *Xenopus laevis* oocytes and support the hypothesis that G4s do exist in cells.<sup>78</sup> The isotopically labelled G4 inside the oocytes had a lifetime of 20 h. Simultaneous incubation of G4 and PDC showed significant changes in the NMR spectrum of the G4 structure, suggesting that an interaction occurs between the G4 and the ligand.

X-ray crystallography is a method used to characterise the molecular structure trapped inside a crystal.<sup>76</sup> This structural determination is obtained at the atomic level by using an X-ray beam with wavelength similar to the interatomic distance inside the crystal. The crystal lattice planes diffract the beam to form a diffraction pattern characteristic of the atomic structure. The advantage of crystallography is to have information on the cations' localisation and on the solvation. An important difficulty is to obtain the crystal which often requires trying different crystallisation conditions.

## 2.3. G-quadruplex genomic localisation

### 2.3.1. *In silico* approach

Guanine-rich sequences can form intramolecular G-quadruplexes when single-stranded. *In silico* approaches were developed to determine and localise those putative G-quadruplex sequences within the genome. Many *in silico* algorithms have been developed but only a few important methodologies are going to be described.

The simplest approach can identify all the sequences containing the " $G_xN_{1-7}G_xN_{1-7}G_xN_{1-7}G_x$ " motif (X number of guanines present in the track and involved in the generation of the G-quartet and N any nucleotide) using an expression-matching algorithm. This strategy has been followed independently by Huppert *et al.*<sup>79</sup> with *Quadparser* and by Todd *et al.*<sup>80</sup> with *Quadruplexes*, who both identified ~ 375,000 Putative G-quadruplex Sequences (PQS) in the human genome. However, this methodology can only identify the more basic G4 sequences and fails to identify more complex sequences that do not match the basic consensus motif. In addition, the analysis of the PQS from *Quadparser* showed an enrichment at the promotor level and even higher enrichment at nuclease hypersensitive sites (NHS), which indicates that G4s are not randomly distributed inside the genomes but are specifically localised in functional regions and mainly at the promotor region of proto-oncogenes.

Another approach utilised was to evaluate the potential of the predicted G-quadruplex sequences with a score. *G4P calculator* by Eddy *et al.*<sup>81</sup> follow this approach by associating to a sequence a score called G4 potential (G4P). The score criteria take into account long G-tracts ( $\geq 3$ ) generating at least four G-quartets. In this case, there is no impact of the loops on the result. This approach aims to map the G-quadruplex potential of a region and not to identify a specific G4.

Transcript sequences reported in the human Reference Sequence (RefSeq) database were analysed using *G4P calculator*. Data analysis highlighted that there is a strong correlation between gene function and G4P, a high G4P score is associated with proto-oncogenes and transcription factors. This result, similar to the one obtained with Quadparser, suggests an important regulatory role of G4s in the transcription.

*G4RNA screener*<sup>82</sup> and *Quadron*<sup>83</sup> present a new approach using machine learning to recognise G-quadruplex sequences. This methodology has for advantage to avoid defining the G4s as a particular consensus motif and letting the machine learning find new criteria to identify new G4 sequences. It could improve the identification of non-canonical G4 forming sequences and give a more accurate picture of the G4 diversity. However, the new predictive features identified by machine learning can be complex to be determined and used.

### 2.3.2. *In vitro* and *in vivo* methodologies

The utilisation of algorithms had predicted ~ 375,000 putative G-quadruplex sequences,<sup>79,80</sup> which is an impressive number that we can compare with the ~ 20,000 coding genes and the ~ 20,000 non-coding genes (i.e. regulatory elements, telomeres, centrosomes, ncRNAs) evaluated for the human genome.<sup>84,85</sup> This would signify that each gene (coding and non-coding) has roughly nine G4s. It is clear that this number is overestimated, and that the G4s actually forming in the genome are much lower. As a matter of fact, the algorithms cannot take into account the chromatin state, which impacts the accessibility of the DNA molecule, and, as a consequence, its ability to fold into a G4. This is why, difference *in vitro* and *in cellulo* methodologies have been developed to identify the forming G-quadruplex sequences.

One of the first approaches reported is the so called **hf2 G4 pull-down sequencing** methodology reported by Lam *et al.* (2013) developed by means of an anti-G4 hf2 antibody.<sup>86</sup> The principle of the method, summarised in Figure 19, is to incubate magnetic beads coated with hf2 antibody in purified genomic extracts to form G4:hf2 complexes. This allows G4 enrichment and their identification by NGS.

### Hf2 G4 pull-down sequencing

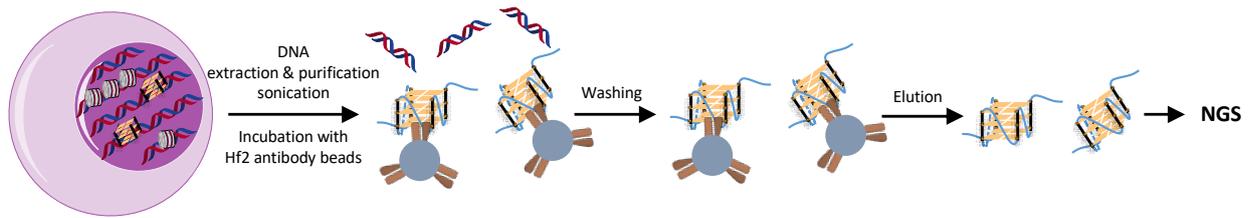


Figure 19: Representation of *hf2* G4 pull-down sequencing methodology by Lam *et al.*<sup>86</sup>

This methodology has been applied to human breast adenocarcinoma (MCF7) genomic DNA extracts and allowed the identification of over 700 G4s. However, only 74% of the sequences were already identified by *G4P calculator*, which means that 26% of them did not match the basic sequence motif. This approach allowed to isolate only 700 G4 sequences, which is a small amount as compared to the one identified with other methodologies. Actually, this limitation comes from the high selectivity of hf2 antibody for specific G4 structures.<sup>87</sup>

**G4-seq** is a methodology developed by Chambers *et al.* (2015) that takes advantage of the ability of G-quadruplexes to stall DNA polymerases.<sup>88</sup> In this experiment, each genomic fragment is sequenced in a negative ( $\text{Li}^+$ ) and in a positive ( $\text{K}^+$ ) G-quadruplex forming condition as represented in Figure 20. The mismatch results between the two conditions indicate the G4 formations.

### G4-seq

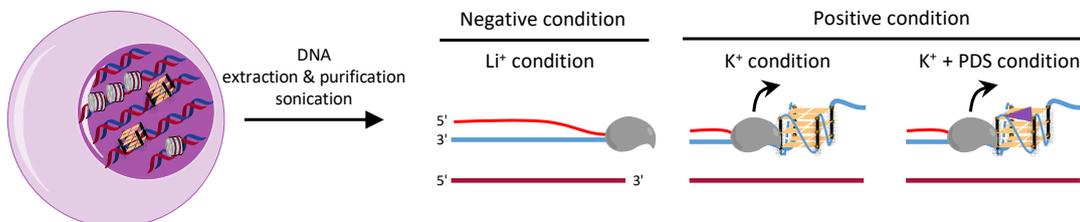


Figure 20: Representation of *G4-seq* methodology by Chambers *et al.*<sup>88</sup>

This methodology has been applied in the human genome of primary human B lymphocytes and the number of observed G-quadruplexes (OQs) was of 525,890 OQs for  $\text{K}^+$  and 716,300 OQs for  $\text{K}^+ + \text{PDS}$  conditions. In this case, the number of OQs exceeded the 375,000 PQS identified by the *in silico* algorithms, with only 60% for  $\text{K}^+$  and 73% for  $\text{K}^+ + \text{PDS}$  conditions matching the PQS. This result can be explained by the important number of OQs that do not match the canonical G4 structures, containing a long loop (24%) or a bulge (30%).

The OQs distribution inside the genome is similar to the *in silico* approaches with an enrichment at gene promoters, 5'UTRs and exons. Similar results were obtained when this **G4-seq** methodology was applied to twelve different species (mammalian, plants, parasite, yeast, and bacteria) to represent widely the living.<sup>89</sup>

A methodology similar to **hf2 G4 pull-down sequencing** was developed using the anti-G4 antibody BG4,<sup>90</sup> which has a high affinity for all G4 topologies, and by fixing the chromatin before isolation to preserve the chromatin state. The **G4 chromatin immunoprecipitation-sequencing (G4 ChIP-seq)** methodology represented in Figure 21 was developed by Hänsel-Hertsch *et al.* (2016) aiming to identify the G-quadruplex forming sequences in chromatin.<sup>91,92</sup>

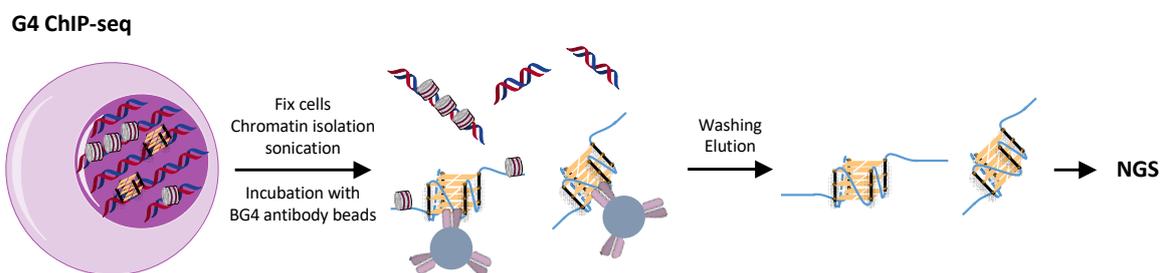


Figure 21: Representation of **G4 ChIP-seq** methodology developed by Hänsel-Hertsch *et al.*<sup>92</sup>

This methodology allowed the identification of over 10,500 G4 sequences inside the human precancerous skin (HaCaT) genome. In addition, this methodology was applied to the genome of different cell lines showing a poor share of G-quadruplex forming sequences (16%), highlighting the importance of the chromatin context in the formation of G-quadruplexes.

Zheng *et al.* (2020) have adapted the G4 ChIP-seq methodology by replacing BG4 with a protein that recognises and cages G-quadruplexes called G4 probe (G4P).<sup>93</sup> **G4P ChIP-seq** has for advantage of knock-in G4P to produce it *in vivo* and making G4P interact with the G-quadruplexes before the fixation as summarised in Figure 22. This methodology was utilised on lung carcinoma epithelial cells (A549) and allowed to identify up to 123,000 G4:G4P interactions of which 43,000 were present in high fold ( $\geq 5$  signal/input) and enriched mainly upstream of the transcription starting site (TSS).

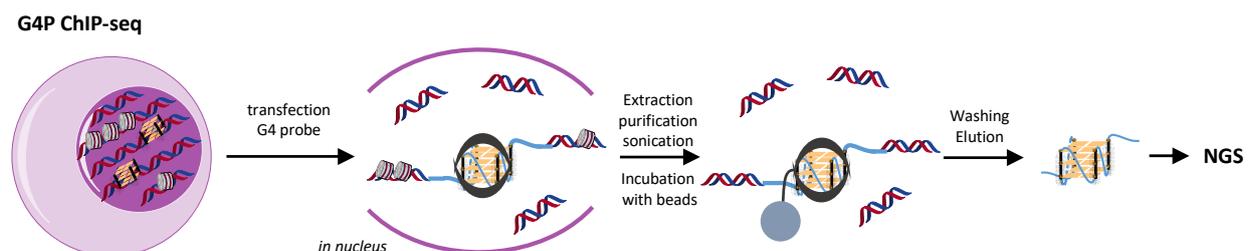


Figure 22: Representation of **G4P ChIP-seq** methodology developed by Zheng *et al.*<sup>93</sup>

Lyu *et al.* (2022) recently reported a new approach adapted from the **cleavage under targets and tagmentation** (CUT&Tag) protocol.<sup>94</sup> The benefit of CUT&Tag is the incubation of the antibody directly in permeabilised cells, avoiding genome fixation and extraction to preserve the chromatin context even more. Another advantage is the tagmentase that cuts specifically at the proximity of the antibody interaction site and ligate adapters to the fragment ends, preparing for NGS analysis. **G4-CUT&Tag** combines the tagmentase with the anti-G4 BG4 antibody to map the native genomic G-quadruplex sites as displayed in Figure 23.

#### G4 CUT&Tag

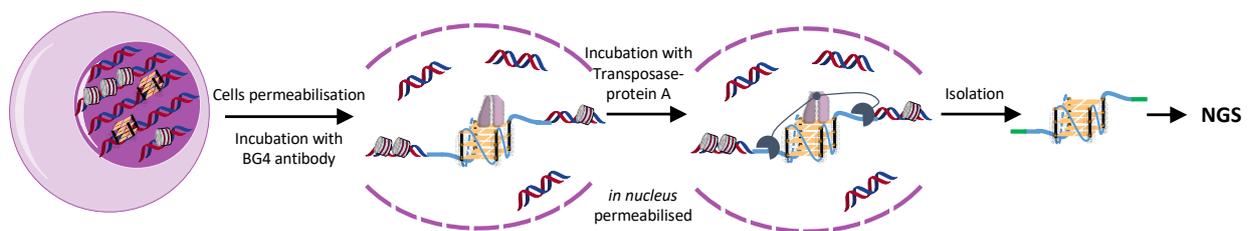


Figure 23: Representation of **G4 CUT&tag** methodology reported by Lyu *et al.*<sup>94</sup>

**G4-CUT&Tag** identified over 10,000 G4s, a number comparable to the one obtained by G4 ChIP-seq. However, only a few (30-40%) of these G-quadruplexes are shared between G4 ChIP-seq and G4P ChIP-seq, demonstrating the impact of the approach on the identification of forming G4 sequences.

Together with methods based on antibody or G4 protein recognition, a different approach based on small synthetic molecules was developed. Muller *et al.* (2010) proposed to use a small molecule that has a high affinity for G-quadruplexes to realise the pull-down.<sup>95</sup> For that, they functionalised a PDS with biotin and realised the first *in vitro* pull-down on fibrosarcoma genome (HT1080), isolating telomere G-quadruplex sequences. It is only later and with another biotin - PDS derivative (PDB ligand) that Zeng *et al.* (2019) applied this concept displayed in Figure 24 to purified epithelial-like cells (HEK 293T) genome extract, identifying ~ 51,000 G4 forming sequences.<sup>96</sup>

#### G4 PDB-Profilig

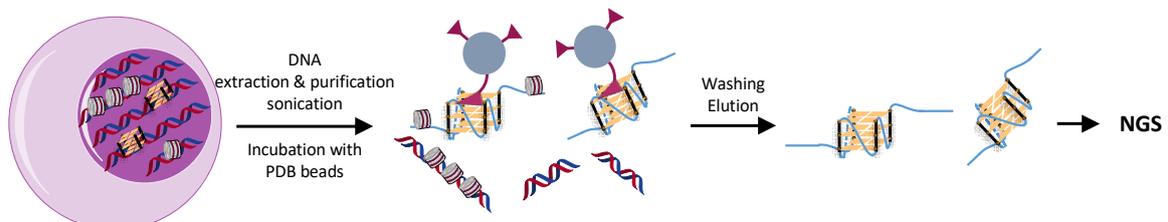


Figure 24: Representation of **G4 PDB-Profilig** methodology developed by Zeng *et al.*<sup>96</sup>

These results have demonstrated the importance of the chromatin context in the formation of the G-quadruplexes which is unique for each cell lines. Another important element is that the different methodologies share only a few G4 structures.

## 2.4. G-quadruplex biological roles

G-quadruplexes are transient structures that form when DNA gets unwind forming single-stranded domain. The first G4 identified were the telomeric G-quadruplexes that have been intensively studied for their impact on the protection and regulation of the telomeres.<sup>97</sup> In addition, many G4s have been identified at promoters and *in silico* approaches confirmed an enrichment in such regulatory elements.<sup>48</sup> Therefore, G-quadruplexes play important biological roles during important DNA transactions in which DNA is present in its single-stranded form, such as during replication, transcription, and telomere protection (Figure 25). Additionally, if the transcribed RNA is G-rich, the G4 formed play a role during translation and RNA processing.<sup>48</sup>

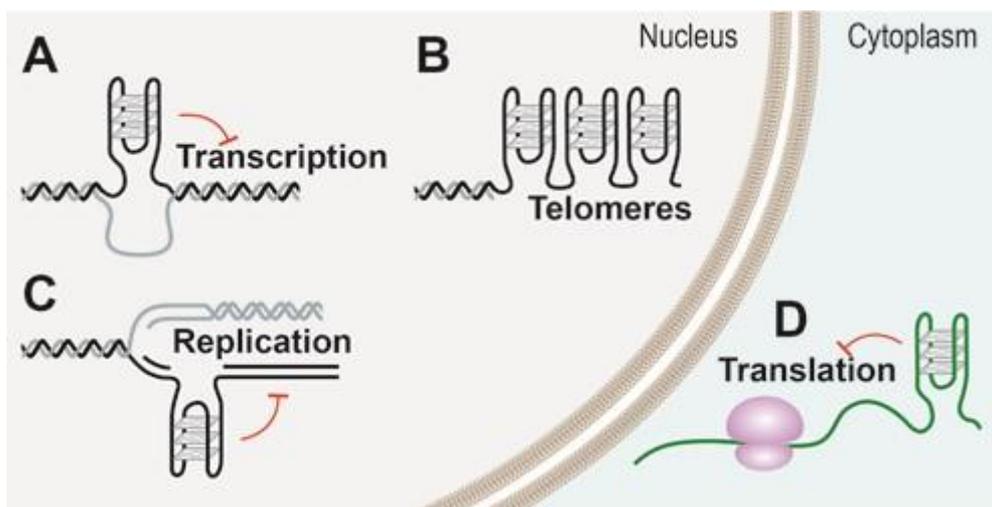


Figure 25: G-quadruplex forms when G-rich region becomes transiently single stranded, as during transcription (A), replication (B), as well in the single stranded telomeric overhangs. In addition, G-quadruplex form in G-rich RNA sequences as well. (Figure from Rhodes and Lipps, 2015)<sup>48</sup>

### 2.4.1. Telomeres

Telomeres were the first identified sequences able to form G-quadruplexes. Telomeres are located at the chromosome ends and are responsible for their protection against deleterious processes, such as homologous fusion of the extremities or degradation by endogenous nucleases. The telomeric sequence comprises a 3'G-rich sequence. Eukaryotic telomeric sequences are highly conserved with only minor variability, for example, G<sub>3</sub>T<sub>2</sub>A for vertebrates, G<sub>4</sub>T<sub>2-4</sub> for ciliates, G<sub>2</sub>T<sub>2</sub>A for insects, G<sub>2</sub>CT<sub>2</sub>A for nematodes, or G<sub>3</sub>T<sub>3</sub>A for plants.<sup>98</sup> The telomeres comprise long proximal double strand repeat ranging from 10 to 15 kb, and a 3'-end distal single-stranded overhang repeat ranging from 100 to 200.<sup>99-101</sup> Telomeres are structured by six proteins to form a nucleoprotein complex named shelterin, responsible for the protection of the telomeres from inappropriate DNA damage response mechanisms. The telomere shapes itself as a long telomeric loop (T-loop), where the 3'G-overhang invades the double-stranded DNA and pairs with the 5' strand forming the displacement loop (D-loop), represented in Figure 26.<sup>102</sup>

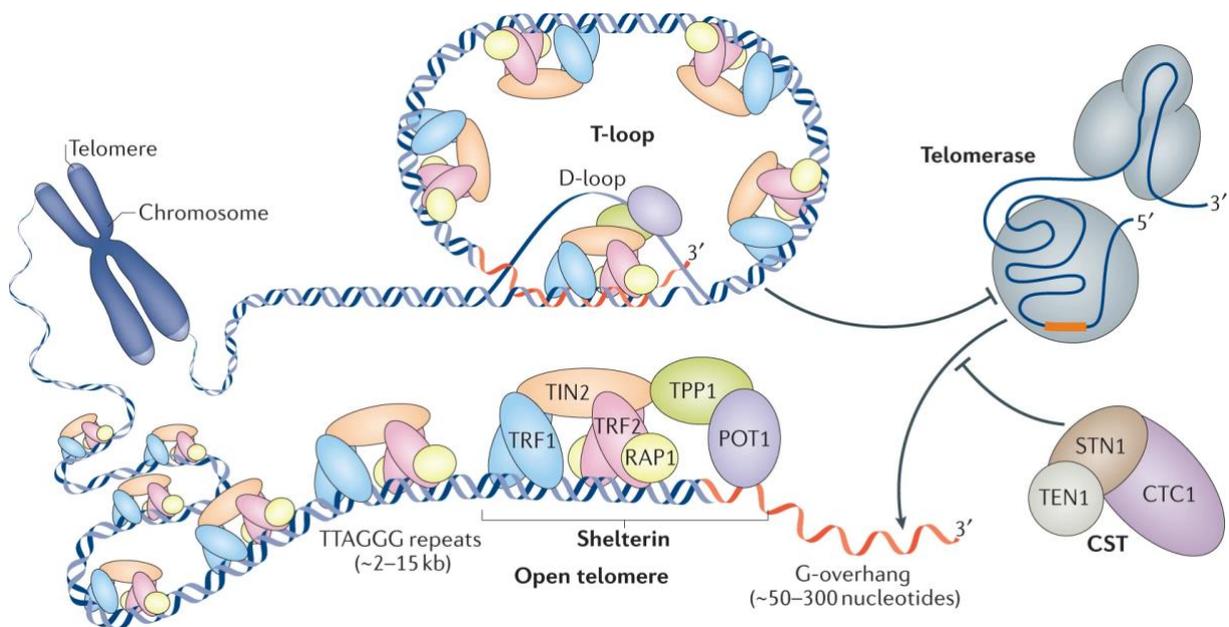


Figure 26: Schematic representation of the telomere organisation composed of a shelterin repeat unit. The 3' G-overhang invades the double-stranded 3' region to form the T-loop and this displacement in single-strand forms the D-loop. (Figure from Lim et Cech, 2021)<sup>102</sup>

The telomeres become shorter at each replication but are essential to protect the integrity of the chromosome. It is why, a cell reaching the Hayflick telomere length limit enters into senescence.<sup>103</sup> The telomere elongation is controlled by the reverse transcriptase telomerase which is a ribonucleoprotein complex composed of a reverse transcriptase (hTERT) and a RNA template (hTR RNA).<sup>104</sup> Normally the telomerase activity is diminished or suppressed in somatic cells, and is active mainly in fast-dividing cells. However, the telomerase is reactivated in most cancer cells during the process of immortalisation.<sup>105</sup> The human telomeric sequence is known to form hybrid G4 structures in K<sup>+</sup>-rich buffer which are not substrates for the telomerase.<sup>97</sup> It is believed that stabilising these G-quadruplexes by using small molecules could prevent the telomerase from binding telomeric DNA, thus, preventing the elongation of the telomeres. This is why, inhibition of the telomerase via G-quadruplex stabilisation was considered a powerful approach to treat cancer.

#### 2.4.2. G-quadruplex as regulatory elements

Promoters are DNA regions located upstream of a gene that recruits RNA polymerase to start the transcription. *In silico* and *in vitro* analyses have demonstrated G4 enrichment at promoters and mainly at promoters of proto-oncogenes, highlighting the possibility of their involvement in the development of tumorigenesis. Several examples of promoters of oncogenes containing G4s have been reported in previous studies: c-myc,<sup>106</sup> c-kit,<sup>107</sup> RAS,<sup>108</sup> and hTERT<sup>109</sup> contributing to the proliferation, VEGF<sup>110</sup> affecting angiogenesis, and Bcl-2<sup>111</sup> preventing apoptosis. This section focuses on the G-quadruplex formed in the c-MYC promoter region.

The c-MYC proto-oncogene encodes a transcription factor critical for regulating a broad range of cellular processes, such as cell growth, proliferation, differentiation, transformation, angiogenesis, and apoptosis.<sup>112</sup> Thus, dysregulation and aberrant overexpression are often associated with malignant tumours. This is why c-MYC proto-oncogene regulation is critical for the development of new anticancer strategies.

c-MYC transcription is regulated by a complex system comprising a variety of promoters (P0, P1, P2, and P3), transcriptional start sites, and several nuclease hypersensitive elements (NHEs) which single-stranded forms either an unwind or non-canonical structure.<sup>113</sup> The NHE III<sub>1</sub> region is a 27-bp sequence upstream of the P1 promoter that controls up to 90% of c-MYC transcription level.<sup>114</sup> Its template strand is a G-rich sequence in equilibrium between a transcriptionally active form (single-strand) and a silencer form (G-quadruplex).<sup>114</sup>

The Purine-rich 27-bp (Pu27) sequence can fold into a G-quadruplex and the folding is promoted by the negative supercoiling generated during transcription. This c-MYC Pu27 can form different polymorphic G4 structures and different G4 models have been used to study its ability to fold into a G4 structure and its biological roles.<sup>115</sup>

One of these model sequences is **Pu24** a 24-nt sequence characterised by NMR. The latter was obtained after substituting a G with a T at positions 4 and 13, and resolved with the help of the analogue sequence Pu24I, where an inosine substituted the T at position 10.<sup>116</sup> It is a parallel three-layer G-quadruplex represented in Figure 27, formed by four G-tracks and a foldback guanine. The structure presents four loops: two one-thymine propeller loops, a diagonal loop, and a snap-back loop. In addition, two triads (G10·G11·A12 and G20·A22·G23) stabilise the structure capping both extremities.

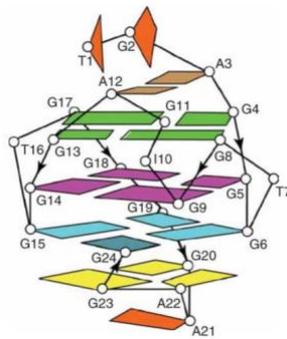


Figure 27: Schematic representation of the Pu24 c-MYC G-quadruplex (TGAGGGTGGIGAGGGTGGGGAAGG). (Figure from Phan et al., 2005)<sup>116</sup>

**cMYC22** is a 22-nt sequence obtained after substituting a G with a T at positions 4, 14, and 23. It forms a canonical parallel G-quadruplex with the 5'- and 3'-flanking regions capping the external G-quartets, represented in Figure 28.<sup>117</sup> The 3'-TAA flanking sequence interacts with G9 and G22. The 5'-TGA flanking region interacts with the external G-quartet only by stacking of the A6

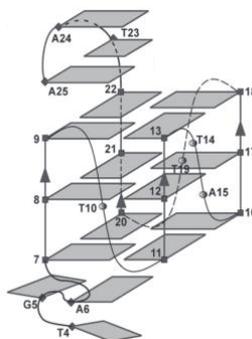


Figure 28: Schematic representation of the cMYC22 G-quadruplex (TGAGGGTGGGTAGGGTGGGTAA). (Figure from Ambrus et al., 2005)<sup>117</sup>

The impact of the G-quadruplex on c-MYC transcription was demonstrated by two different experiments. The Pu27 sequence was mutated to impede G4 folding or stabilise by a ligand to stabilise the G4 structure. In the first case, the researchers observed a significant increase in the c-MYC transcription.<sup>106</sup> Differently, in the second case, stabilisation of G4 using TMPyP4 resulted in a decrease of the c-MYC transcription and downregulation of several c-MYC-regulated gene.<sup>118</sup> These results show the importance of the G-quadruplex forming in the NHE III<sub>1</sub> region acting as a repressor element responsible for downregulating c-MYC transcription. In addition, the nucleolin,<sup>119</sup> a multifunctional nucleolar phosphoprotein known to interact with parallel G-quadruplexes,<sup>120</sup> was observed to promote the formation of the c-MYC G-quadruplex resulting in downregulating the transcription.<sup>121</sup> c-MYC G4 acts as a transcriptional repressor element that can be promoted and stabilised by small molecules and nucleolin. This method to downregulate c-MYC made G4 an interesting target to develop new therapeutic treatments for the modulation of c-MYC expression and an approach to target malignant tumours overexpressing c-MYC.

### 3. G-quadruplex targeting

As explained in the previous section, G-quadruplexes are engaged in diverse biological processes, such as telomere protection, telomerase inhibition, transcription and translation regulation, acting as recognition elements for proteins, etc.<sup>48</sup> To study them, small molecules (i.e. G4 ligands) have been designed to modulate their *in vitro* activities. Overall, the overrepresentation of G4s in proto-oncogene promoters has made them a focus in the development of new anticancer molecules. Initially, G-quadruplex ligands were derived from DNA intercalators by attempting to displace their selectivity for G4s. Indeed, handy G4 ligands must be selective versus the other DNA structures, such as the duplex DNA. Intercalating ligands intercalate through  $\pi$ -stacking between the base pairs, and the same principle is applied to the G4 ligand interacting with one preferential external G-quartet through  $\pi$ -stacking and electrostatic interactions. G-quartets form large aromatic surfaces which may provide binding selectivity to the ligands by maximising its  $\pi$ -stacking overlap, as compared to the smaller two-base aromatic surface in duplex DNA. In addition, positively charged ligands show stronger binding through interaction with the negatively charged sugar-phosphate backbone as observed for nucleic acids targeting in general. The design of new ligands combining rational design and empirical approaches led to synthesising a considerable number of ligands, which requires careful evaluation through interaction studies to identify the most performants.

#### 3.1. G-quadruplex Ligands

In this section, only a handful selection of G4 ligands is described. Indeed, G4LDB 2.2 is a G-quadruplex ligand database listing over 3,000 allegedly G-quadruplex ligands developed during the past two decades.<sup>122</sup> These G4 ligands were developed to modulate and stabilise G4s *in vitro* and to study their impact on cell responses. Another goal was to develop new anticancer molecules due to the overrepresentation of G-quadruplexes in proto-oncogene promoters. Initially, the strategy was to derivate dsDNA probes into specific G4 ligands by increasing their aromatic surface to stack with one preferential external G-quartet and adding positive charges to interact with the sugar-phosphate backbone by electrostatic interaction.



The interaction of **BRACO-19** with the telomeric bimolecular G-quadruplex has been characterised by X-ray crystallography as shown in Figure 30.<sup>127</sup> The crystal unit contained two parallel G4s stacked interfaced via **BRACO-19**, with one G-quartet being a T·A·T·A tetrad formed by both 5'-end. The  $\pi$ - $\pi$  stacking interaction was reported only for two on the fourth G-quartet guanines as well as only one T·A of the T·A·T·A tetrad. In addition, the water molecule network has been demonstrated to be necessary to allows water-mediated interaction for both the positively charged ring nitrogen aligned with the ion channel and one carbonyl atom of a side-chain with the thymidine base T24.

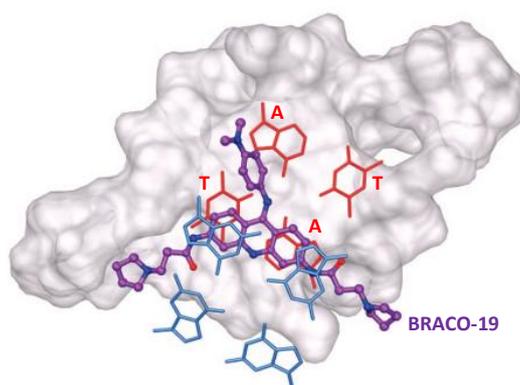


Figure 30: Interaction of **BRACO-19** with the telomeric bimolecular G-quadruplex. **BRACO-19** (purple) interact with the top G-quartet (blue) and the bottom T·A·T·A tetrad (red) (Figure adapted from Campbell et al., 2008).<sup>127</sup>

Another important family of ligands are based on the porphyrin core, containing positive charges by N-methylation and quaternisation of the aza-aromatic moieties, which improve its water solubility. **TMPyP4** (Figure 31) is a tetra cationic porphyrin presenting good stabilisation for G4 ( $\Delta T_{1/2} \approx 17$  °C)<sup>128</sup> and a good affinity value with the human telomeric G-quadruplex ( $k_D \approx 0.68$   $\mu$ M). However, **TMPyP4** is not a selective G4 ligand presenting a similar affinity for hairpin DNA ( $k_D \approx 0.70$   $\mu$ M).<sup>129,130</sup> Its interaction with antiparallel G-quadruplex induces a conversion to the parallel topology.

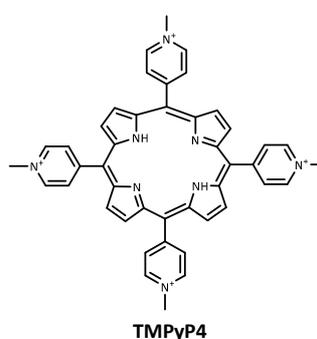
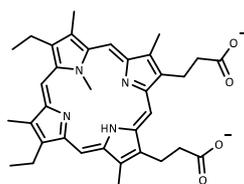


Figure 31: Molecular structure of the G-quadruplex ligand **TMPyP4**.

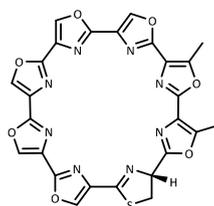
Nonetheless, *in cell* experiments have demonstrated its ability to downregulate the expression of various oncogenes such as c-myc.<sup>118</sup> A strategy to improve **TMPyP4** interaction with G4s was to take advantage of the chelating properties of the porphyrin core, which is positioned above the central channel of the G4 structure when interacting by  $\pi$ -stacking with the external G-quartet. Thus, different metals - such as Cu(II), Ni(II), Mn(III), Pt(II) and Zn(II) - have been coordinated to **TMPyP4** and all the resulting derivatives present binding constants laying in the micromolar range ( $10^6$ - $10^7$  M<sup>-1</sup>).<sup>131</sup>



**NMM**

Figure 32: Molecular structure of **NMM** ligand.

**N-Methyl Mesoporphyrin IX (NMM)** (Figure 32) is another porphyrin derivative reported to be able to stabilise G4 structures. **NMM** is synthesised by N-methylation of the natural mesoporphyrin IX (MPIX) and the presence of the two propionate groups makes the molecule soluble in water. **NMM** shows a medium to low affinity for G-quadruplexes and preferentially binds to the parallel topology.<sup>132</sup>



**Telomestatin**

Figure 33: Molecular structure of **telomestatin**.

Natural products have always been an excellent source for finding new ligands characterised by a large structural variety. A standard method is to screen the natural products against a target to identify a potential candidate core. Kazuo *et al.* have followed this strategy by screening natural products for telomerase inhibition.<sup>133</sup> This screening identified a neutral macrocyclic molecule with a good potential as telomerase inhibitor. It was named **Telomestatin** (Figure 33) and was isolated from *Streptomyces anulatus*. This compound presents a large planar aromatic surface with a size comparable to the one of a G-quartet, which induces a high G4 stabilisation ( $\Delta T_{1/2} \approx 24$  °C) and exhibits a 70-fold selectivity versus duplex DNA.<sup>128</sup>

Following the initial development of G-quadruplex ligands with rigid aromatic cores,<sup>128</sup> new scaffolds characterised by some flexibility were explored with the purpose to further improve the  $\pi$ -stacking with G4s. This has been achieved by breaking up the large aromatic surface into smaller aromatic units linked by amine or amide bridges. This is the case of ligand **12459** (Figure 34-Left) composed of two quinolinium units anchored by amine linkers to a triazine core.<sup>128</sup> **12459** shows weak G4 stabilisation ( $\Delta T_{1/2} \approx 8$  °C).

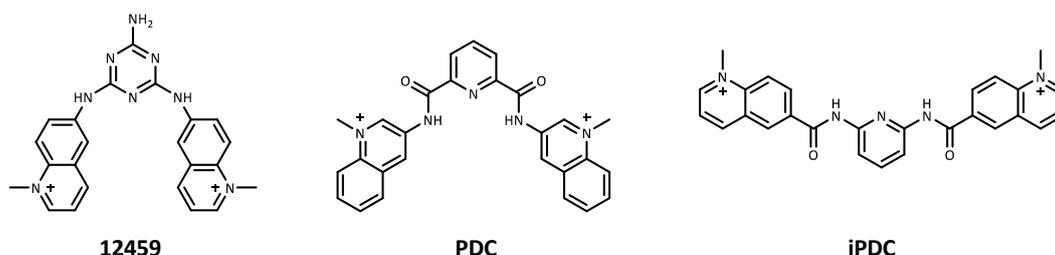


Figure 34: Molecular structures of **12459**, **PDC** and **iPDC**.

The triazine core was subsequently replaced by a pyridine core linked to two quinoliniums by amide bonds giving the PyridoDiCarboxamide family (**PDC** family). This structure features a classic planar electron-rich aromatic surface with extra flexibility from the units. The main representative of this series is **360A** (also called **PDC**) represented in Figure 34, showing high stabilisation of G4 ( $\Delta T_{1/2} \approx 21$  °C) combined with a strong selectivity versus duplex DNA ( $> 150$ -fold).<sup>128</sup> An important feature of the **PDC** is its ability to adopt two geometries: a V shape and a linear shape. The V shape forms when both amides interact by hydrogen bonding with the pyridine core. It has been demonstrated by two independent experiments that the V shape is crucial for its interaction with G-quadruplexes. **iPDC** (Figure 34-right) is a **PDC** structural isomer where the amide bridges are reversed. This prevents the formation of the hydrogen bonds with the pyridine core, thus, **iPDC** adopts a linear shape, which has a much lower affinity for G4s as compared to **PDC**.<sup>134</sup> The second demonstration was to induce the linear **PDC** form by adding a metal which is chelated by the tridentate center. **PDC** binding to G4 was inhibited and G4 unfolded when treated with Cu(II), process which was reported to be reversible if Cu(II) is removed from the solution.<sup>135</sup>

PDC was modified and functionalised to adjoin new properties. For example, PDC was radiolabelled with tritium by Granotier *et al.* to obtain  $^3\text{H}$ -360A (Figure 35) for the detection of the ligand in T98G cells and in metaphase chromosome spreads. It showed preferential binding at the chromosome ends i.e. telomere, and interstitial regions.<sup>136</sup>

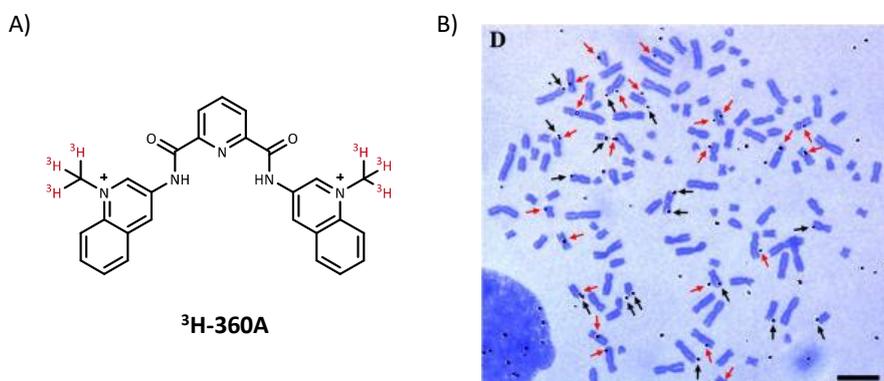


Figure 35: A) Molecular representation of  $^3\text{H}$ -360A with in red tritium, and B) autoradiography results of T98G metaphase chromosomes spreads incubated 48 h with  $^3\text{H}$ -360A. Black arrows represent silver grains on the terminal regions and red arrows silver grains on the interstitial regions. (Figure from Granotier *et al.*, 2005)<sup>136</sup>

More versatile PDCs were developed to introduce readily moieties that adds new functions, for example, **clickable PDCs** (Figure 36-A) functionalised with a side chain bearing a terminal alkyne or azide for post-binding functionalisation by Copper(I)-catalysed alkyne-azide cycloaddition (CuAAC). These **clickable PDCs** were employed in the G-quadruplex Guided Immunofluorescence Staining (G4-GIS) methodology (Figure 36-B) developed in the lab to follow G4 ligand distribution in cells.<sup>137</sup> A 5-BrdU immuno-tag was introduced *in situ* on the **clickable PDCs**, and afterwards recognised by a commercially available antibody which allows performing classical immunofluorescence staining protocol. This new imaging method combines the specific recognition of the ligands for G4s with the fluorescence amplification signal guaranteed by immunofluorescence detection. These ligands mainly localised in the cytoplasm suggesting a selective interaction with G4 RNA.

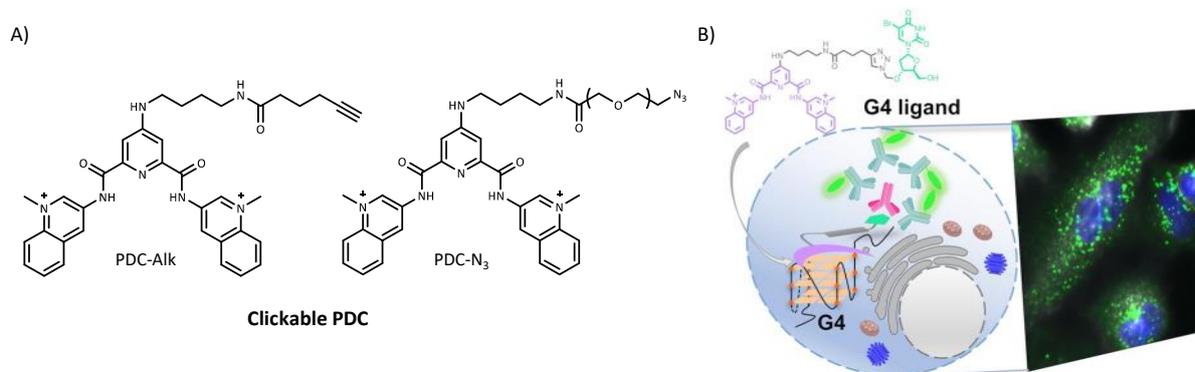


Figure 36: A) Molecular structure of the clickable PDCs and B) the G4-GIS methodology (left) utilising the clickable PDCs to post-functionalise the PDC (in purple) with a 5-BrdU immuno-tag (in green) to realise immunofluorescence staining in fix cells (right). The fluorescence is mainly in the cytoplasm indicating interaction with RNA G4 and with some foci in the nucleus (nucleus in blue) (Figure from Masson *et al.*, 2021).<sup>137</sup>

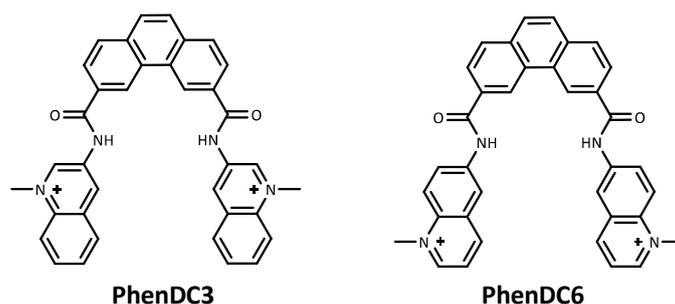


Figure 37: Molecular structure of PhenDC3 and PhenDC6.

Phenanthroline derivatives (**PhenDC series**) are analogues of **PDC** featuring an ortho phenanthroline core in place of the pyridine core.<sup>134</sup> The bended structure of the phenanthroline motif induces a U-shaped conformation as illustrated by **PhenDC3** and **PhenDC6** structures in Figure 37, which results in a perfect match for the G-quartet geometry. As a result, **PhenDC** compounds show a very high affinity and selectivity for G4s, which ranks them amongst the most potent G4 ligands developed so far. Between **PhenDC6** and **PhenDC3**, the second shows the highest affinity and the ability to induce G4 folding. These properties have made **PhenDC3** a reference G4 ligand and has been exploited to study G4s both *in vitro* and *in cells*.

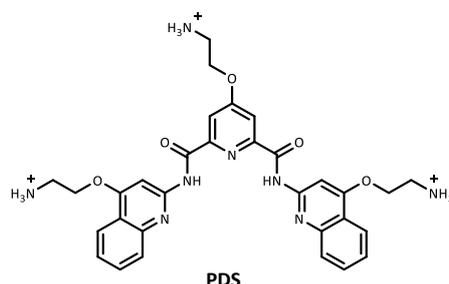


Figure 38: Molecular structure of the PDS.

Pyridostatin (**PDS**) in Figure 38 is a **PDC** analogous developed by Rodriguez *et al.*,<sup>138</sup> differing by the cationic charge localisation and the branching connectivity on the quinoline units. **PDS** is known for its high affinity toward G-quadruplexes and high selectivity versus duplex DNA. As for **PDC** and **PhenDC3**, many studies used **PDS** to stabilise G4 *in cells*.

Another strategy developed was to be inspired by the G-quartet structure and to synthesise G-quartet biomimetics. Nikan and Sherman have developed a lipophilic template-assembled synthetic G-quadruplex (**TASQ**) composed of four guanines grafted to a lipophilic calixarene, which structured itself as a singular mimic G-quartet.<sup>139</sup> However, the lipophilic nature of the compound is incompatible with physiological conditions. This strategy was taken over by Monchaud who developed a series of water-soluble TASQs to resolve this issue. The first reported compound was **DOTASQ** (Figure 39) composed by four guanines grafted on a tetraaza macrocycle (DOTA). However, **DOTASQ** is not able to structure itself as a singular mimic G-quartet in the absence of cation and presents low G4 stabilisation properties with Terbium ( $\Delta T_{1/2} \approx 10\text{ }^{\circ}\text{C}$ ).<sup>140</sup> A new scaffold was designed by switching the DOTA macrocycle to a porphyrin to improve the formation of the G-quartet structure without cation. It corresponds to the **PorphySQ** able to structure itself into a singular mimic G-quartet which showed weak G4 stabilisation properties ( $\Delta T_{1/2} \approx 6.5\text{ }^{\circ}\text{C}$ ).<sup>141</sup> Furthermore, the guanine neutral side chain was replaced by a protonable side chain (PNA) to add electrostatic interaction with the negatively charged DNA backbone and try to improve the interaction with G4s. Improvement was observed for **PNA-DOTASQ** as it does not require terbium to stabilise G4s ( $\Delta T_{1/2} \approx 12\text{ }^{\circ}\text{C}$ ).<sup>142</sup> However, **PNA-PorphySQ** show no significant G4 stabilisation improvement ( $\Delta T_{1/2} \approx 7\text{ }^{\circ}\text{C}$ ).<sup>143</sup> The last scaffold developed was **CyTASQ** using a cyclene macrocycle which conserves the weak G4 stabilisation ( $\Delta T_{1/2} \approx 5\text{ }^{\circ}\text{C}$ ), but an enhanced selectivity versus duplex DNA.<sup>144</sup> These scaffolds have been functionalised with different moieties to subjoin some new properties. One of them was to introduce a biotin to fish the G-quadruplex structures such as **BioTASQ** and **BioCyTASQ**. The biotine in the **BioTASQ** interacts with one of the guanine arms preventing the structuration into a singular mimic G-quartet and abolishing the interaction with the G4. **BioCyTASQ** succeeded to conserve its affinity and selectivity toward G4s, and allowed to identify 57,500 G-quadruplex forming sequence in a new methodology called G4-DNA precipitation and sequencing (G4DP-seq).<sup>145</sup> In addition, more versatile TASQs have been developed by adding an alkyne (**MultiTASQ**) or an azide (**azidoMultiTASQ**) for post-functionalisation by click chemistry.<sup>146</sup>

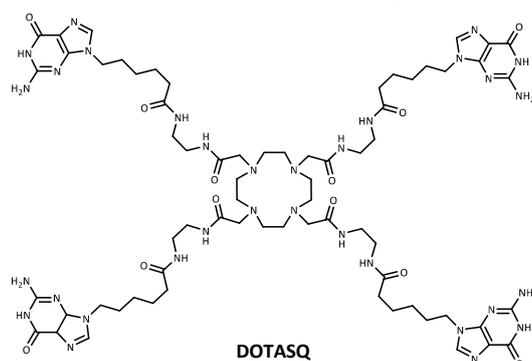


Figure 39: Molecular structure of **DOTASQ**.

G-quadruplexes have been found to be associated to numerous cancer-related functions, such as control of the expression of several oncogenes and tumour suppressors. This is why many G4 ligands were developed focused on antitumoral activities. However, after two decades of research, only a few G4 ligands have reached the clinical trial steps and none is yet available on the market. The main problem is how to target a specific G4 sequence such as c-myc or VEGF by using small molecules, with a  $\pi$ -stacking and electrostatic interactions that are insufficient to drive a sequence specificity. In addition, it is likely that *in vivo* G-quadruplexes are much more structurally complex and polymorphic than those studied *in vitro* which often are mutated and/or truncated to limit the polymorphism. A more promising G4 ligand activity is to induce DNA damages leading to cell death when combined with inhibitors of DNA repair pathway. The quinolone derivatives (**CX-3543** and **CX-5461**) is a family of compounds that have been explored as drugs in clinical trial.<sup>147</sup>

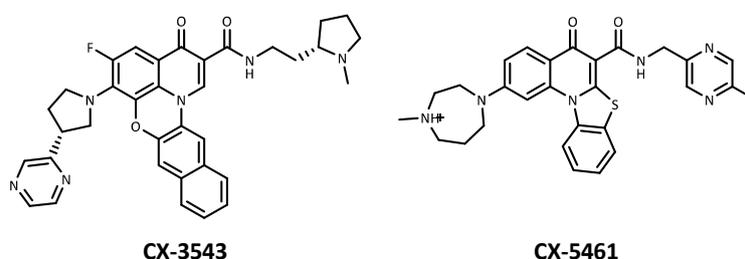


Figure 40: Molecular structure of **CX-3543** and **CX-5461**.

**CX-3543 (Quarfloxin)** (Figure 40-Left) was the first to enter clinical phase 1 in 2005 but the assays have been stopped due to a lack of efficacy. The reported effect was a disruption of the nucleolin interaction with rDNA G-quadruplexes inducing apoptotic death in cancer cells.<sup>148</sup> Currently, efforts are focused on **CX-5461 (Pidnarulex)** (Figure 40-Right) a known rDNA transcription inhibitor with anti-tumorigenic activity developed originally by Cylene Pharmaceuticals as an anticancer drug targeting RNA polymerase I. Recently, **CX-5461** was identified to stabilise G4 ( $\Delta T_{1/2} \approx 20$  °C) and weakly binding duplex DNA.<sup>149</sup> **CX-5461** is actually in phase I clinical trials as a treatment of BRCA1/2-deficient tumours. These tumours particularity is their homologous recombination repair pathway deficiency, which prevents double-strand break repair. **CX-5461** G4 stabilisation should cause DNA breaks leading to cellular death. Hence, these drugs, like others, seem more promising used in combination with inhibitors of repair pathways or other DNA-damaging agents following the principle of synthetic lethality.

### 3.2. Anti-G-quadruplex antibodies

In addition to the small molecules strategy to target G4 structures, another strategy was to develop anti-G4 antibodies taking advantage of the known specificity of antibodies to selectively bind G-quadruplexes.

The two first generated anti-G4 antibodies were reported by Schaffitzel *et al.* in 2001 by screening the synthetic Human Combinatorial Antibody Library (HuCAL) against the *Oxytricha* telomeric G-quadruplex sequence (T<sub>4</sub>G<sub>4</sub> repeat).<sup>150</sup> From the selection, two single-chain antibody fragments (scFvs) **Sty3** and **Sty49** were identified. The latter demonstrated nanomolar affinities for both parallel and antiparallel *Oxytricha* telomeric G4 structures, whereas **Sty3** presents a strong specificity for parallel structures. Globally, both antibodies did not exhibit any affinity toward single-, duplex-, triplex- and hairpin DNA. **Sty49** was used for immunofluorescence staining experiments to visualise the telomeric G-quadruplexes in the macronuclei of *Stylonychia lemnae*. The absence of staining observed when using **Sty3** suggested that *Stylonychia* telomeric G4s fold into the antiparallel structure.

Later in 2008, Fernando *et al.* published an engineered anti-G4 scFv **hf2** generated by phage display against the c-kit2 G-quadruplex.<sup>87</sup> **hf2** antibody has a strong selectivity for c-kit2 with nanomolar affinity (~ 1.6 nM) compared to submicromolar for c-kit1 (~ 0.5 μM) and micromolar for duplex DNA (~ 4.6 μM). The high specificity of **hf2** for particular G4 structures limited its efficiency for pull-down and cell imaging experiments. The same team published in 2013 the scFv **BG4** antibody exhibiting nanomolar affinity toward various DNA and RNA G-quadruplexes.<sup>90</sup> The advantage of **BG4** as compared to **hf2** is the absence of preference for a specific G4 topology or molecularity, displaying strong selectivity versus duplex DNA and i-motif. **BG4** has been utilised in many experiments as it is considered as the reference to target specifically G-quadruplex using antibodies due to its high G4 selectivity, and absence of other versatile and non-topology specific anti-G4 antibodies. For example, **BG4** has been used to visualise DNA G4s within the different cell cycle phases and in various human cell lines.

Other anti-G4 antibodies developed are **1H6** and scFv **D1**. The first is a monoclonal antibody anti-G4 generated by Henderson *et al.* by immunising hybridoma cells against the *Oxytricha* telomeric bimolecular G4 Oxy-2 (T<sub>4</sub>G<sub>4</sub>).<sup>151</sup> **1H6** showed subnanomolar affinity toward Oxy-2 (~ 0.3 nM) and other G4 structures, with no apparent affinities for other DNA structures. **D1** is a parallel-specific anti-G4 generated from the human *BCL2* G4 and has been used for the immunoprecipitation of the human parallel G-quadruplexes.<sup>152</sup>

Antibodies anti-G4 represent useful tools to perform immunofluorescence experiments allowing visualisation/localisation of G-quadruplexes in cells. However, doubts about their selectivity have been raised by recent experiments indicating that other potential targets could cross-react. For example, a strong affinity for some G-rich non-G4 forming single-strand DNA has been found for **BG4**.<sup>153</sup> **1H6** showed a poor affinity for G-quadruplex sequences lacking thymidines and a high affinity for single-strand poly(T).<sup>154</sup>

Recently, **SG4** a nanobody specific for the G-quadruplexes has been generated by Galli *et al.*<sup>155</sup> Nanobodies are smaller, simpler to modify, and can be expressed directly inside cells and fold properly. **SG4** shows nanomolar affinity for G4s, and was utilised in G4 ChIP-seq isolating 5,531 and 10,621 G-quadruplex forming sequences from K562 and U2OS cell lines, which is consistent with the result obtained using BG4. The nanobody was successfully expressed in HEK293T and utilised for live cell imaging.

### 3.3. Interaction studies

The interaction of a ligand with a G-quadruplex must be fully characterised in terms of binding parameters (e.g. affinity and stoichiometry) and in terms of binding selectivity. As well, the determination of binding sites and modes (quartet stacking, loop or groove binding) is of utmost importance to understand the ligand/G4 interaction. Affinity measures the efficiency of a ligand to bind a given G-quadruplex and is quantified by the binding constant (association constant  $K_A$  or dissociation constant  $K_D$ ) when appropriate methods are available or through a relative indicator *via* semi-quantitative methods. Ligand selectivity is the capability of binding G-quadruplexes (or one G4-topology) preferentially versus other secondary nucleic acid structures, such as duplex DNA and eventually hairpins or i-motif. Biophysical methods are the prime choice to study physical interactions and have already been used to characterise dsDNA/ligand interaction. Those methods can be divided into three categories: structural experiments with NMR spectroscopy, X-ray crystallography, and CD spectroscopy. Constant binding experiments including, for example, surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC). Indirect affinity experiments including FRET-, CD- and UV-melting, fluorescence intercalator displacement, and others. In parallel to these classical methods, high-throughput approaches were developed to screen rapidly and efficiently large collections of new ligands, such as FRET-melting and G4-FID.

It is essential to evaluate the G-quadruplex/ligand interaction using several methods since each has advantages and drawbacks. However, it is essential to remember that those experiments are realised *in vitro* and cannot take into account the *in cellulo* complexity, from the uptake and localisation of the ligand, to the chromatin environment filled with competing duplex DNAs, mRNAs and proteins. In this section, the focus will be given to some general and high-throughput methods.

### 3.3.1. FRET-melting

The FRET-melting assay is a thermal denaturation method used to measure the melting temperature of nucleic acid structures quickly and reproducibly. It is one of the first to have been used for the interaction study of G-quadruplex ligands, and was optimised and standardised by De Cian *et al.*<sup>156</sup> This method measures the melting temperature of G-quadruplexes through a FRET (Förster Resonance Energy Transfer) effect and study the interaction with G4 ligands. This experiment requires the functionalisation of both G4 sequence extremities by a donor and an acceptor fluorophore pair, for which the emission spectrum of the donor overlaps with the acceptor absorption spectrum to allow the transfer of excitation energy. The standard FRET pair represented in Figure 41-A is the fluorescein (FAM) in 5' position and the tetramethylrhodamine (TAMRA) in 3' position. The Förster Resonance Energy Transfer (FRET) phenomenon is the non-radiative dipole-dipole transfer of the excitation energy of a donor to an acceptor. Thus, the emission of fluorescence is realised by the acceptor instead of the donor. FRET efficiency depends on the distance between the donor and the acceptor. The donor and the acceptor are close enough to enable the FRET phenomenon when the G-quadruplex is folded, and in this condition, the fluorescein fluorescence is quenched. Unfolding the G-quadruplex increases the distance between the donor and acceptor, thereby preventing FRET and leading to the restoration of the fluorescein fluorescence as displayed in Figure 41-B.

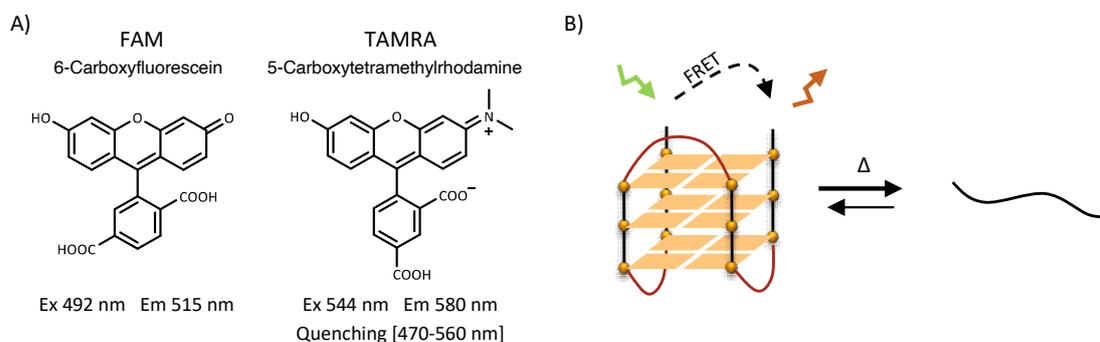


Figure 41: A) Structure and fluorescence parameters of the FAM/TAMRA FRET pair functionalising the 5',3' of the G4 sequence. B) Schematic representation of the FRET-melting experiment.

The experiment consists in increasing the temperature from 25 °C to 95 °C while recording the fluorescein fluorescence emission. The melting temperature ( $T_m/^\circ\text{C}$ ) is the temperature that corresponds to the maximum of the melting curve first derivative. However, sigmoidal FRET-melting curves are often poorly resolved presenting multiple local first derivative maxima, which makes it impossible to determine the correct melting temperature. Therefore,  $T_m$  is approximated using the half maximum fluorescence temperature  $T_{1/2}$ . It corresponds to the temperature at half the maximum fluorescence (i.e. the temperature at a fluorescence of 0.5 when the fluorescein fluorescence is normalised between 0 and 1). A G-quadruplex can be stabilised by a ligand which increases its melting temperature. The difference between the melting temperature of the G4 with and without ligand ( $\Delta T_{1/2}$ ) is the parameter representing the induced stabilisation by the ligand (Figure 42). The  $\Delta T_{1/2}$  is a semi-quantitative evaluation of ligand affinity, hence, high  $\Delta T_{1/2}$  usually reflects a strong ligand/G-quadruplex interaction.  $\Delta T_{1/2}$  allows affinity ranking of ligands when these are compared rigorously in the same conditions. Indeed,  $\Delta T_{1/2}$  is relative to  $T_m$  that itself is strongly dependent on the experimental conditions ( $\text{K}^+$ -buffer) and on the G4 sequence. Hence, comparison with other ligands should be made uniquely for the same G-quadruplex at the same buffer composition, concentration, and ligand/G-quadruplex ratio.

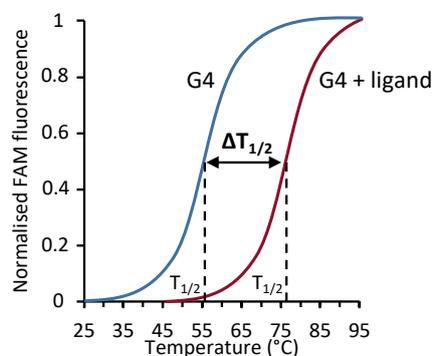


Figure 42: Schematic representation of the FRET-melting curves to measure the  $\Delta T_{1/2}$  values as the difference between the melting temperature of the G-quadruplex (blue) and the G-quadruplex + ligand (red). The melting temperature is estimated at half height of the pseudo sigmoidal curve.

By adding an unlabelled competitor, FRET-melting allows evaluating the selectivity against other nucleic acid structures (e.g. duplex, hairpin, G-quadruplex, i-motif). The standard competitor is ds26, a self-complementary 26-mer duplex strand used as a duplex DNA competitor. A ligand is considered selective when the  $\Delta T_{1/2}$  is not significantly affected by the presence of the competitor.

The experiment is usually carried out in 96-well plates allowing up to 14 ligands to be evaluated for one G-quadruplex in three conditions (without and with 3 and 10  $\mu\text{M}$  competitor) within 2 h to 3 h. Some verifications must be done before running a FRET-melting experiment. When using a new G4 sequence, the 5' and 3' bi-functionalisation can impact the G4 topology, therefore, it should be verified before proceeding with the experiments. FRET efficiency needs to be sufficient or should be improved by reducing the 5' or 3' tail length, e.g. F21T correspond to the 22AG sequence shortened by one nucleotide at the 5'-end. The melting temperature must not be too high to have an optimal range to measure the stabilisation. For this purpose, potassium concentration is modulated to maintain the melting temperature inferior to 65°C. The standard potassium concentration is 1 mM to have the lower G-quadruplex melting temperature and best range to measure the induced stabilisation. It is not always possible, for example, F21CTAT have the same melting temperature at 1 mM and 10 mM  $\text{K}^+$  but requires 10 mM to fold into the antiparallel G-quadruplex.<sup>157</sup> Also, less stable G4s can require 10 mM  $\text{K}^+$  since they are not stable enough (i.e. low melting temperature) at lower potassium concentrations, such as for F21T. For the competitors, it must not be complementary to the G4 sequence and have a melting temperature significantly higher than the G-quadruplex to avoid that the competitor unfold before the G-quadruplex. Some ligands are fluorescents and can overlap the fluorescein emission, preventing the measure of the melting temperature. Using another fluorophore FRET-compatible pair, such as cyanines and BODIPY, allows to overcome this issue.

### 3.3.2. G4-FID

Fluorescence intercalator displacement (FID) studies the displacement of an intercalating molecule from a duplex DNA by a putative ligand. It has been adapted by Monchaud *et al.* to evaluate the interaction of G4 ligands on G-quadruplexes.<sup>158,159</sup> The experiment consists on the displacement of an on-off probe by a G4 putative ligand. Standard G4-FID use Thiazole Orange (TO), a versatile probe that binds nucleic acid structures such as duplex-DNA, G-quadruplex and i-motif without discrimination. TO fluorescence is "off" when free in solution: the excitation energy dissipates due to the free rotation of the monomethine bridge between the dihydrobenzothiazole and the quinolinium moiety as represented in Figure 43-A.

This monomethine bridge cannot rotate when TO interacts with the aromatic surface of any nucleic acid turning the fluorescence emission "on". This results in a low quantum yield when free in solution ( $\Phi_F = 2 \times 10^{-4}$ ) that is enhanced by a factor of 500 to 1000 when bound to nucleic acids.<sup>160</sup> TO interacts with most G-quadruplexes with a 1:1 stoichiometry and an apparent binding constant of about  $10^6 \text{ M}^{-1}$ .<sup>161</sup> The on-off probe must have a modest affinity toward the target to allow neglecting the displacement equilibrium of the probe and to reflect the ligand fixation equilibrium.

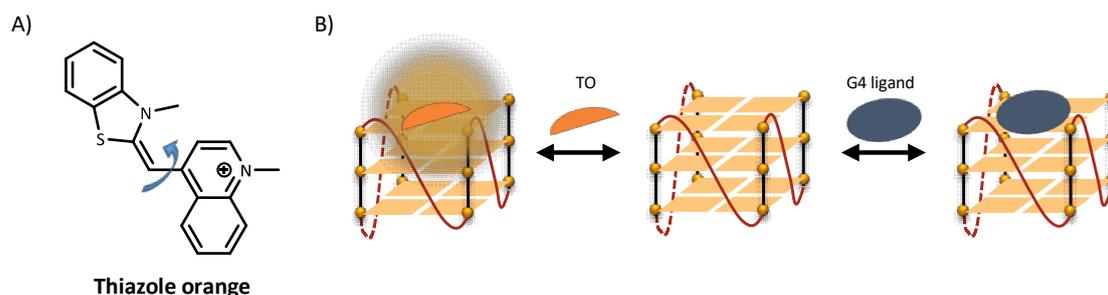


Figure 43: A) Molecular structure of the thiazole orange (TO) with the rotative monomethine bridge (blue arrow). B) Representation of the equilibrium in the G4-FID experiment.

During this experiment, the fluorescence decrease is recorded as a function of the successive addition of a putative G4 ligand that displaces the TO from the G-quadruplex as represented in Figure 43-B. TO displacement is expressed as  $TOD(\%) = 100 (1 - F/F_0)$  with F indicating the fluorescence after addition of the ligand and  $F_0$  representing the initial fluorescence. G4-FID is a semiquantitative assay because the TO displacement adds a step to the ligand interaction with the G-quadruplex. This results in a supplementary chemical equilibrium that impedes the computation of the affinity constant. The  $DC_{50}$  and TOD value are two parameters that can be used to express the interaction of a ligand with a G-quadruplex. The  $DC_{50}$  value corresponds to the ligand concentration required to displace 50% of TO as displayed in Figure 44. Ligands with  $DC_{50}$  lower than  $0.5 \mu\text{M}$  are considered strong binders and ligands with  $DC_{50}$  between  $0.5 \mu\text{M}$  and  $1 \mu\text{M}$  are medium binders. The TO displacement (TOD in %) obtained at a given ligand concentration can also be used as an indicator. The ligand concentration is usually fixed at  $1 \mu\text{M}$  (4 eq. of ligands vs G-quadruplex), and strong ligands display TOD values above 80% (up to 100%) in these conditions.

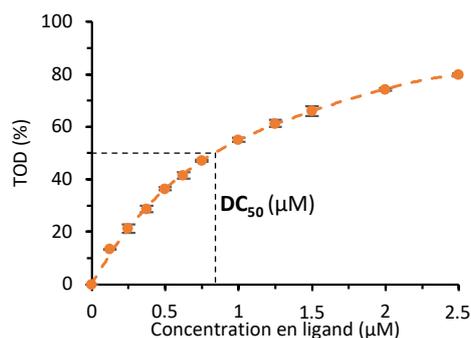


Figure 44: Representation of the displacement curve obtained by plotting TO fluorescence variation (TOD in %) as a function of ligand concentration. The main interaction parameter is the  $DC_{50}$  which corresponds to the concentration of ligand inducing 50% TOD.

G4-FID titration curves can be used to determine the ligand apparent affinity constant uniquely if TO displacement equilibrium can be neglected ( $K_{ligand} \gg K_{TO}$ ). For example, Halder *et al.* determined high association constants for PhenDC3 ( $K_a = 9.4 \cdot 10^8 \text{ M}^{-1}$ ) and PDC ( $K_a = 9.6 \cdot 10^7 \text{ M}^{-1}$ ) on the RNA G-quadruplex ( $G_3U$ )<sub>4</sub>.<sup>162</sup> G4-FID is a convenient and rapid method requiring no specific equipment or modified oligonucleotide. This method has been adapted for high-throughput screening in 96-well plates and similarly to FRET-melting, it allows rapid evaluation and ranking of G4 ligands. TO can also be used with duplex DNA (ds26) enabling evaluation of G-quadruplex vs duplex G4 ligand selectivity. Finally, it is possible to change the TO probe (Figure 45) with a blue-shifted probe (Hoechst 33258) or a red-shifted probe (TO-PRO3) in case of optically incompatible ligands significantly absorbing or emitting at the TO wavelength.<sup>161</sup>

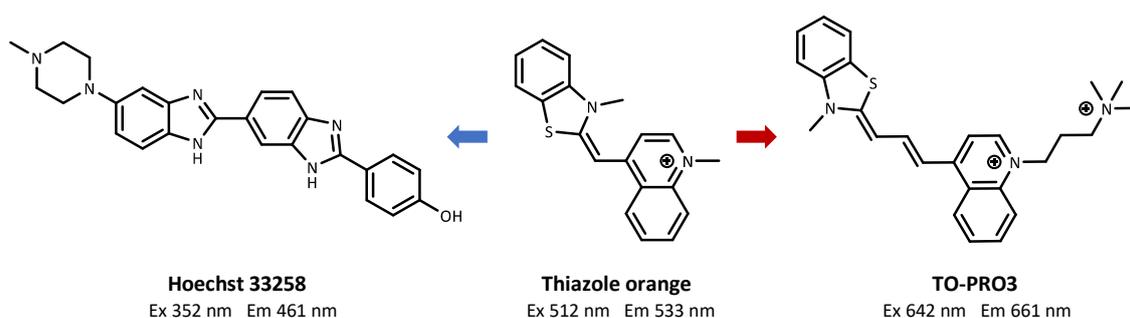


Figure 45: Molecular structures of the different G4-FID compatible probes.

### 3.3.3. CD

CD spectroscopy is a robust method to characterise the G-quadruplex/ligand interaction and a good alternative for faster low-resolution characterisation compared to high-resolution but complex NMR spectroscopy and X-ray crystallography methodologies. Some G4 ligands can induce topology changes upon interaction with the G-quadruplex target. Thus, CD is the preferential method to assign the new topology. This is an interesting feature to control since modification of the G4 topology may impact its function. This phenomenon was observed for both PDC and PhenDC3, able to change the topology of telomeric sequences from hybrid to antiparallel.<sup>60</sup> Marchand *et al.* have reported that structural change of the hybrid human telomeric G4 by PhenDC3 was more complex than just a topology switch to the antiparallel structure, but PhenDC3 could also eject one potassium cation to take its place.<sup>163</sup> In addition to structural characterisation, CD-melting is a typical experiment that allows measuring the induced stabilisation of a G-quadruplex by a ligand. The melting curve is obtained by measuring the decrease of the maximum ellipticity intensity upon temperature increase.

### 3.3.4. UV-visible and fluorescence titration

UV-visible and fluorescence spectroscopy are commonly used to determine the binding affinity of a ligand on a G-quadruplex through classical titration based on the successive addition of a G4 solution at fixed ligand concentration. In the case of UV-visible titration, the ligand absorbance spectrum should vary during the addition of G-quadruplex until reaching a plateau and must not overlap to the G4 absorption band. The curve obtained by plotting the absorbance variation versus the G-quadruplex concentration is then fitted with a model (generally 1:1 or 2:1 depending on the interaction stoichiometry) to determine the affinity constant.

The same approach is applied in the case of fluorescence. The best-case scenario is represented by ligands exhibiting very large fluorescence variation upon binding, typically ON/OFF probes, such as the thiazole orange. The latter is OFF in solution and ON when interacting with the G-quadruplex as explained above. Another example is represented by naphthalene diimide (NDI) which fluorescence is OFF in solution due to self-aggregation and ON when interacting with the G4 target.<sup>164</sup>

In addition, the stoichiometry of interaction between the fluorescent ligand and the G4 can be determined using the continuous variation method (also called Job plot).<sup>165</sup> The fluorescence is recorded as a function of the ligand mole fraction from 0 to 1 at a fixed total concentration with the G-quadruplex. This should give a bell-shaped curve whose maximum indicates the binding stoichiometry.<sup>166</sup>

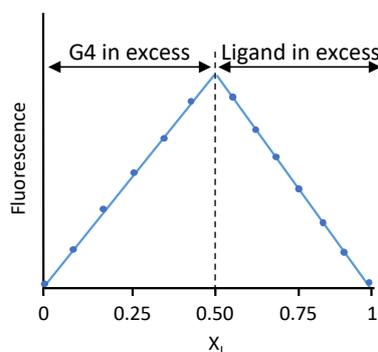


Figure 46: The bell-shaped curve allows to determine the stoichiometry of interaction between the G4 ligand and G4 structure.

The bell-shaped curve as represented in Figure 46 shows that the G-quadruplex is in excess between 0 to the maximum and that the ligand is in excess between the maximum and 1. The rapport between the two gives the binding stoichiometry, for example, a maximum at 0.5 identify a 1:1 stoichiometry.

## 4. G-quadruplex alkylation

### 4.1. Thermally activatable compounds and metalation agents

DNA alkylation is the reaction of transferring an alkyl group to DNA (mainly at nucleobase sites). Nucleobases are subjected to chemical alkylation by several exogenous chemical agents, and often induces DNA damage which, if not repaired, may lead to cell death. Therefore, the use of alkylating agents represents one of the primary therapeutic treatments against cancer. A common strategy used to enhance the recognition of a particular DNA structure is to tether the alkylating moiety to a target-affine ligand, combining both advantages with a more straightforward synthesis pathway than to develop an original 2 in 1 ligand.

