



HAL
open science

Développement d'une chimie combinatoire encodée pour l'identification de ligands spécifiques d'ADN G-quadruplex

Pierre Rieber

► **To cite this version:**

Pierre Rieber. Développement d'une chimie combinatoire encodée pour l'identification de ligands spécifiques d'ADN G-quadruplex. Biochimie [q-bio.BM]. Université Grenoble Alpes [2020-..], 2022. Français. NNT : 2022GRALV016 . tel-03774521

HAL Id: tel-03774521

<https://theses.hal.science/tel-03774521v1>

Submitted on 11 Sep 2022

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Chimie Biologie**

Arrêté ministériel : 25 Mai 2016

Présentée par

PIERRE RIEBER

Thèse dirigée par **Thomas LAVERGNE**, Université
Grenoble Alpes

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

**Développement d'une chimie combinatoire
encodée pour l'identification de ligands
spécifiques d'ADN G-quadruplex.**

**Development of combinatorial encoded
chemistry for the identification of specific
ligands of G-quadruplex DNA structure.**

Thèse soutenue publiquement le **10 Mars 2022**,
devant le jury composé de :

Monsieur Thomas LAVERGNE

CHARGE DE RECHERCHE HDR, CNRS délégation Alpes,
Directeur de thèse

Monsieur Michael SMIETANA

PROFESSEUR DES UNIVERSITES, Université de Montpellier,
Rapporteur

Monsieur David MONCHAUD

DIRECTEUR DE RECHERCHE, CNRS délégation Centre-Est,
Rapporteur

Madame Maria DUCA

CHARGE DE RECHERCHE HDR, CNRS délégation Côte d'Azur,
Examinatrice

Monsieur Anton GRANZHAN

CHARGE DE RECHERCHE HDR, CNRS délégation Paris Centre,
Examineur

Monsieur Eric PEYRIN

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes,
Président



Remerciements

Je tiens tout d'abord à remercier les membres du jury, qui ont accepté de rapporter et/ou examiner ces travaux de thèse, à savoir :

- Dr. **David MONCHAUD**, Directeur de Recherche à l'Institut de Chimie Moléculaire de l'Université de Bourgogne.
- Dr. **Michael SMIETANA**, Professeur à l'Institut des Biomolécules Max Mousseron de l'Université de Montpellier.
- Dr. **Maria DUCA**, Chargée de Recherche à l'Institut de Chimie de Nice de l'Université de Côte d'Azur.
- Dr. **Isabelle BAUSSANNE**, Chargée de Recherche au Département de Pharma-cochimie Moléculaire de l'Université Grenoble-Alpes.
- Dr. **Anton GRANZHAN**, Chargé de Recherche à l'Institut Curie.
- Dr. **Eric PEYRIN**, Professeur au Département de Pharmacochimie Moléculaire de l'Université Grenoble-Alpes.

Je tiens ensuite à remercier chaleureusement mon directeur de thèse, le Dr. **Thomas LAVERGNE**, pour sa disponibilité, son aide et son soutien pendant l'élaboration de mes travaux de thèse et également son implication aux tâches de relectures et corrections, notamment de mon manuscrit. Thomas, merci pour la franchise et la sincérité de nos échanges.

Ce travail n'aurait pas été possible sans les soutiens du ministère de l'Enseignement supérieur, de la recherche et de l'innovation ainsi que du laboratoire d'excellence Arcane de l'Université Grenoble-Alpes. Ces allocations de recherches et aides financières m'ont permis de me consacrer pleinement à mes travaux de recherches et leurs diffusions.

Dans le cadre du projet soutenu par la LABEX Arcane, je tiens particulièrement à remercier **Isabelle BAUSSANNE**, **Camille VAN WESEMAEL**, **Rémy LARTIA** et **Martine DEMEUNYNCK** pour leurs aides précieuses en synthèse, leurs conseils et leurs disponibilités.

Je remercie également l'ensemble de l'équipe I2BM pour leurs accueils, le Dr. **Didier BOTURYN** pour nos échanges et nos moments sportifs, **Régine Rozand** pour sa disponibilité et son efficacité administrative et **Jamila Burlet** pour nos discussions bienveillantes.

Au cours de ces 40 mois passés à Grenoble, j'ai eu l'occasion de croiser le chemin de nombreux « non-permanents » : des post-doctorants (Eugénie L., Silvia, A., Yang, S. Romain, N.), des doctorants (Hugo, S. Rym, K. Sean, M. Jordan, C. Yoann, C. Jonny, C. Biagio, T. Alexandre, D. Benjamin, L.), des ingénieurs d'études (Claire, T. Coelio V.) et des stagiaires (Alicia, H. Samantha, N., Mattéo, G. Rémy. G). Je souhaite tous vous remercier pour les bons moments que nous avons partagés. Bonne réussite et bonne continuation à tous !

Je veux également remercier mes amis de toujours, **Antonin, W.** et **Thomas, P.** de m'avoir épaulé à distance ces dernières années. Les opportunités professionnelles nous ont rapprochés, je ne manquerai pas de rattraper avec vous le temps, et surtout les Picons, perdu.

À vous avec qui j'ai passé plus de temps que personne d'autre sur les bancs de la fac. À vous avec qui j'ai partagé mes années universitaires à former des binômes de TP (parce les trinômes n'étaient malheureusement pas réglementaires). À toi, **Philippe, S.**, qui me fait l'honneur d'une place particulière à ton mariage et à toi, **Brayan, R.**, avec qui les différences nous ont conduits à une amitié immuable, je voulais vous remercier personnellement pour cette amitié que je souhaite pérenne.

Je souhaite également remercier les membres de ma famille et de ma belle-famille qui ne cesse de me soutenir dans mes projets personnels. Particulièrement, je tiens à exprimer ma reconnaissance profonde à mes parents, **Annie et Denis, R.**, et à ma sœur, **Fanny, S.**, pour leur soutien et leur amour au quotidien. Enfin, à ma grand-mère paternelle et mon grand-père maternel qui sont assurément fiers de moi de là où ils sont, je pense à vous à chaque instant.

Je souhaite terminer ces remerciements à celle qui partage ma vie depuis plus de 5 ans, à celle qui a tout quitté de son Espagne natale pour venir me rejoindre, à celle dont je suis le plus fière et dont je porte un amour impérissable. À toi **Desirée, J.**, merci de la confiance et l'amour que tu me portes.

*« Le succès ne dure pas, l'échec ne tue pas.
Ce qui compte, c'est d'avoir le courage de continuer. »*

W.Churchill

Table des matières

CHAPITRE 1 — INTRODUCTION.....	22
PARTIE 1 : L'ACIDE DESOXYRIBONUCLEIQUE. APPARIEMENTS NON-CANONIQUES ET STABILISATION D'EDIFICES NON-USUELS.....	22
<i>Sous-partie 1.1 : Histoire et découverte de l'ADN.....</i>	22
Section 1.2.1 : Structure primaire des acides désoxyribonucléiques, les nucléotides.....	22
Section 1.2.2 : La double hélice d'ADN.....	24
Section 1.2.3 : Structures non-usuelles de l'ADN.....	26
<i>Sous-partie 1.3 : G-quadruplex, structures et fonctions.....</i>	28
Section 1.3.1 : Structures des G-quadruplex.....	28
<i>Paragraphe 1.3.1.a) Formation de motifs G-quadruplex.....</i>	28
<i>Paragraphe 1.3.1.b) Stabilisation par cations métalliques.....</i>	29
Section 1.3.2 : Topologies des G-quadruplex.....	30
<i>Paragraphe 1.3.2.a) Les différentes topologies adoptées par les G-quadruplex.....</i>	30
<i>Paragraphe 1.3.2.b) Orientations des bases.....</i>	31
<i>Paragraphe 1.3.2.c) Nature des boucles et des sillons.....</i>	32
<i>Paragraphe 1.3.2.d) Structures atypiques de G-quadruplex.....</i>	34
Section 1.3.3 : Occurrence et fonctions des G-quadruplex.....	34
<i>Paragraphe 1.3.3.a) Région télomérique.....</i>	35
<i>Paragraphe 1.3.3.b) Régions promotrice de gènes.....</i>	37
<i>Paragraphe 1.3.3.c) Autres régions du génome.....</i>	38
<i>Paragraphe 1.3.3.d) Acide ribonucléique (ARN).....</i>	38
<i>Paragraphe 1.3.3.e) Séquences génomiques non humaines.....</i>	39
<i>Sous-partie 1.4 : Ligands de structures G-quadruplex.....</i>	40
Section 1.4.1 : Systèmes polycycliques/aromatiques fusionnés.....	41
Section 1.4.2 : Systèmes aromatiques non fusionnés.....	44
Section 1.4.3 : Systèmes non aromatiques.....	46
Section 1.4.4 : Systèmes macrocycliques.....	46
Section 1.4.5 : Systèmes organométalliques.....	47
Section 1.4.6 : Biomolécules et dérivés.....	47
<i>Paragraphe 1.4.6.a) Conjugués peptidiques, nucléotidiques et glycosidiques.....</i>	48
<i>Paragraphe 1.4.6.b) « Peptide-nucleic acids ».....</i>	48
<i>Sous-partie 1.5 : Méthodes combinatoires pour l'identification de ligands de structures G-quadruplex.....</i>	49
Section 1.5.1 : Criblage en parallèle et approche in silico.....	49
Section 1.5.2 : Chimie combinatoire.....	51
Section 1.5.3 : « Display technologies ».....	55
PARTIE 2 : DE LA LIBRAIRIE ENCODEE DE POLYPEPTIDES AUX LIBRAIRIES ENCODEES PAR L'ADN.....	58
<i>Sous-partie 2.1 : Méthode alternative pour la formation de librairies encodées.....</i>	60
Section 2.1.1 : Librairie délivrant un pharmacophore unique (en. single-pharmacophore library).....	61
Paragraphe 2.1.1.a) ADN encodé (en. DNA encoded).....	61
Paragraphe 2.1.1.b) ADN matriciel (en. DNA templated).....	63
Section 2.1.2 : Librairie délivrant une paire de pharmacophore (en. dual pharmacophore library).....	67
Paragraphe 2.1.2.a) Librairie encodée par auto-assemblage de l'ADN (en. Encoded self-assembling chemical library- ESAC).....	67
Paragraphe 2.1.2.b) Librairie encodée par assemblage d'ADN et d'acide nucléiques peptidiques.....	69
Section 2.1.3 : Les chimies compatibles.....	69
<i>Sous-partie 2.2 : Sélection, déconvolution et validation.....</i>	69
Section 2.2.1 : Sélection sur cible par affinité.....	70
Section 2.2.2 : Amplification des code-barres ADN.....	73
Section 2.2.3 : Séquençage des code-barres ADN.....	76
Section 2.2.4 : Resynthèse des touches et études d'interactions.....	79
<i>Sous-partie 2.3 : De la touche au candidat médicament.....</i>	80
PARTIE 3 : PROJET DE THESE.....	82
CHAPITRE 2 — ASSEMBLAGE DE LIBRAIRIES COMBINATOIRES ENCODEES PAR L'ADN CIBLANT L'ADN G-QUADRUPLEX....	86
PARTIE 1 : CONCEPTION ET SYNTHÈSE DE LA PREMIÈRE GÉNÉRATION DE LIGANDS ENCODES PAR L'ADN.....	86
<i>Sous-partie 1.1 : Composition, conception et validation des buildings blocks de la DEL1.....</i>	90
<i>Sous-partie 1.2 : Assemblage de la DEL1.....</i>	95
Section 1.1.1 : DEL1-Cycle 1.....	95
Section 1.1.2 : DEL1-Cycle 2.....	99
Section 1.1.3 : DEL1-Cycle 3.....	100

Section 1.1.4 : DEL1-Cycle 4.	103
PARTIE 2 : CONCEPTION ET SYNTHÈSE DE LA DEUXIÈME GÉNÉRATION DE LIGANDS ENCODES PAR L'ADN.	105
Sous-partie 2.1 : Composition, conception et validation des building blocks de la DEL2.	105
Sous-partie 2.2 : Assemblage de la DEL2.	111
Section 1.2.1 : DEL2-Cycle 1.	111
Section 1.2.2 : DEL2-Cycle 2.	112
Section 1.2.3 : DEL2-Cycle 3.	114
Section 1.2.4 : DEL2-Cycle 4.	116
PARTIE 3 : CONCLUSION DES SYNTHÈSES DE LIBRAIRIES CHIMIQUES ENCODEES CIBLANT LES G-QUADRUPLEX.	117
CHAPITRE 3 — SELECTIONS ET ANALYSES BIOINFORMATIQUES DES LIBRAIRIES ENCODEES PAR L'ADN CIBLANT LES G-QUADRUPLEX.	121
PARTIE 1 : PROCESSUS DE SÉLECTION <i>IN VITRO</i> PAR AFFINITÉ.	121
Sous-partie 1.1 : DEL1 : première campagne de sélections.	125
Section 1.1.1 : Processus de sélection <i>in vitro</i>	125
Section 1.1.2 : Génération d'amplicons par PCR.	137
Sous-partie 1.2 : DEL1 : deuxième campagne de sélections.	141
Section 1.2.1 : Processus de sélection <i>in vitro</i>	141
Paragraphe 1.2.1.a) Nouvelles conditions appliquées au promoteur d'oncogène Myc.	143
Paragraphe 1.2.1.b) Expériences de compétitions.	146
Paragraphe 1.2.1.c) Sélection sur i-motif.	149
Section 1.2.2 : Génération d'amplicons par PCR.	152
Sous-partie 1.3 : DEL2 : campagne de sélections.	153
PARTIE 2 : ANALYSE DES SÉLECTIONS RÉALISÉES PAR SÉQUENÇAGE À HAUT-DEBIT.	155
Sous-partie 2.1 : Traitement bio-informatique de la première campagne de sélection.	155
Section 2.1.1 : Analyse de la librairie naïve.	156
Section 2.1.2 : Analyse des motifs après sélections.	159
Paragraphe 2.1.2.a) Myc versus contrôles.	161
Paragraphe 2.1.2.b) Myc versus les autres topologies de structures G-quadruplex.	167
Paragraphe 2.1.2.c) Variation d'enrichissement par applications de conditions stringentes (cible Myc).	172
Paragraphe 2.1.2.d) Myc versus Myc22.	175
Section 2.1.3 : Analyse des molécules uniques après sélections.	177
CHAPITRE 4 — SYNTHÈSE DES TOUCHES ET CONTRÔLES IDENTIFIÉS ET CARACTÉRISATIONS BIOPHYSIQUES.	188
PARTIE 1 : SYNTHÈSE DES COMPOSÉS PEPTIDO-ORGANIQUES.	188
Sous-partie 1.1 : Synthèse de séquences peptidiques.	190
Sous-partie 1.2 : Synthèse des composés peptido-organiques.	194
PARTIE 2 : ÉTUDES BIOPHYSIQUES : STABILISATION ET AFFINITÉ.	197
Sous-partie 2.1 : FRET-Melting.	197
Section 2.2.1 : Effet stabilisant de PDS, PDC3, 360A et leurs dérivés peptido-organiques.	201
Paragraphe 2.2.1.a) Stabilisation des composés vis-à-vis des proto-oncogènes Myc et Kit.	202
Paragraphe 2.2.1.b) Stabilisation des composés vis-à-vis des séquences télomériques.	202
Paragraphe 2.2.1.c) Stabilisation du contrôle <i>Hairpin</i> d'ADN.	203
Section 2.2.2 : Étude de l'effet stabilisant des ligands contrôles (peptides, tronqués, non-enrichis).	203
Paragraphe 2.2.2.a) Dérivés de la famille Pyridostatine.	204
Paragraphe 2.2.2.b) Dérivés de la famille Phénanthroline.	205
Paragraphe 2.2.2.c) Dérivés de la famille Pyridocarboxamide.	205
Sous-partie 2.2 : Titration de fluorescence.	207
Section 2.2.1 : Dérivés de la famille Pyridostatine.	210
Section 2.2.2 : Dérivés de la famille Phénanthroline.	211
Section 2.2.3 : Dérivés de la famille Pyridocarboxamide.	211
Sous-partie 2.3 : Biolayer interferometry (BLI).	213
Section 2.3.1 : Fonctionnalisation du biocapteur.	216
Section 2.3.2 : Analyse de l'interaction G-quadruplex/ligand.	216
Section 2.3.3 : Dissociation partielle du ligand.	216
Section 2.3.4 : Sensorgrammes des couples ligands/G-quadruplex.	217
Sous-partie 2.4 : Comparaison des données obtenues.	219
Sous-partie 2.5 : Modélisation moléculaire.	220
CHAPITRE 5 — CONCLUSION ET PERSPECTIVES.	225
CHAPITRE 6 — MATERIALS AND METHODS.	229
1. EXPERIMENTAL SECTION.	229

1.1 General information, material and instrumentation.....	229
1.1.1 Materials.....	229
1.1.2 Beads and Resins.....	229
1.1.3 Chemical compounds.....	229
1.1.4 Oligonucleotides and Enzymes.....	229
1.1.5 Mass Spectrometry and NMR.....	229
1.1.6 Ethanol Precipitation of DNA.....	230
1.1.7 Agarose gel electrophoresis.....	230
1.2 Synthesis of DEL-1 and DEL-2 building blocks.....	230
DEL1-[Cycle 2].....	230
DEL1-[Cycle 3].....	230
DEL2-[Cycle 2].....	231
DEL2-[Cycle 3].....	236
1.3 Preparation of the 190.440 G4-DEL-1 library.....	236
1.3.1 General synthetic sequence for G4-DEL-1 and barcoding scheme.....	236
1.3.2 General procedures for building block couplings in G4-DEL-1.....	237
1.3.3 Building Blocks Validation.....	239
1.3.4. Assembly of DEL-1:.....	240
Procedure for Cycle 1-DEL1.....	240
Procedure for Cycle 2-DEL1.....	243
Procedure for Cycle 3-DEL1.....	244
Procedure for Cycle 4-DEL1.....	246
1.4 Preparation of the 696.200 G4-DEL-2 library.....	248
1.4.1 General synthetic sequence for G4-DEL-2 and barcoding scheme.....	248
1.4.2 General procedures for building block couplings in G4-DEL-2.....	249
1.4.3 Building Blocks Validation.....	250
1.4.4 Assembly of DEL-2.....	251
Procedure for Cycle 1-DEL2.....	251
Procedure for Cycle 2-DEL2.....	253
Procedure for cycle 3-DEL2.....	254
Procedure for Cycle 4-DEL2.....	256
1.5 Affinity selection of G4-DEL-1 against streptavidin bound G4 structures and controls.....	257
1.5.1. Affinity selection targets.....	257
1.5.2. General procedure for affinity selections.....	258
1/ Target loading.....	258
2/ Beads capping.....	258
3/ Beads splitting and target folding.....	259
4/ Incubation, washes and elution.....	259
1.5.3. Post-selection PCR amplification of the DNA barcodes for illumina HTS sequencing.....	259
1.5.4. List and details of reported selection experiments.....	260
1.6 High Throughput Sequencing.....	262
1.7 Off-DNA re-synthesis of hits and controls.....	271
Synthesis of azido peptides 1, t1, 2, 3, t3, 4 and 5.....	271
Synthesis of compounds 1-PDP, t1-PDP, 2-PDC3, 3-360A-s, 3-360A-l, t3-360A-s, 4-PDP, 5-360A-s.....	276
1.8 Biophysical evaluation of hits and controls.....	282
1.8.1 Fluorescence melting experiments (FRET-melting).....	282
1.8.2 Fluorimetric titration.....	282
1.8.3 Bio Layer Interferometry (BLI).....	287
1.9 Computational Analysis.....	289
2. SUPPLEMENTARY TABLES.....	292
Chemical structure, validation and source of building blocks.....	292
Building blocks [Cycle 1 & 4] – DEL1 & DEL2.....	292
Building blocks [Cycle 2] – DEL1.....	302
Building blocks [Cycle 2] – DEL2.....	303
Building blocks [Cycle 3] – DEL1.....	303
Building blocks [Cycle 3] – DEL2.....	305
Sequence of DNA tags and corresponding building block ID.....	308
DNA Tags [Cycle 1 & 4] – DEL1 & DEL2.....	308
DNA Tags [Cycle 2] – DEL1.....	312
DNA Tags [Cycle 2] – DEL2.....	313
DNA Tags [Cycle 3] – DEL1.....	313
DNA Tags [Cycle 3] – DEL2.....	313

BIBLIOGRAPHIE	315
ANNEXES	330
1. PUBLICATION	330
2. COMMUNICATIONS	330
a) Posters	330
b) Communications orales	330
c) Evènements	331

Table des figures

Figure 1 : Structure schématique et tridimensionnelle d'une nucléobase (PDB id : 6RLS)	22
Figure 2 : Bases puriques et pyrimidiques de l'ADN et de l'ARN	23
Figure 3 : Formation de la liaison phosphodiester entre deux nucléobases	24
Figure 4 : Représentation des conformations possibles de A/ la liaison N-osidique (syn/anti) ou B/ du sucre (C2'/C3'— endo) des nucléobases	24
Figure 5 : Appariements Watson-Crick des bases de l'ADN et ARN	25
Figure 6 : Conformations de la double hélice d'ADN. ADN-A (PDB id : 4OKL). ADN-B (PDB id : 6RLS). ADN-Z (PDB id : 1R4D) ⁹ . Taille des sillons annotés suivant la nature de l'ADN considéré	25
Figure 7 : Exemples de structures secondaires de l'ADN. A/ TWJ (PDB id : 1EKW). B/ Jonction de Holiday (PDB id : 1M6G). C/ Tige-boucle (PDB id : 1MFJ)	27
Figure 8 : Exemples de structures polycaténares de l'ADN. A/ ADN-Triplex (PDB id : 1D3X). B/ i-motif (PDB id : 1I9K)	27
Figure 9 : représentation de la structure des guanines, la formation primaire du quartet de guanines ainsi que l'assemblage de la structure secondaire des G-quadruplex	29
Figure 10 : Positionnement des cations monovalents de sodium (Na ⁺) et de potassium (K ⁺) au sein de G-quartets ou d'un édifice G-quadruplex. Dessin réalisé par D. Jiménez	30
Figure 11 : Représentation exemplifiée d'un G-quadruplex tétramoléculaire (a), bimoléculaire (b) et uni- ou intramoléculaire (c). Dessin réalisé par D. Jiménez	30
Figure 12 : Représentation schématiques des différentes topologies que peuvent adopter les G-quadruplex. De gauche à droite sont représentées les topologies parallèles, antiparallèles et hybrides de structures tétramoléculaires (a), bimoléculaires (b) et unimoléculaires (c). Dessin réalisé par D. Jiménez	31
Figure 13 : Conformation de la liaison N-osidique des bases (a) en fonction de la topologie adoptée par les structures G-quadruplex (b). 32	
Figure 14 : Illustration des différentes conformations des guanines possibles au sein d'un G-quadruplex. Notation de la taille des sillons : m-moyen, l-large, e-étroit	33
Figure 15 : Illustration des différentes boucles identifiables au sein de séquences G-quadruplex. Dessin réalisé par D. Jiménez	33
Figure 16 : A/ Structure G-quadruplex à site vacant. B/ Édifice G-quadruplex à renflement formant un ADN duplex. Schémas tirés des publications respectives (K.-B. Wang ²⁷ ; T. Q. Ngoc Nguyen ²⁹)	34
Figure 17 : Représentation schématique de l'extrémité télomérique riche en guanines des chromosomes humains. Dessin réalisé par D. Jiménez	36
Figure 18 : Ligands de structures G-quadruplex tirés de la review de T.Santos ¹¹⁴	40
Figure 19 : Illustration de certaines classes de ligands ciblant les tétraplexes de guanines	43
Figure 20 : Criblage conventionnel à haut-débit. Découverte du ligand J20 ciblant le G-quadruplex VIH-1 au niveau du promoteur LTR-III ¹⁷¹	50
Figure 21 : Criblage in silico. Découverte des ligands Indolhydrazide (gauche) et TSIZ01 (droite) ciblant le G-quadruplex de la séquence télomérique humaine F21T	51
Figure 22 : Chimie combinatoire / stratégie One-bead-one-compound. Découverte de peptides (modifiés) ciblant la structure télomérique G-quadruplex — hTelo	52
Figure 23 : Chimie combinatoire / stratégie Fragment-based. Synthèse de séquences pluri-aromatiques ciblant le promoteur LTR-III du virus VIH-1	52
Figure 24 : Chimie combinatoire / Méthode par librairie combinatoire dynamique. Synthèse de macrocycles polyoxazoles ciblant le promoteur d'oncogène Myc (Myc22)	54
Figure 25 : Chimie combinatoire / Méthode de Kinetic target-directed synthesis. Synthèse d'un dérivé glycosidique de la Pyridostatine (PDS) afin de cibler l'ADN G-quadruplex télomérique (hTelo)	55
Figure 26 : Représentation des bibliothèques combinatoires encodées A et B ciblant les structures G-quadruplex. Figure issue des travaux de Litovchick et al. ¹⁸⁸	57
Figure 27 : Chimie combinatoire / Technologie display de librairie encodée par l'ADN. Synthèse d'un dérivé linéaire trifonctionnalisé ainsi qu'un composé imidazopyrimidine bifonctionnel ciblant le proto-oncogène Myc (Myc 1234)	58
Figure 28 : Représentation schématique d'une librairie obtenue par phage display et une librairie obtenue par chimie combinatoire encodée par l'ADN. A gauche : interaction d'un phage filamentueux, présentant un anticorps à sa surface, avec une cible protéique. A droite : interaction d'une molécule organique encodée par un fragment d'ADN sur une cible d'intérêt	59
Figure 29 : Représentation schématique des bibliothèques encodées à single-pharmacophore (gauche) ou dual-pharmacophore (droite)	61

Figure 30 : Illustration d'une librairie « recorded » où A/ le fragment chimique est directement lié à l'oligonucléotide B/ lié à l'oligonucléotide par l'intermédiaire d'un espaceur bifonctionnel.	62
Figure 31 : Illustration d'une librairie « recorded » encodée par de l'ADN simple brin via l'utilisation de la polymérisation de Klenow.	62
Figure 32 : Illustration d'une librairie « templated » où A/ l'enrichissement de la librairie se fait, après appariement de code-barres ADN complémentaires, par réaction chimique dû à la proximité des fragments.	63
Figure 33 : Exemple d'une librairie « templated ». Enrichissement de la librairie permise par proximité spatiale de building blocks contrôlés par appariements de code-barres ADN complémentaires.	64
Figure 34 : Illustration d'une librairie « templated » à séquence poly-désoxyinosines. Encodage et fonctionnalisation de la librairie par appariement, ligation enzymatique et couplage.	64
Figure 35 : Exemple d'une librairie « templated » à séquence poly-désoxyinosines présentant un monophosphate en 5' et une séquence connue amino-terminale (vert) en 3'. La fonction chimique pré-encodée par le code-barre ADN (rouge, A) est liée de manière enzymatique à la séquence poly-I en 5'. L'utilisation d'un espaceur photo-clivable présentant une séquence complémentaire à la séquence Poly-I en 3' permet l'introduction du fragment par couplage peptidique. Après clivage du linker la molécule obtenue est prête pour le deuxième tour de la librairie encodée.	65
Figure 36 : Illustration d'une librairie « templated » obtenue par YoctoReactor™.	66
Figure 37 : Exemple d'une librairie « templated » obtenue par technologie YoctoReactor™. L'appariement de séquence double brin d'ADN pré-fonctionnalisé par un building block permet le couplage de ceux-ci dû à leur proximité spatiale. Le clivage final des linkers permettent d'obtenir un composé exprimé à la surface de l'ADN.	66
Figure 38 : Illustration d'une librairie obtenue par DNA routing.	67
Figure 39 : Représentation schématique de la conception de librairie ESAC à l'aide de deux séquences oligonucléotidiques chimiquement encodées.	67
Figure 40 : Méthode d'autoassemblage de librairies encodées.	68
Figure 41 : Assemblage de librairie d'oligonucléotidique chimiquement encodée par polymérisation de Klenow.	68
Figure 42 : Formation de librairie combinatoire encodée via l'utilisation d'acides nucléiques d'origines peptidiques.	69
Figure 43 : Méthodes de DEL dont les expériences in vitro de sélections peuvent être effectuées en solution — IDPCR (C.), IDUP, DPAL (D.). Figure issue de la review de A. Gironda-Martinez ²²⁸	71
Figure 44 : Principe général de l'amplification de fragment d'ADN par la méthode de Polymerase Chain Reaction.	74
Figure 45 : Phases discernées lors de l'amplification d'ADN par Polymerase Chain Reaction.	75
Figure 46 : Visualisation de la progression de l'amplification à l'aide du marqueur fluorescent, le SYBR Green. Image tirée du site de l'institut Agro de Montpellier.	76
Figure 47 : Schéma explicatif de la technologie de séquençage Illumina. Image tirée du livre « Next generation Sequencing – Advances, Applications and Challenges » - Chapitre ²²⁵¹	78
Figure 48 : Méthodes de resynthèses de composés issus de DEL en fonction de la stratégie Single — ou Dual-pharmacophore ¹⁹⁵	79
Figure 49 : Mode de construction de la DEL1 – Formation de 190 440 composés uniques.	83
Figure 50 : Mode de construction de la DEL2 – Formation de 696 200 composés uniques.	83
Figure 51 : Structure de l'espaceur bidirectionnel, l'Headpiece.	86
Figure 52 : Stratégie d'encodage appliquée à la DEL1. Chaque segment d'ADN (T1 à T4) va être enzymatiquement lié à l'Headpiece de départ par l'intermédiaire des leurs extrémités non-cohésives en 3'. Les extrémités 5', elles, exhibent un groupement phosphate libre. La région X des code-barres T1 à T4 représente des oligonucléotides non définis qui coderont pour les entités chimiques utilisées dans chaque cycle de synthèse. La séquence terminatrice (CP) contient une plus longue séquence parmi laquelle on retrouve les variables identifiables par le symbole X souligné (X).	88
Figure 53 : Assemblage d'un code-barre ADN complet après quatre cycles de ligation (T1 à T4) et ajout du closing primer (CP). Augmentation croissante des masses moléculaires de la librairie après ligation enzymatique. Gel à 5% d'agarose UltraPure (disponible commercialement) précoulé avec du GelRed™. Migration à 90V pendant 2 h puis exposition à une lampe UV (GelDoc) pour caractérisation.	89
Figure 54 : Représentation schématique des zones distinctes composant les structures G-quadruplex. A- Tétrades de guanines ou G-quartets. B-Boucles et C-Sillons des structures G-quadruplex.	91
Figure 55 : Structure des ligands de la DEL1 par association de quatre buildings blocks.	91
Figure 56 : Structure générale des séquences pentapeptidiques DEL1 — [Cycle 1/4] retenues.	91
Figure 57 : Structure corps trifonctionnels DEL1 — [Cycle 2].	93
Figure 58 : Structure de l'HP-Ext-GAWRG utilisé pour la validation des 9 cœurs trifonctionnels du [Cycle 2].	94
Figure 59 : Représentation exhaustive des fragments hétéroatomatiques alcynes utilisés dans le cadre de la librairie DEL1.	94
Figure 60 : Synthèse de l'HP-Ext-GAWRG-Ke utilisé pour la validation des 10 ligands-alcynes de structures G-quadruplex [Cycle 3].	95
Figure 61 : Augmentation croissante des masses moléculaires de la librairie après ligation enzymatique des code-barres ADN T1 sur l'HP. Exemple de ligation partiellement complète au centre du gel. Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.	96
Figure 62 : Chromatogramme UPLC du mélange DEL1 — [Cycle 1] avant (gris) et après (bleu) purification suivant la méthode Cap&Catch. Disparition significative de la fraction γ (H ₂ N-HP-T1 non fonctionnalisés) après purification.	98
Figure 63 : Chromatogramme UPLC de DEL1 — [Cycle 1] (46 composés en mélange) après purification et déprotection des amines primaires de leur groupement protecteur Fmoc.	98
Figure 64 : Chromatogramme UPLC du mélange DEL1 — [Cycle 2]. Présence d'ADN code-barres T1, T2 (α) et T1-T2 (β).	99
Figure 65 : Chromatogramme UPLC de DEL1 — [Cycle 2] (414 composés en mélange) après purification HPLC et déprotection du groupement protecteur Fmoc.	100

Figure 66 : Gel retard (EMSA) effectué après incorporation d'un alcyne-PEG ₄ -biotine sur i) l'Headpiece-NH ₂ et ii) la DEL1 — [Cycle 3]. L'incubation en présence de streptavidine montre un retard de gel sur le contrôle Headpiece-biotine (ligne 3). Aucun retard observé sur l'échantillon DEL-1 — [Cycle 3] en présence de streptavidine (ligne 5). Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.	102
Figure 67 : Filtration sur membranes Ø 30 KDa. Séparation de la DEL1 — [Cycle 3] (MW _{moy} =29.487,99 g/mol) des code-barres ADN T3 résiduels.	103
Figure 68 : Gel retard (EMSA) effectué après incorporation d'une biotine-NHS sur i) l'Headpiece-NH ₂ et ii) la DEL1 — [Cycle 4]. L'incubation en présence de streptavidine montre un retard de gel sur le contrôle Headpiece-biotine (ligne 3). Aucun retard observé sur l'échantillon DEL-1 — [Cycle 4] en présence de streptavidine (ligne 5). Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.	104
Figure 69 : Gel d'électrophorèse illustrant les 6 tailles d'ADN, de l'Headpiece à la séquence encodée complète (T1 à T4) liée au closing primer (CP). Gel 5% (Agarose UltraPure), précoulé au GelRed™ ; migration à 90V/2 h.	105
Figure 70 : Structure des ligands de la DEL2 par association de quatre buildings blocks.	106
Figure 71 : DEL2 — [Cycle 1] — Structure des 59 motifs pentapeptidiques.	107
Figure 72 : Série de ligands trifonctionnels (— NH-Fmoc, — COOH, — Alcyne) synthétisés au laboratoire (A/ B/ et E/) et en collaboration (C/ et D/).	108
Figure 73 : Synthèses des indolizine-s et -l à partir de trois synthons (A, B et C).	109
Figure 74 : DEL2 — [Cycle 3] — Structures générales et liste non-exhaustive des azotures (protéogéniques, non-protéogéniques) utilisés dans le cadre du Cycle 3 de la DEL2.	110
Figure 75 : DEL2 — [Cycle 4] — Structure des 59 motifs pentapeptidiques utilisées lors du Cycle 4 de la DEL2.	110
Figure 76 : Chromatogramme UPLC du mélange DEL2 — [Cycle 1] après déprotection du groupement protecteur Fmoc.	111
Figure 77 : Gel retard (EMSA) effectué après Cap&Catch ²⁸⁷ et réalisé indépendamment sur les cœurs trifonctionnels de la DEL2-[Cycle 2]. Quel que soit le cœur considéré, un retard est observable (lignes 2, 4, 6 et 8). Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.	113
Figure 78 : Gel retard (EMSA) effectué après deuxième expérience de catch (exemple du cœur trifonctionnel trioxazole). Aucun effet retard observé après réalisation de 2 catch successifs (catch 2). Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.	113
Figure 79 : Structure générale des ligands de la DEL2 en fin de [Cycle 2]. DEL2 — [Cycle 2] = 236 composés.	114
Figure 80 : Gel retard (EMSA) effectué avant biotinylation et après Cap&Catch sur la librairie en mélange DEL2 — [Cycle 3]. L'effet retard observé sur une partie de la DEL2 (ligne 5 ; « apr biot »). Disparition de la bande après Catch en présence de streptavidine. Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.	115
Figure 81 : Structure générale des ligands de la DEL2 en fin de [Cycle 3]. DEL2 — [Cycle 3] = 11 800 composés.	115
Figure 82 : Structure générale des ligands de la DEL2 en fin de [Cycle 4]. DEL2 — [Cycle 4] = 696 200 composés.	116
Figure 83 : Encodage (Headpiece, code-barres T1 à T4 et Closing primer) et BB utilisé afin de former la DEL2 composée de 696 200 composés en mélange.	117
Figure 84 : Représentation schématique du système de billes magnétiques (Dynabeads™ MyOne™ T1) streptavidines. Fonctionnalisation des billes par des séquences oligonucléotidiques biotinylées et coiffage des sites de la streptavidine par de la biotine libre en solution. L'ADN génomique est également rajouté au système pour prévenir des interactions non-spécifiques.	122
Figure 85 : Expérience de sélection in vitro par affinité réalisée de manière séquentielle : 1) immobilisation des cibles, 2) incubation de la DEL, 3) lavages et applications de stringences et 4) élution des composés les plus affins pour la cible. Dans certains cas, une dernière étape de double — ou contre-sélection 5) peut être effectuée.	125
Figure 86 : Courbes d'amplifications obtenues par qPCR. Représentation de l'intensité de fluorescence en fonction du nombre de cycles d'amplifications. Analyse effectuée pour chaque concentration (10 ¹² à 10 ⁵ copies) de librairie DEL1 non sélectionnée ainsi que pour les contrôles NTC et buffer.	127
Figure 87 : Représentation générale de la séquence complète des code-barres ADN des membres de la DEL1 : de l'Headpiece au Closing primer-1. Hybridation des amorces Illumina au niveau du brin sens du CP1 et du brin anti-sens de l'HP.	138
Figure 88 : A/ Amplification des 29 conditions de sélections sur 27 cycles. Classement des sélections en fonction du nombre de cycles nécessaires afin d'atteindre le plateau : B/ 20 cycles, C/ 23 cycles et D/ 27 cycles.	139
Figure 89 : Quantification massique d'ADN obtenu pour chaque condition de sélection après amplification par qPCR. Quantification réalisée sur gel d'agarose (UltraPure) à 5%. Migration à 90V/2 h.	140
Figure 90 : Purification sur gel d'Agarose 5% (Ultrapure) du mélange équimassique des deux lots de sélections (lot 1 : Sel1 à 28 ; lot 2 : Sel29).	141
Figure 91 : Fonctionnalisation d'oligonucléotides par une biotine en 3' via l'espaceur triéthylène glycol (/3BioTEG/).	145
Figure 92 : Fonctionnalisation d'oligonucléotides par une biotine en 5' via l'espaceur hexa-éthylène glycol (/5Biosg//iSp18/).	146
Figure 93 : A/ Représentation schématique des édifices G-quadruplex (rouge) et i-motif (bleu) — Image tirée de la publication de C.E Kaiser ³⁰² . B/ Formation des structures i-motifs possible en condition acide (dépendant de la séquence considérée).	149
Figure 94 : Etude de la dissociation thermique des structures i-motif par dichroïsme circulaire en fonction de différentes valeurs de pH (5,5 à 7,4).	150
Figure 95 : Quantification massique d'ADN obtenu pour chaque condition de sélection après amplification par qPCR. Quantification réalisée sur gel d'agarose (UltraPure) à 5%. Migration à 90V/2 h.	153
Figure 96 : Migration des quantités équimassiques de sélections (90V/2hrs/5% Agarose UltraPure) suivie de la séparation physique des bandes d'intérêts.	153
Figure 97 : Structure générale du brin sens des code-barres ADN de la DEL1.	156