A well-known dsDNA alkylating agent is the nitrogen mustard chlorambucil (Chl) presenting the chloroethylamine functional group. Each 2-chloro ethyl arm can cyclise intramolecularly by releasing chloride to form the electrophilic unstable aziridinium ion, which can be attacked by the N7 nucleophilic site of the guanine to form a covalent adduct with DNA as represented in Figure 47.<sup>167,168</sup> Displacement of the second chlorine allows chlorambucil to perform the second alkylation within the same strand or with the opposite strand (interstrand).

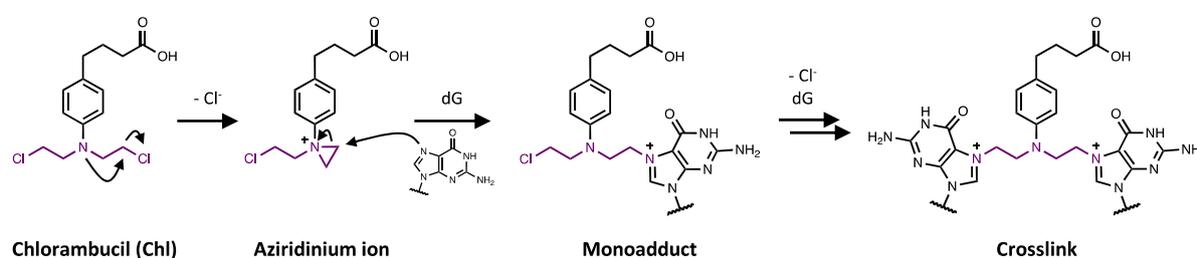


Figure 47: Schematic representation of the alkylation by *chlorambucil* moiety.

Chlorambucil is a drug approved to treat various types of cancer such as chronic lymphocytic leukaemia and Hodgkin and non-Hodgkin lymphomas.<sup>169</sup> Chlorambucil has been tethered to PDS to obtain the alkylating ligand **PDS-Chl** (Figure 48).<sup>170</sup> This ligand was able to alkylate specifically the G4 structures formed by c-MYC22 and 22AG.

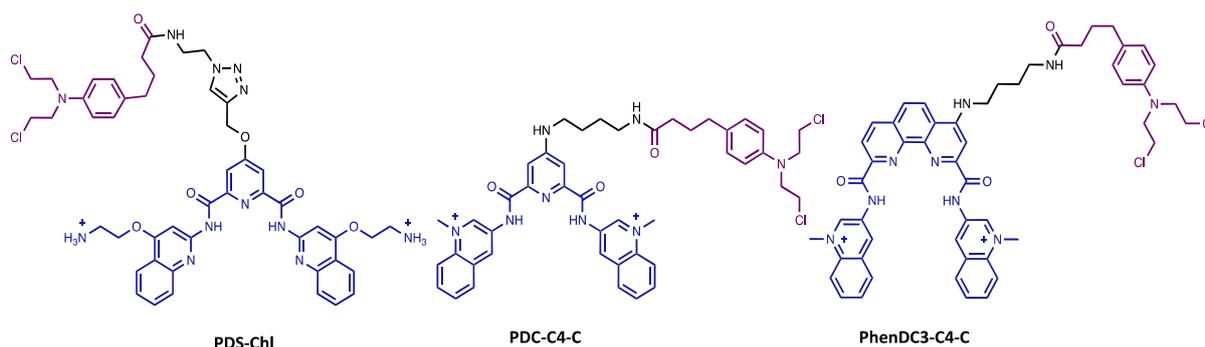


Figure 48: Molecular structure of **PDS-Chl**,<sup>170</sup> **PDC-C4-C** and **PhenDC3-C4-C**.<sup>171</sup>

Later, **PDC-C4-C** and **PhenDC3-C4-C** were synthesised by tethering chlorambucil to PDC and PhenDC3 link through a diaminoalkane spacer. These compounds were used to characterise the mode of binding of the compounds on a G-quadruplex.<sup>171</sup> This functionalisation had an important impact on the binding affinity which is 10-fold lower for the PDC and >100-fold lower for the PhenDC3. Furthermore, **PDC-C4-C** alkylation experiments in the presence of 22AG highlighted the ability of this ligand to form monoalkylation adducts in proximity of the central loop, suggesting binding of the ligand in this region.

Another approach is based on the synthesis of quinone-methide precursor (QMP). Quinone-methide (QM) are electrophilic intermediate capable of reacting with DNA nucleophilic sites (Figure 49).<sup>172</sup> QMP can be activated by specific trigger such as fluoride ions, thermal treatment and UV irradiation.<sup>173</sup> An important aspect of QMs is the reversible nature of the DNA covalent adducts.<sup>174</sup>

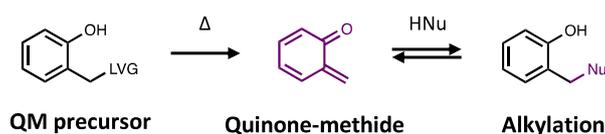


Figure 49: Mechanism of the quinone-methide alkylation.

Naphthalene diimide (NDI) has been tethered to a quaternary ammonium salt of a Mannich base to obtain a new family of quinone methide precursors. **NDI-QMP** (Figure 50) is a thermally activable alkylating G4 ligand that has been demonstrated to form reversible binding with G-quadruplexes.<sup>175</sup>

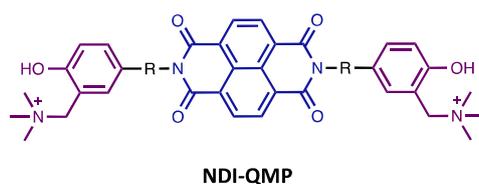


Figure 50: Molecular structure of **NDI-QMP** by Di Antonio et al.<sup>175</sup> R = C1, C2, or C3.

A different family of compounds is represented by metalation agents that has been tethered to different G4 ligands. Cisplatin is a well-known anticancer drug that interacts with the DNA after intracellular activation by substituting of the two chlorines by water to form an aquo complex. The latter hydrated agent reacts mainly with the N7 of the guanine, and less efficiently with adenine, to form a monofunctional or a bifunctional adducts (Figure 51).<sup>176</sup> The bifunctional adduct corresponds to either the intrastrand or interstrand crosslinking products generated between neighbour purine bases.

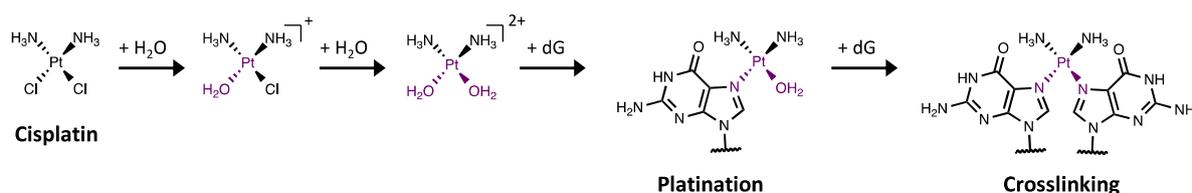


Figure 51: Mechanism of platination by cisplatin.

This strategy was utilised for G4 platination by Bertrand *et al.* They reported **Pt-MPQ** a quinacridine tethered with a monofunctional platinum complex (Figure 52).<sup>177</sup> This compound has been reported to be able to alkylate 22AG selectively on the guanines located on the same external G-quartet. Another derivative was developed by functionalising PDC with a N-heterocyclic carbene-platinum complex (NHC-Pt).<sup>178</sup> **NHC-Pt-PDC** presents a similar 22AG platination pattern as the one reported above, but in addition, was also able to generate metalation bonds with the adenines located in the loops of the structure. This compound was evaluated in cells and showed relevant effects at telomeres.

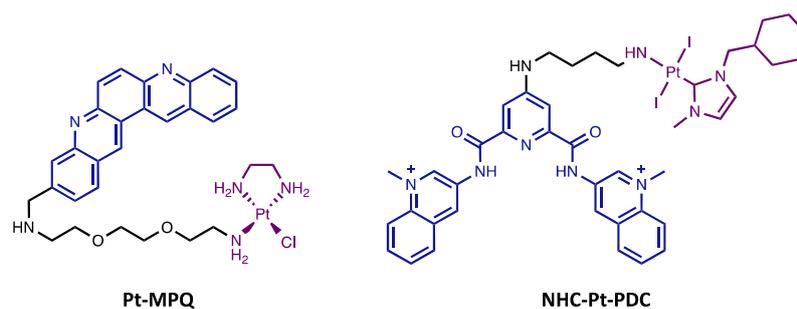


Figure 52: Molecular structures of **Pt-MPQ**<sup>177</sup> and **NHC-Pt-PDC**.<sup>178</sup>

## 4.2. Photoactivatable G4 ligands

The previous methodology lacked the possibility of spatiotemporal control for the alkylation or platination due to the thermal activation of the alkylation moieties. It is why, there has been recent development of photoactivatable G4 ligands, designed in a similar way with a known G4 ligand tethered to a photoactivatable moiety, to control the spatiotemporal alkylation.

Verga *et al.* reported a PDC functionalised by either a benzophenone or a tetrafluorophenyl azide to form a new family of photoactivatable G4 ligands.<sup>179</sup> Both photoactivatable moieties are well-known for their efficiency and have been utilised in photolabelling strategy. The benzophenone (Figure 53-Top) upon UV irradiation generates a triplet transient that has a biradical character, which reacts specifically with thymine to give an oxetane product.<sup>180</sup> The tetrafluorophenyl azide (Figure 53-Bottom) upon UV irradiation generates a singlet carbene transient which reacts by H-abstraction and with nucleophile sites.<sup>181</sup>

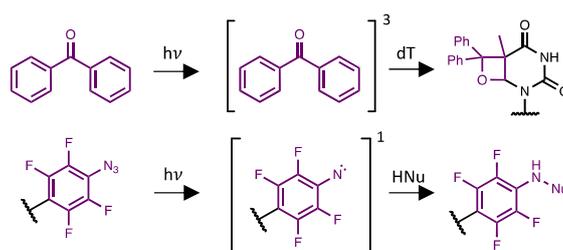


Figure 53: Alkylation mechanisms upon UV irradiation for benzophenone (top) and tetrafluorophenyl azide (bottom).

Each photoactivatable moiety (Figure 54) was tethered to the PDC by a C4, C8 or a PEG spacer. Functionalisation slightly negatively impacted the PDC core affinity for the G-quadruplexes but conserved its selectivity versus duplex DNA. The photochemical reactions displayed good alkylation yields for C4 and PEG derivatives comprised between 20% and 30%. Differently, the C8 derivatives showed poor alkylation (< 12%). This result can be explained by considering the lipophilic character of the linker, which places the alkylating moiety far apart from the G4 structure. Sequencing experiments highlighted a preferential reactivity of benzophenone derivatives for the thymine bases located in the loops of the telomeric G-quadruplex structures. Differently, aryl azide derivatives prefer to react with the guanines of the quartets.

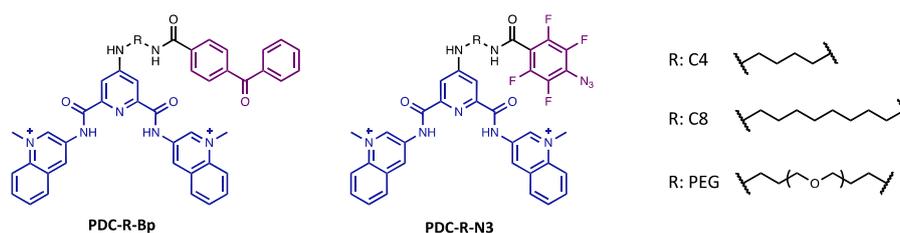


Figure 54: Molecular structure of **PDC-R-Bp** and **PDC-R-N3** developed by Verga *et al.*<sup>179</sup>

Recently, Lena *et al.* have published a new class of photoactivatable G4 ligands using photoactivatable quinone methide precursors (QMP) as photoactivatable moiety and also in the scaffold to interact with the G-quadruplex.<sup>182</sup> These compounds are composed by a benzene tethered to two QMPs, either phenol or 2-naphthol derivatives, to form a broad V-shape structure (Figure 55). These QMPs showed poor G4 stabilisation properties surely due to poor superposition with the G-quartet. The photoactivatable ligands show the ability to generate monoalkylated and crosslinking products with the human telomeric sequence (22AG), but with modest yields and reduced selectivity.

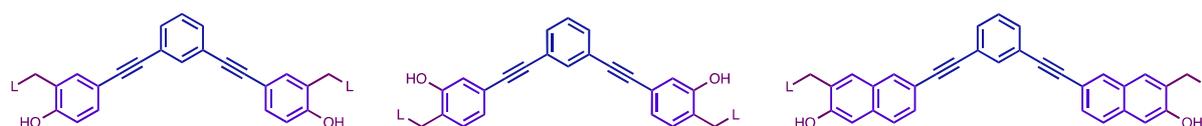


Figure 55: Molecular structures of the new class of photoactivatable G-quadruplex ligands developed by Lena *et al.*<sup>182</sup> L = -NMe<sub>2</sub> and -NMe<sub>3</sub>Cl.

Another approach was developed by using a bimolecular tool comprising a pro-reactive furan tethered to a G4 ligand and a photosensitiser. The latter is triggered by light to generate <sup>1</sup>O<sub>2</sub> which oxidises and activates the furan group as a keto-enol derivative, which can react with nucleophilic sites of a G4 structure. Cadoni *et al.* proposed **PDC-A** (Figure 56) as a PDC functionalised with a furan, which can be activated by methylene blue, capable to alkylate G4s.<sup>183</sup>

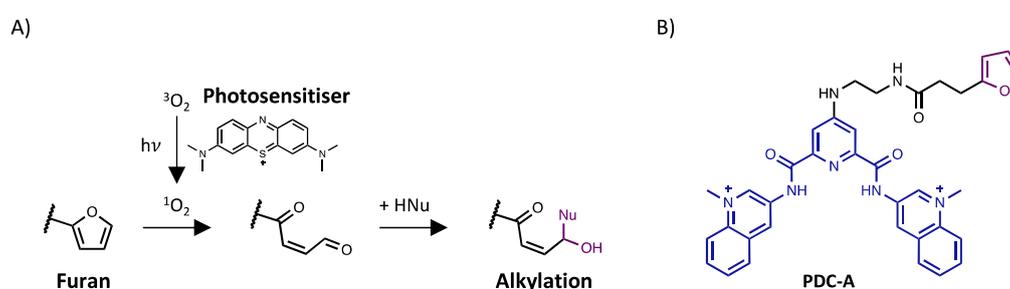


Figure 56: A) Representation of the alkylation mechanism activated by UV irradiation through the photosensitiser. B) Molecular structure of **PDC-A** from Cadoni *et al.*<sup>183</sup>

Recently, the same group followed the same approach by tethering the photoactivatable moiety between a peptide nucleic acid (PNA) and PDC (Figure 57-A) to improve its G4 recognition properties.<sup>184</sup> In addition to the furan derivative, two new photoactivatable moieties (aliphatic diazirine and aryl azide) were evaluated (Figure 57-B). In the presence of the c-kit G-quadruplex, the aliphatic diazirine did not form any alkylated adduct, aryl azide showed a poor conversion, and the furan derivative showed alkylation yields of about 30%. The **PDC-Furan-PNA** compound is a good candidate to alkylate a specific G4 sequence thanks to the PNA allowing specific sequence recognition and the PDC to bring the crosslinker at proximity of the G-quadruplex.

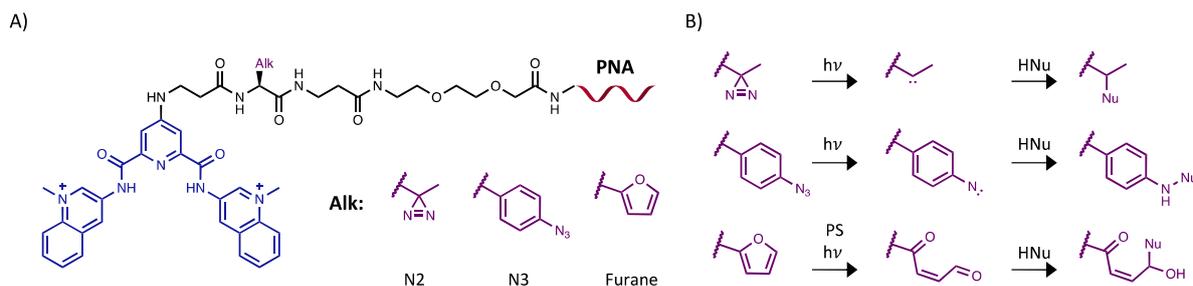


Figure 57: A) Molecular structure of *PDC-Photoalkylating-PNA*,<sup>184</sup> and B) mechanism for each photoactivatable moiety. PS = Photosensitiser.

### 4.3. Photoaffinity labelling

The photoaffinity labelling (PAL) strategy is a method used to study a wide range of interactions between a binder and its target, such as to identify binder-target interactions, localisation of the binding site within the target, and to investigate the structure-function relationships.<sup>185</sup> For this purpose, the PAL strategy converts the reversible non-covalent binding interaction of the binder to its target into a covalent binding interaction through light activation (Figure 58). For that, a photoactivatable moiety must be introduced into the binder scaffold.<sup>186</sup> Upon activation by light, the photoactivatable moiety generates a highly reactive species able to covalently bind its close environment. It can be the surface of a macromolecule during protein-protein interactions or the binding pocket during small molecule-protein interactions.<sup>187</sup>

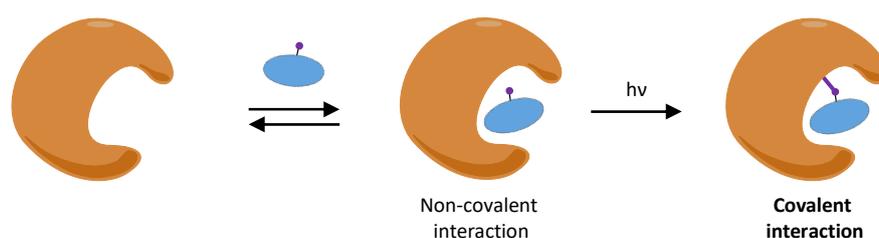


Figure 58: Representation of the photoaffinity labelling strategy.

The modification of the binder scaffold by tethering the photoactivatable moiety must be designed in such a way that the structural features required for the reversible non-covalent binding with the target are retained. It is important to conserve the high binding affinity and functional properties of the binder, as the labelling is the result of the transient species lifetime and the dissociation rate.<sup>188</sup> To schematise, labelling tightens proportionally to the binding of the photoactivatable binder to its binding site. Another important factor is the position of the photoactivatable moiety inside the binder structure as poor positioning relative to the binding site can prevent the efficient labelling.

The photoaffinity labelling lies on the features and limitations of the introduced photoactivatable moiety.<sup>185,187</sup> Ideally, its activation wavelength should be of low energy (>320 nm) to avoid irradiating endogenous chromophores and generating photo-oxidative damage. The generated highly reactive species should be able to react with any chemical groups, and avoid intramolecular rearrangement into less reactive species or quenching through intramolecular reactivity. However, there is no photoactivatable moiety sharing all these qualities, and choosing one should be done in accordance with its compatibility with the application. There are three main photoactivatable moieties used in PAL strategies: aryl azide, benzophenone, and diazirine (Figure 59).<sup>189</sup>

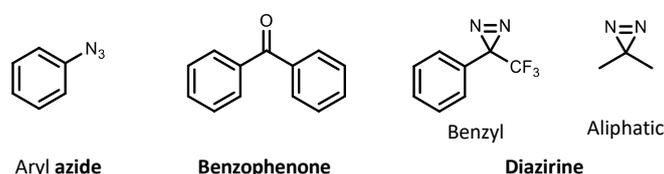


Figure 59: Representation of the four photoactivatable moieties utilised in photoaffinity labelling.

Aryl azide is activated by high energy light (<320 nm) and generates a nitrene transient, which react by C-, N- and O-H insertion.<sup>190</sup> Aryl azide moiety has a relatively small size, therefore, can be introduced in any binder, however, it also reacts by intramolecular ring expansion to generate the less reactive benzazirine and ketenimine transients which react with nucleophiles (Figure 60).<sup>191</sup>

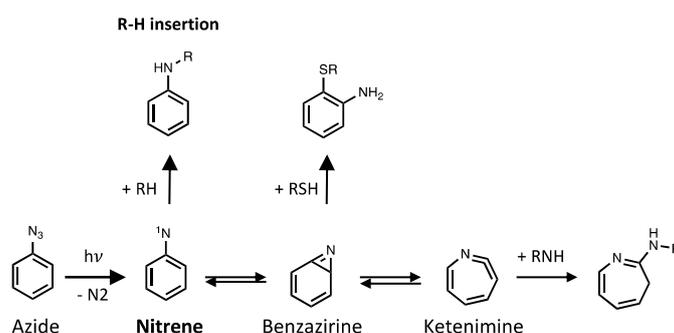


Figure 60: Reactivity of the aryl azide photoactivatable compound.

Benzophenone is activated by lower energy light ( $> 320$  nm) and generates a triplet state having a diradical character.<sup>180</sup> Benzophenone has a preference for C-H insertion, with its main reactivity being the hydrogen abstraction by one radical reacting on the C-H bond generating a new radical on the target carbon, which can couple with the second initial radical to form a C-C bond between the binder and the target (Figure 61).<sup>192</sup> It is also known to react specifically with thymidine in DNA to form oxetane products or to promote thymidine dimerisation by triplet-triplet energy transfer.<sup>193</sup> The benzophenone triplet state has a long lifetime in water and reverses to the ground state when not reacting, allowing it to be repeatedly excited over long irradiation times. Benzophenone is a large moiety increasing the risk for steric clash, however, it has been widely used in pharmaceutical studies.<sup>194</sup>

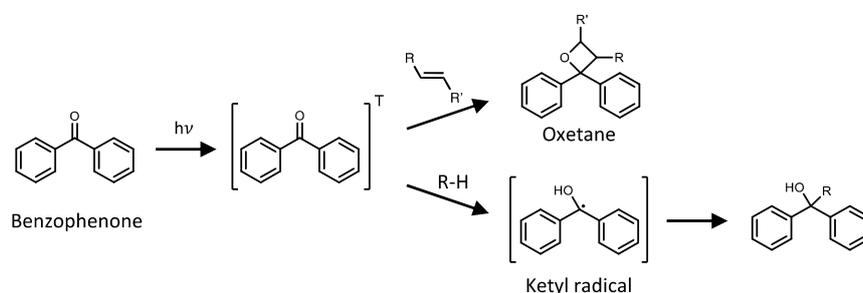


Figure 61: Reactivity of the benzophenone photoactivatable compound.

Diazirine in the benzyl<sup>195,196</sup> or aliphatic<sup>197</sup> form is considered the gold standard. Diazirine is activated by low energy light ( $> 320$  nm) and generates a carbene, a species more reactive than the correspondent nitrene. The generated carbene has been described to react without preference by C-, N-, and O-H insertion and to react with amino acid side chains and backbone (Figure 62).<sup>198–201</sup> Its advantage over the aryl azide and benzophenone is its small size and highly reactive short living carbene, which is responsible for efficient alkylation, with no apparent labelling preference, and high labelling yield.

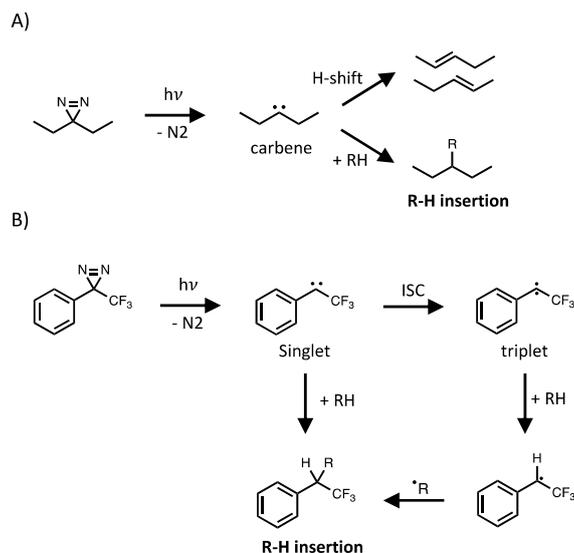


Figure 62: Reactivity of the A) aliphatic and B) benzyl trifluoromethyl diazirine photoactivatable compounds.

The photoaffinity labelling strategy has been applied mainly to protein-protein (introduction of unnatural photoactivatable amino acids) and small molecule-protein interactions, but other targets have been studied such as cofactors, lipids and carbohydrates with photoactivatable proteins.<sup>187</sup> The covalent modifications are visualised by gel electrophoresis and proteins identified by mass spectroscopy. It is now standard to add a tag for recognition and isolation of the covalently bound target, mainly by the addition of a terminal alkyne or a stable azide for post-covalent functionalisation using click chemistry.<sup>187</sup>

Recently, the PAL strategy has been applied for the identification of new G-quadruplex binding proteins through the generation of a tertiary complex. In this case, it is not the direct binding site that is covalently bound but it mediates the photoalkylation towards G4 protein by a small photoactivatable molecule sharing the binding feature for G4 structures. This approach has been developed independently by Su *et al.*<sup>202</sup> and Zhang *et al.*<sup>203</sup> both by tethering an aliphatic diazirine-alkyne chain into the scaffold of a G4 ligand (PDB and PDS respectively). The methodology is similar for both of them (Figure 63): the photoactivatable G4 ligand is incubated in cells and then the irradiation is carried out. The nucleus content is extracted and biotin is introduced by click chemistry. The covalently bound proteins are enriched, digested, and analysed by LC-MS/MS.

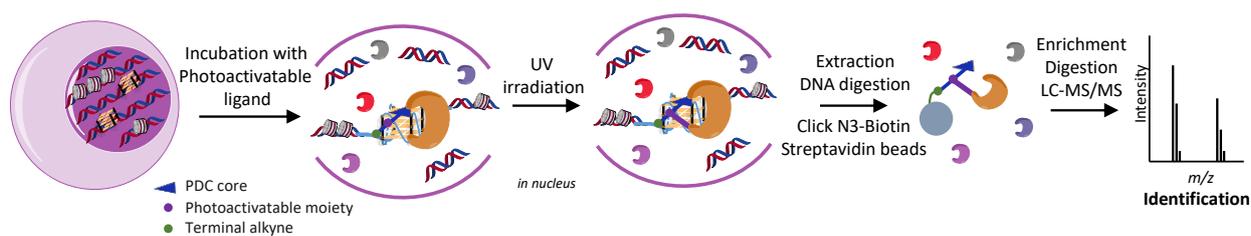


Figure 63: Representation of the PAL strategy utilised by Su *et al.*<sup>202</sup> and Zhang *et al.*<sup>203</sup> to identify new G4 proteins.

The first one allowed to identify 114 and 281 G4 bound proteins in SV589 and MM231 cells respectively,<sup>202</sup> and the second one identified 248 G4 bound proteins in HEK293T cells.<sup>203</sup> Importantly, among well-established G4 binding proteins others unknown partners were identified. The binding of the latter was confirmed by additional *in vitro* experiments.<sup>202</sup> The outcome of these two works confirmed the possibility to employ photoaffinity labelling approaches to identify new G4 binding proteins.

However, the study of small molecules or proteins with nucleic acid was limited due to the absence of powerful methods to identify the covalent binding position on the nucleic acid sequence. Recently, new covalent strategies have been developed to investigate the interaction of small molecules with RNA. Mortison *et al.* determined the binding site of tetracycline analogue Col-3 and doxycycline on 80S ribosomal RNA by tethering the photoactivatable diazirine and stable azide for post-covalent binding functionalisation on the tetracycline structures.<sup>204</sup> Similarly, Mukherjee *et al.* showed the possibility to identify new RNA binders and their binding sites by tethering aryl azide and aliphatic diazirine into the potent RNA binder scaffold.<sup>205</sup> These results demonstrate the possibility of identifying the binding site of a ligand in the nucleic acid sequence. This result opens the possibility to localise non-canonical structures into the RNA and DNA sequences by tethering photoactivatable moiety into non-canonical structure selective binder.



## Aims of the work



G-quadruplexes are DNA and RNA structures reported over 30 years ago.<sup>54,206,207</sup> Accumulation of evidence supports their formation *in vivo*, and they seem to play important roles in regulating biological phenomena such as transcription and translation.<sup>48</sup> Thus, an important question was, and still is, to identify and localise putative G4 sequences (PQS) forming in the genome and to better understand their impact on the organisms. The distribution of G4s determined by *in silico* approaches identified over 350,000 PQS. These studies highlighted that they are distributed non-randomly and non-uniformly in the genome, and are enriched in regulatory regions such as at promoters and in 5'- and 3'-UTR.<sup>208</sup> More recent *in vitro* experiments performed in the presence of purified DNA extracts allowed to identify over 700,000 G4 sequences genome-wide.<sup>88</sup> Recently, *in cellulo* experiments carried out in the presence of a selective G4 antibody (BG4) and with the chromatin organisation preserved identified over 10,000 forming G4s genome-wide, confirming the distribution observed with the *in silico* approaches and demonstrating the importance of the chromatin organisation for the G4s formation.<sup>91</sup> The difference in chromatin organisation between cell lines, such as normal vs. cancer cell lines, could lead to the identification of overrepresented G4s which could become therapeutic markers and targets. The latter methodology is based on immunoprecipitation experiments performed after cell chemical fixation.<sup>92</sup> However, the chemical fixation freezes the genome at one specific state and misses the information associated to the dynamic of G4 folding processes.

Many G4 ligands have been developed to interact selectively with G-quadruplex structures.<sup>122</sup> Quite recently, new approaches have been developed to functionalise well-established G4 ligands with moieties allowing fluorescence detection, covalent bond generation, and antibody/protein recognition.<sup>209</sup> So far, only a few G4 ligands have been developed to isolate and identify G4s.<sup>95,96,145</sup> However, these strategies rely on non-covalent interactions ( $\pi$ -stacking and electrostatic) between the ligand and the G4 structure. Therefore, the generated complex can dissociate during the enrichment and isolation protocol and the ligand can localise on a different genomic site.

The principal objective of this PhD project was to synthesise new chemical biology tools able to pave the way for the development of a new G4 mapping methodology based on the generation of a covalent bond between the ligand and the G4 structure, making impossible the dissociation and relocalisation of the ligand on different genomic sites. We called this method: Chemical immunoprecipitation sequencing (Chem-IP-seq). The strategy is represented in Figure 64: i) cells are incubated with a photoactivatable G4 ligand, ii) UV irradiation allows the generation of a covalent bond between the photoactivatable moiety and the G4 structure, iii) post-binding functionalisation with a tag by copper(I)-catalysed alkyne-azide cycloaddition (CuAAC), and iv) enrichment, isolation and sequencing. The final objective is to map genomic G4s in a dynamic context, avoiding G4 ligand relocalisation to different genomic sites and having a precise view of the accessibility and localisation of G4s genome wide.

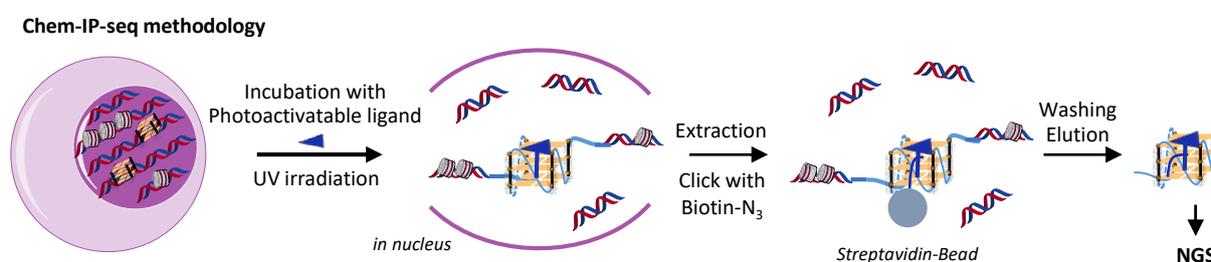


Figure 64: Representation of the **chemical immunoprecipitation sequencing (Chem-IP-seq)** methodology based on photoactivatable G4 ligands.

For this PhD project, two new families of photoactivatable G4 ligands have been synthesised and studied as candidates to be used in the future Chem-IP-seq methodology. These two new families of photoactivatable G4 ligands have been developed by functionalising the aromatic core of PDC (aka 360A) able to interact with high affinity and selectivity with G4s. The PDC core was functionalised with two different moieties: a photoactivatable moiety to trap covalently G-quadruplexes and an alkyne moiety for post-binding functionalisation. The latter will allow adding an hapten by CuAAC reaction necessary to realise the pull-down experiments. Two different families of compounds have been designed and synthesised: we called them **in-line** compounds and **branched** compounds.

The first family was obtained by tethering 4-diazirine-7-octynoate, bearing the photoactivatable aliphatic diazirine and the terminal alkyne, to PDC through a diamino C2, C4 and DEG linker. The second one was obtained by tethering one of the three photoactivatable moieties - benzophenone (Bph), benzyl trifluoromethyl diazirine (N2), and tetrafluorophenyl azide (N3) - with one of the four spacers bearing the alkyne - C1, C3, DEG and TEG -, to PDC through an *N*-azaneylbutyl-4-azaneylbutanamide linker. The two different families of compounds are represented in Figure 65. The synthesis of the different photoactivatable G4 ligands studied in this PhD project was realised by different people: Dr. T. Masson synthesised the **in-line** family, Mrs. C. Landras Guetta synthesised all **branched** G4 ligands except the ones belonging to the DEG series synthesised by myself.

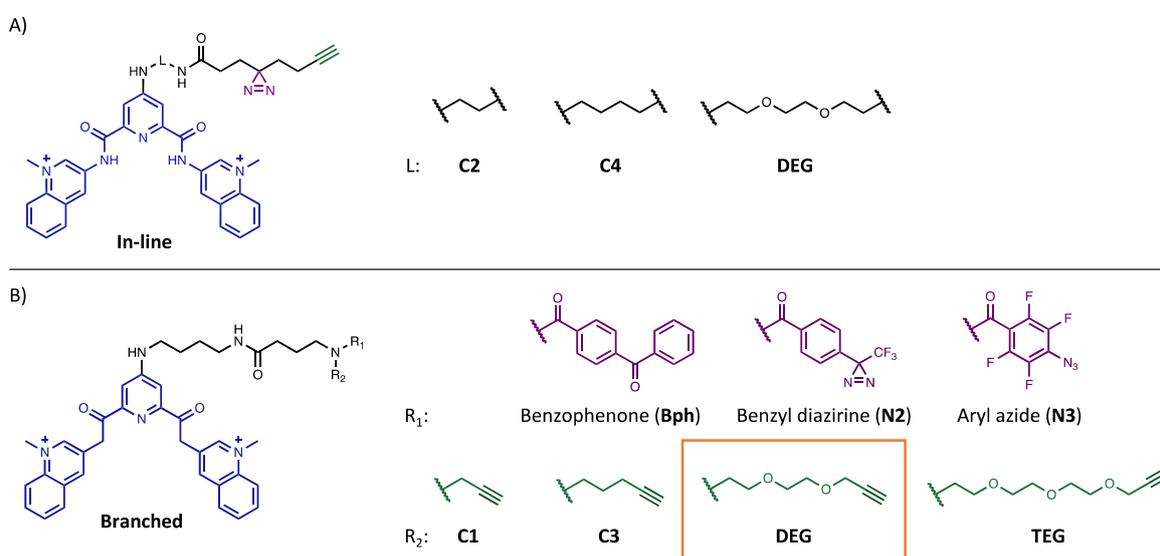


Figure 65: Representation of the molecular structure of the **branched** (top) and **in-line** (bottom) compounds. The orange box corresponds to the DEG series synthesised during this thesis. The photoactivatable moiety is in purple and the alkyne function with (**branched**) or without (**In-line**) spacer in green.

G4 photoalkylation is controlled by two steps represented in Figure 66. The first step is the primordial interaction of the photoactivatable G4 ligand with the G4 structure to form a G4 ligand:G4 complex. The PDC core is known for its ability to selectively and strongly stabilise G4 structures. However, the presence of a functionalisation can be detrimental for its affinity and selectivity toward G4s. This is why, interaction and selectivity of the photoactivatable G4 ligands with G-quadruplexes were assessed by biophysical methods.



Figure 66: Two steps process controlling G-quadruplex photoalkylation.

The second step is the ability of the photoactivatable moiety to form a covalent bond with the G4 structure. The capacity of the photoactivatable G4 ligands to generate covalent adducts was evaluated by denaturing gel electrophoresis. In addition, chemical and enzymatic sequencing experiments were carried out to try to determine the exact G4 alkylation sites.

Finally, the photochemical properties of the compounds and alkylation mechanism were studied in the laboratory of Prof. Mauro Freccero in Italy, taking advantage of the experience of the lab in the characterisation of photoreactive species and photoadducts. To facilitate the experiment, solely the photoactivatable moieties were used for the study. To complete this goal, photoactivatable mimic prototypes (Figure 67-A) were designed by taking into account the nature of the linker tethering the photoactivatable moiety to PDC as the latter could modify their photoreactivity properties. The **in-line** photoactivatable moiety is tethered to PDC through a diamino linker, thus, an *N*-methylamide function was added to the 4-diazirine-7-octynoate precursor. Differently, the **branched** photoactivatable moieties are linked to PDC through a diamino C4 linker and to the terminal alkyne through a spacer. It was decided to replace both linker and spacer with an ethyl group, which results in their *N,N'*-diethylamide derivatives. The prototype structures were used to characterise the nature of the transient species by laser flash photolysis (LFP). Then, the photoreactivity of these compounds was studied by product distribution analysis of reaction carried out in the presence of different nucleosides and the generated adducts were characterised by NMR (Figure 67-B).

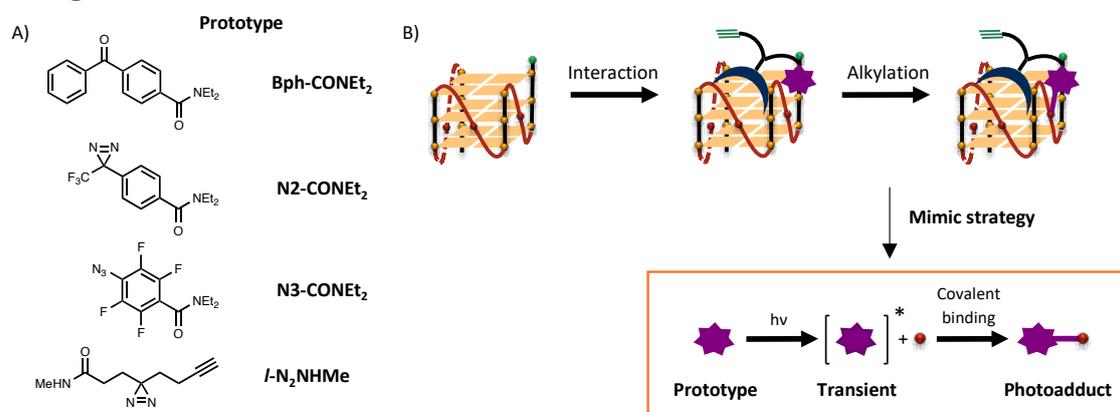


Figure 67: Representation of the mimic strategy using photoactivatable prototypes to study the generation of the transient intermediates and the formation of the photoadducts in the presence of nucleobases. The orange box highlights the mimic strategy.

## Chapter II:

# Synthesis of photoactivatable G4 ligands and biophysical evaluation



To achieve the final aim of this PhD project, we designed and synthesised two different families of G4 ligands. The molecular tools designed for this purpose comprise three motifs: i) a recognition motif able to interact selectively with G4s, ii) a photoactivatable moiety able to trap covalently G4 structures, and iii) a sorting function that enables pull-down experiments. The recognition motif chosen for this work is the PDC core, which is well-known for its high affinity and selectivity for G-quadruplex structures. In addition, **PDC-NH<sub>2</sub>**, its amino derivative, has been synthesised before and is suitable for grafting functions.<sup>179,210</sup> Concerning the photoactivatable motif, we selected a panel of four different compounds known for their high efficiency in photoaffinity labelling approaches: benzophenone (Bph), tetrafluorophenyl azide (N3), aliphatic (*N*-N2) and benzyl trifluoromethyl diazirine (N2).<sup>189</sup> Finally, it was decided to graft a spacer bearing a terminal alkyne, allowing post-binding functionalisation by CuAAC with a hapten (Biotin). The latter can be used for enrichment and isolation with streptavidin-coated beads. This post-binding functionalisation is intended to be used after recognition and alkylation steps: this approach should reduce the ligand bulkiness that could negatively impact the recognition. On this basis, two new families of photoactivatable G4 ligands have been synthesised featuring **in-line** and **branched** compounds represented in Figure 69.

The **in-line** compounds comprise a unique linear photoactivatable-alkyne moiety corresponding to the *N*-4-diazirine-7-octynamide represented in Figure 68. Thus, the aliphatic diazirine (*N*-N2) is the only photoactivatable moiety tethered to the **in-line** compounds. This series' structural variation comes from the three different linkers utilised to connect the *N*-4-diazirine-7-octynamide moiety and PDC: two aliphatic C2 and C4 linkers, and a hydrophilic DEG linker. The synthesis of this family of compounds was conducted by Dr. T. Masson during his PhD and described in more detail in his PhD manuscript.<sup>211</sup>

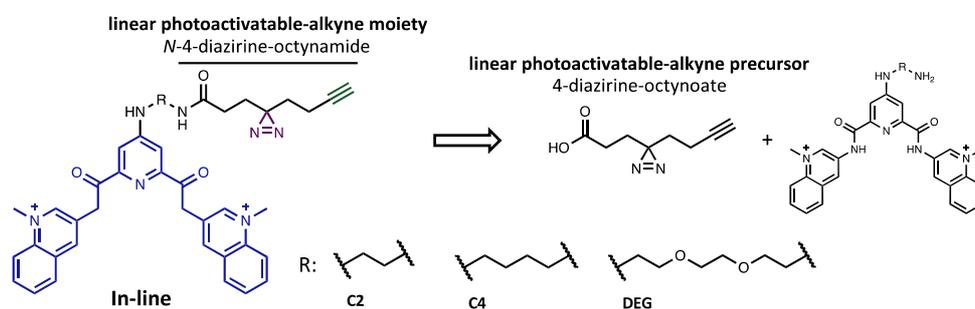


Figure 68: **In-line** G4 ligand retrosynthesis.

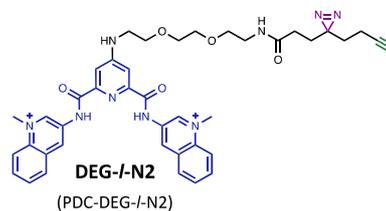
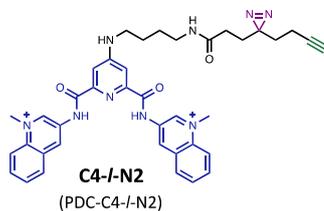
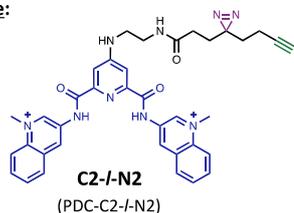
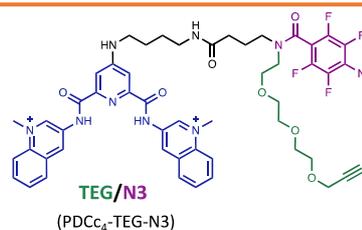
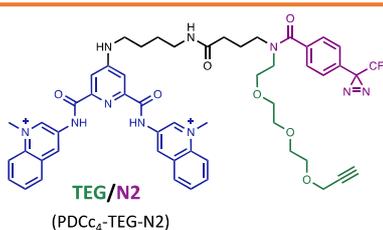
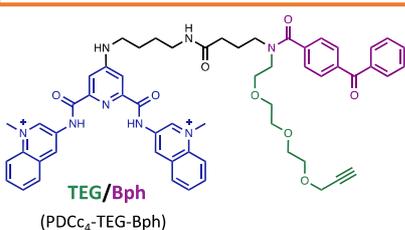
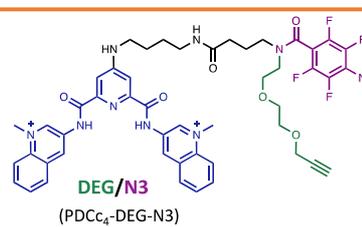
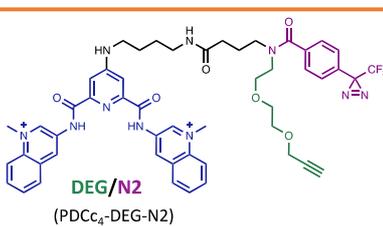
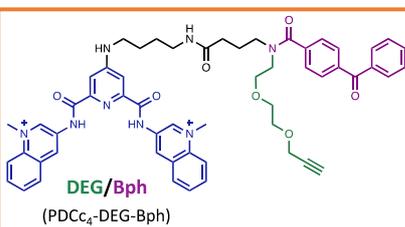
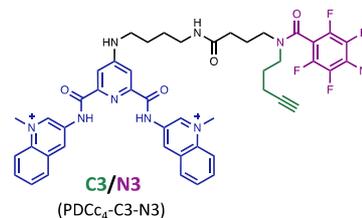
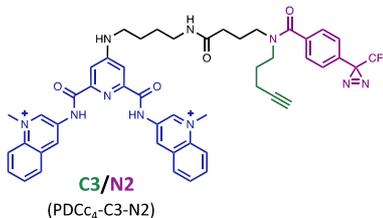
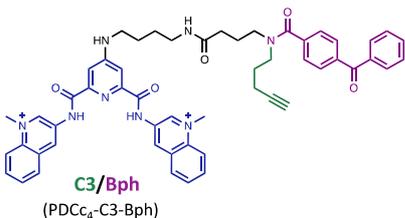
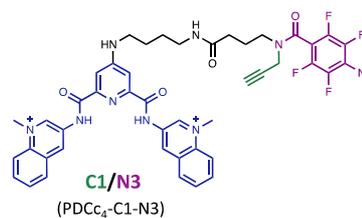
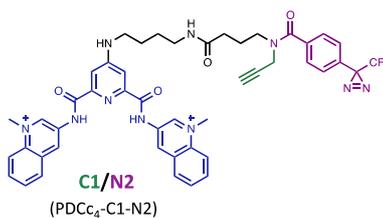
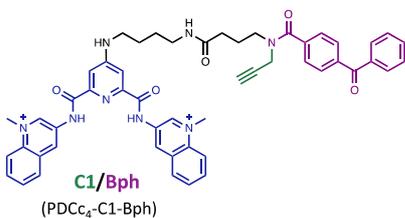
**In-line:****Branched:**

Figure 69: Representation of the photoactivatable G4 ligand library synthesised by the team. These compounds are divided into two different families: **in-line** compounds (**PDC-C2-I-N2**, **PDC-C4-I-N2**, and **PDC-DEG-I-N2**), which vary only for the diamino linker length and nature (C2, C4, and DEG) connecting the PDC core and the diazirine-alkyne moiety; and the branched compounds. The latter present three different photoactivatable moieties (benzophenone (**Bph**) (left), benzyl diazirine (**N2**) (center), and aryl azide (**N3**) (right)) and four different spacers characterised by different length and nature (aliphatic **C1** and **C3**, hydrophilic **DEG** and **TEG**). The orange box put in evidence the series of compounds I synthesised in this work. In each series, the simplified name for each compound corresponds to the length of the variable linker (for **in-line**) or spacer (for **branched**) and the nature of the photoactivatable moiety.

The **branched** compounds share a ramified-type structure where PDC is linked to the terminal alkyne and photoactivatable moiety by an *N*-azaneylbutyl-4-azaneylbutanamide chain represented in Figure 70. The amine in C3 is converted into a tertiary *N,N*-amide by tethering the photoactivatable moiety and the terminal alkyne. The first variability corresponds to the three different photoactivatable moieties (benzophenone, benzyl trifluoromethyl diazirine, and tetrafluorophenyl azide) introduced in the **branched** compounds, aiming to identify the most efficient photoactivatable moiety.

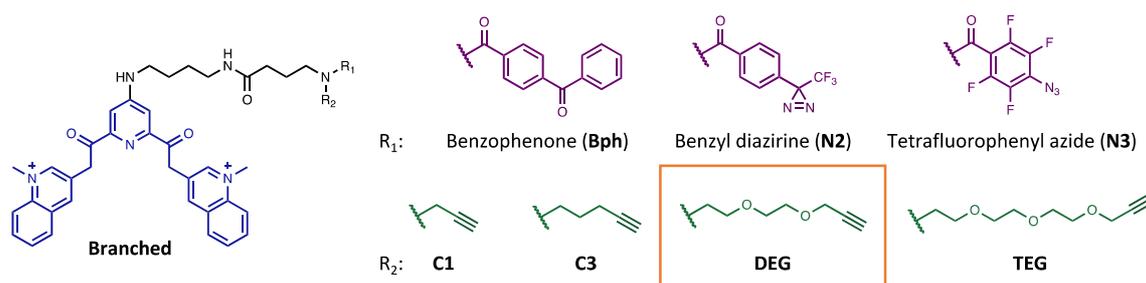


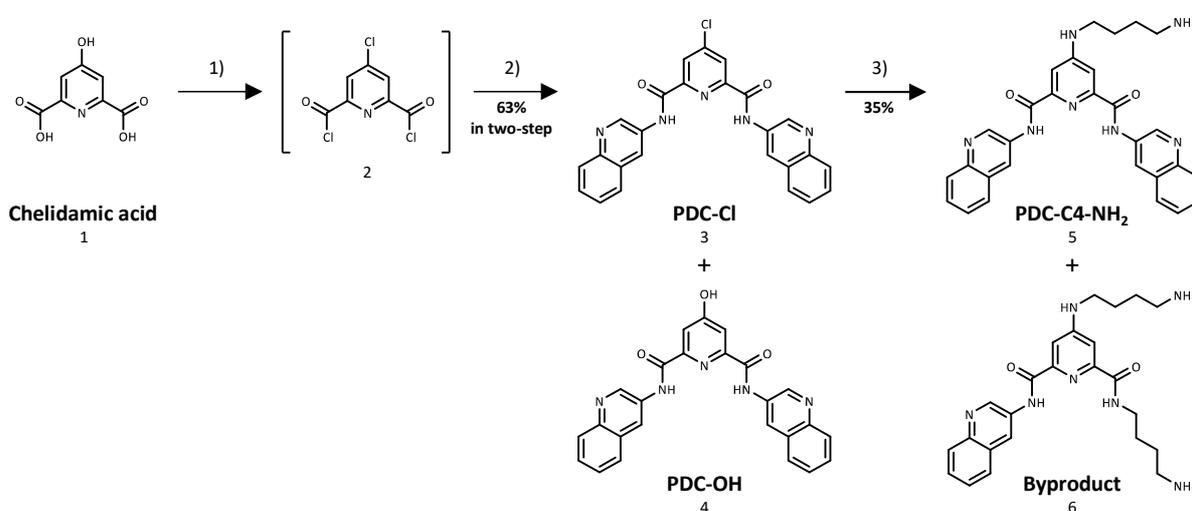
Figure 70: Representation of the branched scaffold and the different combinations of photoactivatable moieties ( $R_1$  in purple) and alkyne spacers ( $R_2$  in green).

The second structural variability corresponds to the nature and length of the spacer bearing the terminal alkyne. The reason for this diversity is to improve its accessibility for post-binding functionalisation by copper-catalysed azide-alkyne cycloaddition (CuAAC) reaction. The four spacers are the aliphatic C1 and C3, and the hydrophilic DEG and TEG. The combination of these structural variations provided a family of 12 compounds. C1-, C3-, and TEG derivatives were synthesised by Mrs. C. Landras Guetta and M. M. Seze, and are described in Dr. T. Masson's PhD thesis manuscript.<sup>211</sup> The remaining DEG series have been synthesised by myself.

# 1. Synthesis

## 1.1. PDC-NH<sub>2</sub> synthesis

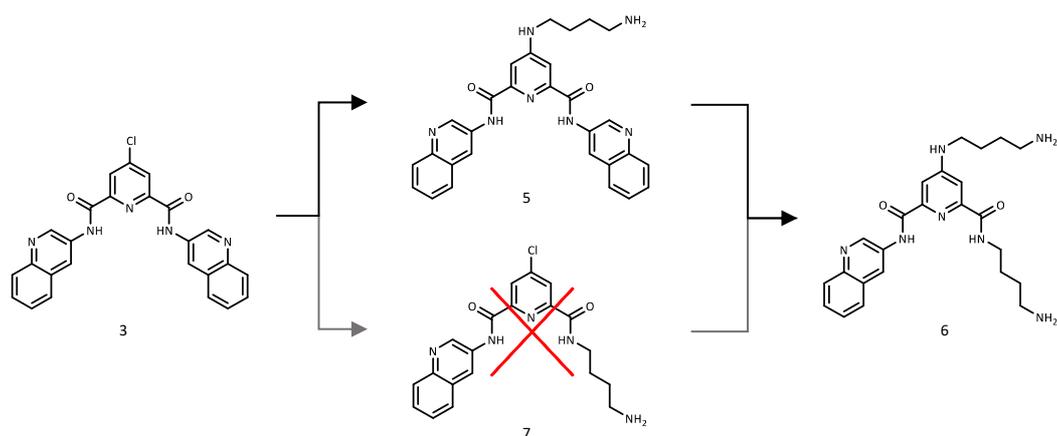
PDC has an excellent affinity and selectivity for G4s and has often been used as a recognition core in many molecular tools.<sup>212</sup> PDC-NH<sub>2</sub> derivatives were designed and synthesised to functionalise easily the PDC core with different moieties. Structurally, PDC-NH<sub>2</sub> were obtained by functionalising the position 4 of the central pyridine with a diamino-alkane linker, thus leaving a primary amine available for functionalisation (Scheme 2).



Scheme 2: Synthesis of PDC-C4-NH<sub>2</sub> (5) from chelidamic acid (1). Conditions: 1) SOCl<sub>2</sub>, DMF, 80 °C, 16 h. 2) 3-aminoquinoline, TEA, DMF, r.t., 16 h. 3) 1,4-diaminobutane, TEA, 90 °C, 16 h.

Various linkers differing by nature and length can be introduced following the same synthetic pathway, and the more common PDC-NH<sub>2</sub> derivative, named PDC-C4-NH<sub>2</sub>, is obtained by aromatic nucleophilic substitution of PDC-Cl in the presence of diamino butane (C4). This is a key compound used in the synthesis of both in-line and branched compounds. This section is going to describe the synthesis of PDC-C4-NH<sub>2</sub>. This compound was synthesised by following the synthetic pathway reported by Dr. Masson in his PhD thesis manuscript.<sup>211</sup>

**PDC-Cl (3)** was synthesised following an improved one-pot procedure published by Xie *et al.*<sup>213</sup> In this synthesis (Scheme 2), 4-chloropyridine-2,6-dicarbonyl chloride (**2**) was obtained by reacting chelidamic acid (**1**) with thionyl chloride. The formed acyl chloride **2** functions were used to introduce the 3-aminoquinoline by nucleophilic acyl substitution reaction to form **PDC-Cl**. **PDC-C4-NH<sub>2</sub>** (**5**) was obtained by following a published protocol.<sup>179</sup> **PDC-Cl** was submitted to an aromatic nucleophilic substitution (S<sub>N</sub>Ar) to replace the chlorine in position 4 of the pyridine by the 1,4-diaminobutane present in excess to generate **PDC-C4-NH<sub>2</sub>**. Specifically, the reaction was carried out overnight at 90 °C using 1,4-diaminobutane as solvent.



Scheme 3: Byproduct **6** is generated from **PDC-C4-NH<sub>2</sub>** (**5**) as the theoretical quinoline substituted derivative **7** is not detected by LC-MS.

In literature, these reactions are reported to be efficient and generate products with high yields. However, I encountered different pitfalls. During the synthesis of **PDC-C4-NH<sub>2</sub>**, I obtained the product with a poor 24% yield and the reaction mixture showed the presence of a byproduct. The latter was characterised by MS and its mass matches byproduct **6** obtained from transamidation reaction of one 3-aminoquinoline by a 1,4-diaminobutane. Intermediate **7** could not be found in the LC-MS chromatogram (Figure 71), thus, the byproduct must have been formed by substitution of one 3-aminoquinoline of **PDC-C4-NH<sub>2</sub>** (Scheme 3).

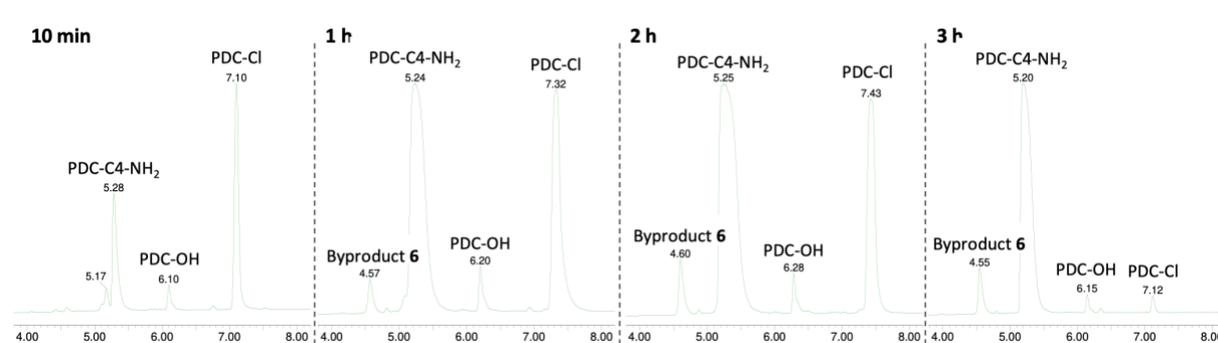


Figure 71: LC-MS chromatograms of the synthesis of **PDC-C4-NH<sub>2</sub>** at 70 °C after 10 min, 1 h, 2 h and 3 h of reaction.

Decreasing the reaction temperature allowed better control of the reaction progression, limiting the formation of the byproduct to an insignificant trace. The reaction occurs in 1-2 h at 90 °C, 3 h at 70 °C and over 6 h at 50 °C. However, the yield (~ 30%) remains low regardless of the temperature. Filtration was not enough to remove completely the byproduct, this is why it was essential to monitor the reaction by LC-MS to avoid forming large quantities of byproduct. In addition, monitoring the substitution of the aromatic hydroxyl group into chlorine by NMR is essential to avoid the formation of PDC-OH (**4**).

## 1.2. Synthesis of **Branched DEG G4 ligands**

The **branched** compounds are bifunctional PDC compounds comprising a photoactivatable moiety and a spacer bearing a terminal alkyne, both introduced on a branched linker coupled to **PDC-C4-NH<sub>2</sub>**. The three photoactivatable moieties used to functionalise this family of compounds are benzophenone (**Bph**), benzyl trifluoromethyl diazirine (**N2**) and tetrafluorophenyl azide (**N3**). The spacers bearing the terminal alkyne are the aliphatic **C1** and **C3**, and the hydrophilic **DEG** and **TEG**. In these derivatives, the position of the photoactivatable moiety is always the same, independently of the nature and spacer series. Thus, the relative distance of the photoactivatable moieties from the G4 structure is fixed. Only the distance between the interaction site and the alkyne changes depending on the spacer. In principle, this should modify the accessibility of the terminal alkyne during CuAAC.

The syntheses of the **C1**, **C3**, and **TEG** series have already been covered in Dr. Masson's thesis manuscript, and only the **DEG** series (Figure 72) will be described in this section. The four series globally share similar synthetic pathways, starting from the introduction of the spacer bearing the terminal alkyne and the photoactivatable moiety into a common branched linker, which can later be tethered to the **PDC**.

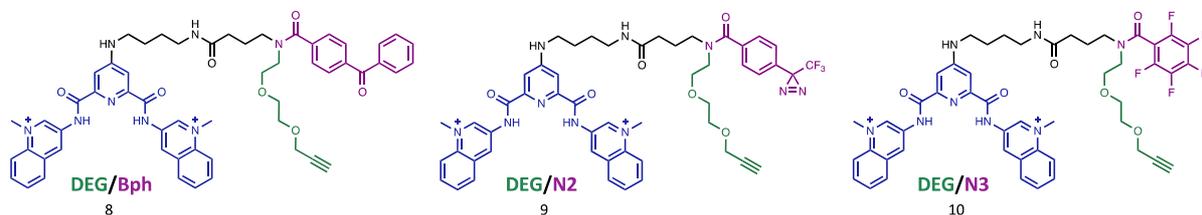
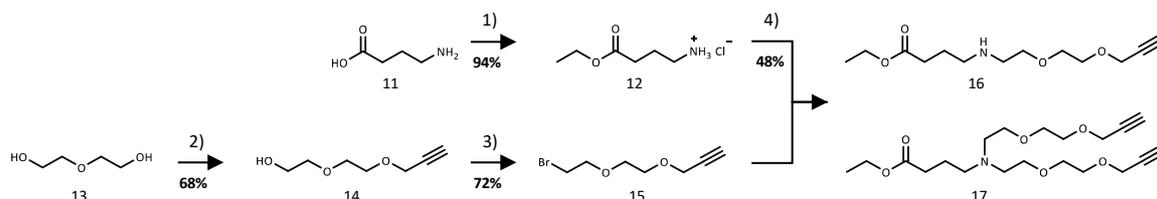


Figure 72: **Branched** compounds of the DEG series (**8-10**) synthesised in this work.

### 1.2.1. Synthesis of the common linker intermediate

The photoactivatable moieties are all introduced on a common intermediate linker bearing a secondary amine. Only after, the functionalised linker is tethered to the PDC core. The common intermediate linker is obtained by nucleophilic substitution between the ester-protected  $\gamma$ -aminobutyric acid **12** and compound **15** (scheme 3). At first, the carboxylic acid precursor **11** is protected in its ester form by using thionyl chloride in ethanol and precipitated into the amine hydrochloride form **12**. Precursor **14** is synthesised by a nucleophilic substitution reaction of one of the free hydroxyl groups of diethylene glycol **13** with propargyl bromide. Compound **15** is generated by Appel reaction conducted on compound **14** using tetrabromomethane and triphenylphosphine. Finally, ethyl 4-aminobutanoate hydrochloride **12** is deprotonated to allow the nucleophilic substitution with compound **15** to form the common intermediate linker **16** with 48% yield.



Scheme 4: Synthesis of the common intermediate linker. Conditions: 1)  $\text{SOCl}_2$ , EtOH, 80 °C, 5 h. 2) 3-Bromopropyne, *t*-BuOK, THF, r.t., 16 h. 3)  $\text{CBr}_4$ ,  $\text{PPh}_3$ , DCM, r.t., 16 h. 4) TEA, ACN, 40 °C, 48 h.

However, during this reaction step, we could observe the generation of a byproduct obtained from a second nucleophilic substitution reaction occurring on compound **16** forming the tertiary amine **17**. These two products have a close polarity, making them not separable by classic silica or alumina chromatography with or without ammonia or triethylamine. RP-HPLC was avoided since the products do not have chromophores to be detected. Therefore, it was decided to use the mixture containing both products in the following step, since the tertiary amine cannot react in such reaction conditions. The mixture was characterised by LC-MS and NMR, bisubstituted byproduct **17** represents 20% of the mixture (NMR proportion, Figure 73)

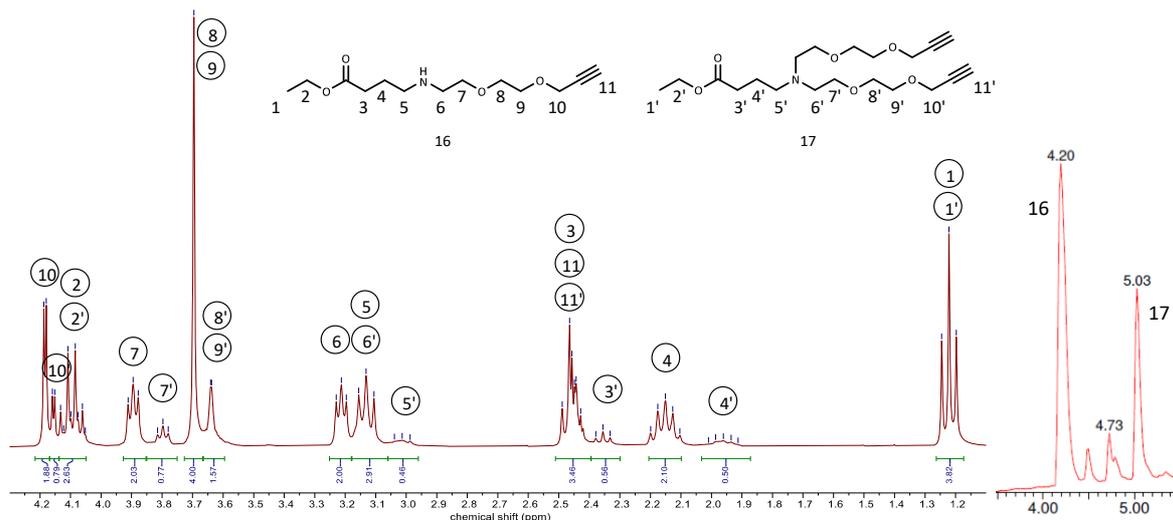
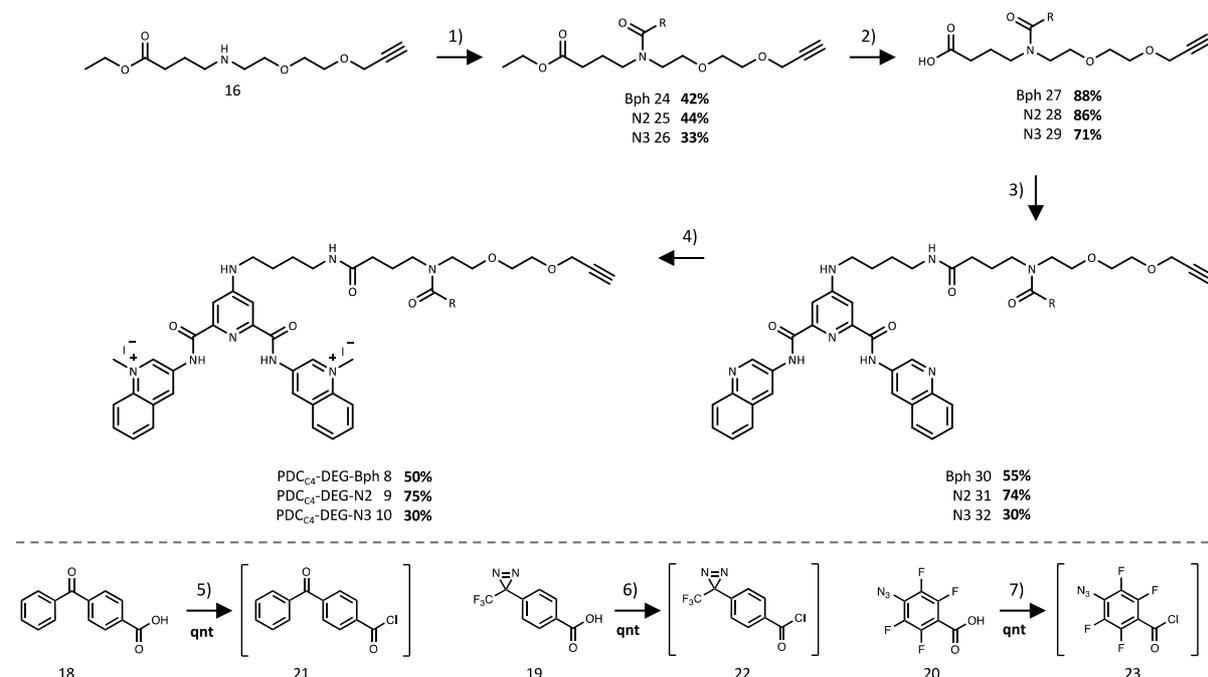


Figure 73: Left)  $^1\text{H}$  NMR (300 MHz) spectrum. Right) MS chromatogram of the mixture of common linker intermediate **16** and byproduct **17**.  $M/z$  (ESI+)  $[M+H]^+$  pic 4.2 min = 258.4 and pic 5.0 min = 384.5, and calcd. for **16**  $[M+H]^+$  = 258.3 and for **17** = 384.5.

## 1.2.2 Synthesis of the photoactivatable G4 ligands

The different photoactivatable moieties are introduced on compound **16** in their acyl chloride forms (**21-23**) (Scheme 5). These compounds **21-23** are obtained by reacting benzophenone acid **18** and benzyl trifluoromethyl diazirine **19** with thionyl chloride, and tetrafluorophenyl azide **20** with phosphorus pentachloride.



Scheme 5: Synthesis of the DEG series photoactivatable G4 ligands. Conditions: 1) Acyl chloride derivative (**21-23**), TEA, DCM, r.t., 16 h. 2) NaOH, DCM/EtOH/H<sub>2</sub>O, r.t., 16 h. 3) PDC-C4-NH<sub>2</sub>, HATU, DIPEA, DMF, 0 °C to r.t., 1 h. 4) MeI, DMF (for **8** and **9**) or ACN (for **10**), 40 °C, 16 h. 5)  $\text{SOCl}_2$ , 80 °C, 2 h. 6)  $\text{SOCl}_2$ , 80 °C, 20 h. 7)  $\text{PCl}_5$ , Et<sub>2</sub>O, r.t., 2 h.

The secondary amine group of compound **16** reacts with the appropriated acyl chloride to form the corresponding *N,N*-disubstituted amides **24-26**. Then, the carboxylic acid function is recovered by saponification reaction, giving the expected products **27-29**.

The next step consists in the peptide coupling reaction carried out with HATU and **PDC-C4-NH<sub>2</sub>** (**5**). To verify the efficiency of the coupling system, HATU activation of the benzophenone derivative **27** was monitored by LC-MS after quenching with methanol. The chromatogram showed a rapid and total conversion. The latter presented a major peak, which mass corresponds to methyl ester **34**, and multiple small peaks, which we associated to degradation products. Product **34** is formed by reaction of the HOBt activated product **33** with the methanol used to quench and prepare the LC-MS sample (Figure 74). It is interesting to note that a similar assay was realised on TEG benzophenone derivative, but the HOBt activated product showed higher stability and slower quenching reaction (data not shown). To avoid any product degradation, it was decided to mix the carboxylic acid precursors (**27-29**) and **PDC-C4-NH<sub>2</sub>** (**5**) before slowly adding HATU to the solution. This protocol gave the expected products (**30-32**).

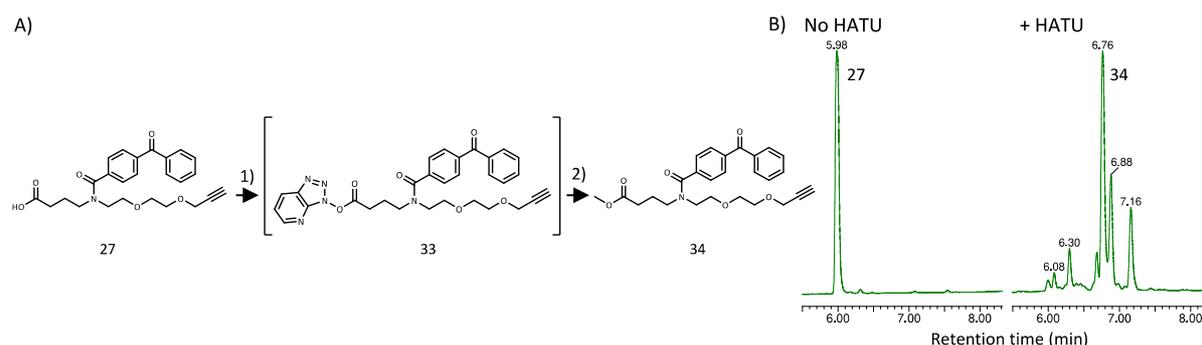


Figure 74: A) scheme of the HOBt intermediate **33** formations by the HATU activation of the carboxylic acid benzophenone derivative **27**, which is quenched by methanol during the preparation of the sample for LC-MS analysis. Conditions: 1) HATU, DIPEA, DMF, r.t., 2) MeOH, r.t. B) LC chromatogram of the reaction before adding HATU (left) and LC-MS ran 5 min after adding HATU (right).

Finally, the quinoline rings are methylated using an excess of iodomethane to obtain the final products **PDC<sub>C4</sub>-DEG-Bph** (**8**), **PDC<sub>C4</sub>-DEG-N2** (**9**) and **PDC<sub>C4</sub>-DEG-N3** (**10**) (Scheme 5). After the workup, we realised that the tetrafluorophenyl azide derivative (**10**) was not pure. LC-MS analysis highlighted the presence of a second peak, which corresponds to the reduction of azide group. Indeed, azides are known to be sensitive to heat, and the removal of the DMF occurs at 40 - 45 °C, which could induce azide reduction. Therefore, DMF was replaced by acetonitrile to avoid heating when removing the solvent and successfully gave the expected product. The conversion could not be pushed to 100%, and traces of mono-methylated product remained in mixture (< 5%). **PDC<sub>C4</sub>-DEG-N3** (**10**) was used without any further purification for the biophysical and biochemical assays.

## 2. Biophysical evaluation of the photoactivatable G4 ligands

After having synthesised all the target compounds, it was essential to ensure that all the functions of the trifunctional molecular tools did not significantly affect PDC binding properties. This section focuses on the propensity of the PDC core to still interact and stabilise G4 structures selectively once embedded into the **in-line** and **branched** architectures. Indeed, it is well-known that adding functional moieties to a G4 ligand may negatively impact its interaction with G-quadruplexes and its selectivity versus duplex DNA.<sup>179</sup> In general, this impact is correlated with the engendered bulkiness of the newly introduced moiety. In our case, **in-line** compound binding properties (stabilisation and selectivity) should not be impacted significantly due to the tiny size of the linear moiety. However, a more critical impact might be expected with the **branched** compounds due to the ramified structure and the bulkied photoactivatable moieties. FRET-melting and G4-FID are two biophysical routines and high-throughput methods described in the introduction (Chapter I.3.3, p. 62), and were utilised to confirm that the stabilisation and selectivity for G4 structures were not affected. FRET-melting measures the stabilisation induced by a compound for a specific G4 structure.<sup>156</sup> It can be used to evaluate the selectivity with DNA or RNA, using a duplex DNA or a transfer RNA (tRNA) as competitors. The G4-FID measures the displacement of the Thiazole orange (TO) probe by a compound in a competition assay.<sup>158,159</sup> These two biophysical methods are complementary and allow to obtain information on the interaction by G4-FID and the stabilisation and selectivity by FRET-melting. Thus, these methods are well adapted to verify the remaining efficiency of the new PDC derivatives.

### 2.1. DNA and RNA sequences used in the biophysical studies

A panel of G4 forming sequences representative of the polymorphism and structural characteristics of the G4s have been selected (Table 1). It comprises six DNA sequences covering the three G4 topologies (parallel, antiparallel, and hybrid), small to long loops and mix-quartet. One RNA sequences have also been selected to evaluate the ligand interaction with RNA G4 by competitive FRET-melting.<sup>214</sup> Most DNA sequences have been utilised in both biophysical methods. However, FRET-melting requires 5'-FAM / 3'-TAMRA functionalisation, and the sequences may be truncated to optimise the energy transfer, thus, making them slightly different from the unmodified ones used for G4-FID.

The tables below summarise the sequences utilised in both FRET-melting and G4-FID experiments with the buffer utilised and associated G4 melting temperature ( $T_{1/2}$ ) measured by FRET-melting, and the thiazole orange/G4 association model and TO association constant ( $K_a$ ) for each sequence for G4-FID.<sup>161,215</sup>

Name	F/T derivative	Topology	Sequence	Buffer	$T_{1/2}$	Ref
<b>DNA G-quadruplexes</b>						
22AG	<b>F21T</b>	Hybrid	5'-AGGGTTAGGGTTAGGGTTAGGG-3'T	K10	57 °C	63
22CTA	<b>F21CTAT</b>	Antiparallel	5'-GGGCTAGGGCTAGGGCTAGGG-3'T	K10	59 °C	69
cMYC22	<b>FMyc22T</b>	Parallel	5'-TGAGGGTGGGTAGGGTGGGTAA-3'T	K1	57 °C	216
CEB25wt	<b>FCEB25wtT</b>	Parallel	5'-AAGGGTGGGTGTAAGTGTGGGTGGGT-3'T	K1	57 °C	218
VEGF	<b>FVEGFT</b>	Parallel	5'-CGGGCGGGCCTTGGGCGGGGT-3'T	K1	66 °C	220
HIV-Pro1	<b>FHIVPRO1T</b>	Antiparallel	5'-TGGCCTGGGCGGGACTGGG-3'T	K1	60 °C	68
<b>RNA G-quadruplexes</b>						
<b>NRAS</b>		Parallel	5'-GGGAGGGGCGGGUCUGGG-3'	K1	n.m.	222
Name	Topology	Sequence	Model	TO/G4 log $K_a$ ( $M^{-1}$ )	Ref	
<b>22AG</b>	Hybrid	5'-AGGGTTAGGGTTAGGGTTAGGG-3'	1:1	6.1	161	
<b>22CTA</b>	Antiparallel	5'-GGGCTAGGGCTAGGGCTAGGG-3'	1:1	6.2*	161	
<b>cMYC22</b>	Parallel	5'-TGAGGGTGGGTAGGGTGGGTAA-3'	1:1	6.7	161	
<b>CEB25wt</b>	Parallel	5'-AAGGGTGGGTGTAAGTGTGGGTGGGT-3'	1:1	5.9	215	
<b>VEGF</b>	Parallel	5'-CGGGCGGGCCTTGGGCGGGGT-3'	1:1	6.2	**	

Table 1: G4 sequences and important parameters in top) FRET-melting and bottom) G4-FID. The  $T_{1/2}$  values are the ones measured in FRET-melting experiment. \* Log  $K_a$  for 21CTA. \*\* Log  $K_a$  for VEGF was determined during this PhD project. n.m. : not measured.

### 2.1.1. G4 DNA sequences and related topologies

The first sequence utilised was the human telomeric sequence that features tandem repeats of TTAGGG. **22AG** is a 22-mer sequence folding into a hybrid 3+1 in potassium buffer<sup>63</sup> and antiparallel in sodium buffer (Figure 75).<sup>64</sup> However, it is known that one of the PDC properties is to convert hybrid G4 into antiparallel form upon binding, even in potassium buffer.<sup>60</sup> In the same way, the photoactivatable G4 compounds convert the 22AG hybrid into its antiparallel form, as demonstrated in Dr. Masson's thesis using CD.<sup>211</sup>



**CEB25*wt*** is a 26-mer sequence found in the human subtelomeric regions called minisatellites, which are known to be genetically unstable and involved in cancer onset.<sup>59,217</sup> This sequence adopts a parallel topology and is characterised by a 9-nt long central loop (Figure 78). It is also described as L191, indicating the number of nucleotides in the loops, i.e. composed of two short propeller loops and one central 9-nt long loop.<sup>218</sup>

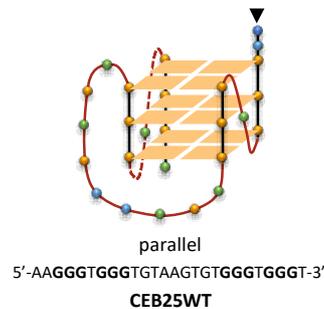


Figure 78: Representation of the parallel G4 structure formed by the CEB25*wt* sequence characterised by a 9-nt long central loop.<sup>218</sup>

The vascular endothelial growth factor (VEGF) is a signal protein stimulating the formation of blood vessels in both vasculogenesis and angiogenesis.<sup>219</sup> It is a common factor overexpressed in cancer cells to increase blood supply for tumour growth. VEGF's proximal promotor region contains five G-repeats that can fold into highly polymorphic G-quadruplexes. **VEGF** Pu22-T12T13 sequence utilised in this study forms a highly stable parallel G4 structure (Figure 79) with a melting temperature of about 66 °C.<sup>220</sup>

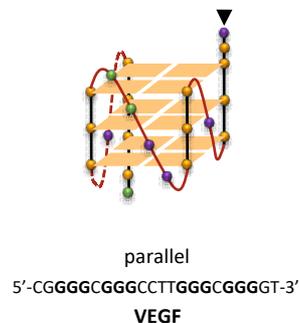


Figure 79: Representation of the highly stable parallel G-quadruplex formed by the VEGF sequence.<sup>220</sup>

The human immunodeficiency virus 1 (HIV-1) has a ssRNA genome characteristic of the retrovirus family to which it belongs. The genomic RNA is reverse transcribed into dsDNA and integrated into the host genome. This integration is helped by the long terminal repeats (LTRs) located at each extremity of the virus genome. These LTR regions contain a G-rich region featuring three superimposed G-quadruplexes: LTR-II, LTR-III, and LTR-IV. As described in the introduction, the 28-mer LTR-III sequence forms a particular 3+1 hybrid G4 comprising a hairpin.<sup>69</sup>

To avoid having two structures (G4 and hairpin) combined in one, the LTR-III sequence was truncated to remove the hairpin part. **HIV-Pro1** is an antiparallel G-quadruplex (Figure 80) with only two G-quartets and one Watson-Crick G-C base pair.<sup>68</sup>

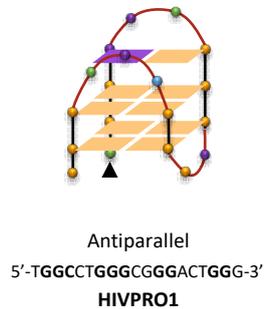


Figure 80: Representation of the HIV-Pro1 antiparallel structure comprising an apical GC base pair.<sup>68</sup>

### 2.1.2. RNA sequence and related topology

**NRAS** is another sequence utilised in this study. This sequence was found at the 5'UTR region of the NRAS transcript and its responsible of the regulation of its translation.<sup>221</sup> NRAS belong to the RAS family, which are translated into GTPase. NRAS is often mutated in human cancer and plays an important role in cell growth and survival. **NRAS** is an 18-mer sequence folding into a parallel RNA G-quadruplex (Figure 81).<sup>222</sup>

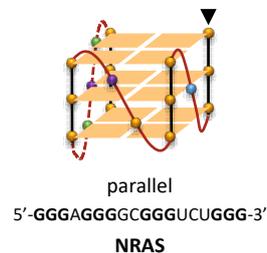


Figure 81: Representation of the parallel RNA G-quadruplex formed by the **NRAS** sequence.<sup>222</sup>

## 2.2. FRET-melting and G4-FID assays

### 2.2.1. FRET-melting protocol

FRET-melting was first conducted using a “double dilution” protocol that the team had already utilised and represented in Figure 82-A. It consists of making the final solution directly inside the well by adding a DNA solution containing the G-quadruplex and the competitors in the appropriate buffer. Then, the pre-diluted-in-buffer ligand solution (200  $\mu\text{M}$  in DMSO  $\rightarrow$  8.33  $\mu\text{M}$  in buffer) was added to make the final solution at the right concentration (8.33  $\mu\text{M}$   $\rightarrow$  1  $\mu\text{M}$  in buffer) inside the well. However, the results using this protocol with 22AG showed unreliable data (Figure 82-Right, in red) with a lack of reproducibility and poor stabilisation for some of the photoactivatable G4 ligands. We hypothesised that precipitation of the compounds occurs between the first and second dilution. To avoid this, we used the “simple dilution” protocol where the solution of each well is prepared in an Eppendorf and the photoactivatable G4 ligands are diluted only once from DMSO to buffer in the final solution (100  $\mu\text{M}$  in DMSO  $\rightarrow$  1  $\mu\text{M}$  in buffer). This overcomes the problem giving reproducible data. (Figure 82-Right, in blue). These results indicate a problem of solubility for some of the branched compounds, which must be kept in mind.

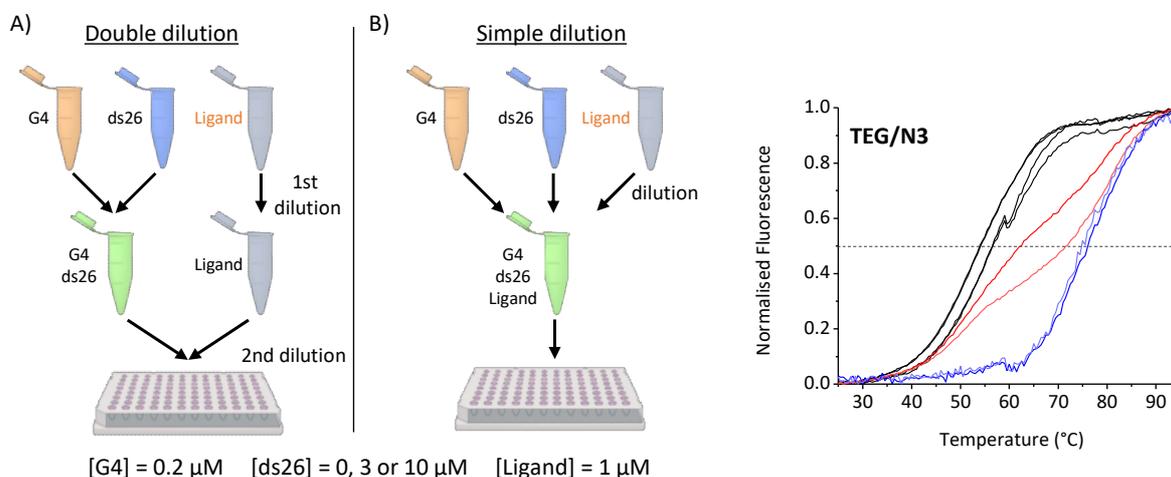


Figure 82: Two protocols have been tested for the FRET-melting experiments: A) the double dilution protocol with a first pre-dilution of the ligand from DMSO (written in orange) into buffer before a second dilution directly into the well. B) The simple dilution protocol was used to overcome this problem by preparing the final solution in an Eppendorf by diluting the ligand only one time.

## 2.2.2. G4-FID protocol

G4-FID method has been first developed in cuvette but has been rapidly adapted to microplate for high-throughput application.<sup>158,161</sup> However, the microplate preparation is similar to the FRET-melting “double dilution” protocol, which could lead to precipitation in our specific case. This is why a panel of photoactivatable G4 ligands was utilised with 22AG to evaluate the results obtained by G4-FID both in cuvette and microplate, the comparison among the two helped us to choose the best methodology. The panel of compounds was selected to cover a range of F21T stabilisation measured by FRET-melting, which will be discussed later, but to summarise: **C4-I-N2** and **C3/N2** are highly stabilising compounds ( $\Delta T_{1/2} \approx 22$  °C and 23 °C respectively), **C1/N2** is a modest stabilising compound ( $\Delta T_{1/2} \approx 10$  °C), and **PDC** used as reference ( $\Delta T_{1/2} \approx 28$  °C). (For the names of the compounds refer to Figure 69, p. 90)

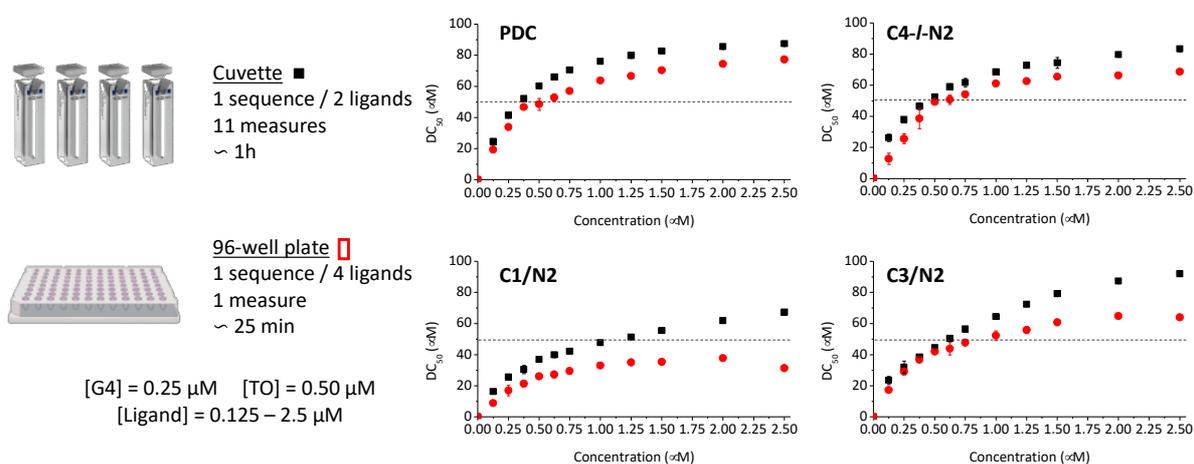


Figure 83: G4-FID experiment can be realised in cuvette (■) or in 96-well microplate (●). Right) The results obtained with both methods in the presence of PDC, C4-I-N2, C1/N2, and C3/N2 evaluated with 22AG.

Strong stabilising compounds such as PDC, C4-I-N2, and C3/N2 present similar displacement curves (Figure 83) with the two methods, with slight increase  $DC_{50}$  values for microplates, as previously reported (0.1-2  $\mu$ M) (Table 2). For the modest stabilising compound C1/N2, the  $DC_{50}$  is not measurable in microplate since the data points do not reach 50% TO displacement as compared to the cuvette measurement. Thus, the microplate protocol is not compatible for our specific case as some photoactivatable G4 ligands might not reach 50% TO displacement preventing to obtain any  $DC_{50}$  values for those compounds. Therefore, all the further results obtained by G4-FID have been realised in cuvettes.

		PDC	C4-I-N2	C1/N3	C3/N2
FRET-melting	$\Delta T_{1/2}$ (°C)	28	22	10	23
Cuvette	$DC_{50}$ ( $\mu$ M)	0.35	0.44	1.12	0.61
96-well plate		0.51	0.54	n.d.	0.77

Table 2:  $DC_{50}$  values obtained in cuvette and 96-well microplate, and  $\Delta T_{1/2}$  values reported as a reference.  $DC_{50}$  value for **C1/N3** was not determinable (n.d.) in 96-well microplate.

### 2.2.3. TO binding to VEGF G-quadruplex

G4-FID is a competitive assay where a G4 ligand displaces a fluorescent probe bound to a G4 structure, as described in the introduction (Chapter 1.3.3.2, p. 65). Thus, it is essential to characterise the interaction of the probe with the G4 structure to better understand the competition process. The stoichiometry of the TO/VEGF complex was determined using the Job's method of continuous variations: TO fluorescence was measured for a range of VEGF molar fractions without modifying the total molar concentration of the two binding partners (1  $\mu$ M). The Job plot (Figure 84-A) presents a maximum between 0.5 and 0.6 indicating a predominant 1:1 stoichiometry. Furthermore, the affinity constant of the probe ( $K_a$ ) was determined by fitting the fluorometric titration curve of TO binding to VEGF using a 1:1 binding model (Figure 84-B), which gave a  $K_a$  value of  $1.68 \cdot 10^6 \text{ M}^{-1}$  ( $\log K_a \approx 6.2$ ). These two results are consistent with those previously found ( $\log K_a$  TO/G4 between 5.9 and 6.7) for the other sequences of the panel (Table 1).<sup>161,215</sup>

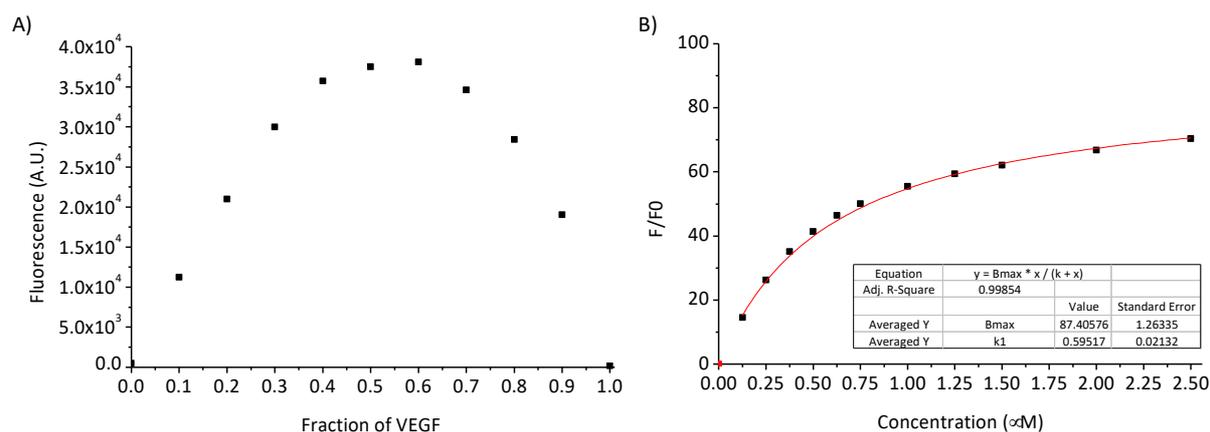


Figure 84: A) Job plot of VEGF/TO with a maximum between 0.5 - 0.6, indicating a 1:1 binding stoichiometry. B) Fluorescent titration of TO by increasing concentration of VEGF.

### 2.3. Compounds evaluation

We performed FRET-melting and G4-FID assays to evaluate the propensity of the PDC core to still interact and stabilise G4 structures selectively once embedded into the **in-line** and **branched** architectures. In FRET-melting, stabilisation values are comparable uniquely using the same sequence. For each ligand, three values are shown corresponding to no-competition (in black), and with 3  $\mu\text{M}$  (red) and 10  $\mu\text{M}$  (blue) of ds26 competitor. In addition, G4-FID gives complementary information on the interaction through the  $\text{DC}_{50}$  value (in black), when FRET-melting informs on the stabilisation induced by the ligand.

#### 2.3.1.22AG/F21T

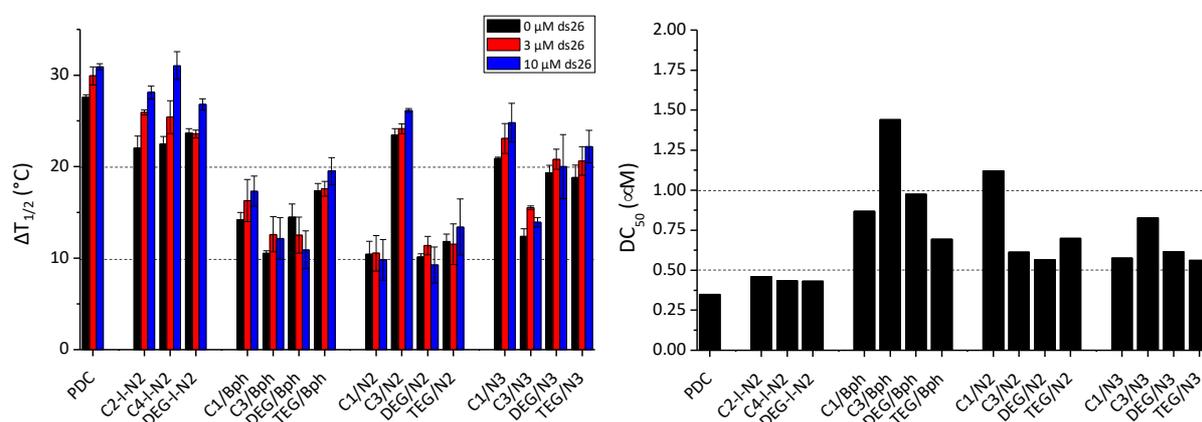


Figure 85: Left)  $\Delta T_{1/2}$  values ( $^{\circ}\text{C}$ ) obtained by the FRET-melting with F21T in K10-buffer plotted for each compound without (black) or with 3  $\mu\text{M}$  (red) or 10  $\mu\text{M}$  (blue) of ds26 competitors. Right)  $\text{DC}_{50}$  ( $\mu\text{M}$ ) values obtained by G4-FID experiment with 22AG plotted for each compound.

Globally, from the results of FRET-melting shown in Figure 85-Left, the compound series can be classified into two groups, with the **in-line** and **N3** compounds presenting a high stabilisation of F21T ( $\Delta T_{1/2} \geq 20$   $^{\circ}\text{C}$ ), and the **Bph** and **N2** compounds presenting a more modest stabilisation ( $\Delta T_{1/2} \approx 10$   $^{\circ}\text{C}$ ). Similarly, G4-FID data (Figure 85-Right) show the **in-line** compounds displacing TO with  $\text{DC}_{50} < 0.50$   $\mu\text{M}$ , the **N2** and **N3** compounds displacing TO with a medium efficiency ( $\text{DC}_{50} \approx 0.60$   $\mu\text{M}$ ), and the **Bph** derivatives displacing TO with a low efficiency ( $\text{DC}_{50} > 0.80$   $\mu\text{M}$ ).

Five compounds are clearly above the others, with first all the **in-line** derivatives presenting high stabilisation ( $\Delta T_{1/2} \approx 22\text{-}24\text{ }^{\circ}\text{C}$ ) slightly lower than the one of PDC ( $\Delta T_{1/2} \approx 28\text{ }^{\circ}\text{C}$ ) and efficient TO displacement ( $DC_{50} < 0.50\text{ }\mu\text{M}$ ). The **C3/N2** derivative is remarkably above in terms of stabilisation ( $\Delta T_{1/2} \approx 23\text{ }^{\circ}\text{C}$ ) compared to the rest of the N2 series ( $\Delta T_{1/2} \approx 10\text{ }^{\circ}\text{C}$ ). The **C1/N3** derivative is slightly better to stabilise F21T ( $\Delta T_{1/2} \approx 21\text{ }^{\circ}\text{C}$ ) than the rest of the **N3** compounds series ( $\Delta T_{1/2} \approx 12\text{-}19\text{ }^{\circ}\text{C}$ ).

The two compounds **C3/Bph** and **C1/N2** seem to present weaker interaction by G4-FID compared to the other members of their series. **C3/Bph** presents a weak interaction displaying a  $DC_{50}$  value of  $1.44\text{ }\mu\text{M}$ , while the rest of Bph derivatives displayed  $DC_{50}$  values  $< 1.00\text{ }\mu\text{M}$ . Similarly, **C1/N2** shows a weak interaction with a  $DC_{50}$  value of  $1.12\text{ }\mu\text{M}$  vs.  $0.57\text{-}0.70\text{ }\mu\text{M}$  for the rest of the N2 series.

### 2.3.2.22CTA/F21CTAT

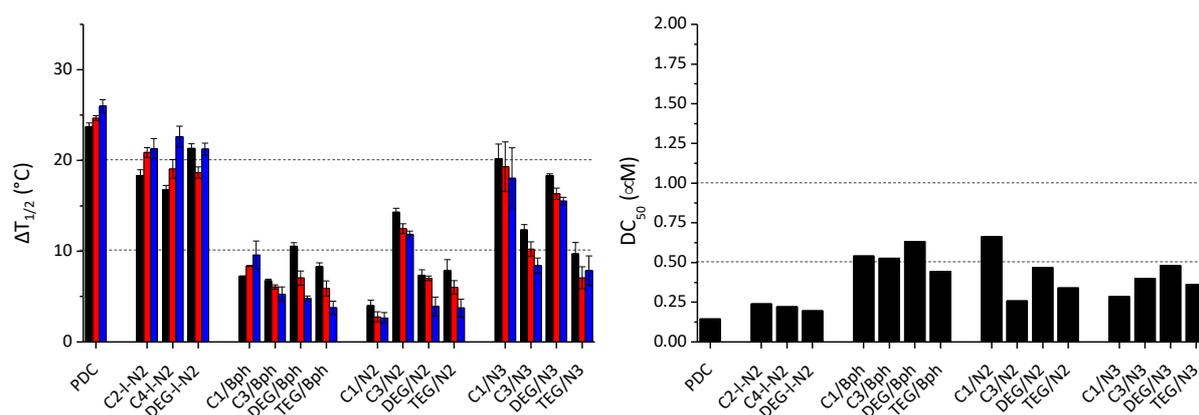


Figure 86: Left)  $\Delta T_{1/2}$  values ( $^{\circ}\text{C}$ ) obtained by FRET-melting with F21CTAT in K10-buffer plotted for each compound without (black) or with  $3\text{ }\mu\text{M}$  (red) or  $10\text{ }\mu\text{M}$  (blue) of ds26 competitors. Right)  $DC_{50}$  ( $\mu\text{M}$ ) values obtained by G4-FID experiment with 22CTA plotted for each compound.

The different series of photoactivable G4 ligands stabilise 22CTA with various efficiency (Figure 86). The interaction of the **in-line** compounds remains excellent with high stabilisation ( $\Delta T_{1/2} \approx 17\text{-}21\text{ }^{\circ}\text{C}$ ) and efficient TO displacement ( $DC_{50} \approx 0.20\text{-}0.24\text{ }\mu\text{M}$ ) close to that of PDC ( $\Delta T_{1/2} \approx 24\text{ }^{\circ}\text{C}$  and  $DC_{50} \approx 0.15\text{ }\mu\text{M}$ ). In addition, three compounds of the branched series stand out. **C1/N3** and **DEG/N3** show performances (respectively  $\Delta T_{1/2} \approx 20\text{ }^{\circ}\text{C}$  and  $18\text{ }^{\circ}\text{C}$ ,  $DC_{50} \approx 0.29\text{ }\mu\text{M}$  and  $0.48\text{ }\mu\text{M}$ ) slightly below the **in-line** derivatives.

The rest of the **branched** compounds stabilises more weakly 22CTA ( $\Delta T_{1/2} \leq 10$  °C) but with  $DC_{50}$  values around 0.50  $\mu\text{M}$ , except for **C3/N2**, which presents a modest stabilisation ( $\Delta T_{1/2} \approx 14$  °C) and an efficient TO displacement ( $DC_{50} \approx 0.26$   $\mu\text{M}$ ). Moreover, most of the **branched** compounds show a lower G4-selectivity, as shown by the slight decrease of  $\Delta T_{1/2}$  value in the presence of the duplex competitor ds26. Finally, the **C1/N2** derivative is the weakest of the series regarding stabilisation ( $\Delta T_{1/2} \approx 4$  °C). However, the TO displacement shows a medium efficiency ( $DC_{50} \approx 0.66$   $\mu\text{M}$ ) that was not expected for such a weak stabilisation value. All compounds showed lower  $DC_{50}$  value with 22CTA as compared to 22AG (e.g.  $DC_{50} \approx 0.15$  vs. 0.35  $\mu\text{M}$  respectively for PDC).

### 2.3.3.cMYC22/FcMYC22T

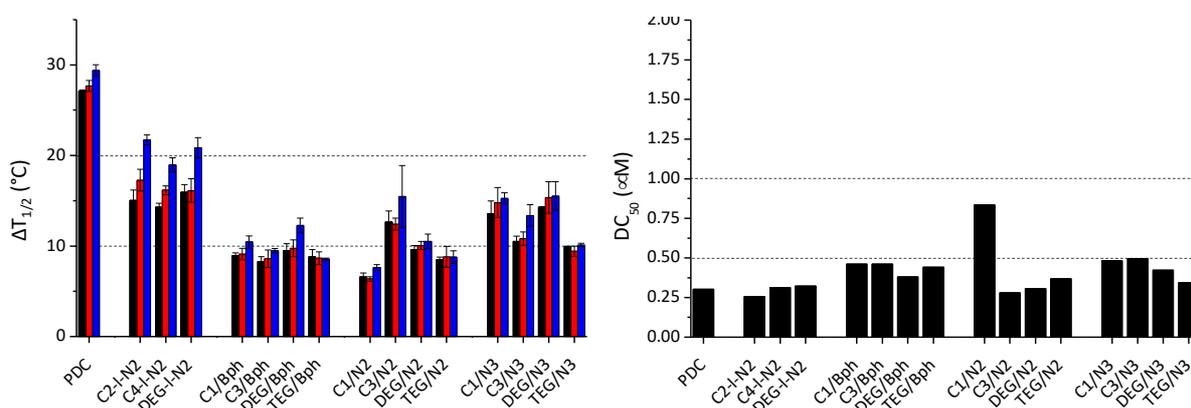


Figure 87: Left)  $\Delta T_{1/2}$  values (°C) obtained by FRET-melting with FcMYC22T in K1-buffer plotted for each compound without (black) or with 3  $\mu\text{M}$  (red) or 10  $\mu\text{M}$  (blue) of ds26 competitors. Right)  $DC_{50}$  ( $\mu\text{M}$ ) values obtained by G4-FID experiment with cMYC22 plotted for each compound.

The interaction of the compounds with cMYC22 presents globally a similar trend: the **in-line** derivatives are the best stabilisers ( $\Delta T_{1/2} \approx 14$ -16 °C) followed by the **N3** derivatives ( $\Delta T_{1/2} \approx 10$ -14 °C) while the **Bph** and **N2** series are weaker ( $\Delta T_{1/2} \approx 8$ -9 °C and 7-13 °C respectively) (Figure 87). However, the **in-line** functionalisation seems to have a more negative impact on the **PDC** core stabilisation ( $\Delta T_{1/2} \approx 27$  °C) with cMYC22 compared to previous data obtained with 22AG and 22CTA. Indeed, the  $\Delta T_{1/2}$  value of the three in-line compounds drops down of approx. 10 °C as compared to that of **PDC** when it is only half this value with 22AG and 22CTA.

Again, the **C3/N2**, **C1/N3**, and **DEG/N3** derivatives present a higher stabilisation than the other members of their respective series, which appears to be a constant feature. Globally, the  $DC_{50}$  data are consistent with  $\Delta T_{1/2}$  data and confirm the main trend, with the most active compounds, i.e. the **in-line** have  $DC_{50} \approx 0.26-0.32 \mu\text{M}$ , whereas all the others show  $DC_{50}$  between 0.31 to 0.50  $\mu\text{M}$ . As well, the high  $DC_{50}$  of **C1/N2** consistent with its low stabilising effect ( $DC_{50} \approx 0.83 \mu\text{M}$ ) confirms the weaker G4 interaction of this compound.

### 2.3.4. CEB25 wt/FCEB25 wt

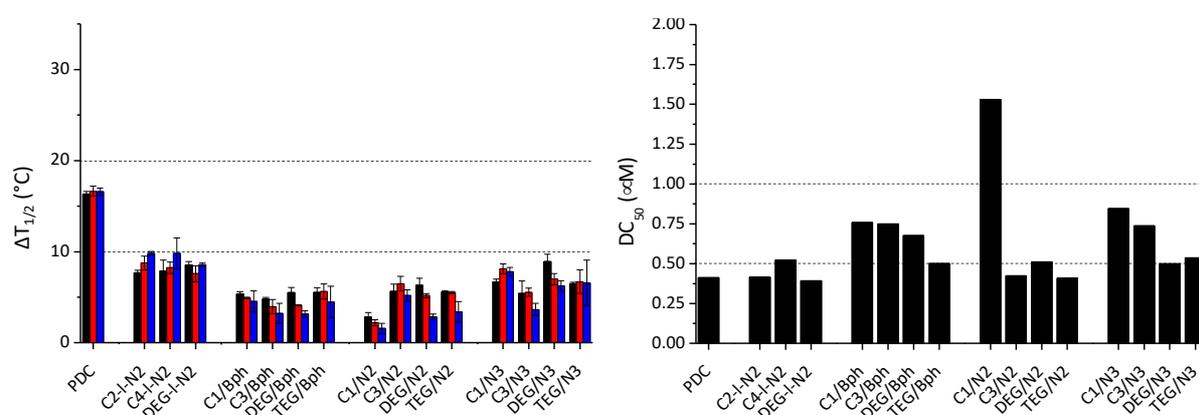


Figure 88: Left)  $\Delta T_{1/2}$  values (°C) obtained by FRET-melting with FCEB25wtT in K1-buffer plotted for each compound without (black) or with 3  $\mu\text{M}$  (red) or 10  $\mu\text{M}$  (blue) of ds26 competitors. Right)  $DC_{50}$  ( $\mu\text{M}$ ) values obtained by G4-FID experiment with CEB25wt plotted for each compound.

The results with CEB25 wt (Figure 88) present a similar trend for the stabilisation induced by the compounds: **In-line** ( $\Delta T_{1/2} \approx 8-9 \text{ }^\circ\text{C}$ ) > **N3** ( $\Delta T_{1/2} \approx 5-9 \text{ }^\circ\text{C}$ ) > **Bph** ( $\Delta T_{1/2} \approx 5-6 \text{ }^\circ\text{C}$ ) and **N2** ( $\Delta T_{1/2} \approx 3-6 \text{ }^\circ\text{C}$ ). Similarly to the behaviour observed with cMYC22, the  $\Delta T_{1/2}$  of the **in-line** series drop-down of approx. 8  $^\circ\text{C}$  as compared to PDC ( $\Delta T_{1/2} \approx 16 \text{ }^\circ\text{C}$ ). Notably, the  $\Delta T_{1/2}$  value of PDC is significantly lower as compared to the other G4 tested, whilst the  $T_{1/2}$  (57  $^\circ\text{C}$ ) is in the same range as F21T or F21CTAT (57-59  $^\circ\text{C}$ ), meaning that globally this G4 matrix accommodates less easily G4 ligands. The less efficient binding of compounds to CEB25 wt is confirmed by the higher range of  $DC_{50}$  values measured for PDC and the **in-line** series between 0.39 and 0.52  $\mu\text{M}$ . Consistently, the **Bph** and **N3** series exhibit higher  $DC_{50}$  between 0.50 and 0.85  $\mu\text{M}$ , whereas the **N2** series (without **C1/N2**) appeared more active with  $DC_{50}$  comparable to the **in-line** compounds ( $DC_{50} \approx 0.41-0.51 \mu\text{M}$ ). Finally, the **C1/N2** derivative confirms its outsider position with weak stabilisation ( $\Delta T_{1/2} \approx 3 \text{ }^\circ\text{C}$ ) combined with a  $DC_{50}$  off the charts > 1.50  $\mu\text{M}$ .

### 2.3.5. VEGF/FVEGFT

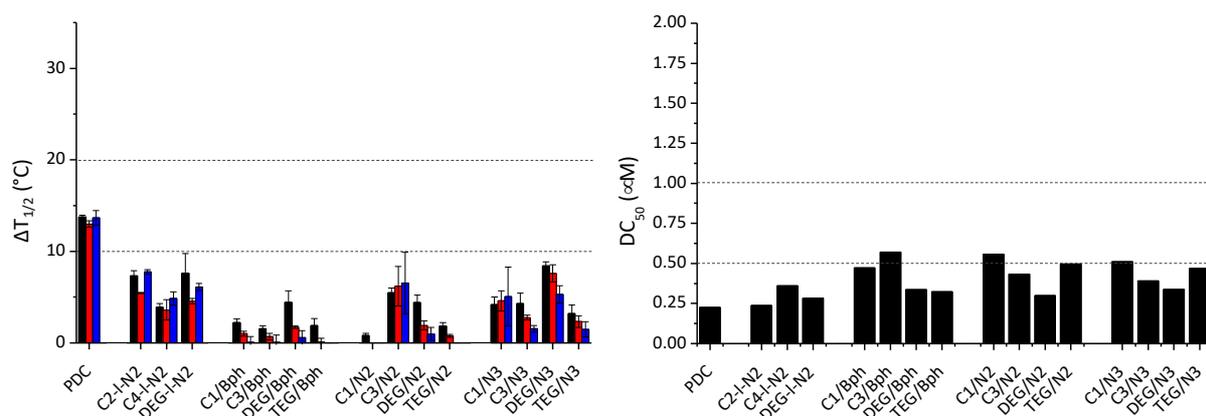


Figure 89: Left)  $\Delta T_{1/2}$  values ( $^{\circ}\text{C}$ ) obtained by FRET-melting with FVEGFT in K1-buffer plotted for each compound without (black) or with 3  $\mu\text{M}$  (red) or 10  $\mu\text{M}$  (blue) of ds26 competitors. Right)  $DC_{50}$  ( $\mu\text{M}$ ) values obtained by G4-FID experiment with VEGF plotted for each compound.

VEGF is a highly stable G-quadruplex with a melting temperature of 66  $^{\circ}\text{C}$  in 1 mM  $\text{K}^{+}$  buffer. Thus, lower  $\Delta T_{1/2}$  values are expected upon G4 ligands binding. This is clearly illustrated by the moderate increase of melting temperature induced by the reference compound PDC ( $\Delta T_{1/2} \approx 14$   $^{\circ}\text{C}$ ), whereas its  $DC_{50}$  is maintained at a low level (0.23  $\mu\text{M}$ ), revealing strong binding (Figure 89). In this context, the **in-line** compounds and two branched derivatives (C3/N2 and DEG/N3) still induce significant stabilisation ( $\Delta T_{1/2} \approx 4\text{-}8$   $^{\circ}\text{C}$  for the **in-line** series and  $\Delta T_{1/2} \approx 5\text{-}8$   $^{\circ}\text{C}$  for C3/N2 and DEG/N3 derivatives) but again reduced of approx. 7  $^{\circ}\text{C}$  as compared to PDC.

Very poor stabilisation is obtained for all the other branched derivatives with  $\Delta T_{1/2}$  values lying  $\leq 5$   $^{\circ}\text{C}$  or barely measurable (e.g. C1/N2). In these conditions, only the **in-line**, C3/N2 and C1/N3 are not strongly affected by the ds26 competitor, whilst all the other branched compounds do not resist the competition. On the contrary, the G4-FID results indicate low  $DC_{50}$  ( $DC_{50} \approx 0.24\text{-}0.36$   $\mu\text{M}$ ) for the in-line compounds similar to PDC ( $DC_{50} \approx 0.23$   $\mu\text{M}$ ), and globally, between 0.30 up to 0.60  $\mu\text{M}$  for all the other derivatives. This indicates that ligands may have strong interaction while inducing poor stabilisation. The strong discrepancy between the two assays points to the limit of FRET-melting and the lower significance of the data when the G4 target has a high thermal stability. This underlines the importance of using systematically a second method like G4-FID to collect more reliable information and have a global picture.

### 2.3.6.FHIVPRO1T

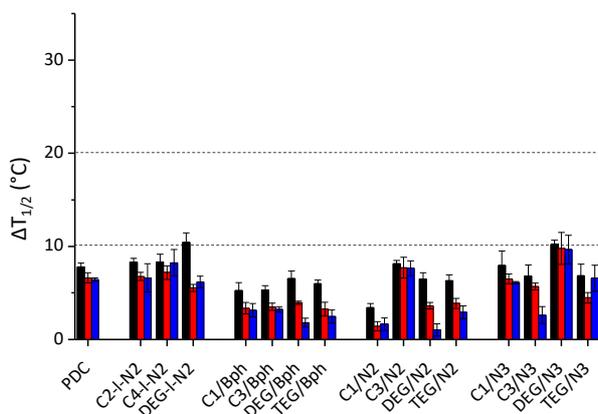


Figure 90:  $\Delta T_{1/2}$  values ( $^{\circ}\text{C}$ ) obtained by FRET-melting with FHIVPRO1T in K1-buffer plotted for each compound without (black) or with 3  $\mu\text{M}$  (red) or 10  $\mu\text{M}$  (blue) of ds26 competitors.

HIV-Pro1 forms a particularly stable ( $T_{1/2} \approx 60$   $^{\circ}\text{C}$  in 1 mM  $\text{K}^+$  condition) antiparallel structure with an apical G-C base pair, which is likely to impact the interaction and hence the stabilisation by G4 ligand, for example, the PDC stabilisation is the lowest from the G4 panel with  $\Delta T_{1/2} \approx 8$   $^{\circ}\text{C}$ . Same trend is observed as with the previous sequences (Figure 90): in-line series, C3/N2 and DEG/N3 derivatives have a stabilising effect comparable to that of PDC and resist pretty well to the competition. All the others (Bph, N2, N3 series) derivatives show slightly weaker stabilisation, which is drastically reduced by the presence of the competitor. Finally, C1/N2 derivative continues to display poor stabilisation ( $\Delta T_{1/2} \approx 3$   $^{\circ}\text{C}$ ).

## 2.4. RNA G-quadruplex in FRET-melting

To study G4 ligand binding to RNA a FRET-melting competition methodology<sup>214</sup> was utilised. The binding of the G4 ligands to DNA G-quadruplex F21T was challenged by 3  $\mu$ M of unlabelled RNA G-quadruplex NRAS used as competitor.

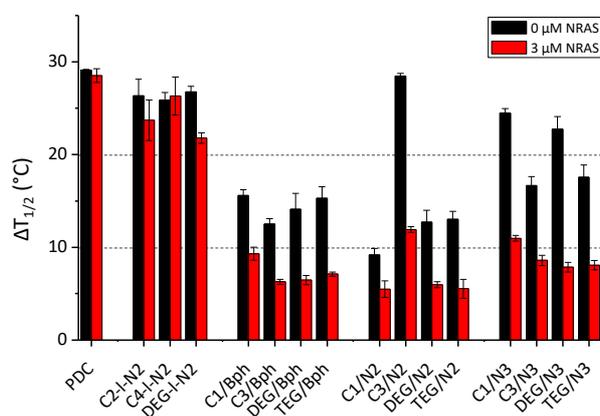


Figure 91:  $\Delta T_{1/2}$  values ( $^{\circ}$ C) obtained by FRET-melting competition experiment using F21T in K10-buffer plotted for each compound without (black) or with 3  $\mu$ M (red) of RNA NRAS competitors.

Two cases are visible from the results (Figure 91): all the **branched** compounds are displaced by NRAS sequence from their G4-DNA target with drastic decrease of  $\Delta T_{1/2}$  (e.g. **C3/N2**, the most striking example, displaying a decrease from 28  $^{\circ}$ C to 12  $^{\circ}$ C). This fully demonstrates the branched compounds' propensity to interact with RNA G-quadruplex. Surprisingly, poor to no competition between NRAS and F21T is observed with the **in-line** compounds. This indicates that F21T sequester too efficiently a significant fraction of the in-line derivatives, indicating a higher affinity for F21T vs. NRAS. To observe a significant variation of the  $T_{1/2}$  values for high affinity compounds such as the in-line derivatives a much lower G4 ligand:G4 ratio should have been used as previously reported.<sup>214</sup>

## 2.5. Conclusion

Each of the different families of photoactivatable G4 ligands stabilises DNA G-quadruplexes differently. Globally, the trend order shows an excellent interaction of the **in-line** and then the **N3** compounds, and a more modest interaction for the **Bph** and **N2** compounds. The **branched** compounds are more impacted by the presence of a duplex DNA competitor when stabilising 22CTA, CEB25 *wt*, VEGF and HIV-pro1. However, there is almost no competition with 22AG and cMYC22, which could indicate a slight preference of the **branched** compounds for some particular G-quadruplexes. All the **in-line** compounds present very similar interactions inside the series, and behaviour close to PDC, with no effect of the ds26 competitor on the stabilisation. The **N3** series comprises two compounds that stand out with **C1/N3** and **DEG/N3**, which reach stabilisation similar to the ones displayed by the **in-line** derivatives, but more affected by the presence of a duplex DNA competitor: effect highly depend on the nature of the G4 sequence. The **N2** series comprises the two interaction ends. **C3/N2** is the compound showing stabilisation properties very close to the **in-line** derivatives, with binding almost not affected by the presence of a competitor and no preference for a specific G4 structure. The same series contains the outsider of all the compounds that is **C1/N2**. It presents the poorest stabilisation over all the sequences and a  $DC_{50}$  always over 0.50  $\mu\text{M}$  and even over 1.50  $\mu\text{M}$  with CEB25 *wt*. The **Bph** series is interesting for its constancy with almost identical stabilisation and  $DC_{50}$  values for all the compounds. There is no impact of the length or type of the terminal alkyne spacer on the stabilisation by the **Bph** series. This remark can be extended to all the **branched** compounds, which do not show a rational impact of the spacer length or type on the stabilisation. Indeed, the three stand-out **branched** compounds comprise three of the four spacers (C1, C3, and DEG). However, even if these results cannot be rationalised, the length and type are still relevant, directly impacting some combination of photoactivatable G4 ligands. The latter can interact with DNA G4 and RNA G4 without selectivity.



## Chapter III:

### Photochemical reactivity with G-quadruplex structures



The second functionality bore by the photoactivatable G4 ligands is the photoactivatable moiety, which has for purpose the covalent capture of G4 structures. This section presents the ability of these new families of photoactivatable compounds to generate covalent adducts with a variety of G-quadruplexes upon UV activation. To validate the alkylation properties of these compounds, the G4 ligand is at first equilibrated with the G4 to form a G4 ligand/G4 complex. Then, UV irradiation activates the photoactivatable moiety to generate a highly reactive species, which rapidly reacts with its close environment. At first, we have studied the efficiency of these photoactivatable G4 compounds to trap different G-quadruplexes. Afterwards, different sequencing methods were applied to determine the exact G4 alkylation sites.

### 1. G4 Alkylation efficiency<sup>a</sup>

The alkylation efficiency of the compounds was studied by Dr. T. Masson on radiolabelled 22AG, cMYC22 and CEB25*wt* by denaturing gel electrophoresis. This technique allows the separation of DNA fragments in denaturing conditions as a function of their mass and charge. Each DNA fragment generates a band whose migration depends on its electrophoretic mobility associated to the total mass and charge, which is revealed by autoradiography. Technically (Figure 92), the photoactivatable G4 compound (2 eq.) was equilibrated with the folded <sup>32</sup>P-5'-labelled G4 for 30 min at room temperature, and then the solution was irradiated up to 1 h at 360 nm in a multirays photoreactor. The adducts obtained by irradiation and the unreacted G4 were separated by denaturing polyacrylamide gel electrophoresis (denaturing-PAGE).

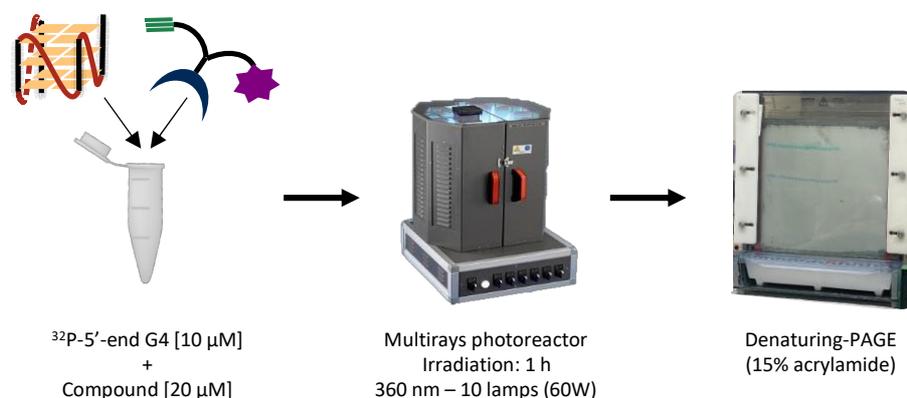


Figure 92: Representation of the alkylation efficiency experiment. The folded <sup>32</sup>P-5'-labelled G4 is mixed with one photoactivatable G4 compound in an Eppendorf tube, which is irradiated at 360 nm with 10 lamps (60 W) up to 1 h. The reaction mixture is then separated by denaturing-PAGE.

<sup>a</sup> Experiment realised by Dr. T. Masson.

The generated covalent adducts should migrate differently compared to the reference unmodified DNA due to the principle on which denaturing gel electrophoresis is based. The photoadducts have a bigger molar mass than the G4 reference sequence (about 3 to 4 dNMP) and introduce two positive charges. Both additional mass and positive charge should make the adduct bands migrate slowly as compared to the unmodified G4.

A summary of Dr. T. Masson's works is presented here as it is essential for the context and global understanding of this PhD work. Much more detailed information on alkylation experiments can be found in Dr. T. Masson's thesis.<sup>211</sup> The alkylation process is regulated by different parameters related to the interaction between the compound and the G4 structure (Figure 93). The first one is associated to the affinity ( $K_a$ ) of the G4 ligand, which was evaluated by biophysical methods and has been reported in the previous chapter. As a matter of fact, the alkylation process should be more efficient for high-affinity compounds, high  $K_a$  values increase the residence time of the compounds on the G4 structure, and as a consequence increase the probability of the alkylation event (see section I.4.3, p. 75). Therefore, **in-line** compounds should display higher alkylation yields due to their strong interaction with G4s. Differently, **N2 and N3 branched** compounds showed more substantial variability in their binding properties, as a function of the nature and length of the spacer bearing the terminal alkyne, and the outcome seemed less obvious. The second important element is associated to the spatial proximity between the photoactivatable moiety and the reactive sites of the G-quadruplex and the positioning and nature of the photoreactive intermediate. All the **branched** compounds present the same distance between the PDC core and the photoactivatable moieties. Therefore, they should present equal spatial proximity and the efficiency of the alkylation should be associated exclusively to the positioning and nature of the reactive intermediate. However, it will be interesting to observe the results obtained with the **in-line** compounds since the distance between the PDC core and the photoactivatable moiety depends on the linker (C2, C4, and DEG). Therefore, the linker may modify the accessibility of the photoactivatable moiety toward the G4 reactive sites.

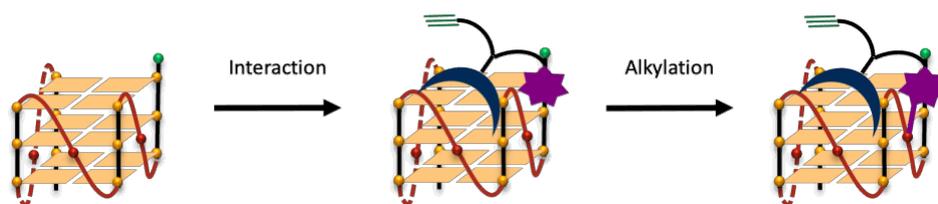


Figure 93: Representation of the molecular steps for the alkylation of the G4 structure. The first step corresponds to the mandatory non-covalent interaction of the PDC core with the external G-quartet. The second step is the formation of the covalent bond with the G4 structure through UV irradiation.

## 1.1. Photoreactivity of *in-line* compounds

In this paragraph, we compare the alkylation efficiency of the *in-line* compounds (Figure 94) in the presence of three different G4 structures after 1 h irradiation at 360 nm. In the presence of the human telomeric sequence 22AG, the alkylation adducts are presented by a band migrating slower than the reference DNA G4 and showed alkylation yield between 57 and 60%. The alkylation adducts obtained in the presence of cMYC22 protooncogene sequence are represented mainly by a faster migrating band characterised by a conversion of 39 to 48%. A similar result was achieved for the alkylation reaction performed in the presence of the minisatellite sequence CEB25 *wt*, the main adducts are represented by faster migrating bands and were obtained with a yield ranging between 64-74%. It is interesting to note that both cMYC22 and CEB25 *wt* G4s adopt a parallel topology, and their alkylated adducts are represented by faster migrating bands in gel electrophoresis experiments. However, cMYC22 shows a lower conversion as compared to CEB25 *wt*. The higher alkylation yield observed for the latter may be explained by the presence of the 9-nt long central loop, which in combination with its dynamics can expose a higher number of accessible alkylation sites as compared to cMYC22. Differently, 22AG forms adducts represented by slow migrating bands and with a conversion close to the one of CEB25 *wt*. This could be explained by the hybrid topology formed by 22AG in potassium conditions, characterised by the presence of lateral loops, which are more accessible for alkylation. Surprisingly, cMYC22 and CEB25 *wt* alkylation adducts are represented by faster migrating bands. This may be due to the formation of stable secondary structures, which remain partially folded also in denaturing conditions. These results highlighted the excellent propensity of the *in-line* compounds to alkylate all the different G4 structures used for this study.

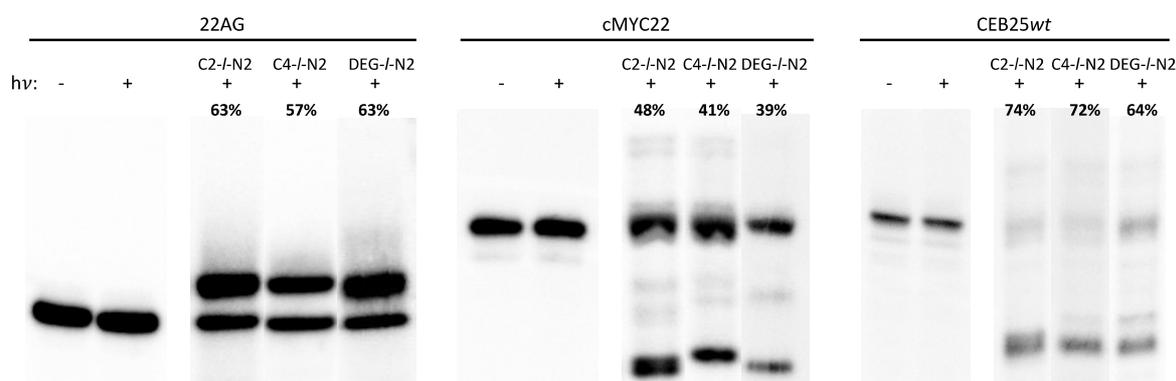


Figure 94: Denaturing gel electrophoresis (15% acrylamide) experiments of the photoalkylation reaction performed in the presence of 22AG, cMYC22 or CEB25wt (10  $\mu$ M) in  $K^+$  buffer (100 mM) with C2-I-N2, C4-I-N2, and DEG-I-N2 (20  $\mu$ M). hv: absence (-) or presence (+) of irradiation at 360 nm for 1 h.

## 1.2. Photoreactivity of the **branched** compounds

As for the **in-line** G4 ligands, **branched** compound reactivity was studied in the presence of the same G4 structures mentioned above. The photochemical reactions performed in the presence of 22AG (Figure 95) showed the formation of covalent adducts characterised by two main bands running slower than the reference DNA G4 in denaturing gel electrophoresis. Globally, the conversion is poor for the three photoactivatable series and similar for all compounds bearing the same photoactivatable moiety: benzophenone derivatives showed the highest conversion from 8 to 18%, aryl azide (N3) derivatives displayed alkylation yields included in the range between 6-9%, and benzyl diazirine (N2) derivatives showed a poor conversion (3-5%). The alkylation of cMYC22 (Figure 96) generated covalent adducts, which showed a faster migration as compared to the reference DNA G4 and displayed high conversion, between 50% to 100% for most of them except for the **C1/N2** derivative, which seems not to react. The alkylation products obtained with CEB25 *wt* are represented by migration bands, which show an electrophoretic mobility almost identical to the reference DNA G4. As the two bands are too close and partially overlapping, it was impossible to clearly determine the alkylation yield, however, it seems closer to 50%. Interestingly, the **branched** compounds showed excellent alkylation yields in the presence of cMYC22 compared to the **in-line** compounds, which showed alkylation yields up to 50%, opposite results were obtained in the presence of 22AG.

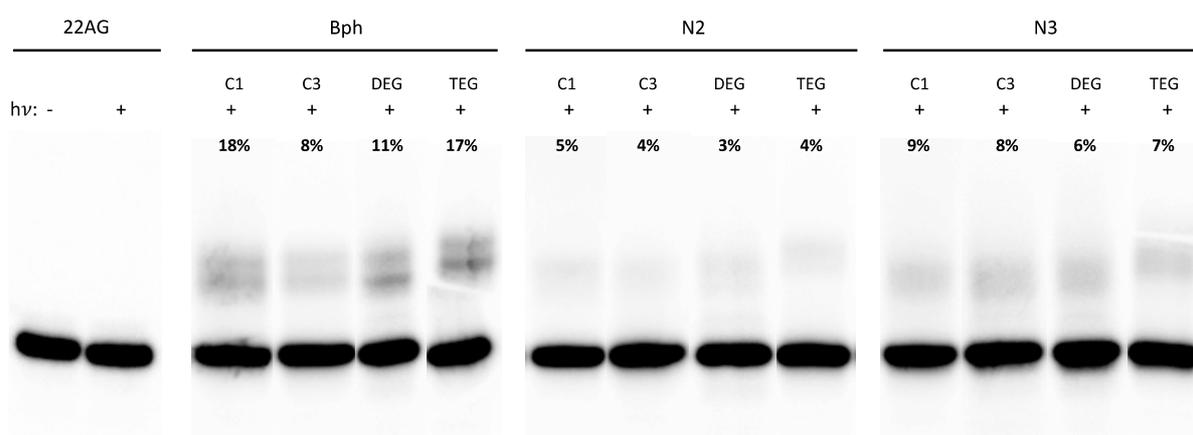


Figure 95: Denaturing gel electrophoresis (15% acrylamide) experiments of the photoalkylation reaction performed in the presence of 22AG (10  $\mu$ M) in  $K^+$  buffer (100 mM) with benzophenone derivatives: **C1/Bph**, **C3/Bph**, **DEG/Bph**, and **TEG/Bph**, benzyl diazirine derivatives: **C1/N2**, **C3/N2**, **DEG/N2**, **TEG/N2**, and aryl azide derivatives: **C1/N3**, **C3/N3**, **DEG/N3**, **TEG/N3** (20  $\mu$ M). *hv*: absence (-) or presence (+) of irradiation at 360 nm for 1 h.

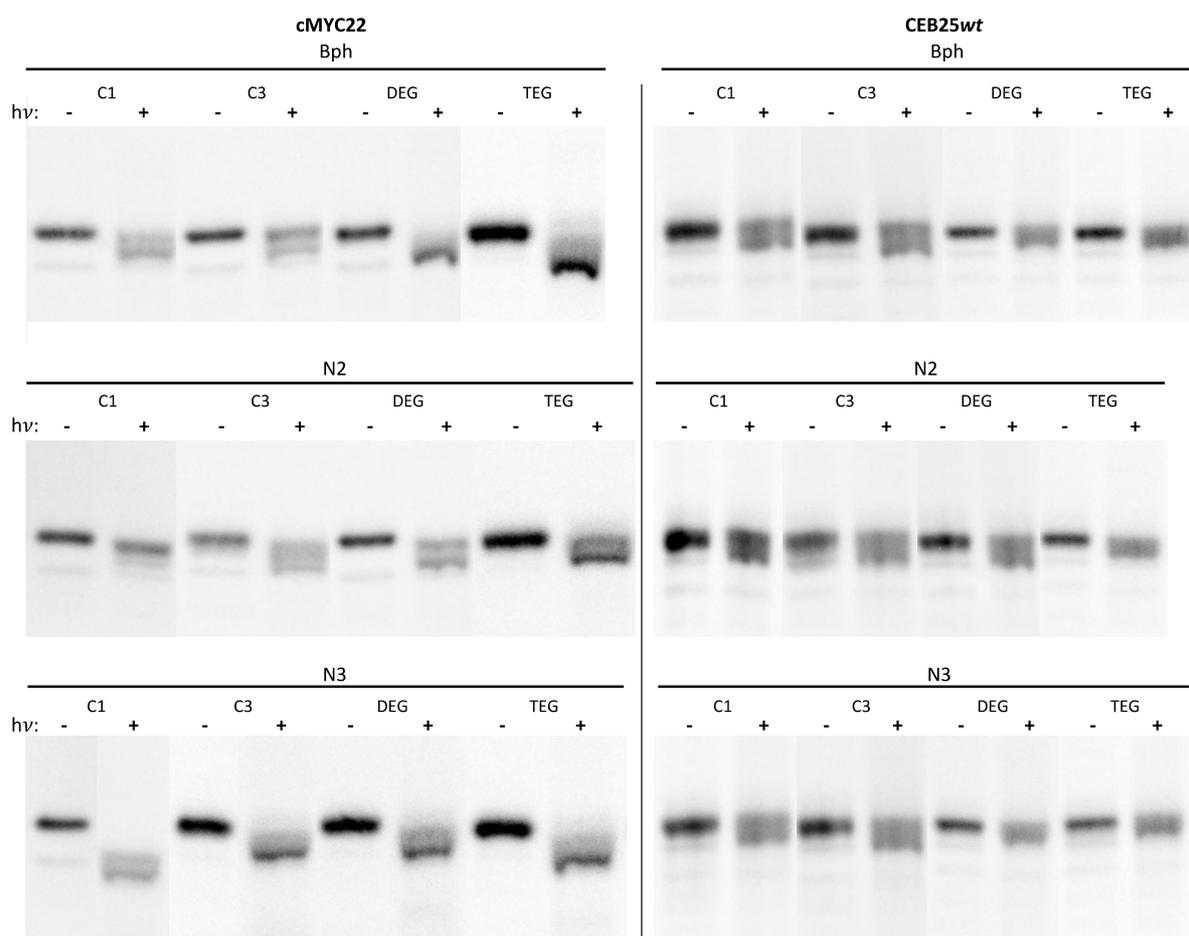


Figure 96: Denaturing gel electrophoresis (15% acrylamide) experiments of the photoalkylation reaction performed in the presence of cMYC22 (left) or CEB25wt (right) (10  $\mu$ M) in K<sup>+</sup> buffer (100 mM) with benzophenone derivatives: **C1/Bph**, **C3/Bph**, **DEG/Bph**, and **TEG/Bph**, benzyl diazirine derivatives: **C1/N2**, **C3/N2**, **DEG/N2**, **TEG/N2**, and aryl azide derivatives: **C1/N3**, **C3/N3**, **DEG/N3**, **TEG/N3** (20  $\mu$ M). *h $\nu$* : absence (-) or presence (+) of irradiation at 360 nm for 1 h.

## 2. Chemical and enzymatic sequencing methods

Identifying the alkylation sites on the G-quadruplex structures was attempted using chemical sequencing methods. DMS footprinting is a chemical sequencing method developed by Maxam and Gilbert and described in detail in Chapter 1 (see section 1.2.2.2, p. 40).<sup>74</sup> The latter is a specific guanine (and adenine to some extent)-sequencing method performed in two steps: i) methylation reaction in the presence of dimethylsulfate (DMS), and then, ii) basic treatment with piperidine and heating to induce depurination and  $\beta$ -elimination reaction to cleave the 2'-deoxyribose backbone into the <sup>32</sup>P-5'- and 3'-fragments. Another sequencing exploited the alkylation of the covalent G4 compounds: the samples were exclusively treated with piperidine. If the covalent modification destabilises the N-glycosidic bond enough, the piperidine treatment should enable  $\beta$ -elimination reaction. These two chemical sequencing methods were applied to the alkylation products obtained from the **in-line** and **branched** benzophenone derivatives, which gave the highest alkylation yields, in the presence of 22AG to try to find the exact alkylation sites. The experiments with the **in-line** derivatives were realised by Dr. T. Masson and the ones with the **branched** benzophenone derivatives by myself.

### 2.1. Chemical sequencing on the in-line alkylated adducts

The chemical sequencing of the unique adducts (1.1, 2.1, and 3.1) generated by the three **in-line** compounds shows a product distribution different from the non-alkylated 22AG (1.0, 2.0, and 3.0) as displayed in Figure 97. Remarkably, the guanosine bands showed a reduced intensity as compared to 22AG. This may indicate a direct alkylation of the guanosines or that the alkylation induced by the G4 ligands keeps folded the G4 structure, protecting it from DMS methylation. In addition, new bands not matching 22AG pattern can be observed: for example, **C2-/N2** adduct (2.1) presents a new fragment just above *G3* (i.e. 5'-AG fragment) that could correspond to a short alkylated fragment, suggesting modifications located at the close proximity of the 5'-end. However, it is interesting to observe that each covalent adduct generated by the three **in-line** derivatives presents different patterns, indicating that the difference in distance between the PDC core interacting with the G-quartet and the aliphatic diazine moiety has an impact on the preferential alkylation positions. Finally, the covalent adducts show very high thermal stability: no significant dealkylation/depurination was observed upon heating at 95 °C.

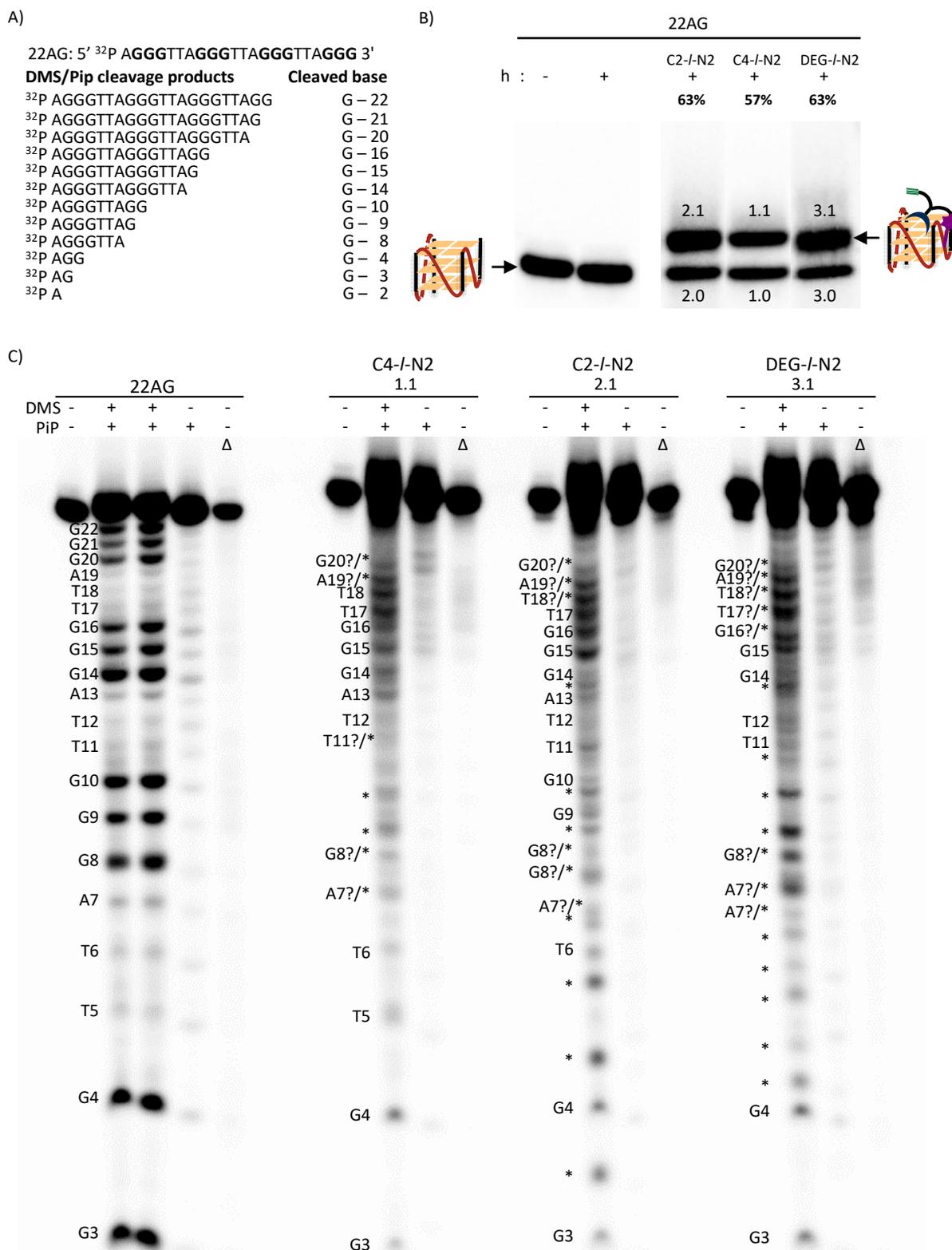


Figure 97: A) Reference fragmentation pattern after DMS/piperidine treatment of the 22AG sequence. B) Denaturing gel electrophoresis (15% acrylamide) experiments of the photoalkylation reaction performed in the presence of 22AG (10 μM) in K<sup>+</sup> buffer (100 mM) with C2-I-N2, C4-I-N2, and DEG-I-N2 (20 μM). h v: presence (+) of irradiation at 365 nm for 1 h. Bands have been numbered with respect to their migration as compared to a control, band X.0: no alkylation, band X.1: alkylation retarded band. C) Denaturing gel electrophoresis (20% acrylamide) of alkylated products of bands 1.1, 2.1, and 3.1 with or without DMS treatment and/or piperidine treatment. 22AG band correspond to band 1.0 with DMS and DMS/piperidine treatment to give a reference scale for fragment migration. DMS: dimethylsulfate treatment; Pip: piperidine treatment; Δ: sample was heated at 90 °C for 25 min; \* correspond to the formation of a new fragment absent in the control DMS/piperidine pattern.

## 2.2. Chemical sequencing on the benzophenone alkylated adducts

The chemical sequencing of the main two adducts generated by three branched benzophenone derivatives, **C1/Bph** (1.2 and 1.3), **DEG/Bph** (4.2 and 4.3) and **TEG/Bph** (3.2 and 3.3), in the presence of 22AG did not allow the identification of the alkylation sites. After treatment, we observed (Figure 98) an almost complete dealkylation and regeneration of the starting G4 sequence. As a consequence, piperidine treatment did not allow us to make any substantial hypothesis about the nature of the alkylation products. Nevertheless, these results highlighted the thermal instability of the covalent adducts generated by the reaction of the benzophenone reactive species and 22AG.

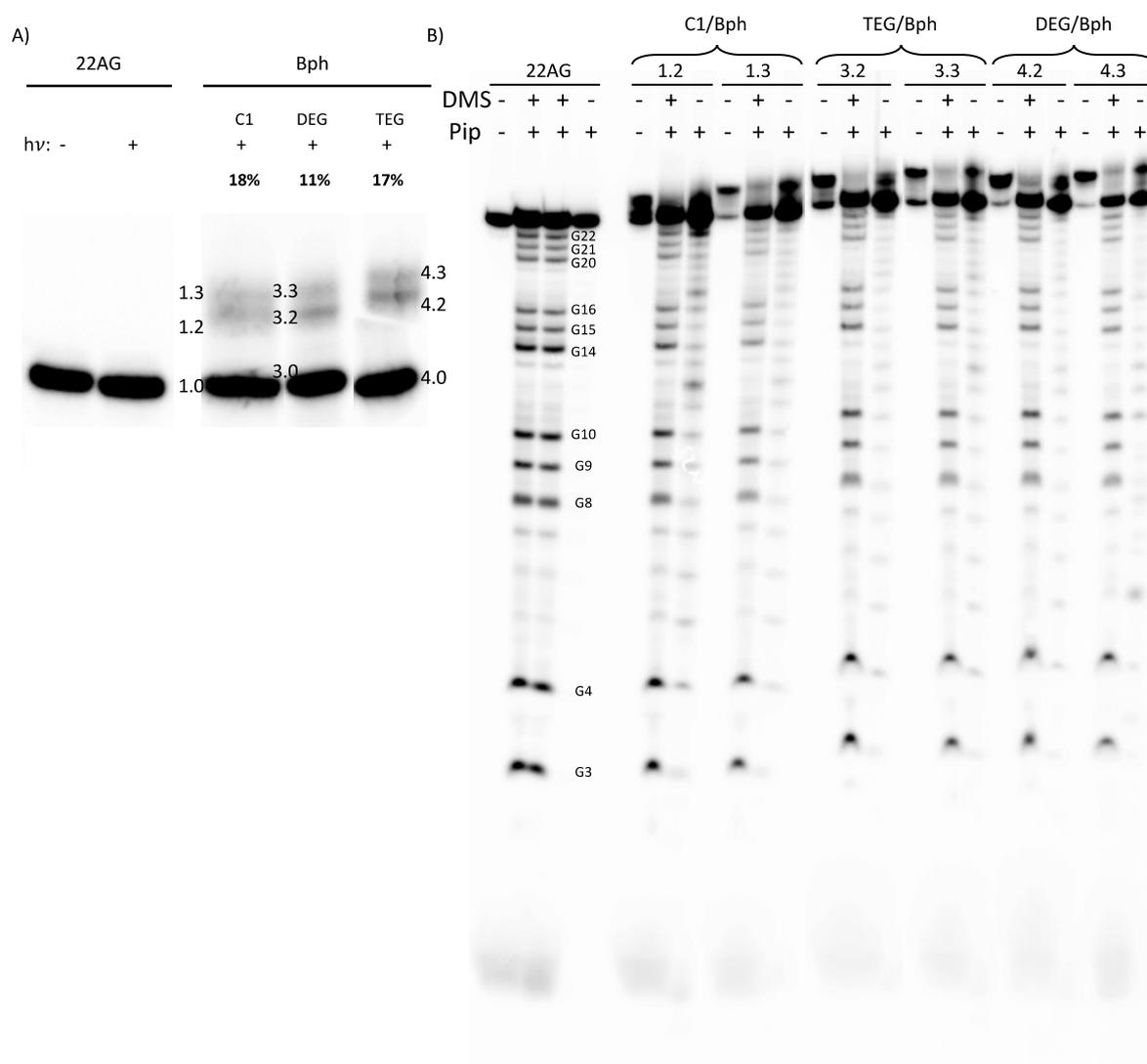


Figure 98: A) Denaturing gel electrophoresis (15% acrylamide) experiments of the photoalkylation reaction performed in the presence of 22AG (10  $\mu$ M) in  $K^+$  buffer (100 mM) with **C1/Bph**, **TEG/Bph**, and **DEG/Bph** (20  $\mu$ M) after irradiation at 365 nm for 1 h. Bands have been numbered with respect to their migration as compared to the control, band **X.0**: no alkylation, band **X.2** and **X.3**: alkylation retarded bands, numbered in the order of the speed of migration. B) Denaturing gel electrophoresis (20% acrylamide) of alkylated products of bands **1.2**, **1.3**, **2.1**, **2.2**, **3.1** and **3.2** with or without DMS and/or piperidine treatment. 22AG band correspond to band **1.0** with DMS and DMS/piperidine treatment to give a reference scale for fragment migration. DMS: dimethylsulfate treatment; Pip: piperidine treatment.