Figure 98 : Séquence « HptoCP » comportant l'Headpiece et le Closing primer-1 ainsi que les code-barres issus de la DEL1 : T1-2, T2-1, T3-1 et T4-2.....	157
Figure 99 : Représentation exhaustive des 10 ligands formant les Building Blocks de la DEL1 — [Cycle 3].	164
Figure 100 : Top 100 des composés sélectionnés contre la cible Myc — WASH1 (25°C). Représentation i) de l'enrichissement des motifs uniques et ii) leurs corrélations pour chaque molécule du top 100.	178
Figure 101 : Composition des motifs les plus représentés au sein du top 100 contre la cible Myc — WASH1 (25°C).	178
Figure 102 : Gauche. Top 100 des composés sélectionnés contre la cible Myc — WASH4 (70°C). Représentation i) de l'enrichissement des motifs uniques et ii) leurs corrélations pour chaque molécule du top 100. Droite. Composition des motifs les plus représentés au sein du top 100 contre la cible Myc — WASH4 (70°C).....	179
Figure 103 : Gauche. Top 100 des composés sélectionnés après double sélection contre la cible Myc — W4/W4 (70°C). Représentation i) de l'enrichissement des motifs uniques et ii) leurs corrélations pour chaque molécule du top 100. Droite. Composition des motifs les plus représentés au sein du top 100 après double sélection contre Myc.	180
Figure 104 : Top 100 des composés sélectionnés contre la cible Streptavidine (gauche) et la cible hp1 (droite) — WASH1 (25°C). Représentation i) de l'enrichissement des motifs uniques et ii) leurs corrélations pour chaque molécule du top 100. En s'intéressant aux molécules uniques interagissant avec la cible hp1 (Figure 104, droite) il a été possible de constater que tous les motifs des [Cycles 2 et 3] de la DEL1 étaient représentés. Toutefois, l'intensité relative de ces motifs varie puisque parmi les cœurs trifonctionnels les motifs majoritairement enrichis sont les dérivés Lysines (K, KG), la proline (P) et le Dap (D). De manière similaire, ce sont les motifs hétéroaromatiques à système fusionné (PDP, 360A (— s/— l) ainsi que la Furopyridazinone et le PDC3 qui ont été enrichis. L'enrichissement des motifs pentapeptidiques a également présenté de grandes variations. Néanmoins, ce sont les motifs GFRWG et GLRLG qui gouvernent l'enrichissement des motifs du [Cycle 4] alors que GFFRG, GRWFG et GAWAG n'ont exhibé que des enrichissements faibles au [Cycle 1].	181
Figure 105 : Top 100 des composés sélectionnés après sélection stringente contre la cible Myc22 — WASH4 (70°C) — ainsi que la composition des motifs les plus représentés (gauche). Top 100 contre la cible 22CTA — WASH2 (50°C) — et que la composition des motifs les plus représentés (droite). Représentation i) de l'enrichissement des motifs uniques et ii) leurs corrélations pour chaque molécule du top 100.	182
Figure 106 : Composés identifiés pour leurs enrichissements individuels à l'égard de différentes sélections (dont contrôles). A/ Dérivé Pyridostatine 1-PDS. B/ Dérivé Pyridocarboxamide 3-360A-s. C/ Dérivé Phénanthroline 2-PDC3.	184
Figure 107 : Composition organo-peptidique des dérivés Pyridostatine 1-PDP et Pyridocarboxamide 3-360A-s.	188
Figure 108 : Résine Rink-Amide Polyéthylène glycol (PEG) 4-mehtylbenzhydrylamine (MBHA) utilisée lors de la synthèse supportée sur phase solide des différentes séquences peptidiques.	190
Figure 109 : Structure de ligands parents Pyridostatine, Pyridocarboxamide et Phénanthroline issus de la littérature.	191
Figure 110 : Structures développées des 5 séquences peptidiques et de leurs tronqués (t1 et t3).	194
Figure 111 : Représentation exhaustive de tous les composés synthétisés. Dans le cas des familles Pyridostatine et Pyridocarboxamide : synthèse du dérivé enrichi 1-PDP et 3-360A-s respectivement ainsi que leurs tronqués t1-PDP et t3-360A-s. Synthèses des deux dérivés non-enrichis 4-PDP et 5-360A-s. Dans la famille Phénanthroline : synthèse d'un seul composé 2-PDC3.	196
Figure 112 : Phénomène de FRET illustré au sein d'une séquence G-quadruplex doublement marquée par une Fluorescéine (FAM) et une Rhodamine (TAMRA). Dissipation de l'effet FRET par dénaturation thermique.	198
Figure 113 : Expérience de titration de fluorescence d'une séquence G-quadruplex. Phénomène d'extinction de fluorescence d'une Cyanine-5 par ajout croissante d'une concentration de ligands. Figure adaptée de D.D. Le ²⁵	208
Figure 114 : Biolayer Interferometry (BLI). Utilisation d'un biocapteur fonctionnalisé par la cible d'intérêt de manière covalente ou non. Image tirée de la thèse du Dr. Eugénie Laigre, avec approbation de l'auteure.	213
Figure 115 : BLI. Mesure interférométriques afin de caractériser l'association et la dissociation d'une interaction G-quadruplex/ligand. Figure de ForteBio adaptée par le Centre d'Interactions Macromoléculaires (CMI) d'Harvard.	214
Figure 116 : BLI. Expérience de BLI caractérisée par une étape d'immobilisation, d'interaction et analyses et de régénération. Figure de ForteBio adaptée par le Centre d'Interactions Macromoléculaires (CMI) d'Harvard.....	215
Figure 117 : A/ Modélisation moléculaire du ligand 3-360A-s en interaction avec la cible Myc22. B/ Détermination des interactions de la chaîne peptidique avec le tétraplexe de guanines. C/ et D/ Caractérisation de contacts polaires, hydrophobes et ioniques.	222
Figure 118 : Comparaison du déploiement tridimensionnel des dérivés Pyridocarboxamides (3-360A-s, 3-360A-l et 5-360A-s) par modélisation moléculaire sur la cible Myc22.	223

Materials & methods section

Figure S1 : Barcoding scheme for DEL-1. The headpiece (HP) containing the forward primer sequence is in black, DNA Tag 1 in blue, DNA Tag 2 in red, DNA Tag 3 in green, DNA Tag 4 in light blue, Closing Primer in gray, ligation overhangs are underlined, the library barcode is boxed.	237
Figure S2. Representative starting material for cycle 2 building block validation	239
Figure S3 : Representative starting material for cycle 3 building block validation	240
Figure S4 : UPLC chromatogram of crude N-Fmoc S1 library	240
Figure S5 : UPLC chromatogram of purified N-Fmoc S1 library	241
Figure S6 : UPLC chromatogram and ESI mass spectrum of purified S1 library.....	242
Figure S7 : UPLC chromatogram of crude N-Fmoc S2 library.....	243
Figure S8 : UPLC chromatogram and ESI mass spectrum of purified S2 library.....	244
Figure S9 : UPLC chromatogram of S3 library.....	245
Figure S10 : S3 coupling efficiencies assessment via streptavidin electrophoretic mobility shift assay.....	245

Figure S11 : UPLC chromatogram and ESI mass spectrum of purified S3 library.....	246
Figure S12 : UPLC chromatogram of crude N-Fmoc S4 library.....	247
Figure S13 : N-Fmoc S4 purity assessment via streptavidin electrophoretic mobility shift assay.	247
Figure S14 : Agarose gel electrophoresis analysis of cycle 1 to cycle 4 G4-DEL1 library.	248
Figure S15 : Barcoding scheme for DEL-2 . The headpiece (HP) containing the forward primer sequence is in black, DNA Tag 1 in blue, DNA Tag 2 in green, DNA Tag 3 in orange, DNA Tag 4 in light blue, Closing Primer in gray, ligation overhangs are underlined, the library barcode is boxed.	249
Figure S16 : Representative starting materials for cycle 3 building block validation	250
Figure S17 : UPLC chromatogram and ESI mass spectrum of purified S1 library.....	252
Figure S18 : UPLC chromatogram of purified N-Fmoc S2 library.	253
Figure S19 : UPLC chromatogram and ESI mass spectrum of purified S2 library.....	254
Figure S20 : UPLC chromatogram of crude S3 library.....	255
Figure S21 : UPLC chromatogram and ESI mass spectrum of purified S3 library.....	256
Figure S22 : UPLC chromatogram of purified N-Fmoc S4 library.	257
Figure S23 : CD analyses (1.67 μ M in 100 mM KCl, 0.01% Tween-20, 1X PBS pH 7.4) of biotin-oligonucleotide targets.	258
Figure S24 : Analysis of high-throughput sequencing results. Actual (dots) vs expected (vertical line) frequency for all library members obtained from naïve library sequencing. This plot shows the distribution of sequence frequencies within the 190 440 sequences of the naïve G4-DEL1 library (x axis) versus the abundance of each frequency in sequence counts (y axis). The blue vertical indicates the theoretical frequency for each sequence assuming equal abundance of each of the 190 400 sequences analyzed by 60 million sequence reads.	263
Figure S25 : Graphical representation of the correlation between the frequency of the 190 440 sequences in the naïve G4-DEL1 library (x axis) and their coefficient of variation (y axis). To determine the coefficient of variation for each sequence, 60 million of the reads were randomly distributed into clusters of 5 million reads and the ratio of standard deviation to the mean was calculated for each sequence from those clusters. Only sequences for which CV was below 60% (180 014 sequences) were considered for statistical analysis of the selections (below the blue horizontal). Only sequences for which CV was below 50% and frequency within the naïve library was above $5 \cdot 10^{-4}$ (175 242 sequences) were considered for the selection of enriched library members and controls to be resynthesized (within the green box).	264
Figure S26 : Schematic representation of the Enrichment (E) in % of each DEL1 building blocks after selections against Myc G4 and across various selection stringencies. Selection wash temperatures are indicated in selection labels (x axis), 2rnds signifies two successive rounds of selections against Myc G4 (Table S1).	265
Figure S27 : Schematic representation of the Enrichment (E) in % of each DEL1 building blocks after selections against wTel26 , 22CTA , Kit1 and Myc G4. Gentle (25°C) wash procedures were used in all selections.....	266
Figure S28 : Schematic representation of the Enrichment (E) in % of each DEL1 building blocks after selections against Myc and Myc22 G4. Harsh (70°C) wash procedures were used in both selections.	267
Figure S29 : Schematic representation of DEL-1 building blocks abundance and correlation within the 100 most enriched compounds identified after two successive rounds (2rnds) of selections against Myc G4 (A) and selection with 22CTA G4 (B). A mild wash procedure (50°C) was implemented in the selection displayed in B whereas a harsh wash procedure (70°C) was implemented in the selection displayed in A (Table S1). Variation of node size and color (gradient from green to red) represent the abundance of individual building blocks across selections. Variation of building block connecting line size and colors represent the abundance of building block correlations across selections. Created with Cytoscape.	268
Figure S30 : HPLC chromatogram of pure peptide 1	271
Figure S31 : UPLC chromatogram of pure peptide t1	272
Figure S32 : HPLC chromatogram of pure peptide 2	273
Figure S33 : HPLC chromatogram of pure peptide 3	273
Figure S34 : UPLC chromatogram of pure peptide t3	274
Figure S35 : UPLC chromatogram of pure peptide 4	275
Figure S36 : UPLC chromatogram of pure peptide 5	275
Figure S37 : UPLC chromatogram of pure 1-PDP	277
Figure S38 : UPLC chromatogram of pure t1-PDP	277
Figure S39 : UPLC chromatogram of pure 2-PDC3	278
Figure S40 : UPLC chromatogram of pure 3-360A-s	279
Figure S41 : UPLC chromatogram of pure 3-360A-l	280
Figure S42 : UPLC chromatogram of pure t3-360A-s	280
Figure S43 : UPLC chromatogram of pure 4-PDP	281
Figure S44 : UPLC chromatogram of pure 5-360A-s	282
Figure S45 : CD analyses (1.6 μ M in 50 mM Tris.HCl, 150 mM KCl, 0.05% v/v, Triton X-100) of 5'Cy5 oligonucleotide targets.	283
Figure S46 : Normalized of fluorescence intensity (λ_{ex} = 610nm, λ_{em} =675nm) upon titrations of ligands (final concentration: 1.8nM to 10 μ M) to 5'-Cy5-labelled oligonucleotides ([c]=2nM in 50mM Tris.HCl, 150mM KCl, 0.05 v/v % Triton X-100). Black lines represent fits of experimental data to a 1:1 or 2:1 binding model. Data are means \pm s.d. from three experiments (N=3),.....	286
Figure S47 : Plots of observed rate constants (k_{obs}) versus ligand concentrations for ligands 1-PDP and 3-360A-s and reference ligands PDS and 360A in interaction with Myc G4 (Myc-biot).....	288
Figure S48 : Adsorption isotherm (square) and fitting curve (line) using a 1:1 Langmuir interaction model for ligands 1-PDP and 3-360A-s and reference ligands PDS and 360A in interaction with Myc G4 (Myc-biot).	288

Figure S48 : Best pose obtained from the molecular dynamic simulation of the interaction of **3-360A-I** (A: side view, B: top view) and **5-360A** (C: side view, D: top view) with Myc22 (PDB:2L7V). Ligands in side views are shown as sticks and Myc22 is shown as spheres with G-tetrads in orange, loop residues in blue and flanking residues in dark grey. Ligands in top views are shown as sticks and Myc22 is displayed as residue-based cartoons and ribbon, G4 loop residues are colored in blue, the 5' flanking sequence dTdGdA has been removed for clarity.

290

Table des tableaux

Tableau 1 : Mesure des interactions G-quadruplex/ligands DEL ¹⁸⁸ par SPR	57
Tableau 2 : Tableau illustratif des composés marquants ciblant des protéines ou des acides nucléiques issus de campagnes de bibliothèques combinatoires encodées	81
Tableau 3 : Tableau associé à l'expérience de gel retard (EMSA) réalisé sur la DEL1 — [Cycle 3] et sur l'Headpiece	101
Tableau 4 : Tableau associé à l'expérience de gel retard (EMSA) réalisé sur la DEL1 — [Cycle 4] et sur l'Headpiece	104
Tableau 5 : Tableau associé au gel d'électrophorèse final de la DEL1. Caractérisation de tous les intermédiaires encodés par l'ADN dans le cadre de la première bibliothèque combinatoire	105
Tableau 6 : Tableau associé à l'expérience de gel retard (EMSA) réalisée sur les cœurs de la DEL2 — [Cycle 2]	112
Tableau 7 : Tableau associé à l'expérience de gel retard (EMSA) après deux expériences de catch-streptavidine successives (exemple du cœur trioxazole — DEL2 — [Cycle 2])	113
Tableau 8 : Tableau associé à l'expérience de gel retard (EMSA) réalisé sur la DEL2 — [Cycle 3]	115
Tableau 9 : Séquences oligonucléotidiques (3' — biotine) utilisées dans le cadre de la première campagne de sélection in vitro par affinité à partir de la DEL1	126
Tableau 10 : Quantification du nombre de copies obtenues après sélection contre la séquence du proto-oncogène Myc en fonction du ratio O/L (oligonucléotide:bibliothèque)	128
Tableau 11 : Quantification du nombre de copies obtenues après sélection contre la séquence du proto-oncogène Myc en fonction du pourcentage de détergent (Tween-20) utilisé	129
Tableau 12 : Première campagne de sélections réalisées à partir de la DEL1. Conditions de lavages décrites sous le tableau	136
Tableau 13 : Indexage unique de l'amorce Illumina Multiplexing en fonction de 29 sélections effectuées	138
Tableau 14 : Séquences oligonucléotidiques (3' — biotine) utilisées dans le cadre de la deuxième campagne de sélection in vitro par affinité à partir de la DEL1	142
Tableau 15 : Deuxième campagne de sélections réalisées à partir de la DEL1. Conditions de lavages décrites sous le tableau	151
Tableau 16 : Indexage unique de l'amorce Illumina Multiplexing en fonction des sélections de la deuxième campagne	152
Tableau 17 : Première campagne de sélections réalisées à partir de la DEL21. Conditions de lavages décrites sous le tableau	154
Tableau 18 : Ordre d'enrichissement des ligands DEL1 — [Cycle 3] en fonction de la cible G-quadruplex	170
Tableau 19 : Facteurs d'enrichissements et classement des trois composés retenus en fonction des différentes cibles	184
Tableau 20 : Enrichissement du dérivé 3-360A à chaîne courte (-s) et chaîne longue (-l) en fonction des cibles utilisés	197
Tableau 21 : Séquences G-quadruplex FAM/TAMRA utilisées lors des expériences de FRET-Melting	200
Tableau 22 : Mesure de la température de dénaturation des cibles FAM/TAMRA, en absence ou en présence de compétiteurs, à une concentration saline fixée	200
Tableau 23 : Mesure du K_D^{app} des ligands en fonction des cibles G-quadruplex par extinction de fluorescence. N=3. K_D^{app} mesurés suivant un mode d'interaction G4 : ligand 1:1 (a), 2:1 (b) ou valeurs non-déterminés (N.D.)	210
Tableau 24 : BLI. Gamme de concentration pour la caractérisation des interactions G-quadruplex/ligand	216
Tableau 25 : BLI. Constantes cinétiques et thermodynamiques obtenues sur la cible Myc. N=2	218
Tableau 26 : Résumé des données obtenues pour la caractérisation des composés par FRET-Melting, Titration de fluorescence et BLI sur les cibles de topologies parallèles (Myc, Myc22 et Kit1), antiparallèles (22CTA), hybride (wtTel26) et le contrôle Hairpin (hp1)	219

Materials & methods section

Table S1 : Characteristics of affinity selection experiments	260
Table S2 : HTS normalized reads pre- and post-selections for selected compounds and controls and corresponding enrichment values	269
Table S3 : DNA sequences used as targets for fluorimetric titrations	282
Table S4 : BB1 & BB4 – DEL1 & DEL2	292
Table S5 : BB2 – DEL1	302
Table S6 : BB2 – DEL2	303
Table S7 : BB3 – DEL1	303
Table S8 : BB3 – DEL2	305
Table S9 : DNA Tag sequences of BB1 & BB4 – DEL1 & DEL2	308
Table S10 : DNA Tag sequences of BB2 – DEL1	312
Table S11 : DNA Tag sequences of BB2 – DEL2	313
Table S12 : DNA Tag sequences of BB3 – DEL1	313
Table S13 : DNA Tag sequences of BB3 – DEL2	313

Table des équations

Équation 1 : Coefficient de variation calculé à partir de l'écart-type des fréquences (σ) divisé par la moyenne des fréquences (μ).	158
Équation 2 : Calcul du Ratio normalisé (RN) relatif à la librairie naïve d'un composé n dans une sélection ($R= x$).....	160
Équation 3 : Calcul du ratio (r) d'un composé n après sélection ($R= x$). $f_{n,R=x}$ décrivant la fréquence du composé n après sélection x et $f_{n,R=0}$ la fréquence du composé n au sein de la librairie naïve.....	160
Équation 4 : Enrichissement (E) d'un composé n après sélection ($R= x$) calculé à partir du $RN_{n,x}$ et du nombre de molécules N composant la librairie après sélection.....	161
Équation 5 : Facteur d'enrichissement (EF) d'un composé n après sélection ($R= x$) calculé à partir du $RN_{n,x}$ et de $E_{n,x}$	161

Table des schémas

Schéma 1 : Distanciation de la partie chimie de l'ADN codant par couplage d'un espaceur N-Fmoc polyéthylène glycol (PEG) nommé N-Fmoc-Ext.	87
Schéma 2 : DEL1 — [Cycle 1] — Incorporation des motifs pentapeptidiques et de leurs code-barres sur l'Headpiece-NH ₂ . DEL1 — [Cycle 1] = 46 composés.	96
Schéma 3 : Purification DEL1 — [Cycle 1] — Méthode Cap&Catch ²⁸⁷ . « Cap » des H ₂ N-HP-T1 non fonctionnalisés par de la biotin-NHS puis « catch » des composés biotinylés sur colonne Streptavidine.	97
Schéma 4 : DEL1 — [Cycle 2] — Incorporation des cœurs trifonctionnels et de leurs code-barres sur les 46 composés encodés du [Cycle 1]. DEL1 — [Cycle 2] = 414 composés.	99
Schéma 5 : DEL1 — [Cycle 3] — Incorporation des ligands G-quadruplex alcynes et de leurs code-barres sur les 414 composés encodés du [Cycle 2]. DEL1 — [Cycle 3] = 4140 composés.	101
Schéma 6 : DEL2 — [Cycle 1] — Incorporation des motifs pentapeptidiques et de leurs code-barres sur l'Headpiece-NH ₂ . DEL1 — [Cycle 1] = 59 composés.	111
Schéma 7 : DEL2 — [Cycle 2] — Incorporation des cœurs trifonctionnels et de leurs code-barres sur les 59 composés encodés du [Cycle 1].....	112
Schéma 8 : DEL2-[Cycle 3] — Incorporation des 50 ligands G-quadruplex azotures et de leurs code-barres ADN.	114
Schéma 9 : Synthèse complète de la première librairie combinatoire encodée DEL1 (190 440 composés). Représentation des cycles 1 à 4 jusqu'à l'incorporation du ClosingPrimer-1. Quantification finale de 5,32 nmoles.	117
Schéma 10 : Synthèse complète de la deuxième librairie combinatoire encodée DEL2 (696 200 composés). Représentation des cycles 1 à 4 jusqu'à l'incorporation du ClosingPrimer-2. Quantification finale de 3 nmoles.	118
Schéma 11 : Synthèse peptidique supportée sur phase solide. 1/ Déprotection. 2/ Activation. 3/ Couplage. Réside utilisée dans pour la synthèse de peptides : Rink-Amide.	189
Schéma 12 : Synthèse de la Fmoc-azidonorleucine par réaction de diazotransfert catalysée par le cuivre.	192
Schéma 13 : Synthèse SPPS des 5 séquences peptidiques.	192
Schéma 14 : Réaction de cycloaddition 1,3 — dipolaire catalysée par Cu (I). Cycloaddition d'un alcyne et d'un azoture afin de former le 1,4 — triazole.....	195
Schéma 15 : Réaction secondaire de la forme oxydée de l'Ascorbate de sodium, la DHA, avec les fonctions guanidines, par exemple, des arginines.	196

Materials & methods section

Scheme S1 : Synthesis of building block ID: ID1 and ID2.....	231
Scheme S2 : Assembly scheme for the four-step split-and-pool synthesis of the DNA-encoded chemical library DEL-1	237
Scheme S3 : Synthesis of the extended headpiece (ext-HP) from the commercially available Headpiece (HP, IDT code: /5Phos/CAT CGA TTT GGC AGT CA/iSp9//iUniAmM//iSp9/T GAC TGC CAA ATC GAT GTG).	239
Scheme S4 : Synthesis of the biotin functionalized extended-headpiece.	245
Scheme S5 : Assembly scheme for the four-step split-and-pool synthesis of the DNA-encoded chemical library G4-DEL-2	249
Scheme S6 : Functionalization of unreacted N ₃ -BB2-BB1-ext-HP-T1s-T2s-T3s with biotin.....	255

Table des graphiques

Graphique 1 : Représentation du taux d'incorporation des 72 composés azotures par chimie Click (Cu ^{II}) sur un Headpiece modifié alcyne. Rendement $\geq 80\%$ observé pour 50 composés distincts.....	110
Graphique 2 : Quantification du nombre de copies obtenues en fonction des étapes du processus de sélection. 1) Input copies, 2) Flow-through, 3) W1 à W5 : lavages successifs, 4) Output copies.....	130
Graphique 3 : Quantification du nombre de copies obtenues en fonction de la stringence des lavages (W1 à W5) appliquée et de la cible G-quadruplex considérée. Sans mention contraire, les lavages ont été réalisés à température ambiante (25°C) et sous agitation pendant 15 min.	131
Graphique 4 : Quantification du nombre de copies obtenues en fonction des conditions d'élutions utilisées. Exemple sur la cible Myc.	132
Graphique 5 : Quantification du nombre de copies obtenues après sélection classique, double-sélection ou contre-sélection sur la cible principale Myc.....	135
Graphique 6 : Quantification du nombre de copies obtenues après renforcement des conditions de stringences. Exemples des cibles A/ Myc et B/ Streptavidine.	144

Graphique 7 : Quantification du nombre de copies obtenues en fonction du taux de fonctionnalisation des billes magnétiques (1, 10 ou 100%).	145
Graphique 8 : Quantification du nombre de copies obtenues en fonction de la quantité de compétiteurs. Expériences réalisées A/ par immobilisation de la cible sur billes magnétiques ou B/ en solution dans des conditions de stringentes (W4@50 °C).	147
Graphique 9 : Quantification du nombre de copies obtenues en fonction de la quantité de compétiteurs ajoutés pour le couple TERRA/wtTel26 (compétiteur) et Myc22/22CTA (compétiteur). Expériences réalisées en conditions stringentes (W4@70 °C).	148
Graphique 10 : Représentation du nombre de molécules en fonction du nombre de copies obtenues pour chaque composé. Valeur moyenne à 173 copies/composés. Amplitude maximale de la sigmoïde à 155 copies/composés.	158
Graphique 11 : Représentation du Coefficient de variation de chaque composé de la DEL1 en fonction de sa fréquence dans la librairie naïve. Valeurs bornées pour un CV \leq 50% et une fréquence \geq 5E-04 (300 copies).	159
Graphique 12 : Enrichissement (%) des 46 pentapeptides DEL1 — [Cycle 1] en fonction des cibles hp1, scr26, Streptavidine et Myc. Valeurs obtenues en condition de sélection WASH1 (25°C). N=2 (exception de scr26).	162
Graphique 13 : Enrichissement (%) des 9 cœurs trifonctionnels DEL1 — [Cycle 2] (gauche) et des 10 ligands de structures G-quadruplex DEL1 — [Cycle 3] (droite) en fonction des cibles hp1, scr26, Streptavidine et Myc. Valeurs obtenues en condition de sélection WASH1 (25°C). N=2 (exception de scr26).	163
Graphique 14 : Enrichissement (%) des 46 pentapeptides DEL1 — [Cycle 4] en fonction des cibles hp1, scr26, Streptavidine et Myc. Valeurs obtenues en condition de sélection WASH1 (25°C). N=2 (exception de scr26).	166
Graphique 15 : Enrichissement (%) des 46 pentapeptides DEL1 — [Cycle 1] en fonction des cibles wtTel26, 22CTA, Kit1 et Myc. Valeurs obtenues en condition de sélection WASH1 (25°C). N=1.	168
Graphique 16 : Enrichissement (%) des 9 cœurs trifonctionnels DEL1 — [Cycle 2] (gauche) et des 10 motifs hétéroaromatiques DEL1 — [Cycle 3] (droite) en fonction des cibles wtTel26, 22CTA, Kit1 et Myc. Valeurs obtenues en condition de sélection WASH1 (25°C). N=1.	169
Graphique 17 : Enrichissement (%) des 46 pentapeptides DEL1 — [Cycle 4] en fonction des cibles wtTel26, 22CTA, Kit1 et Myc. Valeurs obtenues en condition de sélection WASH1 (25°C). N=1.	171
Graphique 18 : Enrichissement (%) des 46 pentapeptides DEL1 — [Cycle 1] (gauche) et DEL1 — [Cycle 4] (droite) en fonction de la stringence appliquée sur la cible Myc. Valeurs obtenues en condition de sélection WASH1, WASH2, WASH4, WASH5 et WASH4/WASH4 (double-sélection). N=1.	173
Graphique 19 : Enrichissement (%) des 9 cœurs trifonctionnels DEL1 — [Cycle 2] (gauche) et des 10 motifs hétéroaromatiques DEL1 — [Cycle 3] (droite) en fonction de la stringence appliquée sur la cible Myc. Valeurs obtenues en condition de sélection WASH1, WASH2, WASH4, WASH5 et WASH4/WASH4 (double-sélection). N=1.	174
Graphique 20 : Enrichissement (%) des 46 pentapeptides DEL1 — [Cycle 1] (gauche), des 9 cœurs trifonctionnels DEL1 — [Cycle 2] (milieu) et des 46 pentapeptides DEL1 — [Cycle 4] (droite) en fonction des cibles Myc22 et Myc. Valeurs obtenues en condition de sélection WASH4 (70°C). N=2.	175
Graphique 21 : Enrichissement (%) des 10 motifs hétéroaromatiques DEL1 — [Cycle 3] en fonction des cibles Myc22 et Myc. Valeurs obtenues en condition de sélection WASH4 (70°C). N=2.	176
Graphique 22 : Courbe de fluorescence normalisée en fonction de la température. Expérience réalisée par qPCR par mesure de la fluorescence d'émission du FAM. Caractérisation de la température de dénaturation (Tm) de l'oligonucléotide en absence (gauche) ou en présence de ligand stabilisateur (droite). Capacité stabilisateur d'un ligand mesuré par calcul du ΔTm .	199
Graphique 23 : Représentation des valeurs de stabilisation (ΔTm °C) de chaque ligands en fonction de la cible considérée, en absence ou présence de compétiteurs (ds26 ou scr26).	201
Graphique 24 : Représentation des valeurs de stabilisation (ΔTm °C) de tous les ligands (peptides, composés peptido-organiques et tronqués) pour chacune des cibles utilisées.	204
Graphique 25 : Etude des valeurs de stabilisation (ΔTm °C) de 360A (gauche) et 3-360A-s (droite) à différente concentration pour toutes les cibles G-quadruplex.	205
Graphique 26 : Représentation araignée des différentes familles des ligands en fonction des différentes cibles.	207
Graphique 27 : Dichroïsme circulaire des séquences G-quadruplex cyanines. Expérience réalisée au sein du tampon de titration.	209
Graphique 28 : Courbes caractéristiques des expériences de titrations de fluorescence. A gauche : Représentation de la baisse des valeurs de fluorescences (RFU) en fonction de la concentration en ligand (ex : [1-PDP] μ M). Au centre. Représentation de la baisse de fluorescence (rouge) ou l'augmentation de l'extinction de fluorescence (noir) en fonction des différentes concentrations en ligands. A droite : Normalisation et ajustement des valeurs de fluorescence en fonction de la concentration en ligand. La valeur de fluorescence à 0,5 permettant de mesurer le K_D^{app} du ligand considéré pour sa cible.	210
Graphique 29 : BLI. Sensorgrammes de l'interaction de PDS/1-PDP et 360/3-360A-s avec la cible Myc.	217
Graphique 30 : BLI. Iso-affinité des composés caractérisés par BLI sur la cible G-quadruplex Myc.	218

Abréviations

3

360A · (PDC) Pyridodicarboxamide bisquinolinium
360A-l-Alk · Pyridodicarboxamide bisquinolinium alcyne (chaîne latérale longue)
360A-s-Alk · Pyridodicarboxamide bisquinolinium alcyne (chaîne latérale courte)

9

96-WP · plaque 96-puits (96-well plate)

A

A · Adénine, Alanine
AA · Acide aminé (Amino Acid)
ADN · Acide désoxyribonucléique
ADNmt · ADN mitochondrial
ADNr · ADN ribosomique
AMBER · Assisted Model Building and Energy Refinement
Amicon · Filtrés à centrifuger pour la purification et la concentration de biomolécules
amoles · atto-moles (1E-18)
ARN · Acide ribonucléique
ARNm · ARN messager

B

B/W · Binding and Washing
BLI · Biolayer interferometry
Boc · tert-butyloxycarbonyl
BSA · Albumine de sérum bovin (Bovine Serum Albumine)
building block · block de construction (Building block, BB)

C

C · Cytosine, Cystéine
CD · Dichroïsme circulaire (Circular dichroism)
colonne Spin-X · Tube de microcentrifugation à membrane filtrante en polypropylène
CP · Séquence terminatrice (Closing Primer)
C-term · Extrémité Carboxy-terminale
CuAAC · Copper(I)-catalyzed alkyne-azide cycloaddition
CuSO₄·5H₂O · Sulfate de cuivre pentahydraté
CV · Coefficient de variation (CV)
Cy · Cyanine
CZ · Carbazole
CZ-Alk · Carbazole alcyne

D

D · Acide aspartique
DBF · Dibenzofulvène

DCC · Dynamic combinatorial chemistry
DCL · Dynamic combinatorial library
DEL · Librairie chimique encodée par l'ADN (DNA-Encoded chemical library)
DHA · Acide déhydroxyascorbique
DIPEA · Diisopropyléthylamine
DMF · Diméthylformamide
DMSO · Diméthylsulfoxyde
dual-pharmacophore library · Librairie à double pharmacophore

E

E · Acide glutamique
EDC · 1-Ethyl-3-(3-diméthylaminopropyl)carbodiimide
EDTA (Na₂EDTA) · Acide éthylènediaminetétraacétique disodium
ELISA · Enzyme-linked immunosorbent assay
EMSA · Electrophoretic Mobility shift assay
enrichment · enrichissement
ESAC library · Librairie encodée par auto-assamblage d'ADN (encoded self-assembling chemical library)
Extended (ou Ext) · N-Fmoc-N"-succinyl-4,7,10-trioxa-1,13-tridecanediamine

F

F · Phénylalanine
FAM · carboxyfluorescéine
FB · Fragment-based
Fmoc · Fluorénylméthoxycarbonyle
fmoles · femto-moles (1E-15)
FP · Furopyridazinone
FP-Alk · Furopyridazinone alcyne
FQ · Fluorescence Quenching
FRET · Transfert d'énergie entre molécules fluorescentes (Förster resonance energy transfer)

G

G · Guanine
G4 · G-quadruplexe
GelRed · 5,5'-(6,22-Dioxo-11,14,17-trioxa-7,21-diazaheptacosane-1,27-diyl)bis(3,8-diamino-6-phenylphenanthridin-5-ium) diiodide

H

H · Histidine
HBTU · 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetraméthyluronium hexafluorophosphate
Headpiece · Précurseur chimique bifonctionnel
HEPES · Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique

HOAt · 1-hydroxy-7-azabenzotriazole
HPLC · Chromatographie liquide à haute performance (high performance liquid chromatography)
HPtoCP · Séquence ADN (80 mers) mimant les code-barres ADN des DELs
HR-MS · Spectroscopie de masse à haute résolution (High resolution mass spectroscopy)
HS-DNA · ADN de sperme d'Harreng (Herring Sperm DNA)
HTS · Criblage à haut-débit (High-throughput screening)

I

I · Isoleucine
input copies · copies d'entrée
ISAHC · imidazole-1-sulfuryl azoture
ITC · Titration calorimétrique isotherme (Isothermal titration calorimetry)

K

K · Lysine
KCl · Chlorure de potassium
 K_d · Constante de dissociation
 K_D^{app} · Constante de dissociation apparente
known · connu (évènement)
KTDS · Kinetic target-directed synthesis

L

L · Leucine
LiCl · Chlorure de lithium

M

M · Méthionine
MBHA · 4-Methylbenzhydrylamine
MeCN · Acétonitrile
MS · Spectrométrie de masse (Mass spectrometry)

N

N · Asparagine
NaAsc · Ascorbate de sodium
NDI · Naphtalène diimide
NDI-Alk · Naphtalène diimide alcyne
NHS · N-hydrosuccinimide
NMM · N-Methylmorpholine
nmoles · nano-moles (1E-09)
N-term · Extrémité Amino-terminale

O

output copies · copies de sortie

P

P · Proline

Pbf · 2,2,4,6,7,-pentaméthyl-2,3-dihydrobenzofuran-5-sulfonyl
PCR · Réaction en chaîne par polymérase (Polymerase Chain Reaction)
PDB · Banque de données sur les protéines (Protein Data Bank)
PDC3 · Phénanthroline dicarboxamide bisquinolinium
PDC3-Alk · Phénanthroline bisquinolinium alcyne
PDP-Alk · Pyridostatine bisquinoline pyrrolidine alcyne
PEG · Polyéthylène glycol
pmoles (1E-12) · pico-moles
PNA · Acide nucléique peptidique (peptid nucleic acid)

Q

Q · Glutamine
Q-I-Alk · Benzofuroquinoline alcyne (chaîne latérale longue)
Q-s-Alk · Benzofuroquinoline alcyne (chaîne latérale courte)

R

R · Arginine
Ra · Ratio ajusté
reads · séquence d'un fragment d'ADN (Illumina)
RFU · Relative fluorescence units
RMN · Résonnance magnétique nucléaire
Rn · Ratio normalisé, Ratio normalisé, Ratio normalisé
RO5 · Règles de Lipinski (Lipinski's rules of five)

S

S · Sérine
SA · Streptavidine
SA bead · Bille de streptavidine
scFv antibody · Anticorps de fragment variable à chaîne unique (single chain variable fragment antibody)
single-pharmacophore library · Librairie à pharmacophore unique
SPPS · Synthèse supportée en phase solide (Solid Phase Peptide Synthesis)
SPR · Résonnance plasmonique de surface (Surface plasmon resonance)

T

T · Thymine, Thréonine
T-20 · Tween-20 (détergent)
TAE · Tris.HCl/Acide acétique/Na₂EDTA (tampon)
TAMRA · carboxytétraméthylrhodamine
TBTA · tris(benzyltriazolylmethyl)amine
tBu · tert-butyle
TEA.HCl · Tris/Acide acétique/EDTA, chlorhydrate
TEAA · Acétate de triéthylammonium (Triethylammonium acetate)
TERRA · Telomeric repeat-containing RNA
TFA · Acide trifluoroacétique
TGS · Target-guided synthesis
THPTA · tris(hydroxypropyltriazolyl methyl)amine
TIS · Triisopropylsilane
Tm (ou
T_{1/2}) · Température de fusion (melting temperature)

TMPyP4 · *Tétra-(N-méthyl-4-pyridyl)porphyrine*
Tris.HCl · *2-amino-2-hydroxyméthylpropane-1,3-diol
chlorhydrate*
TWJ · *Three-Way junctions, Jonction à 3 voies (Three-Way
junctions)*

U

U · *Uracile*
unknown · inconnu (évènement)
UTR · *Région non-traduite (untranslated region)*
UV · *Ultraviolet-visible (spectroscopie)*

V

V · *Valine*
VS · *Criblage virtuel (Virtual screening)*

W

W · *Tryptophane*

Y

Y · *Tyrosine*

λ

λ · *longueur d'onde*

Chapitre 1 — Introduction.