## 2.3. Enzymatic sequencing with 3'-exonuclease

Chemical sequencing did not allow to identify the alkylated sites neither for the **in-line** or the **branched** compounds. This is why we decided to apply to these covalent adducts the 3'-exonuclease enzymatic sequencing.<sup>223</sup> 3'-exonuclease digests single-stranded DNA nucleotide by nucleotide from the 3'- to the 5'-end, and it is unable to recognise modified nucleotide. Thus, the 3'-exonuclease digests the sequence until it encounters the alkylated nucleotide and stops to give a <sup>32</sup>P-5'-fragment (Figure 99). The thermal reversibility of the adducts generated by the **branched** benzophenone derivatives can be used for the identification of the alkylated sites. The migration of the partially digested fragments, after dealkylation, can thus be compared with the migration of the non-alkylated 22AG ladder to identify the exact alkylation sites. The 3'-exonuclease sequencing experiments on the adducts generated by the **in-line** compounds were realised by Dr. T. Masson, differently 3'-exonuclease sequencing on the adducts generated by the **branched** compounds were performed by myself.

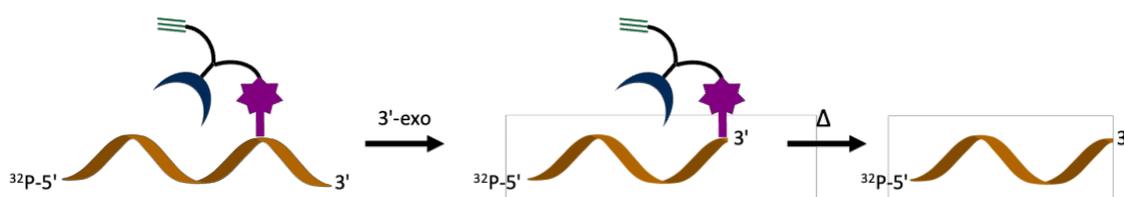


Figure 99: Representation of the 3'-exonuclease digestion and subsequent dealkylation process. In-line compounds cannot be dealkylated upon heating, thus the second step is not realised. The branched benzophenone derivatives can be dealkylated, thus the second steps was realised.

### 2.3.1. 3'-exonuclease treatment of in-line covalent adduct

The covalent adducts generated from the photoreaction of the in-line compounds with 22AG show a product distribution different from the non-alkylated 22AG as displayed in Figure 100, but almost identical for the three **in-line** covalent adducts. Increasing unit of 3'-exonuclease (from 0.0013 to 0.04 U/ $\mu$ L) does not show the formation of new bands. Remarkably, five fragments present a slower migration as compared to the non-treated alkylated band. This is characteristic of very short alkylated fragments where the negative charges are not sufficient to compensate the impact of the PDC core positive charges and corresponding mass. Thus, these bands correspond surely to fragments where the alkylation is located in the first 5'-AGGGTT motif.

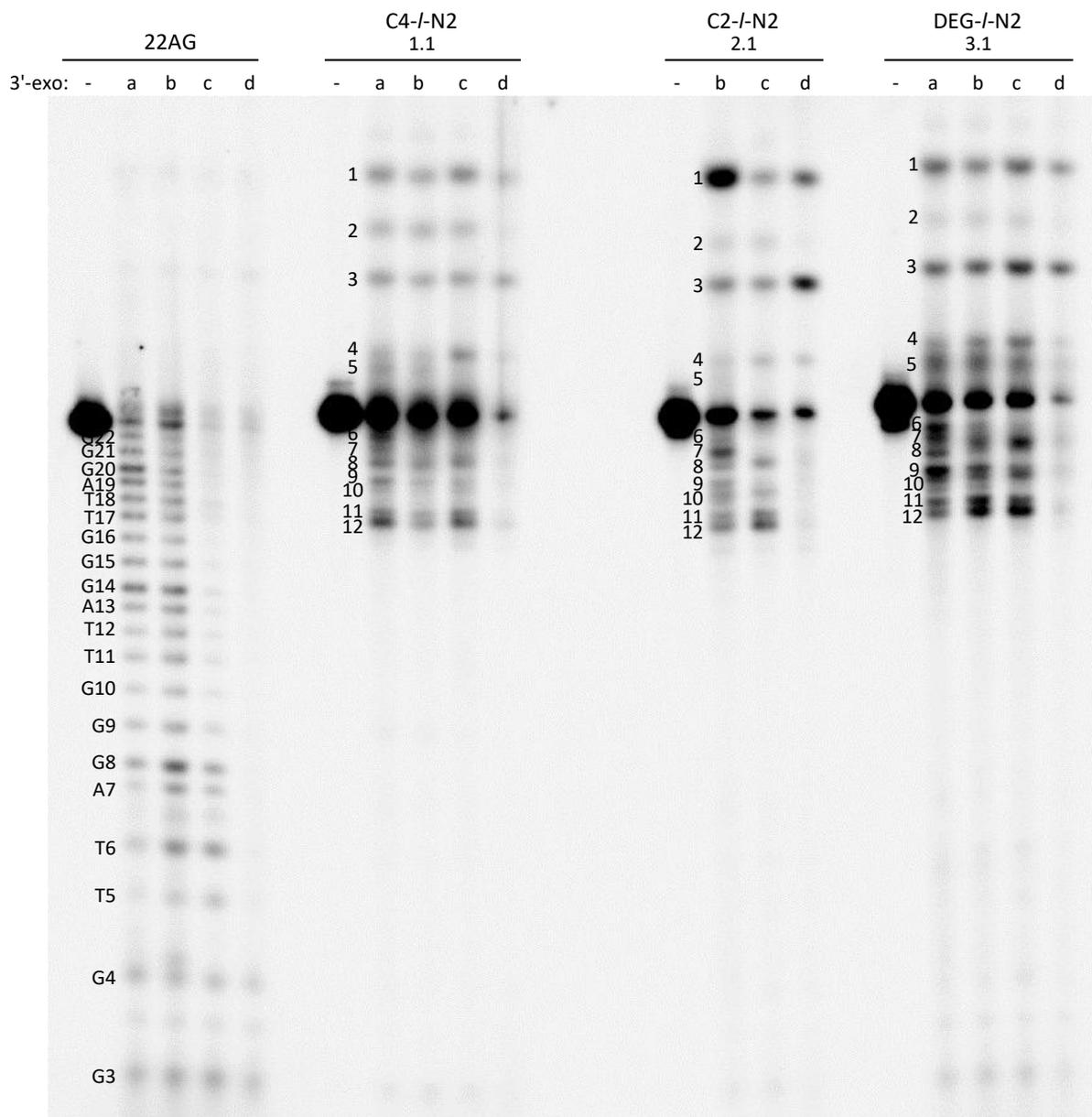


Figure 100: Denaturing gel electrophoresis (20% acrylamide) of 3'-exonuclease digestion of bands 1.0 (22AG), 1.1, 2.1 and 3.1. Enzyme units/ $\mu$ L: absence (-), 0.0013 U/ $\mu$ L (a), 0.002 U/ $\mu$ L (b), 0.004 U/ $\mu$ L (c), and 0.04 U/ $\mu$ L (d).

### 2.3.2. 3'-exonuclease treatment of **benzophenone** covalent adduct

The 3'-exonuclease sequencing was realised on the main covalent adducts generated by the four **branched** benzophenone compounds. Two main slow migrating bands .2 and .3 were isolated for the four derivatives, and the weaker slow migrating band .1 were isolated for **C3/Bph** and **DEG/Bph**, as represented in Figure 101.

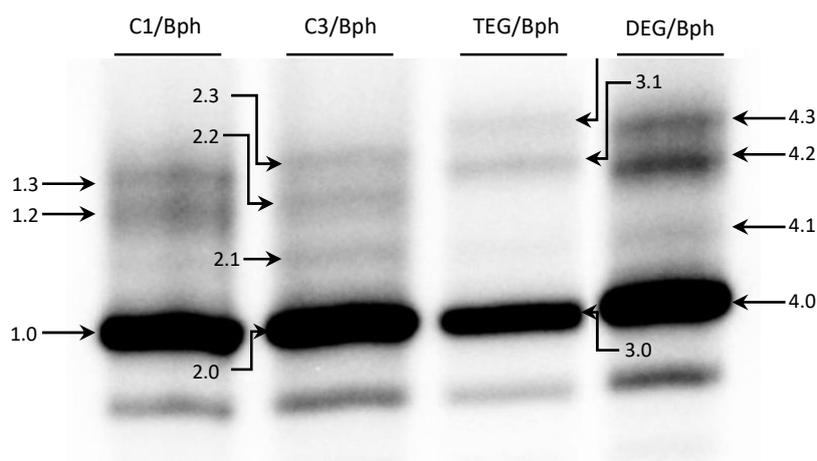


Figure 101: Denaturing gel electrophoresis (15% acrylamide) experiments of the photoalkylation reaction performed in the presence of 22AG (10  $\mu$ M) in  $K^+$  buffer (100 mM) with **C1/Bph**, **C3/Bph**, **TEG/Bph**, and **DEG/Bph** (20  $\mu$ M) after irradiation at 360 nm for 1 h.

Each adduct was treated with different activities of 3'-exonuclease and the generated 5'-fragments migrate faster than the undigested sequence (Figure 102). Globally, all the covalent adducts share similar migrating fragments. The more abundant digested fragments were isolated and heated at 90 °C for 25 min in water to induce partial dealkylation. The resulting product distribution was analysed by denaturing PAGE.

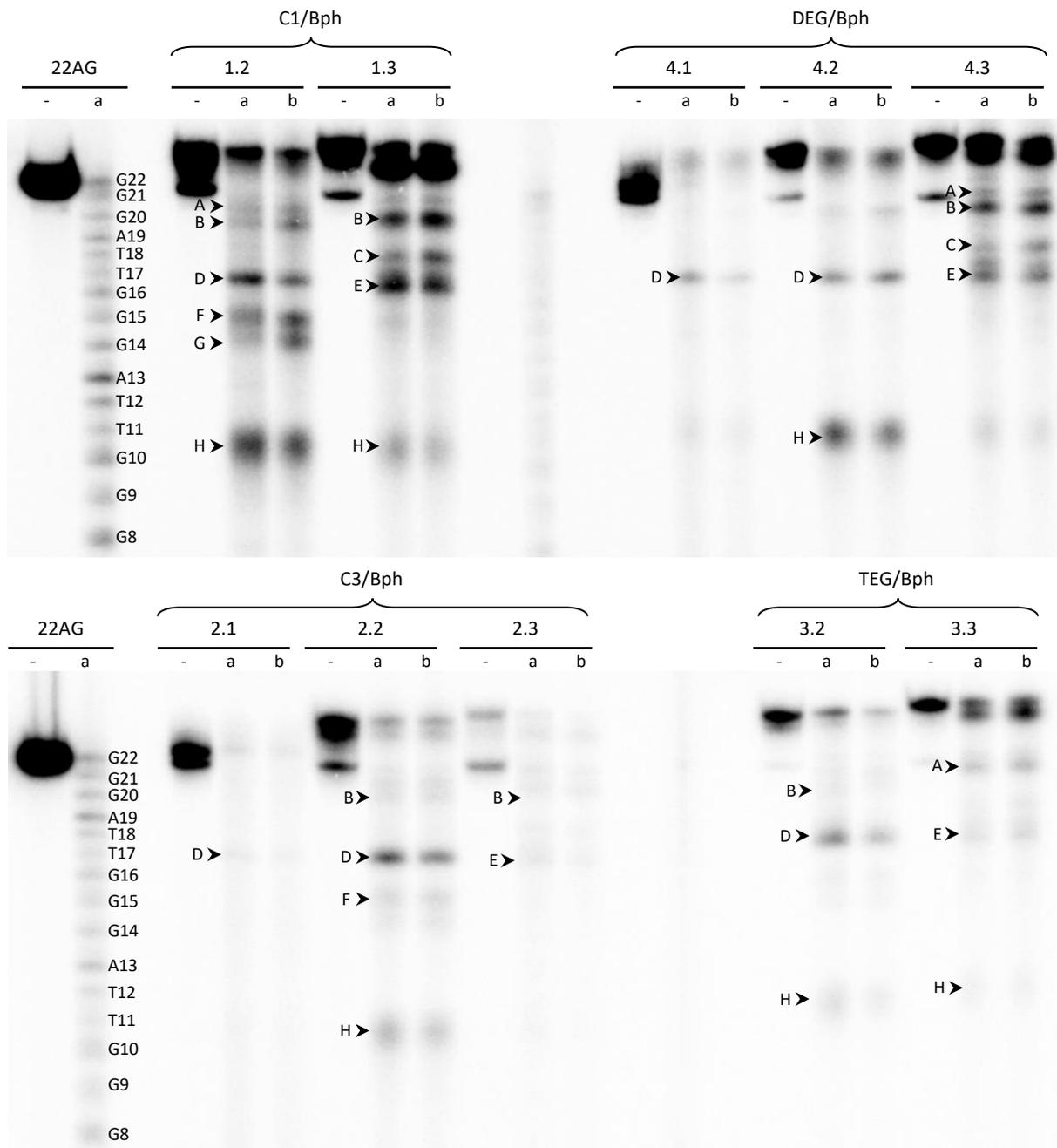


Figure 102: Denaturing gel electrophoresis (20% acrylamide) of 3'-exonuclease digestion of bands 1.0 (22AG), 1.2, 1.3, 2.1, 2.2, 2.3, 3.1 and 3.2. Enzyme units/μL: absence (-), 0.001 U/μL (a), and 0.0013 U/μL (b). Each isolated fragment band is labelled with a letter (A, B, C, D, E, F, G, H) in function of their migration order from top to bottom.

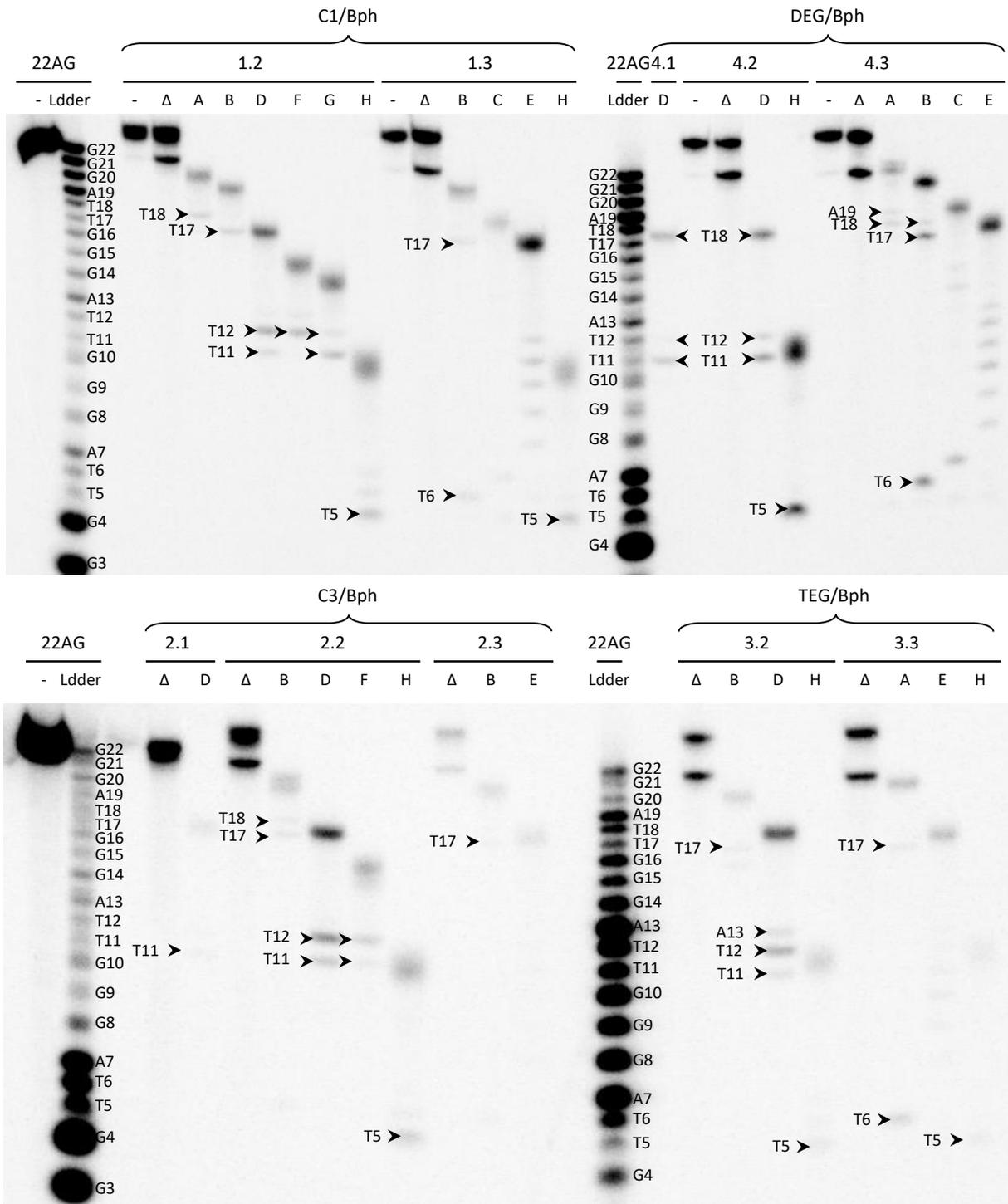


Figure 103: Denaturing gel electrophoresis (20% acrylamide) for the thermal treatment of the digested fragments by heating at 90 °C for 25 min the isolated fragments (A, B, C, D, E, F, G) from Figure 102. Each dealkylation is only partial: the slow migrating band corresponds to the reference alkylated fragment and fast migrating arrests corresponds to the dealkylated fragment. The numbered nucleotides correspond to the digestion dealkylated arrests.

The migration of the faster migrating bands corresponding to the dealkylated fragment (Figure 103) can be compared with the reference 22AG ladder. The slower migrating fragment are used as a reference of the initial alkylated fragment. Globally, all dealkylated fragments correspond to the thymidines located in the different loops of the G4 structure of 22AG, as reported in Table 3.

Arrest	2.1	4.1	1.2	2.2	3.2	4.2	1.3	2.3	3.3	4.3	Main
A			T18						T17/T6*	A19, T18	T18
B			T17	T18, T17	T17		T17/T6*	T17/T6*		T17/T6*	T17/T6*
C							**			**	**
D	T11	T12, T11	T12, T11	T12, T11	A13, T12, T11	T11, T12					T11, T12
E							**	**	**	**	**
F			T12	T12, T11							T11, T12
G			T12, T11								T11
H			T5	T5	T5	T5	T5		T5		T5

Table 3: Summary of the products generated from the dealkylation. \* The T6 position is found in addition to the T17. \*\* Presence of a degradation pattern.

The arrests *A* and *B* gave respectively **T18** and **T17**, the arrests *D*, *F*, and *G* gave both **T12** and **T11**, and the arrest *H* gave **T5**. In addition, arrests *C* and *E* present a partial degradation. Surprisingly, arrest *A* for 3.3 and *B* for 1.3, 2.3 and 4.3 present the formation of two alkylation sites corresponding to **T17** and **T6**. The **T17** site is consistent with the results obtained for arrest *B* in 1.2, 2.2, and 3.2, and in theory, the **T6**-alkylated arrest should be migrating faster and been between arrests *G* and *H*. This additional site may be due to partial degradation as observed in arrests *C* in 1.3 and 4.3 with an intense A7 fragment band. The observed degradations may originate from the numerous treatments we submitted the samples to or to an intrinsic instability of the photoadducts non-thermally reversible.

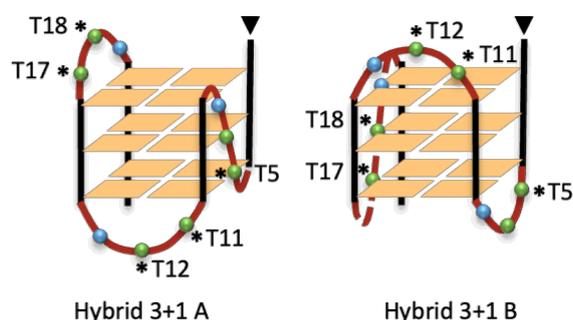


Figure 104: Representation of the alkylation sites (\*) on 22AG structure considering the hybrid structure.

### 3. Conclusion

The **in-line** compounds present excellent alkylation yields with the diverse G-quadruplexes studied, showing alkylation yields in the range between 40 and 70% with no selectivity for a particular G4 topology. Sequencing experiment did not allow the characterisation of the alkylated nucleotide. However, the numerous alkylation fragments observed by chemical sequencing experiments and 3'-exonuclease digestion highlighted a specific 5'-region localisation of the alkylation sites.

The **branched** compounds present an important alkylation selectivity towards specific G4 topologies showing poor alkylation (8-18% Bph, 6-9% N3, and 3-5% N2) with the hybrid structure of 22AG, and high alkylation yields with the parallel topologies of cMYC22 and CEB25 *wt*. It is interesting to note that there is no strict correlation between the ability of the ligand to stabilise a G4 structure and the efficiency of the alkylation process. This indicates that the interaction between ligand and G4 is strong enough to guarantee the alkylation. Thermal treatment of the alkylated products produced by **branched** benzophenone derivatives in the presence of 22AG, obtained by enzymatic sequencing, underlined the thermal reversibility of these products. The latter highlighted the strong selectivity of these G4 ligands to form covalent bonds with the thymidine located in the loops of the G4 structures.

A family of photoactivatable G4 ligands functionalised with benzophenone (Bp) and aryl azide (N3) moieties was previously synthesised and studied in the laboratory by Verga *et al.*<sup>179</sup> These ligands present a similar scaffold as the one used in this study with the exception of the terminal alkyne spacer. In these compounds, the photoactivatable moieties were connected to the PDC core by various linkers (C4, C8, and PEG), as represented in Figure 105.

These benzophenone ligands displayed alkylation yields ranging between 26-33% for PDC-C4-Bp and PDC-PEG-Bp and poor alkylation yield for PDC-C8-Bp (3%). Also in this previous work, alkylation adducts were represented by three slower migrating bands, as observed for our benzophenone derivatives. The new benzophenone series presents a lower alkylation yield (8-18%) as compared to PDC-C4-Bp and PDC-PEG-Bp, but a better yield than the PDC-C8-Bp. This last observation is quite interesting since the new benzophenone series scaffold is structurally closer to the latter (linker with a similar length). The yield improvement could be due to the presence of the central amide adding some polarity and flexibility to the aliphatic chain, and thus to the benzophenone moiety. In addition, 3'-exonuclease sequencing experiments on their benzophenone adducts showed the formation of covalent adducts strongly selective

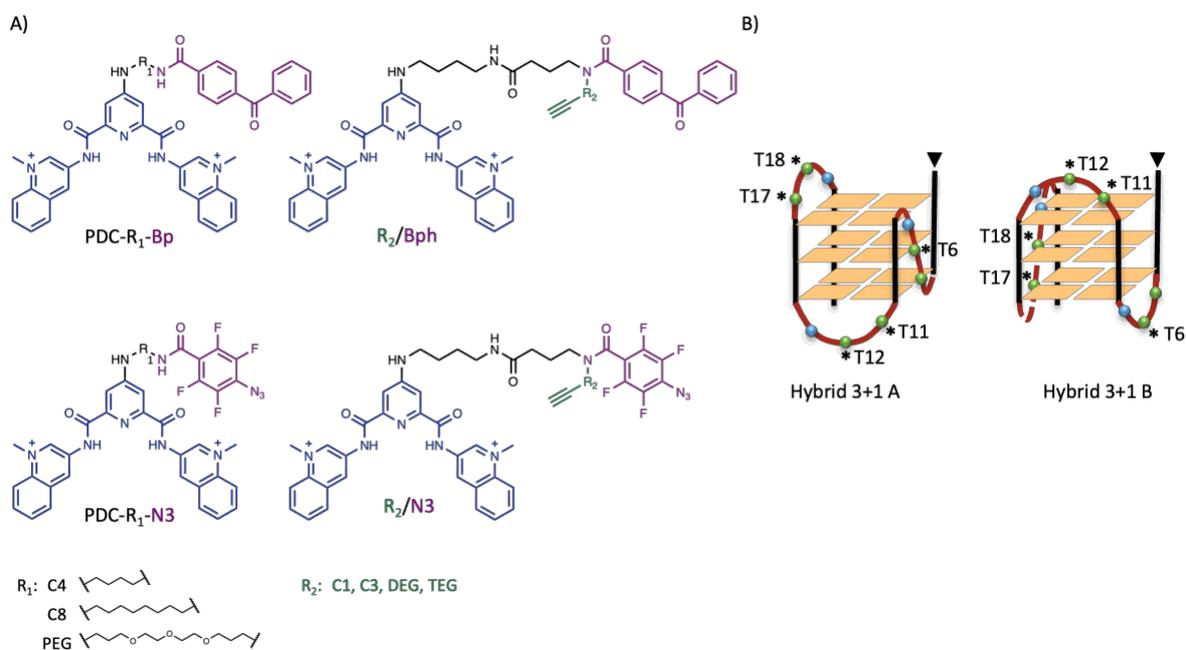


Figure 105: A) Representation of the photoactivatable G4 ligands developed by Verga et al. (2014)<sup>179</sup> (left) and the scaffold benzophenone and aryl azide derivatives studied in this thesis (right). B) Representation of the alkylation sites generated by PDC-PEG-Bp on the two hybrid forms of 22AG.

for thymidine (Figure 105), which corroborates with the results of the previously developed benzophenone series. Alkylation efficiency was also studied with cMYC22, but no alkylation was observed in the presence of PDC-C4-Bp and PDC-PEG-Bp. Differently, the molecules presented in this PhD work showed yields comprised between 50 to 100% with cMYC22, suggesting that the nature of the linker plays a relevant role for these kinds of studies. In a near future, the information collected with these data will help us to design more efficient photoactivatable G4 ligands.

## Chapter IV:

Detection of the reactive species by LFP  
and  
Structural characterisation of the photoadducts by NMR



In this new chapter, we have focused on the detection and characterisation by Laser Flash photolysis (LFP) of the transient species generated by the photoreactive moieties, as well as their reactivity with nucleosides including the characterisation of the reaction kinetics, whenever it is possible. Finally, products distribution, isolation, and NMR characterisation were realised for the adducts generated from the aliphatic diazirine and the benzophenone. Indeed, the investigation was not conducted directly with the PDC conjugated to both the photoactivatable and alkyne moieties (**in-line** compounds and **branched benzophenone** derivatives listed in Figure 69), but instead on prototype compounds to mimic the reactivity of the photoactivatable G-quadruplex ligands, to improve the water-solubility at high concentration (mM range for photochemical experiments vs.  $\mu\text{M}$  range for biophysics and biochemistry experiments).

## 1. Presentation of the photoactivatable moieties

### 1.1. Benzophenone

Benzophenone is a well-known photoactivatable compound known for its ability to form covalent adducts through C-C bond by UV-activation. Benzophenone can be excited by absorption of a low energy photon ( $\lambda \sim 350 \text{ nm}$ ) inducing an electronic transition from the nonbonding n-orbital of the oxygen to the antibonding  $\pi^*$ -orbital of the ketone ( $n \rightarrow \pi^*$  transition).<sup>224</sup> A higher energetic transition is possible ( $\lambda \sim 250 \text{ nm}$ ) inducing an electronic transition from the bonding  $\pi$ -orbital to the antibonding  $\pi^*$ -orbital ( $\pi \rightarrow \pi^*$  transition). Both result in the generation of a singlet biradical excited state ( $S_1$ ) that undergoes efficient and fast ( $t < \text{ns}$ ) intersystem crossing (ISC) to generate a triplet excited state ( $T_1$ ) with a long lifetime ( $t > \mu\text{s}$ ) with biradical character.<sup>225</sup> The quantum yield ( $\phi$ ) of the C-H abstraction is close to 1 for the  $n \rightarrow \pi^*$  triplet state and much lower ( $< 0.1$ ) for the  $\pi \rightarrow \pi^*$  triplet state.<sup>226</sup> This triplet state can react by H-abstraction to form the benzophenone ketyl radical (benzhydryl) transient (Figure 106).<sup>180</sup>

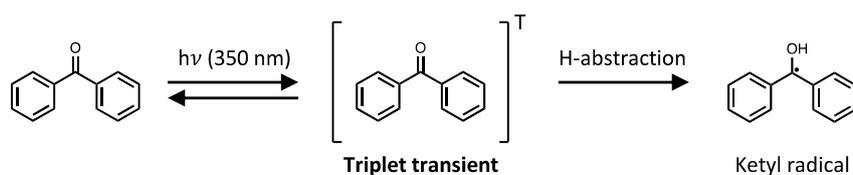


Figure 106: Molecular representation of the benzophenone transient species. The UV activation at 350 nm generate a triplet state transient having a biradical character. During its lifetime, the triplet state transient can react by H-abstraction to generate the ketyl radical transient.

Benzophenone has been exploited as photoactivatable moiety for biolabelling due to the biocompatible activation absorption (330 – 360 nm) with endogenous chromophores ( $\lambda < 300$  nm) and the high quantum yield. The photolabelling is theoretically possible with any biotargets such as proteins and DNA. Benzophenone can react with any amino acids by a first H-abstraction to generate an  $\alpha$ -carbon radical and the benzophenone ketyl radical reacts with the latter to form the photoadduct as displayed in Figure 107.<sup>227,228</sup>

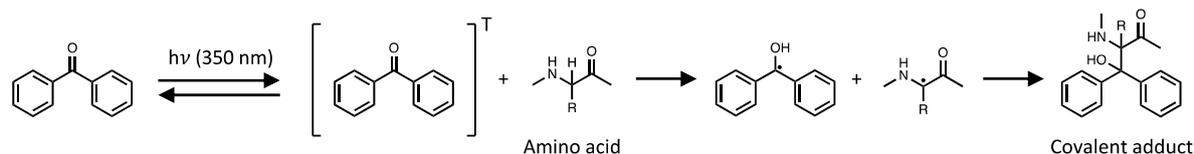


Figure 107: Photolabelling of an amino acid by the activated benzophenone.

Benzophenone photolabelling of DNA shows a selectivity for the thymidine bases over the rest of the purines as observed by the measured quenching kinetic constant in Figure 108 by Charlier and Helene.<sup>180</sup> It reacts with the double bond of the thymine by a [2+2] cycloaddition named Paternò-Büchi reaction, yielding an oxetane.<sup>193</sup> Another described mechanism is the triplet-triplet energy transfer to the thymidine forming thymine-thymine dimers leading to DNA damage.<sup>193</sup> However, a limited number of studies use benzophenone with DNA strands and not only nucleosides and nucleobases. For example, benzophenone derived as nucleotide has been used to crosslink DNA strands,<sup>229</sup> or tethered on PDC to alkylate selectively G-quadruplexes.<sup>179</sup> The benzophenone was incorporated into proteins to investigate protein-protein and protein-DNA interaction.<sup>230,231</sup>

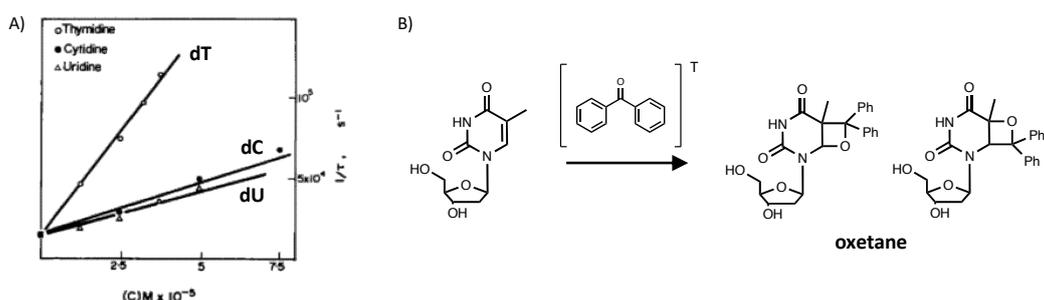


Figure 108: A) Total rate constant for the quenching of benzophenone triplet state by pyrimidine nucleosides in aqueous solution. (Figure modified from Charlier and Helene, 1972).<sup>180</sup> B) Molecular representation of the oxetane adducts formation by the [2+2] cycloaddition of the triplet state benzophenone on the thymidine.

## 1.2. Aryl azide

Aryl azide is a widely used photolabelling moiety generating a nitrene through photochemical activation. The azidobenzene activation is triggered by absorbing high energy light ( $\lambda \sim 250$  nm) and losing  $N_2$  to generate a short living ( $< \mu s$ ) singlet nitrene transient,<sup>190</sup> which can be used for photolabelling. However, the photochemistry of the azidobenzene is complex due to many intermediate species formed shown in Figure 109.<sup>191</sup> The singlet nitrene ( $^1N$ ) can rearrange by ring expansion into a ketenimine from an intermediate bicyclic azirine (benzazirine), both structures can be attacked by a nucleophile. In addition, the singlet nitrene can be converted into the triplet nitrene ( $^3N$ ) by intersystem crossing, which can dimerise into azobenzene.

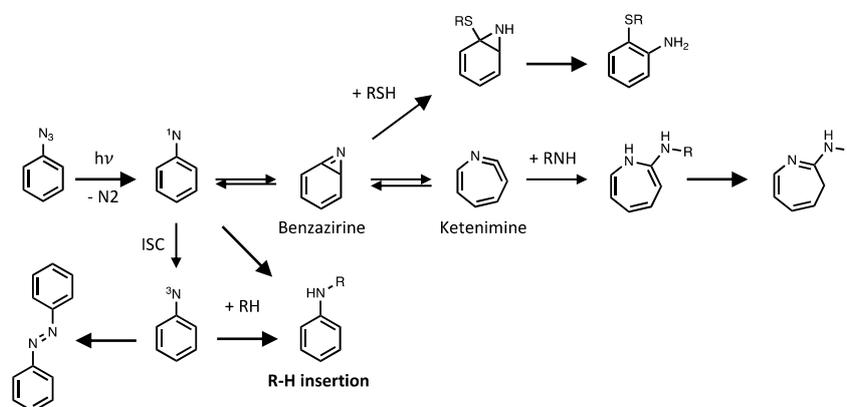


Figure 109: Photolysis pathways of the azidobenzene.

However, irradiation at  $\sim 250$  nm can lead to the excitation of endogenous chromophore and cause substantial damage making it incompatible for biological environment. In addition, phenyl azide is poorly reactive due to the rearrangement to form benzazirine and ketenimine producing undesirable products. Different aryl azide derivatives have been developed to overcome these problems such as the tetrafluorophenylazide moiety exploited in this thesis (Figure 110).<sup>181</sup> Its nitrene can be triggered at lower energy ( $\lambda \sim 350$  nm) and the presence of the halogens raised the energy barrier of the ring expansion reaction and significantly increased the lifetime of the corresponding singlet phenylnitrenes.<sup>232</sup> The longer lifetime increases the probability to react with surrounding molecules.

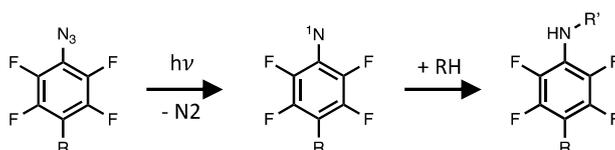


Figure 110: Representation of the tetrafluorophenylazide moiety and the insertion pathway.

### 1.3. Diazirine

Diazirine (N<sub>2</sub>) is a small lipophilic and non-bulky photoactivatable moiety, which upon irradiation at 350 nm generates a diazoisomer transient and a carbene after releasing a dinitrogen molecule. Common diazirines are found in two forms: aliphatic diazirine and phenyl diazirine. The aliphatic diazirine<sup>197</sup> generates a singlet carbene having a zwitterionic character, it presents one fully occupied orbital and one empty orbital. The fully occupied orbital acts as a nucleophile realising H-abstraction on the nucleophile, which reacts with the empty orbital, in a concerted mechanism, by X-H insertion (X = C, O, N). However, the singlet carbene has a short lifetime and reacts quickly by intramolecular H-shift yielding an alkene (Figure 111).

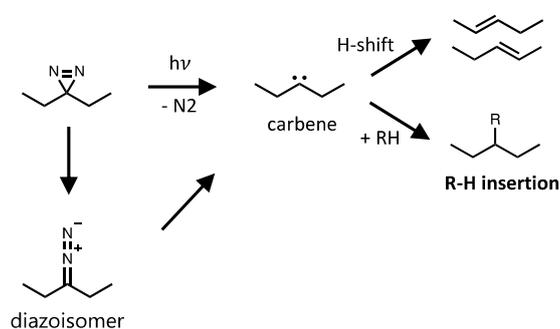


Figure 111: Photolysis pathway of the aliphatic diazirine.

Benzyl diazirines have been developed to reduce these problems, for example, the trifluorophenyldiazirine designed by Brunner *et al.* is a well-known photolabelling moiety.<sup>195</sup> It forms mainly the triplet state carbene, which has a longer lifetime and behaves as a biradical. This feature reduces the competing H-shift. However, trifluorophenyldiazirine also has drawbacks, it can be quenched by additional pathways represented in Figure 112. For example, it can react with an amine by N-H insertion and hydrolyse to generate a ketone, significantly reducing the overall efficiency of the alkylation step.<sup>196</sup>

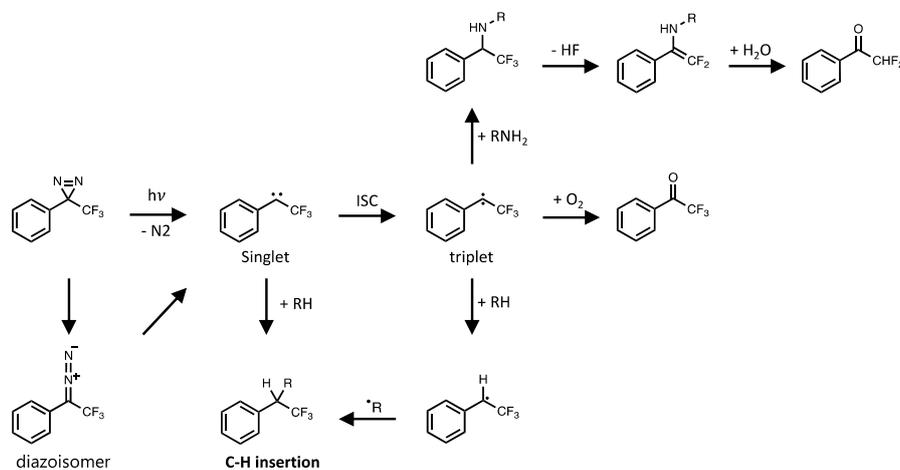


Figure 112: Photolysis pathway of the trifluorophenyldiazirine.

The photochemical activation of the diazirine is considered to generate a carbene and a diazoisomer, the latter is trusted to be converted into the carbene during continuous UV irradiation. Recently, this reactivity has been questioned by O'Brien *et al.* (2020).<sup>233</sup> They succeeded to trap the diazoisomer generated from an aliphatic diazirine using bicyclononyne (BCN) into the 3*H*-pyrazole form with a 22% yield (Figure 113-A). In addition, they observed that the alkenes, lactone and O-H insertion product yields diminished. This indicates that not only the diazoisomer is formed but reacts to generate the alkenes, lactone, and O-H insertion products as well. Finally, they evaluated the labelling of bovine serum albumin (BSA) and observed up to 69% decrease in labelling efficiency when adding the BCN competitor (Figure 113-B), confirming that the main reactive species was the diazoisomer and not the carbene.

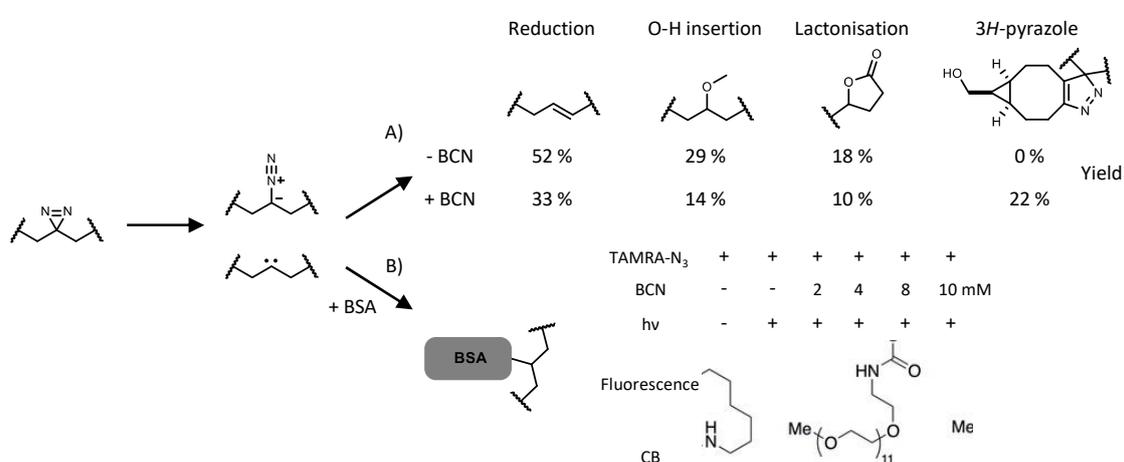


Figure 113: Summary of O'Brien *et al.* results with A) the trapping of the diazoisomer into the 3*H*-pyrazole by BCN and B) competitive labelling of BSA in presence of BCN. (Gel figure from O'Brien *et al.*, 2020)<sup>233</sup>

Similarly, West *et al.* (2021) challenged the statement that diazirine could label amino acids without preference.<sup>198</sup> At first, they demonstrated a difference in the reactivity toward natural amino acids between the aliphatic and benzyl trifluoromethyl diazirine. The first one reacted with only nine amino acids and displayed a strong preference for acidic amino acids, such as tyrosine, glutamic acid, cysteine, aspartic acid and histidine. Differently, the second one reacted with all of the 20 natural amino acids with preference for cysteine, glutamic acid and aspartic acid. They attributed the difference of reactivity to the stability of the diazoisomer generated from the benzyl trifluoromethyl diazirine, in which the negative charge is stabilised by both the aromatic ring and the trifluoromethyl substituent. In addition, they demonstrated that the reaction of the aliphatic diazirine with the acidic amino acids (glutamic acid, histidine, and tyrosine) is pH-dependent.

This reactivity can be explained by the ability of the diazoisomer to react with acidic amino acid through a proton transfer mechanism, producing the amino acids anion and the diazonium ion, which react together to generate the insertion product. Finally, the labelling of BSA is pH-independent using benzyl trifluoromethyl diazirine, but pH-dependent using aliphatic diazirine leading up to 50% labelling decrease when changing pH from 5.8 to 8.0 (Figure 114).

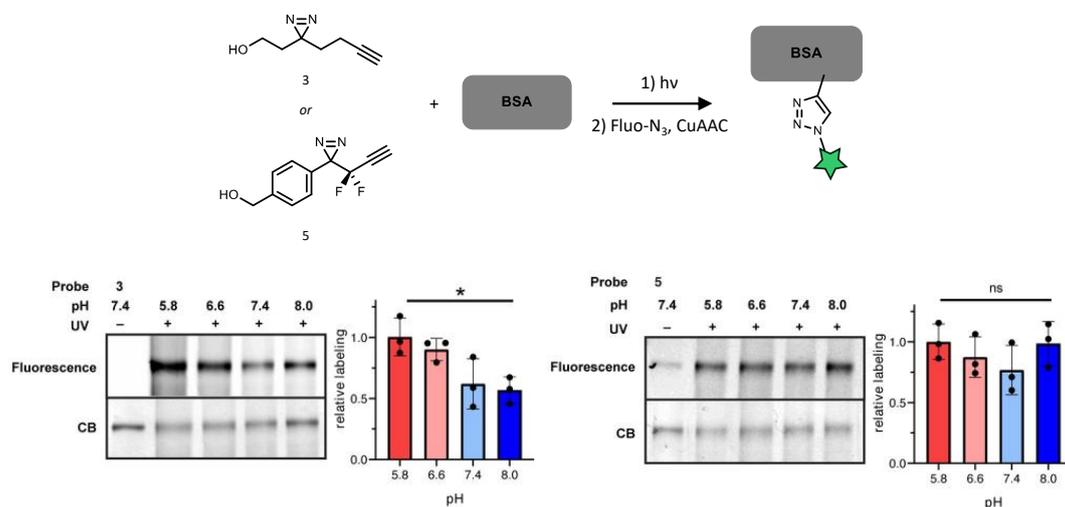
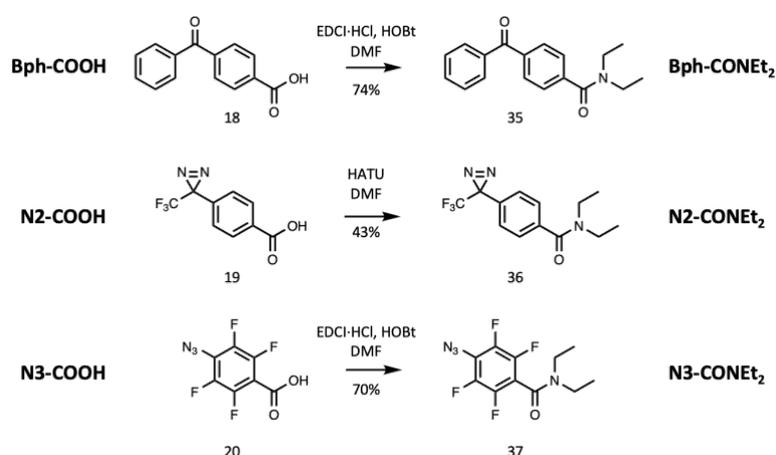


Figure 114: BSA labelling utilising the aliphatic diazirine **3** or benzyl trifluoromethyl diazirine **5** in function of the pH. (Bottom figure from West et al., 2022)<sup>198</sup>

To summarise, aliphatic diazirine generates a carbene and a diazoisomer that are both responsible for the labelling through at least one pH-dependent mechanism. The benzyl trifluoromethyl diazirine seems to generate mainly a carbene (singlet and triplet), which labels through pH-independent mechanisms. Thus, excluding the carbene to follow a similar pH-dependent "ion pair" pathway. It remains to be confirmed as evidence of an "ion pair" pathway, in addition to the concerted O-H insertion and alcohol ylide pathways, which has been described with diarylcarbenes interacting with O-H bonds.<sup>201,234,235</sup>

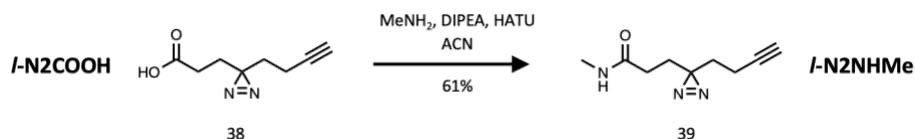
## 2. Synthesis of the photoactivatable moieties

The mimic prototypes have been designed to keep the global structure including the linker functional group tethering the photoactivatable moiety, as its nature could modify their reactivity and properties. The photoactivatable moieties in the **branched** compound are linked to the PDC and the terminal alkyne-spacer by an *N,N*-disubstituted amide, thus, the **branched** prototype was designed by linking an *N,N*-disubstituted amide with two ethyl chains on the photoactivatable moiety. The **branched** photoactivatable precursors are commercially available in their carboxylic acid form, which were used to synthesise the **branched** compounds as detailed in chapter II (see section II.1.2.2, p. 96). A similar approach was followed here by coupling the carboxylic acid photoactivatable moiety with diethylamine (Scheme 6).



Scheme 6: Synthesis of the **branched** mimic prototype (**X-CONEt<sub>2</sub>**) from their corresponding carboxylic acid precursors (**X-COOH**). **X** = photoactivatable moiety.

Precisely, the synthesis of the 4-benzoyl-*N,N*-diethylbenzamide (**35**) and the 4-azido-*N,N*-diethyl-2,3,5,6-tetrafluorobenzamide (**37**) were carried out by adding 1.1 eq. of EDCI·HCl and 1.1 eq. of HOBT to a solution containing the corresponding carboxylic acid precursor (**18** and **20**) in DMF. The reaction was run for 2-3 h at room temperature and products were extracted in the organic phase to give **Bph-CONEt<sub>2</sub> (35)** and **N3-CONEt<sub>2</sub> (37)** with 74% and 70% yields, respectively. For the synthesis of the *N,N*-diethyl-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamide (**36**), the use of EDCI·HCl was not successful with a conversion of 20% determined by LC-MS after stirring overnight. The reaction was repeated by replacing EDCI·HCl and HOBT by 1.5 eq. of HATU to give **N2-CONEt<sub>2</sub> (36)** with 43% yield. Product purities were checked by HPLC with over 95% purity for all three.



Scheme 7: Synthesis of the *I-N<sub>2</sub>NHMe* prototype from the *I-N<sub>2</sub>COOH* precursor.

The **in-line** compounds are composed of the diazirine-alkyne moiety linked to the PDC by a C2, C4 or DEG spacer through an *N*-monosubstituted amide. It has been decided to functionalise the diazirine-alkyne moiety precursor 4-diazirine-octynoic acid *I-N<sub>2</sub>COOH* (**38**) by coupling the methyl amine to form the *N*-methyl amide diazirine-alkyne *I-N<sub>2</sub>NHMe* prototype (**39**). It has been synthesised by Mrs. C. Landras Guetta by mixing the methylamine, DIPEA and HATU with *I-N<sub>2</sub>COOH* (**38**) for 3 h at room temperature to obtain *I-N<sub>2</sub>NHMe* (**39**) with a 61% yield (Scheme 7).

## 2. Photophysical studies

The photoactivatable moieties must absorb light to be activated which correspond to the Grotthus-Draper law. It was further precise by Stark and Einstein that a primary chemical reaction occurs when a molecule is activated by absorbing a single photon. After light absorption, a molecule is generated in its excited state and the process to return to the ground state can follow different pathways, depending on the properties of the specific molecule. Deactivation of an excited state can occur through non-radiative decay, intermolecular energy transfer, fluorescence (S1 -> S0), phosphorescence (T1 -> T0) or by reacting through the transient specie formed. It is the last possibility that interests us as it generates photoadducts. Thus, it is important to know the efficiency and properties of the photoactivatable moieties utilised, as well, to record the compound UV-visible spectra to identify any modification and band superpositions.

### 2.1. Photoactivatable moieties

The spectra of the **branched** precursors and prototypes have been recorded in acetonitrile and they show globally no difference between the precursors and the corresponding prototypes.

The benzophenone spectra in Figure 115 present an intense  $\pi \rightarrow \pi^*$  transition with a maximum at 258 nm more intense for **Bph-COOH** than **Bph-CONEt<sub>2</sub>**. They present also a weak  $n \rightarrow \pi^*$  transition of similar intensity with a maximum at 340 nm, generated by the overlapping of the  $\pi \rightarrow \pi^*$  transition, ranging from 320 nm to 390 nm.

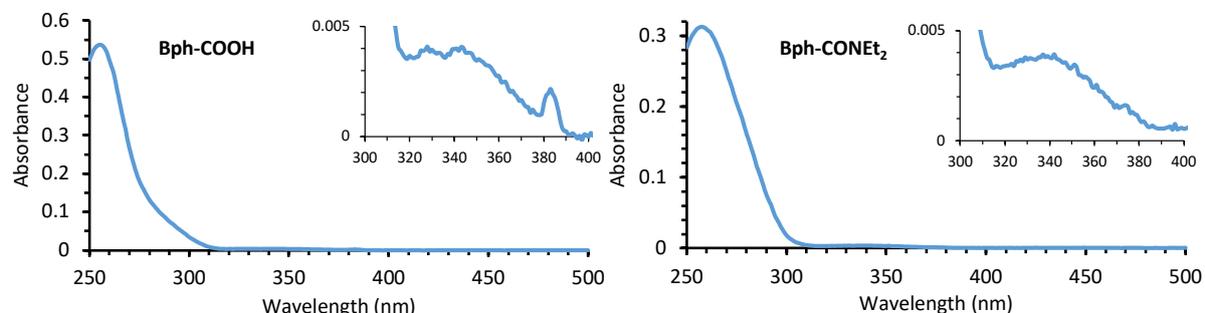


Figure 115: Spectra of the **Bph-COOH** precursor (left) and **Bph-CONEt<sub>2</sub>** prototype (right) at 20  $\mu$ M in acetonitrile.

The benzyl diazirines spectra in Figure 116 present an intense  $\pi \rightarrow \pi^*$  transition with a maximum centred at wavelengths lower than 250 nm and a hypsochromic shift for **N2-CONEt<sub>2</sub>**, and the weak  $n \rightarrow \pi^*$  transition with a maximum at 350 nm, ranging from 310 nm to 390 nm.

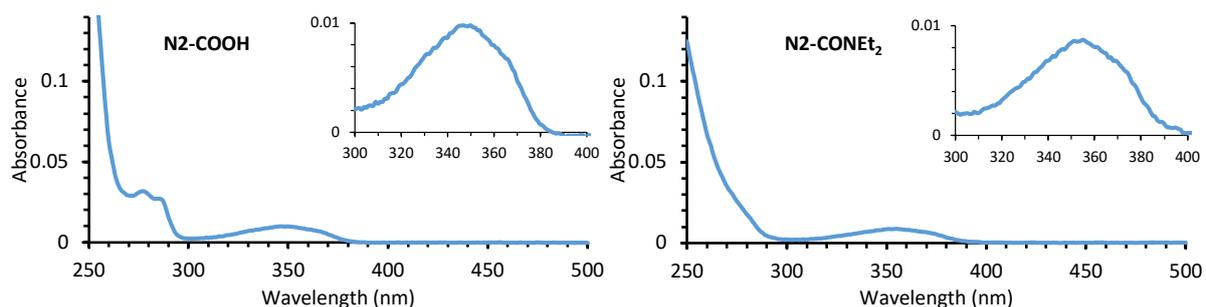


Figure 116: Spectra of the **N2-COOH** precursor (left) and **N2-CONEt<sub>2</sub>** prototype (right) at 20  $\mu$ M in acetonitrile.

The aryl azides spectra in Figure 117 is interesting due to the presence of a unique transition corresponding to the an intense  $\pi \rightarrow \pi^*$  transition with a maximum at 265 nm for **N3-COOH** and a hypsochromic shift for **N3-CONEt<sub>2</sub>** with a maximum at 255 nm, extending to 380 nm.

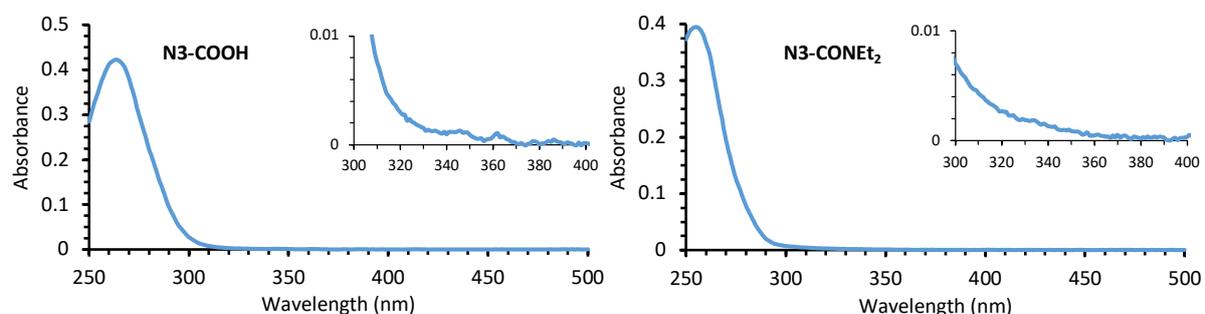


Figure 117: Spectra of the **N3-COOH** precursor (left) and **N3-CONEt<sub>2</sub>** prototype (right) at 20  $\mu$ M in acetonitrile.

The spectrum in Figure 118 of the *I-N<sub>2</sub>NHMe* prototype was also measured and shows an intense  $\pi \rightarrow \pi^*$  transition with a maximum at 210 nm and a weak  $n \rightarrow \pi^*$  transition peak with a maximum centred at 350 nm and ranging from 310 nm to 380 nm.

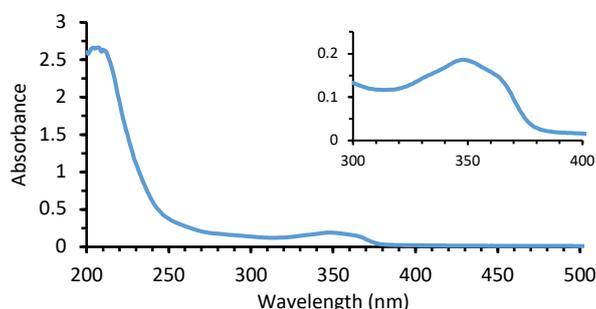


Figure 118: Spectrum of the *I-N<sub>2</sub>NHMe* prototype at 3 mM in acetonitrile.

## 2.2. In-line and branched compounds

The spectra of the **in-line** and **branched** compounds in Figure 119 are very similar and exhibits an intense  $\pi \rightarrow \pi^*$  transition with a maximum at 275 nm and a medium  $n \rightarrow \pi^*$  transition with a maximum at 350 nm with a peak ranging from the  $\pi \rightarrow \pi^*$  transition overlaps around 320 nm to 380 nm.

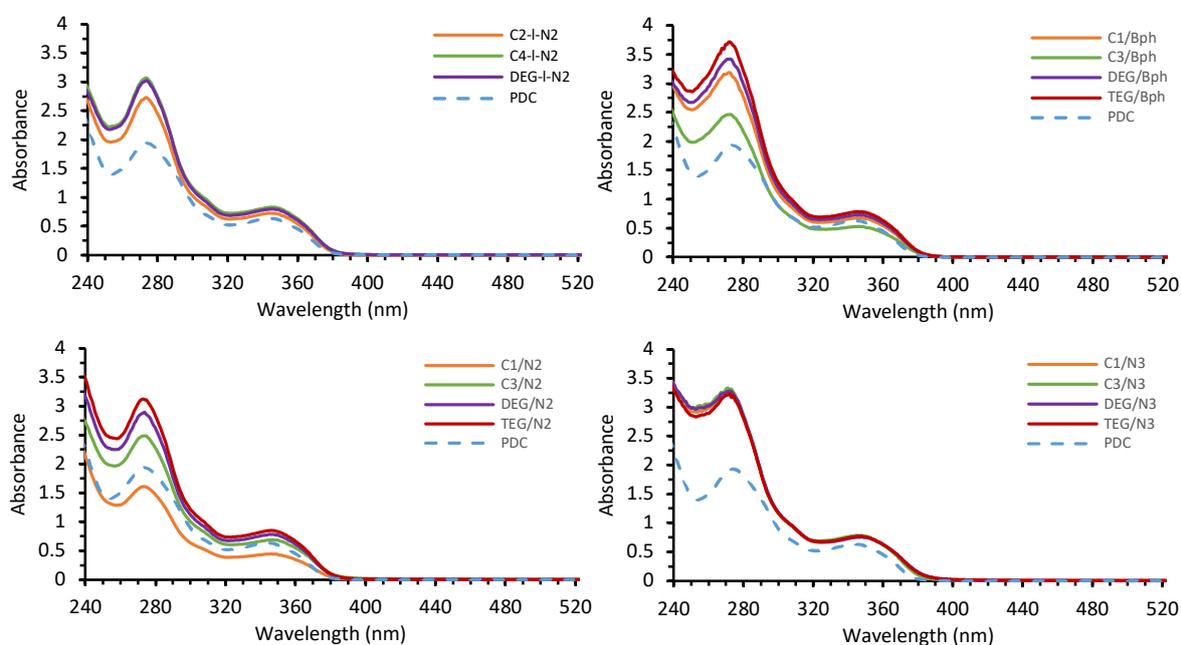


Figure 119: UV-visible spectra of **in-line** and **branched** compounds in water at 50  $\mu$ M. The PDC spectrum in water at 50  $\mu$ M is used as a reference.

To avoid irradiating endogenous chromophores, the irradiation target is the  $n \rightarrow \pi^*$  transition at 366 nm. However, the PDC strongly absorb in this region, as observed from the UV-visible spectrum in Figure 120, compared to the **in-line** and **branched** prototype. Nonetheless, biochemistry studies have shown that the PDC moiety was not preventing the activation of the photoactivatable moieties and the generation of the photoadducts with good conversion. This may be due to an energy transfer from the PDC chromophore to the photoactivatable moiety.

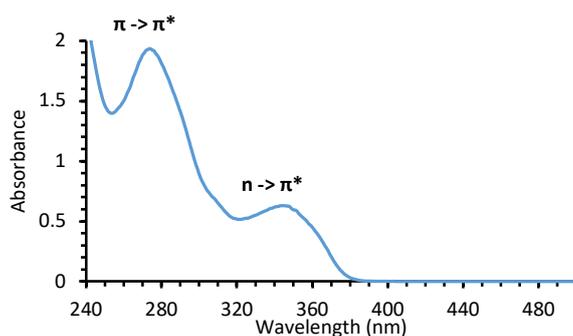


Figure 120: Spectrum of PDC at 50  $\mu$ M in water.

### 3. Laser flash photolysis studies

Laser Flash photolysis (LFP) is a spectroscopy method used to study the formation of transient species (e.g. radical intermediate or short-lived UV-vis-absorbing excited state) or reactive intermediate photochemically generated.<sup>236</sup> However, the formation of these species can be followed only if they absorb in the UV-visible region. LFP begins with a short pulse (<5 ns) of light generating the desired species, which relaxation to the ground state is followed by absorption spectroscopy recorded as a function of time, allowing us to observe the species decay. This technique is used to determine the physical/chemical properties, generation efficiency and reactivity of the desired species, and it can be applied to measure the reaction rate constant with spectroscopic-compatible agents by following the impact on the transient decay.

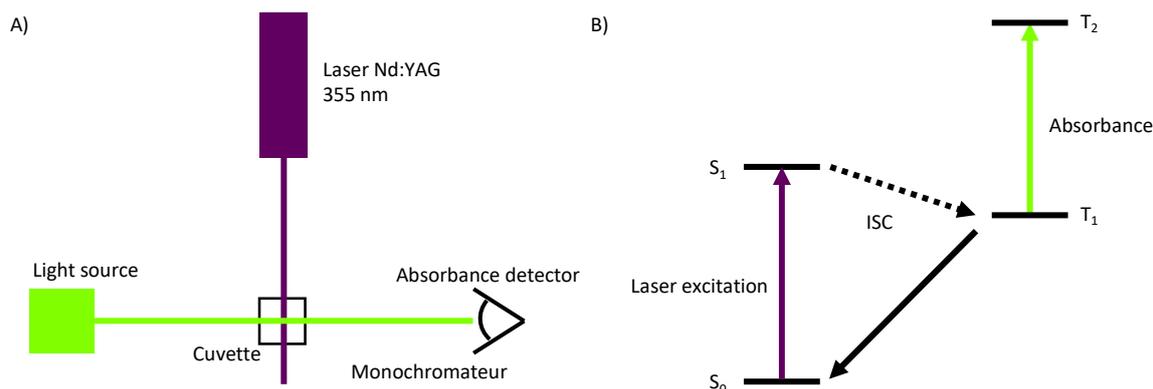


Figure 121: A) Representation of nanosecond Laser Flash Photolysis. A laser pulse irradiates the sample, and a perpendicular light source passing through the sample is recorded in function of time to record the change produced by the laser pulse activation. B) Example of activation of an absorbing excited species by laser excitation: after producing the single excited state ( $S_1$ ),  $S_1$  instantaneously convert by intersystem crossing (ISC) to the triplet state ( $T_1$ ) which absorbs UV light and can be follow through time.

Technically (Figure 121), a high energy laser with nanosecond pulses generates the transient species by populating higher energy electronic states. A second light source, the probe beam, is perpendicular to the laser and produces a continuous light source. The light passes through the sample then the monochromator selects the desired wavelength which is recorded by the detector. In our case, the detector can follow microsecond-to-second processes, which allows us to study different light-triggered processes, such as free radicals or triplet (long-lived) excited states, and to obtain information on lifetimes and reaction kinetics.

### 3.1. Photochemical investigation of benzophenone derivatives

Benzophenone is a well-known photoactivatable moiety already characterised by LFP: it forms a long-living triplet state and its reaction kinetics with nucleobases are known.<sup>180</sup> We decided to study the transient generated from the mimic benzophenone prototype (**Bph-CONEt<sub>2</sub>**) as well as the carboxylic acid benzophenone (**Bph-COOH**) and methylamide benzophenone (**Bph-CONHMe**) synthesised by Dr. A. Benassi (Figure 122).

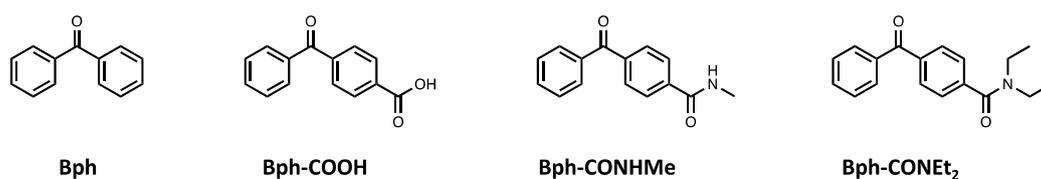


Figure 122: Molecular representation of the different benzophenone derivatives studied in this thesis.

### 3.1.1. Detection and characterisation of the reactive species

The **Bph-COOH** decay in water and in the absence of oxygen (i.e. deaerated by flushing with argon) (Figure 123) shows the formation of a long living ( $t > 10 \mu\text{s}$ ) transient species similar to the one of the benzophenone.<sup>180</sup> The transient decay does not return at an absorbance null but shows the generation of a second more stable species. The first species possess a maximum at 545 nm and one at wavelength higher than 750 nm corresponding to the triplet state of the **Bph-COOH**. This triplet assignment was confirmed by the quenching observed in presence of oxygen (i.e. without argon flushing).<sup>180</sup> The triplet quenching is due to the triplet-triplet interaction between the triplet transient and the oxygen triplet state ( $^3\text{O}_2$ ), which results in the de-excitation to the ground state of **Bph-COOH** and in the excitation of the oxygen molecule to the singlet state ( $^1\text{O}_2$ ). The second long-living species can be attributed to the ketyl radical formation with the maxima centred at 575 nm and 725 nm, which is very similar to that of the benzophenone ketyl radical.<sup>180</sup>

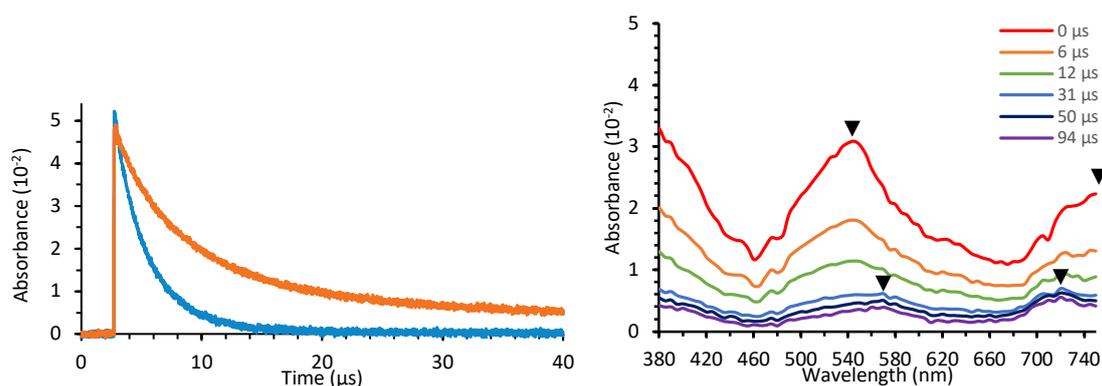


Figure 123: Left) Transient species decay of the **Bph-COOH** precursor (3.5 mM) without (orange) and with (blue) oxygen recorded at 540 nm. Right) Full spectrum of the **Bph-COOH** precursor (3.5 mM) without oxygen recorded in function of the time: the triplet state with the maxima at 545 nm and  $>750$  nm decays after tens of microseconds, and generates the ketyl radical transient with maxima at 575 nm and 725. The black arrows highlight the maxima.

Similarly, the **Bph-CONHMe** decay (Figure 124), recorded in acetonitrile without oxygen, presents a long living ( $t > 10 \mu\text{s}$ ) triplet excited state transient with maxima at 550 nm and  $> 740$  nm, and a more effective quenching in the presence of oxygen compared to **Bph-COOH**. The Ketyl radical is still observed with maxima at 575 nm and 725 nm.

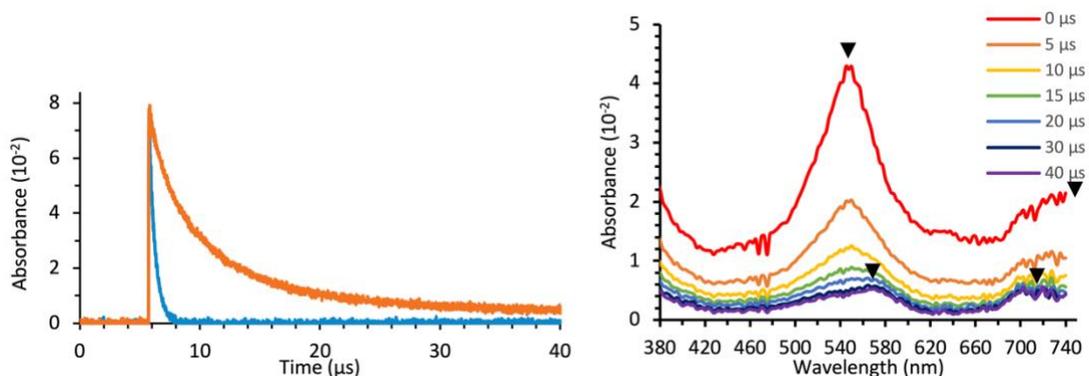


Figure 124: Left) Transient species decay of the **Bph-CONHMe** derivative (4 mM) without (orange) and with (blue) oxygen recorded at 540 nm. Right) Full spectra of the **Bph-CONHMe** derivative (4 mM) recorded in function of the time in the absence of oxygen. The triplet state with maxima at 545 nm and >740 nm decays after tens of microseconds, and generates the ketyl radical transient with maxima at 575 nm and 725 nm. The black arrows highlight the maxima.

On the contrary, the benzophenone prototype **Bph-CONEt<sub>2</sub>** decay (Figure 125) exhibits a very short lifetime ( $\tau_{1/2} < 50$  ns) without oxygen, and addition of oxygen has no effect on the transient decay. This could indicate that the generated transient is not a triplet state, but the lack of the oxygen effect could also be due to the too short living state. In addition, the spectrum is recorded at the limit of the LFP capacity, in the range of the laser pulse (10 ns), which explains this bell-shaped curve instead of an exponential decay. Thus, it is not possible to confirm that the short living transient is a triplet state. Interestingly, only the *N,N*-disubstituted amide is strongly affected by its conjugation compared to the *N*-monosubstituted amide. This difference could be explained by the rotor thermal de-excitation process.

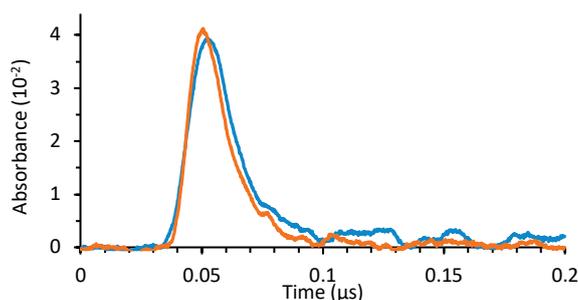


Figure 125: Transient species decay of **Bph-CONEt<sub>2</sub>** prototype (3.5 mM) without (orange) and with (blue) oxygen in acetonitrile recorded at 540 nm.