Partie 1 : L'acide désoxyribonucléique. Appariements non-canoniques et stabilisation d'édifices non-usuels.

Sous-partie 1.1 : Histoire et découverte de l'ADN.

L'essor des acides nucléiques date des premiers travaux de F. Miescher en « chimie physiologique » avec l'identification, en 1869-71¹, d'une « substance cellulaire capable d'être précipitée en milieu acide », aujourd'hui connu sous le nom d'acide désoxyribonucléique ou ADN. Les quatre bases formant la structure de l'ADN à savoir l'adénine (A), la thymine (T), la cytosine (C) et la guanine (G) ont été découvertes en 1929^{2,3} par P. Levene et il a fallu attendre les années 1950 et les travaux de E. Chargaff pour apprendre que les proportions d'A/T et G/C étaient identiques au sein du génome^{4,5}. À partir de ces travaux et de résultats de diffraction de rayons X réalisés sur des cristaux d'ADN^{6,7}, c'est en 1953 que les chercheurs J. Watson et F. Crick décrivent pour la première fois la double hélice d'ADN⁸ en proposant des modes d'appariements de bases complémentaires deux à deux, l'adénine se liant avec la thymine et la guanine avec la cytosine. Les travaux de J. Watson, F. Crick ainsi que ceux de M. Wilkins leur ont valu le prix Nobel de médecine en 1962. Il est à noter que la physicochimiste britannique R. Franklin, décédée en 1958, a également activement contribué à cette découverte. Toutefois, ne pouvant être attribuée à titre posthume, la chercheuse ne figure pas parmi la liste des lauréats du prix Nobel.

Sous-partie 1.2 : Structures de l'ADN.

Section 1.2.1 : Structure primaire des acides désoxyribonucléiques, les nucléotides.

Les séquences d'acides désoxyribonucléiques sont formées par assemblages successifs d'unités nucléotidiques composées d'une base nucléique (ou nucléobase), d'un ose pentacarboné (pentose) et d'un ou plusieurs groupements phosphates (Figure 1).

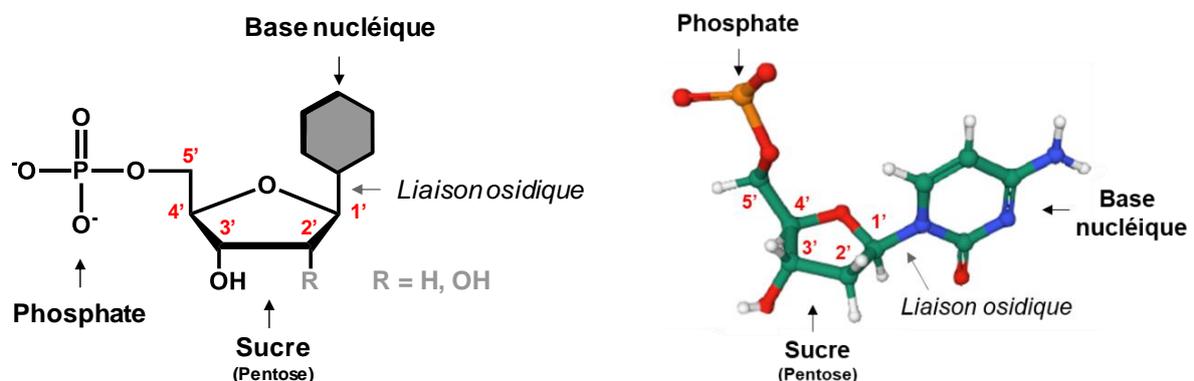


Figure 1 : Structure schématique et tridimensionnelle d'une nucléobase (PDB id : 6RLS).

Les bases nucléiques de l'ADN représentées par leur dénomination A, T, G et C sont au nombre de quatre et représentent respectivement l'Adénine, la Thymine, la Guanine et la

Cytosine (Figure 2). Dans l'univers des acides ribonucléiques (ARN), la Thymine est remplacée par l'Uracile (U). Parmi ces 5 bases nucléiques, deux motifs hétérocycliques peuvent être caractérisés, les motifs purines (A, G) et les motifs pyrimidine (T, C et U). Ces deux classes d'hétérocycles présentent des amines tertiaires et des cétones accepteurs de liaisons hydrogène (Figure 2, hétéroatomes rouges) et, suivant les nucléobases considérées, des amines secondaires et (/ou) primaires donneuses de liaisons hydrogène (Figure 2, hétéroatomes verts).

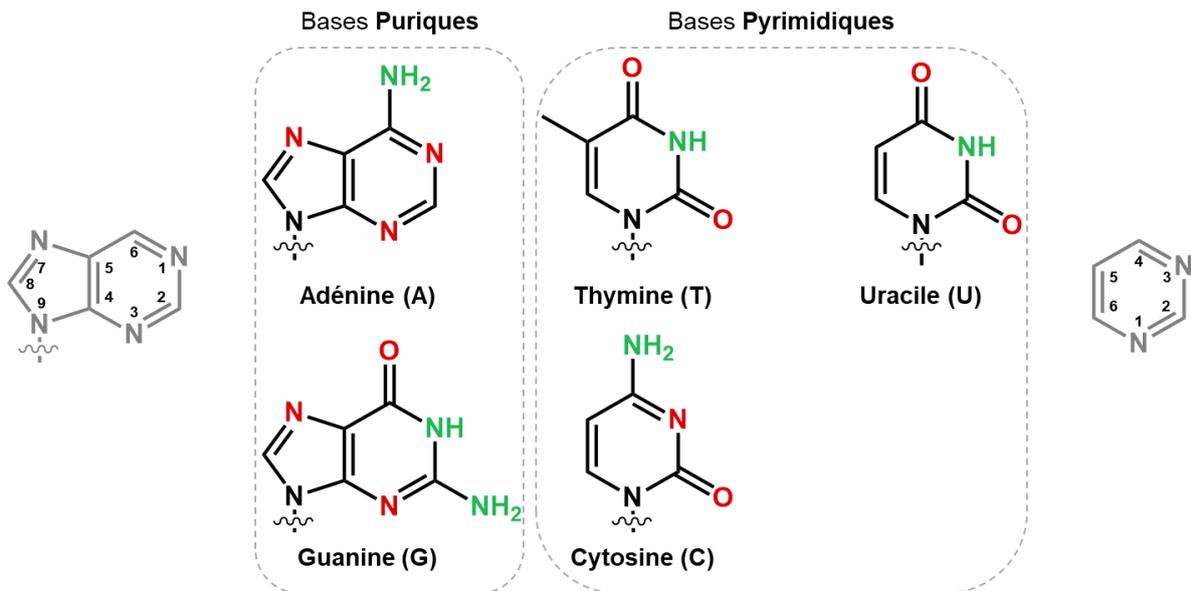


Figure 2 : Bases puriques et pyrimidiques de l'ADN et de l'ARN.

La liaison osidique caractérisant le lien entre les bases nucléiques et le sucre est une liaison de type N-glycosidique. En effet l'azote N1 des cycles puriques, ou l'azote N9 des cycles pyrimidiques, se lie avec le 2' — désoxy-D-ribose en position C1' (Figure 1). Dans le cadre des ARN, la liaison N-glycosidique est réalisée avec un sucre portant un groupement hydroxyle en position 2' du ribose (Figure 1, R=OH).

La fonctionnalisation (poly)phosphate en 5' des nucléotides permet de former des séquences d'acides nucléiques (ADN ou ARN) par formation de liaison phosphodiester. La biosynthèse de ces édifices est réalisée par des enzymes, appelées des polymérases, capables de lier le phosphate en 5' d'un nucléotide « n » avec l'hydroxyle en 3' d'une nouvelle base « n+1 » (Figure 3). Dès lors, il est possible de constater que la molécule d'ADN possède une directionnalité caractérisée par ces extrémités 3' — OH et 5' — Phosphate. La directionnalité des séquences d'ADN est une donnée importante, notamment dans le cas de la réplication de l'ADN car l'élongation de l'ADN se produit toujours du sens 5' vers 3'.

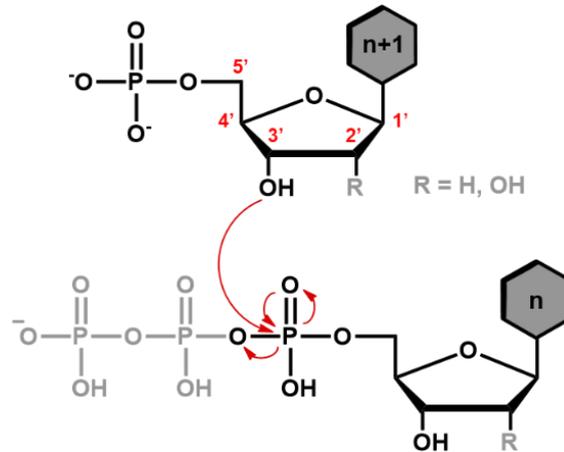


Figure 3 : Formation de la liaison phosphodiester entre deux nucléobases.

Il est également à noter que différentes conformations peuvent être adoptées par un nucléotide. La formation de la liaison N-osidique peut orienter la nucléobase soit à l'opposé du ribose (Figure 4, A/, conformation *anti*) soit vers l'intérieur du ribose (Figure 4, A/, conformation *syn*). Parmi ces deux conformations, la forme prédominante au sein de séquences ADN duplex (ou d'ARN) est la conformation *anti*. En effet, celle-ci représente le moins d'encombrement stérique entre l'hétérocycle de la nucléobase et le 5'-phosphate du sucre.

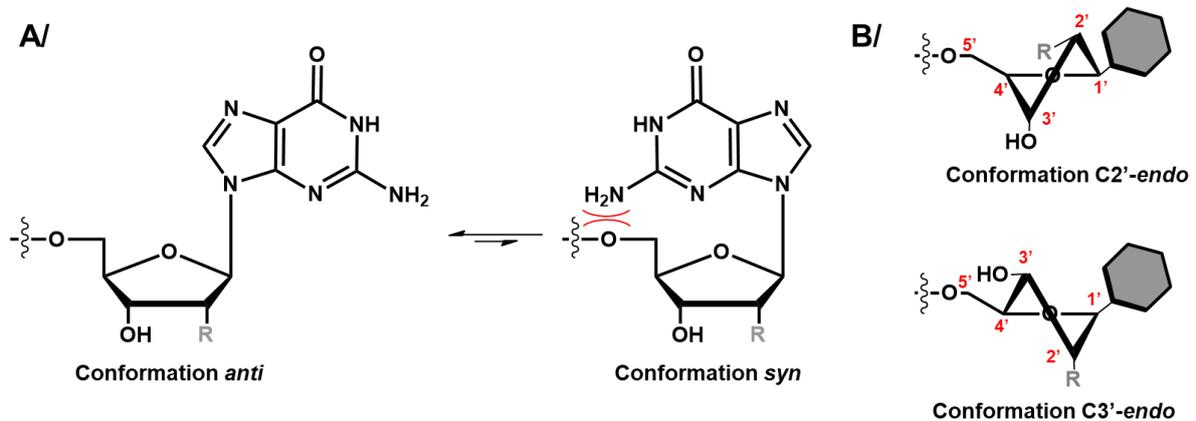


Figure 4 : Représentation des conformations possibles de A/ la liaison N-osidique (*syn/anti*) ou B/ du sucre (C2'/C3'—*endo*) des nucléobases.

De manière similaire, la partie osidique peut également présenter différentes conformations tridimensionnelles. Il est alors possible de retrouver la forme C2'-*endo* et la forme C3'-*endo* (définie selon le plan C1'-O-C4' ; Figure 4, B/). Dans le cas de l'ADN (R=H), les conformations C2'- et C3'-*endo* peuvent coexister. A *contrario*, dans le cadre des ARN (R=OH), la conformation C3'—*endo* est favorisée par des raisons d'encombrement stérique.

Section 1.2.2 : La double hélice d'ADN.

En 1953, J. Watson et F. Crick ont mis en lumière la structure bicaténaire de l'ADN à partir des travaux de cristallographie réalisés par M. Wilkins et R. Franklin. La structure en double hélice résulte de l'appariement de nucléobases deux à deux au travers de liaisons

hydrogène. De plus, la nature hétérocyclique des bases de l'ADN favorise l'empilement des nucléobases et la formation de phénomènes d'empilement π (en. π -stacking).

L'appariement des séquences désoxyribonucléiques complémentaires est régie par l'interaction entre A:T (ou A:U dans le cas des ARN) et G:C par la formation de liaisons hydrogène de type Watson-Crick (Figure 5). Dans le cadre de l'appariement des paires de bases A:T deux liaisons hydrogène sont impliquées alors que les appariements G:C exhibent trois liaisons hydrogène.

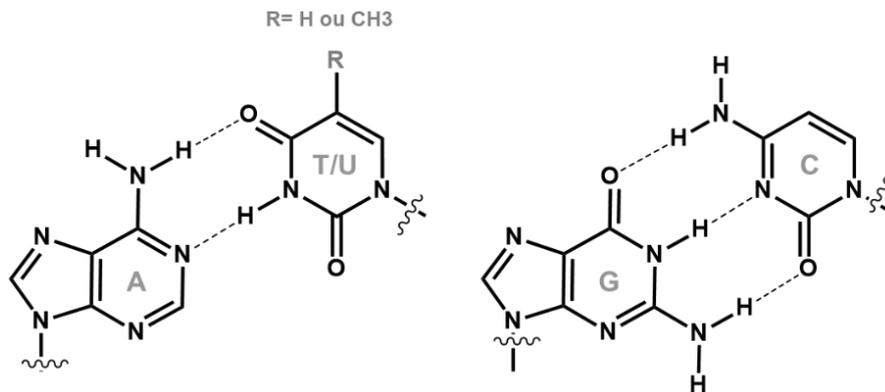


Figure 5 : Appariements Watson-Crick des bases de l'ADN et ARN.

La structure en double hélice d'ADN, formée par l'association de deux brins complémentaires, s'accompagne de la formation de sillons au sein de l'édifice. Étant donné que les brins d'ADN s'associent de façon dissymétrique par rapport à l'axe de la double hélice à cause de la dimension distincte des motifs purines et pyrimidines, deux sillons de tailles inégales sont formés : un grand ainsi qu'un petit sillon (Figure 6).

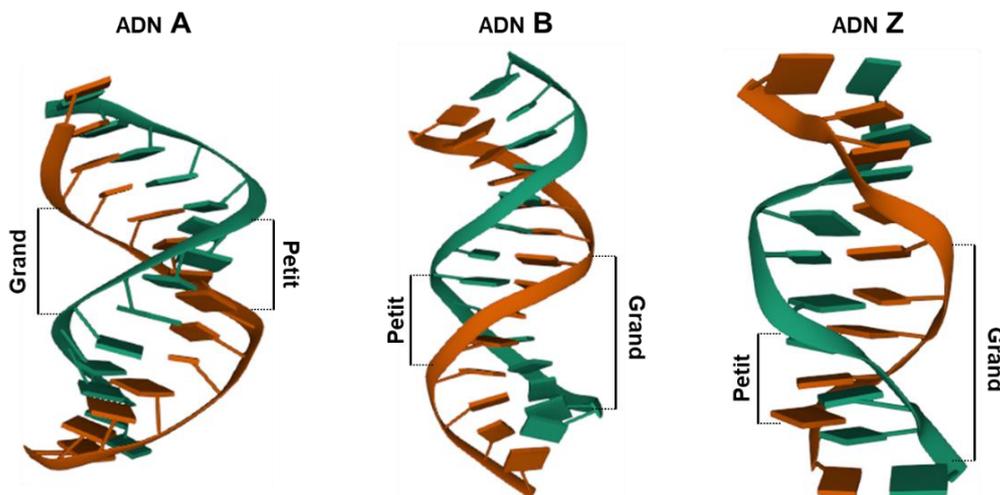


Figure 6 : Conformations de la double hélice d'ADN. ADN-A (PDB id : 4OKL). ADN-B (PDB id : 6RLS). ADN-Z (PDB id : 1R4D)⁹. Taille des sillons annotés suivant la nature de l'ADN considéré.

Les doubles hélices d'ADN existent sous forme de trois conformations distinctes. La première d'entre-elles, l'ADN-A¹⁰, est majoritairement observée au sein d'échantillons présentant de hautes forces ioniques et de faibles taux d'hydratation. Dans le cas de cette conformation, les

bases de l'ADN sont orientées en *anti* et les sucres adoptent la conformation C3'-*endo* (Figure 4). L'hélice droite formée par l'ADN-A exhibe un pas de 11 paires de bases avec un diamètre de 2.3 nm. L'autre conformère adopté par la double hélice d'ADN est connu sous le nom d'ADN-Z¹¹. Sous cette conformation, la structure générale est plus contrainte et majoritairement retrouvée au sein de séquences riches en G:C. À l'inverse de l'ADN-A, l'ADN-Z adopte une hélice gauche (diamètre = 1.8 nm, tour d'hélice = 12 pdb) dont les bases pyrimidiques sont orientées en *anti* avec une conformation des oses sous forme C2'-*endo*, alors que les bases puriques orientent les bases en *syn* et sont composées de sucres adoptant la conformation C3'-*endo* (Figure 4).

La dernière conformation pouvant être adoptée par la double hélice d'ADN est la conformation la plus répandue et connue sous ADN-B. La topologie de cet ADN a été cristallisée pour la première fois au début des années 1980¹² et présente des caractéristiques plutôt similaires à l'ADN-A. En effet, l'ADN-B est une hélice droite de 2.0 nm de diamètre possédant un pas d'hélice de 10,5 résidus par tour. De plus, les substituants des bases sur les résidus osidiques sont orientés en *anti*. La principale différence avec l'ADN-A est au niveau de la torsion du sucre puisque les furanoses de l'ADN-B adoptent une conformation C2'-*endo* (Figure 4).

Section 1.2.3 : Structures non-usuelles de l'ADN.

Au-delà des conformères principaux d'ADN sous forme bicaténaire, des structures secondaires diverses peuvent être obtenues en fonction de l'environnement ou de la nature de la séquence. Sans tous les énumérer, il est possible de faire l'état de séquences d'ADN effilées, comprenant une ou plusieurs régions non complémentaires au sein de la séquence, ou de séquences ramifiées, avec un ou plusieurs partenaires supplémentaires.

Dans le cadre de ces séquences ramifiées, il est possible d'observer des jonctions à 3 voies (en. « *Three-Way Junctions* », TWJ) ou à 4 voies (« *Four-Way Junctions* », FWJ). Les TWJ sont des structures d'ADN composées de trois séquences oligonucléotidiques partiellement complémentaires les unes aux autres (Figure 7, A/). Alors que l'assemblage de duplex d'ADN linéaire ne produit que des structures unidimensionnelles, l'assemblage de motifs TWJ offre un moyen polyvalent pour construire des réseaux bidimensionnels, des réseaux tridimensionnels, des polyèdres, des dendrimères et diverses architectures complexes¹³. Les jonctions quadruples, également connues sous le nom de « Jonction de Holliday » (Figure 7, B/), sont le résultat de la recombinaison homologue de quatre brins d'ADN. Ces jonctions sont généralement symétriques et font partie du maintien de l'intégrité du génome.

Dans un autre registre, il est possible de distinguer des séquences palindromiques capables de se replier afin de former des structures secondaires en épingles à cheveux appelées des « tige-boucle » (en. *stem-loop* ; Figure 7, C/).

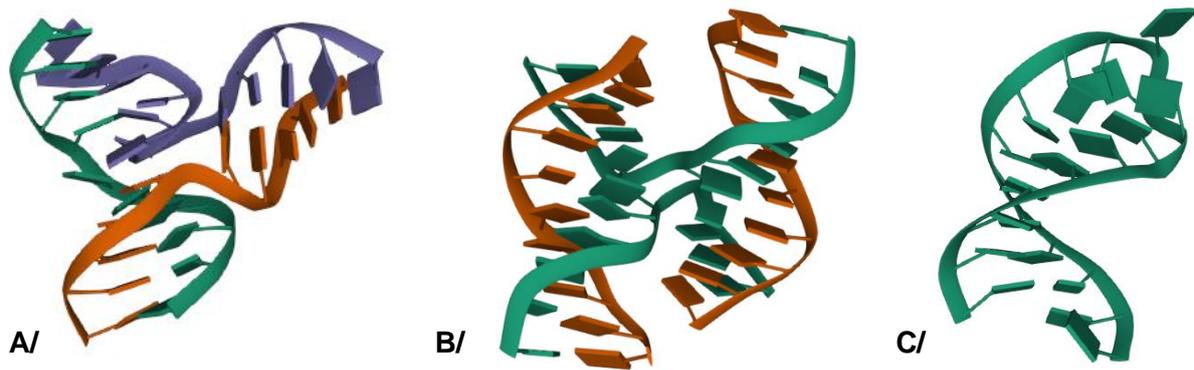


Figure 7 : Exemples de structures secondaires de l'ADN. A/ TWJ (PDB id : 1EKW). B/ Jonction de Holiday (PDB id : 1M6G). C/ Tige-boucle (PDB id : 1MFJ).

Il est également possible de rencontrer des séquences d'ADN polycaténaire capables de former des édifices secondaires particuliers telles que les G-quadruplex (mono-, bi- ou quadricaténaire), les i-motifs ou encore les ADN triplex (tricaténaire). L'ADN G-quadruplex sera défini et décrit en détail dans la suite du manuscrit.

L'ADN triplex est une structure secondaire au sein de laquelle trois brins d'ADN s'enroulent afin de former une triple hélice (Figure 8, A/). Généralement, la formation de cet édifice se fait par insertion d'un brin d'ADN dans le grand sillon de l'ADN-B bicaténaire. Les interactions entre l'ADN simple brin et le duplex se font par appariements non-Watson Crick de type Hoogsteen (voire Hoogsteen inverse). Les i-motifs sont quant à eux des édifices mono-, bi- ou quadricaténaires formés par l'interaction de résidus semi-protonés de cytosines sous forme d'appariements de type Hoogsteen.

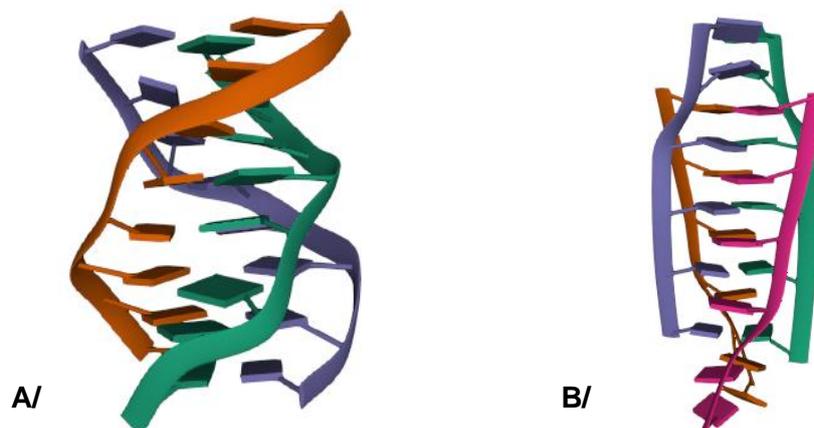


Figure 8 : Exemples de structures polycaténaires de l'ADN. A/ ADN-Triplex (PDB id : 1D3X). B/ i-motif (PDB id : 119K).

Sous-partie 1.3 : G-quadruplex, structures et fonctions.

Section 1.3.1 : Structures des G-quadruplex.

Paragraphe 1.3.1.a) Formation de motifs G-quadruplex.

La résolution cristallographique des quartets de guanines en 1962¹⁴ est le tournant qui a permis aux G-quadruplex, quelques décennies plus tard, de devenir un élément central de la génétique, de la chimie-biologie, de la catalyse ou encore de la chimie supramoléculaire.

L'historique des plateaux de guanines remonte aux années 1870, après la découverte des acides nucléiques par F. Miescher¹⁵. La nature chimique de ces nouvelles entités non-protéiques étant jusqu'alors méconnue, il a fallu attendre les travaux d'Hammersten¹⁶ en 1894 montrant l'identification de l'acide guanylique au sein de cellules pancréatiques bovines. De forts doutes concernant la forme mono- ou polymérique de cet acide a conduit, en 1910, les chercheurs Levene/Jacobs¹⁷ et Bang¹⁸ à en isoler une fraction plus pure permettant, ainsi, l'identification de la forme monomérique de l'acide guanylique sous forme « gélatineuse ». En étudiant la viscosité de gels de guanosines disodiques (5'-GMP et 3'-GMP à pH 5) à basse température¹⁹, M. Gellert et coll. ont obtenu les premiers clichés de cet hydrogel de guanines par diffraction aux rayons-X en 1962. L'équipe a ainsi établi la formation d'une structure hélicoïdale structurée par des plateformes carré-plan formées de guanines appelées tétrade de guanines ou G-quartet. Les recherches et innovations ayant été établies à partir de cette découverte cruciale ont été longuement discutées par J.T. Davis au travers d'une revue près de quarante années plus tard²⁰.

Les tétrades de guanines sont des édifices à quatre bases au sein desquels les guanines se stabilisent par liaisons hydrogène. Les faces Hoogsteen des guanines (Figure 9) présentent à l'interface deux atomes électroattracteurs accepteurs (A) d'hydrogènes, l'azote N₇ et l'oxygène O₆. Sur les faces Watson-Crick on retrouve l'oxygène O₆ mais également des groupements donneurs (D) d'hydrogènes que sont l'azote intracyclique protoné N₁ et le groupement amine secondaire N₂. Au total, un réseau de 8 liaisons hydrogène stabilise ainsi une tétrade. Il a été montré par la suite que les séquences d'acides nucléiques riches en guanine étaient capables de former des structures secondaires appelées G-quadruplex, reposant sur l'empilement par π -stacking de motifs G-quartet (Figure 9). La description de ces séquences d'acides nucléiques montre un enchaînement de bases sous la forme $(X_n G_p X_n)_z$ où X représente les différentes bases nucléiques (Adénine, Thymine, Guanine, Cytosine) à nombre variable (n) ; G représente les guanines dont le nombre (p) varie généralement entre 2 et 4 ; (z) représente le nombre de répétitions du motif dans la séquence nucléotidique.

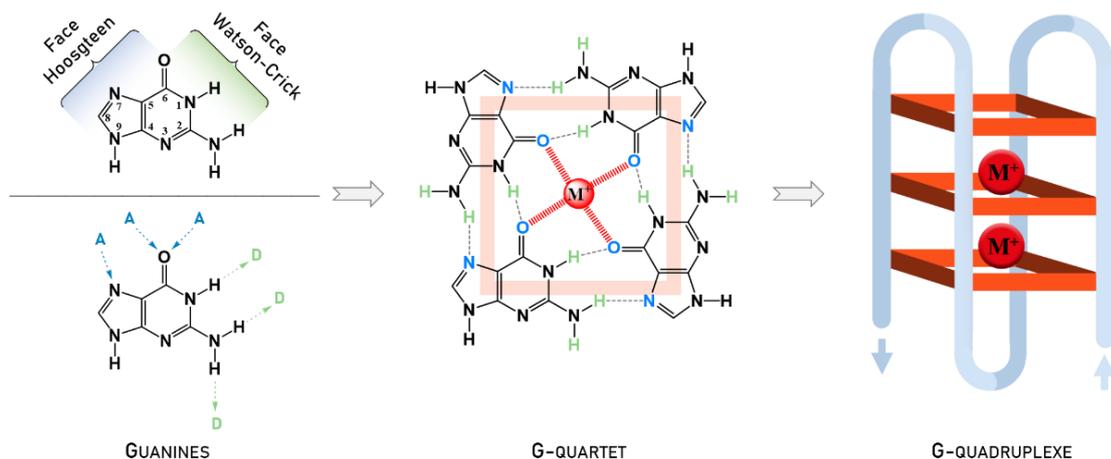


Figure 9 : représentation de la structure des guanines, la formation primaire du quartet de guanines ainsi que l'assemblage de la structure secondaire des G-quadruplex.

Paragraphe 1.3.1.b) Stabilisation par cations métalliques.

Afin de compléter les études structurales de cette nouvelle structure G-quadruplex, le groupe de T. J. Pinnavaia, en 1975, fait l'étude d'édifices hélicoïdaux de solutions concentrées de guanosine 5'-monophosphate ($\text{Na}_2(5')\text{-GMP} > 0.4\text{M}$ dans l'eau) par Résonance Magnétique Nucléaire (RMN) du proton²¹. En 1978, le même groupe démontra que le phénomène d'autoassemblage « dépendait fortement de la nature de l'ion alcalin »²² et que parmi eux, les monocations sodium (Na^+) et potassium (K^+) stabilisent le plus les plans carrés de guanines. En effet, la stabilisation des quartets de guanines par Na^+ ou K^+ s'effectue par coordination des métaux alcalins aux oxygènes O_6 des fonctions carbonyles de chaque guanine formant les G-quartets. Il propose alors que la stabilisation du sel $\text{Na}_2(5')\text{-GMP}$ est dû à la coordination octovalente de l'atome de potassium (K^+) étroitement serré entre deux plateaux de guanines. Quelques années plus tard, l'équipe de C. Bailey publie une liste de cations mono- et divalents stabilisateurs de structures G-quadruplexes²³. Classé par ordre de stabilité croissante, $\text{K}^+ > (\text{Rb}^+ > \text{Cs}^+) > \text{Ca}^{2+} > \text{Na}^+ > \text{Mg}^{2+} > \text{Li}^+$, le potassium apparaît une nouvelle fois le cation de choix pour la stabilisation de tels édifices. En effet, la position du cation métallique au sein des structures G-quadruplex explique cet ordre de stabilité. Un cation monovalent de sodium possède un rayon atomique ($r_{\text{atome}} = 180 \text{ pm}$) suffisamment petit pour coordonner quatre atomes d'oxygène O_6 au centre du carré-plan formé par les guanines. Dans le même ordre d'idée, le monocation lithium peut également se positionner au centre du G-quartet. En revanche, présentant une taille plus faible que le sodium ($r_{\text{atome}} = 145 \text{ pm}$), Li^+ n'est pas capable de coordonner les oxygènes O_6 à proximité, d'où son effet stabilisateur de G-quartet quasiment nul. *A contrario*, le potassium K^+ est un atome trop volumineux ($r_{\text{atome}} = 220 \text{ pm}$) et par conséquent ne se coordonne pas au centre de la tétrade mais entre les empilements de G-quartets, tel que l'a préconisé T.J. Pinnavaia. Cette structuration permet au potassium de coordonner huit atomes d'oxygène et ainsi d'apporter un fort effet stabilisant à la structure (Figure 10).



Figure 10 : Positionnement des cations monovalents de sodium (Na^+) et de potassium (K^+) au sein de G-quartets ou d'un édifice G-quadruplex. Dessin réalisé par D. Jiménez.

En plus d'agir comme agent de stabilisation au sein d'un motif G-quadruplex, les cations influencent également les topologies que peuvent adopter ces édifices.

Section 1.3.2 : Topologies des G-quadruplex.

Paragraphe 1.3.2.a) Les différentes topologies adoptées par les G-quadruplex.

Comme explicité précédemment, l'enchaînement de bases nucléiques $(\text{X}_n\text{G}_p\text{X}_n)_z$ décrit la formation d'édifices sous forme de G-quadruplex. Le nombre de répétitions z du motif G-quartet dans la séquence nucléotidique dicte la plurimolécularité du système.

En effet, si au sein d'une même séquence le motif $\text{X}_n\text{G}_p\text{X}_n$ est unique ($z=1$), la formation d'une structure G-quadruplex est possible par association de quatre brins ; on parlera alors d'arrangement structural intermoléculaire et plus précisément tétramoléculaire (Figure 11, a). De manière similaire, si un motif est répété deux fois au sein d'un brin d'oligonucléotide ($z=2$), alors la formation d'un G-quadruplex se fera par l'interaction de 2 partenaires, formant ainsi un G-quadruplex dit bimoléculaire (Figure 11, b). Enfin, lorsqu'une succession de 4 (ou plus) motifs est observée au sein d'une même séquence oligonucléotidique ($z \geq 4$), un motif G-quadruplex intramoléculaire (ou unimoléculaire) est capable d'être formé (Figure 11, c).

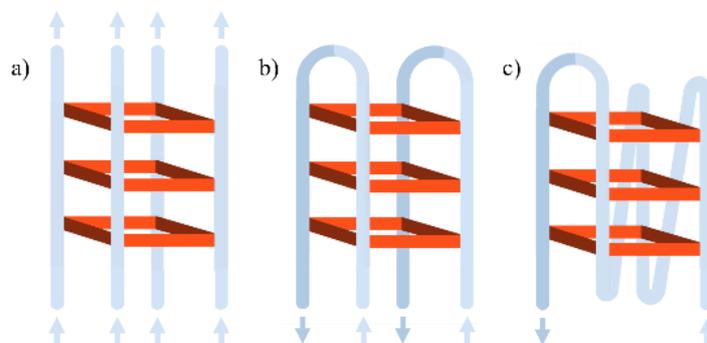


Figure 11 : Représentation exemplifiée d'un G-quadruplex tétramoléculaire (a), bimoléculaire (b) et uni- ou intramoléculaire (c). Dessin réalisé par D. Jiménez.

Au-delà de la plurimolécularité des séquences composant un édifice G-quadruplex, les structures peuvent adopter diverses topologies. Celles-ci sont fonctions de la séquence du/des brins d'acides nucléiques la composant, de la nature et de la concentration du cation métallique, de la température, ou encore des conditions d'hybridation en études *in vitro*, pour ne citer que ces quelques exemples. Les topologies que peuvent adopter les structures G-quadruplex sont au nombre de trois : i) lorsque toutes les séquences d'acides nucléiques sont de mêmes polarités, on parlera de conformation parallèle (Figure 12, a) ; ii) lorsqu'un deux brins composant l'édifice se trouvent de polarité inversée aux deux autres séquences leur faisant face, la structure adoptera une conformation dite antiparallèle (Figure 12, b) ; iii) enfin, la dernière topologie que peut adopter une structure G-quadruplex est appelée topologie hybride (Figure 12, c). Cette dernière assemble 3 séquences orientées uniformément dans le même sens et un dernier brin orienté dans le sens opposé. Ci-dessus sont représentés chacune des topologies pouvant être adopté par un G-quadruplex uni-, bi-, ou tétramoléculaire (Figure 12, représentation de gauche à droite).

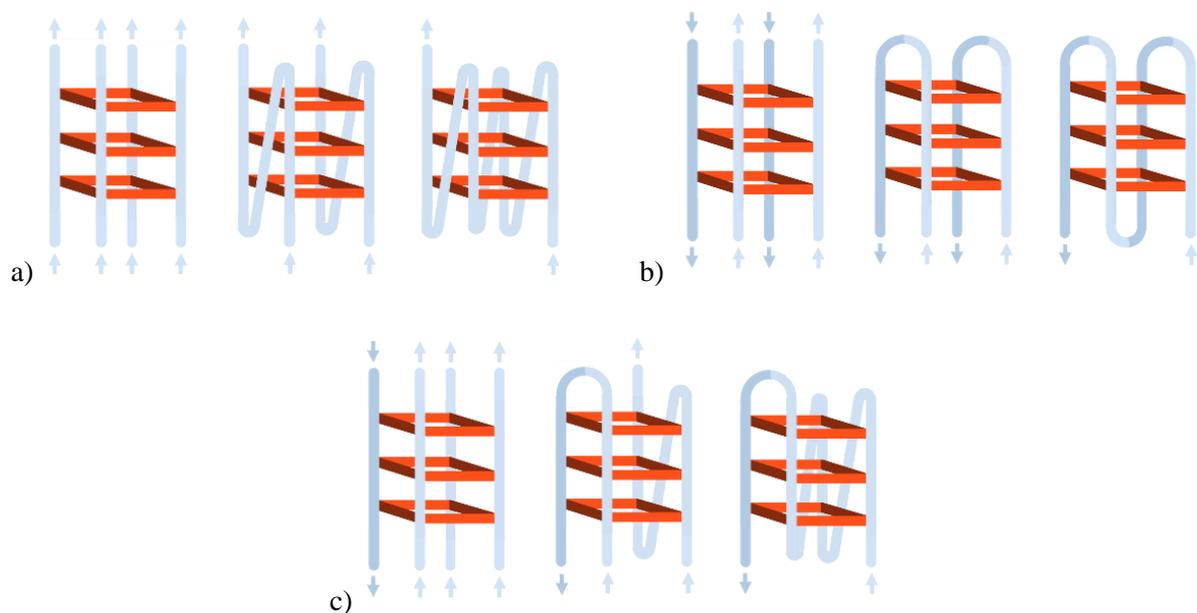


Figure 12 : Représentation schématiques des différentes topologies que peuvent adopter les G-quadruplex. De gauche à droite sont représentées les topologies parallèles, antiparallèles et hybrides de structures tétramoléculaires (a), bimoléculaires (b) et unimoléculaires (c). Dessin réalisé par D. Jiménez.

Paragraphe 1.3.2.b) Orientations des bases.

L'orientation des brins au sein d'une des topologies adoptées par les G-quadruplex est en lien direct avec l'orientation des bases azotées portées par les guanines des tétrades. En effet, l'angle de torsion glycosidique déterminée entre les atomes O4'-C1' et N9-C4 des bases puriques fait état de conformation *syn* ou *anti* des guanosines (Figure 13, a).

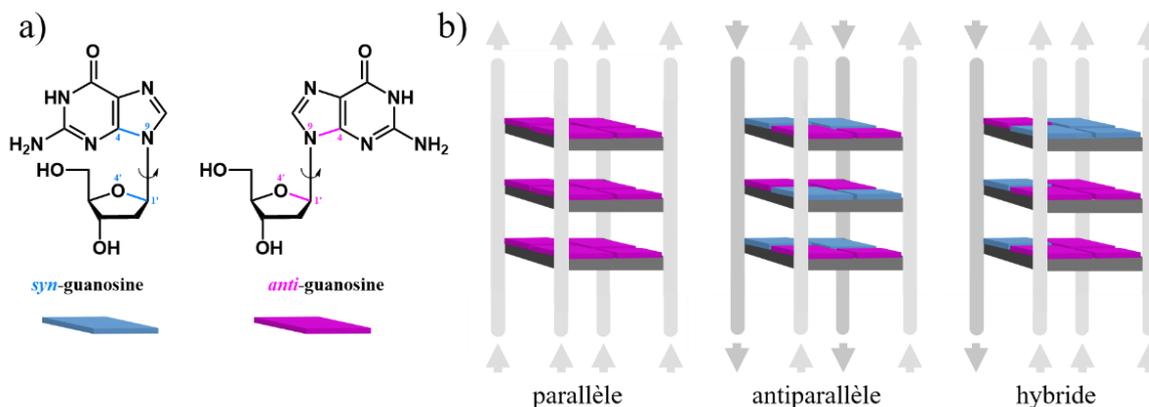


Figure 13 : Conformation de la liaison N-osidique des bases (a) en fonction de la topologie adoptée par les structures G-quadruplex (b).

En prenant le cas des G-quadruplex tétramoléculaires (Figure 13, b), on s'aperçoit qu'une topologie parallèle oriente les guanosines dans une conformation *anti* si les brins sont orientés vers le haut ou *syn* si les brins sont orientés vers le bas (non représenté)²⁴. Au sein d'une topologie antiparallèle, les brins de polarités inversés imposent aux bases se faisant face un angle de torsion opposé. En effet, lorsqu'une guanosine d'un brin orienté vers le haut présente un angle glycosidique en conformation *anti*, la guanosine du brin lui faisant face est orientée vers le bas avec la formation d'un angle de torsion en conformation *syn*. Le même phénomène est ainsi observé pour une conformation hybride (3+1).

Paragraphe 1.3.2.c) Nature des boucles et des sillons.

Au même titre que l'orientation des brins au sein d'un édifice G-quadruplex, la taille et la position de chacun des quatre sillons dépendent également de la conformation du lien glycosidique des guanosines. En effet, dans certains cas la disposition des guanosines va orienter les sucres vers l'extérieur ou vers l'intérieur du plan de la tétrade, et nous parlerons respectivement de sillons *étroit* (e) ou *large* (l). Dans un dernier cas de figure, un sucre peut être orienté vers l'intérieur alors que l'autre pointe vers l'extérieur, nous parlerons ainsi de sillon *moyen* (m). Considérant ces trois types de sillons, une combinatoire de seize arrangements peut théoriquement être observée : 6 arrangements observables avec un arrangement de tétrade comprenant 2 guanosines à angle de torsion glycosidique *syn* et 2 guanosines présentant une liaison en conformation *anti* (Figure 14, 2+2) ; 8 arrangements observables pour une tétrade contenant 3 guanosines *syn* et 1 *anti*, ou inversement (Figure 14, 3+1) ; et 2 arrangements présentant des angles de torsions équivalentes pour chaque guanosine (Figure 14, 1).

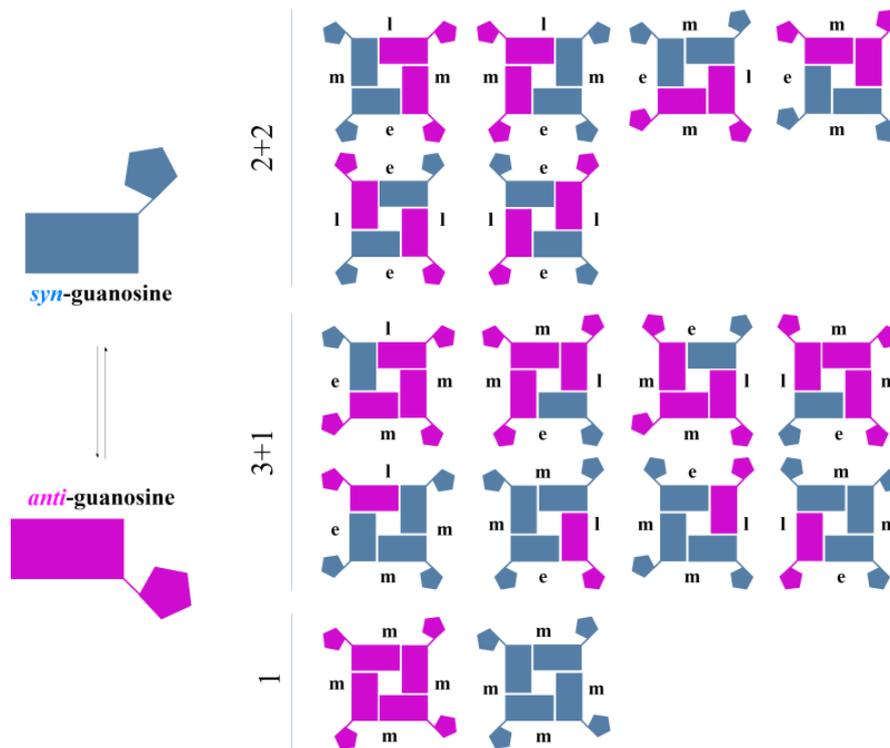


Figure 14 : Illustration des différentes conformations des guanosines possibles au sein d'un G-quadruplex. Notation de la taille des sillons : m-moyen, l-large, e-étroit.

À l'exception des G-quadruplex tétramoléculaires, les édifices à repliements structurés bi- ou intramoléculaires exhibent la formation de boucles au sein même de leurs structures. Selon l'orientation et la position des brins, trois types de boucles peuvent être discernées²⁵. Tout d'abord, lorsqu'une boucle relie deux brins adjacents parallèles, on parle de boucle externe (Figure 15). À l'inverse, lorsqu'une boucle relie deux brins adjacents antiparallèles, on parle de boucle latérale. Enfin, la formation de boucle diagonale est observée lorsqu'une séquence relie deux brins opposés de polarité inversée.

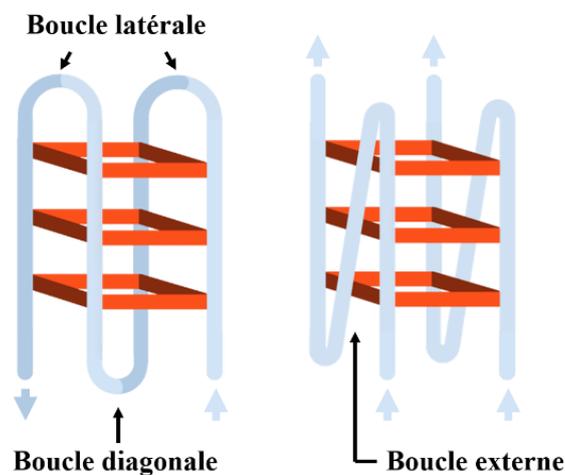


Figure 15 : Illustration des différentes boucles identifiables au sein de séquences G-quadruplex. Dessin réalisé par D. Jiménez

De nombreuses études ont été menées autour de la structure et la longueur de ces boucles. Il a notamment été montré que la longueur de la séquence d'une boucle peut influencer la topologie adoptée par les G-quadruplexes²⁶. En effet, lorsque la séquence d'une boucle porte un nombre de résidus inférieur à deux bases et que la première d'entre elles, en 5', exhibe un angle de torsion glycosidique sous forme *syn*, la configuration de la boucle s'avère être préférentiellement une boucle externe. De manière similaire, les boucles latérales se placent préférentiellement entre deux brins antiparallèles séparés par des sillons étroits portant un à deux résidus tout au plus. Enfin, il a également été montré que des boucles diagonales étaient préférentiellement formées dès lors que la séquence de la boucle dispose d'au moins trois bases.

Paragraphe 1.3.2.d) Structures atypiques de G-quadruplex.

Au-delà des structures G-quadruplex qui répondent aux classifications décrites ci-dessus, plusieurs structures non-usuelles ont été mises en évidence ces dernières années. Notamment, K.-B. Wang *et al.* ont caractérisé par RMN une structure G-quadruplex possédant un site vacant constitué de tétrades et d'une triade²⁷ (Figure 16, A/). L'équipe de A. Tuân Phan a également pu montrer que les G-quadruplex pouvaient être formés par des guanines non contiguës avec apparition d'un renflement (en, *bulge*)²⁸ composés d'une ou plusieurs nucléobases. Plus récemment, T. Q. Ngoc Nguyen *et al.* ont pu caractériser la formation d'ADN duplex au sein de renflements portés par des édifices G-quadruplexes²⁹ (Figure 16, B/).

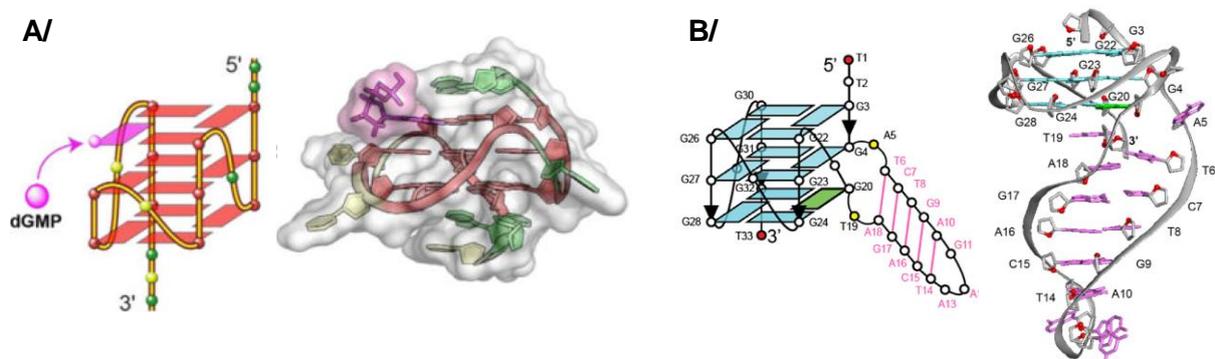


Figure 16 : A/ Structure G-quadruplex à site vacant. B/ Édifice G-quadruplex à renflement formant un ADN duplex. Schémas tirés des publications respectives (K.-B. Wang²⁷ ; T. Q. Ngoc Nguyen²⁹).

Section 1.3.3 : Occurrence et fonctions des G-quadruplex.

La prédiction ou la détermination de l'existence de structures G-quadruplex au sein d'un génome, notamment le génome humain, est une thématique largement documentée dans la littérature. Des outils biologiques ont notamment conduit l'équipe de S. Balasubramanian à la mise en évidence de G-quadruplex au sein de cellules humaines. En effet, l'utilisation de l'anticorps BG4, portant une très haute affinité pour les G-quadruplex ADN ou ARN, a permis de déterminer la présence de structures G-quadruplex ADN dans le processus de réplication³⁰ et de structure G-quadruplex ARN dans la partie cytoplasmique de cellules humaines³¹. Des outils chimiques ont également été utilisés afin d'identifier la présence de G-quadruplex. En

effet, l'équipe de D. Li a utilisé le ligand tétra-(N-méthyl-4-pyridyl) porphyrine, ou TMPyP4, afin de montrer que la stabilisation de G-quadruplex issu de la région promotrice de hnRNPk (famille des ribonucléoprotéines hétérogènes) permettait d'inhiber sa transcription et son expression³². Enfin, des méthodes informatiques (algorithmes) ont été conçues afin de prédire la présence de structures G-quadruplex au sein du génome humain. Parmi ces algorithmes on peut citer *Quadparser*³³, développé dans l'équipe de S. Balasubramanian, *G4P Calculator*³⁴, développé par Maizels et al., et enfin *G4Hunter*³⁵, mise au point par l'équipe de J. L. Mergny. Ces méthodes informatiques ont permis de prédire des séquences G-quadruplex au sein de gènes codant pour des facteurs de transcription ou des facteurs de croissance mais également au sein de gènes codant pour des phénomènes de signalisation cellulaire.

Parmi ces méthodes (biologique, chimique ou informatique), plusieurs d'entre-elles ont également montré l'implication des G-quadruplex dans des processus d'expression génique³⁶, dans la traduction de protéines³⁷, au sein de voies protéolytiques³⁸ ainsi que dans la modulation de la transcription génique³⁹. Chez l'Homme, les G-quadruplex sont capables de se former dans diverses régions, y compris les (sous-)télomères, le corps des gènes, les régions régulatrices de gènes³⁹, les régions promotrices d'oncogènes⁴⁰. Les fonctions biologiques des G-quadruplex ainsi que les mécanismes biologiques associés sont variés et riches au sein de la littérature. Il n'est pas question dans cette thèse d'en faire la synthèse exhaustive. Toutefois, des régions intéressantes et parfois ciblées dans le cadre de ce projet sont discutées.

Paragraphe 1.3.3.a) Région télomérique.

Les chromosomes, porteurs de gènes et support de l'information génétique, sont des structures de plus de trois milliards de paires de bases qui, mises bout à bout, s'assemblent sur environ un mètre de long. Réparti sous forme de 23 paires de chromosomes ($2n = 46$), le mètre d'ADN est confiné dans un noyau cellulaire de 1 à 10 μm environ alors que le repliement de chaque molécule d'ADN formerait une pelote de 100 μm de diamètre en tampon physiologique (150 mM de sels monovalents)⁴¹. Ainsi, pour assurer un degré de compaction suffisant afin que toutes les molécules d'ADN puissent siéger à l'intérieur du noyau, chaque molécule s'assemble en une structure complexe appelée chromatine, elle-même formée par l'interaction de l'ADN et des protéines associées connues sous le nom d'histones et de protéines dites « non-histones ». Ces protéines non-histones, de taille légèrement inférieure aux histones, regroupent différentes classes de protéines parmi lesquelles on peut citer les protéines 1 de l'hétérochromatine (HP1)⁴² ou encore les protéines SMC (en. *structural maintenance of chromosomes*)⁴³. La condensation en structures très ordonnées et individualisées des chromatines forme les chromosomes (souvent distingués par sa forme singulière de « X ») dont on peut y distinguer deux chromatides sœurs, c'est-à-dire portant la même information génétique. Chacune des chromatides est jointée par l'intermédiaire d'un élément ADN appelé centromère, jouant un rôle important dans la ségrégation des chromosomes, et présente à leurs extrémités des séquences répétitives d'ADN appelées télomères chromosomiques (Figure 17).

Mis en évidence par B. McClintock en 1939⁴⁴, les télomères sont connus pour protéger les extrémités chromosomiques de nombreux processus délétères. Parmi ces processus non permis, on retrouve la fusion des extrémités homologues (fusion des chromosomes), la dégradation par des nucléases endogènes ou encore la formation par recombinaison dicentrique⁴⁵. La découverte du mécanisme de protection des télomères, et plus généralement des chromosomes, ainsi que le rôle de la télomérase dans la maintenance de la structure a reçu le prix Nobel de médecine en 2009 et a été décerné à E. Blackburn, C. Greider et J. Szostak.

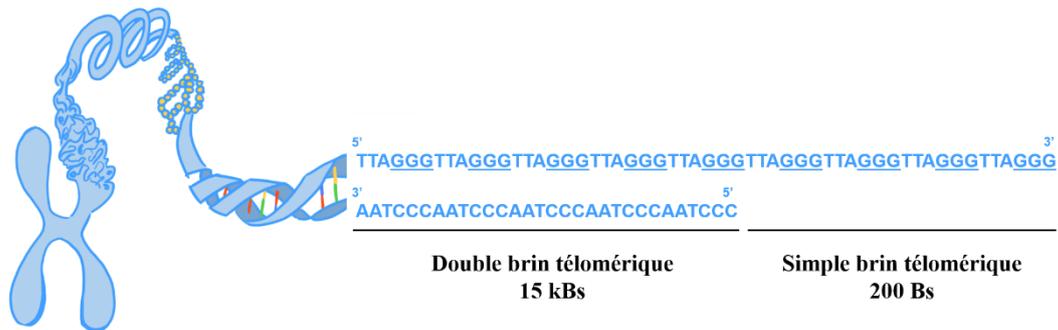


Figure 17 : Représentation schématique de l'extrémité télomérique riche en guanines des chromosomes humains. Dessin réalisé par D. Jiménez.

Chez les vertébrés, la séquence nucléotidique répétitive retrouvée au sein des télomères est 5' — TTAGGG-3' associée à sa séquence complémentaire 3' — AATCCC-5'. La longueur formée par cette répétition de séquences peut varier de 6 à 8 kBs au sein de cellules somatiques et de 10 à 20 kBs au sein de cellules germinales. L'extrémité 3' de l'ADN télomérique présente également une extension simple brin de 100 à 200 Bs⁴⁶ appelé « G-overhang » (Figure 17).

Dans les cellules somatiques saines, la télomérase est strictement régulée et le niveau d'expression est très faible, à l'exception de certains tissus, de cellules germinales mâles, certains lymphocytes activés ou encore certaines populations de cellules souches⁴⁷⁻⁴⁹. En revanche, il a été montré que dans plus de 80-85% de cellules cancéreuses, la ribonucléoprotéine est surexprimée⁴⁷. La conséquence directe de cette surexpression est le maintien de la longueur de la séquence télomérique et l'empêchement de la sénescence de la cellule.

Dû à cette répétition de séquences riches en guanines et l'implication de cette séquence télomérique au sein de diverses pathologies majeures, divers groupes de recherches se sont penchés sur l'éventuelle formation d'édifices G-quadruplex. Après que des groupes comme ceux de S. Neidle⁵⁰, G.N. Parkinson⁵¹ ou encore D.J. Patel⁵² aient démontré leurs formations, plusieurs équipes ont mis en évidence l'effet positif de la stabilisation d'un tel édifice G-quadruplexe⁵³⁻⁵⁵ sur l'inhibition de l'activité de la télomérase.

Lors de la mise en évidence de structures G-quadruplex au sein du génome de différents organismes⁵⁶ (dont l'humain⁵⁰), deux équipes de recherches^{57,58} ont été capable de démontrer

que l'ADN télomérique peut être transcrit en ARN télomérique à séquences répétitives dans des cellules de mammifères (humains et rongeurs). Cette séquence, appelée TERRA (en. *Telomeric repeat-containing RNA*) n'est, en revanche, pas la seule séquence ARN pouvant former un édifice G-quadruplex dans la région télomérique. En effet, le groupe de H. Martadinata a visualisé un motif ARN capable de former de tel édifice dans la sous-unité hTERC de la télomérase. Cette protéine est une transcriptase inverse composée de deux sous-unités⁵⁹, une partie catalytique appelée hTERT (1131 acides aminés chez l'homme) et une sous-unité ARN matricielle (451 nucléotides) appelée hTERC. Durant cette thèse, nous nous sommes plus particulièrement intéressés à la séquence télomérique connue sous le nom de wtTel26, de séquence d (TTAGGGTTAGGGTTAGGGTTAGGGTT).

Paragraphe 1.3.3.b) Régions promotrice de gènes.

Au sein des chromosomes, plus de 40% des régions promotrices de gènes présentent des séquences surreprésentant des structures G-quadruplexes⁶⁰. Considérées comme des cibles thérapeutiques en oncologie due à leurs propriétés physicochimiques et leurs caractéristiques structurales, ces séquences peuvent agir comme répresseurs transcriptionnels d'oncogènes par stabilisation des structures G-quadruplex.

De nombreuses séquences promotrices d'oncogènes ont aujourd'hui été identifiées *in vitro*, parmi lesquelles on retrouve chez l'humain le gène c-Myc⁶¹, c-Kit⁶², KRAS⁶³⁻⁶⁵, BCL-2 (B cell lymphoma 2)⁶⁶, VEGF (vascular endothelial growth factor)⁶⁷, Hif-1a⁶⁸, C-Myb (transcription factor)⁶⁹, PDGF-A (platelet-derived growth factor α polypeptide)⁷⁰, PDGFR β (PDGF receptor β polypeptide)⁷¹, RB1 (retinoblastoma protein 1)⁷², hTERT (human telomerase reverse transcriptase)⁷³ et RET. Ces régions promotrices sont des séquences surexprimées ou mutées au sein de cellules cancéreuses et la formation de structures quadruplex permet l'inactivation, dans la plupart des cas, de la traduction des gènes faisant ainsi de ces régions des cibles thérapeutiques de choix.

Très largement étudié et documenté, le promoteur d'oncogène c-Myc est impliqué dans la régulation et l'expression des gènes. Au sein de pathologies cancéreuses, il a été montré que le promoteur d'oncogènes c-Myc est surexprimé dans près de 80% des cas. De plus, au début des années 2000, l'équipe de L.Hurley a mis en évidence que la formation de structures quadruplex était identifiée dans la région NHE III₁ (en. *nuclease hypersensitivity element III₁*) promotrice du gène c-Myc⁷⁴. De séquence putative d(TGGGG(1)AGGG(2)TGGGG(3)AGGG(4)TGGGG(5)AAGG), la séquence c-Myc NHE III₁ présente un enchaînement de cinq motifs riches en guanines permettant la formation de deux G-quadruplex intramoléculaires distincts appelés Myc1234 et Myc2345, en référence à la numérotation 1 à 5 inscrite au sein de la séquence présentée précédemment.

Parmi les autres séquences promotrices contenant une structure G-quadruplex, il est possible de citer l'oncogène c-Kit présentant une structure quadruplex de topologie intramoléculaire parallèle. Le gène c-Kit code pour un récepteur de type tyrosine kinase très largement surexprimé dans les tumeurs gastro-intestinales. La protéine BCL-2, dont la région codante correspondante porte également un G-quadruplex de conformation mixte

parallèle/antiparallèle, est un inhibiteur de l'apoptose cellulaire. Le facteur de croissance de l'endothélium vasculaire, ou VEGF, dont le gène adopte une structure de type quadruplex intramoléculaire parallèle est connu pour être un régulateur de l'angiogenèse et participe activement à la survie des tumeurs cancéreuses. Intervenant également dans l'angiogenèse, le gène KRAS dispose d'un G-quadruplex intramoléculaire parallèle. Conjointement au gène Kit, le gène RET code pour un récepteur tyrosine kinase dont la surexpression est principalement retrouvée au sein des tumeurs thyroïdiennes. Le gène Retinoblastoma (Rb) code pour une phosphoprotéine nucléaire agissant comme suppresseur de tumeur et le gène chl1 code pour une hélicase capable de dérouler la séquence portant un motif G-quadruplex.

Agissant comme régulateur transcriptionnel, les régions promotrices de gènes semblent être des séquences intéressantes à cibler pour le développement de nouvelles stratégies thérapeutiques et anticancéreuses ciblant les G-quadruplex.

Paragraphe 1.3.3.c) Autres régions du génome.

Comme discuté auparavant, les régions télomériques ou promotrices de gènes sont des régions chromosomiques présentant une forte présence de structures quadruplex. Néanmoins, d'autres régions génomiques présentent également de telles structures.

En 2015, le groupe de Chambers et *al.*⁷⁵ ont mis en évidence plus de 700 000 *locus* chromosomiques capables de former des séquences G-quadruplex. Parmi ceux-ci, les gènes ribosomiques font partie des *loci* les plus riches en G-quadruplex du génome eucaryote⁷⁶. Dans la publication de A. Datta⁷⁷, les chercheurs ont discuté de la possibilité des G-quadruplex à stabiliser l'ADN ribosomique (ADNr) afin de conduire à la sénescence des cellules en altérant l'interaction de l'ADNr avec les hélicases. L'ADNr peut ainsi représenter une cible intéressante dans des approches létales synthétiques afin de lutter contre des processus cancéreux.

Dans le cadre de la recherche contre le cancer, de plus en plus de preuves soutiennent la capacité de l'ADN mitochondrial (ADNmt) à former des G-quadruplex dans les cellules cancéreuses⁷⁸. Dans les travaux de M. Falabella et *al.*⁷⁸, le ligand RHPS4 a été utilisé contre une séquence G-quadruplex ADNmt antiparallèle afin d'altérer la stabilité de la structure quadruplex et d'induire des perturbations dans le métabolisme mitochondrial. Finalement, l'étude indique que la perturbation mitochondriale induite a un rôle inhibiteur direct de la réplication du génome, de la transcription de l'ARN prémitochondriale et les fonctions respiratoires de cellules normales⁷⁹.

Paragraphe 1.3.3.d) Acide ribonucléique (ARN).

Les ARN riches en guanines forment, *in vitro*, des G-quadruplex de topologies parallèles particulièrement stables. Ces ARN G-quadruplex ont également montré leur implication au sein de plusieurs activités régulatrices. Des méthodes de prédictions algorithmiques ainsi que du séquençage structure-dépendant ont permis de détecter et localiser les structures ARN G-quadruplex au niveau du transcriptome avec une grande sensibilité⁸⁰. De plus, ces édifices ont été associés à de nombreux processus biologiques tels que des événements transcriptionnels

et post-transcriptionnel⁸¹ ainsi que dans des processus de traduction d'ARN messenger (ARNm)⁸²⁻⁸⁴.

De nombreuses études génomiques ont également indiqué que la majorité des ARN G-quadruplex était formée au sein de régions non-traduites (en. *untranslated region*, UTR) des ARNm, les régions 5'-UTR^{85,86}, 3'-UTR^{87,88}. Ces régions sont capables de recruter des protéines de traduction spécifiques et de réguler l'expression d'un gène. Toutefois, à l'exemple de la région 5' — UTR de l'oncogène *NRAS*, il a été montré que la formation de structure G-quadruplex était capable d'engendrer l'inhibition de l'expression des gènes^{89,90}. D'autres études ont également montré que les séquences G-quadruplex étaient retrouvées au sein de sites d'épissages alternatifs (en. *alternative splicing (AS) sites*) et qu'elles étaient capables de réguler des processus post-transcriptionnel⁹¹. Des études concernant le sixième intron du gène *hTERT* (en. *human Telomerase reverse transcriptase*) ont notamment montré la réduction des phénomènes d'épissage par formation de structure G-quadruplexe⁹². Les exons peuvent également présenter des G-quadruplex et participer à la régulation de processus d'épissage alternatif, à l'exemple de l'étude sur le gène *FMR1* (syndrome de l'X fragile) où la formation de deux ARN G-quadruplex induit une augmentation de l'activité d'épissage⁹³.

Des séquences ARN G-quadruplex ont également été retrouvées au sein des télomères. En effet, l'ARN long non codant (en. *long non-coding RNA*, lncRNA) de la séquence télomérique appelée TERRA (en. *Telomeric Repeat-containing RNAs*) est issue de la transcription du brin télomérique riche en C et adopte une topologie parallèle⁹⁴⁻⁹⁶. TERRA contribue à l'homéostasie télomérique par interaction dépendante avec l'ADN G-quadruplex et la protéine télomérique TRF2⁹⁷. Également présent au sein de la télomérase humaine, le lncARN G-quadruplex appelé TERC (en. *Telomerase RNA Component*)⁹⁸ a montré *in vivo* son implication dans le recrutement de l'hélice RHAU⁹⁹. Cette étude a ainsi montré l'implication de l'ARN TERC au sein de l'activité télomérique de l'enzyme¹⁰⁰.

Paragraphe 1.3.3.e) Séquences génomiques non humaines.

Au-delà des études réalisées sur le génome humain, les structures G-quadruplex ont également été retrouvées dans les génomes de protozoaires, de bactéries ou encore de virus.

Chez les protozoaires, de nombreuses séquences G-quadruplex ont été détectées et étudiées. C'est notamment le cas de *Plasmodium Falciparum*¹⁰¹ où les édifices G-quadruplex ont été retrouvés au sein de séquences (sub)télomériques, impliqués dans des processus de variation antigénique et de pathogenèse, ainsi qu'au sein du gène *COX1* (génome organellaire mitochondrial), impliqué notamment dans des processus de biosynthèse d'hème. Des études ont également été effectuées chez *Stylonychia lemnae*¹⁰² avec la production de deux anticorps, Sty3 ciblant les G-quadruplex parallèles et antiparallèles et Sty49 présentant une spécificité pour les G-quadruplex antiparallèles. Les G-quadruplex bactériens ont principalement été recensés au niveau des ARN de transfert ainsi que de l'ARN ribosomal¹⁰³. Dans les cas d'*Escherichia coli* et *Trypanosoma brucei*, des études ont montré l'implication des séquences G-quadruplex en tant qu'agent anticancéreux¹⁰⁴ et en tant qu'agent antiparasitaire¹⁰⁵, respectivement. Les virus possèdent également des génomes capables de

proposer des séquences G-quadruplex. Au sein des virus à ADN il est possible de citer le virus de l'Herpes¹⁰⁶ ou encore le papillomavirus¹⁰⁷ et dans le transcriptome, les virus ARN des flavivirus (virus Zika¹⁰⁸) ou encore du SARS coronavirus (SARS-CoV-2¹⁰⁹). Les tétraplexes de guanines peuvent également être retrouvés au sein de certains rétrovirus comme le virus de l'immunodéficience humaine de type 1 (HIV-1^{110,111}).

L'emploi de ligands de structures G-quadruplex connus, ou dans certains cas d'anticorps spécifiques, a fourni des informations intéressantes sur la régulation de nombreux processus biologiques. La présence établie de structures G-quadruplex au sein de cellules vivantes³⁹ ainsi que leurs implications au sein de multiples fonctions biologiques font de ces édifices des cibles intéressantes justifiant l'élaboration de stratégies de synthèses de composés chimiques (ligands) capables de cibler spécifiquement ces structures^{60,112}.

Sous-partie 1.4 : Ligands de structures G-quadruplex.

Au cours des vingt dernières années, de nombreuses molécules interagissant avec les ADN (ou ARN) G-quadruplex ont été rapportées et, à ce jour, plus de 3000 ligands ont été déposés dans la base de données « G4 Ligands Database 2.2¹¹³ » (<http://www.g4ldb.com/>). Certaines de ces structures sont mises en évidence dans la Figure 18 ci-dessous.

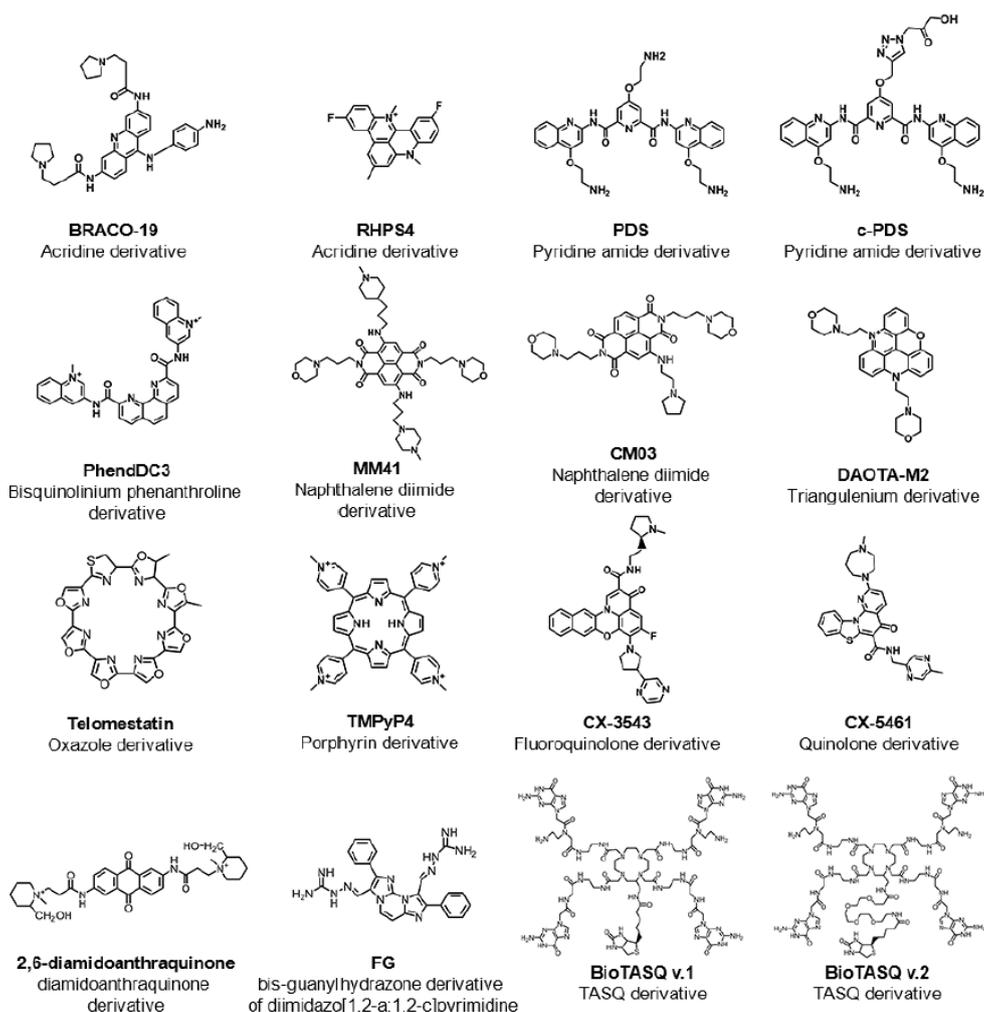


Figure 18 : Ligands de structures G-quadruplex tirés de la review de T.Santos¹¹⁴.