### 3.1.2. Reactivity study with nucleosides

As a photoactivatable moiety, Benzophenone is known to be thymidine-selective, forming a substituted oxetane derivative. Similarly, the biochemistry results indicate that the **branched** benzophenone conserves this thymidine-selectivity over the rest of the nucleobases (see section III.2.3.2, p. 127). However, the benzophenone **Bph-CONEt<sub>2</sub>** prototype cannot be used to measure the rate constant of the transient quenching reaction by nucleobases due to its too short lifetime. That is why, the rate constant has been evaluated for **Bph-COOH** and **Bph-CONHMe**. The quenching decay by a nucleobase can be followed by LFP and results in a shortening of the transient half-life time. The second order rate constant for the reactivity with a nucleoside can be calculated using the pseudo-first order principle considering the generated transient concentration being negligible compared to the nucleoside concentration. Thus,  $k_{obs} = \ln(2)/\tau_{1/2}$  can be calculated for each nucleoside concentration by measuring by LFP the transient half-life time. Finally, the second order rate constant  $k_2$  corresponds to the slope of the  $k_{obs}$  vs [nucleoside] curve. **Bph-COOH** triple state quenching was measured at 3.5 mM titrated with a rising concentration of nucleoside from 0.2 to 1.4 mM.

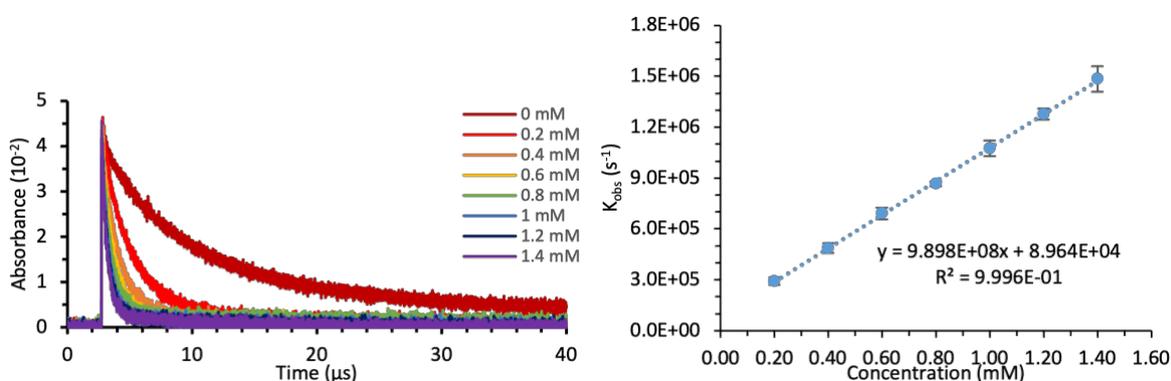


Figure 126: Left) Decay quenching of the **Bph-COOH** triple state recorded in function of the increasing concentration of thymidine (dT) in aqueous solution. Right) Plot of the  $K_{obs}$  versus the concentration. The  $K_{obs}$  is calculated using the half-life time obtained from the quenching decays. The slope of the linear curve  $K_{obs}$  versus concentration correspond to the total rate quenching constant  $k_2$ .

The quenching with the thymidine (dT) represented in Figure 126 is very efficient with an effective quenching for a thymidine concentration starting at 0.2 mM. The computation of the second order rate constant gave a fast quenching of  $k_{dT} = 9.90 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$ , which is 2.6-fold slower than the quenching rate measured with unsubstituted benzophenone.<sup>180</sup> The quenching was also measured with deoxycytidine (dC), deoxyuridine (dU) and uridine (U) which is summarise in Figure 127.

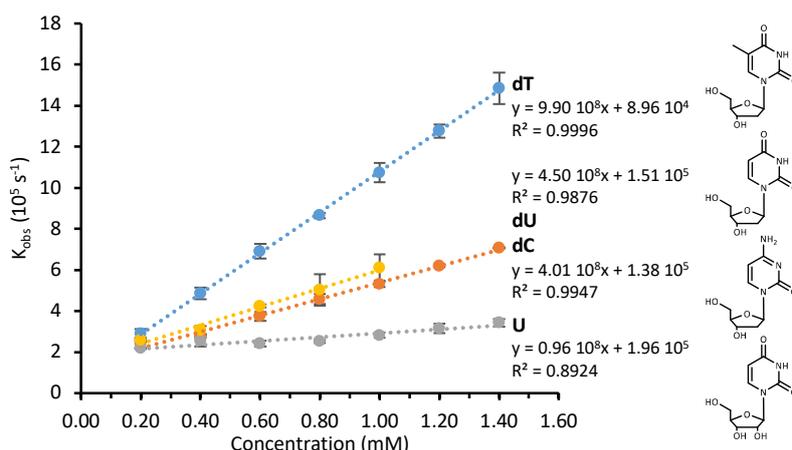


Figure 127: Summary of the total rate constants for the quenching of the **Bph-COOH** precursor (3.5 mM) triple state by thymidine (**dT**), 2'-deoxycytidine (**dC**), 2'-deoxyuridine (**dU**), and uridine (**U**) in aqueous solution ( $n = 3$ ). The  $K_{obs}$  vs concentration slope of the linear plot corresponds to the total rate quenching constant  $k_2$ . The **dU** solubility is limited in water to 1 mM, thus, no data above this concentration has been measured.

The **dT** quenching rate constant can be compared with **dC** and **dU** with a value of respectively  $k_{dC} = 4.01 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$  and  $k_{dU} = 4.50 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$ , which is roughly half the quenching rate of **dT**. **dC** and **dT** structurally differ only from a methyl in position 5 of the pyrimidine, but its presence results in an increase of the quenching rate by more than two, which highlight the importance of the 5'-methyl to stabilise the radical during the [2+2] cycloaddition. The uridine quenching rate was measured with a value of  $k_U = 0.96 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$ , which is five times lower than with **dU**. Interestingly, the 2' position seems to take part of a reactional pathway that is inhibited in the presence of the 2'-hydroxyl, and not negligible to understand the total reactivity of the benzophenone. However, the main pathway remains with thymidine and its [2+2] cycloaddition into the oxetane product. Similarly, the same experiments were realised with **Bph-CONHMe** (Figure 128), which shows a similar trend with the subsequent selectivity: **dT** ( $k_{dT} = 4.12 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$ ) > **dC** ( $k_{dC} = 1.87 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$ ) > **U** ( $k_{dU} = 0.68 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$ ). However, the reactivity is decreased by two-fold for **dT** and **dC**, and slightly less for **U**.

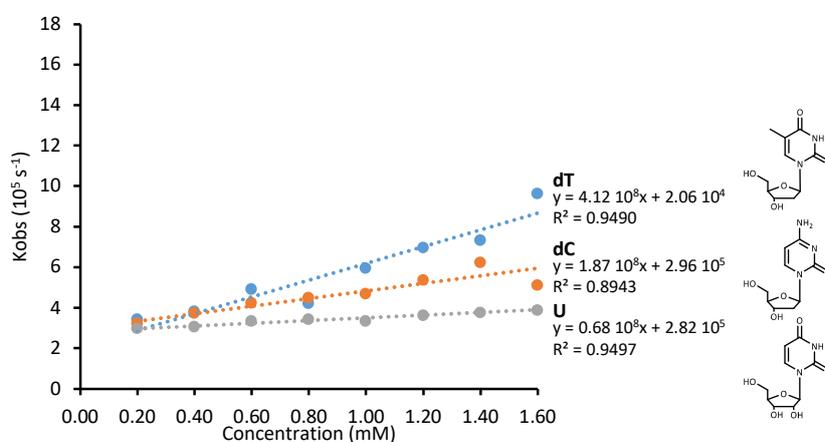


Figure 128: Summary of the total rate constants for the quenching of **Bph-CONHMe** (4 mM) triplet state by thymidine (**dT**), 2'-deoxycytidine (**dC**), and uridine (**U**) in acetonitrile ( $n = 1$ ). The  $K_{obs}$  vs concentration slope of the linear plot corresponds to the total rate quenching constant  $k_2$ .

In these experiments, only four nucleobases have been evaluated for their quenching properties and more precisely, only pyrimidines have been evaluated. In fact, for purines (**dA** and **dG**), **Bph-COOH** triplet excited state was quenched very rapidly ( $t < 1 \mu\text{s}$ ) generating a secondary transient with a maximum absorbance at 570 nm (Figure 129), which can be assigned, by comparing with previously reported data,<sup>4</sup> to the formation of **Bph-COOH** ketyl radical resulting from an overall H-abstraction. Therefore, it was not possible to evaluate the quenching with purines due to the too fast formation of the ketyl radical, which overlaps with the **Bph-COOH** transient at 540 nm.

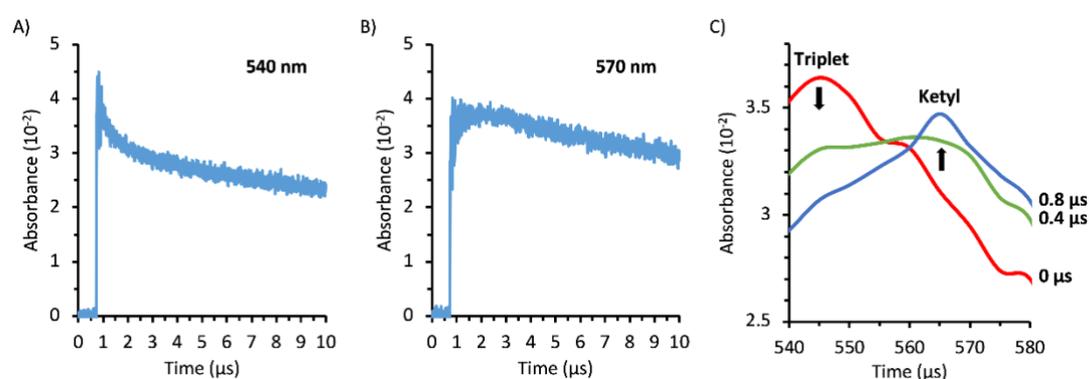


Figure 129: A) **Bph-COOH** (3.5 mM) triplet state transient decay recorded without oxygen at 540 nm in the presence of **dG** (0.5 mM). B) Generated ketyl radical decay monitored at 570 nm. C) Spectra between 540 nm and 580 nm recorded in function of the time. **Bph-COOH** triplet state is the main species observed just after irradiation (0  $\mu\text{s}$ ), then a rapid conversion is observed with a mixture at 0.4  $\mu\text{s}$  and the main species become the ketyl radical at 0.8  $\mu\text{s}$ .

### 3.2. Photochemical characterisation of the **In-line** prototype

The **in-line**  $I\text{-N}_2\text{NHMe}$  prototype forms a carbene upon irradiation, which is not UV active and thus cannot be directly detected by LFP. However, it is possible to trap the carbene by using pyridine to form the corresponding pyridine ylide (Figure 130).<sup>237</sup>

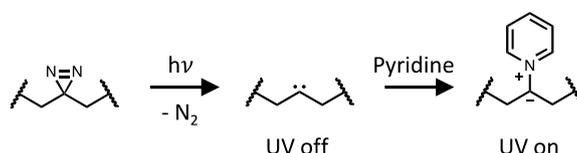


Figure 130: Mechanism to trap the carbene using pyridine.

The decay of the pyridine ylide from  $I\text{-N}_2\text{NHMe}$  was recorded in acetonitrile with an excess of pyridine. It showed the formation of a short living transient ( $< 50$  ns), which has been assigned to the related pyridine ylide (Figure 131), based on a striking similarity of the absorption spectra to a similar ylide.<sup>238</sup> This short life time is at the limit of the LFP device, which explains the bell-shaped curve.

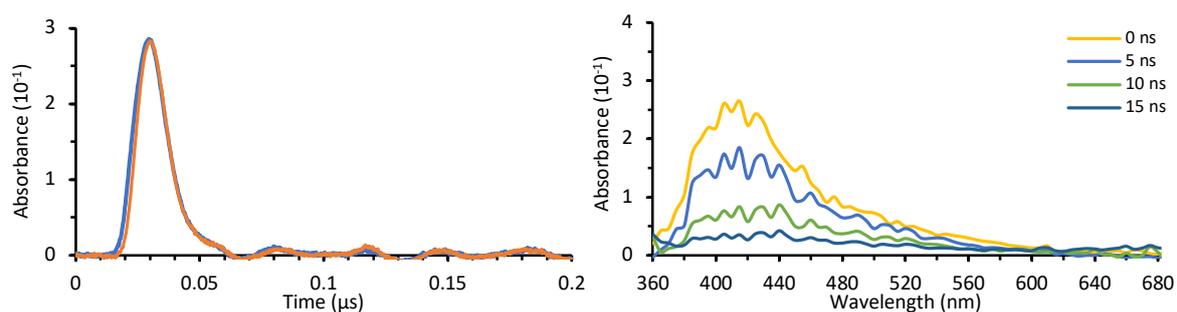


Figure 131:  $I\text{-N}_2\text{NHMe}$  (6 mM) transient decay without (orange) and with (blue) oxygen recorded at 425 nm. Right) Full spectrum of  $I\text{-N}_2\text{NHMe}$  (6 mM) without oxygen recorded in function of time. This presents a band between 370 – 460 nm with a maximum at 425 nm corresponding to the pyridine ylide, which decay after tens of nanoseconds.

## 4. Benzophenone photoadducts distributions

### 4.1. Product distribution

The reaction of **Bph-COOH** and **Bph-CONEt<sub>2</sub>** with thymidine was followed by HPLC. The mixture of **Bph-COOH** (3 mM) and **dT** (3 mM) was flushed with argon and irradiated with four lamps centred at 366 nm in a multirays photoreactor for 1 h. The reaction presents a rapid conversion with 40% of **dT** and 66% of **Bph-COOH** consumed to generate new products (Figure 132-in blue). The difference in consumption between the **dT** and **Bph-COOH** is due to a fraction of **Bph-COOH** being reduced. This was confirmed by the photolysis of **Bph-COOH** alone in a 1:1 mixture water:isopropanol that generated three new peaks (Figure 132-in red) with two having a faster retention time and one a slower retention time compared to **Bph-COOH** peak. These three products were not isolated and characterised but should correspond, in theory, to the reduction in alcohol, and dimerisation of the ketyl radical. This reaction should produce a couple of diastereoisomers, corresponding to two peaks at 24.5 min and 25.5 min in the chromatogram reported in Figure 132. We still observe the formation of these three adducts in the reaction between **dT** and **Bph-COOH**. In the reaction mixture, we detected the formation of at least 7 additional photoadducts.

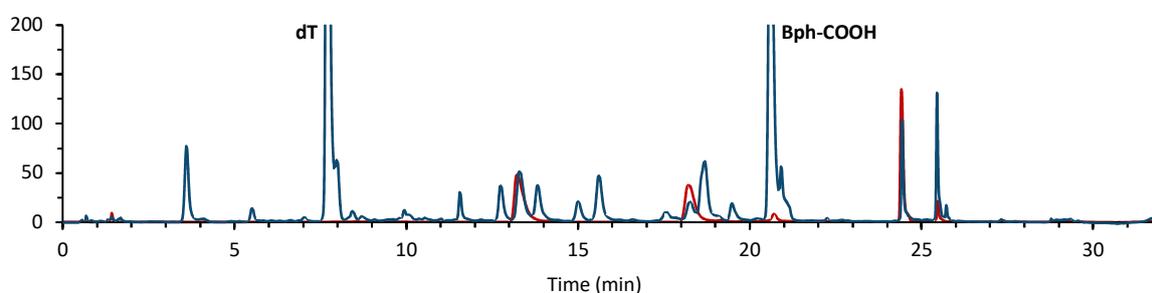


Figure 132: HPLC chromatogram (blue) of the reaction mixture between **thymidine** (3 mM) and **Bph-COOH** (3 mM) after 1 h irradiated at 366 nm. The reduction (red) of **Bph-COOH** (3 mM) was studied in water/isopropanol 1:1 upon irradiation for 1 h at 366 nm.

Preliminarily, the same condition has been applied for the reaction between **Bph-CONEt<sub>2</sub>** and **dT**, but we did not detect any photoadducts. Considering the shorted lifetime of the reactive **Bph-CONEt<sub>2</sub>** excited state, we decided to modify the reaction condition by increasing **dT** concentrations from 10 mM to 100 mM and extending the irradiation time to 38 h. The resulting chromatogram (Figure 133) exhibits seven or six peaks when flushed (red) or not (blue) with argon, respectively. We have decided to isolate and characterise the adducts corresponding to the seven peaks, generated by the reaction between **Bph-CONEt<sub>2</sub>** and thymidine, flushed with argon before irradiation.

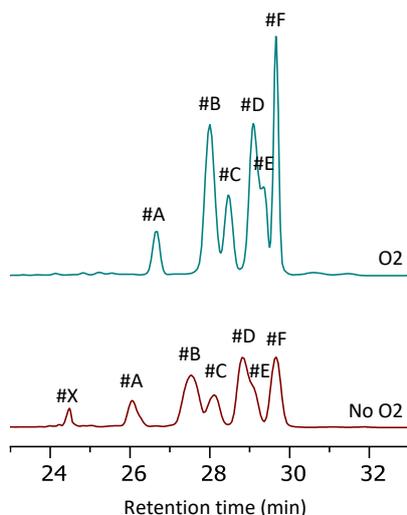


Figure 133: HPLC chromatogram of the reaction mixture between **Bph-CONEt<sub>2</sub>** (10 mM) and **dT** (100 mM) after 38 h irradiated with 4 lamps at 366 nm in the presence (top) or absence (bottom) of oxygen.

## 4.2. Structural characterisation of the photoadducts

The mixture of products was purified as one and <sup>1</sup>H NMR spectrum of the mixture was recorded as displayed in Figure 134-B. The main change in comparison to the <sup>1</sup>H NMR spectrum of **dT** (Figure 134-A) is the strong shielding of the **H6** protons to the upfield 6.00 – 5.00 ppm region, with eight new singlets and a weaker shielding of the **Me7** methyl group, formed after irradiation. This <sup>1</sup>H NMR features suggest a mixture of eight new oxetanes.

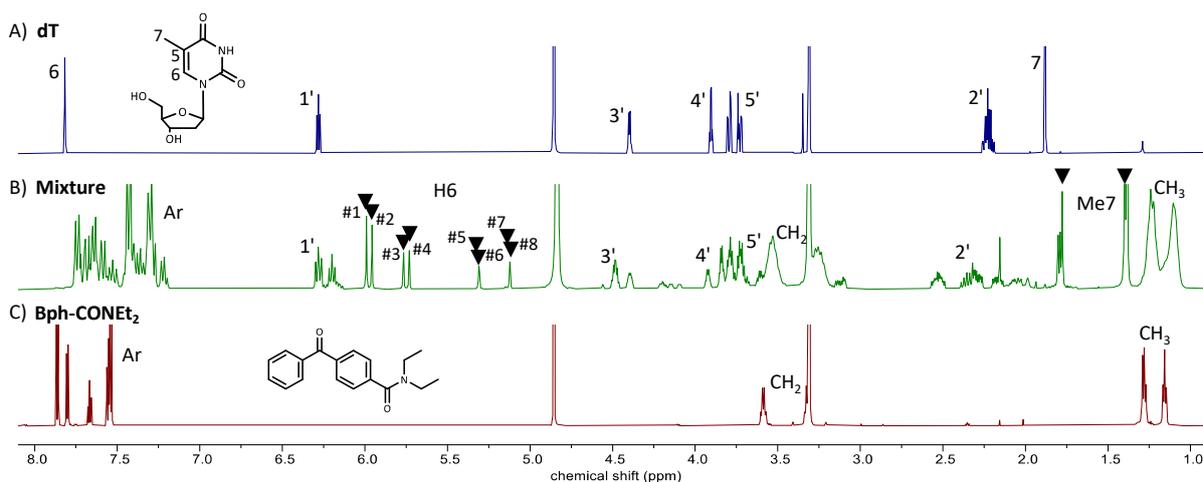


Figure 134: <sup>1</sup>H NMR spectra of: A) **thymidine**, B) mixture of products (**#A-#X**) purified from the reaction between **Bph-CONEt<sub>2</sub>** and **dT**, and C) **Bph-CONEt<sub>2</sub>**.

A closer look reveals eight **H6** protons identified in the 6.00 – 5.00 ppm oxetane region and their integration sum up for 1 (normalised on the reference anomeric protons as shown in Figure 135), which indicates the formation of two regioisomers, each of them existing in four different stereoisomers. The oxetanes obtained are a family of two regioisomers that differ depending on the oxygen position: the "hemiaminal ether" oxetane, exhibiting **H6** resonance downfield at 6.00 – 5.70 ppm and the "ether" oxetane with **H6** signals upfield at 5.40 – 5.00 ppm. The main oxetanes are the hemiaminal ethers with 71%, and only 29% for the ether oxetanes. This may be explained as the hemiaminal ether isomer reduces the steric clash with the 2'-deoxyribose, and the ternary C<sub>5</sub> carbon stabilises more the intermediate radical than the secondary C<sub>6</sub> carbon.

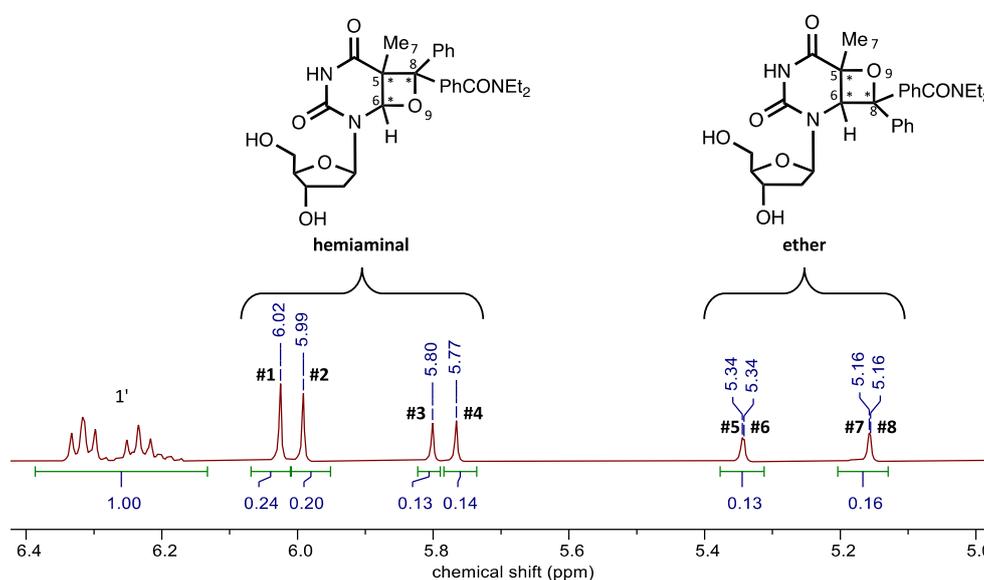


Figure 135: Zoom on the 1' and oxetane **H6** region of the <sup>1</sup>H NMR spectrum of the purified mixture (#A-#X) from **Bph-CONEt<sub>2</sub>** and **dT** reaction.

In addition, the peaks are again divided into two groups of two peaks, i.e., group at 6.00 ppm composed of subdivided peaks #1 and #2 and the group at 5.79 ppm composed of subdivided peaks #3 and #4. The division into two groups is due to the absolute configuration of the C<sub>6</sub>-H and C<sub>5</sub>-Me<sub>7</sub>, i.e. (5R, 6S) or (5S, 6R), as previously reported in the literature.<sup>239</sup> The subdivided peaks (6.02 vs 5.99 ppm and 5.80 vs 5.77 ppm) are due to the third stereocenter (C<sub>8</sub>) in the oxetane ring bound to the different aryl groups (-phenyl (Ph) and -PhCONEt<sub>2</sub>). Each HPLC peak has been purified and <sup>1</sup>H NMR spectra were recorded as shown in Figure 136, except for #X, which does not present an oxetane profile and will be describe later. To define the relative and absolute (using the configuration of the 2'-deoxyribose protons as reference)<sup>240</sup> configuration of each oxetane isomer, Nuclear Overhauser Effect Spectroscopy (NOESY) experiments were performed on each adducts from purified HPLC peaks and are resumed in Figure 137 centred on the **H6** protons.

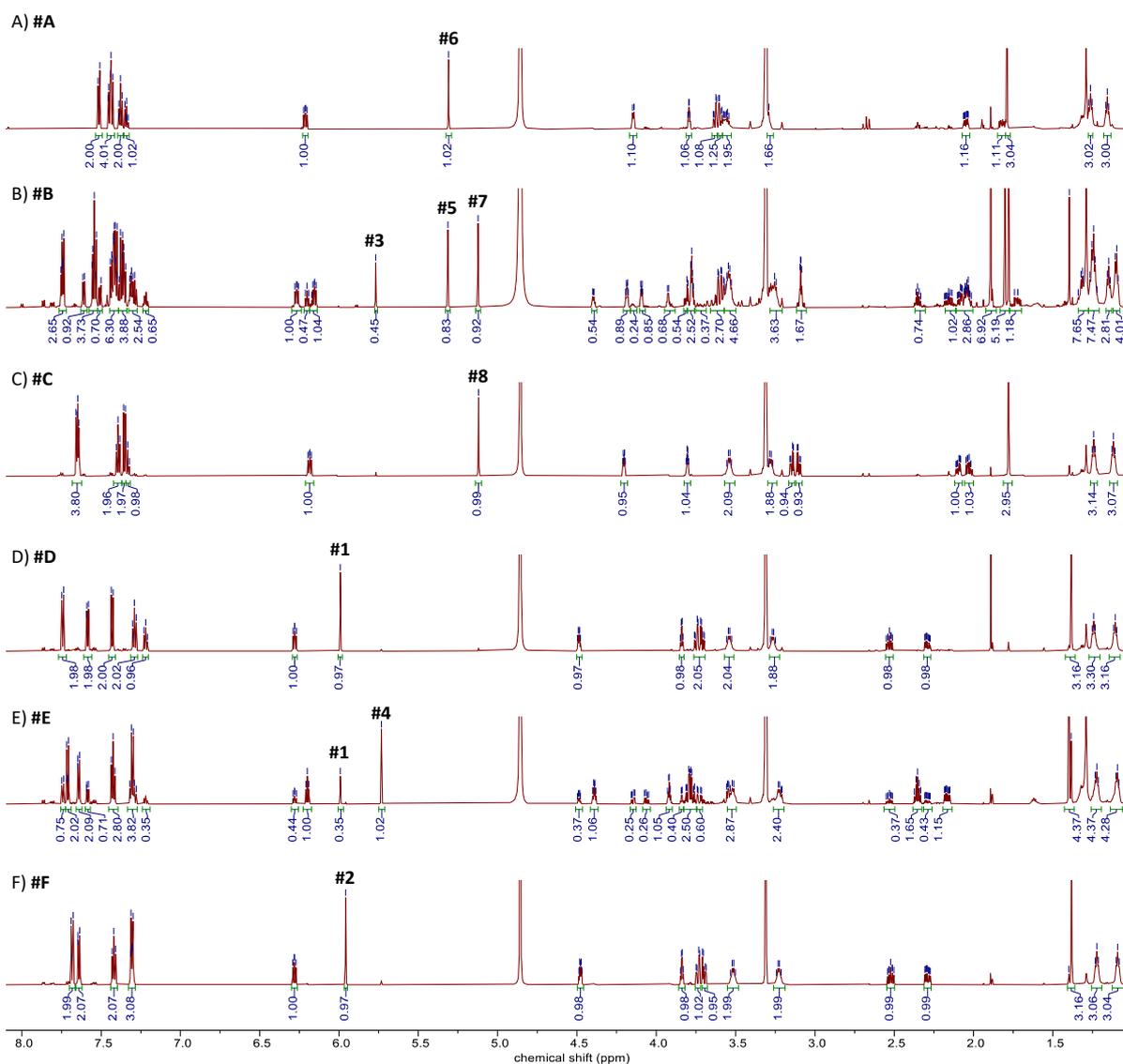


Figure 136:  $^1\text{H}$  NMR spectra (700 MHz) of the six HPLC peaks - A) #A, B) #B, C) #C, D) #D, E) #E, F) #F) - comprising oxetane **H6** peaks.

NOESY experiments allow to identify protons that interact together by correlating through space. It means that the two protons must be at a spatial proximity to correlate without requiring them to be bound at proximity. We focused the NOESY on the characteristic **H6** proton expecting to identify the stereochemistry of the **C<sub>6</sub>-H/C<sub>5</sub>-Me7** carbon atom vs. the 2'-deoxyribose carbons vs. the C atom bound to the aryl groups (Ph or PhCONEt<sub>2</sub>). As expected for a [2+2] Paternò-Büchi cycloaddition, **H6** and **Me7** correlate together indicating that the two groups exhibit a *cis* relative configuration. However, it was not possible to define the absolute configuration for **C<sub>6</sub>-H/C<sub>5</sub>-Me7** from NOESY experiments as the eight **H6** protons correlate with the 2'-deoxyribose protons.<sup>240</sup> For the aryl relative configuration, it was possible to assign the *cis/trans* configuration from the spatial correlation with the **H6** proton, excepted when aromatic peaks were overlapping such as for the mixture of oxetanes **#3/#5/#7** and for oxetane **#8**.

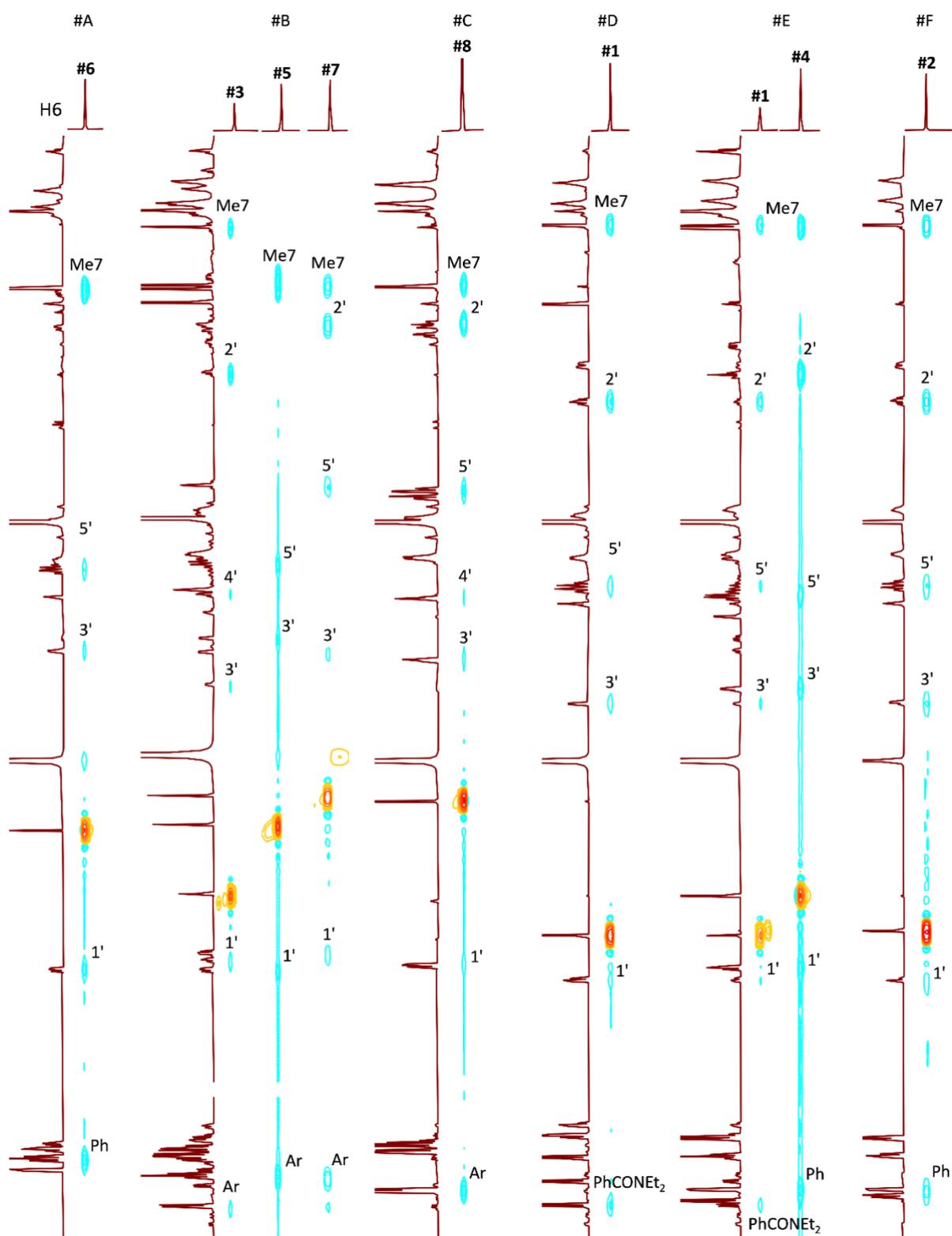


Figure 137: NOESY experiments (700 MHz NMR) for each purified [2+2] adduct centred on the **H6** proton.

Nevertheless, according to the assignments published by Miranda for the hemiaminal ether oxetanes obtained between benzophenone and thymidine,<sup>239,240</sup> the (5R, 6S) stereoisomer exhibits the most deshielded C<sub>6</sub>-H proton, among the [2+2] photoadducts. Therefore, compounds **#1** and **#2** (Figure 135) should exhibit the same (5R, 6S) configuration. Additionally, we have been able to assign the absolute R/S configuration to the C<sub>8</sub>-Aryl carbon for the hemiaminal ether thanks to their relative *cis/trans* configuration vs. **H6/Me7** groups. The information obtained from the analysis of the NMR spectra and NOESY experiments is summarised in Table 4.

HPLC peak	Adduct	Regioisomers	C <sub>6</sub> -H/C <sub>5</sub> -Me7 configuration	H6 vs Aryl spatial coupling	J <sub>H6</sub>
#A	<b>#6</b>	Ether	*cis	Ph	0.06**
	<b>#3</b>	Hemiaminal	5S, 6R	n.d.	0.13
#B	<b>#5</b>	Ether	*cis	n.d.	0.06**
	<b>#7</b>	Ether	*cis	n.d.	0.08**
#C	<b>#8</b>	Ether	*cis	n.d.	0.08**
#D	<b>#1</b>	Hemiaminal	5R, 6S	PhCOEt <sub>2</sub>	0.24
#E	<b>#4</b>	Hemiaminal	5S, 6R	Ph	0.14
#F	<b>#2</b>	Hemiaminal	5R, 6S	Ph	0.20

Table 4: Summary of the NMR data analysis for each adduct from the purified HPLC peaks. n.d.: not determinate. \* Only the relative *cis/trans* configuration is indicated obtained from NOESY experiments as it was not possible to determine the absolute configuration. \*\* The integral was divided by two by considering each **H6** peaks been of equal intensity.

From these results, we can attribute for **#3** and **#5** the opposite *cis* configuration between the **H6** and the aryl groups from their corresponding C<sub>8</sub> diastereoisomers **#4** and **#6**. The latter oxetanes **H6** protons are in *cis* configuration with the phenyl groups, thus, **#3** and **#5**'s **H6** protons are in *cis* configuration with the PhCONEt<sub>2</sub> group. In addition, we can observe a pattern with the *cis* PhCONEt<sub>2</sub> downfield and *cis* phenyl upfield, which we assume should be similar for **#7** and **#8** as summarised in Figure 138.

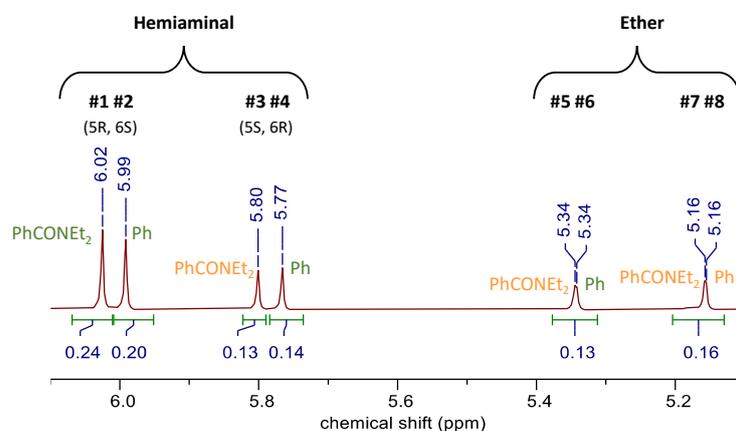


Figure 138: Summary of the assigned peaks based on knowledge with in green the experimental results and in orange the ones-assigned by elimination.

Overall, it has been possible to assign the absolute configurations of the four hemiaminal ether oxetanes, as reported in Figure 139.

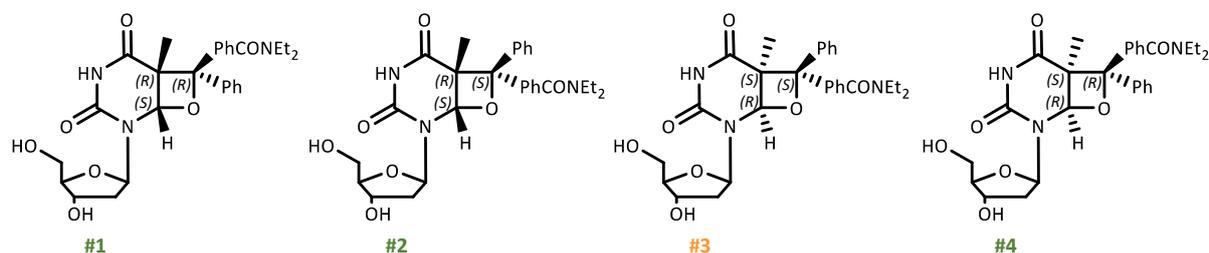


Figure 139: Hemiaminal ether oxetane structures with their absolute configuration.

Oxetane are known to be thermally reversible and biochemical experiments from Chapter III (see section III.2.3.2, p. 127) show mostly the generation of thermally reversible photoadducts as well as some irreversible photoadducts. To confirm the thermal reversibility, the pure hemiaminal ether oxetane **#2** collected as HPLC fraction #F and pure ether oxetane **#6** collected as HPLC fraction #A were heated in water at 95 °C for 2 h. The results are presented in Figure 140. Hemiaminal ether oxetane **#2** undoubtedly shows thermal reversibility with regeneration of **Bph-CONEt<sub>2</sub>** and **dT**, as expected for a reverse [2+2] cycloaddition. However, the ether oxetane **#6** is a thermally stable adducts.

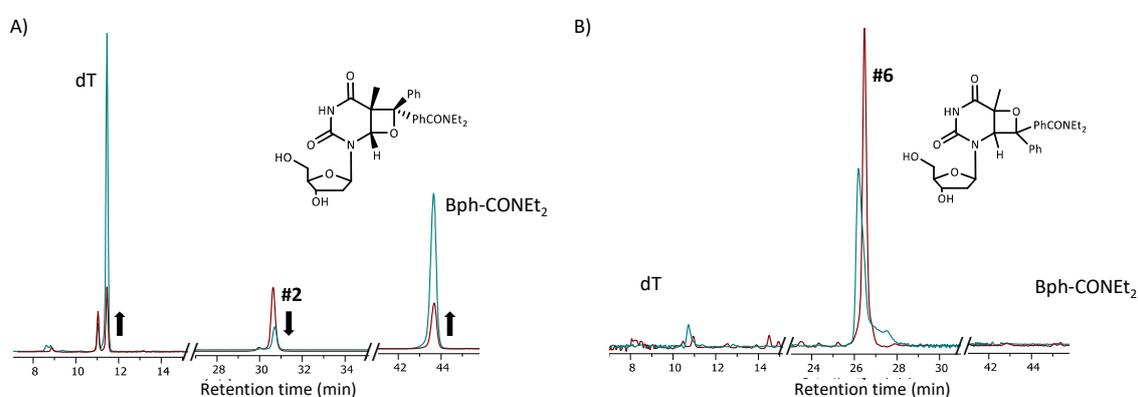


Figure 140: HPLC chromatogram of A) hemiaminal ether **#2** and B) ether **#6** before (red) and after (blue) heating treatment for 2 h at 95 °C.

$^1\text{H}$  NMR spectrum (Figure 141) of the adduct from the HPLC fraction #X does not present the characteristic oxetane chemical shift. It was obtained in low amount, and therefore, difficult to fully characterise. The 2'-deoxyribose peaks (1', 2', 4' and 5') are still present with the 5' peaks deshielded. **Bph-CONEt<sub>2</sub>** aromatic and ethyl groups are still present, which confirms the formation of a photoadduct.

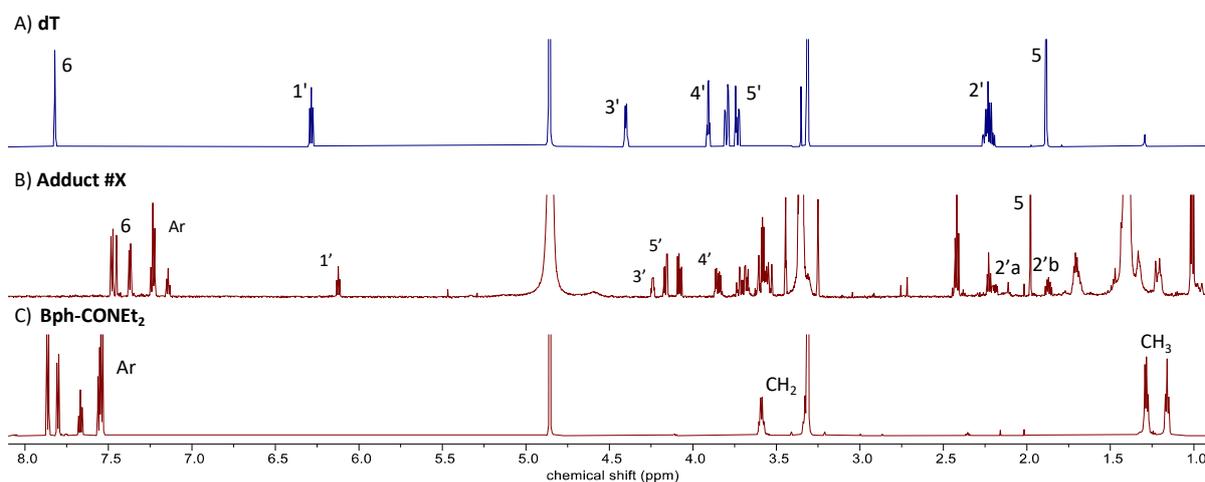


Figure 141: NMR spectra (700 MHz) of the A) thymidine, B) adduct #X and C) **Bph-CONEt<sub>2</sub>** prototype.

It is possible that this photoadduct corresponds to the insertion of the prototype transient into the 5'-OH bond as the signal shows an important downfield shift, thus, it could be the result of an O-H insertion into the 5'-OH bond (Figure 142). However, the concentration prevents to record a carbon spectrum and  $^1\text{H}$ - $^{13}\text{C}$  2D experiments (HSQC/HMBC) even by using a 700 MHz NMR, making it impossible to confirm our hypothesis using the carbon chemical shift to confirm the hemiketal formation and HMBC to confirm the 5'-OH insertion through coupling between the two H5' protons and the hemiketal quaternary carbon.

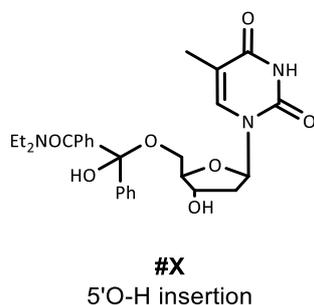


Figure 142: Hypothesised structure of photoadduct #X isolated by HPLC.

## 5. Study of the photoreactivity of aliphatic diazirine

### 5.1. Photolysis study

The reactivity of the  $\beta$ - $\text{N}_2\text{COOH}$  precursor and the  $\beta$ - $\text{N}_2\text{NHMe}$  prototype were investigated using GC-MS by adapting the protocol of O'Brien *et al.*,<sup>233</sup> which have studied similar ester and  $N$ -monosubstituted amide aliphatic diazirines (60 mM) in methanol irradiated at 350 nm for 2 to 16 h. Their photolysis experiments, resumed in Figure 143, showed mainly the formation of the alkene reduction product by [1,2] H-shift and hydroxyl product by O-H insertion from methanol, and the lactone product by lactonisation uniquely with the ester derivative for the latter.

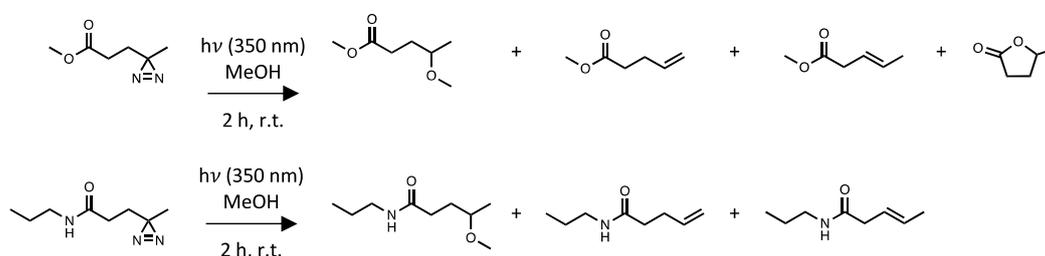


Figure 143: Results reported by O'Brien *et al.*,<sup>233</sup> products were identified after 2 h irradiation at 350 nm of ester (top) and  $N$ -monosubstituted amide (bottom) aliphatic diazirine (60 mM) in methanol.

We have adapted this methodology to study the photolysis of the  $\beta$ - $\text{N}_2\text{COOH}$  precursor and the  $\beta$ - $\text{N}_2\text{NHMe}$  prototype at 5 mM in methanol and/or water and irradiating at 366 nm for 10 to 30 min. Using O'Brien *et al.* results, it was possible to predict the results taking in account the additional terminal alkyne, which is the main difference between our and O'Brien *et al.* aliphatic diazirines. Indeed, the carbene could react with the alkyne through an intramolecular cyclisation to form a cyclopentene, which would prevent post-covalent functionalisation by click chemistry. The different photoproducts we might obtain according to O'Brien *et al.* are described in Figure 144. The reactions were analysed by GC-MS using electrospray ionization method, which generated a high level of fragmentations with loss of the reference molecular ion for small molecules. Thus, it is not possible to attribute the product from the reference molecular ion, instead, they were identified depending on their fragmentation profile, retention time, and predicted products.

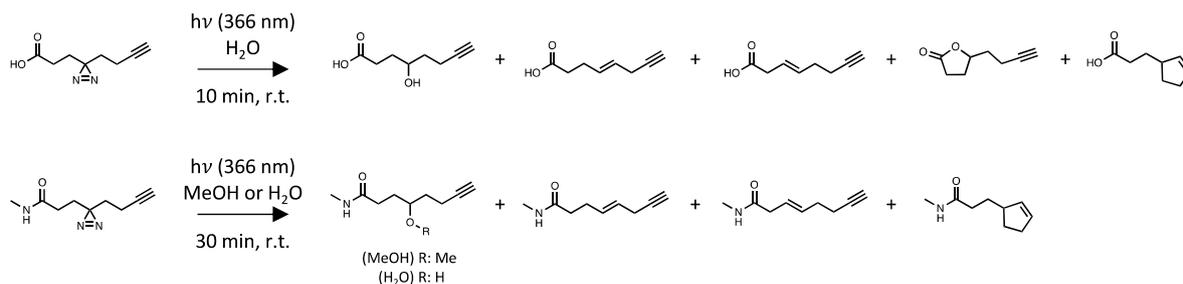


Figure 144: Theoretical products we might obtain after photolysis of *I-N<sub>2</sub>COOH* precursor and *I-N<sub>2</sub>NHMe* prototype according to O'Brien et al.,<sup>233</sup> and the cyclopentene product from the intramolecular cyclisation with the alkene.

Photolysis of the *I-N<sub>2</sub>COOH* precursor in water shows the formation of a unique product with a retention time of 9.64 min (Figure 145), with almost total consumption of the precursor after 10 min of irradiation. The product presents the absence of the carboxylic acid fragment indicating the formation of the lactone product. No O-H insertion with the water was observed, indicating that *I-N<sub>2</sub>COOH* precursor is not a good candidate to isolate photoadducts for photoaffinity labelling as the main and only pathway seems to be the intramolecular lactonisation.

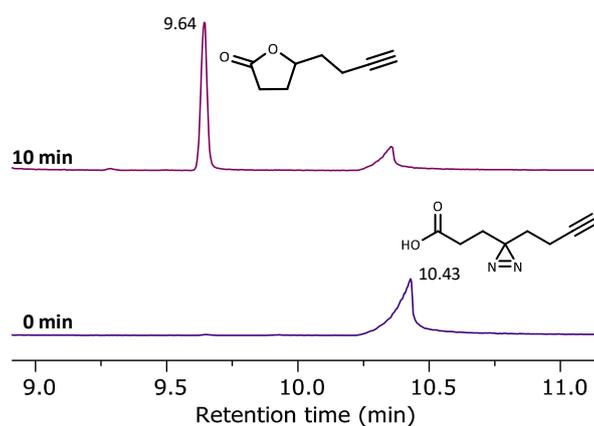


Figure 145: GC-MS chromatogram of *I-N<sub>2</sub>COOH* photolysis in water without irradiation (0 min, bottom) and after 10 min irradiation at 366 nm (top). *I-N<sub>2</sub>COOH* (10.43 min) is converted into a unique product (9.64 min) corresponding to the lactone product.

The *I-N<sub>2</sub>NHMe* prototype in theory cannot form a lactone product and thus could be a better candidate. Photolysis of the *I-N<sub>2</sub>NHMe* prototype in water shows the formation of three products with total conversion after 30 min (Figure 146-Left). The two first peaks (11.60 and 11.74 min) have a short retention time, while the third one (13.81 min) has a longer retention time. The latter has a fragmentation compatible with O-H insertion from water, and the two first peaks have been attributed to the two alkenes reduction products.

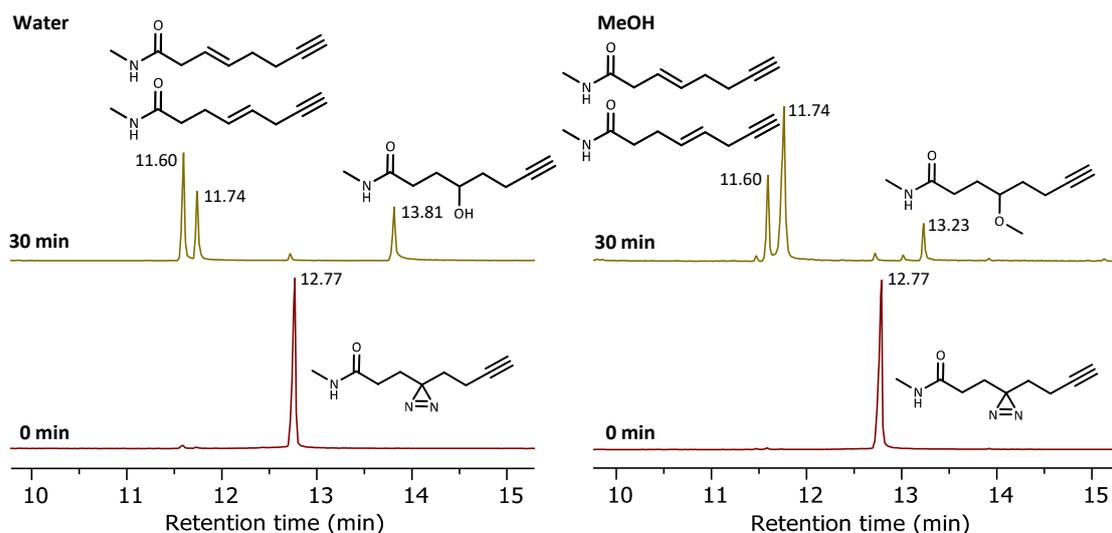


Figure 146: GC-MS chromatogram of the photolysis of the *I-N<sub>2</sub>NHMe* prototype in water (left) and in methanol (right) after 30 min irradiation at 366 nm. The *I-N<sub>2</sub>NHMe* prototype (12.77 min) is converted into two alkene products (11.60 and 11.74 min) and one O-H insertion with the solvent (13.81 min for the water insertion and 13.23 min for the methanol insertion).

In methanol, a similar profile is observed after 30 min of irradiation with the formation of the two alkene reduction products, having the same retention time at 11.60 and 11.74 min observed in water. In addition, there is an O-H insertion from methanol at 13.23 min (Figure 146-Right). Thus, the *I-N<sub>2</sub>NHMe* prototype is a good candidate for intermolecular reaction with nucleoside. In addition, the intramolecular cyclisation involving the alkyne was not observed, making it still available for post-functionalisation after the photoalkylation.

## 5.2. Photoproduct distribution

Taking into account the preliminary evaluation of the photoreactivity of the diazirines described in the previous section, the *l*-N<sub>2</sub>NHMe prototype has been chosen to generate photoadducts in the presence of different nucleosides since the photolysis experiments suggest the formation of the O-H insertion product, and also because the prototype is structurally closer to the **in-line** photoactivatable G4 ligands developed and studied in this thesis. *l*-N<sub>2</sub>NHMe was irradiated in water/acetonitrile 1:1 at 10 mM for 30 min at 366 nm in the presence of **dT**, **dC**, **dG**, and **dA** at 50 mM, and the resulting chromatograms are shown in Figure 147.

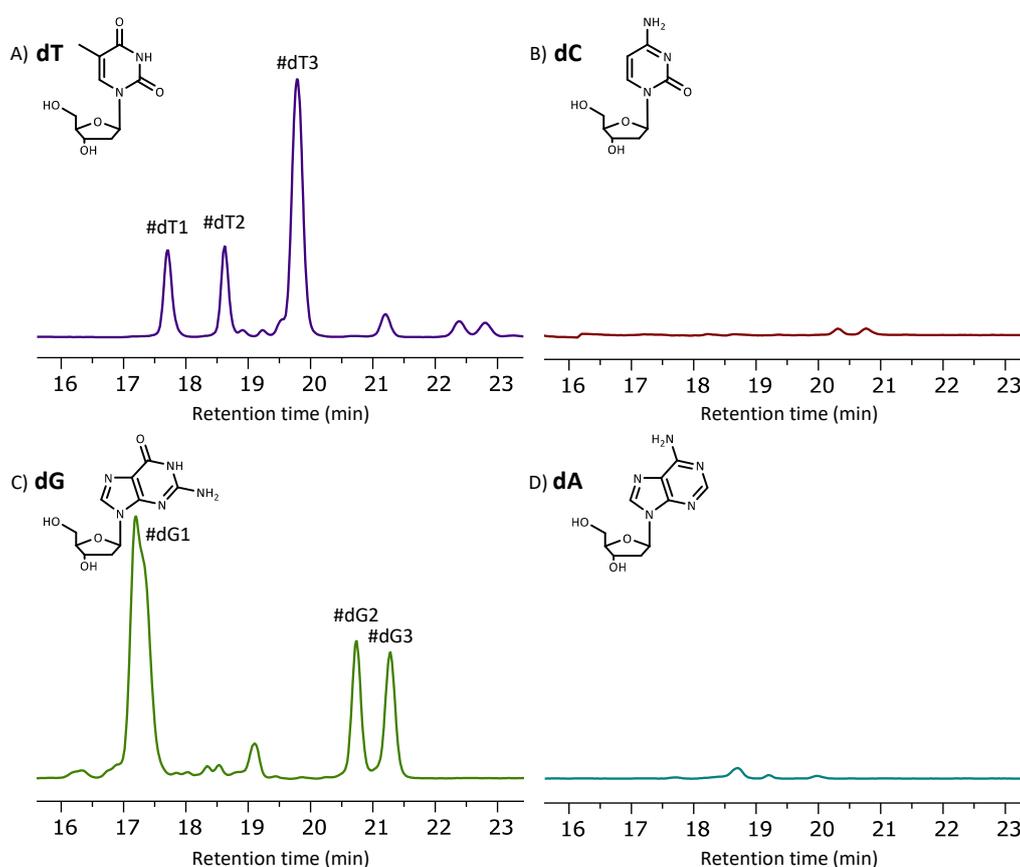


Figure 147: HPLC chromatogram of the reaction between *l*-N<sub>2</sub>NHMe prototype (10 mM) in the presence of A) **dT**, B) **dC**, C) **dG**, and D) **dA** (50 mM) in water/acetonitrile after 30 min irradiation with four lamps (15 W) at 366 nm. The chromatogram is zoomed in the photoproducts region (16 - 23 min). The relative intensity is the same for all four chromatogram. The full chromatogram is available in the annex section.

The photoreaction between *l*-N<sub>2</sub>NHMe and either **dC** or **dA** did not generate any detectable photoadduct. However, the irradiation at 366 nm in the presence of **dT** or **dG** generated at least three peaks (Figure 147-A/C) with a major peak (**#dT3** and **#dG1**) and two minor peaks (**#dT1/#dT2** and **#dG2/#dG3**) for each. In addition, **dG** major peak is large with a shoulder and most likely corresponds to at least two overlapping photoadducts.

It is evident by the absence of reactivity with **dA** and **dC** that the 2'-deoxyribose is not the target of the photogenerated carbene intermediate from *N*-N2NHMe, and thus, the nucleobase cores have to be involved. However, the imidazole ring can be eliminated as no photoadduct is generated with **dA**. Thus the carbene reacts with the pyrimidine ring to generate the photoadducts. In addition, **dG** exocyclic amino group and **dT** double bond can be ruled out as there is no reaction on **dA** exocyclic amino group and on **dC** double bond.

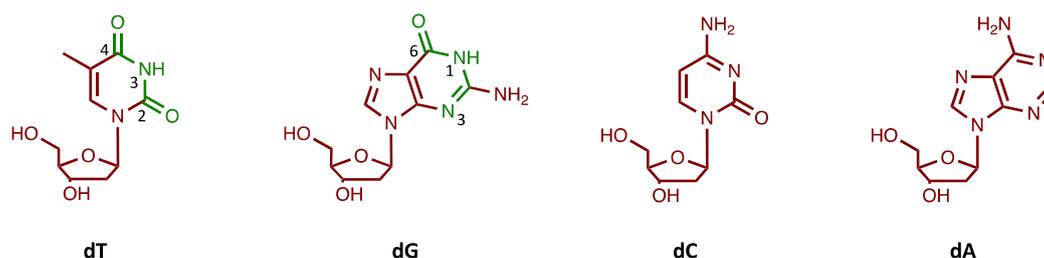


Figure 148: Molecular structure of **dT**, **dG**, **dC**, and **dA**: in red are represented the non-alkylation sites and in green the possible alkylating sites.

The above considerations drastically reduce the number of reactive sites, represented in green in Figure 148 and we can assume that the photoadduct formation results from an R-H insertion for R = N or O atom, the latter accessible from the tautomeric enol forms, although slightly populated.<sup>241</sup> Thus, it is reasonable to assume that the reactive site corresponds to the N3 for **dT** and N1 for **dG**. We hypothesise that the main photoadduct is the N-H insertion forming the main HPLC peak (**#dT3** and **#dG1**), and the two minor HPLC peaks (**#dT1/#dT2** for **dT** and **#dG2/#dG3** for **dG**) could arise by an N- or O-H insertion from the tautomeric forms. This would generate photoadducts with N3, O2, and O4 for **dT**, and N1, O6, and N3 for **dG**. The following section is going to describe the characterisation, after isolation and purification, of the photoadducts **#dT1-#dT3** and **#dG1-#dG3** to confirm the above hypothesis.

## 5.3. Structural characterisation of the photoadducts

### 5.3.1. dT adducts

We first have purified and characterised the main adduct exhibiting an HPLC peak labelled #d**T3** (Figure 147), which  $^1\text{H}$  NMR spectrum is displayed in Figure 149-B. The  $^1\text{H}$  NMR spectrum shows the conservation of all the 2'-deoxyribose protons (1', 2', 3', 4', and 5') and H6/Me7 thymine protons from d**T**, as well for the aliphatic protons (A, D, E, G, H, and J) from the *L*-**N<sub>2</sub>NHMe** moiety. In addition, two new quintet peaks (F, Figure 149-B) are present at 5.09 and 4.94 ppm, integrating for 0.6 and 0.4, respectively, which should correspond to the abstracted proton by the carbene on the thymidine reactive site.

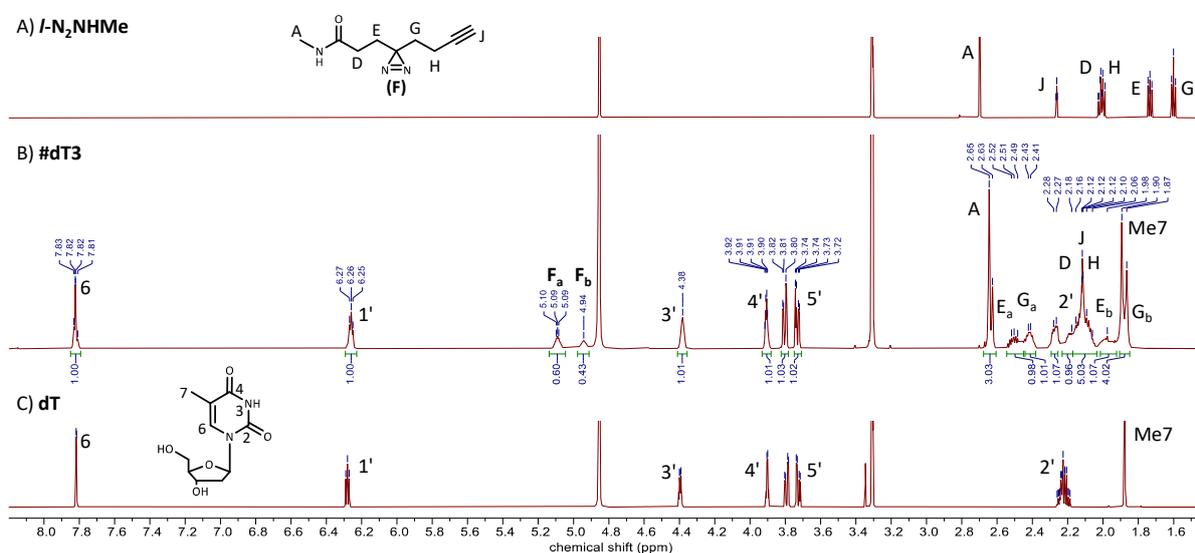


Figure 149:  $^1\text{H}$  NMR spectra (700 MHz) of A) *L*-**N<sub>2</sub>NHMe** prototype, B) purified #d**T3** product, and C) thymidine.

To confirm our hypothesis that this photoadduct (as a mixture of two diastereoisomers) is the result of the N-H insertion reaction, HSQC and HMBC experiments (Figure 150) have been recorded to assign the proton-carbon correlation, and to determine structural correlation between the protons of the *L*-**N<sub>2</sub>NHMe** moiety and the thymidine, respectively. The HSQC indicates the one-bond coupling between protons and carbons ( $^1J_{\text{CH}}$ ). The two peaks **F<sub>a</sub>** and **F<sub>b</sub>** correlate to two different carbons at 54 ppm and 56 ppm, which are carbon chemical shift values characteristic of a C-N bond. The HMBC indicate proton-carbon coupling separated from multiple bonds. **F<sub>a</sub>** proton is coupled in equal intensity with both carbonyl C<sub>2</sub> and C<sub>4</sub> at 152 ppm and 167 ppm, respectively, indicating that the proton is in equidistance from them.

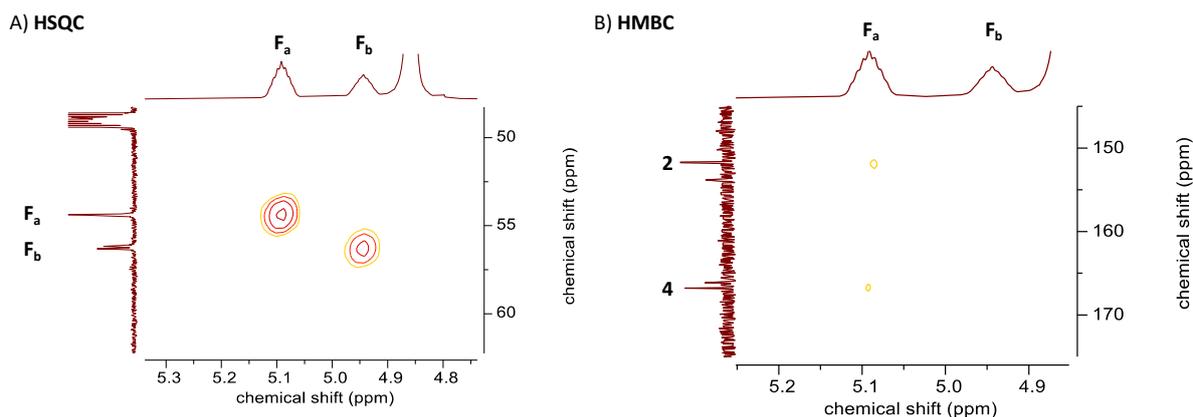


Figure 150: A) HSQC experiment of the purified **#dT3** product centred to the **F** peaks. B) HMBC experiment of the purified **#dT3** product centred to the two protons **F** interacting with the two carbonyl carbones 2 and 4.

These two elements confirm that **#dT3** is a mixture of two diastereoisomers differing only by the carbon **F** absolute configuration, generated from the N-H insertion reaction as predicted, which structures are presented in Figure 151.

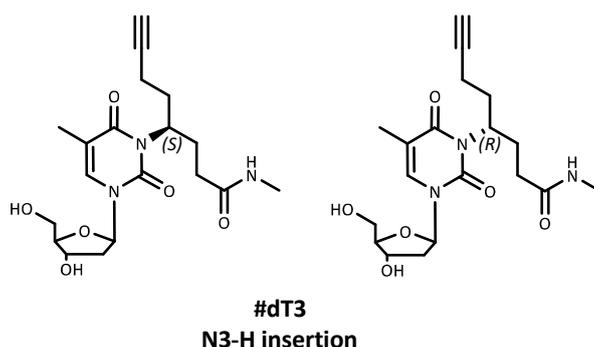


Figure 151: Structure of the two diastereoisomers produced by N-H insertion corresponding to the **#dT3** peak on the chromatogram.

Then **#dT1** was analysed by NMR spectroscopy, but the  $^1\text{H}$  spectrum (Figure 152) was not helpful for a structural assignment as the concentration in the sample was too low. Thus, this photoadduct has yet to be characterised.

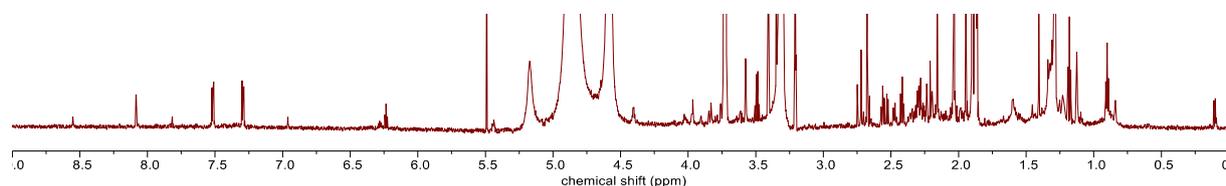


Figure 152:  $^1\text{H}$  NMR spectrum (700 MHz) of the purified **#dT1** adduct.

The photoadduct detected as **#dT2** peak was characterised by NMR, and  $^1\text{H}$  NMR spectrum is displayed in Figure 153. Similarly to **#dT3** photoadduct, all the thymidine (2'-deoxyribose and thymine) and aliphatic protons from the prototype are conserved. A new quintet signal (**F**) at 5.44 ppm integrating for 1 proton appeared in the  $^1\text{H}$  NMR spectrum, which should correspond to the inserted H atom. This **F** proton is slightly downfield compared to **#dT3** **F** protons (5.44 ppm vs 4.9-5.1 ppm for **#dT3**), which might result from an O-H insertion reaction we predicted.

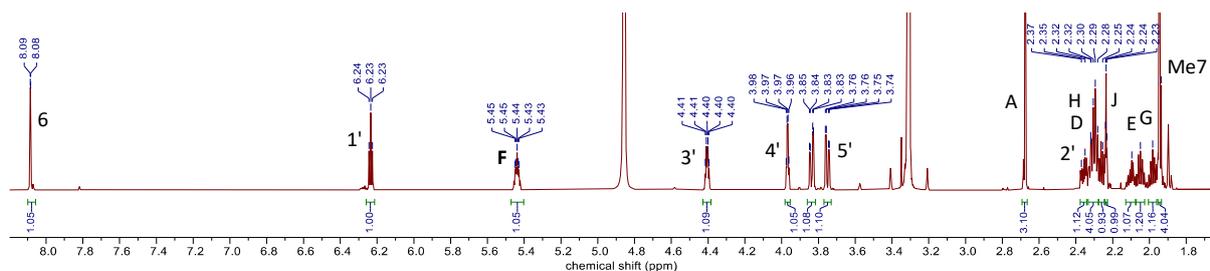


Figure 153:  $^1\text{H}$  NMR spectrum (700 MHz) of the purified **#dT2** adduct.

To confirm this hypothesis, HSQC and HMBC experiments have been recorded, but HMBC experiment did not provide any coupling information for the **F** proton. However, HSQC experiment (Figure 154) shows the coupling of the **F** proton with a carbon at 80 ppm, which is a chemical shift characteristic to C-O bond.

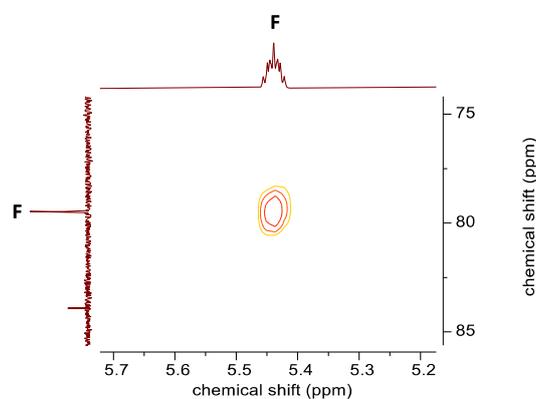


Figure 154: HSQC experiment of the **#dT2** product centred to the **F** proton and **F** carbon peaks.

Thus, we can confirm that **#dT2** product is a new photoadduct resulting from an O-H insertion as predicted. However, it is not possible to assign on which oxygen of **dT** the insertion took place, as no HMBC coupling between the carbonyl  $\text{C}_2$  or  $\text{C}_4$  with the inserted proton were observed. The structures of the two possible photoadducts are drawn in Figure 155.

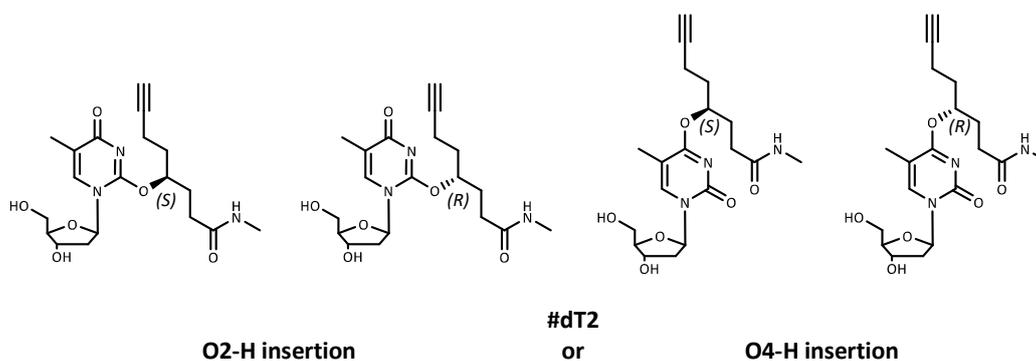


Figure 155: Representation of the two possible O-H insertion for the product #dT2.

### 5.3.2. dG adducts

We first have purified and characterised the main adduct exhibiting an HPLC peak labelled #dG1 (Figure 147), which <sup>1</sup>H NMR spectrum is displayed in Figure 156. The <sup>1</sup>H NMR spectrum shows the conservation of all the 2'-deoxyribose and guanine proton (H8) from dG, as well for the aliphatic protons from the *L*-N<sub>2</sub>NHMe moiety. However, there are five H8 peaks and multiple H1' peaks overlapping. This indicates the presence of more than one photoadduct. In addition, two new peaks at 5.70 ppm (F<sub>a</sub>) and 4.44 ppm (F<sub>b</sub>) integrating respectively for 0.17 and 0.54 are present, which should correspond to the H-abstracted proton of two different photoadducts. The integration sums up to 0.71, indicating that other photoadducts may be generated but their H-abstracted protons could not be found.