La conception rationnelle de ligands sélectifs aux structures G-quadruplex nécessite une connaissance accrue des modes d'interactions entre les ligands et la structure G-quadruplex¹¹⁵. Ainsi, les liaisons principales entre un ligand et sa cible se font, généralement, par i) empilement sur les tétrades externes supérieure ou inférieure, ii) par interactions hydrophobes ou électrostatiques au niveau des boucles ou des sillons^{116,117}, iii) par intercalation au sein de la structure quadruplex ou enfin iv) par combinaison de ces modes de liaisons¹¹⁸. En accord avec ces modes de liaisons, la plupart des ligands de structure G-quadruplex partagent certaines caractéristiques communes. En effet, les ligands présentent principalement des systèmes aromatiques plan capables de s'empiler sur les tétrades de guanines par interaction π - π . Ces unités aromatiques peuvent être fusionnées avec d'autres surfaces rigides, ou conjuguées avec des fragments aromatiques flexibles par le biais de liaisons rotatives. Dans un deuxième temps, afin de pouvoir proposer des interactions électrostatiques additionnelles avec les sillons ou les boucles des structures G-quadruplex, des chaînes latérales cationiques sont couramment ajoutées aux ligands. En effet, ces entités chimiques chargées positivement permettent d'exploiter des interactions électrostatiques avec les groupements phosphates des boucles et/ou sillons des structures ADN (ou ARN) G-quadruplex¹¹⁹.

La structure des ligands doit également répondre à certains paramètres biophysiques essentiels. La constante de dissociation (K_d) est un paramètre d'affinité primordiale caractérisant l'interaction entre le ligand et le G-quadruplex. Généralement, les K_d des ligands observés vis-à-vis des structures G-quadruplex sont inférieurs à 10^{-6} mol.L⁻¹ (μ M). De plus, en supplément des paramètres d'affinité ou de stabilisation, les ligands de structures G-quadruplex doivent présenter une certaine sélectivité vis-à-vis d'ADN duplex largement majoritaire au sein des cellules. Toutefois, parmi les ligands largement étudiés et utilisés dans la littérature, certains ne présentent qu'une faible sélectivité vis-à-vis des duplexes d'ADN (BRACO-19¹²⁰, TMPyP4¹²¹). Des méthodes d'optimisations de propriétés physicochimiques de ligands dans le but d'équilibrer le rapport affinité/sélectivité¹²² ainsi que des méthodes de développement de ligands spécifiques et sélectifs n'altérant pas l'affinité¹²³ des molécules ont été largement documentées.

La comparaison directe des effets des ligands en fonction des cibles ADN (ou ARN) G-quadruplex est complexe tant les résultats obtenus dépendent des séquences exactes utilisées, des conditions expérimentales (concentration en ions par exemple) ou des méthodes utilisées. L'objectif de cette sous-partie est de faire mention des différentes classes de ligands de structures quadruplex, catégorisés en fonction de la structure des ligands, en présentant notamment ceux utilisés dans le cadre du projet de thèse.

Section 1.4.1 : Systèmes polycycliques/aromatiques fusionnés.

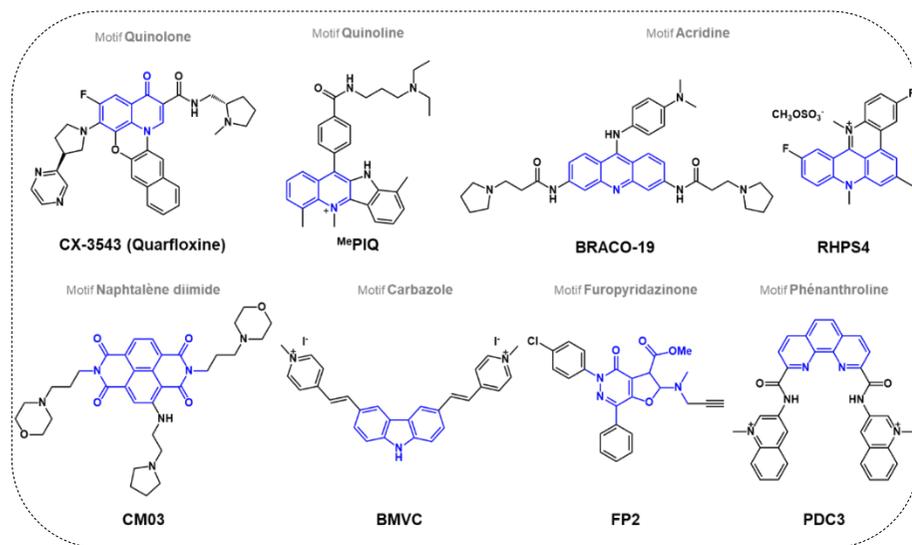
Ces dernières années, plusieurs classes de composés présentant de larges surfaces aromatiques ont été synthétisées et certains d'entre eux sont devenus des candidats prometteurs pour des applications thérapeutiques (Figure 19, A).

Les dérivés quinolines/quinolones sont des hétérocycliques azotés grandement utilisés dans la littérature pour leurs propriétés antiparasitaires et antibactériennes, respectivement. Des études réalisées sur le dérivé fluoroquinolone ont montré leurs capacités à se lier à la fois à la topoisomérase(II) mais également à des édifices G-quadruplex¹²⁴. Ces recherches ont conduit le groupe pharmaceutique *Cylene* à breveter le **CX-3543**, ou Quarfloxine, en tant que premier dérivé fluoroquinolone capable d'interagir avec les structures G-quadruplex afin de générer des activités anticancéreuses. A noter que ce composé a atteint des essais cliniques de phase II (NCT00780663) pour le traitement des carcinomes neuroendocriniens¹²⁵. Les dérivés quinolines ont également été considérés en tant que ligand de structure quadruplexe après des études réalisées sur la cryptolepine, un dérivé indolo[3,2-b]quinoline, pour ces capacités d'inhibiteur de la télomérase et suppresseur d'expression d'oncogènes¹²⁶⁻¹²⁸. À l'exemple du composé **MePIQ** publié par A. Funke¹²⁹, le groupe a pu montrer par FRET-Melting que le composé présentait un effet stabilisateur de plus de 30°C contre la cible Myc et une bonne sélectivité vis-à-vis du duplex ADN ds28. D'autre part, ce composé présente une bonne affinité pour les séquences G-quadruplex avec une constante d'affinité de 1-2 μ M (titration UV/Vis).

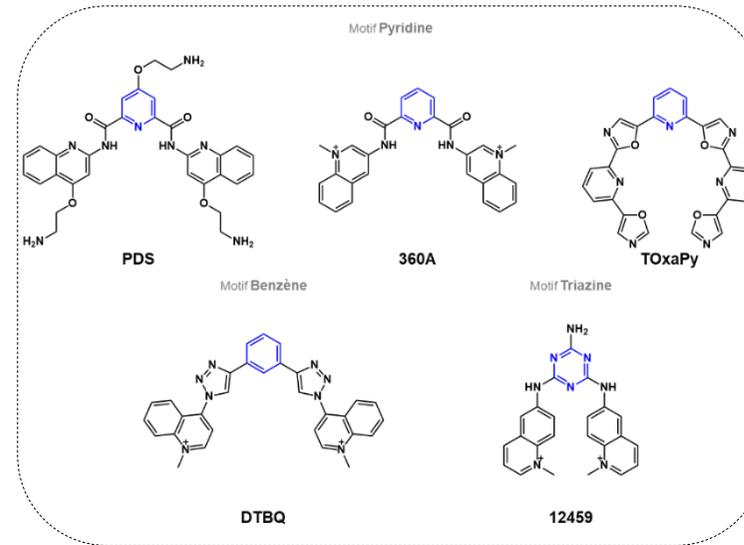
La famille des acridines, composés pyridine entourés par deux noyaux benzènes sur le plan horizontal, présente des molécules aromatiques planes généralement symbolisées par le ligand BRACO-19 ou encore RHPS4. **BRACO-19** est une acridine polydérivée présentant trois groupements amines capables d'effectuer des interactions avec les boucles ou les sillons des structures G-quadruplex. Le plan large formé par le noyau acridine permet également d'associer des interactions par empilement π avec les tétrades de guanines des édifices. Des études d'affinités de BRACO-19 sur la cible ADN télomérique humaine a permis de quantifier un K_d de 62 nM¹³⁰ avec des activités inhibitrices de croissance tumorale par entrave de l'interaction télomérase/télocromosome¹³¹. Le groupe de Z. Huang¹³² a synthétisé une librairie de sels de quinoacridium dans laquelle le composé **RHSP4** est apparu comme le plus sélectif de cellules tumorales avec une activité inhibitrice de télomérase de 330 nM (IC_{50}).

D'autres châssis moléculaires ont également été utilisés afin de cibler les structures G-quadruplex, à savoir les naphthalènes diimides (NDI), les carbazoles (Z) ou encore les Furopyridazinone (FP). Les naphthalènes diimides sont spécialement connus pour leurs propriétés optiques. Toutefois, leurs plateaux étendus électroniquement pauvres en électron ainsi que leurs surfaces planaires ont fait de ces structures des composés intéressants pour des interactions avec les cibles d'acides nucléiques, notamment les G-quadruplex. Le premier exemple, décrit par le groupe de S. Neidle, montre la possibilité d'un NDI tétrasubstitué (éthylpyrrolidine) de présenter des valeurs de stabilisation de 29°C contre la cible Kit1 avec une bonne spécificité vis-à-vis d'ADN duplex¹³³. Plus récemment le même groupe a synthétisé un dérivé NDI trisubstitué, nommé **CM03**, capable d'inhiber la croissance de cellules pancréatiques cancéreuses ($GI_{50} = 11$ nM) par interaction avec les structures G-quadruplex formés au sein du gène PDAC (en. *Human pancreatic ductal adenocarcinoma*)¹³⁴.

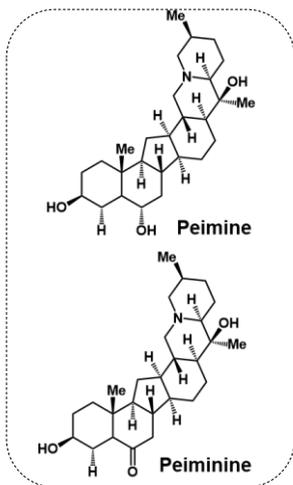
A/ Systèmes polycycliques/aromatiques fusionnés



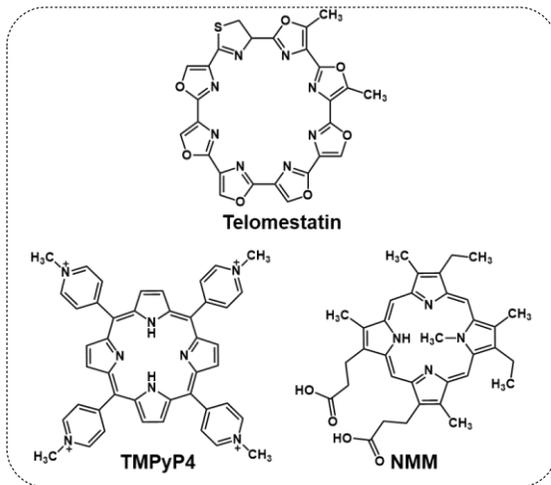
B/ Systèmes aromatiques non fusionnés



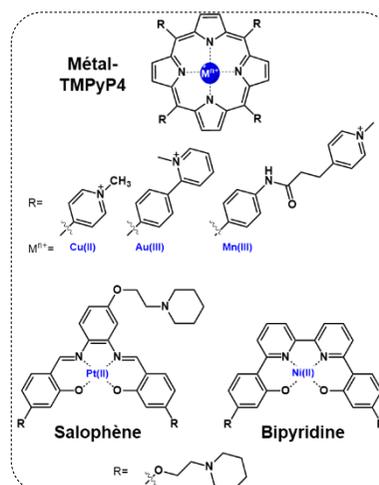
C/ Systèmes non aromatiques



D/ Systèmes Macrocycliques



E/ Systèmes organométalliques



F/ Biomolécules et dérivés

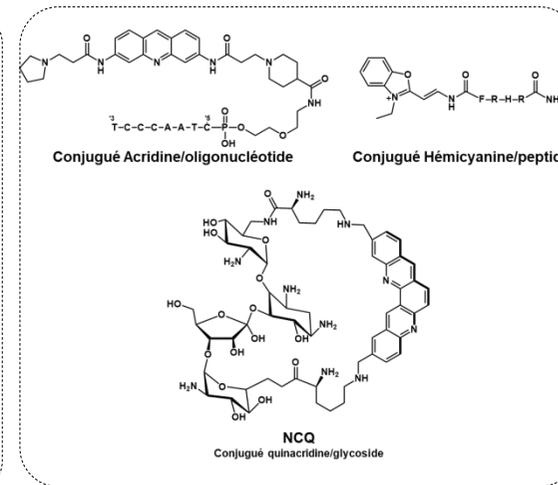


Figure 19 : Illustration de certaines classes de ligands ciblant les tétraplexes de guanines.

Les composés carbazoles en tant que ligand de structure G-quadruplex ont été introduits pour la première fois par Chang *et al.* en 2003^{135,136}. Les carbazoles (CZ) sont des motifs hétérocycliques riches en électrons, composés d'un noyau pyrrolidine, décorés de deux noyaux benzène et possédant des propriétés stabilisatrices de structures G-quadruplex ainsi que des propriétés de biosenseurs d'acides nucléiques. En effet, des études *in vitro* ont montré que le ligand **BMVC** était capable de stabiliser le G-quadruplex issu de la séquence télomérique humaine à plus de 13°C¹³⁶. De plus, BMVC est capable d'augmenter ces propriétés intrinsèques de fluorescence de plus de 90 fois lorsqu'il est en interaction avec une structure G-quadruplex¹³⁵. Parmi les autres variétés de ligands de structure G-quadruplex, il est possible de considérer les Furopyridazinones (FP). Ces motifs générés par réactions multicomposants, sont présents dans un bon nombre de ligands biologiquement actifs (antihypertenseur, antithrombotique, anti-inflammatoire, anticancéreux). Le groupe de B. Pagano a mis au point la synthèse de furo[2,3-*d*]pyridazin-4(5H)-one^{137,138} par réaction multicomposant monotope (en. *one pot*) et a adapté la synthèse à la recherche de composé ciblant les G-quadruplex¹³⁹. Le composé **FP2** décrit présente de faibles valeurs de stabilisations ($\Delta T_m \sim 4^\circ\text{C}$) mais montre néanmoins une spécificité préférentielle pour la séquence G-quadruplex BCL2 vis-à-vis des proto-oncogènes (Myc et Kit1) ainsi que la séquence télomérique (Tel24). Le groupe a également pu montrer que FP2 permettait de réguler la transcription et l'expression de BCL-2 au sein de cellules cancéreuses (Jukats).

Parmi les ligands fréquemment décrit et couramment utilisé dans la littérature pour l'étude de structures G-quadruplex, il est nécessaire de mentionner le composé issu de la famille des phénanthrolines, le **PDC3**. Les phénanthrolines possèdent de larges motifs aromatiques et plans dont les dérivés assurent une bonne affinité pour les G-quadruplex, spécialement par interaction par empilements π . Décrit en 2007 par A. De Cian¹⁴⁰, le PDC3 (ou PhenDC3) est issu de la réaction de l'acide 1,10-phénanthroline-2,9-dicarbonique avec deux motifs amino-méthylquinolinium. Les interactions hydrogènes des paires libres des atomes d'azotes (phénanthroline) avec les protons des deux amides formés permettent de bloquer la conformation de la structure et ainsi de couvrir une large surface riche en électrons. Ces propriétés ont notamment permis d'obtenir de grandes valeurs de stabilisation contre la séquence G-quadruplex télomérique (F21T, $\Delta T_m = 29.7^\circ\text{C}$) avec une haute spécificité pour le duplex d'ADN. D'autres équipes ont fait la synthèse de dérivés de PDC3 capables de discriminer les séquences G-quadruplex présentent au niveau des régions promotrices d'oncogènes (telle que Myc ou Kit) par rapport à celles issues des séquences télomériques. Parmi ces composés, le dérivé phénanthroline *Phen-Pr* présente une stabilisation d'environ 20°C pour les séquences promotrices d'oncogènes, une stabilisation de 1°C pour l'ADN G-quadruplex télomérique, et une affinité de l'ordre de la dizaine de micromolaire¹⁴¹.

Section 1.4.2 : Systèmes aromatiques non fusionnés.

Les ligands de structures G-quadruplex les plus utilisés comportent en leur centre des systèmes aromatiques polycycliques capables d'apporter une surface d'empilement π très large. Toutefois, plusieurs ligands décrits dans la littérature disposent d'un châssis principal

monocyclique aromatique (Benzène, Pyridine, Triazine, etc). L'empilement hydrophobe recherché avec ce genre de composé passe donc par une dérivatisation riche en systèmes aromatiques mono- ou polycycliques (Figure 19, B/).

Les cycles phényle et pyridine, dans un modèle de substitution (1,3), sont les éléments les plus courants pour le noyau central, avec une préférence marquée pour la pyridine car elle peut fournir une stabilisation supplémentaire via des interactions dipôles-cations. Les modèles (1,4) — pyridine/benzène ont également été explorés mais se sont avérés inefficaces en termes d'interaction avec les G-quadruplex (voire les ADN duplex)¹⁴². La dérivatisation de ces motifs centraux se fait généralement par liaison sp^2-sp^2 afin d'induire une géométrie particulière à la molécule. Dans le cadre des motifs pyridines, les structures les plus couramment retrouvées sont les composés pyridine-2, 6-carboxiamide, ou pyridocarboxamide. La dérivation des motifs pyridines par des groupements amides permet non seulement de former une liaison sp^2-sp^2 , et ainsi fixer la géométrie, mais permet également d'introduire deux fonctions acceptrices et donneuses de liaisons hydrogène. De manière similaire à PDC3, les dérivés pyridocarboxamides exploitent la formation de liaisons hydrogène entre les fonctions NH(R) de l'amide, la paire libre du motif pyridine et une molécule d'eau afin d'empêcher la libre rotation de la liaison sp^2-sp^2 . Dans l'optique d'obtenir des ligands de structures quadruplex, les pyridocarboxamides sont régulièrement accompagnés de motifs quinolines/quinoliniums afin d'apporter de la densité électronique et donc des propriétés d'empilement π . Le ligand **360A** (ou *PDC*), développé en 2005 par G. Pennarun et al.¹⁴³, est un composé pyridocarboxamide bisquinolinium capable d'inhiber l'accès de la séquence télomérique à la télomérase par stabilisation de la structure G-quadruplex et ainsi d'induire l'apoptose des cellules. L'introduction de chaînes latérales chargées sur les motifs quinolines ou sur la pyridine centrale permet d'augmenter les interactions électrostatiques et les liaisons hydrogène avec les structures G-quadruplex. Ce rationnel a conduit à la synthèse du composé **PDC**, un pyridocarboxamide bisquinoline trisubstitué par des motifs éthanolamines¹⁴⁴. Outre les dérivés pyridocarboxamide bis-quinoline/quinolinium, les motifs pyridines ont également été fonctionnalisés par des chaînes hétérocycliques en série, à l'exemple du **TOxAPy** décrit par le groupe de M.P. Teulade-Fichou¹⁴⁵. La formation de liaison sp^2-sp^2 entre la pyridine et les séquences « oxazole-pyridine-oxazole » présente un composé heptacyclique capable non seulement d'être sélectif des édifices G-quadruplex mais également de discriminer certaines topologies de structures G-quadruplex. En effet, des études sur la séquence télomérique F21T ont montré que TOxAPy était capable de stabiliser la structure G-quadruplex formée en présence d'ions Na^+ ($\Delta T_m = 10.8^\circ C$) et non celle formée en présence d'ions K^+ ($\Delta T_m < 1^\circ C$). Des isomères de l'heptacycles ont également été analysés et des constantes de dissociations de 60 nM ont été mesurées sur la séquence télomérique 22AG (Na^+) et 200 nM sur le proto-oncogène Myc22 (K^+)¹⁴⁶.

Dans la série à cœur benzénique, il est possible de retrouver des composés portant des chaînes latérales quinoliniums à l'image du composé **DTBQ** (en. *1,3-DiTriazolylBenzene Quinolinium*). Illustré par E. Mendes et al.¹⁴⁷ lors de l'étude de nouveaux composés ciblant les G-quadruplex, DTBQ rassemble les motifs DTB (Ditriazolylbenzène) capables d'interagir avec

la tétrade externe de la séquence télomérique humaine de structure parallèle¹⁴⁸ ainsi que le motif méthylquinolinium retrouvé un bon nombre de fois au sein de ligands de structures G-quadruplex afin d'apporter des interactions hydrophobes et électrostatiques¹⁴⁹. Ce composé aux propriétés très « *drug-like* » a montré des capacités de stabilisation d'environ 15°C vis-à-vis des cibles Telo21 et Myc19 ainsi qu'une bonne sélectivité pour le duplex d'ADN. Le motif hétérocyclique triazine a également fait l'objet d'attention particulière pour la synthèse de composé ciblant les édifices G-quadruplex. La triazine **12459** portant une nouvelle fois les motifs méthylquinoliniums comme chaînes latérales, a montré i) sa capacité d'inhibition de l'activité télomérasique par stabilisation d'édifices G-quadruplex télomériques et ii) son moyen d'altérer sélectivement le phénomène d'épissage au sein de hTERT⁹² par stabilisation de structures G-quadruplex au sein de l'intron 6.

Section 1.4.3 : Systèmes non aromatiques.

La description des ligands de structures G-quadruplex fait généralement intervenir de larges composés aromatiques tels que ceux décrits précédemment. Dans des cas plus rares, des structures ignorant les empilements π (ou tout au moins présentant des empilements π moins importants) peuvent reconnaître les édifices G-quadruplex (Figure 19, C/). La reconnaissance est principalement effectuée au sein des sillons et est principalement décrite par des interactions de types donneuses/acceptrices de liaisons hydrogène, des interactions électrostatiques ou encore des forces de Van der Waals. En titre d'exemple, des dérivés Peimine et Peiminine ont été décrits et brevetés par Tang et *al.* en 2009¹⁵⁰. Il a été montré que ces composés étaient capables de stabiliser la structure G-quadruplex télomérique [d(TTAGGGT)₄], appelée Telo7 par les auteurs, de 13 et 14°C respectivement. Des études cellulaires sur cellules cancéreuses HepG-2 ont également montré que ces composés étaient capables d'inhiber 50 à 80% de leur prolifération à une concentration de 50 $\mu\text{g/mL}$ ¹¹⁸.

Section 1.4.4 : Systèmes macrocycliques.

Certains ligands prometteurs de structures G-quadruplex sont des macrocycles plans dont la taille représente approximativement celle d'une tétrade de guanines. Ces ligands peuvent être classés en trois catégories principales (Figure 19, D/), i) la telomestatin, ces dérivés et des macrocycles formés par formation de liens amides, ii) les porphyrines et ces analogues et enfin iii) les dérivés de phénanthroline.

La **telomestatine** est un composé naturel cyclique isolé de *Streptomyces anulatus* (3533-SV4)¹⁵¹. Ce macrocycle de 24 chaînons possède 7 motifs oxazoles ainsi qu'un motif dihydrothiazole et fait partie des inhibiteurs les plus puissants de la télomérase¹⁵² par stabilisation de la séquence G-quadruplex télomérique¹⁵³. De nombreux analogues ont été synthétisés parmi lesquels des liaisons amides ont été insérées afin de conférer une plus grosse flexibilité au système (HXDV/L2G2-6OTD¹⁵⁴, L1H1-7OTD¹⁵⁵). Ces changements ont notamment permis d'augmenter l'hydrophobicité des molécules et d'améliorer les propriétés d'interactions des ligands¹⁵⁴ pour leurs cibles G-quadruplex.

Les porphyrines, et plus généralement les structures comportant des polymotifs pyrroles, sont caractérisées par une haute hydrophilicité ainsi qu'une grande capacité d'empilement π . Le composé de référence dans cette famille est la porphyrine **TMPyP4**, capable à la fois d'interagir avec l'ADN G-quadruplex télomérique avec un IC_{50} de 6 μ M et de réguler à la baisse l'expression de plusieurs oncogènes (Myc⁶¹, Kit1¹³³, RET¹⁵⁶, KRAS⁶¹). Toutefois, des études ont montré que ce ligand présentait une faible sélectivité pour les structures G-quadruplex vis-à-vis de l'ADN duplex¹⁵⁷. De nombreux dérivés ont par la suite été synthétisés afin d'améliorer les caractéristiques de la porphyrine. En ce sens, la tetrapyridinoporphyrazines (TMPyPz) est un analogue phthalocyanine non symétrique présentant une affinité 100 fois supérieure que TMPyP4 pour la séquence télomérique G-quadruplex (170 \pm 20 nM) et une meilleure sélectivité pour la séquence G-quadruplex par rapport au duplex d'ADN (facteur 30)¹⁵⁸.

Section 1.4.5 : Systèmes organométalliques.

Au sein de la famille de ligands de structures G-quadruplex, il est nécessaire de mentionner les complexes métalliques (Figure 19, E/). Ceux-ci présentent généralement de larges surfaces planaires coordinées avec des métaux cationiques permettant d'améliorer les propriétés d'association, mais également la délivrance des composés¹⁵⁹. Suivant la nature des métaux coordinés aux ligands, les modes d'interactions peuvent être i) directs, par co-coordination du métal au ligand mais également à la structure G-quadruplex, ou ii) indirects, par interactions du ligand coordonné à son métal avec les différentes régions des structures G-quadruplex (tétrades, sillons, boucles). Ayant mentionnée les ligands porphyrines et plus particulièrement le ligand **TMPyP4**, celui-ci peut également coordiner des métaux tels que le cuivre, l'or ou le manganèse afin de présenter des constantes d'associations bien meilleures que son composé parent. En effet, le remplacement des quatre groupements méthylpyridinium du TMPyP4 par des groupements cationiques flexibles (Figure 19, E/ Mn (III) complexe) en présence du **Manganèse** (Mn³⁺) a permis d'augmenter la sélectivité relative de TMPyP4 d'un facteur 10000 avec une constante d'affinité de l'ordre de 10 nM¹⁶⁰. De manière similaire, le dérivé de la porphyrine coordonné par de l'or (Au³⁺; Figure 19, E/ Au (III) complexe) présente un ΔT_m de 32,4°C avec une IC_{50} de 40 nM (réalisée par une méthode de TRAP-G4).¹⁶¹

D'autres complexes métalliques ne faisant pas intervenir des macrocycles ont également été recensés. Ainsi, des ligands polydentates à l'exemple du salophène et des bipyridines ont été décrits comme ligands de structures G-quadruplex. Le **salophène de platine** (Pt²⁺) a montré une capacité stabilisatrice de 25°C vis-à-vis de la séquence télomérique F21T et la **bipyridine de Nickel** (Ni²⁺) a montré une certaine sélectivité pour les G-quadruplex présentant une topologie parallèle ou hybride (EC_{50} = 0,9 μ M)¹⁶².

Section 1.4.6 : Biomolécules et dérivés.

Outre les composés organiques, ou organométalliques, présentés précédemment, la spécificité et la sélectivité des G-quadruplex peuvent être adressées en incorporant des motifs additionnels tels que des peptides, des nucléotides, des glycosides ou encore des structures originales telles que des acides nucléiques peptidiques (Figure 19, F/).

Paragraphe 1.4.6.a) Conjugués peptidiques, nucléotidiques et glycosidiques.

Les conjugués peptidiques, nucléotidiques ou glycosidiques ont été décrits afin de pouvoir essayer de mieux discriminer les différentes conformations de structures G-quadruplex¹⁶³. Des **conjugués peptidiques d'acridines**^{164,165} ou **d'hémicyanine**¹⁶⁶ ont montré des valeurs de stabilisation à l'échelle du micromolaire avec des sélectivités pour les séquences G-quadruplex (hTelo et/ou Kit1, Kit2, N-RAS) supérieures d'un facteur 40 par rapport au duplex d'ADN (Figure 19, F). Les acridines ont également été utilisées afin de synthétiser une librairie de conjugués oligonucléotidiques capables d'interagir avec les édifices G-quadruplex de la séquence télomérique humaine. L'étude réalisée par J. Casals *et al.* vise à faire hybrider les séquences oligonucléotidiques complémentaires des conjugués à la séquence riche en guanines des télomères afin d'apporter le ligand stabilisateur, l'acridine, près de la séquence G-quadruplex¹⁶⁷. Le groupe a ainsi montré que l'oligonucléotide **CTAACCT**, lié par un espaceur bis (ethylene glycol), à une **acridine** fonctionnalisée (acide isonipecotique et une pyrrolidine) était capable de stabiliser à hauteur de 5°C des séquences G-quadruplex. Enfin, les G-quadruplex peuvent également interagir avec des dérivés glycosidiques à l'exemple du composé **NCQ**, une quinacridine fonctionnalisée par la néomycine¹⁶⁸. Le groupe de M. Kaiser *et al.* a montré que ce composé était préférentiellement capable de se lier à des structures G-quadruplex exhibant des boucles avec des valeurs de stabilisation d'environ 14°C vis-à-vis de la séquence télomérique F21T.

Paragraphe 1.4.6.b) « Peptide-nucleic acids ».

Afin de proposer des ligands apportant une meilleure reconnaissance pour sa cible G-quadruplex, les dérivés acides nucléiques peptidiques (en. *peptid nucleic acid*, PNA) ont également été discutés dans la littérature. Dans le cas des acridones, le groupe de S. Ladame a montré que la conjugaison du polyhétérocycle avec une structure PNA de guanines permettait leur interaction avec le fragment télomérique d'ADN simple brin (topologie hybride type 3+1) par formation d'un hybride PNA/ADN¹⁶⁹. Dans le cas de cette étude, le site d'interaction du PNA-acridone est effectué au niveau des deux tétrades externes de la structure hybride G-quadruplex. Ce mode d'interaction a permis de mesurer une bonne stabilisation ($\Delta T_m = 13^\circ\text{C}$) dans des conditions faiblement salines.

Les familles non exhaustives de composés décrites précédemment ont grandement contribué à la compréhension des fonctions biologiques des structures G-quadruplex. Les différentes familles de ligands, ainsi que les ligands eux-mêmes, présentent différents niveaux de sélectivité. Il est en effet possible de discuter de sélectivité d'un ligand pour une structure G-quadruplex par rapport à d'autres formes d'ADN (notamment l'ADN duplex) ou, de la sélectivité d'un ligand par rapport aux différentes topologies (parallèle, antiparallèle, hybride) de G-quadruplex. La prise en compte de la sélectivité, l'amélioration des propriétés de stabilisation et l'affinité du ligand pour les G-quadruplex sont autant de paramètres à prendre en compte afin de synthétiser de nouveaux composés ciblant les G-quadruplex. Les méthodes de synthèses classiques se heurtent parfois à la difficulté de prendre en compte tous ces paramètres afin de faire émerger de nouveaux ligands possédant toutes les propriétés

souhaitées. Dans ce contexte, les méthodes combinatoires permettant de cribler de vastes diversités chimiques à plus grande échelle et à plus haut débit suscitent un intérêt croissant.

Sous-partie 1.5 : Méthodes combinatoires pour l'identification de ligands de structures G-quadruplex.

Traditionnellement, le développement de ligands passe par la synthèse et l'évaluation individuelle de plusieurs composés vis-à-vis d'une cible biologique donnée. Ce processus est non seulement couteux et chronophage mais il ne peut être effectué que pour un nombre de composés raisonnable, restreignant l'espace chimique exploré. Le criblage biologique de composés permet d'identifier rapidement des entités chimiques intéressantes. Toutefois, cela nécessite des tests biologiques fiables et rapides ainsi qu'une librairie de composé. En ce sens, la « chimie combinatoire » tend à assembler des bibliothèques de composés en mélange afin d'être criblé à haut débit sur des cibles biologiques pour identifier de nouveaux pharmacophores. Les méthodes des criblages à haut-débit et de chimie combinatoire décrites ci-après récapitulent les différentes approches utilisées afin d'identifier de nouveaux ligands des cibles ADN (ou ARN) G-quadruplex.

Section 1.5.1 : Criblage en parallèle et approche in silico.

La découverte de nouvelles molécules capables d'interagir avec les structures ADN (ou ARN) G-quadruplex peut être effectuée par le criblage intensif (haut-débit) de molécules connues. Celles-ci proviennent généralement de bibliothèques nationales, de bibliothèques commercialisées par différents groupes industriels ou de bibliothèques « *in-house* » propres à des unités de recherches. Ces criblages à haut-débit peuvent être réalisés de différentes manières suivant la taille de la bibliothèque considérée. Il est ainsi possible d'explorer des interactions ligands/G-quadruplex en utilisant par exemple de la RMN, la MS, la SPR, ou encore l'ITC. Toutefois, les techniques les plus utilisées reposent sur les méthodes utilisant de la fluorescence à l'exemple du FRET ou d'étude sur microarray. En 2019, K. Wang *et al.*¹⁷⁰ ont criblé pas moins de 50k molécules en utilisant une méthode de FRET à haut-débit. Ces composés (10µM) dérivants de deux bibliothèques existantes (*Diversity-oriented small molecule library* et *know drugs, bioactive and natural products library*¹⁷⁰) ont ainsi été testés pour leurs propriétés stabilisatrices en plaque 384-puits contre la cible G-quadruplex doublement marqué Bpu28F (0.5µM, ⁵-BHQ1-TGGGGAGGGTGGGGAG GGTGGGGAAGGTT-FAM-³). Après analyse des résultats de fluorescences de plus de 150 plaques 384-puits, 329 composés sont apparus comme étant capables de stabiliser l'édifice G-quadruplex formé par Bpu28F. Pas plus d'informations ont été transmises sur l'identité des ligands trouvée.

La recherche des propriétés de ligands vis-à-vis d'une cible quadruplex par méthodes de fluorescences ou par des méthodes nécessitant l'immobilisation des cibles sur surfaces peut entraîner un biais à la caractérisation tant les modifications engendrées peuvent altérer ou masquer certains sites d'interactions. Une manière de dépasser ces limitations est d'utiliser la spectroscopie de masse pour du criblage de haut-débit. Le groupe de S.-N. Richter¹⁷¹ a mis au point une méthode d'analyse directe leur permettant de cribler 40k composés contre deux

séquences issues du promoteur LTR du VIH-1 (LTR-III et LTR-IV). Ces études ont montré que le composé J20 (Figure 20) était non seulement capable de reconnaître spécifiquement le G-quadruplex du promoteur LTR-III ($\Delta T_m=3.9^\circ\text{C}$, $K_d=11\mu\text{M}$) mais qu'il était également capable d'induire une activité anti-VIH-1.

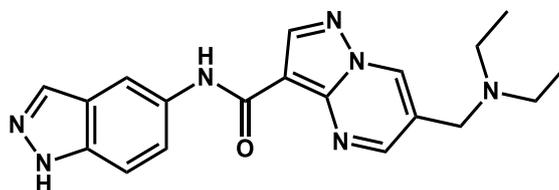


Figure 20 : Criblage conventionnel à haut-débit. Découverte du ligand J20 ciblant le G-quadruplex VIH-1 au niveau du promoteur LTR-III¹⁷¹.

Parallèlement au criblage conventionnel réalisé *in vitro* sur les cibles biologiques d'intérêts, du criblage virtuel (en. *virtual screening*, VS) est également une technique de choix pour l'identification de nouveaux ligands. Historiquement utilisée dans le cas de la recherche de molécules ciblant les protéines, le VS converge ces efforts depuis une vingtaine d'années vers le ciblage de petites molécules interagissant avec des structures d'acides nucléiques dont les structures G-quadruplex.

L'utilisation de bibliothèques dans des campagnes de criblages conventionnels à haut-débit engendre nécessairement des coûts de productions élevées ainsi que l'utilisation de procédés techniques onéreux. Dans le cas du VS *in silico*, le criblage à haut débit s'affranchit des coûts précédemment cités tout en explorant une diversité chimique encore plus complète pour l'identification de composés d'intérêts. Cette méthode nécessite néanmoins quelques prérequis primordiaux tels que la structure hautement résolue de la cible (RMN, cristallographie aux rayons X, etc.) et/ou des informations structurales de(s) ligand(s), endogène(s) ou non, biologiquement actif(s) pour la cible. Lorsque le VS est réalisé à partir d'informations connues sur une cible, on parlera d'une approche structurale. A l'inverse, lorsque le VS se base sur l'identité du ligand connu pour sa cible, on parlera de VS par approche pharmacophore.

Le VS par approche structurale peut être réalisée à partir des données recueillies en ligne (PDB, NDB). Cette méthode a conduit D.L. Ma *et al.*¹⁷² à s'intéresser à des ligands ciblant la structure ADN G-quadruplex télomérique (PDB id : 1KF1). Plus de 100k ligands dérivant d'une bibliothèque *drug-like* (suivant les règles de Lipinski) ont été scannés sur l'édifice G-quadruplex. Parmi les composés interagissant avec la structure G-quadruplex, le composé 1 h-pyrazole-3 — carboxy-4 — methyl-5 — phényl-(1H-indol-3-ylmethylene) hydrazide (Figure 21, gauche) a obtenu le meilleur score. L'étude *in vitro* de ce composé par FRET-Melting a pu confirmer qu'il était capable de stabiliser la structure télomérique avec un ΔT_m de $+18^\circ\text{C}$ tout en adressant une excellente sélectivité vis-à-vis de l'ADN duplex.