A) #dG1

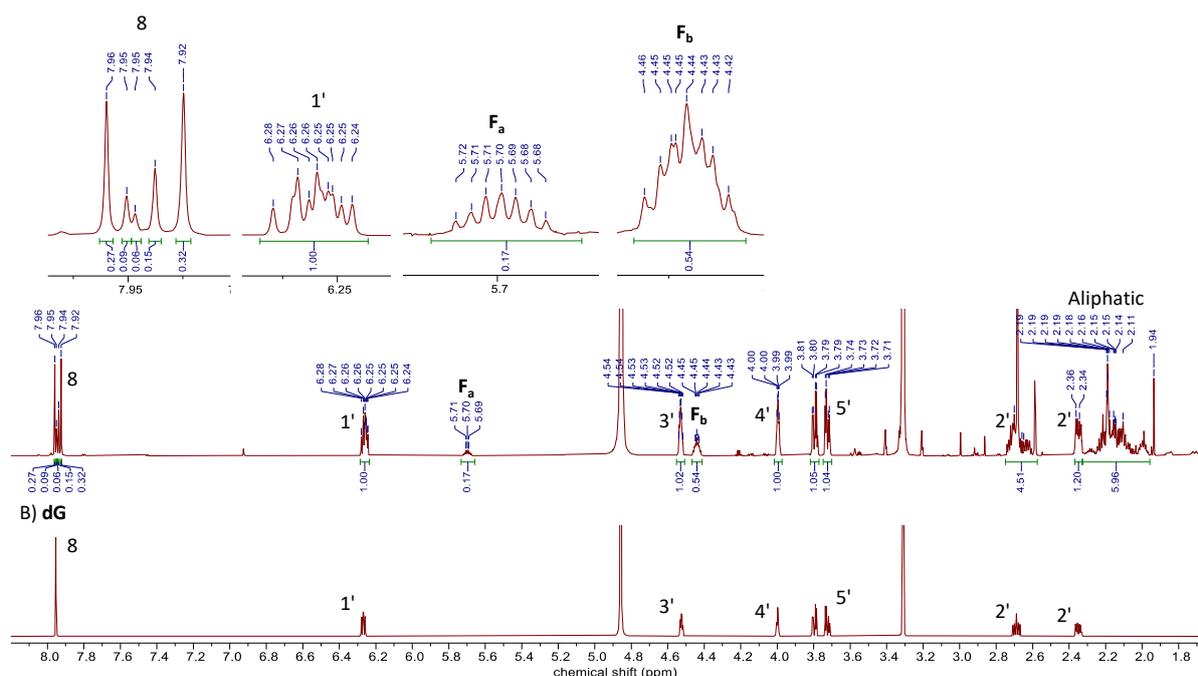


Figure 156: <sup>1</sup>H NMR spectrum (700 MHz) of A) #dG1 adduct with peak region 8, 1', F<sub>a</sub> and F<sub>b</sub> magnified (top), and B) 2'-deoxyguanosine reference (dG).

We focused on identifying the two photoadducts **#dG1a** and **#dG1b** from **F<sub>a</sub>** and **F<sub>b</sub>** peaks, respectively. The first one has an H-abstracted proton (**F<sub>a</sub>**) with a strong downfield effect with a chemical shift at 5.70 ppm, which is a very high chemical shift that corresponds more to an O-H insertion than to the N-H insertion predicted. The HSQC experiment (Figure 157) indicates a coupling with a carbon at 52 ppm that is not visible in <sup>13</sup>C spectrum, surely because of its too low proportion in the mixture (**F<sub>a</sub>** proton integrated for only 0.17). This carbon chemical shift is not compatible with an O-H insertion but is for an N-H insertion. As the chemical shift of the proton and carbon **F** are not compatible together, confirming an N-H insertion over an O-H is complicated.

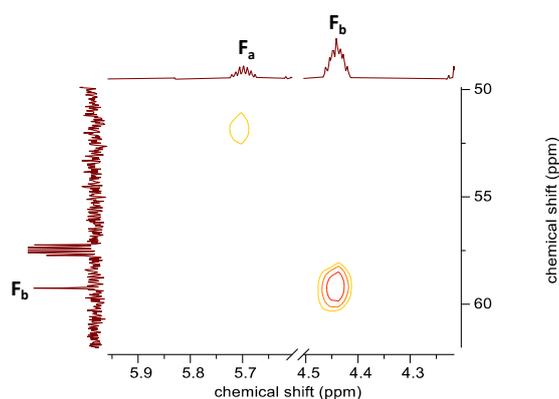


Figure 157: HSQC experiment for the adduct **#dG1** centred to the H-abstracted **F<sub>a</sub>** and **F<sub>b</sub>** peaks.

For the photoadduct **#dG1b**, the H-abstracted proton (**F<sub>b</sub>**) at 4.44 ppm has a chemical shift matching a CH-N proton. In addition, HSQC experiment (Figure 157) allows to identify the carbon at 58 ppm matching a C-N chemical shift. Thus, **#dG1b** correspond to the N-H insertion product, in adequacy with the proton and carbon chemical shift and prediction.

However, the HMBC experiment did not show any coupling with the pyrimidine carbon, making it impossible to determine in which NH the insertion reaction occurred. Since only N1 has a labile proton in the main guanosine tautomer (lactam), the **#dG1b** as the major photoadduct should be generated from the N1-H insertion. For **#dG1a**, we hypothesis the formation of the photoadduct from N3-H insertion, but it is uncertain as the **F<sub>a</sub>** proton is unusually deshielded and remain to be confirmed. The possible structures of the photoadducts are drawn in Figure 158, which should be confirmed by purifying the two photoadducts. Unfortunately, the retention time is almost identical, and so far, we have failed in the purification.

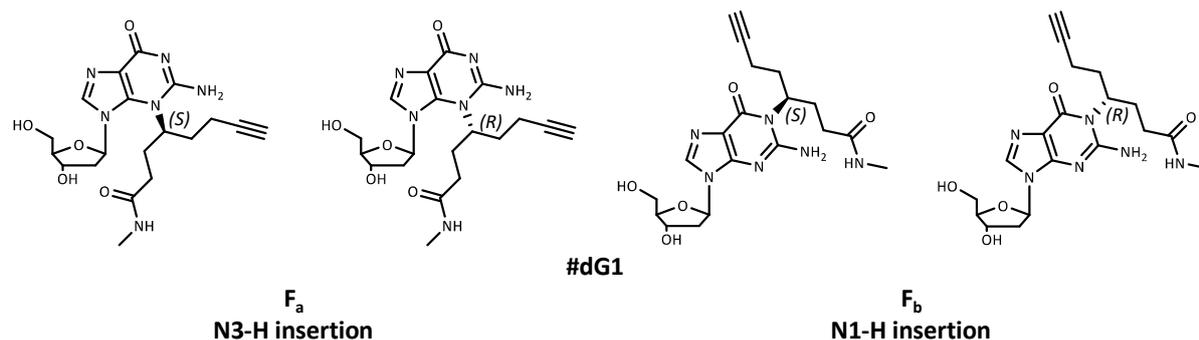


Figure 158: Hypothetic structures of #dG1 from NMR and in adequation with the identified adducts with dG.

Photoadducts of #dG2 and #dG3 were characterised by <sup>1</sup>H NMR and spectra are displayed in Figure 159. It is interesting to observe that both purified #dG2 and #dG3 compounds exhibit overlapping <sup>1</sup>H NMR spectra. Both spectra show the conservation of all the 2'-deoxyguanosine protons and aliphatic protons. In addition, the 5.70-5.50 ppm region presents the formation of multiple peaks with at least one H-abstracted proton (F) at 5.60 ppm integrating for 1 proton and overlapping with impurities. The chemical shift corresponds to an O-H insertion, which correspond to the predicted photoadduct.

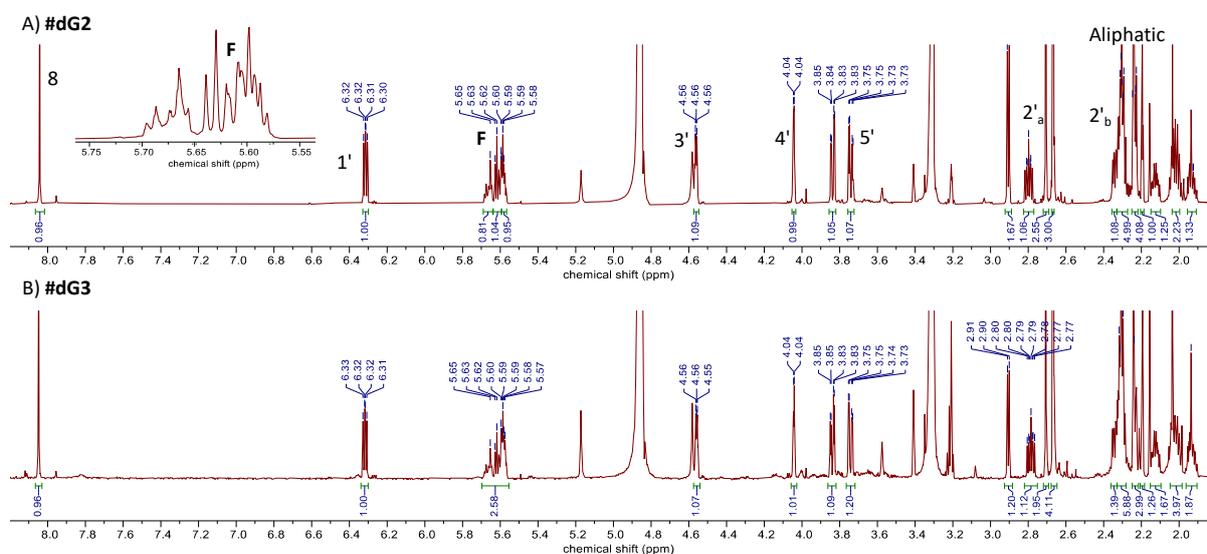


Figure 159: NMR spectra (700 MHz) of the compounds A) #dG2 and B) #dG3.

To confirm that the photoadduct is the one generated from an O-H insertion, HSQC experiment (Figure 160) has been recorded. The 5.60 ppm proton (F) correlates to the carbon at 76 ppm. This chemical shift confirms a C atom bound to an oxygen, which is O6 as it is the only one available.

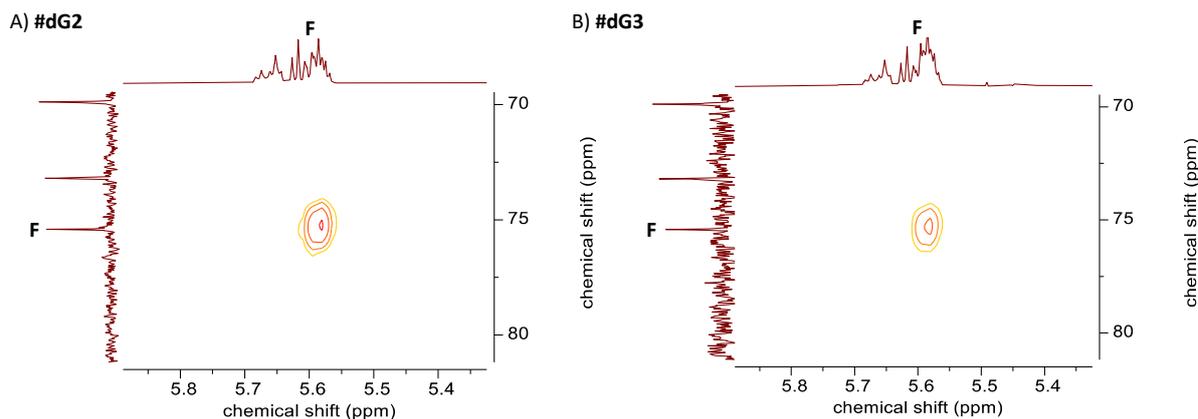


Figure 160: HSQC experiments for A) **#dG2** and B) **#dG3** centred to the **F** peaks.

In addition, to confirm our structural assignment and also that the other peaks are due to impurities, we have decided to conduct a DOSY experiment. This experiment allows to separate peaks of different molecules according to their diffusion coefficient.<sup>242</sup> The DOSY experiment of **#dG3** is displayed in Figure 161 and shows a good separation of at least two species. The first one (dash line) corresponds to the photoadduct matching all the 2'-deoxyguanosine (2'-deoxyribose and H8) and aliphatic protons, and allows to confirm the presence of a unique photoadduct having for H-abstracted the peaks at 5.60 ppm. The second (bottom) species does not have many peaks but corresponds at least to one impurity visible in the 5.70-5.50 ppm region.

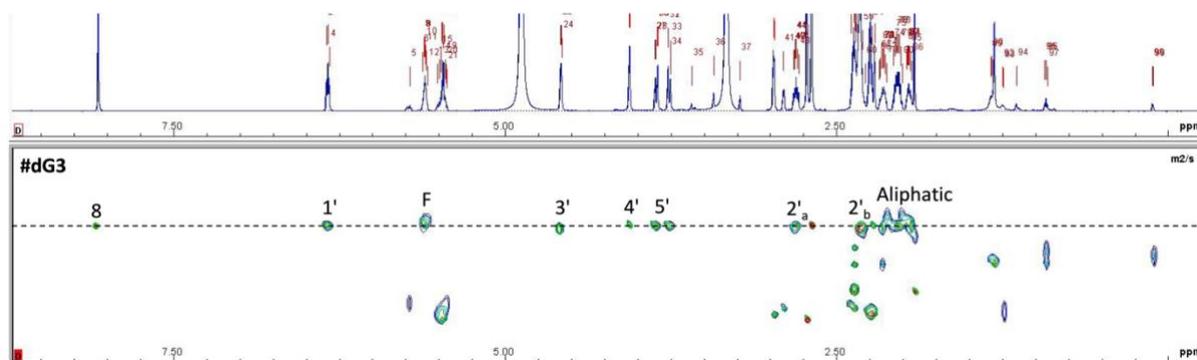


Figure 161: DOSY experiment of **#dG3** compound.

The photoadduct tagged as **#dG2** and **#dG3** are both generated by an O-H insertion reaction on O6 and most likely they are a couple of diastereoisomers only differing for the configuration of the C-atom **F**, as draw in Figure 162.

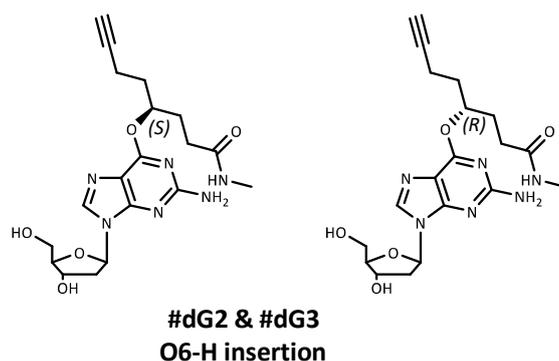


Figure 162: Hypothetical structures of #dG2 and #dG3 photoadduct.

#### 5.4. Competitive product distribution

We have decided to investigate the selectivity of *I-N<sub>2</sub>NHMe* prototype by performing a competition reaction. **dA** and **dC** were ruled out of the assay as only **dT** and **dG** generate photoadducts. The same condition as previous (see section 5.2., p. 164) was used by mixing **dT** and **dG** at 50 mM with *I-N<sub>2</sub>NHMe* prototype at 10 mM, and the solution was irradiated at 366 nm for 30 min. The resulting HPLC chromatogram is presented in Figure 163-B and shows the formation of all six peaks: the three from **dT** and the three from **dG** with similar retention time compared to the non-competitive reactions but in different proportion compared to the references (Figure 163-A). We can observe that the main photoadduct is #dT3 corresponding to the N-H insertion for 46%. The HPLC peaks corresponding to the mixture of N-H insertion from **dG** (#dG1a and #dG1b) correspond to only 19%. The proportion of minor peaks (#dT1/#dT2 for **dT** and #dG2/#dG3 for **dG**) is similar compared to the references (1/3 of the relative proportion in total for the minor peaks). To summarise, this indicates a clear preference of the generated carbene to react with **dT** (68%) over **dG** (32%).

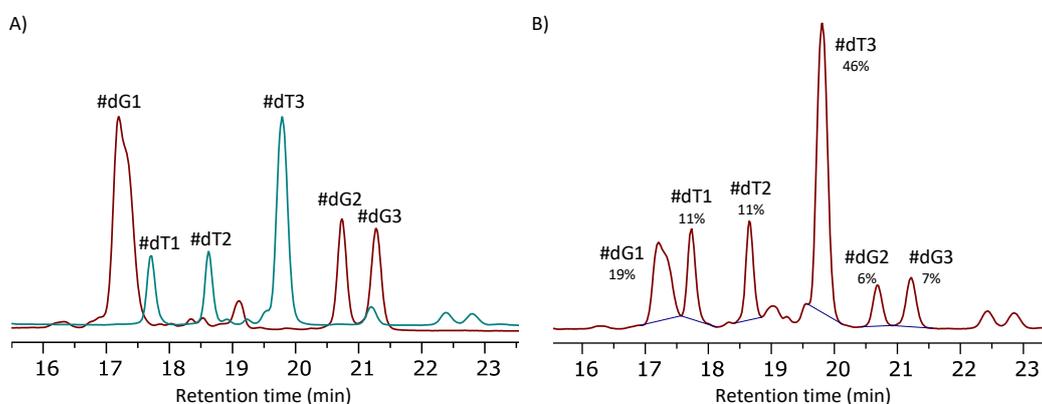


Figure 163: HPLC chromatogram of A) overlapped **dT** (in blue) and **dG** (in red) reaction with *I-N<sub>2</sub>NHMe* and B) the competition reaction with **dT** and **dG** in presence of *I-N<sub>2</sub>NHMe* after 30 min irradiation with 4 lamps (15 W) at 366 nm.

## 6. Conclusion

### 6.1. Conclusion on the benzophenone derivatives

From all the benzophenone derivatives studied, **Bph-COOH** precursor and **Bph-CONHMe** derivative exhibit a similar triplet excited state compared to the benzophenone, with a long lifetime ( $t > 5 \mu\text{s}$ ) and a good selectivity for thymidine. The **Bph-CONEt<sub>2</sub>** prototype mimics more effectively the conjugated functionalisation present in the photoactivatable G4 ligands and allows to better understand the activation and properties of the photoactivatable moiety present in our benzophenone series of compounds. LFP investigation allowed us to discover that this specific conjugation in an *N,N*-disubstituted amide has a negative impact on the lifetime of the benzophenone triplet excited state reducing its lifetime below 50 ns. It was not possible to precisely measure the transient lifetime as it is below the detecting time limit of our LFP. Such shorter lifetime negatively impacts the reaction efficiency, explaining why the irradiation time must be extended to tens of hours to observe the formation of photoadducts, and reach full conversion after days of irradiation with 4 lamps of 15 W centred at 366 nm. On the contrary, the benzophenone and **Bph-COOH** exhibiting a much longer lifetime require just minutes to generate photoadducts and hours to reach full conversion. Despite this low efficiency, we have confirmed that the reaction pathway remains identical to the formation of oxetane by [2+2] cycloaddition with the thymidine (Figure 164). The oxetanes are a family of two regioisomers that do not have the same thermal stability. The "ether" oxetane is thermally stable, and the "hemiaminal ether" oxetane is thermally reversible by undergoing a reverse [2+2] cycloaddition, regenerating the **Bph-CONEt<sub>2</sub>** and thymidine reagent upon heating.

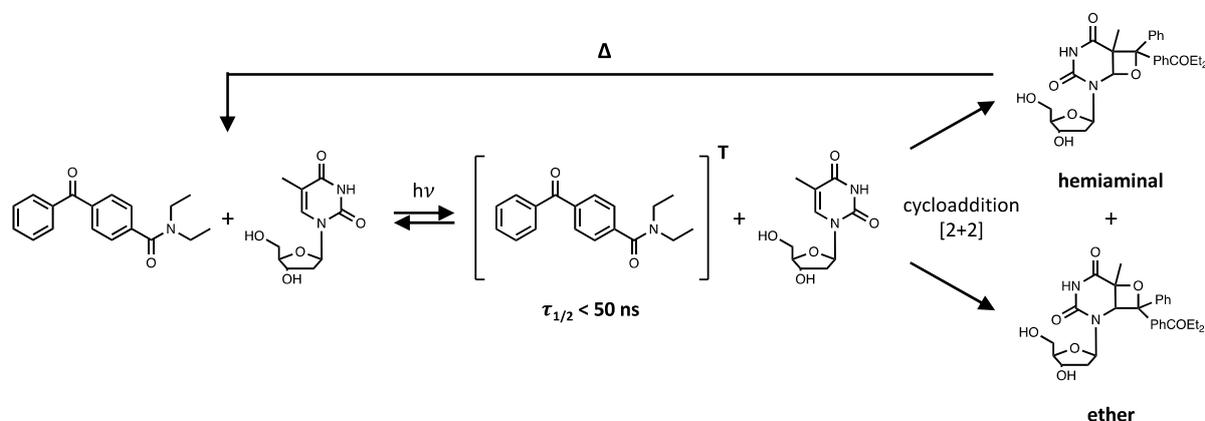


Figure 164: Summary of **Bph-CONEt<sub>2</sub>** reactivity and photoreaction pathway with **dT**.

## 6.2. Conclusion on the *in-line* prototype

The *I-N<sub>2</sub>NHMe* prototype mimics the structure of the *in-line* series synthesised and investigated in this thesis. The Laser Flash Photolysis investigation allows us to indirectly detect the transient species as a pyridine-ylide, which showed a very short lifetime ( $t < 50$  ns). The steady-state photolysis experiment of the *I-N<sub>2</sub>COOH* precursor and the *I-N<sub>2</sub>NHMe* prototype indicates that only *I-N<sub>2</sub>NHMe* was able to react with the environment when *I-N<sub>2</sub>COOH* only formed the lactam intramolecular product. *I-N<sub>2</sub>NHMe* was fully consumed after 30 min of irradiation with 4 lamps (15 W) centred at 366 nm, and generated mainly the intramolecular [1,2] H-shift and an O-H insertion with water and methanol. The photolysis experiment in the presence of nucleosides (50 mM) with *I-N<sub>2</sub>NHMe* (10 mM) showed absence of photoadducts with **dA** and **dC**, but reacted with **dT** and **dG** to generate photoadducts. The main adduct (**#dT3** and **#dG1b**) correspond to the N-H insertion into N3 for **dT** and N1 for **dG**, which are the unique labile protons available. The minor adducts **#dT2** and **#dG2/#dG3** correspond to an O-H insertion into O2 or O4 (it remains unknown on which oxygen the insertion takes place) for **dT** and O6 for **dG**, which correspond to the reaction with the enol form of both **dT** (O2H-dT or O4H-dT) and **dG** (O6H-dT). Another minor adduct, **#dG1a**, is uncertain and requires a better purification to be separated from **#dG1b**, our hypothesis is an N-H insertion from the N3 on the imine form (N3H-dG) but remains to be confirmed. The reactivity of *I-N<sub>2</sub>NHMe* with **dT** and **dG** is summarised in Figure 165.

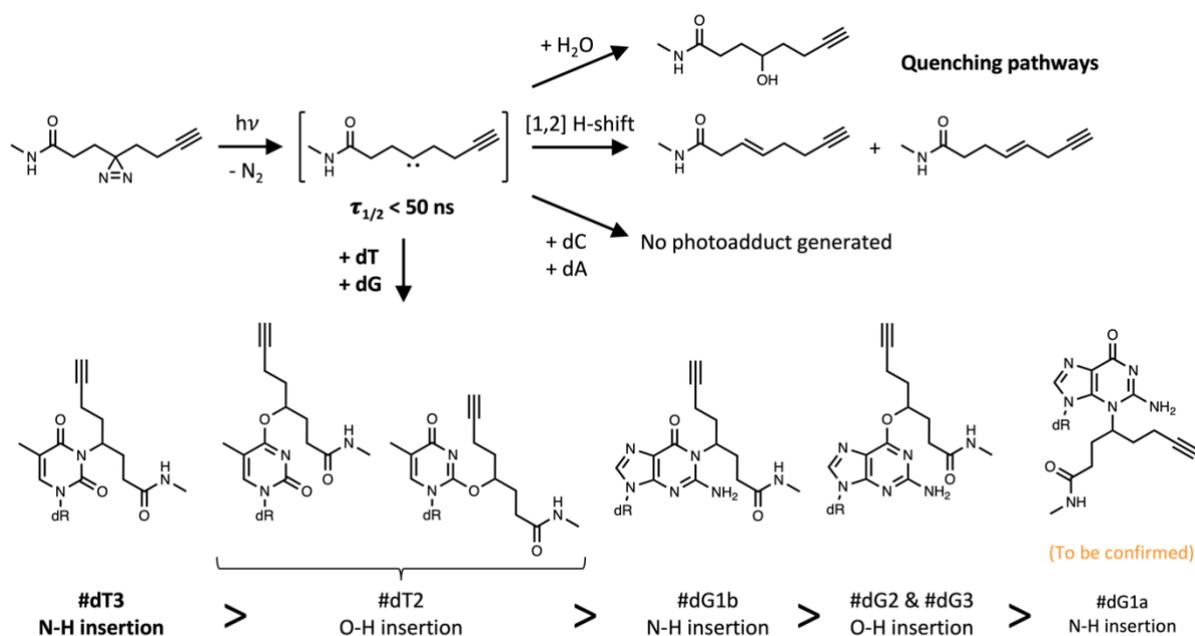


Figure 165: Summary of *I-N<sub>2</sub>NHMe* reactivity and photoreaction pathways. The chevrons indicate the proportion from the *dT/dG* competition reaction.

In terms of proportion, there is 2/3 of the main N-H insertion adduct (**#dT3** and **#dG1b**) and 1/3 of the minor adducts, which is an important proportion of enol and imine forms alkylated vs. the keto forms. We hypothesize that the tautomerisation may be promoted by the alkylation as the carbene has a zwitterionic character, which would explain the important proportion of enol/imine forms alkylated. To finish, a competitive assay on **/N<sub>2</sub>NHMe** in presence of both **dT** and **dG** showed a clear preference to react with **dT** (68%) over **dG** (32%). This competition shows no variation in the relative proportion of main adduct (2/3) vs. the minor adducts (1/3). Thus, the main generated photoadduct was from the N3-H insertion of **dT** (46%).

## Conclusions and perspectives



The accumulation of evidence from *in vitro* and *in cellulo* experiments supports the *in vivo* formation of G-quadruplexes,<sup>48</sup> and BG4-seq experiments identified over 10,000 forming G4 sequences in the human genome.<sup>91,94</sup> These G4 forming sequences are not distributed randomly within the genome and play important roles in regulating processes such as replication, transcription and translation.<sup>208</sup> However, most of the *in vitro* methodologies used to identify G4s rely on the use of chemically fixed genomes and anti-G4 antibodies.<sup>92</sup> It is in this context that we proposed the development and study of a new class of photoactivatable G4 ligands to pave the way for developing a new G4 mapping methodology, named chemical immunoprecipitation sequencing (Chem-IP-seq). It is based on the preliminary alkylation of the G4 structures *in cellulo*, then, isolation and enrichment of the labelled G4 sequences by pull-down experiments and sequencing, allowing to map the accessible G4 structures inside the genome. This methodology has the advantage of not requiring chemical fixation allowing the ligand to recognise and bind G4s in a dynamic system, and the formation of a covalent bond avoids relocalisation of the ligand during treatment. This methodology requires a trifunctional photoactivatable G4 ligand with the ability to i) specifically interact with the G-quadruplexes, ii) generate a covalent bond with the G4s upon UV-activation, and iii) be post-covalent binding functionalised with a hapten for enrichment. The first part of the PhD work focused on the synthesis and study of these new trifunctional photoactivatable G4 ligands.

These photoactivatable G4 ligands are based on the PDC (360A) core, selected for its high affinity and selectivity for G4 structures.<sup>179,210</sup> The PDC core was functionalised with two different moieties: a photoactivatable group to trap covalently the G-quadruplexes and a terminal alkyne for post-covalent binding functionalisation. The latter will allow the introduction of the hapten by CuAAC. The team has designed and synthesised 15 new photoactivatable G4 ligands divided in two different families: 3 compounds belong to the “**in-line**” family and 12 to the “**branched**” family.

The first family has been obtained by tethering the 4-diazirine-7-octynoic acid, bearing the photoactivatable moiety, to the PDC core through a diamino C2, C4 and DEG linker. The second was obtained by tethering one of the three photoactivatable moieties (benzophenone (Bph), benzyl trifluoromethyl diazirine (N2), and tetrafluorophenyl azide (N3)) and one of the four spacers bearing the terminal alkyne (C1, C3, DEG and TEG) to the PDC core through a *N*-azaneylbutyl-4-azaneylbutanamide linker. The synthesis of the different photoactivatable G4 ligands studied in this PhD project has been successfully realised by three different people: Dr. T. Masson, Mrs. C. Landras Guetta, and myself.

The ability of the photoactivatable G4 ligands to interact and stabilise the G4 structures was confirmed by using FRET-melting and G4-FID assays. The two families presented different affinities for the G4 structures due to the difference steric hindrance. The linear linker and the compact diazirine bore by the PDC core in the **in-line** derivatives limit the steric hindrance during interaction with the G4s. This less bulkier functionalisation allows the **in-line** ligands to conserve most of the PDC core stabilisation properties with only a reduction of about 10% to 20% of the  $\Delta T_{1/2}$  values measured in the presence of all the G4 prototypes chosen for this study, and independently of the **in-line** linker nature or length. Accordingly, the  $DC_{50}$  values ( $< 0.5 \mu\text{M}$ , similar to ones displayed by PDC) obtained from the G4-FID experiments showed the high propensity of the **in-line** derivatives to displace thiazole orange in the presence of all the studied G4s. The **branched** ligands present a ramified-type structure that is bulkier and therefore may affect the interaction. This higher hindrance induces a more important negative impact on the interaction and stabilisation with G4s. It results in a loss of about 50 to 80% of the PDC core stabilisation properties. Surprisingly, the  $DC_{50}$  values are less negatively impacted and remain between 0.5 and 1.0  $\mu\text{M}$ , corresponding to medium affinity interacting G4 ligands. The **branched** derivatives present a much more structural diversity, presenting three different photoactivatable moieties and four spacers bearing the terminal alkyne. However, we did not observe a correlation between the different functionalisation and the stabilisation, and the compounds presented the subsequent ranking: **C3/N2**, **C1/N3**, **DEG/N3**  $>$  **N3**  $>$  **N2** and **Bph**  $>$  **C1/N2**. This trend shows that the **N3** functionalisation has a slightly lower negative impact on the PDC stabilisation than the **N2** and **Bph** functionalisation. This could be explained by the smaller size of the tetrafluorophenyl azide moiety as compared to the two others. There is no rational impact of the spacer nature and length on the stabilisation, as shown by the three stand-out branched compounds (**C1/N3**, **C3/N2**, **DEG/N3**) comprising three of the four spacers (C1, C3, and DEG). Regarding selectivity, the **in-line** ligands conserve the selectivity of PDC for the G4s against duplex DNA, differently the **branched** ligands are more affected by the presence of a duplex competitor, mainly the ones displaying the lower stabilisation. To sum up, both **in-line** and **branched** derivatives interact and stabilise G4 structures with different efficiency, following this ranking: **in-line**  $>$  **C3/N2**, **C1/N3**, **DEG/N3**  $>$  **N3**  $>$  **N2** and **Bph**  $>$  **C1/N2**.

The next step focused on the validation of the alkylation properties of these photoactivatable G4 ligands: we studied the ability of these compounds to generate a covalent bond with various G4 structures upon UV activation. Their abilities to generate covalent adducts were evaluated using three G4 structures representing a variety of loop lengths and conformations: 22AG forms a hybrid G4 characterised by short propeller and lateral loops, cMYC22 forms a parallel G4 with only short (1-2 nt) propeller loops, and CEB25 *wt* forming a parallel G4 with two short (1 nt) and a 9-nt long propeller loops.

**In-line** derivatives present excellent alkylation yield with the three G4s (60% with 22AG, 40% with cMYC22, and 70% with CEB25 *wt*), showing no selectivity for a particular G4 topology. It is interesting to note the alkylation efficiency disparity between cMYC22 and CEB25 *wt*, both having a parallel topology with only propeller loops, but presenting an important difference in alkylation efficiency. We hypothesise that the dynamic of the 9-nt long loop of CEB25 *wt*, with less demanding conformational restraint, increases the number of accessible alkylation sites.

The **branched** derivatives showed a marked selectivity for specific G4 structures: the alkylation yield is low with 22AG (< 20%) and high with cMYC22 and CEB25 *wt* ( $\geq 50\%$ ). In these compounds, the position of the photoactivatable moiety position remains the same. We observed only a negligible variation in the alkylation efficiency between the different photoactivatable moieties when compared with the same G4. This indicates that the main parameter controlling the alkylation is the proximity between the activated photoactivatable moiety and the alkylation site. These results highlight the importance of the design of the photoactivatable G4 ligands: identifying the most suitable position where to introduce the photoactivatable moiety will allow to not perturb the G4 interaction and maximise the accessibility of the alkylation sites.

It is intriguing that the alkylation efficiency of 22AG is high with **in-line** derivatives but low with **branched** derivatives. Further investigation is necessary to uncover the reason for this topology specificity of the **branched** ligands. Recent studies have shown that the switch from hybrid to antiparallel induced by the parented PhenDC3 modified the mode of interaction from stacking to the external G-quartet to intercalation between two G-quartets.<sup>243</sup> It would be worthwhile to investigate whether a similar change in interaction mode is likely to occur with PDC and its derivatives and whether it could potentially affect the alkylation process for the **branched** compounds.

To conclude, the **in-line** derivatives show a high alkylation efficiency with all the G4s used for this study and should be able to trap any G4 structure with which the PDC core interacts with. The **branched** derivatives showed alkylation efficiencies which are more dependent on the G4 topology. This is why, the **in-line** compounds are used right now by the research team to develop the Chem-IP-seq methodology.

To understand more about the alkylation properties of these compounds, the second part of this PhD work consisted of determining the exact alkylation sites on the G4 structure, and improving our understanding of the alkylation process, from the generation and characterisation of the reactive species to the elucidation of the structural nature of the photoproducts obtained with nucleosides. Although work has been completed in this context for benzophenone derivatives,<sup>240</sup> the chemical nature of the covalent modification of nucleosides by aliphatic diazirine is a new addition to the current literature. This work focused on both the **in-line** and **branched** benzophenone derivatives, with the successful synthesis of prototype mimics that uniquely conserve the tethering function on the photoactivatable moiety to investigate their photochemical properties.

The case of benzophenone is interesting, but is not particularly novel, as it has already been applied for the alkylation of dsDNA<sup>229</sup> and G4s.<sup>179</sup> However, the **Bph-CONEt<sub>2</sub>** prototype has never been exploited for photoaffinity labelling. Unfortunately, this compound generates a very short-living transient ( $\leq 50$  ns) which cannot be studied by our nanosecond LFP due to its detection limit. This outcome is surprising when compared to the long-lived transient ( $> 5$   $\mu$ s) generated by irradiation of **Bph-COOH** and **Bph-CONHMe**, which is similar to unsubstituted benzophenone both exhibiting T<sub>1</sub> ( $n\pi^*$ ) nature. We hypothesise that the shortening of the lifetime could be produced by a rotor thermal relaxation process arising from the *N,N*-disubstituted amide. This shorter lifetime does not seem to affect the alkylation efficiency observed between the **Bph** series and the G4s, nor the thymidine selectivity as confirmed by the enzymatic sequencing experiment. The only affected element is the time required for the intermolecular reactivity due to the lower efficiency of the alkylation event, which passes from hours for the benzophenone to days for **Bph-CONEt<sub>2</sub>** when irradiating with four lamps at 366 nm.

The reaction in the presence of thymidine generates solely oxetane photoproducts with 71% of hemiaminal ether regioisomers and 29% of the ether regioisomers. Both chemical and enzymatic sequencing demonstrated the thermal reversibility of the main photoadducts, and thermal stability experiment showed that only the hemiaminal ether regioisomers were sensible and reversed upon heating.

In conclusion, LFP did not confirm the triplet state  $T_1$  ( $n\pi^*$ ) nature of the reactive excited state. Nevertheless, the generation of oxetane photoadducts after extended irradiation should be possible exclusively from a short-lived triplet state. We believe that it is the first time that a short-lived triplet excited state transient, reverting to the ground state in the absence of its target has been described and introduced into a ligand for photoaffinity labelling strategy. This photoactivatable moiety opens up to new opportunities allowing to bypass the main drawbacks faced by short-lived transient species that often react through quenching pathways with other unwanted partners or intramolecularly as deactivation to their ground state is not possible, while **Bph-CONEt<sub>2</sub>** reverts to its ground state with no by-product formation in the absence of its targets. In addition, the main product generated with DNA is the hemiaminal ether oxetane which is thermally reversible and thus can release the undamaged original DNA sequence. Following the prototype results, the labelling of thymidine by the **branched benzophenone** series is summarised in the Figure 166.

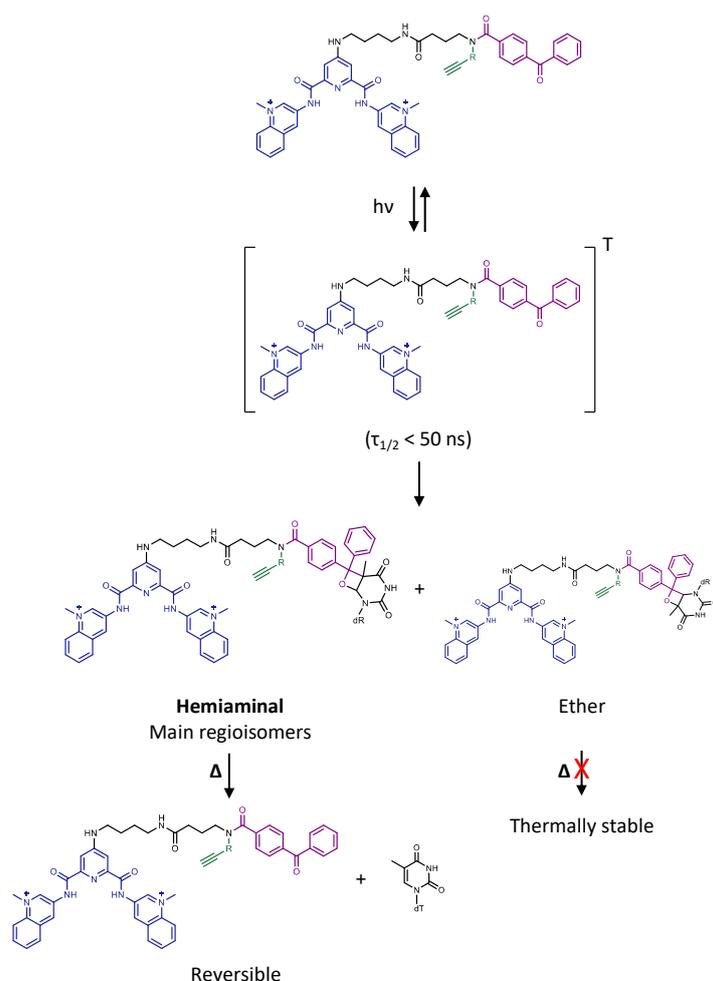


Figure 166: Reactivity of the **Bph** series based on the prototype results.

The aliphatic diazirine has often been studied with amino acids to describe the photoaffinity chemistry on peptides,<sup>198</sup> but its reactivity toward nucleosides has yet to be evaluated. We sought to characterise the photochemical properties of the *L*-**N<sub>2</sub>NHMe** prototype by LFP through trapping of the generated carbene into a pyrimidine ylide transient. The pyrimidine ylide transient was formed, however, presented a too-short lifetime (< 50 ns) to be studied with our nanosecond-LFP. The *L*-**N<sub>2</sub>NHMe** photolysis in water is rapid, showing a total conversion in 30 min, which is close to the reaction time to reach maximum G4 alkylation with the **in-line** derivatives.<sup>211</sup> The prototype upon irradiation at 366 nm forms the alkene product from the intramolecular [1,2] H-shift and the hydroxyl product from O-H insertion with the water. It is important to point out the absence of alkyne reduction, the latter remaining available for post-covalent binding functionalisation.

The intermolecular reaction in the presence of either **dC** or **dA** did not generate any detectable photoadducts. However, the reaction in the presence of **dT** allowed to identify and characterise two photoadducts: the N3-H insertion (**#dT3**) and O-H insertion (**#dT2**) products. The reaction in the presence of **dG** allowed to identify, and partially characterise, at least three photoadducts: the N1-H insertion (**#dG1b**) and the O6-H insertion (**#dG2** & **#dG3**) products, and we hypothesise the formation of the N3-H insertion (**#dG1a**) product, which remains to be confirmed.

It is interesting to observe the absence of other insertions products, although the aliphatic diazirine is considered to be able to react with any C-, O- and N-H bonds.<sup>187</sup> In addition, only functional groups bearing an acidic proton were able to react with the reactive intermediate. Indeed, **dC** and **dA** lack acidic protons, instead, **dT** and **dG** have one acid proton, with *pKa* of 9.93 (N3H/N3<sup>-</sup>) and 9.42 (N1H/N1<sup>-</sup>) respectively. Following the conclusion of O'Brien *et al.*<sup>233</sup> and West *et al.*<sup>198</sup>, the requirement of an acid proton indicates that the reaction arises from the diazoisomer intermediate following an "ion pair" pathway and not from the carbene. It would be interesting to confirm this pathway by reproducing the reactions in the presence of BCN, to trap the diazoisomer intermediate, and we could perform the reaction in non-physiological basic conditions (pH > 10) as the deprotonation of the acidic NH should decrease the alkylation yield.<sup>198,233</sup> As well, it would be interesting to reproduce these reactions using the benzyl trifluoromethyl diazirine prototype as the diazoisomer should be stabilised and not react by the "ion pair" mechanism as proposed by West *et al.*<sup>198</sup>

In addition, the ion pair pathway could explain the important proportion of O2- or O4-H insertion on **dT** and O6- and N3-H insertion on **dG**, which cannot be explained uniquely by the keto-enol tautomerisation. We hypothesise that the ion pair formation promotes the secondary photoadduct formation by the negative charge being delocalised between N3, O2, and O4 on **dT**, and N1, O6, and N3 on **dG**.

Based on the results obtained with the prototype and our hypothesis, the **in-line** derivatives nucleic acids' labelling mechanism is summarised in Figure 167.

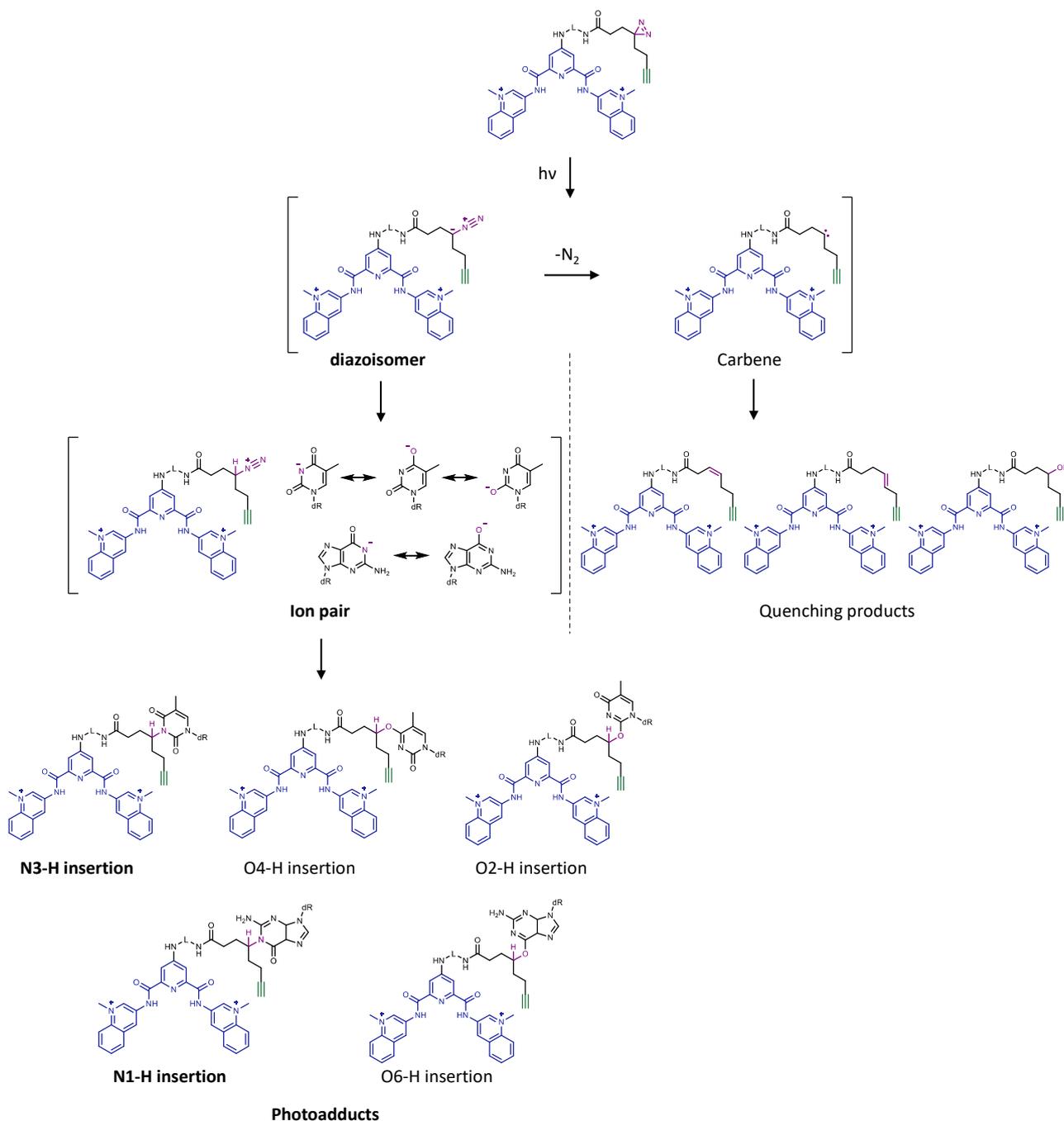


Figure 167: Reactivity of the in-line derivatives based on the prototype results and our hypothesis.

The study carried out in the presence of the prototype compound and **dT** and **dG** involves intermolecular reactions when the reactions between the **in-line** ligands and the G4 structures are pseudo-intramolecular (complex formation). The latter were studied using chemical and enzymatic sequencing experiments, but did not allow to determine the exact alkylation sites. Based on the aforementioned study, we hypothesised that both accessible guanines and thymidines can be alkylated in the G4 structure. However, it would be interesting to tether the aliphatic diazirine prototype to nucleosides to study the carbene reactivity that is a significant part of the intramolecular reactivity.

The **in-line** derivatives have been demonstrated to alkylate the 5'-end of G4 structures, and taking into account the photochemical studies carried out with the prototype system, we can hypothesise both thymidine and guanosine alkylation. The prototype studies indicate that the main alkylation sites are the N1H and O6, both facing the cationic canal inside a G4. This confirms the dynamics of G4s, which are able to breathe and open up as already demonstrated by previous studies, allowing the alkylation sites to be accessible for the labelling.<sup>179,244</sup> It would be interesting to observe the impact of the alkylation on the folding and stability of the G4s.

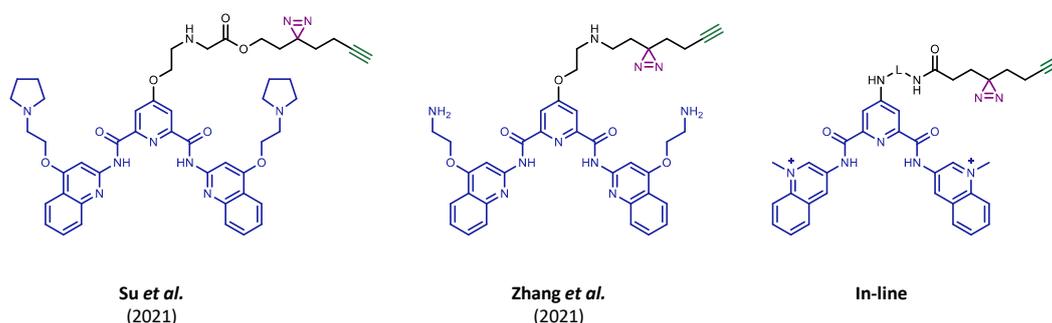


Figure 168: Structure of the *in-line* shared scaffold.

Recently, aliphatic diazirine and terminal alkyne were both tethered to PDB and PDS (Figure 168).<sup>202,203</sup> These ligands were employed in a photoaffinity labelling strategy to investigate the interaction between G4 proteins and G4 structures. However, the potential binding competition between the ligand and the proteins was not considered, which could have biased to a certain degree the obtained results. Additionally, they completely neglected the question of the direct G4 structure labelling. It would be interesting to evaluate the competitive labelling of G4 structures in presence of G4 proteins. The Chem-IP-seq methodology remains to be fully developed, which will allow us to identify the exact G4s the ligands can access in a dynamic system, such as the one of live cells, and will also provide insights into G4 formation as a function of the cellular type and state.

## Appendix

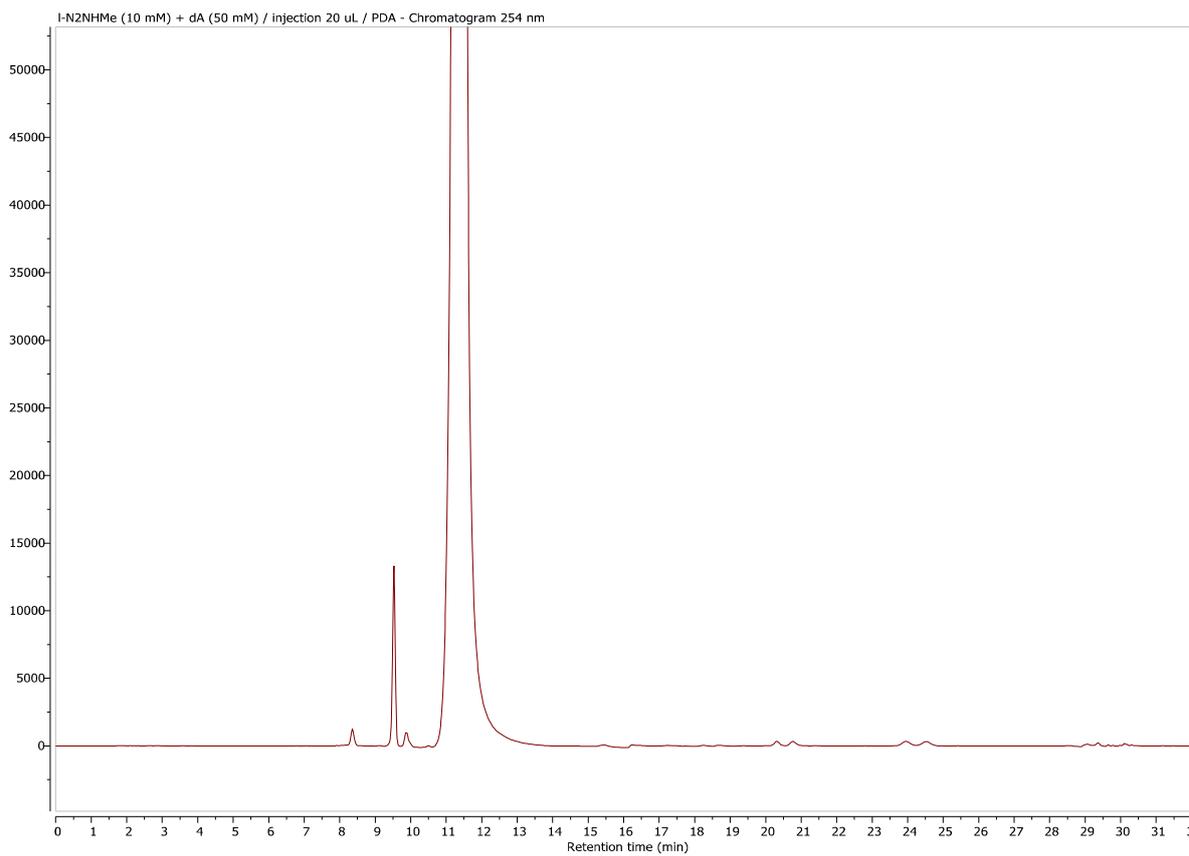


Figure S1: Chromatogram of the reaction between *I-N<sub>2</sub>NHMe* (10 mM) and *dA* (50 mM) after 30 min irradiation at 366 nm.

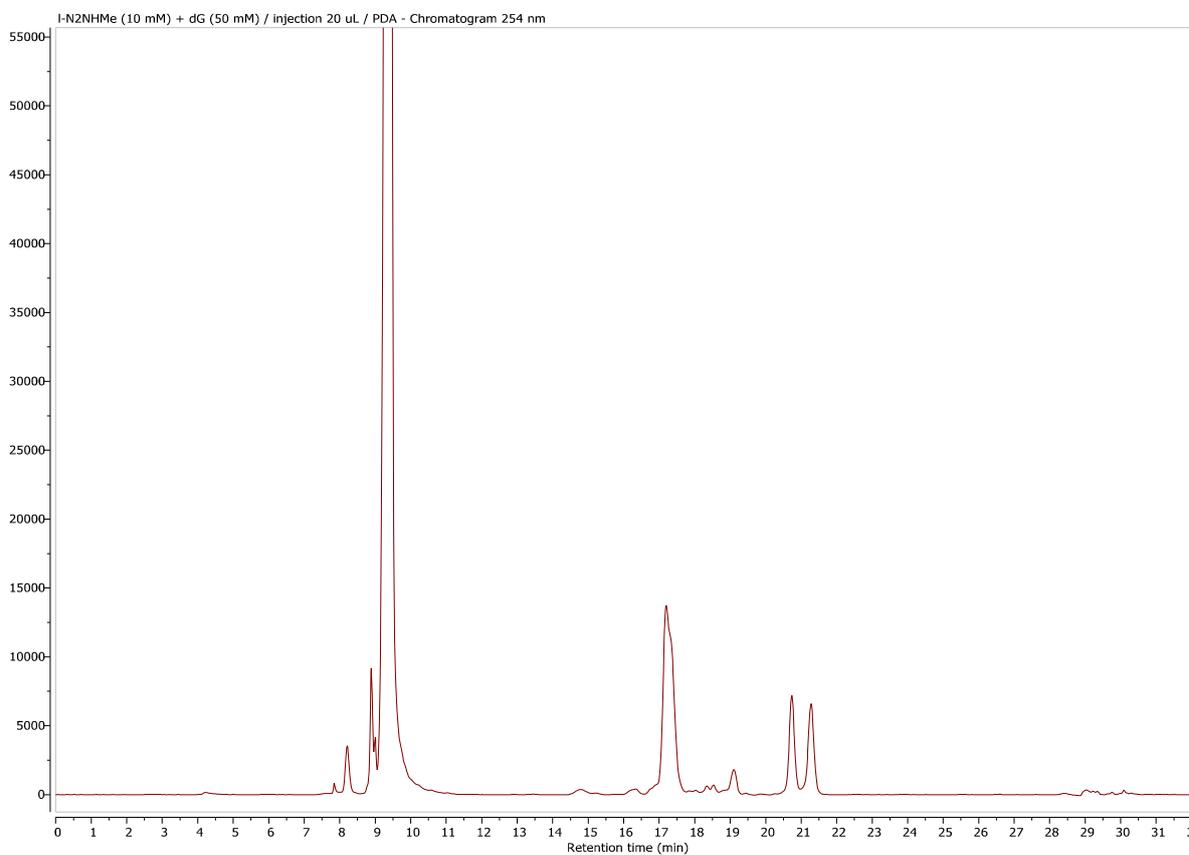


Figure S2: Chromatogram of the reaction between *I-N<sub>2</sub>NHMe* (10 mM) and *dG* (50 mM) after 30 min irradiation at 366 nm.

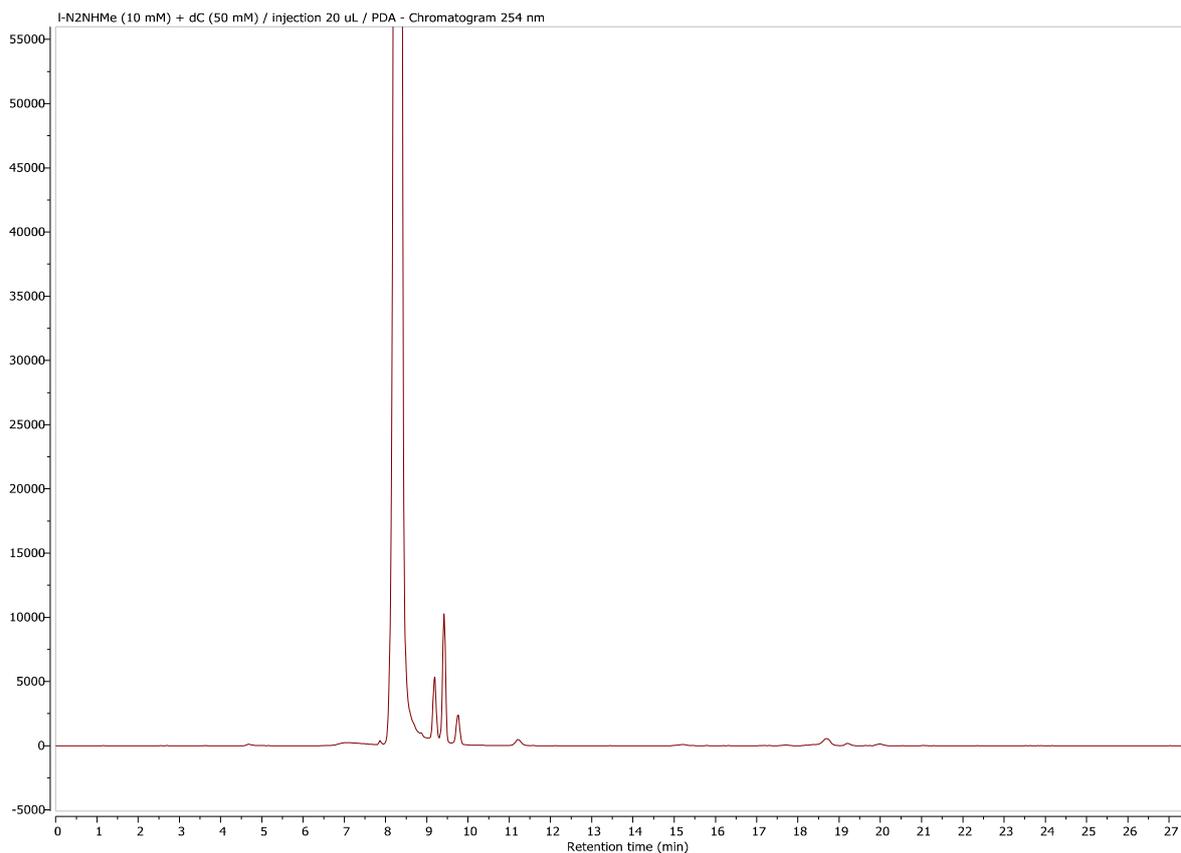


Figure S3: Chromatogram of the reaction between *I-N<sub>2</sub>NHMe* (10 mM) and *dC* (50 mM) after 30 min irradiation at 366 nm.

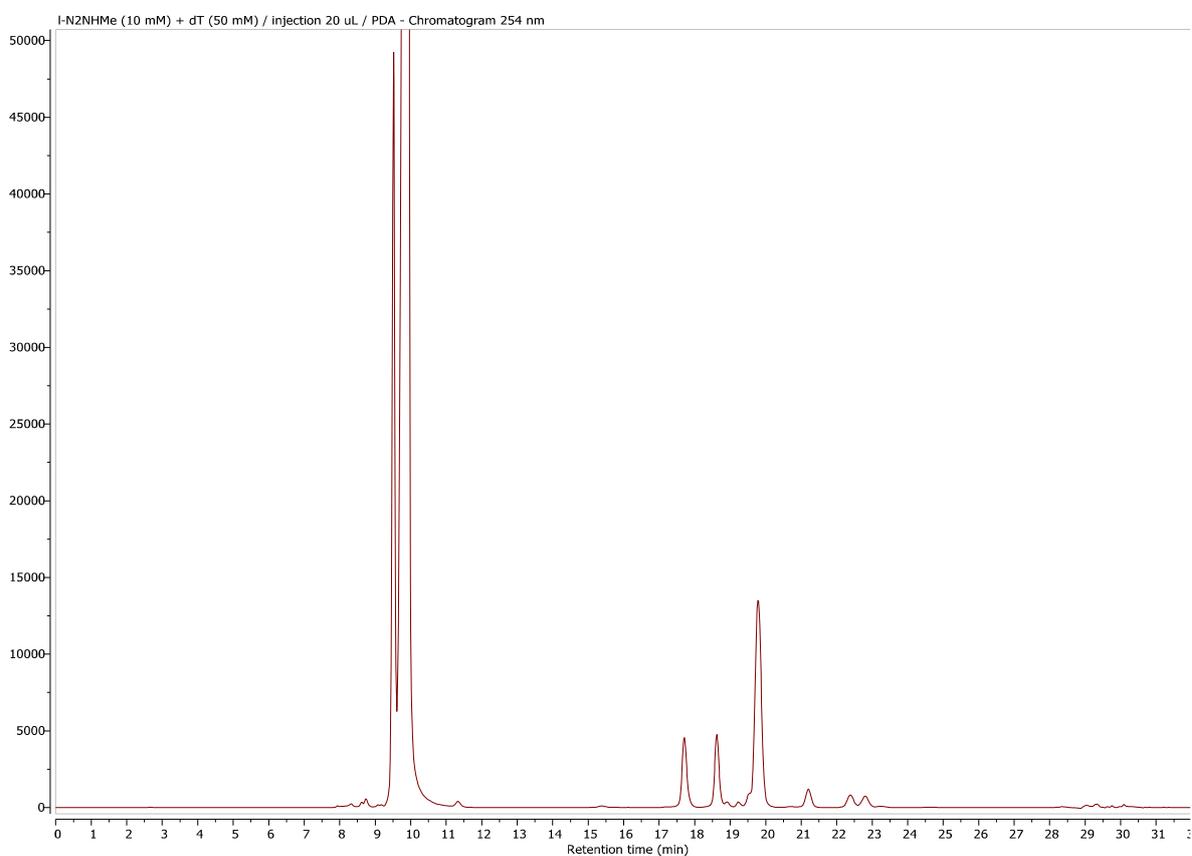


Figure S4: Chromatogram of the reaction between *I-N<sub>2</sub>NHMe* (10 mM) and *dT* (50 mM) after 30 min irradiation at 366 nm.



## Experimental section



## Synthesis

### **Synthesis and characterisation**

#### *For the photoactivatable G4 ligands and prototypes synthesis:*

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. 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 visualisation at 254 and 366 nm. Flash chromatography was performed with silica gel 60 (40-63  $\mu\text{m}$ , Merk). Deuterated solvents DMSO- $d_6$  and  $\text{CDCl}_3$  were purchased from Eurisotop. All  $^1\text{H}$  NMR and APT  $^{13}\text{C}$  NMR spectra were recorded on a Bruker Advance 300 MHz spectrometer using deuterated solvents and TMS as internal standard. The spectra are reported in ppm and referenced to deuterated DMSO (2.50 ppm for  $^1\text{H}$ , 39.52 ppm for  $^{13}\text{C}$ ) or deuterated  $\text{CHCl}_3$  (7.26 ppm for  $^1\text{H}$ , 77.16 ppm for  $^{13}\text{C}$ ). Data for  $^1\text{H}$  NMR are reported as follows: chemical shift ( $\delta$  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 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  $\mu\text{m}$ ) and coupled to a low resolution mass spectrometry (ESI-MS) micromass ZQ 2000 (Waters).

Gradient HPLC injection: flow 0.6 mL/min; gradient elution: eluent A% H<sub>2</sub>O 0.1% FA (formic acid), eluent B% ACN 0.1% FA: 0 min (95/5), 7 min (5/95), 10 min (95/5), 11 min (95/5).

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](http://www.icsn.cnrs-gif.fr)).

#### *For the analytic and preparative photoreaction experiments:*

All commercially available chemicals were bought from Sigma-Aldrich, Fisher Scientific and TCI as reagent grade and were used without further purification. NMR spectra were recorded on a Bruker Avance Neo 700MHz NMR spectrometer equipped with a 5 mm helium-cooled triple resonance cryoprobe and using deuterated methanol  $\text{CD}_3\text{OD}$  purchased from Aldrich-sigma. The spectra are reported in ppm and reference to deuterated methanol (3.31 ppm for  $^1\text{H}$  and 49.00 ppm for  $^{13}\text{C}$ ). Data for  $^1\text{H}$  NMR are reported as follows: chemical shift ( $\delta$  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. GC-MS, analytical and preparative

photoreaction were realised inside glass test tube. Analytical experiments were carried out on i) an Agilent system SERIES 1260 unit equipped with an Xselect HSS T3 C18 column (50x4.6 mm) for Bph-COOH photoreaction (method A), and ii) a Waters delta 600 equipped with a Phenomenex Synergi™ Fusion-RP C18 column (250x10 mm, 4 µm) for Bph-CONEt<sub>2</sub> (method B) and *l*-N2NHMe (method C) photoreaction. Preparative experiments were carried out on i) a Shimadzu Prominence Modular HPLC for Bph-CONEt<sub>2</sub> (method D), ii) on a Waters delta 600 for *l*-N2NHMe with thymidine (method E), and iii) on an Agilent system SERIES 1260 for *l*-N2NHMe with 2'-deoxyguanosine (method F), equipped with a Phenomenex Synergi™ Fusion-RP C18 column (250x10 mm, 4 µm). UV-visible spectra were recorded on a Cary Eclipse 300 spectrophotometer with the photoactivatable G4 ligands and on a Cary Eclipse 60 spectrophotometer with the photoactivatable precursor and prototype.

Gradient HPLC injection A: flow 1 mL/min; gradient elution: eluent A% - aqueous buffer with 25 mM ammonium acetate, eluent B% - Acetonitrile: 0 min (100/0), 1 min (100/0), 6 min (90/10), 8 min (90/10), 16 min (60/40), 26 min (60/40), 30 min (0/100), 32 min (100/0), 34 min (100/0), 36 min (100/0).

Gradient HPLC injection B: flow 6 mL/min; gradient elution: eluent A% - aqueous buffer with 25 mM ammonium acetate, eluent B% - Acetonitrile: 0 min (100/0), 2 min (100/0), 20 min (70/30), 40 min (60/40), 44 min (90/10), 46 min (90/10), 48 min (100/0), 52 min (100/0).

Gradient HPLC injection C: flow 6 mL/min; gradient elution: eluent A% - aqueous buffer with 25 mM ammonium acetate, eluent B% - Acetonitrile: 0 min (100/0), 2 min (100/0), 5 min (90/10), 10 min (90/10), 15 min (80/20), 20 min (80/20), 25 min (75/25), 27 min (20/80), 28 min (20/80), 30 min (100/0), 32 min (100/0).

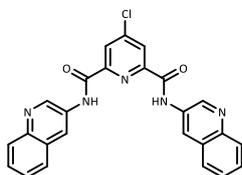
Gradient HPLC injection D: flow 5 mL/min; gradient elution: eluent A% - aqueous buffer with 25 mM ammonium acetate, eluent B% - Acetonitrile: 0 min (100/0), 2 min (100/0), 20 min (70/30), 40 min (60/40), 44 min (90/10), 46 min (90/10), 48 min (100/0), 52 min (100/0).

Gradient HPLC injection E: flow 8 mL/min; gradient elution: eluent A% - aqueous buffer with 25 mM ammonium acetate, eluent B% - Acetonitrile: 0 min (100/0), 2 min (100/0), 5 min (90/10), 10 min (90/0), 15 min (50/50), 25 min (50/50), 27 min (20/80), 28 min (20/80), 30 min (100/0), 32 min (100/0).

Gradient HPLC injection F: flow 8 mL/min; gradient elution: eluent A% - aqueous buffer with 25 mM ammonium acetate, eluent B% - Acetonitrile: 0 min (100/0), 2 min (100/0), 5 min (90/10), 10 min (90/10), 15 min (80/20), 20 min (80/20), 25 min (75/25), 27 min (20/80), 28 min (20/80), 30 min (100/0), 32 min (100/0).

### Photoactivatable G4 ligands synthesis:

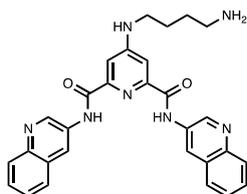
#### 4-chloro-*N*<sup>2</sup>,*N*<sup>6</sup>-bis(3-quinolyl)pyridine-2,6-dicarboxamide (PDC-Cl) (**3**)



Compound **3** was synthesised by following a previously reported method.<sup>213</sup> Under argon atmosphere, dry DMF (1 eq., 0.27 mL, 3.5 mmol) was added dropwise to a suspension of vacuum-dried chelidamic acid monohydrate (1 eq., 625 mg, 3.41 mmol) in SOCl<sub>2</sub> (32 eq., 8 mL, 110 mmol). The mixture was stirred under reflux for 16 h. The volatiles were removed in vacuo, giving the crude 4-chloropyridine-2,6-dicarbonyl chloride **2** as an off-white solid. This intermediate was dissolved in dry DMF (4.5 mL) and added in small portions to a mixture of 3-aminoquinoline (2.6 eq., 1.28 g, 8.88 mmol) and triethylamine (2.6 eq., 1.23 mL, 8.87 mmol) in dry DMF (8 mL), stirred at 0 °C. The mixture was stirred at r.t. overnight and poured into water (200 mL). The white solid was collected by filtration, thoroughly washed with water and dried in vacuo, to give compound **3** (981 mg, 63%) as a white solid.

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ (ppm): 11.45 (s, 2H), 9.36 (d, *J* = 2.4 Hz, 2H), 8.98 (d, *J* = 2.1 Hz, 2H), 8.12 – 8.01 (m, 4H), 7.74 (t, *J* = 6.8 Hz, 2H), 7.65 (t, *J* = 7.1 Hz, 2H).

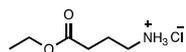
#### 4-[(4-aminobutyl)amino]-*N*<sup>2</sup>,*N*<sup>6</sup>-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (PDC-C4-NH<sub>2</sub>) (**5**)



Compound **5** was synthesised by modifying a previously reported method.<sup>179</sup> PDC-Cl **3** (1 eq., 193 mg, 0.43 mmol) was suspended in 1,4-butanediamine (82 eq., 0.88 mL, 34.8 mmol) and TEA (5 eq., 0.29 mL, 2.08 mmol) was next added. The resulting mixture was stirred at 70 °C for 3.5 h under argon atmosphere. After that time, the reaction was allowed to cool down to r.t. and concentrated under reduced pressure. An aqueous solution of Na<sub>2</sub>CO<sub>3</sub> 1% was added and the precipitate was filtered and washed with water. The afforded solid was suspended in a 1:1 mixture EtOAc/Et<sub>2</sub>O and concentrated three times, before proceeding with the filtration to obtain the expected product as a beige solid (76 mg, 35%).

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ (ppm): 11.31 (s, 2H), 9.37 (d, *J* = 2.1 Hz, 2H), 8.97 (d, *J* = 1.7 Hz, 2H), 8.04 (d, *J* = 8.3 Hz, 3H), 7.72 (t, *J* = 7.3 Hz, 2H), 7.64 (t, *J* = 7.4 Hz, 2H), 7.60 – 7.53 (m, 2H), 2.63 (t, *J* = 6.8 Hz, 2H), 1.66 (quint, *J* = 6.6 Hz, 2H), 1.50 (quint, *J* = 7.0 Hz, 2H).

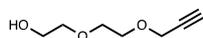
### 4-ethoxy-4-oxobutan-1-aminium (12)



Compound **12** was synthesised by following a previously reported method.<sup>245</sup> Thionyl chloride (2.5 eq., 5.3 mL, 72.7 mmol) was added dropwise at 0 °C to a stirred suspension of commercially available 4-aminobutyric acid **11** (1 eq., 3.0 g, 29.0 mmol) in EtOH (54 mL). The pale-yellow homogeneous reaction mixture was heated to reflux for 5 h. Volatiles were removed under vacuum and the resulting solid was washed with Et<sub>2</sub>O, yielding the expected product as a white solid (4.6 g, 94%).

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) 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).

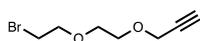
### 2-(2-(prop-2-yn-1-yloxy)ethoxy)ethan-1-ol (14)



Compound **14** was synthesised by following a previously reported method.<sup>246</sup> To a suspension of t-BuOK (0.5 eq., 598 mg, 5.34 mmol) in dry THF (30 mL) was added diethylene glycol **13** (1 eq., 1.11 g, 10.47 mmol) at 0 °C under argon atmosphere. The reaction mixture was allowed to stir at r.t. for 30 min then Propargyl Bromide (0.5 eq., 0.58 mL, 5.34 mmol) in dry THF (3 mL) was added dropwise. The heterogeneous reaction mixture was left to stir for 16 h at r.t. The reaction mixture was diluted with THF and filtered through a celite pad. The crude mixture was concentrated under vacuum and purified by SiO<sub>2</sub> gel chromatography eluted with DCM/EtOAc 1:1 to give the expected product as a yellowish oil (520 mg, 68%).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 4.22 (d, *J* = 2.4 Hz, 2H), 3.75 – 3.65 (m, 6H), 3.63 – 3.56 (m, 2H), 3.10 (t, *J* = 5.7 Hz, 1H), 2.50 (t, *J* = 2.4 Hz, 1H).

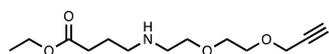
### 3-(2-(2-bromoethoxy)ethoxy)prop-1-yne (15)



Compound **15** was synthesised by following a previously reported method.<sup>247</sup> Previously prepared compound **14** (1 eq., 0.41 g, 2.87 mmol) was dissolved in dry DCM (10 mL) and the solution was cooled to 0 °C under argon atmosphere. PPh<sub>3</sub> (1.2 eq., 0.9 g, 3.44 mmol) was added and the solution was stirred for 15 min at 0 °C. CBr<sub>4</sub> (1.2 eq., 1.14 g, 3.44 mmol) was added and the solution was stirred for additional 30 min at 0 °C before being allowed to warm up to r.t. The reaction mixture was stirred overnight and volatiles were removed under vacuum. The crude mixture was purified by SiO<sub>2</sub> gel chromatography eluted with hexane/EtOAc 9:1 to give the expected product as a yellowish oil (427 mg, 72%).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 4.22 (d, *J* = 2.3 Hz, 2H), 3.82 (t, *J* = 6.3 Hz, 2H), 3.71 (s, 4H), 3.48 (t, *J* = 6.3 Hz, 2H), 2.45 (t, *J* = 2.2 Hz, 1H).

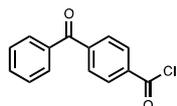
#### Ethyl 4-((2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)amino)butanoate (**16**)



To a mixture of previously prepared compound **12** (3 eq., 1.06 g, 6.36 mmol) and TEA (4 eq., 1.18 mL, 8.49 mmol) in dry ACN (20 mL) was added dropwise at 0 °C, a solution of **15** (1 eq., 439 mg, 2.12 mmol) in dry ACN (2 mL) 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 the crude mixture was filtered with THF, then purified by SiO<sub>2</sub> gel chromatography eluted with DCM/EtOH from 0 to 4% to give the expected product as a yellowish oil (263 mg, 48%).

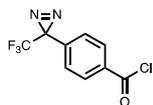
<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 6.36 (s, 1H), 4.21 (q, *J* = 8.3 Hz, 2H), 4.14 (dt, *J* = 7.1 Hz, 5.7 Hz, 2H), 3.97 – 3.90 (t, *J* = 5.0 Hz, 2H), 3.76 – 3.66 (m, 4H), 3.26 (t, *J* = 5.0 Hz, 2H), 3.18 (t, *J* = 7.2 Hz, 2H), 2.56 – 2.45 (m, 3H), 2.20 (quint, *J* = 7.5 Hz, 2H), 1.27 (t, *J* = 7.1 Hz, 3H).; <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 172.84, 79.32, 75.06, 70.15, 68.97, 65.91, 60.94, 58.42, 47.50, 47.23, 31.43, 21.25, 14.17.

#### 4-benzoylbenzoyl chloride (**21**)



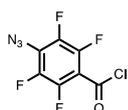
A mixture of commercially available carboxylic acid benzophenone derivative **18** (1 eq.) in thionyl chloride (70 eq.) was heated at reflux for 2 h under argon atmosphere. Volatiles were removed under vacuum yielding quantitatively the expected product without any further purification.

#### 4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoyl chloride (**22**)



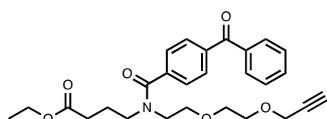
4-[3-(trifluoromethyl)-3H-diazirin-3-yl]benzoic acid **19** (1 eq.) was dissolved in SOCl<sub>2</sub> (27.5 eq.) and the mixture was stirred for 20 h under argon atmosphere. Volatiles were removed under vacuum yielding quantitatively the expected product without any further purification.

#### 4-azido-2,3,5,6-tetrafluorobenzoyl chloride (23)



The commercially available carboxylic acid **20** (1 eq.) was dissolved in Et<sub>2</sub>O and was reacted with PCl<sub>5</sub> (1.1 eq.) under argon atmosphere for 2 h. Volatiles were removed under vacuum yielding quantitatively to the expected product without any further purification.

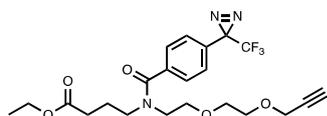
#### Ethyl 4-(4-benzoyl-N-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)benzamido)butanoate (24)



To a mixture of **16** (1 eq., 492 mg, 1.91 mmol) and TEA (2 eq., 0.53 mL, 3.82 mmol) in dry DCM (8 mL) was added dropwise a solution of freshly prepared **21** (2 eq., 936 mg, 3.82 mmol) in dry DCM (8 mL) at 0 °C under argon. The mixture was stirred at 0 °C for 1 h before being allowed to warm to r.t. The mixture was stirred for 16 h under argon atmosphere before being cooled down to 0 °C. Water was added dropwise followed by addition of a saturated solution of NaHCO<sub>3</sub>. The organic layer was separated and the aqueous phase was extracted several times with DCM. The organic layers were combined and dried over MgSO<sub>4</sub>, filtrated and concentration under vacuum. The mixture was purified by SiO<sub>2</sub> gel chromatography eluted with DCM/EtOAc from 100:0 to 80:20 to give the expected product as a pale-yellow oil (370 mg, 42%).

Presence of two conformers: <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 7.82 (t, *J* = 7.1 Hz, 4H), 7.61 (t, *J* = 7.3 Hz, 1H), 7.51 (q, *J* = 7.4 Hz, 3H), 4.18 (t, *J* = 8.6 Hz, 3H), 4.05 (d, *J* = 6.9 Hz, 1H), 3.79 – 3.53 (m, 8H), 3.47 (s, 1H), 3.39 (d, *J* = 7.1 Hz, 1H), 2.42 (d, *J* = 6.1 Hz, 2H), 2.15 (t, *J* = 7.0 Hz, 1H), 2.03 (t, *J* = 7.4 Hz, 1H), 1.86 (t, *J* = 7.0 Hz, 1H), 1.24 (m, 3H).; <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 196.2, 171.5, 171.2, 171.1, 140.7, 140.6, 138.4, 138.2, 137.3, 133.0, 132.7, 132.6, 130.4, 130.1, 128.7, 128.4, 127.3, 126.9, 126.8, 126.5, 126.4, 76.2, 70.5, 70.4, 69.6, 69.2, 68.9, 60.7, 58.6, 49.9, 48.5, 44.8, 31.9, 29.9, 23.9, 22.8, 14.3.; LC/MS: Retention time 6.58 min, m/z (ESI+) = 466.3 [M+H]<sup>+</sup>.

#### Ethyl 4-(N-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamido) butanoate (25)

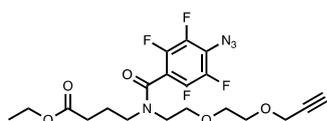


To a mixture of **16** (1 eq., 200 mg, 0.78 mmol) and TEA (2 eq., 0.22 mL, 1.56 mmol) in dry DCM (8 mL) was added dropwise a solution of freshly prepared **22** (1.5 eq., 290 mg, 1.17 mmol) in

dry DCM (2 mL) at 0 °C under argon. The mixture was stirred at 0 °C for 1 h before being allowed to warm to r.t. The mixture was stirred for 16 h under argon atmosphere before being cooled down to 0 °C. Water was added dropwise followed by addition of a saturated solution of NaHCO<sub>3</sub>. The organic layer was separated and the aqueous phase was extracted several times with DCM. The organic layers were combined and dried over MgSO<sub>4</sub>, filtrated and concentration under vacuum. The mixture was purified by SiO<sub>2</sub> gel chromatography eluted with DCM/EtOAc 9:1 to give the expected product as a pale-yellow oil (160 mg, 44%).

Presence of two conformers: <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 7.50 – 7.35 (m, 2H), 7.22 (d, *J* = 8.0 Hz, 2H), 4.23-4.10 (m, 3H), 4.07 – 3.97 (m, 1H), 3.80-3.47 (m, 8H), 3.41 (br s, 1H), 3.32 (br s, 1H), 2.48-2.35 (m, 2H), 2.13 (s, 1H), 1.99 (s, 1H), 1.83 (s, 1H), 1.32-1.12 (m, 3H).; <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 173.3, 171.1, 138.2, 130.1, 127.7, 127.3, 127.2, 126.7, 123.9, 120.3, 79.6, 74.8, 70.5, 70.4, 69.2, 68.6, 60.6, 58.5, 49.6, 48.9, 44.9, 44.7, 31.6, 31.1, 28.7, 28.2, 23.8, 22.7, 14.3.; LC/MS: Retention time 6.94 min, *m/z* (ESI+) = 470.5 [M+H]<sup>+</sup>.

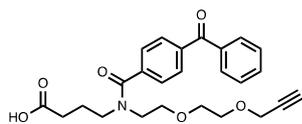
#### **Ethyl 4-(4-azido-2,3,5,6-tetrafluoro-*N*-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)benzamido)butanoate (26)**



To a mixture of **16** (1 eq., 310 mg, 1.2 mmol) and TEA (2 eq., 0.35 mL, 2.5 mmol) in dry DCM (10 mL) was added dropwise a solution of freshly prepared **23** (1.5 eq., 458 mg, 1.8 mmol) in dry DCM (5 mL) at 0 °C under argon. The mixture was stirred at 0 °C for 1 h before being allowed to warm to r.t. The mixture was stirred 16 h under argon atmosphere before being cooled down to 0 °C. Water was added dropwise followed by a saturated solution of NaHCO<sub>3</sub>. The organic layer was separated and the aqueous phase was extracted several times with DCM. The organic layers were combined and dried over MgSO<sub>4</sub>, filtrated and concentration under vacuum. The mixture was purified by SiO<sub>2</sub> gel chromatography eluted with DCM/EtOAc 0 to 3% to give the expected product as a pale-yellow oil (187 mg, 33%).

Presence of two conformers: <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 4.22 - 4.10 (m, 3H), 4.05 (q, *J* = 7.1 Hz, 1H), 3.83 - 3.46 (m, 8H), 3.44 – 3.31 (m, 2H), 2.46 – 2.34 (m, 2H), 2.20 (t, *J* = 6.8 Hz, 1H), 1.99 (quint, *J* = 7.2 Hz, 1H), 1.83 (quint, *J* = 7.3 Hz, 1H), 1.24 (m, 3H).; <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 144.54, 144.27, 142.30, 142.27, 138.91, 80.26, 74.59, 70.66, 70.52, 69.17, 69.10, 68.57, 60.72, 58.51, 49.31, 48.88, 45.43, 31.28, 30.40, 23.35, 22.60, 14.38.; LC/MS: Retention time 6.81 min, *m/z* (ESI+) = 475.3 [M+H]<sup>+</sup>.

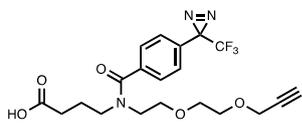
#### 4-(4-benzoyl-N-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)benzamido)butanoic acid (27)



To the previously synthesised compound **24** (1 eq., 337 mg, 0.72 mmol) in EtOH (2 mL) / DCM (2 mL) was added dropwise a solution of NaOH (2 eq., 60 mg, 1.49 mmol) in water (1 mL) at 0 °C. The reaction mixture was allowed to warm to r.t. and stirred 16 h. Volatiles were removed under vacuum and the residue was solubilised in water, washed 3 times with EtOAc and the aqueous layer was then acidified to pH = 1 by addition of an aqueous solution of 10% HCl. The aqueous layer was extracted several times with EtOAc and the organic layers were dried over MgSO<sub>4</sub>, filtered and concentrated to give the expected product as a colourless oil (277 mg, 88%).

Presence of two conformers: <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 7.81 (t, *J* = 7.1 Hz, 4H), 7.61 (t, *J* = 7.4 Hz, 1H), 7.51 (q, *J* = 7.2 Hz, 4H), 4.19 (d, *J* = 7.7 Hz, 2H), 3.82 – 3.35 (m, 10H), 2.44 (s, 2H), 2.19 (s, 1H), 2.02 (s, 1H), 1.85 (s, 1H).; <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 196.1, 171.7, 171.2, 140.3, 138.2, 137.1, 132.8, 130.1, 128.4, 127.0, 126.5, 70.4, 69.1, 68.7, 58.4, 49.4, 48.9, 44.9, 44.7, 29.7, 23.5, 22.6.; LC/MS: Retention time 5.84 min, *m/z* (ESI+) = 438.19 [M+H]<sup>+</sup>.

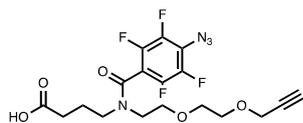
#### 4-(N-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamido) butanoic acid (28)



To the previously synthesised compound **25** (1 eq., 160 mg, 0.34 mmol) in EtOH (1.1 mL) / DCM (1.1 mL) was added dropwise a solution of NaOH (2 eq., 27 mg, 0.68 mmol) in water (0.6 mL) at 0 °C. The reaction mixture was allowed to warm to r.t. and stirred 16 h. Volatiles were removed under vacuum and the residue was solubilised in water, washed 3 times with EtOAc and the pH of the aqueous layer was adjusted to pH = 1 by adding an aqueous solution of 10% HCl. The aqueous layer was extracted several times with EtOAc and the organic layers were dried over MgSO<sub>4</sub>, filtered and concentrated to give the expected product as a colourless oil (130 mg, 86%).

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ (ppm): 7.56 – 7.42 (m, 2H), 7.31 (d, *J* = 7.8 Hz, 2H), 4.13 (s, 2H), 3.62 – 3.37 (m, 9H), 3.28 (s, 2H), 3.17 (s, 1H), 2.28 (s, 1H), 2.03 (s, 1H), 1.81 (s, 1H), 1.67 (s, 1H).; <sup>13</sup>C NMR (75 MHz, DMSO-d<sub>6</sub>) δ (ppm): 174.3, 169.9, 138.9, 130.0, 128.0, 127.9, 127.3, 126.3, 123.7, 120.0, 116.4, 80.3, 79.1, 77.1, 69.5, 68.5, 67.5, 57.5, 48.2, 43.7, 31.2, 30.5, 28.3, 27.8, 23.4, 22.4.; LC/MS: Retention time 6.24 min, *m/z* (ESI+) = 442.5 [M+H]<sup>+</sup>.

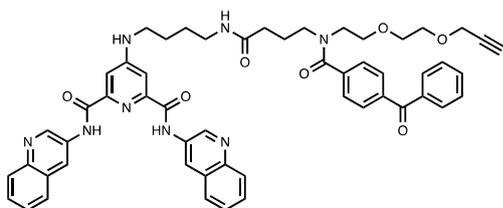
**4-(4-azido-2,3,5,6-tetrafluoro-N-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethyl)benzamido)butanoic acid (29)**



To the previously synthesised compound **26** (1 eq., 179 mg, 0.38 mmol) in EtOH (1.4 mL) / DCM (1.4 mL) was added dropwise a solution of NaOH (2 eq., 30.2 mg, 0.75 mmol) in water (0.7 mL) at 0 °C. The reaction mixture was allowed to warm to r.t. and stirred 16 h. Volatiles were removed under vacuum and the residue was solubilised in water, washed 3 times with EtOAc and the pH of the aqueous layer was adjusted to pH = 1 by adding a aqueous solution of 10% HCl. The aqueous layer was extracted several times with EtOAc and the organic layers were dried over MgSO<sub>4</sub>, filtered and concentrated to give the expected product as a colourless oil (120 mg, 71%).

Presence of two conformers: <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 4.18 (dd, *J* = 11.1 Hz, 2.4 Hz, 2H), 3.76 (d, *J* = 3.4 Hz, 2H), 3.70 – 3.62 (m, 4H), 3.57 (dd, *J* = 5.8 Hz, 2.8 Hz, 1H), 3.51 (t, *J* = 5.0 Hz, 1H), 3.44 – 3.32 (m, 2H), 2.50 – 2.41 (m, 2H), 2.31 (t, *J* = 6.5 Hz, 1H), 2.01 (quint, *J* = 7.1 Hz, 1H), 1.84 (quint, *J* = 7.0 Hz, 1H).; <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 178.1, 177.9, 177.5, 159.4, 144.5, 142.4, 139.0, 133.4, 77.4, 74.8, 70.7, 70.5, 69.3, 69.2, 69.1, 68.6, 58.5, 49.0, 45.8, 45.1, 30.9, 29.9, 23.1, 22.4.; LC/MS: Retention time 6.26 min, *m/z* (ESI+) = 447.4 [M+H]<sup>+</sup>.

**4-((10-(4-benzoylbenzoyl)-14-oxo-4,7-dioxa-10,15-diazanonadec-1-yn-19-yl)amino)-N<sup>2</sup>,N<sup>6</sup>-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (30)**

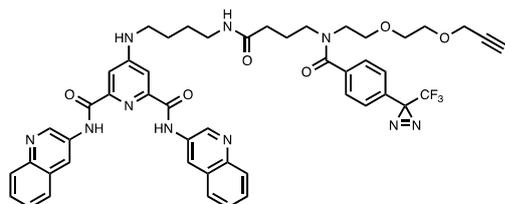


Compound **27** (1.2 eq., 91 mg, 0.21 mmol) was sheltered from light and dissolved in DMF (1 mL). DIEA (3.1 eq., 0.09 mL, 0.54 mmol) and PDC-C4-NH<sub>2</sub> **5** (1 eq., 103 mg, 0.17 mmol) was added under argon at 0 °C. The mixture was stirred 5 min and HATU (1.5 eq., 99 mg, 0.26 mmol) in DMF (0.7 mL) was slowly added. The reaction mixture was allowed to warm to r.t. and stirred 1 h. The volatiles were removed under vacuum. The crude was purified by SiO<sub>2</sub> gel chromatography eluted with DCM/EtOH from 0 to 8% to give the expected product as a white solid (88 mg, 55%).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 10.35 (s, 2H), 9.10 (s, 2H), 8.71 (s, 2H), 7.98 (d, *J* = 8.3 Hz, 2H), 7.63 (m, 9H), 7.40 (m, 10H), 5.96 (s, 1H), 4.16 – 4.02 (m, 2H), 3.74 – 3.48 (m, 6H), 3.46 – 3.38 (m, 3H), 3.38 – 3.27 (m, 3H), 3.16 (s, 3H), 2.39 (t, *J* = 2.3 Hz, 1H), 2.25 (d, *J* = 6.3 Hz, 2H), 2.05 – 1.89 (m, 3H), 1.42 (d, *J* = 7.5 Hz, 1H).; <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 195.3, 172.4,

171.6, 171.0, 170.1, 169.9, 163.2, 156.3, 146.1, 144.7, 141.0, 140.7, 137.1, 136.8, 132.9, 132.0, 129.7, 128.7, 128.6, 128.3, 127.9, 127.8, 127.2, 127.0, 126.5, 124.1, 80.3, 77.1, 69.5, 68.5, 67.6, 59.2, 58.9, 57.6, 53.6, 48.7, 48.3, 48.1, 43.9, 41.8, 39.5, 38.1, 32.9, 32.2, 31.3, 29.1, 28.8, 26.8, 25.6, 23.3, 22.1.; **LC/MS**: Retention time 6.85 min,  $m/z$  (ESI+) = 925.7 [M+H]<sup>+</sup>.

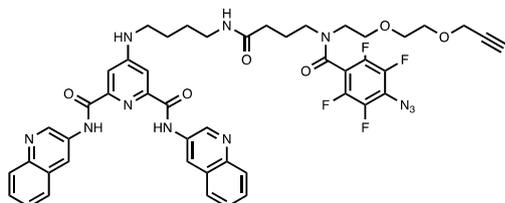
**4-((14-oxo-10-(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoyl)-4,7-dioxo-10,15-diazanonadec-1-yn-19-yl)amino)-N<sup>2</sup>,N<sup>6</sup>-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (31)**



Compound **28** (1.2 eq., 91 mg, 0.21 mmol) was sheltered from light and dissolved in DMF (1.5 mL). DIEA (3 eq., 0.085 mL, 0.52 mmol) and PDC-C4-NH<sub>2</sub> **5** (1 eq., 87 mg, 0.17 mmol) was added under argon at 0 °C. The mixture was stirred 5 min and HATU (1.5 eq., 98 mg, 0.26 mmol) in DMF (0.5 mL) was slowly added. The reaction mixture was allowed to warm to r.t. and stirred 1 h. The volatiles were removed under vacuum. The crude product was purified by SiO<sub>2</sub> gel chromatography eluted with DCM/EtOH from 0 to 8% to give the expected product as a white solid (117 mg, 74%).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 10.34 (s, 2H), 9.10 (s, 2H), 8.74 (s, 2H), 7.98 (d, *J* = 8.3 Hz, 2H), 7.67 (d, *J* = 8.0 Hz, 2H), 7.59 (t, *J* = 7.5 Hz, 2H), 7.47 (t, *J* = 7.2 Hz, 2H), 7.42 – 7.21 (m, 4H), 7.00 (d, *J* = 7.9 Hz, 2H), 6.01 (s, 1H), 4.14 - 4.03 (m, 2H), 3.66 – 3.24 (m, 13H), 3.14 (s, 2H), 2.43 (t, *J* = 2.4 Hz, 1H), 2.25 (s, 2H), 1.93 (s, 2H), 1.63 (s, 4H).; <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 173.1, 172.3, 163.5, 156.3, 149.1, 145.1, 145.1, 137.6, 131.7, 130.4, 128.9, 128.4, 128.1, 127.6, 126.7, 124.8, 124.0, 120.4, 79.7, 77.4, 70.6, 69.3, 68.5, 58.7, 49.1, 45.0, 42.9, 39.4, 33.9, 30.0, 28.8, 28.3, 27.4, 26.2, 23.8.; **LC/MS**: Retention time 6.95 min,  $m/z$  (ESI+) = 929.7 [M+H]<sup>+</sup>.

**4-((10-(4-azido-2,3,5,6-tetrafluorobenzoyl)-14-oxo-4,7-dioxo-10,15-diazanonadec-1-yn-19-yl)amino)-N<sup>2</sup>,N<sup>6</sup>-di(quinolin-3-yl)pyridine-2,6-dicarboxamide (32)**

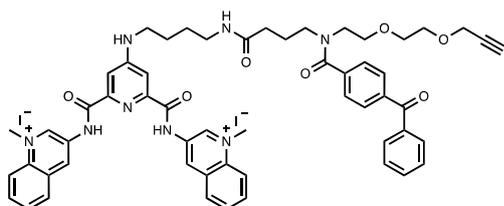


Compound **29** (1.2 eq., 117 mg, 0.26 mmol) was sheltered from light and dissolved in DMF (1.4 mL). DIEA (3 eq., 0.11 mL, 0.66 mmol) and PDC-C4-NH<sub>2</sub> **5** (1 eq., 110 mg, 0.22 mmol) was added under argon at 0 °C. The mixture was stirred 5 min and HATU (1.5 eq., 124 mg, 0.33 mmol) in DMF (0.8 mL) was slowly added. The reaction mixture was allowed to warm to r.t.

and stirred 1 h. The volatiles were removed under vacuum. The crude was purified by SiO<sub>2</sub> gel chromatography eluted with DCM/EtOH from 0 to 5% to give the expected product as a white solid (61 mg, 30%).

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ (ppm): 10.39 (s, 2H), 9.17 (d, *J* = 1.7 Hz, 2H), 8.86 (d, *J* = 1.5 Hz, 2H), 8.54 (dd, *J* = 4.4 Hz, 1.3 Hz, 1H), 8.22 (dd, *J* = 8.4 Hz, 1.2 Hz, 1H), 8.06 (t, *J* = 6.7 Hz, 2H), 7.68 (m, 3H), 7.52 (t, *J* = 7.2 Hz, 1H), 7.35 (d, *J* = 6.0 Hz, 1H), 7.27-7.19 (m, 1H), 5.98 (s, 1H), 4.10 (d, *J* = 2.3 Hz, 2H), 3.75 (d, *J* = 6.6 Hz, 2H), 3.64 – 3.09 (m, 13H), 2.39 (t, *J* = 2.3 Hz, 1H), 2.22 (t, *J* = 6.8 Hz, 1H), 1.42 (m, 6H).; <sup>13</sup>C NMR (75 MHz, DMSO-d<sub>6</sub>) δ (ppm): 171.8, 171.5, 163.6, 156.7, 151.4, 146.5, 145.1, 140.1, 135.1, 132.4, 129.3, 129.1, 128.8, 128.3, 128.2, 127.6, 124.5, 80.6, 69.9, 68.9, 68.8, 68.1, 58.0, 54.0, 48.9, 48.4, 44.9, 44.8, 42.3, 38.6, 33.0, 27.2, 26.1, 23.5, 22.6.; LC/MS: Retention time 6.89 min, *m/z* (ESI+) = 934.6 [M+H]<sup>+</sup>.

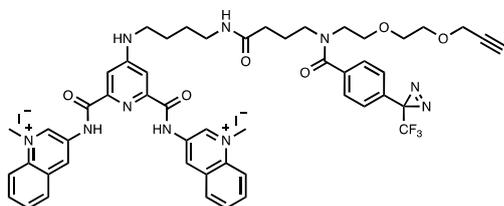
**3,3'-((4-((10-(4-benzoylbenzoyl)-14-oxo-4,7-dioxa-10,15-diazanonadec-1-yn-19-yl)amino)pyridine-2,6-dicarbonyl)bis(azanediyl))bis(1-methylquinolin-1-ium) iodide (PDC<sub>4</sub>-DEG-Bph (DEG/Bph)) (8)**



Mel (100 eq., 0.53 mL, 8.51 mmol) was added slowly to a solution of **30** (1 eq., 77.5 mg, 0.084 mmol) in DMF (0.65 mL) under argon atmosphere sheltered from the light. The reaction mixture was stirred for 16 h at 40 °C. Volatiles were removed under vacuum and the resulting solid was suspended in EtOH, washed several times with EtOH and dried with Et<sub>2</sub>O, affording compound **8** as a yellow solid (50.9 mg, 50%).

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ (ppm): 11.70 (s, 1H), 10.13 (s, 1H), 9.64 (s, 2H), 8.55 (dd, *J* = 8.2 Hz, 5.1 Hz, 4H), 8.22 (t, *J* = 7.9 Hz, 2H), 8.07 (t, *J* = 7.6 Hz, 2H), 7.72 (m, 7H), 7.54 (m, 6H), 4.78 (s, 6H), 4.14 (d, *J* = 10.2 Hz, 2H), 3.54 (m, 9H), 3.30 – 3.12 (m, 4H), 3.02 (d, *J* = 5.2 Hz, 1H), 2.55 (s, 1H), 2.16 (d, *J* = 6.9 Hz, 1H), 1.89 - 1.70 (m, 3H).; <sup>13</sup>C NMR (75 MHz, DMSO-d<sub>6</sub>) δ (ppm): 195.3, 171.6, 170.1, 163.2, 156.5, 144.7, 141.0, 137.1, 136.7, 135.7, 134.1, 133.9, 132.9, 132.4, 130.4, 129.9, 129.7, 129.2, 128.6, 127.0, 126.5, 119.3, 80.3, 77.1, 69.5, 68.5, 67.6, 57.5, 46.2, 41.8, 38.1, 34.4, 32.9, 26.2, 25.5.; LC/MS: Retention time 5.11 min, *m/z* (ESI+) = 954.1 [M-H]<sup>+</sup>.; HRMS (ESI, *m/z*) calcd. for formula C<sub>56</sub>H<sub>58</sub>N<sub>8</sub>O<sub>7</sub><sup>+</sup> [M-I]<sup>+</sup> = 1081.3468; found = 1081.3521.

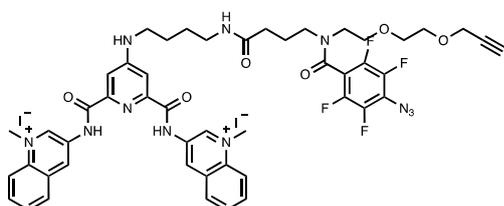
**3,3'-((4-((14-oxo-10-(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoyl)-4,7-dioxo-10,15-diazanonadec-1-yn-19-yl)amino)pyridine-2,6-dicarbonyl)bis(azanediy))bis(1-methylquinolin-1-ium) iodide (PDC<sub>4</sub>-DEG-N2 (DEG/N2)) (9)**



Mel (100 eq., 0.71 mL, 11.4 mmol) was added slowly to a solution of **31** (1 eq., 105 mg, 0.11 mmol) in DMF (0.9 mL) under argon atmosphere sheltered from the light. The reaction mixture was stirred for 16 h at 35 °C. Volatiles were removed under vacuum and the resulting solid was suspended in EtOH, washed several times with EtOH and dried with Et<sub>2</sub>O, affording compound **9** as a yellow solid (103 mg, 75%).

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ (ppm): 11.72 (s, 2H), 10.14 (d, *J* = 1.7 Hz, 2H), 9.66 (s, 2H), 8.56 (dd, *J* = 8.4 Hz, 2.7 Hz, 4H), 8.24 (t, *J* = 8.0 Hz, 2H), 8.08 (t, *J* = 7.6 Hz, 2H), 7.91 (d, *J* = 13.2, 1H), 7.75 (m, 2H), 7.59 – 7.43 (m, 3H), 7.31 (d, *J* = 8.0 Hz, 2H), 4.79 (s, 6H), 4.13 (s, 2H), 3.65 – 3.39 (m, 13H), 3.29 (s, 3H), 3.15 (s, 2H), 3.02 (s, 1H), 2.56 (t, *J* = 5.6 Hz, 1H), 2.15 (t, *J* = 7.4 Hz, 1H), 1.86 (m, 2H).; <sup>13</sup>C NMR (75 MHz, DMSO-d<sub>6</sub>) δ (ppm): 171.6, 169.8, 163.2, 156.6, 144.8, 138.9, 138.6, 135.7, 134.1, 133.9, 132.4, 130.4, 129.9, 129.2, 127.9, 127.3, 126.3, 123.6, 120.0, 119.3, 77.1, 69.4, 68.5, 67.6, 67.4, 57.5, 48.1, 46.2, 43.9, 41.8, 34.4, 32.8, 28.3, 26.7, 26.2, 25.5, 23.2.; LC/MS: Retention time 5.33 min, *m/z* (ESI<sup>+</sup>) = 957.8 [M-H]<sup>+</sup>.; HRMS (ESI, *m/z*) calcd. for formula C<sub>51</sub>H<sub>53</sub>F<sub>3</sub>IN<sub>10</sub>O<sub>6</sub><sup>+</sup> [M-I]<sup>+</sup> = 1085.3141; found = 1085.3181.

**3,3'-((4-((10-(4-azido-2,3,5,6-tetrafluorobenzoyl)-14-oxo-4,7-dioxo-10,15-diazanonadec-1-yn-19-yl)amino)pyridine-2,6-dicarbonyl)bis(azanediy))bis(1-methylquinolin-1-ium) iodide (PDC<sub>4</sub>-DEG-N3 (DEG/N3)) (10)**



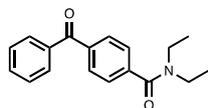
Mel (100 eq., 0.33 mL, 5.3 mmol) was added slowly to a solution of **32** (1 eq., 48 mg, 0.05 mmol) in ACN (4 mL) under argon atmosphere sheltered from the light. The reaction mixture was stirred for 16 h at 40 °C. Volatiles were removed under vacuum and the resulting solid was suspended in EtOH, washed several times with EtOH and dried with Et<sub>2</sub>O, affording compound **10** as a yellow solid (22 mg, 35%).

<sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ (ppm): 11.72 (s, 2H), 10.13 (s, 2H), 9.65 (s, 2H), 8.55 (d, *J* = 8.4 Hz, 4H), 8.23 (t, *J* = 7.4 Hz, 2H), 8.08 (t, *J* = 7.3 Hz, 3H), 7.93 – 7.52 (m, 3H), 4.78 (s, 6H), 4.11 (t, *J* = 3.0 Hz, 2H), 3.74 – 2.99 (m, 22H), 2.14 (t, *J* = 7.6 Hz, 1H), 1.96 (t, *J* = 6.6 Hz, 1H), 1.82 (t, *J*

= 7.3 Hz, 1H).; **<sup>13</sup>C NMR** (75 MHz, DMSO-d<sub>6</sub>) δ (ppm): 171.3, 171.1, 163.2, 156.5, 144.7, 135.7, 134.1, 133.9, 132.4, 130.4, 129.9, 129.2, 119.2, 77.1, 69.5, 68.4, 67.6, 67.5, 57.5, 47.9, 46.2, 44.5, 44.3, 41.8, 38.1, 38.0, 32.4, 26.7, 25.5, 23.0; **LC/MS**: Retention time 5.27 min, m/z (ESI+) = 962.9 [M-H]<sup>+</sup>; **HRMS** (ESI, m/z) calcd. for formula C<sub>49</sub>H<sub>49</sub>F<sub>4</sub>IN<sub>11</sub>O<sub>6</sub><sup>+</sup> [M-I]<sup>+</sup> = 1090.2843; found = 1090.2888.

## Prototypes synthesis:

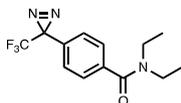
### 4-benzoyl-*N,N*-diethylbenzamide (Bph-CONEt<sub>2</sub>) (35)



To a solution of 4-benzoylbenzoic acid **18** (1 eq., 600 mg, 2.65 mmol), diethylamine (1.1 eq., 0.3 mL, 2.91 mmol) and HOBt (1.1 eq., 394 mg, 2.92 mmol) in DMF (7 mL) was added EDCI-HCl (1.1 eq., 560 mg, 2.92 mmol) at 0 °C under argon atmosphere. The reaction mixture was allowed to warm to room temperature and was stirred 2 h. The mixture was poured into a sat. aq. NaHCO<sub>3</sub> solution then extract with EtOAc. The organic phase was washed three times with water. Dried by MgSO<sub>4</sub> then concentrated to give the expected product as a white powder (554 mg, 74%).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 7.82 (t, *J* = 7.5 Hz, 4H), 7.61 (t, *J* = 7.4 Hz, 1H), 7.50 (t, *J* = 7.6 Hz, 4H), 3.58 (d, *J* = 6.7 Hz, 2H), 3.26 (d, *J* = 6.6 Hz, 2H), 1.27 (d, *J* = 6.5 Hz, 3H), 1.13 (d, *J* = 6.5 Hz, 3H).; <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 196.2, 170.3, 141.0, 138.1, 137.2, 132.8, 130.2, 130.1, 128.4, 126.2, 43.3, 39.3, 14.3, 12.9.; **LC/MS**: Retention time 6.22 min, *m/z* (ESI+) = 282.19.; **HRMS** (ESI, *m/z*) calcd. for formula C<sub>18</sub>H<sub>20</sub>NO<sub>2</sub><sup>+</sup> [M+H]<sup>+</sup> = 282.1489; found = 282.1503.

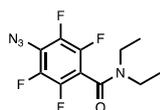
### *N,N*-diethyl-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamide (N2-CONEt<sub>2</sub>) (36)



To a solution of 4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoic acid **19** (1 eq., 300 mg, 1.3 mmol) and diethylamine (2 eq., 0.27 mL, 2.6 mmol) in DMF (3 mL) was added dropwise a solution of HATU (1.5 eq., 740 mg, 1.95 mmol) in DMF (1.5 mL) at 0 °C in 30 min. Then the solution was stirred at r.t. overnight. The mixture was poured into a sat. aq. NaHCO<sub>3</sub> solution and extracted with EtOAc. The organic phase was washed three times with water, dried over MgSO<sub>4</sub> and then concentrated to give the expected product as a white powder (161 mg, 43%).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 7.41 (d, *J* = 8.5 Hz, 2H), 7.23 (d, *J* = 8.1 Hz, 2H), 3.54 (d, *J* = 5.5 Hz, 2H), 3.21 (d, *J* = 5.5 Hz, 2H), 1.24 (s, 3H), 1.10 (s, 3H).; <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 170.1, 138.7, 130.1, 127.0, 126.7, 123.9, 120.3, 43.4, 39.5, 14.4, 13.0.; **LC/MS**: Retention time 6.60 min, *m/z* (ESI+) = 286.17.; **HRMS** (ESI, *m/z*) calcd. for formula C<sub>13</sub>H<sub>15</sub>F<sub>3</sub>N<sub>3</sub>O<sup>+</sup> [M+H]<sup>+</sup> = 286.1162; found = 286.1158.

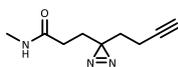
### 4-azido-*N,N*-diethyl-2,3,5,6-tetrafluorobenzamide (N3-CONEt<sub>2</sub>) (37)



To a solution of 4-azido-2,3,5,6-tetrafluorobenzoic acid **20** (1 eq., 600 mg, 2.55 mmol), diethylamine (1.5 eq., 0.4 mL, 3.88 mmol) and HOBt (1.1 eq., 379 mg, 2.81 mmol) in DMF (6.6 mL) was added EDCI·HCl (1.3 eq., 638 mg, 3.3 mmol) at 0 °C under argon atmosphere. The reaction mixture was allowed to warm to room temperature and was stirred overnight. The mixture was poured into a sat. aq. NaHCO<sub>3</sub> solution then extracted with EtOAc. The organic phase was washed three times with water, dried by MgSO<sub>4</sub> and then concentrated to give the expected product as a brown oil (520 mg, 70%).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 3.59 (q, *J* = 7.1 Hz, 2H), 3.23 (q, *J* = 7.1 Hz, 2H), 1.26 (t, *J* = 7.1 Hz, 3H), 1.14 (t, *J* = 7.2 Hz, 3H); <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 58.0, 144.3, 142.3, 142.1, 141.0, 139.0, 138.7, 43.3, 39.8, 14.1, 12.8.; LC/MS: Retention time 6.48 min, *m/z* (ESI+) = 291.05.; HRMS (ESI, *m/z*) calcd. for formula C<sub>11</sub>H<sub>11</sub>F<sub>4</sub>N<sub>4</sub>O<sup>+</sup> [M+H]<sup>+</sup> = 291.0864; found = 291.0875.

### 3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)-*N*-methylpropanamide (I-N<sub>2</sub>NHMe) (39)



To a solution of 3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)propanoic acid<sup>211</sup> (1.0 eq, 182 mg, 1.095 mmol) in dry ACN (22 mL) protected from the light, a commercially available 2M solution of CH<sub>3</sub>NH<sub>2</sub> in THF (1.8 eq, 0.99 mL, 1.97 mmol) and DIPEA (1.3 eq, 0.25 mL, 1.42 mmol) were added at r.t. under argon atmosphere. Then HATU (1.3 eq, 541 mg, 1.42 mmol) was added portion wise and the resulting reaction mixture was stirred for 3 h. Volatiles were removed under vacuum and the resulting mixture solubilised in EtOAc. The organic phase was washed 2 times with water, brine and dried over MgSO<sub>4</sub>. After filtration and evaporation, the crude product was purified by SiO<sub>2</sub> gel chromatography eluted with DCM/MeOH from 0 to 1 % leading to **39** as a pale beige solid (120 mg, 61 %).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ (ppm): 5.41 (s, 1H), 2.81 (d, *J* = 4.7 Hz, 3H), 2.06 – 1.97 (m, 2H), 1.96 – 1.81 (m, 4H), 1.65 (t, *J* = 7.2 Hz, 2H). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ (ppm): 172.1, 82.8, 69.3, 32.5, 30.3, 28.6, 28.0, 26.6, 13.4.

## Analytical experiments

### *Bph-COOH and Bph-CONEt<sub>2</sub> experiments*

1) Bph-COOH (3 mM) or 2) Bph-CONEt<sub>2</sub> (10 mM) in presence of thymidine (1) 3 mM or 2) 100 mM) in 1) water or 2) 1:1 water/ACN was flushed vigorously with argon for 1 min and capped to protect from oxygen. The flushed solution were irradiated in a Multirays photoreactor with four lamps (15 W) at 366 nm for 1) 1 h or 2) 38 h. The mixtures were analysed by HPLC.

### *-N<sub>2</sub>NHMe analytical experiments*

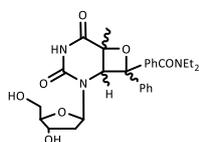
A mixture of *-N<sub>2</sub>NHMe* (10 mM) and 50 mM of one of the four nucleosides (Thymidine, 2'-deoxyguanosine, 2'-deoxycytosine, and 2'-deoxyadenosine) in 1:1 water/ACN was irradiated in a Multirays photoreactor with four lamps (15 W) at 366 nm for 30 min. The resulting mixture was analysed by HPLC injecting 20  $\mu$ L of irradiated solution without dilution.

## Preparative photosynthesis

### **Bph-CONEt<sub>2</sub> and thymidine photoreaction**

Bph-CONEt<sub>2</sub> (10 mM) was solubilised in a solution of thymidine (100 mM) in 1:1 water/ACN and was vigorously flushed with argon for 1 min before sealing with a cap. The solution was irradiated in a Multirays photoreactor with four lamps (15 W) at 366 nm for 38 h. The resulting mixture was purified by HPLC.

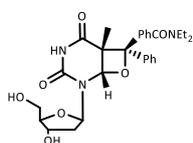
### ***N,N*-diethyl-4-(2-((2*R*,4*S*,5*R*)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-6-methyl-3,5-dioxo-8-phenyl-7-oxa-2,4-diazabicyclo[4.2.0]octan-8-yl)benzamide (#6 and #8)**



**#6** Unknown stereochemistry. <sup>1</sup>H NMR (700MHz, MeOD-d<sub>4</sub>)  $\delta$  (ppm): 7.51 (d,  $J$  = 8.4 Hz, 2H), 7.44 (dd,  $J$  = 10.6, 8.1 Hz, 4H), 7.38 (t,  $J$  = 7.5 Hz, 2H), 7.34 (dd,  $J$  = 8.3, 6.1 Hz, 1H), 6.21 (dd,  $J$  = 9.1, 5.9 Hz, 1H), 5.31 (s, 1H), 4.14 (d,  $J$  = 6.0 Hz, 1H), 3.79 (q,  $J$  = 3.4 Hz, 1H), 3.63 (dd,  $J$  = 11.6, 3.5 Hz, 1H), 3.60 (dd,  $J$  = 11.6, 3.5 Hz, 1H), 3.58 – 3.52 (m, 2H), 3.29 (m, 2H), 2.05 (ddd,  $J$  = 13.1, 5.9, 2.0 Hz, 1H), 1.82 (ddd,  $J$  = 13.1, 5.1, 1.1 Hz, 1H), 1.79 (s, 3H), 1.26 (t,  $J$  = 6.9 Hz, 3H), 1.16 (t,  $J$  = 6.8 Hz, 3H).; <sup>13</sup>C NMR (176 MHz, MeOD-d<sub>4</sub>)  $\delta$  (ppm): 173.2, 172.9, 153.5, 143.4, 140.6, 137.8, 129.8, 129.6, 128.4, 127.9, 127.1, 92.5, 87.6, 86.4, 78.4, 73.1, 63.7, 63.6, 45.2, 39.0, 22.3, 14.4, 13.1.

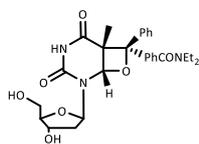
**#8** Unknown stereochemistry.  $^1\text{H NMR}$  (700MHz, MeOD- $d_4$ )  $\delta$  (ppm): 7.65 (dd,  $J = 7.7, 5.3$  Hz, 4H), 7.39 (t,  $J = 7.6$  Hz, 2H), 7.35 (d,  $J = 7.3$  Hz, 2H), 7.33 (t,  $J = 7.3$  Hz, 1H), 6.18 (dd,  $J = 8.3, 5.7$  Hz, 1H), 5.12 (s, 1H), 4.20 (dt,  $J = 4.9, 2.2$  Hz, 1H), 3.80 (td,  $J = 4.3, 2.7$  Hz, 1H), 3.54 (q,  $J = 6.8$  Hz, 2H), 3.28 (q,  $J = 7.2$  Hz, 2H), 3.15 (dd,  $J = 11.4, 4.1$  Hz, 1H), 3.10 (dd,  $J = 11.4, 4.1$  Hz, 1H), 2.10 (ddd,  $J = 13.3, 5.8, 2.2$  Hz, 1H), 2.03 (ddd,  $J = 14.3, 8.3, 5.8$  Hz, 1H), 1.78 (s, 3H), 1.24 (t,  $J = 6.8$  Hz, 3H), 1.12 (t,  $J = 6.7$  Hz, 3H).;  $^{13}\text{C NMR}$  (176 MHz, MeOD- $d_4$ )  $\delta$  (ppm): 173.3, 172.6, 152.4, 146.5, 139.6, 137.6, 129.2, 129.0, 128.0, 127.9, 127.0, 92.6, 88.6, 86.8, 78.3, 72.9, 63.0, 63.0, 45.0, 40.9, 40.0, 22.5, 14.4, 13.0.

***N,N*-diethyl-4-((1*S*,6*R*,8*S*)-2-((2*R*,4*S*,5*R*)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-6-methyl-3,5-dioxo-8-phenyl-7-oxa-2,4-diazabicyclo[4.2.0]octan-8-yl)benzamide (#1)**



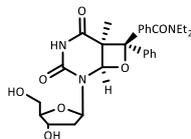
$^1\text{H NMR}$  (700MHz, MeOD- $d_4$ )  $\delta$  (ppm): 7.74 (d,  $J = 8.4$  Hz, 2H), 7.58 (d,  $J = 7.3$  Hz, 2H), 7.43 (d,  $J = 8.4$  Hz, 2H), 7.29 (t,  $J = 7.8$  Hz, 2H), 7.22 (t,  $J = 7.4$  Hz, 1H), 6.28 (dd,  $J = 7.6, 6.3$  Hz, 1H), 5.99 (s, 1H), 4.49 (dt,  $J = 6.6, 3.5$  Hz, 1H), 3.84 (q,  $J = 3.6$  Hz, 1H), 3.75 (dd,  $J = 12.0, 3.5$  Hz, 1H), 3.71 (dd,  $J = 12.0, 3.4$  Hz, 1H), 3.54 (q,  $J = 7.4$  Hz, 2H), 3.26 (q,  $J = 7.2$  Hz, 2H), 2.53 (ddd,  $J = 13.8, 7.6, 6.3$  Hz, 1H), 2.29 (ddd,  $J = 13.6, 6.2, 3.6$  Hz, 1H), 1.38 (s, 3H), 1.24 (t,  $J = 6.5$  Hz, 3H), 1.11 (t,  $J = 6.6$  Hz, 3H).;  $^{13}\text{C NMR}$  (176 MHz, MeOD- $d_4$ )  $\delta$  (ppm): 173.2, 171.8, 153.0, 144.2, 143.2, 137.3, 128.9, 128.7, 127.4, 127.3, 126.7, 90.0, 87.7, 85.2, 83.8, 72.2, 63.0, 55.5, 45.0, 40.9, 40.4, 19.5, 14.3, 13.0.

***N,N*-diethyl-4-((1*S*,6*R*,8*R*)-2-((2*R*,4*S*,5*R*)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-6-methyl-3,5-dioxo-8-phenyl-7-oxa-2,4-diazabicyclo[4.2.0]octan-8-yl)benzamide (#2)**



$^1\text{H NMR}$  (700MHz, MeOD- $d_4$ )  $\delta$  (ppm): 7.68 (d,  $J = 8.6$  Hz, 2H), 7.64 (d,  $J = 7.4$  Hz, 2H), 7.42 (t,  $J = 7.8$  Hz, 2H), 7.31 (dt,  $J = 6.8, 2.0$  Hz, 3H), 6.28 (dd,  $J = 7.6, 6.4$  Hz, 1H), 5.96 (s, 1H), 4.48 (dt,  $J = 6.6, 3.5$  Hz, 1H), 3.84 (q,  $J = 3.6$  Hz, 1H), 3.74 (dd,  $J = 12.0, 3.5$  Hz, 1H), 3.69 (dd,  $J = 11.9, 3.6$  Hz, 1H), 3.52 (q,  $J = 6.8$  Hz, 2H), 3.23 (q,  $J = 7.1$  Hz, 2H), 2.52 (ddd,  $J = 13.8, 7.6, 6.3$  Hz, 1H), 2.29 (ddd,  $J = 13.6, 6.1, 3.5$  Hz, 1H), 1.38 (s, 3H), 1.22 (t,  $J = 6.6$  Hz, 3H), 1.09 (t,  $J = 6.6$  Hz, 3H).;  $^{13}\text{C NMR}$  (176 MHz, MeOD- $d_4$ )  $\delta$  (ppm): 173.2, 171.9, 153.0, 145.2, 142.0, 137.1, 129.5, 128.9, 127.1, 126.9, 126.8, 89.9, 87.7, 85.1, 83.9, 72.2, 63.1, 55.7, 44.9, 40.8, 40.4, 19.5, 14.4, 13.0.

***N,N*-diethyl-4-((1*R*,6*S*,8*S*)-2-((2*R*,4*S*,5*R*)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-6-methyl-3,5-dioxo-8-phenyl-7-oxa-2,4-diazabicyclo[4.2.0]octan-8-yl)benzamide (#4)**

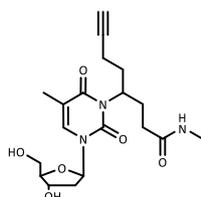


**<sup>1</sup>H NMR** (700MHz, MeOD-*d*<sub>4</sub>) δ (ppm): 7.71 (d, *J* = 8.5 Hz, 2H), 7.64 (d, *J* = 7.4 Hz, 2H), 7.42 (t, *J* = 7.9 Hz, 2H), 7.32 – 7.28 (m, 3H), 6.20 (t, *J* = 6.9 Hz, 1H), 5.73 (s, 1H), 4.39 (dt, *J* = 6.8, 3.7 Hz, 1H), 3.92 (q, *J* = 4.2 Hz, 1H), 3.80 (m, 1H), 3.78 (m, 1H), 3.52 (m, 2H), 3.23 (d, *J* = 6.8 Hz, 2H), 2.35 (quint, *J* = 6.7 Hz, 1H), 2.17 (m, 1H), 1.40 (s, 3H), 1.22 (t, *J* = 6.6 Hz, 3H), 1.09 (t, *J* = 6.6 Hz, 3H).; **<sup>13</sup>C NMR** (176 MHz, MeOD-*d*<sub>4</sub>) δ (ppm): 173.2, 171.8, 153.2, 145.1, 142.0, 137.1, 129.5, 128.9, 127.2, 126.9, 126.8, 89.8, 88.2, 86.8, 85.6, 72.6, 63.6, 56.1, 44.9, 40.8, 38.9, 19.2, 14.4, 13.0.

***l*-N2NHMe and thymidine photoreaction**

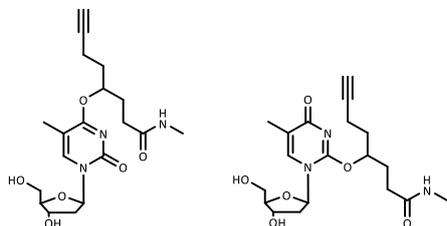
*l*-N2NHMe (10 mM) was solubilised in a solution of thymidine (100 mM) in water. The resulting solution was irradiated in a Multirays photoreactor with four lamps (15 W) at 366 nm for 30 min. The resulting mixture was purified by HPLC.

**4-(3-((2*R*,4*S*,5*R*)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-methyl-2,6-dioxo-3,6-dihydropyrimidin-1(2*H*)-yl)-*N*-methyloct-7-ynamide (#dT3)**



Diastereoisomers mixture. **<sup>1</sup>H NMR** (700 MHz, CD<sub>3</sub>OD-*d*<sub>4</sub>) δ (ppm): 7.82 (dd, *J* = 8.6, 1.1 Hz, 1H), 6.26 (t, *J* = 6.5 Hz, 1H), 5.09 – 4.94 (quint, *J* = 4.7 Hz, 1H), 4.38 (s, 1H), 3.91 (q, *J* = 3.4 Hz, 1H), 3.81 (dd, *J* = 12.0, 3.3 Hz, 1H), 3.74 (dd, *J* = 12.0, 3.4 Hz, 1H), 2.64 (d, *J* = 13.1 Hz, 3H), 2.51 – 2.49 (m, 1H), 2.48 – 2.40 (m, 1H), 2.30 – 2.25 (m, 1H), 2.22 – 2.06 (m, 6H), 2.03 – 1.96 (m, 1H), 1.88 (m, 4H).; **<sup>13</sup>C NMR** (176 MHz, CD<sub>3</sub>OD-*d*<sub>4</sub>) δ (ppm): 175.7, 175.6, 166.8, 166.1, 153.8, 151.7, 136.8, 136.7, 136.4, 136.2, 111.4, 111.3, 110.2, 110.1, 88.9, 88.8, 87.6, 87.4, 86.7, 84.0, 83.9, 72.0, 71.9, 70.0, 69.7, 69.6, 62.7, 62.5, 56.3, 56.2, 54.4, 41.4, 34.2, 34.0, 31.5, 31.3, 28.7, 28.4, 26.3, 16.8, 16.7, 13.5, 13.1.

**4-((1-((2*R*,4*S*,5*R*)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-methyl-2-oxo-1,2-dihydropyrimidin-4-yl)oxy)-*N*-methyloct-7-ynamide or 4-((1-((2*R*,4*S*,5*R*)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-methyl-4-oxo-1,4-dihydropyrimidin-2-yl)oxy)-*N*-methyloct-7-ynamide (#dT2)**



Diastereoisomers mixture. <sup>1</sup>H NMR (700 MHz, CD<sub>3</sub>OD-d<sub>4</sub>) δ (ppm): 8.08 (d, *J* = 1.1 Hz, 1H), 6.23 (t, *J* = 6.4 Hz, 1H), 5.44 (quint, *J* = 4.6 Hz, 1H), 4.40 (dt, *J* = 6.5, 3.7 Hz, 1H), 3.97 (q, *J* = 3.4 Hz, 1H), 3.84 (dd, *J* = 12.1, 3.4 Hz, 1H), 3.75 (dd, *J* = 12.2, 3.3 Hz, 1H), 2.68 (s, 3H), 2.37 – 2.25 (m, 6H), 2.24 (t, *J* = 2.6 Hz, 1H), 2.11 (m, 1H), 2.05 (m, 1H), 1.98 (m, 1H), 1.95 (m, 4H).; <sup>13</sup>C NMR (176 MHz, CD<sub>3</sub>OD-d<sub>4</sub>) δ (ppm): 175.6, 174.9, 156.5, 136.7, 117.6, 89.4, 87.9, 83.9, 79.5, 71.7, 70.3, 62.5, 42.3, 33.6, 32.6, 30.5, 26.3, 15.4, 13.6.

#### ***I*-N2NHMe and 2'-deoxyguanosine photoreaction**

*I*-N2NHMe (10 mM) was solubilised in a solution of 2'-deoxyguanosine (50 mM) in 1:1 water/ACN. The resulting solution was irradiated in a Multirays photoreactor with four lamps (15 W) at 366 nm for 30 min. The resulting mixture was purified by HPLC.

## **Biophysical experiments**

A series of oligonucleotides covering a range of possible G4 conformations was used for G4-FID and FRET-melting assays. Oligonucleotides were purchased from Eurogentec as dried samples purified by HPLC-RP. The oligonucleotides were solubilised in MiliQ water and stored at -20 °C. The exact concentrations were determined by measuring the absorbance at 260 nm via a Cary Eclipse 300 spectrophotometer.

### **Förster Resonance Energy Transfer (FRET)-melting assay**

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 for 16 h. Duplex DNA was heated at 95 °C for 10 min and left to cool down to room temperature over a period of at least 3 h. Buffer conditions are specified below for each G4 sequence. The stabilisation of ligands on G4 structure was monitored via FRET-melting assay performed in 96-well plates on the real-time PCR apparatus 7900HT Fast Real-Time PCR system. This assay was carried out with G4-forming sequences labelled with FRET partners at each end (5'-FAM (donor) and 3'-TAMRA (acceptor)) in the presence of G4 ligands. To determine ligand G4 versus duplex selectivity, an unlabelled DNA competitor (ds26) was used. Each well was filled with 0.20 µM (1 eq.) of labelled G4 DNA, 1 µM (5 eq.) of ligand, and dsDNA competitor at 0, 3 and 10 µM (0, 15 and 50 mol eq.) respectively, for a total volume of 25 µL. To determine ligand G4 versus RNA G4 selectivity, an unlabelled RNA competitor NRAS was used. Each well was filled with 0.20 µM (1 eq.) of labelled F21T, 1 µM (5 eq.) of ligand, and NRAS competitor at 0 and 3 µM (0 and 15 eq.) respectively, for a total volume of 25 µL. Solutions were prepared in buffer containing 10 mM Li.Caco pH 7.3, and i) 10 mM KCl completed with 90 mM LiCl for F21T and F21CTAT, and ii) 1 mM KCl, completed with 99 mM LiCl for all the others G-quadruplex sequences. The FAM channel was used to collect the fluorescence signal. Stabilisation of compounds with G4-structures was monitored as follows: 5 min at 25 °C and then increase of 0.5 °C every minute until reaching 95 °C. The stabilisation was evaluated by determining the temperature at half denaturation of the G4 (obtain at the half height of the sigmoidal curve,  $\Delta T_{1/2}$ ) in the absence and in the presence of ligand, duplex DNA, by using Origin pro 8.0 non-linear fitting method BiDoseResp.

#### *FRET-melting sequences:*

F21T: 5'-FAM-GGGTTAGGGTTAGGGTTAGGG-TAMRA-3'

FMYC22T: 5'-FAM-TTGAGGGTGGGTAGGGTGGGTAA-TAMRA-3'

F21CTAT: 5'-FAM-GGGCTAGGGCTAGGGCTAGGG-TAMRA-3'

FCEB25wtT: 5'-FAM-AAGGGTGGGTGTAAGTGTGGGTGGGT-TAMRA-3'

FVEGFT: 5'-FAM-CGGGGCGGGCCTTGGGCGGGGT-TAMRA-3'

FHIV-PRO1T: 5'-FAM-TGGCCTGGGCGGGACTGGG-TAMRA-3'

ds26: 5'-CAATCGGATCGAATTCGATCCGATTG-3'

NRAS: 5'-GGGAGGGGCGGGUCUGGG-3'

### **Fluorescent intercalator displacement assay (G4-FID)**

The experiment was carried out on pre-folded G-quadruplex structures: the sequence was heated at 95 °C for 10 min (in 10 mM Li.Caco pH 7.3 and 100 mM KCl) and left to cool down to room temperature over 3 h.

#### *In cuvette:*

G4-FID experiments were performed on a Cary Eclipse spectrofluorometer, using 1.5 mL fluorescence cuvette. 0.25  $\mu$ M of folded DNA was prepared in a solution of 10 mM Li.Caco pH 7.3 and 100 mM KCl mixed with 0.50  $\mu$ M Thiazole Orange (TO) to obtain a total volume of 1 mL. Each addition of ligand (from 0.5 to 10 eq.) was followed by a 3 min equilibration time and fluorescence emission spectra were recorded from 510 to 750 nm ( $\lambda_{exc} = 495$  nm, 1.0 nm increment, 0.1 s integration time and 10 nm excitation/emission slits). The percentage of TO displacement was calculated as follows:  $TOD(\%) = 100 - (F/F_0) \times 100$ , where  $F_0$  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. The  $DC_{50}$  value was obtained by fitting the curve using Origin pro 8.0 non-linear fitting method Hill.

#### *In 96-well plate:*

G4-FID experiment was carried in a 96-well Black Flat bottom Polystyrene NBS microplates from Corning and measurement was performed on a FLUOstar Omega microplate reader from BMG Labtech. Microplate was filled as follows: from 190 to 90  $\mu$ L of K+100 buffer (to reach a final volume of 200  $\mu$ L), 10  $\mu$ L of a prefolded DNA solution (5  $\mu$ M) and TO (10  $\mu$ M for G4-DNA and 15  $\mu$ M for ds26), from 0 to 100  $\mu$ L of ligand (5  $\mu$ 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  $\mu$ M. Finally, the microplate was gently shaken for 5 min at 500 rpm and emission of TO fluorescence in each well was measured at 520 nm after excitation at 485 nm. TO displacement was plotted and  $DC_{50}$  obtained as previously mentioned.

*G4-FID sequences:*

22AG: 5'-A-GGGTTAGGGTTAGGGTTAGGG-3'

22CTA: 5'-A-GGGCTAGGGCTAGGGCTAGGG-3'

cMYC22: 5'-TTGAGGGTGGGTAGGGTGGG TAA-3'

CEB25wt: 5'-AAGGGTGGGTGTAAGTGTGGGTGGG T-3'

VEGF: 5'-CGGGGCGGGCCTTGGGCGGGGT-3'

### **Fluorimetric titration**

Each titration was performed in a 1.5 ml fluorescence cuvette, in 10 mM lithium cacodylate buffer pH 7.3, 100 mM KCl, in a total volume of 1 ml. Titrations are performed with the addition of TO (0.5  $\mu\text{M}$ ) followed by the successive addition of oligonucleotide (up to 2.5  $\mu\text{M}$ ), and fluorescence emission spectra were recorded from 510 to 750 nm ( $\lambda_{\text{exc}} = 495$  nm, 1.0 nm increment, 0.1 s integration time and 10 nm excitation/emission slits).

### **Continuous variation method (Job plot)**

A series of solutions with varying mole fractions of TO and VEGF were used for the experiments: the sum of TO and VEGF concentrations was always 1  $\mu\text{M}$  in a total volume of 1 mL. Each solution fluorescence emission spectra were recorded from 510 to 750 nm ( $\lambda_{\text{exc}} = 495$  nm, 1.0 nm increment, 0.1 s integration time and 10 nm excitation/emission slits). The Job plot was drawn as the fluorescence values versus the ligand mole fraction.

## **Biochemistry experiments**

### **Sample preparation**

Oligonucleotide sequences were purchased from Eurogentec as dried samples purified by PAGE, dissolved in MilliQ water, and stored at -20 °C. [ $\gamma$ - $^{32}\text{P}$ ] ATP was purchased from Perkin Elmer as a 25  $\mu\text{L}$  or 50  $\mu\text{L}$  solution for a total radioactivity of 9.25 MBq or 18.5 MBq respectively. Oligonucleotide sequences were 5'-radiolabeled as followed: 1  $\mu\text{L}$  of 20  $\mu\text{M}$  PAGE-purified oligonucleotide was mixed with 2  $\mu\text{L}$  MilliQ water, 1  $\mu\text{L}$  T4-PNK buffer 10x, 3  $\mu\text{L}$  [ $\gamma$ - $^{32}\text{P}$ ] ATP and 3  $\mu\text{L}$  of T4-PNK enzyme. The reaction mixture was incubated at 37 °C for 40 min. The radiolabelled 22AG and cMYC22 sequences were purified by 20% denaturing gel electrophoresis. The radiolabelled CEB25wt sequence was diluted with MilliQ water, purified through a G25 Sephadex column and used without further purification.

### Photoalkylation experiments

Photochemical experiments were carried out in 1.5 mL Eppendorf tubes containing a prefolded solution of 10  $\mu\text{M}$  DNA and 1.0 to 2.5  $\mu\text{L}$  of 5'-end radiolabeled sequence depending on the amount of radioactivity (Total counts for sample 50,000) in K+100 buffer for a total volume of 18  $\mu\text{L}$ . Finally, 2  $\mu\text{L}$  of ligand at 2 mM in DMSO was added and allowed to equilibrate in the dark for 30 min. After equilibration time, samples were irradiated using 10 lamps at 360 nm of a Multirays photoreactor from Helios Quartz. After irradiation, 5  $\mu\text{L}$  of loading solution (bromophenol blue and xylene cyanol in formamide/water 8/2) was added. Samples were concentrated 15 min at 30  $^{\circ}\text{C}$  using a SpeedVac, loaded on a 15% denaturing gel electrophoresis and gels were run for 5 to 6 h at 20 W. After that time, a europium poisoned screen was exposed to the gel for a time determined by the total radioactivity. The screen was finally analysed using a Typhoon apparatus and images were treated using ImageQuant TL v8.0. The percentage of alkylation was determined by calculating the ratio of the total counts of alkylated bands divided by the total counts of the alkylated bands + the non-alkylated band corresponding to the starting material.

### Chemical sequencing experiments

After their separation by gel electrophoresis, the different products of alkylation were eluted from the gel using an aqueous solution of NaCl at 0.15 M, and precipitated with cold ethanol. The non-alkylated and alkylated oligonucleotides were treated with dimethylsulfate (DMS)/piperidine (probe of the free N7 of the guanine base) in Maxam-Gilbert sequencing conditions: the oligonucleotides were dissolved in 19  $\mu\text{L}$  of water and incubated with 1  $\mu\text{L}$  of DMS for 1 min 30 sec at 37  $^{\circ}\text{C}$ . The alkylation products and the non-alkylated oligonucleotides, after DMS/piperidine treatments and precipitation, were incubated with 50  $\mu\text{L}$  of 1 M piperidine aqueous solution at 90  $^{\circ}\text{C}$  for 25 min. After evaporation of the piperidine, the samples were washed four times with 100  $\mu\text{L}$  of water before being solubilised in 4  $\mu\text{L}$  of water and 3  $\mu\text{L}$  of loading dye solution. Samples were loaded on a 20% denaturing gel and ran for 3 h at 20 W.

### Enzymatic sequencing

3'-exonuclease is usually employed to determine the alkylation sites of oligonucleotides since its digestion stops at alkylation sites. Non-alkylated and alkylated oligonucleotides, isolated by gel electrophoresis, were solubilised in 3  $\mu\text{L}$  of water, 1  $\mu\text{L}$  of buffer composed by Tris-HCl buffer (pH 8.0) at 10 mM, 2 mM  $\text{MgCl}_2$  and 0.5 mg/mL tRNA, and treated with 1  $\mu\text{L}$  of the non-diluted or diluted 3'-exonuclease phosphodiesterase I from *Crotalus adamanteus* venom (UBS) at 0.04 U/ $\mu\text{L}$  diluted 10, 20, 30, 40 times for 30 min at 37  $^{\circ}\text{C}$  for the alkylated samples. The partial digestion of nonalkylated oligonucleotide was run with 0.001 U/ $\mu\text{L}$  for 20 min at 37  $^{\circ}\text{C}$ . As alkylation could occur at several positions on the same oligonucleotide, the alkylated

products were treated with an increasing concentration of enzyme which enables to bypass the first alkylated base encountered at the 3'-end. The treatment was stopped by putting the sample into ice and by adding loading dye. The samples were conserved at -20 °C. 3 µL of loading solution was added before purifying the digested fragment on a 20% denaturing gel electrophoresis and gels were run for 3 h at 20 W. The migration of the fragments depends on the presence of the alkylating agents, and the branched benzophenone compounds can be dealkylated by heating, therefore, each of them was eluted from the gel, precipitated and solubilised in 90 µL of water to be dealkylated by treatment at 90 °C for 30 min. Samples were concentrated 40 min at 30 °C using a SpeedVac before adding the loading solution. The dealkylated fragments were loaded on 20% denaturing gel and migrated for 3 h at 20 W. Their migrations were compared with the fragments obtained by partial digestion of the starting oligonucleotide. From their migration, it is possible to determine their length and consequently the alkylated base where the 3'-exonuclease stopped its digestion.

## **Photochemical experiments**

### **Laser Flash Photolysis (LFP) Experiments**

The experiments were performed on an LP 920 Edinburgh Instrument photolysis to carry out the detection of the transient species. The minimum response time of the detection system was about 10 ns. The laser beam (Nd/YAG laser operated using the third harmonic  $\lambda = 355$  nm) was focused on a 3 mm wide circular area of the cell and the first 5 mm in depth were analysed at a right-angle geometry. The incident pulse energies used were 5 - 8 mJ per pulse. Solutions for analysis were placed in a 10 mm fluorescence cuvette in which the sample absorbance at 355 nm was comprised between 0.4-0.6 over 1 cm. When required, oxygen was removed from the solutions by vigorously bubbling with argon for 1 min before sealing the cuvette with its cap. Acquisition and processing of the absorption signals were performed by LP900 7.0.2 (Build 0) software. Nanosecond laser flash photolysis spectra were reconstructed point from time profiles taken each 5 nm.

### **Quenching rate constant experiments**

Quenching rate constants were obtained for pyridine nucleosides with Bph-COOH (3.5 mM) and Bph-CONHMe (4 mM) in water by recording the decay spectra of the triplet absorption maximum (540 nm) with successive addition of the nucleoside (from 0.2 to 1.4 mM for all except for the deoxyuridine from 0.2 to 1.0 mM) followed by oxygen removal. The quenching rate constant corresponds to a second order rate constant determined by using the pseudo-first order approximation where  $k_{obs} = \ln(2)/t_{1/2}$  and  $t_{1/2}$  obtained using the second order non-linear fitting procedure on the decay curves. The slope of  $k_{obs}$  versus the concentration of nucleoside corresponds to the quenching rate constant.

### **GC-MS Photolysis**

The in-line *l*-N<sub>2</sub>COOH precursor and *l*-N<sub>2</sub>NHMe prototype at 10 mM were solubilised in methanol (both) or in water (for *l*-N<sub>2</sub>NHMe). The solutions were irradiated in a Multirays photoreactor with four lamps (15 W) at 366 nm for between 10 to 30 min. The resulting products were analysed using ISQTM 7610 Single Quadrupole GC-MS. The GC parameters were as follows: starting temperature 60 °C per 4 min, then a rate of 8.75°C/min, final temperature 200 °C.