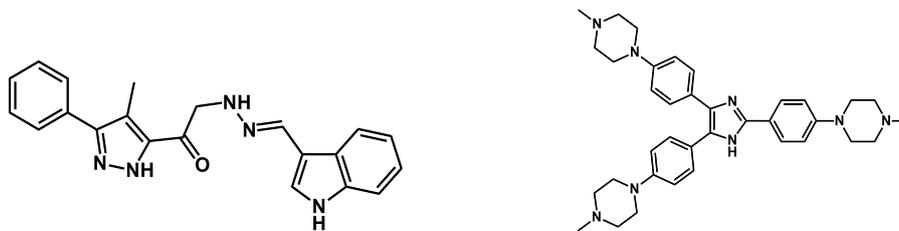


Figure 21 : Criblage *in silico*. Découverte des ligands Indolhydrazide (gauche) et TSIZ01 (droite) ciblant le G-quadruplex de la séquence télomérique humaine F21T.

La technique de VS par approche structurale souffre de limitation dès lors que la cible étudiée peut adopter plusieurs conformations, à l'exemple de la séquence ADN G-quadruplex du télomère humain (topologie parallèle, PDB ID: 1KF1 ; antiparallèle, PDB ID : 143D ; hybride, PDB ID : 2HY9). Afin de contrecarrer cette limitation, le VS est réalisé par l'approche pharmacophore. En effet, alors que l'approche structurale nécessite une connaissance poussée de la structure de la cible, l'approche pharmacophore nécessite des connaissances précises du ligand. En se basant sur les travaux de M.J. Moore *et al.* axés sur la synthèse d'acridines trisubstituées¹³⁰ ciblant les G-quadruplex télomériques, l'équipe de S.-B. Chen *et al.*¹⁷³ a scanné 5000 composés d'origines naturels ainsi que leurs dérivés. Parmi cette librairie, 31 molécules possèdent les mêmes types d'interactions que peuvent exhiber les acridines trisubstituées (empilement π , liaisons hydrogène, etc.). Une nouvelle classe de composé a alors émergé avec la découverte du composé triaryl-imidazole, TSIZ01 (Figure 21, droite). Des expériences complémentaires de caractérisations biophysiques (CD, SPR) ont été réalisées. TSIZ01 montre une capacité de stabilisation de 23,5°C contre la cible F21T, une très bonne sélectivité pour le duplex d'ADN, et une constante d'association mesurée de 463 nM.

Section 1.5.2 : Chimie combinatoire.

Outre les méthodes de criblages conventionnelles, réalisées à partir de librairies disponibles, ou de criblages *in silico*, nécessitant de solides compétences en chémoinformatique, d'autres techniques combinatoires peuvent être utilisées afin de synthétiser et cribler à haut débit des librairies de composés sur les G-quadruplex.

Parmi les techniques de synthèse combinatoire il est possible de mentionner la méthode *one-bead-one-compound* (OBOC)¹⁷⁴. Ce concept est basé sur de la synthèse en phase solide, telle que décrite par Merrifield en 1963. En effet, des composés uniques vont être individuellement synthétisés sur support solide (généralement des billes) *via* l'utilisation de synthèse fractionnée appelée *split & mix synthesis* (Furka, 1995)¹⁷⁵. L'originalité de cette méthode est que l'identification des propriétés de chaque ligand de la librairie peut être interrogée individuellement par mesure biochimique ou spectroscopique.

La technique OBOC a notamment été utilisée par J.E. Redman¹⁷⁶ pour la synthèse de polyamides non naturels ciblant la structure ADN G-quadruplex télomérique (hTelo). Dans cette étude, trois librairies de 36 séquences peptidiques non naturelles immobilisées sur billes ont été criblées et les expériences de titration ELISA ont permis de caractériser de fortes affinités pour 8 composés distincts. Des études complémentaires d'affinités ont permis

d'obtenir par SPR des K_d de 2 μM pour le composé 8 (WKKR) et 4 μM pour 4 (Y-Im-Py-R ; Figure 22). L'effet stabilisateur, mesuré par FRET-Melting, a montré que le composé 8 ($\Delta T_m=21,6^\circ\text{C}$) stabilisait plus fortement hTelo que le dérivé peptidique non naturel 4 ($\Delta T_m=15,1^\circ\text{C}$). Néanmoins, malgré de bonnes caractéristiques vis-à-vis des G-quadruplex, les composés 8 et 4 ne présentent pas de grande sélectivité vis-à-vis de l'ADN duplex.

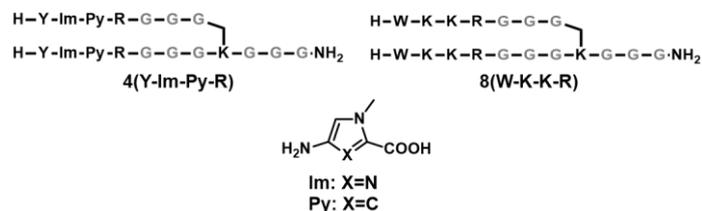


Figure 22 : Chimie combinatoire / stratégie One-bead-one-compound. Découverte de peptides (modifiés) ciblant la structure télomérique G-quadruplex — hTelo.

Grandement utilisée dans le cadre de la recherche de ligands à hautes affinités ciblant les sites actifs des protéines ou l'interface protéine-protéine, le criblage par fragment (en. *Fragment-based*, FB) ciblant les acides nucléiques reste une technique largement sous-exploité¹⁷⁷.

Cette méthode repose, au préalable, sur la sélection de différents fragments chimiques (masse moléculaire généralement <300Da, $\log P \leq 3$, accepteur/donneur de liaisons H ≤ 3) par criblage haut-débit d'une librairie de composé sur la cible d'intérêt. Après identification des fragments les plus affins (RMN, SPR, ITC, etc.), ceux-ci peuvent être étendus et/ou combinés afin de générer des molécules avec de plus hautes propriétés d'associations (affinité et/ou spécificité). Une connaissance structurale précise de la cible étudiée peut dans certains cas orienter la combinaison des fragments.

Après adaptation de la méthode de criblage par fragment contre la cible G-quadruplex ADN du promoteur d'oncogène Myc¹⁷⁷, plusieurs groupes se sont intéressés à cette technique afin d'isoler de nouveaux composés intéressants. Les travaux de M. Tassinari *et al.*¹⁷⁸ ont notamment permis de trouver deux dérivés pyridines-amidoximes-oxadiazole (PAO1 et PAO2, respectivement ; Figure 23) ciblant le G-quadruplex formé au niveau du promoteur LTR-III du virus VIH-1 (Figure X).

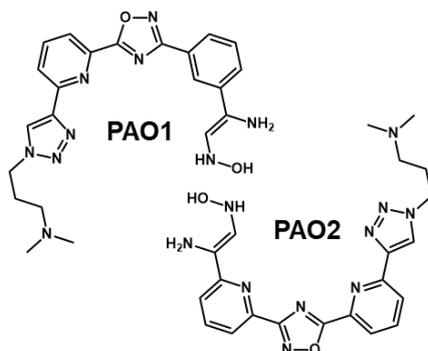


Figure 23 : Chimie combinatoire / stratégie Fragment-based. Synthèse de séquences pluri-aromatiques ciblant le promoteur LTR-III du virus VIH-1.

En effet, la première partie de ces travaux a nécessité la mesure des propriétés de stabilisation d'une librairie de 150 fragments, d'origines commerciales ou obtenues par synthèse, par expérience de fluorescence (FRET-Melting). Compte tenu des entités chimiques criblées, seules 5 d'entre elles ont été retenues. Des valeurs de stabilisation variant de 1 à 6°C ont ainsi pu être mesurées tout en exhibant une bonne sélectivité pour l'ADN duplex. Les motifs retenus présentent des cycles mono (hétéro) aromatiques (pyridine ou benzène) bi- ou trisubstitués par un acide carboxylique, ou son bioisostère amidoxime, un alcyne ou encore un dérivé pyrrolidine. À partir de la combinaison des 5 fragments identifiés, 17 composés uniques ont pu être synthétisés dont deux présentant une stabilisation respectivement de 10,8 et 12,8°C (4000 éq.) vis-à-vis de la séquence G-quadruplex LTR-III (0,25µM). Une dernière étape d'optimisations structurales a conduit aux composés PAO1 et PAO2 présentant respectivement des ΔT_m de 9 et 14,1°C (100 éq.). Des études biochimiques complémentaires ont été réalisées sur le composé PAO1 et ont montré sa capacité à inhiber la Taq polymérase lors d'une étude de *Taq Polymerase stop assay*. Cette dernière information conforte la forte capacité de PAO1 à stabiliser l'édifice G-quadruplex de LTR-III.

Un nombre important de groupes de recherche, industrielle ou académique, ont exploré la synthèse chimique ciblée (en. *Target-guided synthesis*, TGS) afin d'identifier de nouveaux composés intéressants. Complémentaire aux outils traditionnels de *drug discovery*, cette approche vise également à réduire le nombre de composés inactifs au sein de campagne de criblage tout en compensant le manque de prédictibilité d'un criblage *in silico*. Au sein de cette approche, deux catégories peuvent être discutées. La première catégorie permet à la cible de déplacer l'équilibre chimique des motifs d'une librairie en mélange afin de ne former thermodynamiquement que les composés les plus affins pour cible étudiée. Cette méthode est connue sous le nom de chimie combinatoire dynamique (en. *dynamic combinatorial chemistry*, DCC). Dans un deuxième temps, la cible peut être utilisée en tant que catalyseur cinétique de composés en mélange. Cette méthode connaît le nom de synthèse cinétique par la cible ou *Kinetic target-directed synthesis* (KTDS).

La DCC est une approche de chimie combinatoire permettant la génération de librairies de composés (en. *Dynamic combinatorial libraries*, DCL) par la réaction réversible d'entités chimiques et réalisées sous contrôle thermodynamique. Cette méthode a été adaptée plusieurs fois pour la recherche de ligands de structures G-quadruplex, à l'exemple des travaux réalisés il y a près de 15 ans par A. Bugaut *et al.*¹⁷⁹ ou plus récemment par O. Reznichenko *et al.*¹⁸⁰.

Dans cette première étude, un peptide macrocyclique hétéroaromatique portant une fonction thiol a été utilisé afin de cibler spécifiquement deux séquences G-quadruplex intramoléculaires (c-Kit21 et c-Myc22). Considéré seul, le macrocycle exhibe des constantes d'affinités (SPR) respectives de 67,5 et 82,5 µM. L'approche par DCC a eu pour intérêt de former les meilleurs équilibres entre le macrocycle et différents dérivés benzylmercaptan afin d'améliorer les propriétés d'affinités du ligand. Le macrocycle, les dérivés benzylmercaptan ainsi que la structure G-quadruplex ont été incubés en présence d'un tampon permettant les échanges de ponts disulfures. La composition de la DCL a été analysé par UV-HPLS-MS et les ligands les

plus prometteurs ont été resynthétisés et caractérisés par SPR. Ainsi, l'édifice G-quadruplex de la cible c-Myc22 a permis de caractériser par DCC la formation du composé M1 avec une affinité de 6,8 μM . une baisse d'affinité de plus d'un facteur 12 a ainsi été obtenue par formation d'un pont disulfure entre le motif 1 — (4-(mercaptométhyl)benzyl) guanidine et le macrocycle de départ (Figure 24, gauche). *A contrario*, la cible c-Kit21 a présenté la meilleure interaction avec le macrocycle M2 exhibant le motif N-carbamimidoyl-4 — (mercaptométhyl) benzamide ($K_D=6.6 \mu\text{M}$; Figure 24, droite).

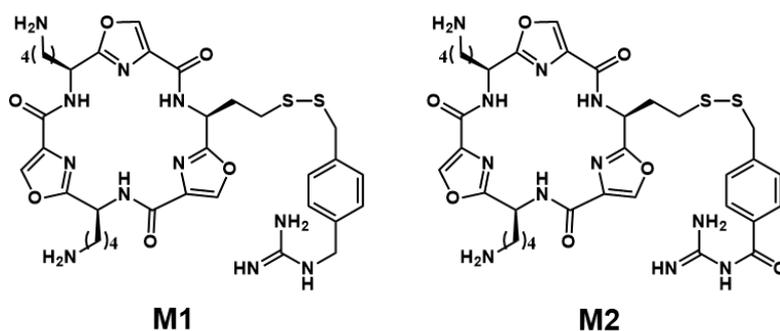


Figure 24 : Chimie combinatoire / Méthode par librairie combinatoire dynamique. Synthèse de macrocycles polyoxazoles ciblant le promoteur d'oncogène Myc (Myc22).

La méthode de synthèse chimique ciblée par contrôle cinétique, ou KTDS, utilise la cible afin de pouvoir faire réagir de manière irréversible deux entités chimiques en mélange. Dans cette approche, les réactifs s'associent au sein d'un site spécifique de la cible et la réaction irréversible est permise par proximité spatiale des deux entités. La réaction irréversible la plus utilisée au sein de la méthode KTDS est la réaction *in situ* de cycloaddition click entre un azoture et un alcyne. Cette méthode a été illustrée dans les travaux de M. Di Antonio *et al*⁸¹. Au cours de cette étude, le groupe a utilisé deux dérivés Pyridostatine-alcyne (le dérivé Amino-PDS (alk) et le dérivé Pyrrolidine-PDS (alk), noté PDS₁ et PDS₂ respectivement) et 6 dérivés azotures possédant les motifs éthylamine, propylamine, glucopyrannose, éthylpyrrolidine, acide éthanoïque et acide benzoïque. Dans les conditions classiques de cycloaddition 1,3 — dipolaire, la réaction est catalysée par le Cu (I) afin de ne former que l'isomère 1,4. Toutefois dans cette expérience, la catalyse est effectuée par l'édifice télomérique G-quadruplex *via* le rapprochement spatial des deux réactifs. Ainsi, 24 molécules distinctes sont susceptibles d'être obtenues (PDS₁/PDS₂, 6 azotures et la formation des deux régioisomères 1,4 et 1,5 dus à l'absence de cuivre dans la réaction). Après incubation de tous les partenaires dans les conditions décrites précédemment, seul un dérivé a été spécifiquement formé par catalyse de l'ADN G-quadruplex télomérique, le glucopyrannose-PDS₁. L'ajout de cuivre au milieu réactionnel a permis d'observer la formation des 24 ligands. Les propriétés du ligand glucopyrannose-PDS₁ ont été évaluées par FRET-Melting et une valeur de stabilisation de 28°C a été mesurée avec une sélectivité totale vis-à-vis de l'ADN duplex.

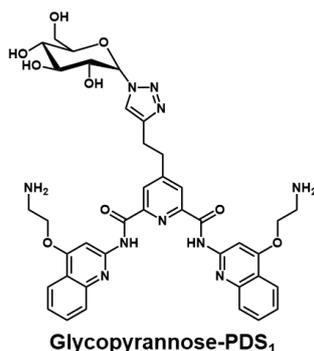


Figure 25 : Chimie combinatoire / Méthode de Kinetic target-directed synthesis. Synthèse d'un dérivé glycosidique de la Pyridostatine (PDS) afin de cibler l'ADN G-quadruplex télomérique (*hTelo*).

Des études complémentaires de ce ligand ont montré qu'il était capable de réduire le nombre de *foci* TRF1 ($IC_{50} = 1,3 \mu M$) permettant ainsi d'affirmer l'interaction du Glycopyrannose-PDS₁ avec l'ADN G-quadruplex télomérique (Figure 25).

Section 1.5.3 : « Display technologies ».

Les technologies *display* englobent une collection de techniques permettant la production de bibliothèques macromoléculaires complexes incluant des peptides, des anticorps monoclonaux ou encore des aptamères¹⁸²⁻¹⁸⁴. Ces méthodes reposent sur l'évolution dirigée (en. *directed evolution*) des membres de la bibliothèque par criblage itératif ainsi que sur la mise en relation d'un phénotype à un génotype. Indépendamment de sa nature, les composés distincts d'une bibliothèque en mélange porteront une entité génomique unique. Celle-ci est de nature variée à l'exemple de phage, de ribosome, d'ADN ou d'ARN. Ci-après sont représentées des techniques *display* permettant l'identification de composés ciblant les structures G-quadruplex.

Le phage *display* est une technologie cellulaire permettant de produire des bibliothèques de peptides ou des protéines à la surface de bactériophages. Ces derniers sont des virus génétiquement modifiés capables d'infecter des bactéries et de contenir le matériel génétique nécessaire à la production de bibliothèque de peptides (ou protéines). Dans le cas du phage *display*, les protéines de l'enveloppe du phage sont génétiquement modifiées pour exprimer de manière hétérologue des peptides à la surface de la cellule virale tout en contenant l'information génétique qui code pour ces protéines à l'intérieur de la capsid du virus, reliant ainsi le phénotype au génotype.

En 2013, G. Biffi *et al.*³⁰ a utilisé une bibliothèque issue de l'institut Sanger¹⁸⁵ contenant $2.3E+10$ phages afin de cibler des ADN G-quadruplex humains intra- et intermoléculaires. Ces phages, exhibant à leur surface une variété d'anticorps à chaîne unique (en. *single-chain antibody*), ont été sujets à deux tours successifs de sélection contre des structures G-quadruplex immobilisés sur billes magnétiques. Parmi les milliards de composés, le meilleur anticorps appelé BG4 a été isolé. Des méthodes de caractérisations par test ELISA ont montré de bonnes affinités pour les G-quadruplex intra- et intermoléculaires (0.5-1.6 et 2 nM, respectivement). De plus, aucune affinité n'a pu être mesurée contre des *hairpins* d'ARN, des ADN simples-brins ou encore l'ADN duplexe. Démontrant une sélectivité indiscutable vis-à-vis des autres structures

d'acides nucléiques, la sélectivité de BG4 a été évaluée pour les différentes confirmations de structures G-quadruplex. Les cibles parallèles des proto-oncogènes Myc et Kit ainsi que la cible antiparallèle TBA ou encore la structure hybride G-quadruplex télomérique ont été utilisés. Néanmoins, avec des constantes d'affinités similairement proches pour toutes les topologies de structures G-quadruplex, BG4 ne semble pas capable d'interagir de manière sélective avec les G-quadruplex de topologies variés. Après les études réalisées sur BG4, H. — Y. Liu *et al.*¹⁸⁶ a criblé 1.37E+08 phages contre les antigènes G-quadruplex parallèles Myc et BCL-2, la cible antiparallèle TBA et la séquence hybride télomérique humaine. Trois étapes de sélections d'anticorps successives ont été réalisées sur chaque cible et au total 14 anticorps à chaîne unique ont pu être identifiés par test ELISA. Parmi ceux-ci, l'anticorps D1 était capable de reconnaître spécifiquement les G-quadruplex parallèles au détriment des autres topologies de G-quadruplex avec une constante de dissociation de 17 nM. En complément, une excellente spécificité vis-à-vis des autres structures secondaires de l'ADN (ADN duplexe, *Hairpin*, ADN simple brin) a pu être observée. Les deux exemples de sélection d'anticorps par phage *display* a montré qu'il était possible de distinguer d'intéressantes molécules (K_d au nM) au sein de la famille des biomolécules et qu'il a été possible d'apporter de la sélectivité entre les différentes topologies que peuvent adopter les structures G-quadruplex.

Dans le cadre de la recherche de molécules ciblant les structures G-quadruplex, l'évolution dirigée peut être adressée en utilisant une technique de ribosome *display*. Contraire au phage *display*, le ribosome *display* lie les phénotypes à un complexe ribosomique composé d'ARNm et du ribosome. Simplement, une librairie d'ADN codant pour différentes séquences d'anticorps est transcrite *in vitro* en ARNm. Ceux-ci vont être traduits en protéines (*i. e* anticorps) par l'activité enzymatique du ribosome. Toutefois, les ARNm traduits sont démunis de codon-stop, impliquant ainsi la liaison du complexe ternaire ARNm-ribosome-anticorps. Ce complexe va ainsi être utilisé afin de réaliser du criblage à haut-débit sur la cible d'intérêt.

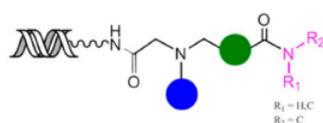
L'équipe de A. Plückthun¹⁸⁷ a utilisé cette technique pour générer des anticorps spécifiques des ADN G-quadruplex télomériques chez les ciliés (*Stylonychia lemnae*). À partir de 2E +09 séquences codant pour différents anticorps à chaîne unique et six sélections successives contre la cible parallèle télomérique dT (T₄G₄)₄ le groupe a pu caractériser deux anticorps, nommé Sty3 et Sty49, hautement spécifique des séquences G-quadruplex. La mesure des K_d par BIAcore a permis d'obtenir des valeurs d'affinités de 126 pM et 2,96 nM contre la cible parallèle télomérique décrite précédemment. Néanmoins, les deux anticorps ont également présenté des affinités de l'ordre du nM contre la cible télomérique antiparallèle d (G₄T₄)₄G₄ montrant ainsi un manque de spécificité entre topologies de structures G-quadruplex.

Les technologies *display* permettent la synthèse de biomolécules avec un criblage à grande échelle pour l'identification *in vitro* rapide de composés d'intérêts. Cependant, les approches décrites ci-dessus ne sont pas compatibles avec l'assemblage de librairies de petites molécules. Une technologie émergente et qui permettant d'associer synthèse organique et encodage génotypique est la technologie connue sous le nom de chimie encodée par l'ADN (en. DNA *Encoded chemistry*). Imaginée au début des années 90, cette technique est aujourd'hui largement utilisée dans le processus de découverte de composés actifs. Ciblant

essentiellement les protéines, certaines bibliothèques chimiques encodées (en. *DNA-Encoded chemical Library*, DEL) ont été utilisées afin de cibler les acides nucléiques, notamment les cibles G-quadruplex.

A ce jour, seul *X-Chem Pharmaceuticals* a rapporté l'utilisation de chimie encodée par l'ADN afin de cibler les structures G-quadruplexes¹⁸⁸. Dans cette étude, le groupe a réutilisé 2 des 33 DEL déjà confectionnées afin de cibler des protéines telles que des tyrosines kinases ou encore des récepteurs couplés aux protéines G. La première bibliothèque, appelée bibliothèque A, est une bibliothèque linéaire de trisynthons dotée de 217.6 millions de membres. La deuxième bibliothèque, appelée bibliothèque B, est une bibliothèque d'imidazopyrimidine dérivatisée en deux points et composée de 1.275 millions de molécules uniques (Figure 26).

Library A: Linear w/ Reversed Formyl Acids



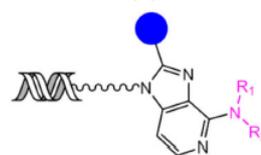
Cycle B: 1024 primary amines/anilines

Cycle C: 85 Formyl Acids

Cycle D: 2500 Amines

217,600,000 trisynthons

Library B: Imidazopyrimidine



R₁, R₂ = C

Cycle B: 1051 Secondary Amines

Cycle C: 1213 Aldehydes

1,274,863 disynthons

Figure 26 : Représentation des bibliothèques combinatoires encodées A et B ciblant les structures G-quadruplex. Figure issue des travaux de Litovchick et al.¹⁸⁸.

Les bibliothèques ont ensuite été criblées, de manière indépendante, contre différentes cibles G-quadruplex issues de la séquence télomérique humaine ou encore des promoteurs d'oncogènes Myc et Kit. Deux tours de sélections successifs ont ainsi permis de sélectionner les composés les plus affins vis-à-vis des différentes cibles. L'identité des composés sélectionnés a ensuite pu être déterminée par amplification des code-barres ADN et séquençage à haut-débit selon un processus que nous décrivons plus loin.

Dans chaque bibliothèque, un composé unique portant les meilleures caractéristiques a été resynthétisé en large quantité (DEL (A)-1 et DEL (B)-1 respectivement ; Figure 27) afin de pouvoir évaluer les constantes d'affinités par SPR.

Tableau 1 : Mesure des interactions G-quadruplex/ligands DEL¹⁸⁸ par SPR.

Cibles	DEL (A)-1	DEL (B)-1
C-myc G4 Pu-27	1.05 μM	328 nM
C-myc G4 1234	338 nM	59 nM
C-myc G4 2345	308 nM	197 nM
C-kit1 G4	1.1 μM	807 nM
C-kit2 G4	1.7 μM	1.4 μM
Télomère G4	8.4 μM	5.25 μM
HP006	nd	nd
Pu27rev	nd	nd

Figure 27 : Chimie combinatoire / Technologie display de librairie encodée par l'ADN. Synthèse d'un dérivé linéaire trifonctionnalisé ainsi qu'un composé imidazopyrimidine bifonctionnel ciblant le proto-oncogène Myc (Myc 1234).

DEL (A)-1 exhibe un K_d de l'ordre de 300 nM pour les cibles G-quadruplex Myc comportant les tétrades 1234 ou 2345 (Pu27 : ^{5'} — TGGGG (1)AGGG (2)TGGGG (3)AGGG (4)TGGGG (5) AAGG-^{3'}) alors qu'une affinité au micromolaire a été observée pour la cible putative G-quadruplex Pu27. Cette différence d'affinité peut être expliquée par la présence de conformation multiple au sein de la séquence G-quadruplex Pu27. De manière similaire, le composé DEL (B)-1 présente une meilleure affinité pour les G-quadruplex à 4 plateaux de guanines. De plus, une variation d'affinité est observable suivant la séquence à 4 plateaux considérés. En effet, un K_d de 59 nM a été mesuré pour la séquence 1234 alors qu'une affinité de 197 nM a été retrouvée vis-à-vis de la séquence 2345. En comparant les résultats obtenus sur les autres topologies de structures G-quadruplex il est possible de constater que les ligands DEL (A/B)-1 semblent présenter une certaine sélectivité pour la cible Myc.

Cette étude, réalisée en parallèle de ces travaux de thèse, a démontré l'efficacité de l'approche DEL appliquée aux structures nucléiques et plus spécifiquement aux structures G-quadruplex. Dans le cadre de ce projet de thèse, cette approche associant synthèse organique et encodage unique par l'ADN a été utilisée afin d'adresser la synthèse de ligand spécifique de structures G-quadruplex. La partie 2 de ce chapitre introductif fait l'évaluation exhaustive de toutes les techniques sous-jacentes à cette technologie d'encodage par l'ADN.

Partie 2 : De la librairie encodée de polypeptides aux librairies encodées par l'ADN.

L'identification de nouveaux ligands organiques ciblant les protéines est généralement la première réflexion lors d'un processus de découverte de médicaments (en. *drug discovery*). Malheureusement, certain manque d'informations structurales de cibles biologiques génère couramment l'impossibilité de prédire les modes de liaisons d'un ligand et ceux malgré un avancement important des méthodes et moyens informatiques. Ainsi, depuis de nombreuses années, le processus de criblage à haut débit (en. *High-throughput screening – HTS*) est utilisé afin de cribler un très grand ensemble de molécules, couramment appelé librairie chimique. Cette technique de criblage permet de construire, stocker et tester chaque individualité d'une librairie vis-à-vis d'une cible souhaitée en utilisant des méthodes informatiques et automatiques¹⁸⁹⁻¹⁹². Il est également bon de rappeler que la technique de criblage à haut débit n'a évidemment pas supplanté les campagnes de criblages conventionnelles¹⁹³.

Le criblage à haut débit est une technique onéreuse, principalement utilisée par les compagnies pharmaceutiques depuis plusieurs décennies, permettant de cribler des librairies chimiques de plusieurs millions de molécules organiques les unes après les autres. En effet, il n'est pas envisageable de mixer tous les composés avec la cible d'intérêt puisqu'il serait impossible de déconvoluer spécifiquement les ligands attachés à leur cible sans s'emparer des individualités non souhaitées. En 2014, l'étude de R.A. Goodnow a montré que la confection de vastes librairies chimiques ainsi que leur criblage individuel coûtait approximativement 0,4 à 2 milliards de dollars, soit près de 1 000€ par composé¹⁹⁴.

En accord avec ces précédentes considérations, les chercheurs ont consacré des efforts considérables dans la découverte et le développement de méthodes facilitant l'identification de molécules ciblant les protéines. En ce sens, les bibliothèques chimiques encodées par l'ADN permettent la construction et le criblage de bibliothèque de taille sans précédent et permettent, par conséquent, la découverte de nouveaux ligands organiques pour un coût fortement réduit. L'accroissement successif des membres de la bibliothèque encodée engendre un effondrement des concentrations de chacune des molécules jusqu'à un certain point où les techniques d'analyse modernes ne seraient même plus capables de les détecter. En revanche, les code-barres ADN indexés à chacune des molécules de la bibliothèque vont permettre leur amplification, leur identification et leur quantification¹⁹⁵.

Comme discuté au précédent chapitre, la technique de phage *display* développée en 1985 est un outil important pour l'identification des interactions protéines-protéines puisqu'un criblage simultané de tous les phages peut être effectué en incubant la bibliothèque avec une cible d'intérêt afin d'extraire, par sélection d'affinité, les ligands les plus affins (Figure 28, gauche). Contrairement au HTS classique, les ligands sélectionnés peuvent être caractérisés par amplification et séquençage des code-barres ADN correspondant. De cette technique mêlant phénotype et génotype, respectivement anticorps de fragment variable à chaîne unique (en *single chain variable fragment* – *scFv* – *antibodies*) et plasmides, a découlé la première expérience de chimie combinatoire encodée en 1992¹⁹⁶. Les professeurs S.A. Brenner et R. A. Lerner ont eu l'idée de combiner de la diversité chimique avec la puissance de l'encodage génétique pour outrepasser les limitations du criblage à haut débit et assurer une méthode versatile et puissante pour la découverte de drogues (Figure 28, droite).

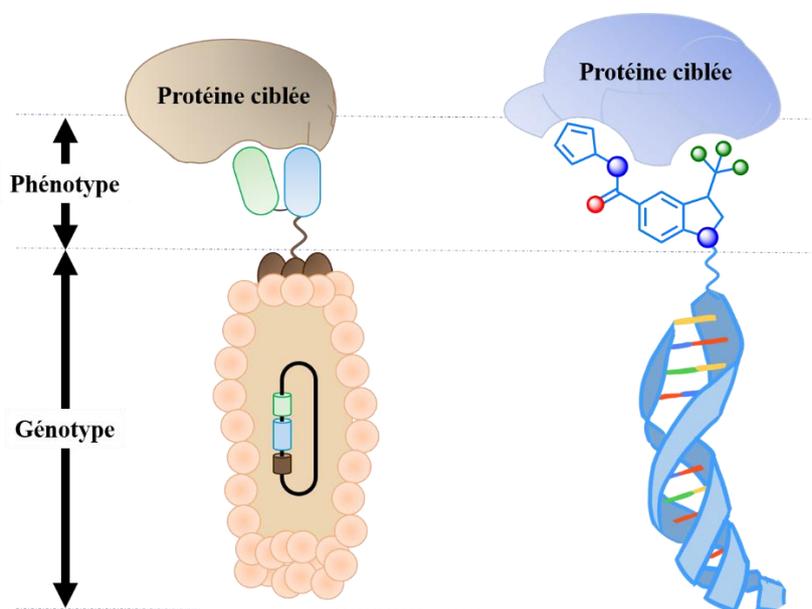


Figure 28 : Représentation schématique d'une bibliothèque obtenue par phage display et une bibliothèque obtenue par chimie combinatoire encodée par l'ADN. A gauche : interaction d'un phage filamentueux, présentant un anticorps à sa surface, avec une cible protéique. A droite : interaction d'une molécule organique encodée par un fragment d'ADN sur une cible d'intérêt.

Lors de ce premier pas vers la chimie encodée, les auteurs ont envisagé la synthèse simultanée de séquences polypeptidiques distinctes et d'oligonucléotides sur billes en utilisant des procédés de chimie orthogonale et des savoirs de chimie combinatoire (notamment la stratégie « split-and-pool »). Il a alors rapidement été imaginé la possibilité de remplacer les billes par un lien (en. *linker*) chimique robuste capable de connecter et préserver le phénotype et le génotype. Ainsi dans cette stratégie, il est possible d'associer à divers fragments (en. *building block*) peptidiques une séquence d'oligonucléotide qui, après itérations successives, va définir un brin d'ADN. Une fois formé, ce brin d'ADN va uniquement servir de séquence identifiant la chaîne peptidique associée et ne va en aucun cas servir de phénotype contre la cible étudiée. De ce fait, après sélection par affinité sur une cible de choix, les ligands les plus affins pourront être identifiés grâce à leur code-barre ADN par amplification puis séquençage.

L'année suivante, l'équipe de Needels¹⁹⁷ met en pratique le postulat de librairie encodée proposé par S.A. Brenner et R. A. Lerner en préparant, par synthèse parallèle de chimie combinatoire, une librairie supportée sur billes de $8,2 \times 10^5$ séquences peptidiques. La collection de billes portant les séquences heptapeptidiques a ensuite été criblée contre un anticorps anti-peptide fluorescent et les meilleurs ligands ont été identifiés par cytométrie en flux (en. *fluorescence activated cell sorting*, FACS). La même année, l'équipe de Janda¹⁹⁸ a publié un papier dans lequel une librairie de polypeptides contenant l'épitope β -endorphine a pu être couplés à des billes CPG (en. *control pore glass*) au travers d'un *linker* chimique bifonctionnel et lier successivement par les code-barres ADN. Malgré une preuve de concept dument démontré dans les années 1990, il a fallu attendre le début des années 2000 pour enfin voir arriver un nombre important de publications utilisant cette technique. En 2004, plusieurs groupes ont reportés la construction de petites librairies de molécules encodées par l'ADN en utilisant non plus des supports solides mais un *linker* chimique afin d'associer phénotype et génotype¹⁹⁹⁻²⁰¹.

À partir de 2004, de réels progrès et innovations ont été effectués autour du développement de librairies combinatoires encodées avec notamment l'essor de techniques inédites pour allier un fragment amplifiable à un motif organique.

Sous-partie 2.1 : Méthode alternative pour la formation de librairies encodées.

Les librairies combinatoires encodées par l'ADN peuvent être arrangées en deux catégories structurellement différentes¹⁹⁵. Tout d'abord, on distingue les librairies où le motif pharmacophore est couplé de manière covalente à une ou deux séquences oligonucléotidiques (Figure 29, gauche). Ces librairies sont connues sous le nom de *single-pharmacophore library*. A noter qu'en 2009, le groupe de M.A. Clarck à GlaxoSmithKline a développé un précurseur bifonctionnel sur lequel il est possible de lier enzymatiquement les code-barres double brin d'ADN et de coupler les synthons organiques afin d'amplifier la diversité de la librairie²⁰². Ce précurseur appelé *Headpiece* en anglais sera plus amplement discuté en partie 6 de ce rapport.

Par opposition au *single-pharmacophore*, la méthode de librairie *dual-pharmacophore* repose sur le fait que chaque brin d'ADN va porter en partie terminale un fragment unique de la

librairie. L'appariement des brins d'ADN (Figure 29, droite) va permettre de former un couple de fragments interagissant conjointement sur la cible mise en présence.



Figure 29 : Représentation schématique des bibliothèques encodées à single-pharmacophore (gauche) ou dual-pharmacophore (droite).

Section 2.1.1 : Librairie délivrant un pharmacophore unique (en. single-pharmacophore library)

Paragraphe 2.1.1.a) ADN encodé (en. DNA encoded)

Le cas le plus généralement utilisé en chimie combinatoire encodée est la construction d'une librairie dite encodée (en. *DNA encoded*) encore connue sous le nom de librairie d'ADN par enregistrement, ou *DNA recorded library* en anglais. Conceptuellement, la confection de cette librairie est celle-là plus proche des suggestions faites par S.A. Brenner et R. A. Lerner au début des années 1990. En effet, la librairie est simplement construite par ligation enzymatique séquentielle de code-barres ADN référant l'identité des composés organiques associés au corps du pharmacophore. Afin d'exemplifier cette technique, il est bon de considérer¹⁹⁵ deux bibliothèques, L1 et L2, contenant chacune q entités organo-génomiques différentes. La librairie L1 voit ces structures organiques (A) liées au corps génomique (code A) par l'intermédiaire d'un seul brin d'ADN (Figure 30, A) alors qu'ils seront couplés aux deux brins d'ADN au sein de la librairie L2 par l'usage d'un espaceur (en. *linker*) bidirectionnel (Figure 30, B/).

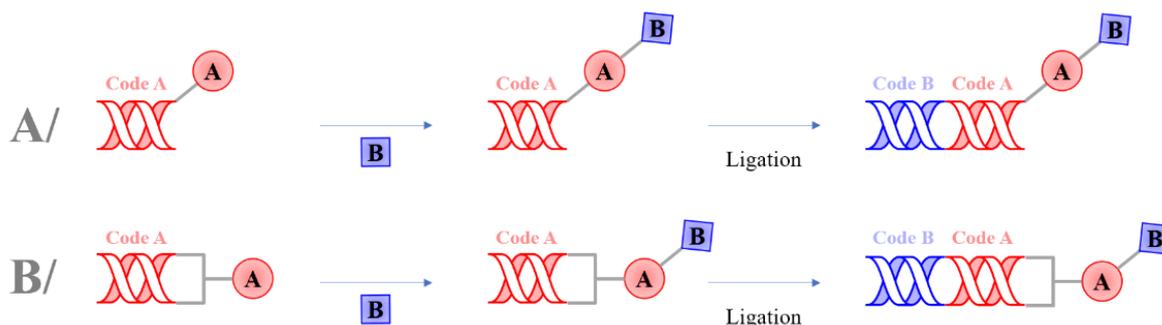


Figure 30 : Illustration d'une librairie « recorded » où A/ le fragment chimique est directement lié à l'oligonucléotide B/ lié à l'oligonucléotide par l'intermédiaire d'un espaceur bifonctionnel.

Dans les deux cas, comme chacun des composés chimiques ($A_1, A_2, \dots, A_\alpha$) est couplés à un code-barre ADN unique (code $A_1, A_2, \dots, A_\alpha$), les α composés organiques encodés par l'ADN peuvent être mixés sans ambiguïté. Une fois rassemblé, chacune des librairies peuvent être subdivisée en β lots différents afin d'y incorporer de nouveaux composés organiques notés B (B_1, B_2, \dots, B_β). S'en suit la ligation d'un nouveau code-barre d'ADN double-brin distinct, noté code B (code B_1, B_2, \dots, B_β). Ainsi, en utilisant $\alpha \times \beta$ composés organiques uniques (A et B) et $\alpha \times \beta$ code-barres ADN distincts (code A et code B), les librairies L1 et L2 vont être composées de $\alpha \times \beta$ entités organiques encodées. Évidemment cette méthode peut être réitérée plusieurs fois afin d'atteindre le nombre souhaité de composés au sein d'une librairie ($\alpha \times \beta \times \gamma$ pour 3 tours de *DNA-recorded* etc...).

En considérant la synthèse d'une telle librairie, il est intéressant de se rendre compte que pour n produits organiques au sein de la librairie, il est nécessaire de $2n$ séquences oligonucléotidiques à cause de la structure hétérodimérique de l'ADN. Afin d'outrepasser la notion de code-barres ADN double brin et d'alléger le coût total de production d'une telle librairie, le groupe de D. Neri a décrit une procédure d'encodage à n oligonucléotides^{203–208}.

Dans cette nouvelle méthode, il est à considérer une nouvelle librairie L3 portant α fragments organiques A différents. Chacun des *building blocks* ($A_1, A_2, \dots, A_\alpha$) est attaché du côté 5' à un code-barre ADN simple brin par l'intermédiaire d'un linker présentant un groupement amine primaire libre (Figure 31, C/).

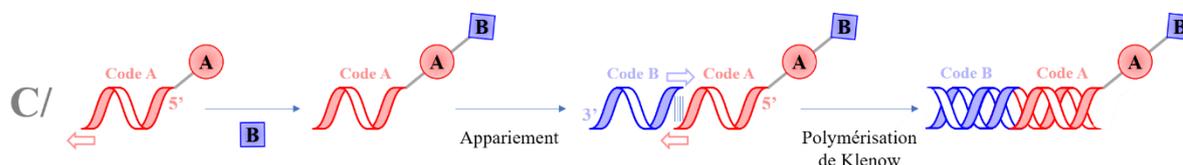


Figure 31 : Illustration d'une librairie « recorded » encodée par de l'ADN simple brin via l'utilisation de la polymérisation de Klenow.

Les séquences oligonucléotidiques utilisées sont spécialement conçues pour avoir une séquence globale commune dont seules les bases centrales diffèrent, permettant ainsi d'encoder individuellement les composés organiques associés. Ainsi, après unification de tous

les membres de la librairie suivie de division équivalente en β lots, un nouvel ensemble de composés organiques est couplé (B_1, B_2, \dots, B_β). L'identité de chacun des couplages B est enregistrée (en. *recorded*) par appariement de nouvelles séquences code-barres B (code B_1, B_2, \dots, B_β). Chacune de ces nouvelles séquences présente une partie partiellement complémentaire à l'ADN simple brin A. Cet appariement est effectué par utilisation du fragment dit « de Klenow » provenant du clivage enzymatique de l'ADN polymérase I par la subtilisine. En effet, le traitement de l'ADN Poly I (103 KDa) par cette protéase à Sérine engendre la formation de deux fragments, dont le fragment de Klenow (68 KDa) qui maintient l'activité polymérase de $5' \rightarrow 3'$ et l'activité d'exonucléase et de relecture (en. *exonucleases and proofreading activities*) de $3' \rightarrow 5'$ au même titre que l'ADN poly I. En revanche, contrairement à cette polymérase, le fragment de Klenow perd l'activité exonucléasique de $5' \rightarrow 3'$. C'est bien cette dernière propriété ôtée qui permet la formation de l'encodage à double brin à partir de code-barres (code A et B) simples brins.

Paragraphe 2.1.1.b) ADN matriciel (en. *DNA templated*)

Une méthode alternative à la formation de librairie à pharmacophore unique est l'utilisation de l'ADN matriciel (en. *DNA templated*) inventé par le groupe de David Liu au début des années 2000²⁰⁹. Dans cette approche, la formation d'ADN sous forme d'hétéroduplex par appariement de bases est utilisée afin d'accélérer la réaction entre deux entités chimiques présentes en parties terminales de chacune des séquences d'oligonucléotides (Figure 32, A/).



Figure 32 : Illustration d'une librairie « templated » où A/ l'enrichissement de la librairie se fait, après appariement de code-barres ADN complémentaires, par réaction chimique dû à la proximité des fragments.

En effet, la méthode de *DNA templated synthesis* débute par la synthèse de conjugués oligonucléotidiques préalablement fonctionnalisés en partie terminale par un fragment A ; la séquence d'oligonucléotide associée comportant autant de régions codantes (codon) qu'il n'y a de diversité structurale à introduire. En d'autres termes, pour une librairie composée de 2 entités A et B, la séquence d'oligonucléotide matricielle comprendra 2 régions codantes, code A et code B. Evidemment, si l'entité A comporte un ensemble de α *buildings blocks* et que l'entité B comporte un ensemble de β *buildings blocks*, il sera nécessaire de partir d'une combinatoire de $\alpha \times \beta$ conjugués oligonucléotidiques préalablement fonctionnalisés par les α *buildings blocks* de l'entité A. Une fois préparé, l'ADN matriciel est apparié à une séquence complémentaire du code-barre B (anticodon) préalablement fonctionnalisée par l'entité B. La proximité spatiale des entités chimiques due à l'appariement de l'anticodon B permet la réaction entre A et B afin d'enrichir la diversité structurale de la librairie.

Le schéma ci-après représente un exemple de chemin réactionnel dérivé d'un des travaux de D. Liu¹⁹⁹ pour la synthèse d'une librairie à trois diversités fonctionnelles (Figure 33). La stratégie de synthèse de *DNA templated* utilise ici deux espaceurs chimiques, un espaceur dit

auto-immolable (en. *autocleaving linker*) en vert ainsi qu'un espaceur non cicatriciel (en. *scarless linker*) en bleu, permettant l'incorporation de diversités fonctionnelles R_2 et R_3 , respectivement, sur un ancrage moléculaire comportant déjà une fonctionnalisation R_1 . Une représentation plus exhaustive de toutes les stratégies utilisables pour contrôler la réactivité chimique après appariement de séquences d'ADN est discutée en détail par X. Li et D. Liu dans une publication de fonctionnalisation de molécules synthétiques²¹⁰.

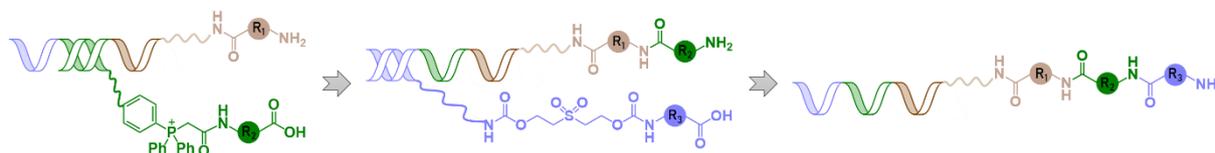


Figure 33 : Exemple d'une librairie « templated ». Enrichissement de la librairie permise par proximité spatiale de building blocks contrôlés par appariements de code-barres ADN complémentaires.

Le transfert de fragments sur une structure moléculaire naissante doit être convenablement suivi par une étape de clivage contrôlée afin de permettre une continuité de synthèse. Pour ce faire, les petites séquences oligonucléotidiques complémentaires sont généralement biotinyllées en partie terminale opposée au fragment. Ainsi, l'utilisation de billes fonctionnalisées par une (strept) avidine permet de purifier efficacement le système (Figure 32, A/) afin d'obtenir une entité finale simple brin et déprivatisée par l'ajout de divers *building blocks*.

Cette approche de synthèse de librairie *DNA templated* à particulièrement prouvée son efficacité pour la synthèse de complexe macrocyclique^{199,211}. En revanche, la principale limitation d'un tel système est la reconnaissance hautement spécifique de l'appariement de la séquence matricielle aux oligonucléotides complémentaires.

Afin de pallier cette limitation, le groupe de X. Li²¹² a proposé une optimisation du système d'encodage, associant au sein de la séquence matricielle non plus des régions codantes complémentaires aux code-barres ADN mais des régions dites universelles formées par des séquences poly-désoxyinosines (Poly-I, Figure 34, B/). En effet, les désoxyinosines monophosphates (dIMP) sont des nucléosides à base purique généralement obtenue par désamination de la désoxyadénosine monophosphate (dAMP) capable d'indifférencier les quatre bases de l'ADN. Cela permet ainsi une hybridation avec de multiples séquences d'oligonucléotides.

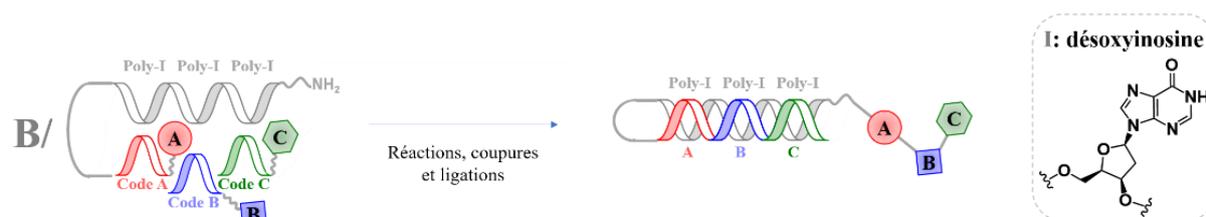


Figure 34 : Illustration d'une librairie « templated » à séquence poly-désoxyinosines. Encodage et fonctionnalisation de la librairie par appariement, ligation enzymatique et couplage.

Initialement, la synthèse *DNA templated* utilise des composés chimiques pré-encodés par des code-barres ADN connus qui, par appariement de leurs séquences complémentaires au brin matriciel, vont délivrer leurs fragments chimiques avant d'être dénaturé. Ceci permet ainsi l'obtention de conjugués oligonucléotidiques simple brin et chimiquement fonctionnalisés. Dans ce système d'oligonucléotidique poly-inosinylé, les code-barres ADN préfonctionnalisés vont être liés enzymatiquement à la séquence Poly-I avant de délivrer leur fragment chimique par l'intermédiaire d'un espaceur clivable (Figure 35).

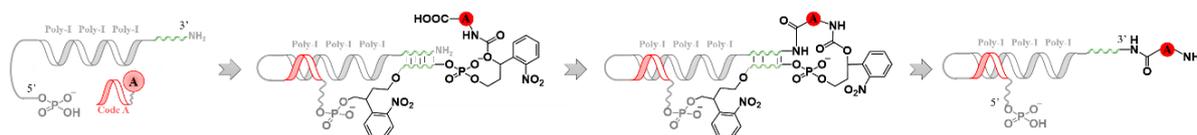


Figure 35 : Exemple d'une librairie « templated » à séquence poly-désoxyinosines présentant un monophosphate en 5' et une séquence connue amino-terminale (vert) en 3'. La fonction chimique pré-encodée par le code-barre ADN (rouge, A) est liée de manière enzymatique à la séquence poly-I en 5'. L'utilisation d'un espaceur photo-clivable présentant une séquence complémentaire à la séquence Poly-I en 3' permet l'introduction du fragment par couplage peptidique. Après clivage du linker la molécule obtenue est prête pour le deuxième tour de la librairie encodée.

Une autre méthode de librairie *templated* appelée *YoctoReactor*TM a été proposée par la compagnie danoise Vipergen en 2009²¹³. En biologie, les jonctions à trois (ou quatre) brins d'ADN représentent des états transitoires importants notamment dans les mécanismes de réplifications et de recombinaison²¹⁴. La capacité de l'ADN à former de telles jonctions forme au centre de l'édifice un espace d'une capacité volumique de 10^{-24} L, soit un yoctolitre. L'équipe danoise a alors utilisé la capacité de l'ADN double brin à former de telles jonctions tridimensionnelles afin de générer des librairies combinatoires de composés organiques.

La formation de jonctions triples est possible grâce à l'interaction spécifique d'oligonucléotides spécialement conçus pour spontanément s'autoassembler en de tel édifice. Chacun des bras du *YoctoReactor* est composé : i) d'une séquence ADN stabilisée sous forme de double brin (éliminant ainsi les problèmes d'association et d'appariement de l'ADN simple brin) et codant spécifiquement pour les divers *buildings blocks* insérés dans la librairie, et ii) une région non variable qui régit l'autoassemblage du *YoctoReactor*. Les *buildings blocks* permettant l'enrichissement de la librairie sont reliés au bras du *YoctoReactor* par l'intermédiaire d'espaceurs clivables ou covalents et spécifiquement positionnés le long du bras de la jonction triple. En effet, afin d'engendrer l'interaction la plus efficace possible, les *buildings blocks* se doivent d'interagir au centre du *YoctoReactor*. La capacité volumique centrale d'une jonction triple étant particulièrement petite, un yoctolitre, les concentrations locales avoisine ainsi le millimolaire ce qui permet de générer une réaction chimique efficace.

La première étape de formation d'une telle librairie nécessite la mise en présence de deux séquences d'ADN préfonctionnalisées (séquences rouge et bleu fonctionnalisées par le

building block A et B respectivement ; Figure 36, C/) ainsi qu'une petite séquence d'oligonucléotide dite « d'appui » afin de former une structure à 3 membres.

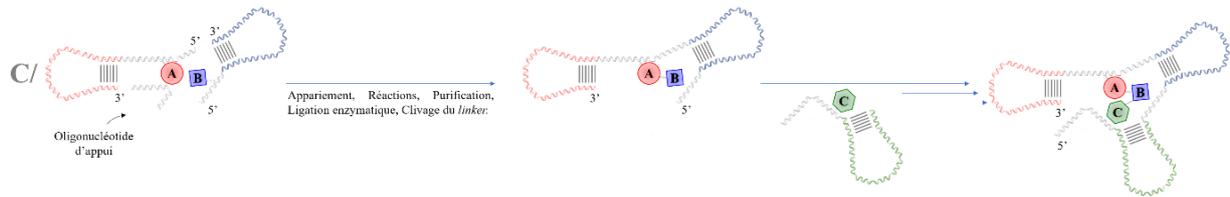


Figure 36 : Illustration d'une librairie « templated » obtenue par YoctoReactor™.

L'appariement spécifique de ce nouvel édifice va permettre d'installer spatialement les fragments et donc de les faire réagir entre eux. Une fois l'édifice formé et purifié, les deux bras du YoctoReactor sont liés enzymatiquement et les espaceurs sont clivés afin d'obtenir une séquence d'ADN encodée et enrichie structurellement. Notons que la seconde étape d'enrichissement ne nécessite plus la présence de l'oligonucléotide d'appui puisque la formation de la jonction triple sera effectuée par la séquence précédemment obtenue, présentant déjà deux bras du YoctoReactor, ainsi que le dernier bras fonctionnalisé (séquence verte fonctionnalisée par le building block C).

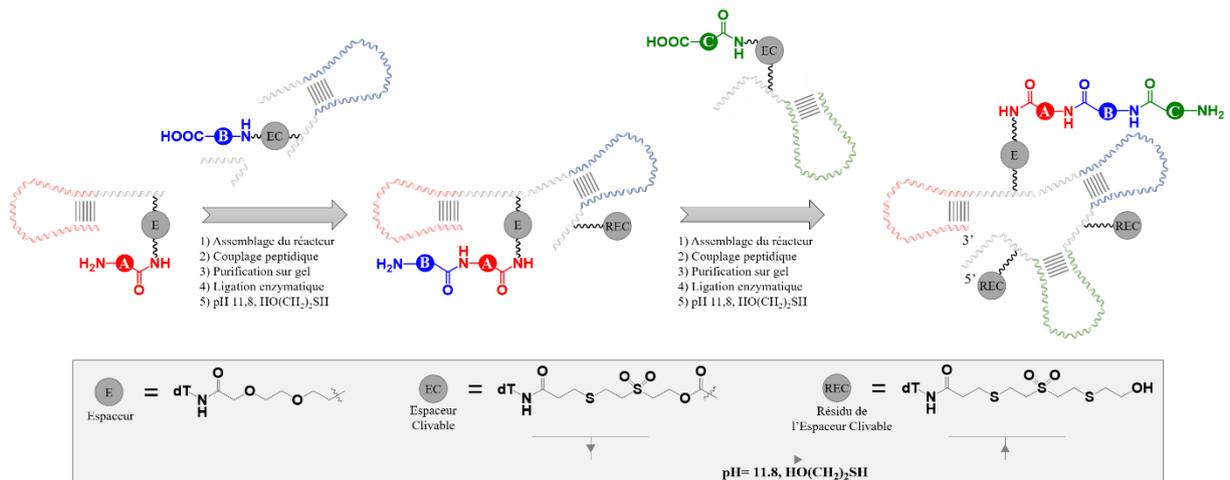


Figure 37 : Exemple d'une librairie « templated » obtenue par technologie YoctoReactor™. L'appariement de séquence double brin d'ADN pré-fonctionnalisé par un building block permet le couplage de ceux-ci dû à leur proximité spatiale. Le clivage final des linkers permettent d'obtenir un composé exprimé à la surface de l'ADN.

La dernière méthodologie utilisant l'encodage par *DNA templated* est une technologie développée en 2004 par le groupe de P.Harbury connue sous le nom de *DNA routing*²⁰¹. Utilisant également la propriété d'appariement de séquences complémentaires, cette méthode permet l'alternance d'étapes d'hybridations et de couplages chimiques afin de former *in fine* une librairie combinatoire enrichie par divers fragments organiques.

De manière analogue aux travaux de D. Liu sur la technologie *DNA templated*, le *DNA routing* utilise de longues séquences d'ADN simple brin formé par l'alternance successive de régions codantes (code A, B et C, Figure 38, D/) et de régions non codantes (non illustrées) d'environs

une vingtaine de bases chacune. A une des extrémités de ces séquences matricielles est incorporé un espaceur chimique (généralement des hydrocarbures saturés ou des chaînes polyéthylène glycol) portant une amine primaire terminale libre dans le but d'avoir un ancrage chimique disponible afin d'enrichir la librairie par divers motifs moléculaires.

La fonctionnalisation de la librairie par les fragments la composant se fait par méthode de chimie combinatoire *split and mix* dirigé par l'hybridation de l'ADN. En effet, l'incubation d'anticodon complémentaire lié de manière covalente à une résine de diéthylaminoéthyl-sépharose (DEAE, code B *figure _ D*) permet l'appariement spécifique d'une des régions codantes de la séquence matricielle et ainsi la fonctionnalisation indexée du *building block* (B) par réaction chimique, généralement par l'utilisation de la chimie des peptides^{215–217}.

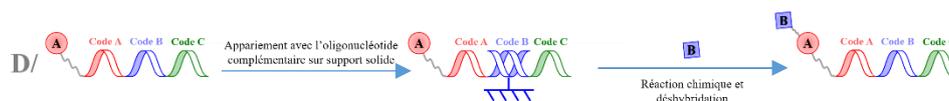


Figure 38 : Illustration d'une librairie obtenue par DNA routing.

Section 2.1.2 : Librairie délivrant une paire de pharmacophore (en. dual pharmacophore library)

À l'instar des conceptions de librairies à pharmacophore unique, les librairies à *dual pharmacophore* portent un *building block* à l'extrémité de chacun des simples brins d'ADN. Chaque oligonucléotide simple brin possède ainsi une région variable, distincte pour chacun fragment chimique, ainsi qu'une région plus conservatrice, essentielle à l'autoassemblage par appariement de séquences complémentaires. Basé sur assemblage automatique des brins conservateurs, cette technique d'encodage est née sous le nom de librairie encodée par autoassemblage de l'ADN, soit *encoded self-assembling chemical library* en anglais — ou ESAC.

Paragraphe 2.1.2.a) Librairie encodée par auto-assemblage de l'ADN (en. Encoded self-assembling chemical library- ESAC)

La conception d'une librairie ESAC a initialement été conçue pour être compatible avec des techniques de séquençage de l'ADN comme le microarray²⁰⁰ et repose sur le simple appariement de séquences d'ADN complémentaires individuellement indexées (code A et code B ; Figure 39, A/). L'Hybridation des séquences complémentaires va ainsi permettre à deux fragments d'être spatialement proche, sans interagir, afin de stabiliser à elles seules une cible donnée d'origine protéique voire oligonucléotidique.

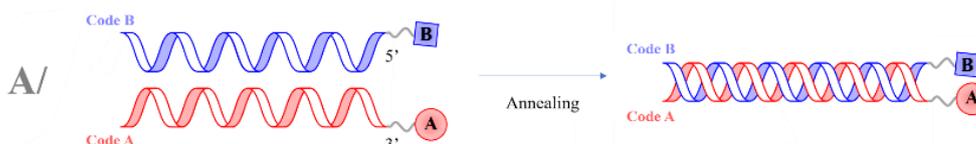


Figure 39 : Représentation schématique de la conception de librairie ESAC à l'aide de deux séquences oligonucléotidiques chimiquement encodées.

Le point négatif de cette approche est le décodage des séquences d'ADN. En effet, le séquençage de ce genre de *dual*-motif oligonucléotidiques n'apporte malheureusement aucune indication sur leurs facteurs d'enrichissement. Étant donné qu'un même codon peut être retrouvé plusieurs fois au sein d'une paire de conjugué oligonucléotidique, il est impossible de quantifier l'enrichissement d'une paire de motifs donnée.

Ainsi, afin de pallier ce manque d'informations, le groupe de D. Neri a breveté une nouvelle procédure de synthèse de librairie à autoassemblage utilisant le séquençage à haut-débit comme moyen de lecture des brins d'ADN (Figure 40, B/).

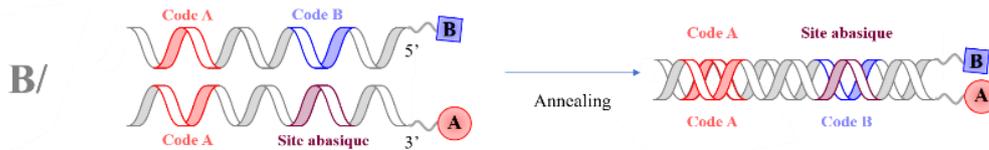


Figure 40 : Méthode d'autoassemblage de bibliothèques encodées.

Dans cette configuration, une séquence nucléotidique doublement indexée par un codon A (code A) et une séquence abasique est individuellement couplée en 3' à différentes entités chimiques composant le fragment A (A). En parallèle, une autre séquence nucléotidique est doublement indexée, premièrement par un des codons B (code B) et deuxièmement par un anticodon de A (code A), et présente une extrémité 5' fonctionnalisée par un nouveau jeu de fragments (B). La présence simultanée du codon/anticodon A va permettre un appariement sélectif des deux séquences complémentaires. Le codon B sera quant-à-lui apparié en face de la séquence abasique permettant ainsi la formation d'un conjugué ADN double brin indexé et présentant un enrichissement chimique A et B.

Un dérivé de cette dernière technique a également été décrit dans la littérature et propose l'utilisation de la polymérisation de Klenow comme expliqué en point 2.01.a). En effet, la librairie ESAC va être synthétisée à partir d'une séquence oligonucléotidique fonctionnalisée en 3' par les fragments A (A) et composé d'une séquence abasique. Au conjugué oligonucléotidique fonctionnalisé va être liée en 5' une petite séquence simple brin contenant l'index du fragment A par l'intermédiaire d'une amorce. Le nouvel oligonucléotide doublement indexé, ressemblant fortement à la séquence décrite dans le paragraphe précédent, va constituer un élément d'une première sous-librairie. En effet, la synthèse en parallèle de séquences indexées portant le fragment B et regrouper sous le terme de sous-librairie « 2 » va ensuite pouvoir être mise en présence d'un des membres de la sous-librairie 1 et, par la polymérisation de Klenow, il est possible d'obtenir une séquence ADN double brin doublement fonctionnalisée (Figure 41, C/).

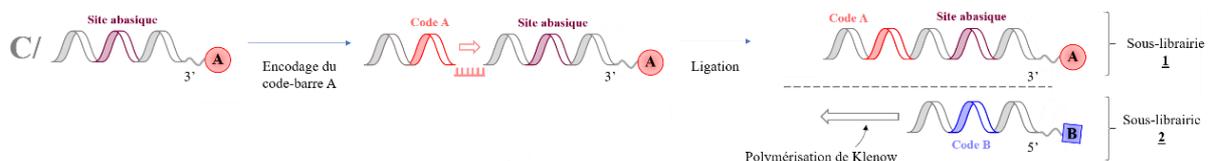


Figure 41 : Assemblage de librairie d'oligonucléotide chimiquement encodée par polymérisation de Klenow.

Paragraphe 2.1.2.b) Librairie encodée par assemblage d'ADN et d'acides nucléiques peptidiques.

L'assemblage de structures multimériques peut également être généré en combinant des séquences ADN et des acides nucléiques d'origine peptidique (en. *Peptide nucleic acids, PNA*). En effet, plusieurs équipes²¹⁸ ont utilisé la synthèse encodée par PNA (en. *PNA-encoded synthesis, PES*) afin d'isoler divers ligands ciblant des protéines (lectines, intégrines, etc.). Dans ce type de librairie, les diversités chimiques sont couplées à l'extrémité du PNA (Figure 42, D/) et les conditions dans lesquels sont effectuées ces réactions peuvent suivre des conditions de chimie organique classique. En effet, au sein des stratégies d'encodage de librairie par l'ADN les réactions applicables sont celles compatibles avec la présence d'ADN (milieu aqueux, nucléobases non protégées, etc.) et limite parfois la diversité de réactions possibles. En revanche, contrairement aux librairies ADN, les librairies de PNA ne peuvent pas suivre des procédures d'amplification classique (i. e PCR). Ainsi, l'utilisation d'une librairie d'ADN simple brin portant la combinaison de deux séquences complémentaires aux PNA permet à la fois l'assemblage de diversités chimiques et l'amplification par PCR des meilleures touches après tests de sélections par affinité sur une cible donnée.

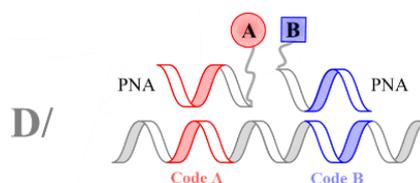


Figure 42 : Formation de librairie combinatoire encodée via l'utilisation d'acides nucléiques d'origines peptidiques.

Section 2.1.3 : Les chimies compatibles

Le développement de librairies combinatoires encodées sollicite des conditions de synthèses particulières. En effet, celles-ci doivent être effectuées en présence d'un co-solvant aqueux, au sein d'un milieu fortement dilué, en préservant l'intégrité de l'information génétique des séquences ADN. Pour ces raisons, les librairies encodées ont longtemps été synthétisées à partir de la chimie des carbonyles, de substitutions nucléophiles aromatiques ou par la formation de lien amides^{202,204,219}. Depuis 2015, le nombre de publications sur des méthodes de synthèses appliquées à la DEL ne cesse de progresser à l'exemple de la publication de P. R. Fitzgerald *et al.*²²⁰ et de K. Götte *et al.*²²¹. Parmi les réactions prépondérantes et couramment utilisées lors de campagne de DEL, il est possible de retrouver la chimie des amides, les substitutions nucléophiles, les aminations réductrices, des couplages pallado-catalysés de type Suzuki-Miyaura ou Buchwald-Hartwig ou encore la CuAAC.

Sous-partie 2.2 : Sélection, déconvolution et validation.

Les méthodes présentées plus tôt dans ce manuscrit permettent l'obtention de librairies chimiques encodées comprenant plusieurs milliers voire millions ou milliards de molécules distinctes. L'intérêt d'une telle prouesse est de pouvoir mettre en compétition les molécules d'une librairie avec une cible d'intérêt biologique par simple incubation. L'incubation *in vitro* de ligands sur une cible permet de ne sélectionner que les composés les plus affins au détriment

de ceux ne présentant aucune reconnaissance pour la cible. Les molécules sélectionnées par ce processus peuvent alors être identifiées grâce aux code-barres ADN unique à chacune des molécules. Il est nécessaire de garder à l'esprit que les entités chimiques d'une librairie ne sont individuellement présentes qu'à l'échelle de la femtomole. Ainsi, l'identification des code-barres ADN par séquençage à haut-débit ne sera possible qu'après un apport suffisant en matériel biologique. Une étape d'amplification des séquences ADN est donc préalablement nécessaire. Une fois amplifiées, les séquences pourront alors être quantitativement identifiées par séquençage à haut-débit et les composés associés aux séquences identifiées comme prépondérantes pourront alors être resynthétisés sans code-barres et évalués par des approches biophysiques, biochimiques et cellulaires classiques.

Section 2.2.1 : Sélection sur cible par affinité.

De manière analogue aux technologies « d'exposition de protéines », ou *protein-display* en anglais, telles que le *phage-display*²²²⁻²²⁶ et le *ribosome-display*²²⁷, les librairies chimiques encodées sont criblées au travers de processus de sélection par affinité. En comparaison avec les méthodes traditionnelles de criblage à haut débit, la technologie d'encodage par l'ADN offre la possibilité de cribler plusieurs millions de composés en une simple et unique expérimentation. Différentes méthodes de criblage ont été mises en œuvres afin de pouvoir sélectionner et identifier les meilleures touches contre une cible donnée. Il est usuel de voir des cibles d'intérêt immobilisées sur support solide mais des sélections en solution ont également été décrites²²⁸.

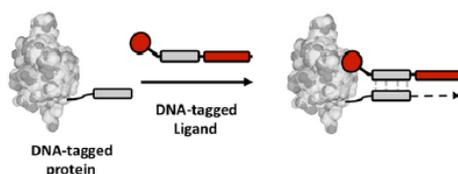
Les sélections par affinité sur phase solide repose sur l'immobilisation d'une cible sur un support solide, support qui est ensuite incubé avec la librairie combinatoire encodée²²⁹⁻²³⁴. Les ligands préférentiellement enrichis contre une cible seront séparés des autres membres de la librairie après capture par affinité, par application de lavages plus ou moins stringents. Parmi les supports utilisables dans les processus de sélection par affinité on peut retrouver les billes magnétiques^{204,208} ou des embouts chargés de résines (en. *Resin-filled tips*). L'immobilisation des cibles, d'origine protéique ou d'acide nucléique, peut être effectuée de manière covalente ou par l'intermédiaire de méthode de capture non covalente telle que l'utilisation de biotine, de tag polyhistidine, etc. Dans le cadre de sélection sur support solide, l'utilisation de différentes conditions expérimentales peut drastiquement influencer les résultats de sélection^{208,235}. En effet, les conditions de lavages (composition des tampons, stringence des lavages) ainsi que l'utilisation de détergent peuvent jouer un rôle non négligeable dans les résultats obtenus.

Les supports solides et les groupes moléculaires d'accrochage utilisés peuvent également amener de l'interaction non spécifique avec d'éventuels ligands, augmentant le niveau du signal-sur-bruit de la sélection et entravant la discrimination entre les artefacts et les vraies touches²³⁶. Afin de prévenir de tels artefacts il est routinier d'utiliser des agents de blocage des interactions non-spécifiques tel que l'ADN génomique (sperme de Saumon), du sérum d'albumine bovin, de la biotine ou encore de l'imidazole.

Malgré les désavantages mentionnés auparavant, les sélections par affinité sur phase solide restent une méthode de choix, dans leur simplicité méthodologique, pour l'identification de touches obtenues dans le cadre de campagne de criblage de librairie combinatoire encodée.

Idéalement, les expériences d'affinité sur support solide devraient laisser leur place à des expériences de sélections en solution afin d'éviter bon nombre d'effets indésirables et potentiellement non contrôlés. L'équipe de D. Liu décrit en 2010 une méthode d'identification de ligands encodés par l'ADN via l'utilisation de protéines marquées par une séquence oligonucléotidique de référence. Cette technique est connue sous le nom de PCR par interaction dépendante (en. interaction-dependent PCR, IDPCR)²³⁷. Dans cette technique, la librairie combinatoire est encodée par un ADN simple brin capable de reconnaître par complémentarité une séquence oligonucléotidique couplée de manière covalente à la protéine d'intérêt. Ce marquage protéique à un double intérêt, le premier étant de stabiliser l'interaction protéine-ligand et le second étant de servir d'amorce à l'amplification par PCR des meilleurs ligands présents au sein de la librairie²³⁸ (Figure 43, C.). Quelques années après l'apparition de cette première technique de sélection en solution, le même groupe a mis au point une méthodologie appelée détermination d'interaction par utilisation de protéines non purifiées (en. Interaction determination using unpurified proteins, IDUP)²³⁹ qui est une extension de la technique IDPCR. Cette nouvelle technique a permis l'identification de ligands issus de librairies encodée au sein d'un mélange complexe de protéines, dans un lysat cellulaire. En effet, au lieu d'attacher un oligonucléotide à une protéine d'intérêt afin d'affiner l'interaction protéine-ligand, le groupe de D. Liu a couplé de manière covalente la librairie d'ADN à un anticorps ou une protéine de fusion portant le marquage SNAP (en. SNAP-tag) permettant ainsi de cibler spécifiquement certaines protéines d'intérêts au sein du lysat cellulaire. Une autre méthodologie du groupe de D. Liu a également date puisqu'elle ne nécessite ni préimmobilisation de la cible ni prémarquage, c'est le marquage par photoaffinité d'ADN programmé (en. DNA-programmed photoaffinity labeling, DPAL)²⁴⁰⁻²⁴² (Figure 43, D.). Dans cette méthode, une librairie de molécules encodée par de l'ADN simple brin est appariée avec une séquence d'oligonucléotide préalablement couplée à un groupement photo-réactif afin de former un hétéroduplexe d'ADN stable. En présence de la protéine d'intérêt et après irradiation, les ligands les plus affins seront capturés sur la cible par une réaction de photo-réticulation (en. photo-crosslinking).

C. Interaction Dependent PCR



D. DNA Photoaffinity Labeling

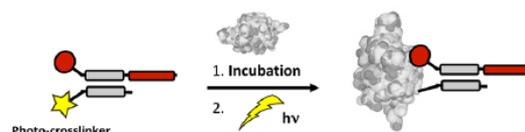


Figure 43 : Méthodes de DEL dont les expériences *in vitro* de sélections peuvent être effectuées en solution — IDPCR (C.), IDUP, DPAL (D.). Figure issue de la review de A. Gironda-Martinez²²⁸.

De nombreuses approches de sélections sur support solide ou en solution ont été décrites à ce jour permettant l'isolation de nombreuses molécules d'intérêts contre différentes cibles principales d'origines protéiques. En revanche, certaines limitations concernant ces cibles sont à considérer. En effet, dans les conditions d'expérimentations, il est parfois compliqué d'isoler une cible dans sa conformation d'origine voire même de la stabiliser. De plus, certaines cibles peuvent perdre les activités biologiques en dehors de leurs conditions cellulaires. Compte tenu de ces considérations, plusieurs groupes se sont intéressés à l'exécution de sélection de bibliothèques encodées au sein d'environnement naturel. En 2015, les scientifiques de *GlaxoSmithKline* (GSK) ont décrit la première campagne de criblage de près de 15 milliards de molécules encodées contre des cellules surexprimant le récepteur de la tachykinine (neurokinin-3 ou NK3) au niveau de leur membrane²⁴³. Des optimisations de sélections, permettant notamment d'éviter des interactions non spécifiques et l'internalisation des molécules, ont ainsi permis l'identification de différents agonistes, comparable en terme d'efficacité et de spécificité à des ligands connus de la littérature. D'autres groupes, dont celui de C. Krusemark, ont reporté des sélections de bibliothèques encodées en milieu cellulaire. Dans son travail de 2019²⁴⁴, le groupe de C. Krusemark a criblé une bibliothèque encodée de 96 peptides (cycliques) de pénétration cellulaire (en. *cell-penetrating peptide*, CPP)²⁴⁵ capable d'interagir avec des protéines membranaires et des protéines cytosoliques. L'interaction spécifique entre les CPP et les cibles d'intérêts a été fixée par une réaction chimique de photo-réticulation, comme exposé en (Figure 43, D.).