## Bibliography



- (1) Miescher, F. Ueber Die Chemische Zusammensetzung Der Eiter- Zellen. *Med.-Chem. Unters.* **4**, 441–460.
- (2) Dahm, R. Friedrich Miescher and the Discovery of DNA. *Developmental Biology* **2005**, *278* (2), 274–288. <https://doi.org/10.1016/j.ydbio.2004.11.028>.
- (3) Lamm, E.; Harman, O.; Veigl, S. J. Before Watson and Crick in 1953 Came Friedrich Miescher in 1869. *Genetics* **2020**, *215* (2), 291–296. <https://doi.org/10.1534/genetics.120.303195>.
- (4) Kossel, A. Zur Chemie Des Zellkerns. *Hoppe-Seyl. z. physiol. Chem.* **1883**, *7*, 7–22. <https://doi.org/10.1515/bchm1.1883.7.1.7>.
- (5) Kossel, A. Zur Chemie Des Zellkerns. *Hoppe-Seyl. z. physiol. Chem.* **1886**, *10*, 248–264.
- (6) Kossel, A. Zur Chemie Des Zellkerns. *Hoppe-Seyl. z. physiol. Chem.* **1896**, *22*, 176–187.
- (7) Chargaff, E.; Lipshitz, R.; Green, C. Composition of the Desoxypentose Nucleic Acids of Four Genera of Sea-Urchin. *J Biol Chem* **1952**, *195* (1), 155–160.
- (8) Elson, D.; Chargaff, E. On the Desoxyribonucleic Acid Content of Sea Urchin Gametes. *Experientia* **1952**, *8* (4), 143–145. <https://doi.org/10.1007/BF02170221>.
- (9) Urry, L. A.; Cain, M. L.; Wasserman, S. A.; Minorsky, P. V.; Reece, J. B.; Campbell, N. A. *Campbell Biology*, Eleventh edition.; Pearson Education, Inc: New York, NY, 2017.
- (10) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA Dodecamer: Conformation and Dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78* (4), 2179–2183. <https://doi.org/10.1073/pnas.78.4.2179>.
- (11) Watson, J. D.; Crick, F. H. Molecular Structure of Nucleic Acids; a Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737–738. <https://doi.org/10.1038/171737a0>.
- (12) Pauling, L.; Corey, R. B. A Proposed Structure For The Nucleic Acids. *Proc Natl Acad Sci U S A* **1952**, *39* (2), 84–97. <https://doi.org/10.1073/pnas.39.2.84>.
- (13) Cobb, M.; Comfort, N. What Rosalind Franklin Truly Contributed to the Discovery of DNA's Structure. *Nature* **2023**, *616* (7958), 657–660. <https://doi.org/10.1038/d41586-023-01313-5>.
- (14) Ussery, D. W. DNA Structure: A-, B- and Z-DNA Helix Families. In *eLS*; John Wiley & Sons, Ltd, Ed.; Wiley, 2002. <https://doi.org/10.1038/npg.els.0003122>.
- (15) Griffith, F. The Significance of Pneumococcal Types. *J. Hyg.* **1928**, *27* (2), 113–159. <https://doi.org/10.1017/S0022172400031879>.
- (16) Downie, A. W. Pneumococcal Transformation-A Backward View Fourth Griffith Memorial Lecture. *Journal of General Microbiology* **1972**, *73* (1), 1–11. <https://doi.org/10.1099/00221287-73-1-1>.
- (17) Avery, O. T.; Macleod, C. M.; McCarty, M. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *J Exp Med* **1944**, *79* (2), 137–158. <https://doi.org/10.1084/jem.79.2.137>.
- (18) Hershey, A. D.; Chase, M. INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. *Journal of General Physiology* **1952**, *36* (1), 39–56. <https://doi.org/10.1085/jgp.36.1.39>.
- (19) Svartman, M.; Stone, G.; Stanyon, R. Molecular Cytogenetics Discards Polyploidy in Mammals. *Genomics* **2005**, *85* (4), 425–430. <https://doi.org/10.1016/j.ygeno.2004.12.004>.
- (20) *Molecular Biology of the Cell. Hauptbd.*, 4. ed.; Alberts, B., Ed.; Garland: New York, 2002.
- (21) Yang, E.; van Nimwegen, E.; Zavolan, M.; Rajewsky, N.; Schroeder, M.; Magnasco, M.; Darnell, J. E. Decay Rates of Human mRNAs: Correlation With Functional Characteristics and Sequence Attributes. *Genome Res* **2003**, *13* (8), 1863–1872. <https://doi.org/10.1101/gr.1272403>.
- (22) Ross, J. The Turnover of Messenger RNA. *Scientific American* **1989**, *260* (4), 48–55.
- (23) Yoon, J.-H.; Kim, J.; Gorospe, M. Long Noncoding RNA Turnover. *Biochimie* **2015**, *117*, 15–21. <https://doi.org/10.1016/j.biochi.2015.03.001>.
- (24) Li, Y.; Breaker, R. R. Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2'-Hydroxyl Group. *J. Am. Chem. Soc.* **1999**, *121* (23), 5364–5372. <https://doi.org/10.1021/ja990592p>.
- (25) *RNA Worlds: From Life's Origins to Diversity in Gene Regulation*; Atkins, J. F., Gesteland,

- R. F., Cech, T., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y, 2011.
- (26) Heinemann, U.; Roske, Y. Symmetry in Nucleic-Acid Double Helices. *Symmetry* **2020**, *12* (5), 737. <https://doi.org/10.3390/sym12050737>.
- (27) Lima, S.; Hildenbrand, J.; Korostelev, A.; Hattman, S.; Li, H. Crystal Structure of an RNA Helix Recognized by a Zinc-Finger Protein: An 18-Bp Duplex at 1.6 Å Resolution. *RNA* **2002**, *8* (7), 924–932. <https://doi.org/10.1017/S1355838202028893>.
- (28) Davis, R. R.; Shaban, N. M.; Perrino, F. W.; Hollis, T. Crystal Structure of RNA-DNA Duplex Provides Insight into Conformational Changes Induced by RNase H Binding. *Cell Cycle* **2015**, *14* (4), 668–673. <https://doi.org/10.4161/15384101.2014.994996>.
- (29) Zenkin, N.; Naryshkina, T.; Kuznedelov, K.; Severinov, K. The Mechanism of DNA Replication Primer Synthesis by RNA Polymerase. *Nature* **2006**, *439* (7076), 617–620. <https://doi.org/10.1038/nature04337>.
- (30) Krahn, N.; Fischer, J. T.; Söll, D. Naturally Occurring tRNAs With Non-Canonical Structures. *Front. Microbiol.* **2020**, *11*, 596914. <https://doi.org/10.3389/fmicb.2020.596914>.
- (31) Chan, C. W.; Badong, D.; Rajan, R.; Mondragón, A. Crystal Structures of an Unmodified Bacterial tRNA Reveal Intrinsic Structural Flexibility and Plasticity as General Properties of Unbound tRNAs. *RNA* **2020**, *26* (3), 278–289. <https://doi.org/10.1261/rna.073478.119>.
- (32) Venkadesh, S.; Mandal, P. K.; Gautham, N. The Structure of a Full Turn of an A-DNA Duplex d(CGCGGGTACCCGCG)<sub>2</sub>. *Biochemical and Biophysical Research Communications* **2011**, *407* (2), 307–312. <https://doi.org/10.1016/j.bbrc.2011.03.007>.
- (33) Franklin, R. E.; Gosling, R. G. The Structure of Sodium Thymonucleate Fibres. I. The Influence of Water Content. *Acta Cryst* **1953**, *6* (8), 673–677. <https://doi.org/10.1107/S0365110X53001939>.
- (34) Harteis, S.; Schneider, S. Making the Bend: DNA Tertiary Structure and Protein-DNA Interactions. *IJMS* **2014**, *15* (7), 12335–12363. <https://doi.org/10.3390/ijms150712335>.
- (35) Luo, Z.; Dauter, M.; Dauter, Z. Phosphates in the Z-DNA Dodecamer Are Flexible, but Their P-SAD Signal Is Sufficient for Structure Solution. *Acta Crystallogr D Biol Crystallogr* **2014**, *70* (7), 1790–1800. <https://doi.org/10.1107/S1399004714004684>.
- (36) Herbert, A. Z-DNA and Z-RNA in Human Disease. *Commun Biol* **2019**, *2* (1), 7. <https://doi.org/10.1038/s42003-018-0237-x>.
- (37) Weisenseel, J. P.; Reddy, G. R.; Marnett, L. J.; Stone, M. P. Structure of the 1, N<sup>2</sup>-Propanodeoxyguanosine Adduct in a Three-Base DNA Hairpin Loop Derived from a Palindrome in the *Salmonella Typhimurium hisD3052* Gene. *Chem. Res. Toxicol.* **2002**, *15* (2), 140–152. <https://doi.org/10.1021/tx010107f>.
- (38) Svoboda, P.; Cara, A. Di. Hairpin RNA: A Secondary Structure of Primary Importance. *Cell. Mol. Life Sci.* **2006**, *63* (7–8), 901–908. <https://doi.org/10.1007/s00018-005-5558-5>.
- (39) Rhee, S.; Han, Z.; Liu, K.; Miles, H. T.; Davies, D. R. Structure of a Triple Helical DNA with a Triplex–Duplex Junction. *Biochemistry* **1999**, *38* (51), 16810–16815. <https://doi.org/10.1021/bi991811m>.
- (40) Jain, A.; Wang, G.; Vasquez, K. M. DNA Triple Helices: Biological Consequences and Therapeutic Potential. *Biochimie* **2008**, *90* (8), 1117–1130. <https://doi.org/10.1016/j.biochi.2008.02.011>.
- (41) Altona, C.; Pikkemaat, J. A.; Overmars, F. J. Three-Way and Four-Way Junctions in DNA: A Conformational Viewpoint. *Current Opinion in Structural Biology* **1996**, *6* (3), 305–316. [https://doi.org/10.1016/S0959-440X\(96\)80048-0](https://doi.org/10.1016/S0959-440X(96)80048-0).
- (42) Thiviyathan, V.; Luxon, B. A.; Leontis, N. B.; Illangasekare, N.; Donne, D. G.; Gorenstein, D. G. Hybrid-Hybrid Matrix Structural Refinement of a DNA Three-Way Junction from 3D NOESY-NOESY. *Journal of Biomolecular NMR* **1999**, *14* (3), 209–221. <https://doi.org/10.1023/A:1008330011425>.
- (43) Nuthanakanti, A.; Ahmed, I.; Khatik, S. Y.; Saikrishnan, K.; Srivatsan, S. G. Probing G-Quadruplex Topologies and Recognition Concurrently in Real Time and 3D Using a Dual-App Nucleoside Probe. *Nucleic Acids Research* **2019**, *47* (12), 6059–6072. <https://doi.org/10.1093/nar/gkz419>.

- (44) Gehring, K.; Leroy, J.-L.; Guéron, M. A Tetrameric DNA Structure with Protonated Cytosine-Cytosine Base Pairs. *Nature* **1993**, *363* (6429), 561–565. <https://doi.org/10.1038/363561a0>.
- (45) Zhang, Y.; El Omari, K.; Duman, R.; Liu, S.; Haider, S.; Wagner, A.; Parkinson, G. N.; Wei, D. Native de Novo Structural Determinations of Non-Canonical Nucleic Acid Motifs by X-Ray Crystallography at Long Wavelengths. *Nucleic Acids Research* **2020**, *48* (17), 9886–9898. <https://doi.org/10.1093/nar/gkaa439>.
- (46) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-Motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nature Chem* **2018**, *10* (6), 631–637. <https://doi.org/10.1038/s41557-018-0046-3>.
- (47) Abou Assi, H.; Garavís, M.; González, C.; Damha, M. J. I-Motif DNA: Structural Features and Significance to Cell Biology. *Nucleic Acids Research* **2018**, *46* (16), 8038–8056. <https://doi.org/10.1093/nar/gky735>.
- (48) Rhodes, D.; Lipps, H. J. G-Quadruplexes and Their Regulatory Roles in Biology. *Nucleic Acids Res* **2015**, *43* (18), 8627–8637. <https://doi.org/10.1093/nar/gkv862>.
- (49) Largy, E.; Mergny, J.-L.; Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. In *The Alkali Metal Ions: Their Role for Life*; Sigel, A., Sigel, H., Sigel, R. K. O., Eds.; Metal Ions in Life Sciences; Springer International Publishing: Cham, 2016; Vol. 16, pp 203–258. [https://doi.org/10.1007/978-3-319-21756-7\\_7](https://doi.org/10.1007/978-3-319-21756-7_7).
- (50) Bang, I. Untersuchungen Über Die Guanylsäure. *Biochem. Z* **1910**, *26*, 293–311.
- (51) Gellert, M.; Lipsett, M. N.; Davies, D. R. Helix Formation by Guanylic Acid. *Proc Natl Acad Sci U S A* **1962**, *48* (12), 2013–2018. <https://doi.org/10.1073/pnas.48.12.2013>.
- (52) Zimmerman, S. B.; Cohen, G. H.; Davies, D. R. X-Ray Fiber Diffraction and Model-Building Study of Polyguanylic Acid and Polyinosinic Acid. *Journal of Molecular Biology* **1975**, *92* (2), 181–192. [https://doi.org/10.1016/0022-2836\(75\)90222-3](https://doi.org/10.1016/0022-2836(75)90222-3).
- (53) Arnott, S.; Chandrasekaran, R.; Marttila, C. M. Structures for Polyinosinic Acid and Polyguanylic Acid. *Biochemical Journal* **1974**, *141* (2), 537–543. <https://doi.org/10.1042/bj1410537>.
- (54) Sundquist, W. I.; Klug, A. Telomeric DNA Dimerizes by Formation of Guanine Tetrads between Hairpin Loops. *Nature* **1989**, *342* (6251), 825–829. <https://doi.org/10.1038/342825a0>.
- (55) Nishio, M.; Tsukakoshi, K.; Ikebukuro, K. G-Quadruplex: Flexible Conformational Changes by Cations, pH, Crowding and Its Applications to Biosensing. *Biosensors and Bioelectronics* **2021**, *178*, 113030. <https://doi.org/10.1016/j.bios.2021.113030>.
- (56) Simonsson, T. G-Quadruplex DNA Structures Variations on a Theme. *Biological Chemistry* **2001**, *382* (4). <https://doi.org/10.1515/BC.2001.073>.
- (57) Webba da Silva, M. Geometric Formalism for DNA Quadruplex Folding. *Chem. Eur. J.* **2007**, *13* (35), 9738–9745. <https://doi.org/10.1002/chem.200701255>.
- (58) Karsisiotis, A. I.; O’Kane, C.; Webba Da Silva, M. DNA Quadruplex Folding Formalism – A Tutorial on Quadruplex Topologies. *Methods* **2013**, *64* (1), 28–35. <https://doi.org/10.1016/j.ymeth.2013.06.004>.
- (59) Piazza, A.; Adrian, M.; Samazan, F.; Heddi, B.; Hamon, F.; Serero, A.; Lopes, J.; Teulade-Fichou, M.; Phan, A. T.; Nicolas, A. Short Loop Length and High Thermal Stability Determine Genomic Instability Induced by G-quadruplex-forming Minisatellites. *EMBO J* **2015**, *34* (12), 1718–1734. <https://doi.org/10.15252/embj.201490702>.
- (60) Ma, Y.; Iida, K.; Nagasawa, K. Topologies of G-Quadruplex: Biological Functions and Regulation by Ligands. *Biochemical and Biophysical Research Communications* **2020**, *531* (1), 3–17. <https://doi.org/10.1016/j.bbrc.2019.12.103>.
- (61) Zacchia, M.; Abategiovanni, M. L.; Stratigis, S.; Capasso, G. Potassium: From Physiology to Clinical Implications. *Kidney Dis (Basel)* **2016**, *2* (2), 72–79. <https://doi.org/10.1159/000446268>.
- (62) Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **2016**, *4*. <https://doi.org/10.3389/fchem.2016.00038>.
- (63) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. Human Telomeric Sequence Forms a Hybrid-Type Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel Strands in Potassium Solution. *Nucleic Acids Research* **2006**, *34* (9), 2723–2735. <https://doi.org/10.1093/nar/gkl348>.

- (64) Wang, Y.; Patel, D. J. Solution Structure of the Human Telomeric Repeat d[AG3(T2AG3)3] G-Tetraplex. *Structure* **1993**, *1* (4), 263–282. [https://doi.org/10.1016/0969-2126\(93\)90015-9](https://doi.org/10.1016/0969-2126(93)90015-9).
- (65) Kettani, A.; Bouaziz, S.; Gorin, A.; Zhao, H.; Jones, R. A.; Patel, D. J. Solution Structure of a Na Cation Stabilized DNA Quadruplex Containing G·G·G·G and G·C·G·C Tetrads Formed by G-G-G-C Repeats Observed in Adeno-Associated Viral DNA 1 Edited by I. Tinoco. *Journal of Molecular Biology* **1998**, *282* (3), 619–636. <https://doi.org/10.1006/jmbi.1998.2030>.
- (66) Zhang, N.; Gorin, A.; Majumdar, A.; Kettani, A.; Chernichenko, N.; Skripkin, E.; Patel, D. J. Dimeric DNA Quadruplex Containing Major Groove-Aligned A-T-A-T and G-C-G-C Tetrads Stabilized by Inter-Subunit Watson-Crick A-T and G-C Pairs. *J Mol Biol* **2001**, *312* (5), 1073–1088. <https://doi.org/10.1006/jmbi.2001.5002>.
- (67) Lim, K. W.; Alberti, P.; Guédin, A.; Lacroix, L.; Riou, J.-F.; Royle, N. J.; Mergny, J.-L.; Phan, A. T. Sequence Variant (CTAGGG)<sub>n</sub> in the Human Telomere Favors a G-Quadruplex Structure Containing a G·C·G·C Tetrad. *Nucleic Acids Research* **2009**, *37* (18), 6239–6248. <https://doi.org/10.1093/nar/gkp630>.
- (68) Amrane, S.; Kerkour, A.; Bedrat, A.; Vialet, B.; Andreola, M.-L.; Mergny, J.-L. Topology of a DNA G-Quadruplex Structure Formed in the HIV-1 Promoter: A Potential Target for Anti-HIV Drug Development. *J. Am. Chem. Soc.* **2014**, *136* (14), 5249–5252. <https://doi.org/10.1021/ja501500c>.
- (69) Butovskaya, E.; Heddi, B.; Bakalar, B.; Richter, S. N.; Phan, A. T. Major G-Quadruplex Form of HIV-1 LTR Reveals a (3 + 1) Folding Topology Containing a Stem-Loop. *J. Am. Chem. Soc.* **2018**, *140* (42), 13654–13662. <https://doi.org/10.1021/jacs.8b05332>.
- (70) Bishop, G. R.; Chaires, J. B. Characterization of DNA Structures by Circular Dichroism. *Current Protocols in Nucleic Acid Chemistry* **2002**, *11* (1), 7.11.1-7.11.8. <https://doi.org/10.1002/0471142700.nc0711s11>.
- (71) Carvalho, J.; Queiroz, J. A.; Cruz, C. Circular Dichroism of G-Quadruplex: A Laboratory Experiment for the Study of Topology and Ligand Binding. *J. Chem. Educ.* **2017**, *94* (10), 1547–1551. <https://doi.org/10.1021/acs.jchemed.7b00160>.
- (72) Paramasivan, S.; Rujan, I.; Bolton, P. H. Circular Dichroism of Quadruplex DNAs: Applications to Structure, Cation Effects and Ligand Binding. *Methods* **2007**, *43* (4), 324–331. <https://doi.org/10.1016/j.ymeth.2007.02.009>.
- (73) Maxam, A. M.; Gilbert, W. A New Method for Sequencing DNA. *Proceedings of the National Academy of Sciences* **1977**, *74* (2), 560–564. <https://doi.org/10.1073/pnas.74.2.560>.
- (74) Maxam, A. M.; Gilbert, W. [57] Sequencing End-Labeled DNA with Base-Specific Chemical Cleavages. In *Methods in Enzymology*; Nucleic Acids Part I; Academic Press, 1980; Vol. 65, pp 499–560. [https://doi.org/10.1016/S0076-6879\(80\)65059-9](https://doi.org/10.1016/S0076-6879(80)65059-9).
- (75) Onel, B.; Wu, G.; Sun, D.; Lin, C.; Yang, D. Electrophoretic Mobility Shift Assay and Dimethyl Sulfate Footprinting for Characterization of G-Quadruplexes and G-Quadruplex-Protein Complexes. In *G-Quadruplex Nucleic Acids: Methods and Protocols*; Yang, D., Lin, C., Eds.; Methods in Molecular Biology; Springer: New York, NY, 2019; pp 201–222. [https://doi.org/10.1007/978-1-4939-9666-7\\_11](https://doi.org/10.1007/978-1-4939-9666-7_11).
- (76) Neidle, S.; Parkinson, G. N. Quadruplex DNA Crystal Structures and Drug Design. *Biochimie* **2008**, *90* (8), 1184–1196. <https://doi.org/10.1016/j.biochi.2008.03.003>.
- (77) Adrian, M.; Heddi, B.; Phan, A. T. NMR Spectroscopy of G-Quadruplexes. *Methods* **2012**, *57* (1), 11–24. <https://doi.org/10.1016/j.ymeth.2012.05.003>.
- (78) Salgado, G. F.; Cazenave, C.; Kerkour, A.; Mergny, J.-L. G-Quadruplex DNA and Ligand Interaction in Living Cells Using NMR Spectroscopy. *Chem. Sci.* **2015**, *6* (6), 3314–3320. <https://doi.org/10.1039/C4SC03853C>.
- (79) Huppert, J. L. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Research* **2005**, *33* (9), 2908–2916. <https://doi.org/10.1093/nar/gki609>.
- (80) Todd, A. K. Highly Prevalent Putative Quadruplex Sequence Motifs in Human DNA. *Nucleic Acids Research* **2005**, *33* (9), 2901–2907. <https://doi.org/10.1093/nar/gki553>.
- (81) Eddy, J.; Maizels, N. Gene Function Correlates with Potential for G4 DNA Formation in the Human Genome. *Nucleic Acids Research* **2006**, *34* (14), 3887–3896.

<https://doi.org/10.1093/nar/gkl529>.

(82) Garant, J.-M.; Perreault, J.-P.; Scott, M. S. Motif Independent Identification of Potential RNA G-Quadruplexes by G4RNA Screener. *Bioinformatics* **2017**, *33* (22), 3532–3537. <https://doi.org/10.1093/bioinformatics/btx498>.

(83) Sahakyan, A. B.; Chambers, V. S.; Marsico, G.; Santner, T.; Di Antonio, M.; Balasubramanian, S. Machine Learning Model for Sequence-Driven DNA G-Quadruplex Formation. *Sci Rep* **2017**, *7* (1), 14535. <https://doi.org/10.1038/s41598-017-14017-4>.

(84) Willyard, C. New Human Gene Tally Reignites Debate. *Nature* **2018**, *558* (7710), 354–355. <https://doi.org/10.1038/d41586-018-05462-w>.

(85) Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A. V.; Mikheenko, A.; Vollger, M. R.; Altemose, N.; Uralsky, L.; Gershman, A.; Aganezov, S.; Hoyt, S. J.; Diekhans, M.; Logsdon, G. A.; Alonge, M.; Antonarakis, S. E.; Borchers, M.; Bouffard, G. G.; Brooks, S. Y.; Caldas, G. V.; Chen, N.-C.; Cheng, H.; Chin, C.-S.; Chow, W.; de Lima, L. G.; Dishuck, P. C.; Durbin, R.; Dvorkina, T.; Fiddes, I. T.; Formenti, G.; Fulton, R. S.; Functammasan, A.; Garrison, E.; Grady, P. G. S.; Graves-Lindsay, T. A.; Hall, I. M.; Hansen, N. F.; Hartley, G. A.; Haukness, M.; Howe, K.; Hunkapiller, M. W.; Jain, C.; Jain, M.; Jarvis, E. D.; Kerpedjiev, P.; Kirsche, M.; Kolmogorov, M.; Korlach, J.; Kremitzki, M.; Li, H.; Maduro, V. V.; Marschall, T.; McCartney, A. M.; McDaniel, J.; Miller, D. E.; Mullikin, J. C.; Myers, E. W.; Olson, N. D.; Paten, B.; Peluso, P.; Pevzner, P. A.; Porubsky, D.; Potapova, T.; Rogae, E. I.; Rosenfeld, J. A.; Salzberg, S. L.; Schneider, V. A.; Sedlazeck, F. J.; Shafin, K.; Shew, C. J.; Shumate, A.; Sims, Y.; Smit, A. F. A.; Soto, D. C.; Sović, I.; Storer, J. M.; Streets, A.; Sullivan, B. A.; Thibaud-Nissen, F.; Torrance, J.; Wagner, J.; Walenz, B. P.; Wenger, A.; Wood, J. M. D.; Xiao, C.; Yan, S. M.; Young, A. C.; Zarate, S.; Surti, U.; McCoy, R. C.; Dennis, M. Y.; Alexandrov, I. A.; Gerton, J. L.; O'Neill, R. J.; Timp, W.; Zook, J. M.; Schatz, M. C.; Eichler, E. E.; Miga, K. H.; Phillippy, A. M. The Complete Sequence of a Human Genome. *Science* **2022**, *376* (6588), 44–53. <https://doi.org/10.1126/science.abj6987>.

(86) Lam, E. Y. N.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. G-Quadruplex Structures Are Stable and Detectable in Human Genomic DNA. *Nat Commun* **2013**, *4*, 1796. <https://doi.org/10.1038/ncomms2792>.

(87) Fernando, H.; Rodriguez, R.; Balasubramanian, S. Selective Recognition of a DNA G-Quadruplex by an Engineered Antibody. *Biochemistry* **2008**, *47* (36), 9365–9371. <https://doi.org/10.1021/bi800983u>.

(88) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat Biotechnol* **2015**, *33* (8), 877–881. <https://doi.org/10.1038/nbt.3295>.

(89) Marsico, G.; Chambers, V. S.; Sahakyan, A. B.; McCauley, P.; Boutell, J. M.; Antonio, M. D.; Balasubramanian, S. Whole Genome Experimental Maps of DNA G-Quadruplexes in Multiple Species. *Nucleic Acids Research* **2019**, *47* (8), 3862–3874. <https://doi.org/10.1093/nar/gkz179>.

(90) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nature Chem* **2013**, *5* (3), 182–186. <https://doi.org/10.1038/nchem.1548>.

(91) Hänsel-Hertsch, R.; Beraldi, D.; Lensing, S. V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; Tannahill, D.; Balasubramanian, S. G-Quadruplex Structures Mark Human Regulatory Chromatin. *Nat Genet* **2016**, *48* (10), 1267–1272. <https://doi.org/10.1038/ng.3662>.

(92) Hänsel-Hertsch, R.; Spiegel, J.; Marsico, G.; Tannahill, D.; Balasubramanian, S. Genome-Wide Mapping of Endogenous G-Quadruplex DNA Structures by Chromatin Immunoprecipitation and High-Throughput Sequencing. *Nat Protoc* **2018**, *13* (3), 551–564. <https://doi.org/10.1038/nprot.2017.150>.

(93) Zheng, K.; Zhang, J.; He, Y.; Gong, J.; Wen, C.; Chen, J.; Hao, Y.; Zhao, Y.; Tan, Z. Detection of Genomic G-Quadruplexes in Living Cells Using a Small Artificial Protein. *Nucleic Acids Research* **2020**, *48* (20), 11706–11720. <https://doi.org/10.1093/nar/gkaa841>.

(94) Lyu, J.; Shao, R.; Kwong Yung, P. Y.; Elsässer, S. J. Genome-Wide Mapping of G-Quadruplex Structures with CUT&Tag. *Nucleic Acids Research* **2022**, *50* (3), e13.

<https://doi.org/10.1093/nar/gkab1073>.

- (95) Müller, S.; Kumari, S.; Rodriguez, R.; Balasubramanian, S. Small-Molecule-Mediated G-Quadruplex Isolation from Human Cells. *Nature Chem* **2010**, *2* (12), 1095–1098. <https://doi.org/10.1038/nchem.842>.
- (96) Zeng, W.; Wu, F.; Liu, C.; Yang, Y.; Wang, B.; Yuan, Y.; Wang, J.; Chen, Y.; Fu, B.; Wu, Z.; Zhou, X. Small-Molecule-Based Human Genome G4 Profiling Reveals Potential Gene Regulation Activity. *Chem. Commun.* **2019**, *55* (16), 2269–2272. <https://doi.org/10.1039/C8CC10052G>.
- (97) Bryan, T. M. G-Quadruplexes at Telomeres: Friend or Foe? *Molecules* **2020**, *25* (16), 3686. <https://doi.org/10.3390/molecules25163686>.
- (98) Wellinger, R. J.; Sen, D. The DNA Structures at the Ends of Eukaryotic Chromosomes. *European Journal of Cancer* **1997**, *33* (5), 735–749. [https://doi.org/10.1016/S0959-8049\(97\)00067-1](https://doi.org/10.1016/S0959-8049(97)00067-1).
- (99) Harley, C. B.; Futcher, A. B.; Greider, C. W. Telomeres Shorten during Ageing of Human Fibroblasts. *Nature* **1990**, *345* (6274), 458–460. <https://doi.org/10.1038/345458a0>.
- (100) Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W. Normal Human Chromosomes Have Long G-Rich Telomeric Overhangs at One End. *Genes Dev.* **1997**, *11* (21), 2801–2809. <https://doi.org/10.1101/gad.11.21.2801>.
- (101) Blackburn, E. H. Structure and Function of Telomeres. *Nature* **1991**, *350* (6319), 569–573. <https://doi.org/10.1038/350569a0>.
- (102) Lim, C. J.; Cech, T. R. Shaping Human Telomeres: From Shelterin and CST Complexes to Telomeric Chromatin Organization. *Nat Rev Mol Cell Biol* **2021**, *22* (4), 283–298. <https://doi.org/10.1038/s41580-021-00328-y>.
- (103) Olovnikov, A. M. Telomeres, Telomerase, and Aging: Origin of the Theory. *Experimental Gerontology* **1996**, *31* (4), 443–448. [https://doi.org/10.1016/0531-5565\(96\)00005-8](https://doi.org/10.1016/0531-5565(96)00005-8).
- (104) Srinivas, N.; Rachakonda, S.; Kumar, R. Telomeres and Telomere Length: A General Overview. *Cancers* **2020**, *12* (3), 558. <https://doi.org/10.3390/cancers12030558>.
- (105) Hahn, W. C.; Meyerson, M. Telomerase Activation, Cellular Immortalization and Cancer. *Annals of Medicine* **2001**, *33* (2), 123–129. <https://doi.org/10.3109/07853890109002067>.
- (106) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proceedings of the National Academy of Sciences* **2002**, *99* (18), 11593–11598. <https://doi.org/10.1073/pnas.182256799>.
- (107) Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. Putative DNA Quadruplex Formation within the Human *c-Kit* Oncogene. *J. Am. Chem. Soc.* **2005**, *127* (30), 10584–10589. <https://doi.org/10.1021/ja050823u>.
- (108) Cogo, S.; Xodo, L. E. G-Quadruplex Formation within the Promoter of the KRAS Proto-Oncogene and Its Effect on Transcription. *Nucleic Acids Research* **2006**, *34* (9), 2536–2549. <https://doi.org/10.1093/nar/gkl286>.
- (109) Monsen, R. C.; DeLeeuw, L.; Dean, W. L.; Gray, R. D.; Sabo, T. M.; Chakravarthy, S.; Chaires, J. B.; Trent, J. O. The hTERT Core Promoter Forms Three Parallel G-Quadruplexes. *Nucleic Acids Research* **2020**, *48* (10), 5720–5734. <https://doi.org/10.1093/nar/gkaa107>.
- (110) Sun, D. Facilitation of a Structural Transition in the Polypurine/Polypyrimidine Tract within the Proximal Promoter Region of the Human VEGF Gene by the Presence of Potassium and G-Quadruplex-Interactive Agents. *Nucleic Acids Research* **2005**, *33* (18), 6070–6080. <https://doi.org/10.1093/nar/gki917>.
- (111) Dexheimer, T. S.; Sun, D.; Hurley, L. H. Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the *Bcl-2* P1 Promoter. *J. Am. Chem. Soc.* **2006**, *128* (16), 5404–5415. <https://doi.org/10.1021/ja0563861>.
- (112) Marcu, K. B.; Bossone, S. A.; Patel, A. J. Myc Function and Regulation. *Annu Rev Biochem* **1992**, *61*, 809–860. <https://doi.org/10.1146/annurev.bi.61.070192.004113>.
- (113) Wierstra, I.; Alves, J. The C-myc Promoter: Still MysterY and Challenge. In *Advances in Cancer Research*; Academic Press, 2008; Vol. 99, pp 113–333. [https://doi.org/10.1016/S0065-230X\(07\)99004-1](https://doi.org/10.1016/S0065-230X(07)99004-1).

- (114) González, V.; Hurley, L. H. The C- MYC NHE III<sub>1</sub>: Function and Regulation. *Annu. Rev. Pharmacol. Toxicol.* **2010**, *50* (1), 111–129. <https://doi.org/10.1146/annurev.pharmtox.48.113006.094649>.
- (115) Le, H. T.; Miller, M. C.; Buscaglia, R.; Dean, W. L.; Holt, P. A.; Chaires, J. B.; Trent, J. O. 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 TMPyP4. *Org. Biomol. Chem.* **2012**, *10* (47), 9393–9404. <https://doi.org/10.1039/C2OB26504D>.
- (116) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. Small-Molecule Interaction with a Five-Guanine-Tract G-Quadruplex Structure from the Human MYC Promoter. *Nat Chem Biol* **2005**, *1* (3), 167–173. <https://doi.org/10.1038/nchembio723>.
- (117) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex Stabilization. *Biochemistry* **2005**, *44* (6), 2048–2058. <https://doi.org/10.1021/bi048242p>.
- (118) Grand, C. L.; Han, H.; Muñoz, R. M.; Weitman, S.; Von Hoff, D. D.; Hurley, L. H.; Bearss, D. J. The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo<sup>1</sup>. *Molecular Cancer Therapeutics* **2002**, *1* (8), 565–573.
- (119) Ugrinova, I.; Petrova, M.; Chalabi-Dchar, M.; Bouvet, P. Chapter Four - Multifaceted Nucleolin Protein and Its Molecular Partners in Oncogenesis. In *Advances in Protein Chemistry and Structural Biology*; Donev, R., Ed.; Protein-Protein Interactions in Human Disease, Part B; Academic Press, 2018; Vol. 111, pp 133–164. <https://doi.org/10.1016/bs.apcsb.2017.08.001>.
- (120) Santos, T.; Salgado, G. F.; Cabrita, E. J.; Cruz, C. Nucleolin: A Binding Partner of G-Quadruplex Structures. *Trends in Cell Biology* **2022**, *32* (7), 561–564. <https://doi.org/10.1016/j.tcb.2022.03.003>.
- (121) González, V.; Guo, K.; Hurley, L.; Sun, D. Identification and Characterization of Nucleolin as a C-Myc G-Quadruplex-Binding Protein. *Journal of Biological Chemistry* **2009**, *284* (35), 23622–23635. <https://doi.org/10.1074/jbc.M109.018028>.
- (122) Wang, Y.-H.; Yang, Q.-F.; Lin, X.; Chen, D.; Wang, Z.-Y.; Chen, B.; Han, H.-Y.; Chen, H.-D.; Cai, K.-C.; Li, Q.; Yang, S.; Tang, Y.-L.; Li, F. G4LDB 2.2: A Database for Discovering and Studying G-Quadruplex and i-Motif Ligands. *Nucleic Acids Research* **2022**, *50* (D1), D150–D160. <https://doi.org/10.1093/nar/gkab952>.
- (123) Read, M. A.; Wood, A. A.; Harrison, J. R.; Gowan, S. M.; Kelland, L. R.; Dosanjh, H. S.; Neidle, S. Molecular Modeling Studies on G-Quadruplex Complexes of Telomerase Inhibitors: Structure–Activity Relationships. *J. Med. Chem.* **1999**, *42* (22), 4538–4546. <https://doi.org/10.1021/jm990287e>.
- (124) Harrison, R. J.; Gowan, S. M.; Kelland, L. R.; Neidle, S. Human Telomerase Inhibition by Substituted Acridine Derivatives. *Bioorganic & Medicinal Chemistry Letters* **1999**, *9* (17), 2463–2468. [https://doi.org/10.1016/S0960-894X\(99\)00394-7](https://doi.org/10.1016/S0960-894X(99)00394-7).
- (125) Harrison, R. J.; Cuesta, J.; Chessari, G.; Read, M. A.; Basra, S. K.; Reszka, A. P.; Morrell, J.; Gowan, S. M.; Incles, C. M.; Tanius, F. A.; Wilson, W. D.; Kelland, L. R.; Neidle, S. Trisubstituted Acridine Derivatives as Potent and Selective Telomerase Inhibitors. *J. Med. Chem.* **2003**, *46* (21), 4463–4476. <https://doi.org/10.1021/jm0308693>.
- (126) Read, M.; Harrison, R. J.; Romagnoli, B.; Tanius, F. A.; Gowan, S. H.; Reszka, A. P.; Wilson, W. D.; Kelland, L. R.; Neidle, S. Structure-Based Design of Selective and Potent G Quadruplex-Mediated Telomerase Inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (9), 4844–4849. <https://doi.org/10.1073/pnas.081560598>.
- (127) Campbell, N. H.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. Structural Basis of DNA Quadruplex Recognition by an Acridine Drug. *J. Am. Chem. Soc.* **2008**, *130* (21), 6722–6724. <https://doi.org/10.1021/ja8016973>.
- (128) Monchaud, D.; Teulade-Fichou, M.-P. A Hitchhiker’s Guide to G-Quadruplex Ligands. *Org. Biomol. Chem.* **2008**, *6* (4), 627–636. <https://doi.org/10.1039/B714772B>.
- (129) Liu, T.; Wu, Y.; Qin, L.; Luo, Q.; Li, W.; Cheng, Y.; Tu, Y.; You, H. Nonselective Intercalation of G-Quadruplex-Targeting Ligands into Double-Stranded DNA Quantified by Single-

- Molecule Stretching. *J. Phys. Chem. B* **2023**, *127* (26), 5859–5868. <https://doi.org/10.1021/acs.jpcc.3c03031>.
- (130) Arora, A.; Maiti, S. Effect of Loop Orientation on Quadruplex–TMPyP4 Interaction. *J. Phys. Chem. B* **2008**, *112* (27), 8151–8159. <https://doi.org/10.1021/jp711608y>.
- (131) Cao, Q.; Li, Y.; Freisinger, E.; Qin, P. Z.; Sigel, R. K. O.; Mao, Z.-W. G-Quadruplex DNA Targeted Metal Complexes Acting as Potential Anticancer Drugs. *Inorg. Chem. Front.* **2017**, *4* (1), 10–32. <https://doi.org/10.1039/C6QI00300A>.
- (132) Vummidi, B. R.; Alzeer, J.; Luedtke, N. W. Fluorescent Probes for G-Quadruplex Structures. *ChemBioChem* **2013**, *14* (5), 540–558. <https://doi.org/10.1002/cbic.201200612>.
- (133) 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.* **2002**, *124* (10), 2098–2099. <https://doi.org/10.1021/ja017308q>.
- (134) 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.* **2007**, *129* (7), 1856–1857. <https://doi.org/10.1021/ja067352b>.
- (135) Monchaud, D.; Yang, P.; Lacroix, L.; Teulade-Fichou, M.-P.; Mergny, J.-L. A Metal-Mediated Conformational Switch Controls G-Quadruplex Binding Affinity. *Angew. Chem. Int. Ed.* **2008**, *47* (26), 4858–4861. <https://doi.org/10.1002/anie.200800468>.
- (136) Granotier, C. Preferential Binding of a G-Quadruplex Ligand to Human Chromosome Ends. *Nucleic Acids Research* **2005**, *33* (13), 4182–4190. <https://doi.org/10.1093/nar/gki722>.
- (137) Masson, T.; Landras Guetta, C.; Laigre, E.; Cucchiari, A.; Duchambon, P.; Teulade-Fichou, M.-P.; Verga, D. BrdU Immuno-Tagged G-Quadruplex Ligands: A New Ligand-Guided Immunofluorescence Approach for Tracking G-Quadruplexes in Cells. *Nucleic Acids Res* **2021**, *49* (22), 12644–12660. <https://doi.org/10.1093/nar/gkab1166>.
- (138) Rodriguez, R.; Müller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, F.; Balasubramanian, S. A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. **2009**.
- (139) Nikan, M.; Sherman, J. C. Cation-Complexation Behavior of Template-Assembled Synthetic G-Quartets. *J. Org. Chem.* **2009**, *74* (15), 5211–5218. <https://doi.org/10.1021/jo9001245>.
- (140) Stefan, L.; Guédin, A.; Amrane, S.; Smith, N.; Denat, F.; Mergny, J.-L.; Monchaud, D. DOTASQ as a Prototype of Nature-Inspired G-Quadruplex Ligand. *Chem. Commun.* **2011**, *47* (17), 4992. <https://doi.org/10.1039/c0cc04960c>.
- (141) Xu, H.-J.; Stefan, L.; Haudecoeur, R.; Vuong, S.; Richard, P.; Denat, F.; Barbe, J.-M.; Gros, C. P.; Monchaud, D. Porphyrin-Templated Synthetic G-Quartet (PorphySQ): A Second Prototype of G-Quartet-Based G-Quadruplex Ligand. *Org. Biomol. Chem.* **2012**, *10* (27), 5212. <https://doi.org/10.1039/c2ob25601k>.
- (142) Haudecoeur, R.; Stefan, L.; Denat, F.; Monchaud, D. A Model of Smart G-Quadruplex Ligand. *J. Am. Chem. Soc.* **2013**, *135* (2), 550–553. <https://doi.org/10.1021/ja310056y>.
- (143) Laguerre, A.; Desbois, N.; Stefan, L.; Richard, P.; Gros, C. P.; Monchaud, D. Porphyrin-Based Design of Bioinspired Multitarget Quadruplex Ligands. *ChemMedChem* **2014**, *9* (9), 2035–2039. <https://doi.org/10.1002/cmcd.201300526>.
- (144) Sperti, F. R.; Charbonnier, T.; Lejault, P.; Zell, J.; Bernhard, C.; Valverde, I. E.; Monchaud, D. Biomimetic, Smart, and Multivalent Ligands for G-Quadruplex Isolation and Bioorthogonal Imaging. *ACS Chem. Biol.* **2021**, *16* (5), 905–914. <https://doi.org/10.1021/acscchembio.1c00111>.
- (145) Feng, Y.; He, Z.; Luo, Z.; Sperti, F. R.; Valverde, I. E.; Zhang, W.; Monchaud, D. Side-by-Side Comparison of G-Quadruplex (G4) Capture Efficiency of the Antibody BG4 versus the Small-Molecule Ligands TASQs. *iScience* **2023**, *26* (6), 106846. <https://doi.org/10.1016/j.isci.2023.106846>.
- (146) Monchaud, D. Template-Assembled Synthetic G-Quartets (TASQs): multiTASQing Molecular Tools for Investigating DNA and RNA G-Quadruplex Biology. *Acc. Chem. Res.* **2023**, *56* (3), 350–362. <https://doi.org/10.1021/acs.accounts.2c00757>.
- (147) Xu, H.; Hurley, L. H. A First-in-Class Clinical G-Quadruplex-Targeting Drug. The Bench-to-Bedside Translation of the Fluoroquinolone QQ58 to CX-5461 (Pidnarulex). *Bioorganic & Medicinal*

*Chemistry Letters* **2022**, *77*, 129016. <https://doi.org/10.1016/j.bmcl.2022.129016>.

- (148) Drygin, D.; Siddiqui-Jain, A.; O'Brien, S.; Schwaebe, M.; Lin, A.; Bliesath, J.; Ho, C. B.; Proffitt, C.; Trent, K.; Whitten, J. P.; Lim, J. K. C.; Von Hoff, D.; Anderes, K.; Rice, W. G. Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Research* **2009**, *69* (19), 7653–7661. <https://doi.org/10.1158/0008-5472.CAN-09-1304>.
- (149) Xu, H.; Di Antonio, M.; McKinney, S.; Mathew, V.; Ho, B.; O'Neil, N. J.; Santos, N. D.; Silvester, J.; Wei, V.; Garcia, J.; Kabeer, F.; Lai, D.; Soriano, P.; Banáth, J.; Chiu, D. S.; Yap, D.; Le, D. D.; Ye, F. B.; Zhang, A.; Thu, K.; Soong, J.; Lin, S.; Tsai, A. H. C.; Osako, T.; Algara, T.; Saunders, D. N.; Wong, J.; Xian, J.; Bally, M. B.; Brenton, J. D.; Brown, G. W.; Shah, S. P.; Cescon, D.; Mak, T. W.; Caldas, C.; Stirling, P. C.; Hieter, P.; Balasubramanian, S.; Aparicio, S. CX-5461 Is a DNA G-Quadruplex Stabilizer with Selective Lethality in BRCA1/2 Deficient Tumours. *Nat Commun* **2017**, *8* (1), 14432. <https://doi.org/10.1038/ncomms14432>.
- (150) Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Plückthun, A. *In Vitro* Generated Antibodies Specific for Telomeric Guanine-Quadruplex DNA React with *Stylonychia Lemnae* Macronuclei. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (15), 8572–8577. <https://doi.org/10.1073/pnas.141229498>.
- (151) Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh, R. M., Jr; Sen, D.; Lansdorp, P. M. Detection of G-Quadruplex DNA in Mammalian Cells. *Nucleic Acids Research* **2014**, *42* (2), 860–869. <https://doi.org/10.1093/nar/gkt957>.
- (152) Liu, H.-Y.; Zhao, Q.; Zhang, T.-P.; Wu, Y.; Xiong, Y.-X.; Wang, S.-K.; Ge, Y.-L.; He, J.-H.; Lv, P.; Ou, T.-M.; Tan, J.-H.; Li, D.; Gu, L.-Q.; Ren, J.; Zhao, Y.; Huang, Z.-S. Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chemical Biology* **2016**, *23* (10), 1261–1270. <https://doi.org/10.1016/j.chembiol.2016.08.013>.
- (153) Javadekar, S. M.; Nilavar, N. M.; Paranjape, A.; Das, K.; Raghavan, S. C. Characterization of G-Quadruplex Antibody Reveals Differential Specificity for G4 DNA Forms. *DNA Research* **2020**, *27* (5), dsaa024. <https://doi.org/10.1093/dnares/dsaa024>.
- (154) Kazemier, H. G.; Paeschke, K.; Lansdorp, P. M. Guanine Quadruplex Monoclonal Antibody 1H6 Cross-React with Restrained Thymidine-Rich Single Stranded DNA. *Nucleic Acids Research* **2017**, *45* (10), 5913–5919. <https://doi.org/10.1093/nar/gkx245>.
- (155) Galli, S.; Melidis, L.; Flynn, S. M.; Varshney, D.; Simeone, A.; Spiegel, J.; Madden, S. K.; Tannahill, D.; Balasubramanian, S. DNA G-Quadruplex Recognition In Vitro and in Live Cells by a Structure-Specific Nanobody. *J. Am. Chem. Soc.* **2022**, *144* (50), 23096–23103. <https://doi.org/10.1021/jacs.2c10656>.
- (156) Decian, A.; Guittat, L.; Kaiser, M.; Sacca, B.; Amrane, S.; Bourdoncle, A.; Alberti, P.; Teuladefichou, M.; Lacroix, L.; Mergny, J. Fluorescence-Based Melting Assays for Studying Quadruplex Ligands. *Methods* **2007**, *42* (2), 183–195. <https://doi.org/10.1016/j.ymeth.2006.10.004>.
- (157) De Rache, A.; Mergny, J.-L. Assessment of Selectivity of G-Quadruplex Ligands via an Optimised FRET Melting Assay. *Biochimie* **2015**, *115*, 194–202. <https://doi.org/10.1016/j.biochi.2015.06.002>.
- (158) 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 & Medicinal Chemistry Letters* **2006**, *16* (18), 4842–4845. <https://doi.org/10.1016/j.bmcl.2006.06.067>.
- (159) Monchaud, D.; Allain, C.; Teulade-Fichou, M.-P. Thiazole Orange: A Useful Probe for Fluorescence Sensing of G-Quadruplex–Ligand Interactions. *Nucleosides, Nucleotides and Nucleic Acids* **2007**, *26* (10–12), 1585–1588. <https://doi.org/10.1080/15257770701548212>.
- (160) Nygren, J.; Svanvik, N.; Kubista, M. The Interactions between the Fluorescent Dye Thiazole Orange and DNA. *Biopolymers* **1998**, *46* (1), 39–51. [https://doi.org/10.1002/\(SICI\)1097-0282\(199807\)46:1<39::AID-BIP4>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0282(199807)46:1<39::AID-BIP4>3.0.CO;2-Z).
- (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* **2011**, *400* (10), 3419–3427. <https://doi.org/10.1007/s00216-011->

- (162) Halder, K.; Largy, E.; Benzler, M.; Teulade-Fichou, M.-P.; Hartig, J. S. Efficient Suppression of Gene Expression by Targeting 5'-UTR-Based RNA Quadruplexes with Bisquinolinium Compounds. *ChemBioChem* **2011**, *12* (11), 1663–1668. <https://doi.org/10.1002/cbic.201100228>.
- (163) Marchand, A.; Granzhan, A.; Iida, K.; Tsushima, Y.; Ma, Y.; Nagasawa, K.; Teulade-Fichou, M.-P.; Gabelica, V. Ligand-Induced Conformational Changes with Cation Ejection upon Binding to Human Telomeric DNA G-Quadruplexes. *J. Am. Chem. Soc.* **2015**, *137* (2), 750–756. <https://doi.org/10.1021/ja5099403>.
- (164) Zuffo, M.; Doria, F.; Botti, S.; Bergamaschi, G.; Freccero, M. G-Quadruplex Fluorescence Sensing by Core-Extended Naphthalene Diimides. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2017**, *1861* (5), 1303–1311. <https://doi.org/10.1016/j.bbagen.2016.11.034>.
- (165) Beauvineau, C.; Guetta, C.; Teulade-Fichou, M.-P.; Mahuteau-Betzer, F. PhenDV, a Turn-off Fluorescent Quadruplex DNA Probe for Improving the Sensitivity of Drug Screening Assays. *Org. Biomol. Chem.* **2017**, *15* (34), 7117–7121. <https://doi.org/10.1039/C7OB01705G>.
- (166) Furlong, W. R.; Rubinski, M. A.; Indralingam, R. The Method of Continuous Variation: A Laboratory Investigation of the Formula of a Precipitate. *J. Chem. Educ.* **2013**, *90* (7), 937–940. <https://doi.org/10.1021/ed3004337>.
- (167) Owen, W. R.; Stewart, P. J. Kinetics and Mechanism of Chlorambucil Hydrolysis. *Journal of Pharmaceutical Sciences* **1979**, *68* (8), 992–996. <https://doi.org/10.1002/jps.2600680819>.
- (168) Kundu, G. C.; Schullek, J. R.; Wilson, I. B. The Alkylating Properties of Chlorambucil. *Pharmacology Biochemistry and Behavior* **1994**, *49* (3), 621–624. [https://doi.org/10.1016/0091-3057\(94\)90078-7](https://doi.org/10.1016/0091-3057(94)90078-7).
- (169) Catovsky, D.; Else, M.; Richards, S. Chlorambucil—Still Not Bad: A Reappraisal. *Clinical Lymphoma Myeloma and Leukemia* **2011**, *11*, S2–S6. <https://doi.org/10.1016/j.clml.2011.02.006>.
- (170) Di Antonio, M.; McLuckie, K. I. E.; Balasubramanian, S. Reprogramming the Mechanism of Action of Chlorambucil by Coupling to a G-Quadruplex Ligand. *J. Am. Chem. Soc.* **2014**, *136* (16), 5860–5863. <https://doi.org/10.1021/ja5014344>.
- (171) Paul, D.; Marchand, A.; Verga, D.; Teulade-Fichou, M.-P.; Bombard, S.; Rosu, F.; Gabelica, V. Probing Ligand and Cation Binding Sites in G-Quadruplex Nucleic Acids by Mass Spectrometry and Electron Photodetachment Dissociation Sequencing. *Analyst* **2019**, *144* (11), 3518–3524. <https://doi.org/10.1039/C9AN00398C>.
- (172) Pande, P.; Shearer, J.; Yang, J.; Greenberg, W. A.; Rokita, S. E. Alkylation of Nucleic Acids by a Model Quinone Methide. *J. Am. Chem. Soc.* **1999**, *121* (29), 6773–6779. <https://doi.org/10.1021/ja990456k>.
- (173) Minard, A.; Liano, D.; Wang, X.; Di Antonio, M. The Unexplored Potential of Quinone Methides in Chemical Biology. *Bioorganic & Medicinal Chemistry* **2019**, *27* (12), 2298–2305. <https://doi.org/10.1016/j.bmc.2019.04.001>.
- (174) Rokita, S. E. Reversible Alkylation of DNA by Quinone Methides. In *Quinone Methides*; Rokita, S. E., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2009; pp 297–327. <https://doi.org/10.1002/9780470452882.ch9>.
- (175) Di Antonio, M.; Doria, F.; Richter, S. N.; Bertipaglia, C.; Mella, M.; Sissi, C.; Palumbo, M.; Freccero, M. Quinone Methides Tethered to Naphthalene Diimides as Selective G-Quadruplex Alkylating Agents. *J. Am. Chem. Soc.* **2009**, *131* (36), 13132–13141. <https://doi.org/10.1021/ja904876q>.
- (176) Sarmah, A.; Roy, R. K. Understanding the Preferential Binding Interaction of Aqua-Cisplatin with Nucleobase Guanine over Adenine: A Density Functional Reactivity Theory Based Approach. *RSC Adv.* **2013**, *3* (8), 2822. <https://doi.org/10.1039/c2ra23223e>.
- (177) Bertrand, H.; Bombard, S.; Monchaud, D.; Teulade-Fichou, M.-P. A Platinum–Quinacridine Hybrid as a G-Quadruplex Ligand. *J Biol Inorg Chem* **2007**, *12* (7), 1003–1014. <https://doi.org/10.1007/s00775-007-0273-3>.
- (178) Betzer, J.-F.; Nuter, F.; Chtchigrovsky, M.; Hamon, F.; Kellermann, G.; Ali, S.; Calméjane, M.-A.; Roque, S.; Poupon, J.; Cresteil, T.; Teulade-Fichou, M.-P.; Marinetti, A.; Bombard, S. Linking of Antitumor Trans NHC-Pt(II) Complexes to G-Quadruplex DNA Ligand for Telomeric Targeting.

- Bioconjugate Chem.* **2016**, *27* (6), 1456–1470. <https://doi.org/10.1021/acs.bioconjchem.6b00079>.
- (179) Verga, D.; Hamon, F.; Poyer, F.; Bombard, S.; Teulade-Fichou, M.-P. Photo-Cross-Linking Probes for Trapping G-Quadruplex DNA. *Angew. Chem. Int. Ed.* **2014**, *53* (4), 994–998. <https://doi.org/10.1002/anie.201307413>.
- (180) Charlier, M.; Helene, C. PHOTOCHEMICAL REACTIONS OF AROMATIC KETONES WITH NUCLEIC ACIDS AND THEIR COMPONENTS I— PURINE AND PYRIMIDINE BASES AND NUCLEOSIDES. *Photochem Photobiol* **1972**, *15* (1), 71–87. <https://doi.org/10.1111/j.1751-1097.1972.tb06224.x>.
- (181) Michalak, J.; Zhai, H. B.; Platz, M. S. The Photochemistry of Various Para-Substituted Tetrafluorophenyl Azides in Acidic Media and the Formation of Nitrenium Ions. *J. Phys. Chem.* **1996**, *100* (33), 14028–14036. <https://doi.org/10.1021/jp961100n>.
- (182) Lena, A.; Benassi, A.; Stasi, M.; Saint-Pierre, C.; Freccero, M.; Gasparutto, D.; Bombard, S.; Doria, F.; Verga, D. Photoactivatable V-Shaped Bifunctional Quinone Methide Precursors as a New Class of Selective G-quadruplex Alkylating Agents. *Chemistry A European J* **2022**, *28* (35). <https://doi.org/10.1002/chem.202200734>.
- (183) Cadoni, E.; Manicardi, A.; Fossépré, M.; Heirwegh, K.; Surin, M.; Madder, A. Teaching Photosensitizers a New Trick: Red Light-Triggered G-Quadruplex Alkylation by Ligand Co-Localization. *Chem. Commun.* **2021**, *57* (8), 1010–1013. <https://doi.org/10.1039/D0CC06030E>.
- (184) Cadoni, E.; De Paepe, L.; Colpaert, G.; Tack, R.; Waegeman, D.; Manicardi, A.; Madder, A. A Red Light-Triggered Chemical Tool for Sequence-Specific Alkylation of G-Quadruplex and I-Motif DNA. *Nucleic Acids Research* **2023**, *51* (9), 4112–4125. <https://doi.org/10.1093/nar/gkad189>.
- (185) Bayley, H.; Knowles, J. R. [8] Photoaffinity Labeling. In *Affinity labeling; Methods in Enzymology*; Academic Press, 1977; Vol. 46, pp 69–114. [https://doi.org/10.1016/S0076-6879\(77\)46012-9](https://doi.org/10.1016/S0076-6879(77)46012-9).
- (186) A. Fleming, S. Chemical Reagents in Photoaffinity Labeling. *Tetrahedron* **1995**, *51* (46), 12479–12520. [https://doi.org/10.1016/0040-4020\(95\)00598-3](https://doi.org/10.1016/0040-4020(95)00598-3).
- (187) Burton, N. R.; Kim, P.; Backus, K. M. Photoaffinity Labelling Strategies for Mapping the Small Molecule–Protein Interactome. *Org. Biomol. Chem.* **2021**, *19* (36), 7792–7809. <https://doi.org/10.1039/D1OB01353J>.
- (188) Chapter 4 The Photoaffinity Labeling Experiment. In *Laboratory Techniques in Biochemistry and Molecular Biology*; Work, T. S., Burdon, R. H., Eds.; Photogenerated Reagents in Biochemistry and Molecular Biology; Elsevier, 1983; Vol. 12, pp 66–111. [https://doi.org/10.1016/S0075-7535\(08\)70436-4](https://doi.org/10.1016/S0075-7535(08)70436-4).
- (189) Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J.-S. Photo-Affinity Labeling (PAL) in Chemical Proteomics: A Handy Tool to Investigate Protein-Protein Interactions (PPIs). *Proteome Sci* **2016**, *15* (1), 1–34. <https://doi.org/10.1186/s12953-017-0123-3>.
- (190) Gritsan, N. P.; Gudmundsdóttir, A. D.; Tigelaar, D.; Platz, M. S. Laser Flash Photolysis Study of Methyl Derivatives of Phenyl Azide. *J. Phys. Chem. A* **1999**, *103* (18), 3458–3461. <https://doi.org/10.1021/jp984624r>.
- (191) Leyva, E.; Platz, M. S.; Moctezuma, E. Investigation of Phenyl Azide Photochemistry by Conventional and Time-Resolved Spectroscopy. Elucidation of Intermediates and Reaction Mechanisms. *Journal of Photochemistry and Photobiology* **2022**, *11*, 100126. <https://doi.org/10.1016/j.jpap.2022.100126>.
- (192) Campbell, P.; Gioannini, T. L. THE USE OF BENZOPHENONE AS A PHOTOAFFINITY LABEL, LABELING IN p-BENZOYLPHENYLACETYL CHYMOTRYPSIN AT UNIT EFFICIENCY. *Photochem Photobiol* **1979**, *29* (5), 883–892. <https://doi.org/10.1111/j.1751-1097.1979.tb07787.x>.
- (193) Cuquerella, M. C.; Lhiaubet-Vallet, V.; Cadet, J.; Miranda, M. A. Benzophenone Photosensitized DNA Damage. *Acc. Chem. Res.* **2012**, *45* (9), 1558–1570. <https://doi.org/10.1021/ar300054e>.
- (194) Smith, E.; Collins, I. Photoaffinity Labeling in Target- and Binding-Site Identification. *Future Medicinal Chemistry* **2015**, *7* (2), 159–183. <https://doi.org/10.4155/fmc.14.152>.
- (195) Brunner, J.; Senn, H.; Richards, F. M. 3-Trifluoromethyl-3-Phenyldiazirine. A New Carbene

- Generating Group for Photolabeling Reagents. *J Biol Chem* **1980**, 255 (8), 3313–3318.
- (196) Chen, Y.; Topp, E. M. Photolytic Labeling and Its Applications in Protein Drug Discovery and Development. *Journal of Pharmaceutical Sciences* **2019**, 108 (2), 791–797. <https://doi.org/10.1016/j.xphs.2018.10.017>.
- (197) Das, J. Aliphatic Diazirines as Photoaffinity Probes for Proteins: Recent Developments. *Chem. Rev.* **2011**, 111 (8), 4405–4417. <https://doi.org/10.1021/cr1002722>.
- (198) West, A. V.; Muncipinto, G.; Wu, H.-Y.; Huang, A. C.; Labenski, M. T.; Jones, L. H.; Woo, C. M. Labeling Preferences of Diazirines with Protein Biomolecules. *J. Am. Chem. Soc.* **2021**, 143 (17), 6691–6700. <https://doi.org/10.1021/jacs.1c02509>.
- (199) Dubinsky, L.; Krom, B. P.; Meijler, M. M. Diazirine Based Photoaffinity Labeling. *Bioorganic & Medicinal Chemistry* **2012**, 20 (2), 554–570. <https://doi.org/10.1016/j.bmc.2011.06.066>.
- (200) Doyle, M. P.; Duffy, R.; Ratnikov, M.; Zhou, L. Catalytic Carbene Insertion into C–H Bonds. *Chem. Rev.* **2010**, 110 (2), 704–724. <https://doi.org/10.1021/cr900239n>.
- (201) Xue, J.; Luk, H. L.; Platz, M. S. Direct Observation of a Carbene-Alcohol Ylide. *J. Am. Chem. Soc.* **2011**, 133 (6), 1763–1765. <https://doi.org/10.1021/ja111029g>.
- (202) Su, H.; Xu, J.; Chen, Y.; Wang, Q.; Lu, Z.; Chen, Y.; Chen, K.; Han, S.; Fang, Z.; Wang, P.; Yuan, B.-F.; Zhou, X. Photoactive G-Quadruplex Ligand Identifies Multiple G-Quadruplex-Related Proteins with Extensive Sequence Tolerance in the Cellular Environment. *J. Am. Chem. Soc.* **2021**, 143 (4), 1917–1923. <https://doi.org/10.1021/jacs.0c10792>.
- (203) Zhang, X.; Spiegel, J.; Martínez Cuesta, S.; Adhikari, S.; Balasubramanian, S. Chemical Profiling of DNA G-Quadruplex-Interacting Proteins in Live Cells. *Nat. Chem.* **2021**, 13 (7), 626–633. <https://doi.org/10.1038/s41557-021-00736-9>.
- (204) Mortison, J. D.; Schenone, M.; Myers, J. A.; Zhang, Z.; Chen, L.; Ciarlo, C.; Comer, E.; Natchiar, S. K.; Carr, S. A.; Klaholz, B. P.; Myers, A. G. Tetracyclines Modify Translation by Targeting Key Human rRNA Substructures. *Cell Chemical Biology* **2018**, 25 (12), 1506-1518.e13. <https://doi.org/10.1016/j.chembiol.2018.09.010>.
- (205) Mukherjee, H.; Blain, J. C.; Vandivier, L. E.; Chin, D. N.; Friedman, J. E.; Liu, F.; Maillet, A.; Fang, C.; Kaplan, J. B.; Li, J.; Chenoweth, D. M.; Christensen, A. B.; Petersen, L. K.; Hansen, N. J. V.; Barrera, L.; Kubica, N.; Kumaravel, G.; Petter, J. C. PEARL-Seq: A Photoaffinity Platform for the Analysis of Small Molecule-RNA Interactions. *ACS Chem. Biol.* **2020**, 15 (9), 2374–2381. <https://doi.org/10.1021/acscchembio.0c00357>.
- (206) Henderson, E.; Hardin, C. C.; Walk, S. K.; Tinoco, I.; Blackburn, E. H. Telomeric DNA Oligonucleotides Form Novel Intramolecular Structures Containing Guanine·guanine Base Pairs. *Cell* **1987**, 51 (6), 899–908. [https://doi.org/10.1016/0092-8674\(87\)90577-0](https://doi.org/10.1016/0092-8674(87)90577-0).
- (207) Sen, D.; Gilbert, W. Formation of Parallel Four-Stranded Complexes by Guanine-Rich Motifs in DNA and Its Implications for Meiosis. *Nature* **1988**, 334 (6180), 364–366. <https://doi.org/10.1038/334364a0>.
- (208) Puig Lombardi, E.; Londoño-Vallejo, A. A Guide to Computational Methods for G-Quadruplex Prediction. *Nucleic Acids Research* **2020**, 48 (1), 1–15. <https://doi.org/10.1093/nar/gkz1097>.
- (209) O’Hagan, M. P.; Morales, J. C.; Galan, M. C. Binding and Beyond: What Else Can G-Quadruplex Ligands Do? *Eur J Org Chem* **2019**, 2019 (31–32), 4995–5017. <https://doi.org/10.1002/ejoc.201900692>.
- (210) 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. Chem. Int. Ed.* **2009**, 48 (12), 2188–2191. <https://doi.org/10.1002/anie.200805613>.
- (211) Masson, T. Nouveaux ligands fonctionnels des G-quadruplexes pour l’imagerie cellulaire et la capture covalente d’ADN génomique, Université Paris-Saclay, Paris-Saclay, 2021.
- (212) Mendes, E.; Aljnadi, I. M.; Bahls, B.; Victor, B. L.; Paulo, A. Major Achievements in the Design of Quadruplex-Interactive Small Molecules. *Pharmaceuticals* **2022**, 15 (3), 300. <https://doi.org/10.3390/ph15030300>.
- (213) Xie, X.; Reznichenko, O.; Chaput, L.; Martin, P.; Teulade-Fichou, M.-P.; Granzhan, A.

- Topology-Selective, Fluorescent “Light-Up” Probes for G-Quadruplex DNA Based on Photoinduced Electron Transfer. *Chemistry – A European Journal* **2018**, *24* (48), 12638–12651. <https://doi.org/10.1002/chem.201801701>.
- (214) Luo, Y.; Granzhan, A.; Verga, D.; Mergny, J. FRET-MC: A Fluorescence Melting Competition Assay for Studying G4 Structures in Vitro. *Biopolymers* **2021**, *112* (4). <https://doi.org/10.1002/bip.23415>.
- (215) Lefebvre, J. Outils moléculaires pour l'étude des G-quadruplex au sein du génome, Université Paris-Saclay, Institut Curie - Orsay, 2017.
- (216) 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 G-Quadruplex Stability. *Nucleic Acids Research* **2011**, *39* (20), 9023–9033. <https://doi.org/10.1093/nar/gkr612>.
- (217) Puig Lombardi, E.; Holmes, A.; Verga, D.; Teulade-Fichou, M.-P.; Nicolas, A.; Londoño-Vallejo, A. Thermodynamically Stable and Genetically Unstable G-Quadruplexes Are Depleted in Genomes across Species. *Nucleic Acids Research* **2019**, *47* (12), 6098–6113. <https://doi.org/10.1093/nar/gkz463>.
- (218) Amrane, S.; Adrian, M.; Heddi, B.; Serero, A.; Nicolas, A.; Mergny, J.-L.; Phan, A. T. Formation of Pearl-Necklace Monomorphic G-Quadruplexes in the Human CEB25 Minisatellite. *J. Am. Chem. Soc.* **2012**, *134* (13), 5807–5816. <https://doi.org/10.1021/ja208993r>.
- (219) Matsumoto, K.; Ema, M. Roles of VEGF-A Signalling in Development, Regeneration, and Tumours. *Journal of Biochemistry* **2014**, *156* (1), 1–10. <https://doi.org/10.1093/jb/mvu031>.
- (220) Agrawal, P.; Hatzakis, E.; Guo, K.; Carver, M.; Yang, D. Solution Structure of the Major G-Quadruplex Formed in the Human VEGF Promoter in K<sup>+</sup>: Insights into Loop Interactions of the Parallel G-Quadruplexes. *Nucleic Acids Research* **2013**, *41* (22), 10584–10592. <https://doi.org/10.1093/nar/gkt784>.
- (221) Morgan, R. K.; Batra, H.; Gaerig, V. C.; Hockings, J.; Brooks, T. A. Identification and Characterization of a New G-Quadruplex Forming Region within the KRAS Promoter as a Transcriptional Regulator. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2016**, *1859* (2), 235–245. <https://doi.org/10.1016/j.bbagr.2015.11.004>.
- (222) 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* **2007**, *3* (4), 218–221. <https://doi.org/10.1038/nchembio864>.
- (223) Redon, S.; Bombard, S.; Elizondo-Riojas, M.; Chottard, J. Platinum Cross-linking of Adenines and Guanines on the Quadruplex Structures of the AG 3 (T 2 AG 3 ) 3 and (T 2 AG 3 ) 4 Human Telomere Sequences in Na<sup>+</sup> and K<sup>+</sup> Solutions. *Nucleic Acids Research* **2003**, *31* (6), 1605–1613. <https://doi.org/10.1093/nar/gkg259>.
- (224) Hassan, M. M.; Olaoye, O. O. Recent Advances in Chemical Biology Using Benzophenones and Diazirines as Radical Precursors. *Molecules* **2020**, *25* (10), 2285. <https://doi.org/10.3390/molecules25102285>.
- (225) Marazzi, M.; Mai, S.; Roca-Sanjuán, D.; Delcey, M. G.; Lindh, R.; González, L.; Monari, A. Benzophenone Ultrafast Triplet Population: Revisiting the Kinetic Model by Surface-Hopping Dynamics. *J. Phys. Chem. Lett.* **2016**, *7* (4), 622–626. <https://doi.org/10.1021/acs.jpcclett.5b02792>.
- (226) Yabumoto, S.; Sato, S.; Hamaguchi, H. Vibrational and Electronic Infrared Absorption Spectra of Benzophenone in the Lowest Excited Triplet State. *Chemical Physics Letters* **2005**, *416* (1–3), 100–103. <https://doi.org/10.1016/j.cpllett.2005.09.025>.
- (227) Battacharyya, S. N.; Das, P. K. Photoreduction of Benzophenone by Amino Acids, Aminopolycarboxylic Acids and Their Metal Complexes. A Laser-Flash-Photolysis Study. *J. Chem. Soc., Faraday Trans. 2* **1984**, *80* (9), 1107. <https://doi.org/10.1039/f29848001107>.
- (228) Deseke, E.; Nakatani, Y.; Ourisson, G. Intrinsic Reactivities of Amino Acids towards Photoalkylation with Benzophenone – A Study Preliminary to Photolabelling of the Transmembrane Protein Glycophorin A. *Eur. J. Org. Chem.* **1998**, *1998* (2), 243–251. [https://doi.org/10.1002/\(SICI\)1099-0690\(199802\)1998:2<243::AID-EJOC243>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1099-0690(199802)1998:2<243::AID-EJOC243>3.0.CO;2-I).
- (229) Jakubovska, J.; Tauraitė, D.; Meškys, R. A Versatile Method for the UVA-Induced Cross-Linking of Acetophenone- or Benzophenone-Functionalized DNA. *Sci Rep* **2018**, *8* (1), 16484.

<https://doi.org/10.1038/s41598-018-34892-9>.

(230) Farrell, I. S.; Toroney, R.; Hazen, J. L.; Mehl, R. A.; Chin, J. W. Photo-Cross-Linking Interacting Proteins with a Genetically Encoded Benzophenone. *Nat Methods* **2005**, *2* (5), 377–384. <https://doi.org/10.1038/nmeth0505-377>.

(231) Lee, H. S.; Dimla, R. D.; Schultz, P. G. Protein–DNA Photo-Crosslinking with a Genetically Encoded Benzophenone-Containing Amino Acid. *Bioorganic & Medicinal Chemistry Letters* **2009**, *19* (17), 5222–5224. <https://doi.org/10.1016/j.bmcl.2009.07.011>.

(232) Cheng, M.; Guo, C.; Gross, M. L. The Application of Fluorine-Containing Reagents in Structural Proteomics. *Angew Chem Int Ed* **2020**, *59* (15), 5880–5889. <https://doi.org/10.1002/anie.201907662>.

(233) O'Brien, J. G. K.; Jemas, A.; Asare-Okai, P. N.; am Ende, C. W.; Fox, J. M. Probing the Mechanism of Photoaffinity Labeling by Dialkyldiazirines through Bioorthogonal Capture of Diazoalkanes. *Org. Lett.* **2020**, *22* (24), 9415–9420. <https://doi.org/10.1021/acs.orglett.0c02714>.

(234) Peon, J.; Polshakov, D.; Kohler, B. Solvent Reorganization Controls the Rate of Proton Transfer from Neat Alcohol Solvents to Singlet Diphenylcarbene. *J. Am. Chem. Soc.* **2002**, *124* (22), 6428–6438. <https://doi.org/10.1021/ja017485r>.

(235) Wang, J.; Kubicki, J.; Gustafson, T. L.; Platz, M. S. The Dynamics of Carbene Solvation: An Ultrafast Study of p-Biphenyltrifluoromethylcarbene. *J. Am. Chem. Soc.* **2008**, *130* (7), 2304–2313. <https://doi.org/10.1021/ja077705m>.

(236) Caldwell, R. A. Laser Flash Photolysis Studies of Intersystem Crossing in Biradicals and Alkene Triplets. In *Kinetics and Spectroscopy of Carbenes and Biradicals*; Platz, M. S., Ed.; Springer US: Boston, MA, 1990; pp 77–116. [https://doi.org/10.1007/978-1-4899-3707-0\\_4](https://doi.org/10.1007/978-1-4899-3707-0_4).

(237) Moss, R. A.; Tian, J.; Chu, G.; Sauers, R. R.; Krogh-Jespersen, K. New Mechanisms Centered on Reactive Intermediates: Examples from Diazirine and Carbene Chemistry. *Pure and Applied Chemistry* **2007**, *79* (6), 993–1001. <https://doi.org/10.1351/pac200779060993>.

(238) Platz, M. S.; White, W. R.; Modarelli, D. A.; Celebi, S. Time Resolved Spectroscopy of Carbenepyridene Ylides: Distinguishing Carbenes from Diazirine Excited States. *Res. Chem. Intermed.* **1994**, *20* (2), 175–193. <https://doi.org/10.1163/156856794X00180>.

(239) Lhiaubet-Vallet, V.; Encinas, S.; Miranda, M. A. Excited State Enantiodifferentiating Interactions between a Chiral Benzophenone Derivative and Nucleosides. *J. Am. Chem. Soc.* **2005**, *127* (37), 12774–12775. <https://doi.org/10.1021/ja053518h>.

(240) Encinas, S.; Belmadoui, N.; Climent, M. J.; Gil, S.; Miranda, M. A. Photosensitization of Thymine Nucleobase by Benzophenone Derivatives as Models for Photoinduced DNA Damage: Paterno–Büchi vs Energy and Electron Transfer Processes. *Chem. Res. Toxicol.* **2004**, *17* (7), 857–862. <https://doi.org/10.1021/tx034249g>.

(241) Singh, V.; Fedeles, B. I.; Essigmann, J. M. Role of Tautomerism in RNA Biochemistry. *RNA* **2015**, *21* (1), 1–13. <https://doi.org/10.1261/rna.048371.114>.

(242) Barjat, H.; Morris, G. A.; Smart, S.; Swanson, A. G.; Williams, S. C. R. High-Resolution Diffusion-Ordered 2D Spectroscopy (HR-DOSY) - A New Tool for the Analysis of Complex Mixtures. *Journal of Magnetic Resonance, Series B* **1995**, *108* (2), 170–172. <https://doi.org/10.1006/jmrb.1995.1118>.

(243) Ghosh, A.; Trajkovski, M.; Teulade-Fichou, M.-P.; Gabelica, V.; Plavec, J. Phen-DC3 Induces Refolding of Human Telomeric DNA into a Chair-Type Antiparallel G-Quadruplex through Ligand Intercalation. *Angewandte Chemie International Edition* **2022**, *61* (40), e202207384. <https://doi.org/10.1002/anie.202207384>.

(244) Ourliac-Garnier, I.; Elizondo-Riojas, M.-A.; Redon, S.; Farrell, N. P.; Bombard, S. Cross-Links of Quadruplex Structures from Human Telomeric DNA by Dinuclear Platinum Complexes Show the Flexibility of Both Structures. *Biochemistry* **2005**, *44* (31), 10620–10634. <https://doi.org/10.1021/bi050144w>.

(245) Satz, A. L.; Cai, J.; Chen, Y.; Goodnow, R.; Gruber, F.; Kowalczyk, A.; Petersen, A.; Naderi-Oboodi, G.; Orzechowski, L.; Strebel, Q. DNA Compatible Multistep Synthesis and Applications to DNA Encoded Libraries. *Bioconjugate Chem.* **2015**, *26* (8), 1623–1632. <https://doi.org/10.1021/acs.bioconjchem.5b00239>.

- (246) Percec, V.; Leowanawat, P.; Sun, H.-J.; Kulikov, O.; Nusbaum, C. D.; Tran, T. M.; Bertin, A.; Wilson, D. A.; Peterca, M.; Zhang, S.; Kamat, N. P.; Vargo, K.; Mook, D.; Johnston, E. D.; Hammer, D. A.; Pochan, D. J.; Chen, Y.; Chabre, Y. M.; Shiao, T. C.; Bergeron-Brlek, M.; André, S.; Roy, R.; Gabius, H.-J.; Heiney, P. A. Modular Synthesis of Amphiphilic Janus Glycodendrimers and Their Self-Assembly into Glycodendrimersomes and Other Complex Architectures with Bioactivity to Biomedically Relevant Lectins. *J. Am. Chem. Soc.* **2013**, *135* (24), 9055–9077. <https://doi.org/10.1021/ja403323y>.
- (247) Cagnoni, A. J.; Varela, O.; Gouin, S. G.; Kovensky, J.; Uhrig, M. L. Synthesis of Multivalent Glycoclusters from 1-Thio- $\beta$ -d-Galactose and Their Inhibitory Activity against the  $\beta$ -Galactosidase from *E. Coli*. *J. Org. Chem.* **2011**, *76* (9), 3064–3077. <https://doi.org/10.1021/jo102421e>.