En 2021, le groupe de X. Li a amélioré les performances de la sélection de bibliothèques encodées en milieu cellulaire en utilisant les travaux de sélection en solution décrite par D. Liu²³⁷. En effet, ce groupe a montré la possibilité de fonctionnaliser des protéines membranaires par une séquence ADN simple brin. Ce marquage, servant également de guide, a permis l'hybridation spécifique des membres de la bibliothèque encodée et l'augmentation de la concentration locale de bibliothèque encodée à proximité de la cellule. Le groupe a ainsi pu mettre en évidence des ligands spécifiques après plusieurs étapes de lavages, récupération des ligands encodés, amplification de l'ADN code-barre et séquençage.

Encore plus récemment, la société ViperGen a décrit une nouvelle technologie de criblage à haut-débit en milieu cellulaire. Cette nouvelle méthodologie de criblage utilise des oocytes de grenouilles afin d'exprimer de manière spécifique les protéines d'intérêts dans le cytosol de la cellule. Largement plus grand qu'une cellule somatique normale (100 000 fois), l'utilisation d'oocytes (présentant un volume approximant les 1 μ L) permet l'injection directe d'une bibliothèque encodée de 194 millions de membres en s'affranchissant des problèmes liés à la pénétration des composés de la bibliothèque au sein de criblage *in vivo*²⁴⁶. L'injection de composés encodés au sein de ce système cellulaire entraîne la liaison des membres de la bibliothèque sur la protéine d'intérêt, spécialement exprimée au sein de l'oocyte, mais également des liaisons non souhaitées aux protéines endogènes de la cellule. Afin de dépasser cette limitation, le groupe a fusionné à la protéine d'intérêt une petite protéine nommée *prey*. L'avantage de *prey* est de reconnaître spécifiquement une molécule appelée *bait* [3-(cyclooctylamino)-2,5,6-trifluoro-4-((2-hydroxyéthyl)sulfonyl)benzenesulfonamide] installée sur une séquence ADN. L'objectif du

complexe *bait – prey* est d'assurer uniquement l'amplification de ligands de la librairie interagissant avec la cible d'intérêt.

Quelles que soient les méthodes de sélections utilisées, la détermination des performances de sélection par affinité peut être quantifiée. En effet, les éluats obtenus *post-sélection* sont composés des entités chimiques les plus affines pour la cible ainsi que des code-barres ADN propre à chaque molécule élue. Il est alors possible d'utiliser une méthodologie d'amplification d'ADN afin de quantifier la performance individuelle des méthodes de sélections.

Section 2.2.2 : Amplification des code-barres ADN.

L'amplification de fragments d'ADN *in vitro* est réalisée par une méthode dite de polymérisation en chaîne (en. *Polymerase Chain Reaction*, PCR). Cette technique permet d'obtenir une quantité importante d'ADN spécifiques et repose sur la réplication d'une matrice d'ADN double brin. On peut décomposer l'amplification d'une séquence d'ADN sous forme de trois phases (Figure 44) :

- La première phase est une étape de dénaturation de l'ADN double brin en un couple de séquence simple brin. Cette étape de séparation est généralement effectuée à haute température (95°C) afin d'obtenir une dissociation complète des deux brins.
- Après séparation, deux amorces de quelques nucléotides de long sont ajoutées en large excès au milieu. En effet, la première amorce va être spécifiquement complémentaire d'une séquence présente sur le brin direct (5' vers 3') de l'ADN dénaturé alors que la seconde va reconnaître le brin indirect (3' vers 5'). Cette phase de fixation du couple d'amorces est appelée la phase d'hybridation.

La température d'hybridation des amorces dépendra directement de la composition en bases des amorces. Pour des amorces inférieures à 30 nucléotides, il est possible d'utiliser la formule suivante pour calculer leur température d'hybridation :

$T^{\circ}_{\text{hybridation}} (\text{°C}) = 2 (A+T) + 4 (G+C)$, où A, T, G et C représentent respectivement le nombre de chacune des bases au sein de la séquence de l'amorce.

- Après hybridation des amorces, le processus d'amplification continue par une phase dite d'élongation. Durant cette phase d'extension des amorces, les brins complémentaires aux deux séquences d'ADN simple brin vont être synthétisés. L'allongement des séquences d'ADN simple brin est permise par l'intervention d'une enzyme qui catalyse l'incorporation des nucléotides lors de la synthèse, la Taq polymérase (à noter que d'autres types de polymérase peuvent être utilisés et ne seront pas détaillés ici). Extraite de la bactérie *Thermus aquaticus*, un microorganisme thermophile à Gram négatif, cette ADN polymérase thermostable présente une activité polymérase de 5' vers 3' ainsi qu'une activité exonucléase de 5' vers 3'. En revanche, aucune activité exonucléase n'est possible de 3' vers 5'. Ceci implique que la correction d'erreur liée à de mauvais appariements n'est pas possible. Malgré tout, le Taq

polymérase présente un faible taux de mauvaise incorporation puisque cette valeur est estimée à 1 pour 10 000 bases. Afin que l'activité polymérase puisse être à sa capacité maximale, une température de 68 à 72°C doit être appliquée, permettant ainsi un taux d'incorporation de 2 à 4 kilobases par minute.

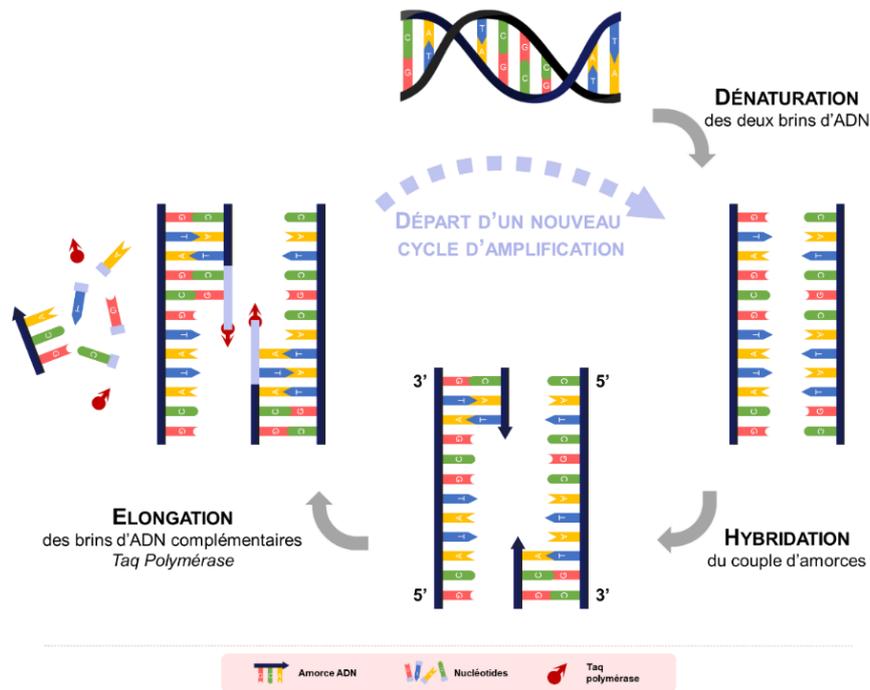


Figure 44 : Principe général de l'amplification de fragment d'ADN par la méthode de Polymerase Chain Reaction.

Une réalisation d'un cycle complet de PCR le milieu est en présence de deux copies du fragment d'ADN de départ. Ainsi, par itération successive de cycle d'amplification, le nombre de copies d'ADN est directement égal à 2^n , où « n » est le nombre de cycles d'amplification effectué.

L'aboutissement d'un cycle de PCR nécessite des variations ponctuelles et précises de températures. Le thermocycleur est un appareil classique de PCR capable d'être programmé afin de délivrer une température donnée pendant une durée déterminée par l'expérimentateur. Une expérimentation sur thermocycleur est généralement rapide puisqu'il ne faut que quelques heures (2 à 3 heures) pour réaliser une PCR avec 30 cycles d'amplification (soit l'obtention de 2^{30} molécules d'ADN).

L'augmentation de la quantité d'ADN lors d'une amplification par un thermocycleur peut également être scindée en trois phases : une phase exponentielle, une phase transitoire et une phase stationnaire (Figure 45). Lors de la phase exponentielle, les nucléotides sont rapidement consommés pour faire face à une augmentation croissante de la quantité d'ADN. Lors de la phase transitoire, l'efficacité de l'amplification diminue fortement due à l'épuisement relatif du milieu réactionnel. Enfin, lorsque tout le milieu réactionnel est consommé et que l'ADN n'est plus amplifiable on atteint la phase dite stationnaire.

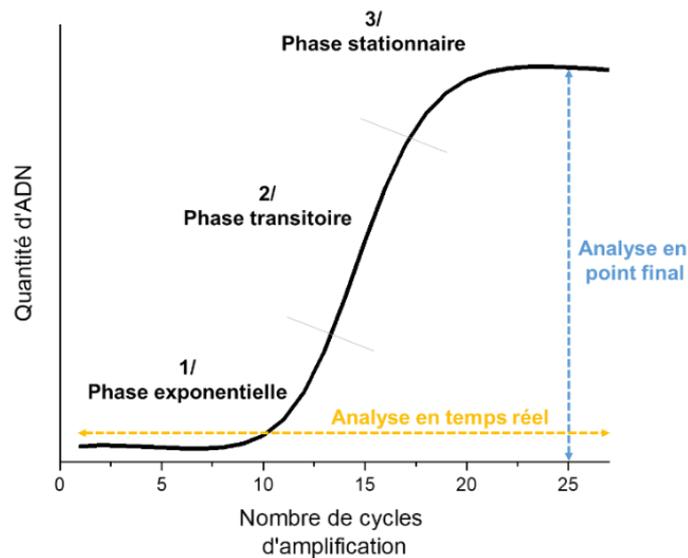


Figure 45 : Phases discernées lors de l'amplification d'ADN par Polymerase Chain Reaction.

L'analyse des résultats d'amplification et la quantification des échantillons sont généralement effectuées à la fin du processus d'amplification, c'est-à-dire lorsque les échantillons se trouvent en phase stationnaire. L'analyse en *point final* présente néanmoins l'inconvénient de ne pas être quantitative. En effet, la détermination de la quantité d'ADN obtenue au sein de plusieurs échantillons sera sensiblement la même à partir du moment où chaque échantillon a atteint sa phase stationnaire. Par conséquent, s'il est admis que l'efficacité de l'amplification est la même quel que soit l'échantillon considéré, il ne sera pas possible de distinguer les échantillons ou l'ADN matrice au début de l'expérience était présente en plus grosse ou plus faible quantité. L'analyse en point final n'est ainsi nullement appropriée pour la détermination d'ADN dans un échantillon.

Néanmoins, si l'on souhaite utiliser la PCR comme méthode de choix pour la quantification d'ADN dans un échantillon, il est nécessaire d'analyser l'amplification de l'ADN lorsqu'il se trouve dans la première phase de l'amplification, c'est-à-dire dans sa phase exponentielle. Pour ce faire, il est souhaitable de déterminer la quantité d'ADN produite par la réaction de PCR après chaque cycle d'amplification. Une possibilité est d'ajouter au milieu réactionnel un agent fluorescent intercalant de l'ADN. Cet agent doit avoir des propriétés de fluorescence différente suivant s'il est en interaction avec un ADN double brin ou libre dans la solution. Un de ces agents grandement utilisés en PCR est une cyanine asymétrique connue sous le nom de SYBR Green. La quantité de SYBR Green intercalée dans le double brin d'ADN est directement proportionnelle à la quantité et la longueur de la double hélice présente dans la solution. Ainsi, la mesure de la fluorescence permet de rendre compte de la quantité d'ADN présents en solution à chaque moment.

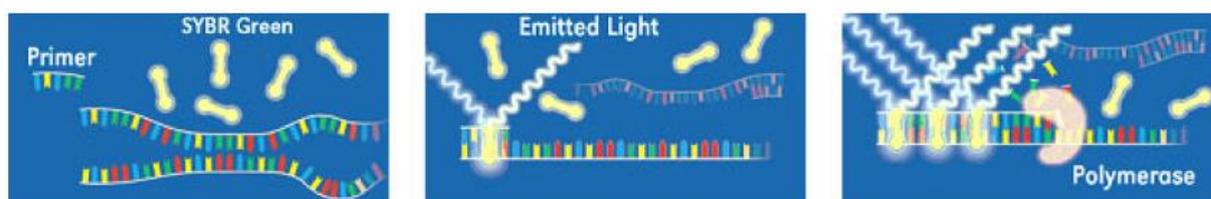


Figure 46 : Visualisation de la progression de l'amplification à l'aide du marqueur fluorescent, le SYBR Green. Image tirée du site de l'institut Agro de Montpellier.

Les appareils capables de faire de la PCR dite quantitative en temps réel sont des thermocycleurs possédant la capacité de quantifier la fluorescence *in situ* des échantillons à la fin de chaque cycle d'amplification. Il est alors possible de tracer des courbes d'amplification afin de définir les phases d'exponentielles, transitoires et stationnaires (Figure 46).

Section 2.2.3 : Séquençage des code-barres ADN.

Une fois les code-barres ADN amplifiés par PCR, l'objectif est de connaître la séquence précise des nucléotides composant l'ADN afin d'en déduire l'identité des composés chimiques sélectionnés ainsi que la valeur d'enrichissement de ces molécules après un processus de sélection. Cette étape de séquençage des ADN code-barres est une étape cruciale dans la technologie d'encodage de librairie chimique par l'ADN.

Les premières bibliothèques chimiques encodées par l'ADN étaient composées de quelques centaines de composés et les code-barres étaient décodés par des technologies de puces à ADN (en. *DNA-microarray*), par amplification bactérienne ou par séquençage classique selon la méthode de Sanger^{200,203,206,247}. Avec l'apparition et la génération de nouvelles méthodes de séquençage à haut-débit, telles que la technologie 454 ou le séquençage Illumina, de bibliothèques plus imposantes ont pu être synthétisées au vu d'être séquencées.

La technologie Roche/454, développée par la société Roche Diagnostics et introduit commercialement en 2004, est une technologie mêlant une première étape de PCR en émulsion puis une deuxième étape de pyroséquençage.

Dans cette approche, la librairie d'ADN simple brin est préparée en incorporant deux adaptateurs 454 aux fragments d'ADN, un en 3' et l'autre en 5'. La librairie est ensuite incubée avec des billes d'agarose préalablement fonctionnalisées par les séquences complémentaires des adaptateurs 454. Chaque bille est ainsi associée à un fragment d'ADN unique. Les complexes fragments-billes sont ensuite isolés individuellement dans des micelles (microréacteurs) contenant les réactifs de PCR par émulsion en milieu hétérogène huile : eau. Le cyclage thermique des micelles (*i. e* la PCR en émulsion) produit ainsi plusieurs millions de copies de chaque fragment d'ADN à la surface de chaque bille. Les molécules uniques amplifiées sont ensuite séquencées en masse. Pour ce faire, les billes sont déposées sur une plaque PTP (en. *Pico Titer Plate*). Chacune bille prise séparément est alors insérée dans un des milliers de puits présents sur la plaque, permettant ainsi de fournir un emplacement fixe au suivi de chaque séquençage.

Les enzymes catalysant les étapes de réaction de pyroséquençage sont ensuite ajoutées au PTP. Les nucléotides, eux, vont être ajoutés les uns après les autres dans chaque microréacteur. L'incorporation du nucléotide va envoyer un signal chimie luminescent qui va différer suivant la nature du nucléotide (A, T, G ou C). Une caméra positionnée en face du PTP enregistre l'émission de luminescence de chaque puits, permettant ainsi, après traitement des données, de remonter à la séquence de l'ADN amplifié.

Cette technologie a été utilisée à plusieurs reprises lors de séquençage de code-barre ADN issus de bibliothèques encodées^{204,205,248}. En revanche, alors qu'il est nécessaire d'utiliser une plaque PTP entière pour une bibliothèque de plus d'un million de séquences à code-barre unique^{205,248,249}, l'utilisation de la technologie 454 pour des bibliothèques présentant plus grande s'avère coûteuse et peut ainsi devenir un facteur limitant à l'utilisation de cette méthode de séquençage^{202,213}.

La technologie de séquençage Illumina (Figure 47) utilise l'approche de séquençage par synthèse (en. *sequencing-by-synthesis*) où les nucléotides (dNTP) ainsi que l'ADN polymérase sont ajoutés simultanément sur les lignes fluidiques afin d'être incorporés.

De manière similaire à la méthode Roche/454, la bibliothèque d'ADN double brin utilisée doit être encodée par des adaptateurs Illumina de chaque côté. Après dénaturation, les ADN simples brins sont hybridés aléatoirement sur une surface solide, également appelée *Flow-cell*, composés de lamelles de verre comportant des lignes fluidiques où s'effectue la chimie du séquençage. Une fois hybridé avec un des adaptateurs présents sur la *flow-cell*, le brin complémentaire de l'ADN simple brin va être synthétisé.

La synthèse de ce brin complémentaire va se faire en plusieurs étapes successives. En effet, les dNTP utilisés portent un marquage fluorescence unique à chaque base ainsi qu'un groupement 3' — OH chimiquement inerte, c'est-à-dire protégé. Cette protection permet ainsi d'enregistrer une incorporation unique à chacune des séquences présentes sur la *flow-cell*. Les dNTP n'ayant pas réagi ainsi que l'ADN polymérase en excès sont retirés de la *flow-cell* par des lavages successifs. L'ajout d'un tampon d'analyse (en. *scan buffer*) est ensuite ajouté afin que le système optique de détection de fluorescence puisse enregistrer l'incorporation du nucléotide à chaque position de la *flow-cell*. Après imagerie complète de la plaque, des entités chimiques capables de cliver les marqueurs fluorescents ainsi que de déprotéger le groupement hydroxyle en 3' sont ajoutées sur la *flow-cell*. Après dernier lavage, la plaque est prête pour l'ajout de nouveaux dNTP ainsi que de l'ADN polymérase pour l'incorporation fluorescente du nucléotide suivant²⁵⁰.

Par ce processus itératif d'incorporation de nucléotide et d'imagerie, il est possible de suivre la synthèse du brin complémentaire à l'ADN matrice simple brin. Celui-ci est ensuite dénaturé afin que l'extrémité du brin complémentaire néo-synthétisé puisse former un pont (en. *bridge amplification*) avec le deuxième adaptateur présent sur la *flow-cell*. La synthèse du brin matriciel va alors permettre l'obtention de deux brins attachés de façon covalente à la cellule de flux. Cette série d'étapes va être perpétuée selon convenance de l'utilisateur afin de

généraliser des clusters de fragments identiques. Le séquençage est alors effectué à chaque amplification d'ADN simple brin.

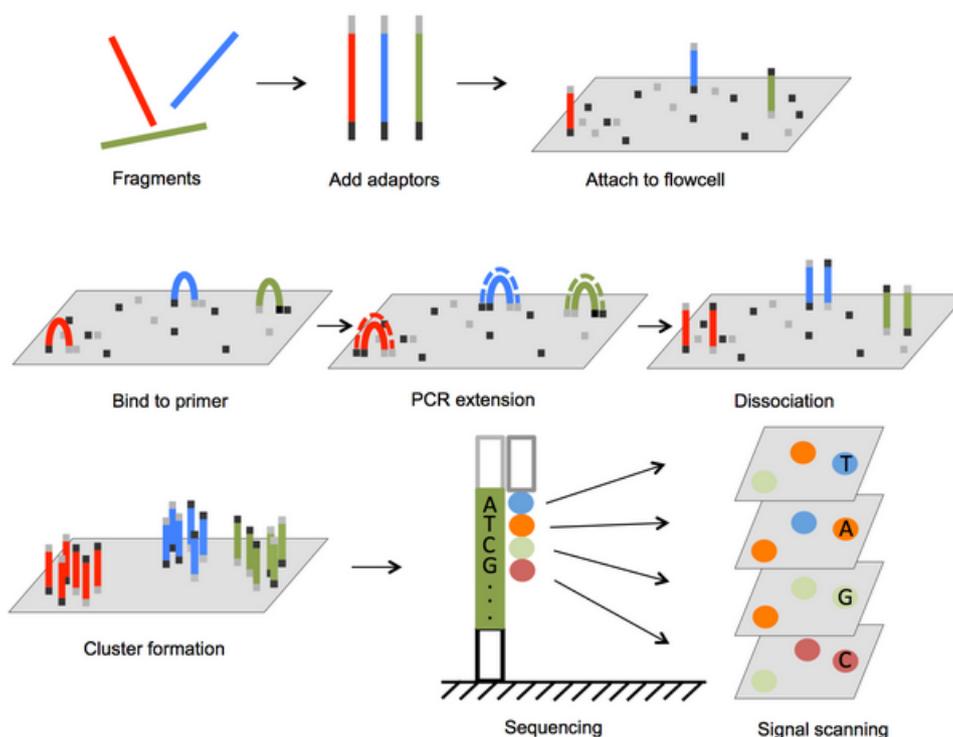


Figure 47 : Schéma explicatif de la technologie de séquençage Illumina. Image tirée du livre « Next generation Sequencing – Advances, Applications and Challenges » - Chapitre 2²⁵¹.

Le séquençage Illumina de librairie encodée par l'ADN a également été grandement décrit dans la littérature^{204,205,252}. La comparaison entre les deux méthodes de séquençage à haut-débit a finalement été réalisée par le groupe de D. Neri en 2010²⁵². L'étude, menée sur une librairie encodée de 4000 composés, montre que les profils d'enrichissement obtenus sont similaires, quelle que soit la méthodologie considérée. Il est néanmoins à noter que l'utilisation du séquençage Illumina permet le décodage de librairie de plus d'un million de composés, permettant l'application générale de ce séquençage à des librairies de tailles variables.

Le traitement informatique des événements enregistrés lors du séquençage permet ainsi d'identifier la fréquence (*i. e.*, le nombre de fois que la séquence d'ADN code-barre est lu lors d'une expérience de séquençage à haut-débit) d'un membre de la librairie au sein même de celle-ci, avant ou après l'application de processus de sélections par affinité. L'application de filtres aux fréquences enregistrées permet la visualisation des meilleures molécules au sein d'une expérimentation (*i. e.*, les composés les plus enrichis après une procédure de sélection). *In fine*, le volume considérable d'information structure-activité obtenue permet, au-delà de la molécule unique, d'identifier des groupes moléculaires favorables à l'interaction ainsi que des familles de ligands. Parallèlement, la spécificité des composés identifiée peut être identifiée grâce à de multiples sélections contrôles.

Section 2.2.4 : Resynthèse des touches et études d'interactions.

L'ADN est un précieux marquage pour l'encodage et le décodage de bibliothèques chimiques. Toutefois, il est désirable de caractériser les ligands identifiés par sélections de DEL en l'absence de leur code-barre ADN.

Dans la plupart des cas, la confirmation des touches sélectionnées lors d'un processus de sélection passe par la resynthèse des composés les plus enrichies sans leur code-barre ADN. Après resynthèse, la caractérisation des propriétés d'association contre une, ou plusieurs, cible d'intérêt est effectuée par l'intermédiaire d'essais biochimiques et/ou biophysiques. La synthèse et la purification de molécules organiques peuvent parfois s'avérer compliquées suivant la structure des composés et le nombre de molécules sélectionnées à resynthétiser.

Les cibles étudiées peuvent elles aussi être un problème lors de mesure, par exemple, de constante d'affinité. En effet, si la structure de la cible n'est pas maintenue lors de l'analyse, si l'activité de la cible est perdue suite à l'immobilisation sur un support solide, alors les caractérisations biophysiques peuvent présenter des difficultés. Dans ce genre de cas, il est souvent recommandé de synthétiser les molécules sélectionnées en fonctionnalisant non plus la molécule par un ADN code-barre mais bien par un marqueur fluorescent. Ceci permet ainsi l'utilisation de la molécule au sein d'expériences de polarisation de fluorescence par exemple²⁰⁴.

Dans certains cas, il est pratique de faire la synthèse d'une molécule sélectionnée en présence de son code-barre ADN^{253,254}. En effet, l'équipe de G. Zimmermann a décrit des méthodes de resynthèse de molécules basées sur l'hybridation de conjugués oligonucléotidiques avec des brins fonctionnalisés par un fluorophore ou encore une autre diversité chimique (Figure 48).

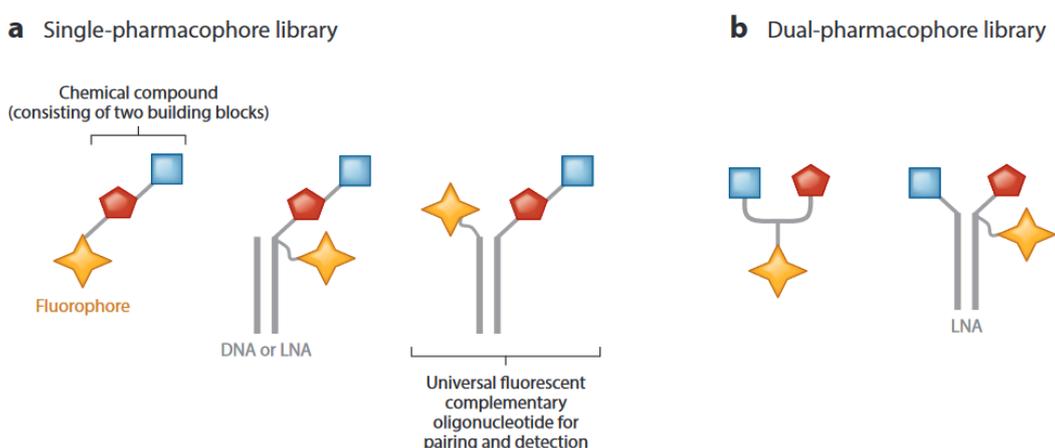


Figure 48 : Méthodes de resynthèses de composés issus de DEL en fonction de la stratégie Single — ou Dual-pharmacophore¹⁹⁵.

L'avantage de la resynthèse en présence d'ADN (code-barre) est que les étapes de synthèses sont connues et découlent des étapes de construction de la bibliothèque chimiquement encodée. De plus, la présence de séquences oligonucléotidiques facilite la purification des conjugués ainsi que leur caractérisation par spectrométrie de masse notamment.

Sous-partie 2.3 : De la touche au candidat médicament

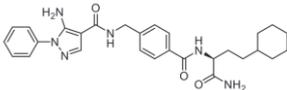
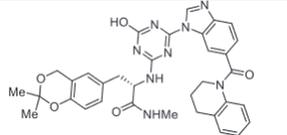
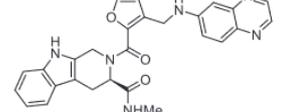
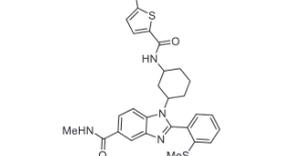
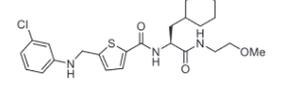
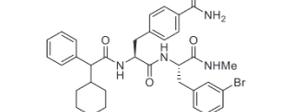
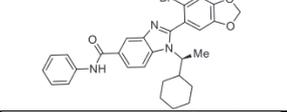
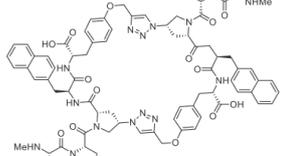
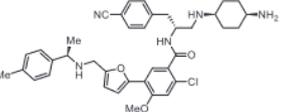
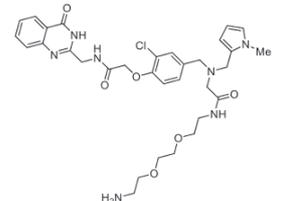
L'utilisation de différentes méthodologies d'encodage par l'ADN couplé aux concepts de sélection, d'amplification et de séquençage décrits plus haut dans ce manuscrit ont permis à de nombreux groupes de faire émerger des molécules à haute affinité pour de larges gammes de protéines d'intérêts pharmaceutiques^{208,243,255–257,257–268,268,269} ainsi que des cibles nucléiques^{188,270}.

En considérant la synthèse des bibliothèques combinatoires encodées de ces vingt dernières années, il est possible de noter deux tendances générales concernant la diversité des ligands encodés. La première tendance est la synthèse de composés *drug-like* respectant les règles de Lipinski (en. *Lipinski's rule of five*, RO5^{271,272}). De manière générale, les bibliothèques décrites dans la littérature présentent deux ou trois diversités chimiques. Parmi celles-ci, seule une voire deux entités chimiques contribue à la reconnaissance par la cible. Ainsi, même si la somme de toutes les entités chimiques présentes sur une molécule dépasse les 500 Da, les entités chimiques prises indépendamment ne dépassent pas les RO5. En effet, les ligands possédant généralement deux fragments présentent une masse moyenne par fragment légèrement supérieur à 200 Da, alors que les ligands présentant trois fragments présentent une masse moyenne en dessous de 160 Da par fragment chimique^{255,273,274}.

La deuxième tendance générale dans la synthèse de ligands encodés est bien évidemment la synthèse de ligands ne respectant pas les RO5 décrites par Lipinski. Parmi les molécules décrites on retrouve notamment la synthèse de macrocycle (d'origine peptidique ou de produits naturels) capable de faciliter les événements de reconnaissance avec certaines cibles dus à large surface généralement plane^{268,269,275,276}.

Dans le tableau 2, une liste non-exhaustive de ligands identifiés contre des cibles protéiques (Kinases, Aminotransférases, Phosphatases, récepteurs couplés aux protéines G (RCPG), facteurs de transcription, interaction protéine-protéine) et acides nucléiques est présentée. À ce jour, 5 composés issus des technologies de sélection DEL sont en [phases cliniques](#).

Tableau 2 : Tableau illustratif des composés marquants ciblant des protéines ou des acides nucléiques issus de campagnes de bibliothèques combinatoires encodées.

Cible	Famille	Nom/Molécule	Caractéristiques
Protéique	Kinase (p38 MAPK inhibitor)	 VPC00628	Vipergen ²⁵⁸ DEL : YoctoReactor/ 12.6 M IC ₅₀ = 7 nM
Protéique	Kinase (p38 MAPK inhibitor)		Praecis Pharmaceuticals (GSK) ²⁰² DEL: DNA recorded/ 800 M IC ₅₀ ≤ 7 nM
Protéique	Kinase (BKT inhibitor)		X-Chem ²⁷⁷ DEL: DNA recorded / 110 M IC ₅₀ = 0.55 nM
Protéique	Aminotransferase (BCAT inhibitor)		GSK ²⁷⁸ DEL: DNA recorded/ 117 M IC ₅₀ = 251 nM
Protéique	Phosphatase (Wip1 inhibitor)		GSK ²⁰² DEL: DNA recorded/ 800 M IC ₅₀ = 13 nM
Protéique	RCPG (β2-adrenergic allosteric inhibitor)		Lefkowitz group /Nuevolution ²⁷⁹ DEL: DNA recorded/ 190 M K _d = 1.7 μM
Protéique	RCPG (PAR2 antagonist)		X-Chem/AstraZeneca ^{280,281} DEL: DNA recorded/ 7.1 M IC ₅₀ = 90 nM
Protéique	Interaction protéine/protéine (PPI) (XIAP antagonist)		Ensemble/Bristol-Myers Squibb ²⁸² DEL: DNA directed chemistry/ 160 K IC ₅₀ = 24 nM
Protéique	Facteur de transcription (Bromodomain ATAD2 inhibitor)	 Optimized hit from DEL campaign	Bayer/X-Chem/Edeltris ²⁸³ DEL: DNA recorded / 110 M IC ₅₀ = 157 nM K _d = 115 nM
Nucléique	G-quadruplexe (c-Myc G4)		X-Chem ¹⁸⁸ DEL: DNA recorded / 218 M K _d = 1 μM

Partie 3 : Projet de thèse.

Près de trois décennies après sa conceptualisation en 1992, la synthèse de bibliothèques chimiques encodées par l'ADN (DEL) est devenue une technologie de choix pour l'identification de pharmacophores d'intérêts dans la découverte de candidats médicaments. Cette technique innovante permet la synthèse, le criblage et l'analyse rapide et économique de collections de composés encodés contre des cibles biologiques *via* l'utilisation de techniques de chimie combinatoire, de biologie moléculaire et de séquençage à haut-débit. Depuis une vingtaine d'années, le nombre de publications scientifiques illustrant l'utilisation efficace de DEL au sein de campagnes d'identifications de *hits* n'a cessé d'augmenter. Cette technique de *drug discovery*, largement répandue au sein de groupe pharmaceutique (GSK, Sanofi, etc.) mais également au cœur des laboratoires de recherches publiques, est aujourd'hui principalement utilisée et largement documentée pour le ciblage de protéines d'intérêts. À la date d'écriture de ce manuscrit, le ciblage de structures d'acides nucléiques par la technologie DEL reste très largement inexploré. Seule, la compagnie *X-Chem Pharmaceuticals*¹⁸⁸ (2019) a montré qu'il était possible de réutiliser des DELs préalablement conçues (*i.e* synthétisées dans le cadre de campagnes d'identification contre diverses cibles protéiques) afin d'identifier de petites molécules capables de cibler des structures G-quadruplex.

Ainsi, ce projet de thèse a pour objectif d'initier **la conception de librairies de ligands chimiques encodées par l'ADN spécialement dédiées à la reconnaissance spécifique de structures d'acides nucléiques et en particulier d'ADN G-quadruplex**. Cette réalisation sans précédent a donc également pour but de démontrer que la technologie DEL peut être étendue à un champ nouveau de cibles d'intérêts biologiques et thérapeutiques.

Au cours de cette thèse deux librairies spécifiques aux structures G-quadruplex, appelées DEL1 et DEL2, ont été synthétisées. Ces librairies ont été choisies à la fois pour leur pertinence vis-à-vis des cibles G4, mais également pour la faisabilité de leur assemblage. La librairie DEL1 a notamment été conçue afin de contenir des molécules présentant une forte probabilité d'affinité pour les structures G-quadruplex et ceux afin d'augmenter les chances de succès lors des premières campagnes de sélection. Cette première librairie a ainsi été conçue à partir de certains motifs (poly)hétéroaromatiques déjà connus dans la littérature pour leurs propriétés d'associations avec les séquences G-quadruplex. Parmi ces motifs il est possible de citer la Phénanthroline (PDC3) ou encore la Pyridostatine (PDS). Afin d'améliorer leurs propriétés de reconnaissance des structures G-quadruplex, ces motifs (BB3 ; Figure 49) ont été associés à des éléments de reconnaissances d'origine peptidique (BB1 et BB4 ; Figure 49) dont l'objectif attendu est d'effectuer des interactions complémentaires avec les boucles et/ou les sillons des structures G-quadruplex. L'association des trois entités chimiques (BB1, BB3 et BB4) est réalisée par l'intermédiaire d'un cœur trifonctionnel (BB2 ; Figure 49) et permet ainsi d'obtenir une librairie de 190 440 composés distincts (Figure 49).

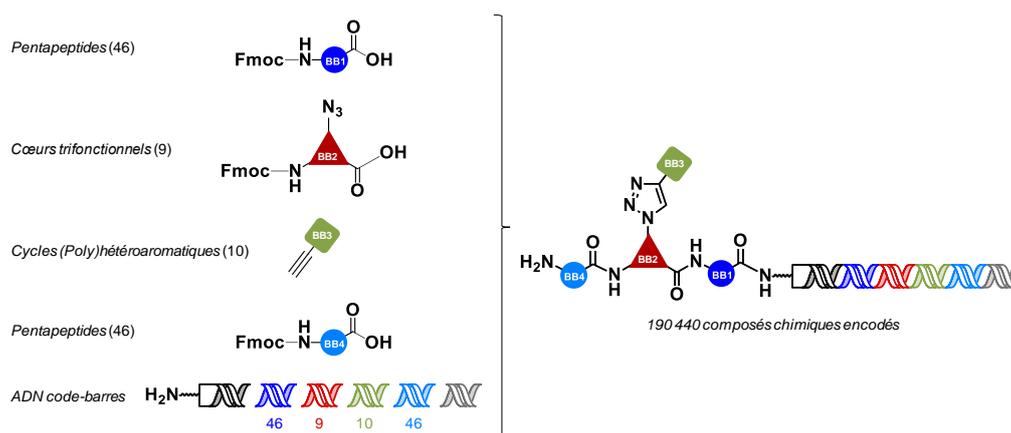


Figure 49 : Mode de construction de la DEL1 – Formation de 190 440 composés uniques.

Dans un deuxième temps, une seconde librairie encodée (DEL2) a été assemblée. Cette librairie est également composée de fragments susceptibles d'interagir avec les G-quadruplex mais ne contient pas les motifs hétéroaromatiques π -stacker de plus haute affinité pour les tétrades de guanine, et ce afin d'explorer d'autres modes possibles d'interactions. Cette librairie est alors formée par l'assemblage itératif de séquences pentapeptidiques (BB1 et BB4 ; Figure 50), de motifs poly- et macrocycliques trifonctionnels (BB2 ; Figure 50) et de motifs chimiques additionnels (BB3 ; Figure 50) d'une grande variété. La DEL2, composé de 696 200 ligands encodés uniques, dispose de toutes les fonctions pouvant permettre la reconnaissance de structures G-quadruplex (Figure 50).

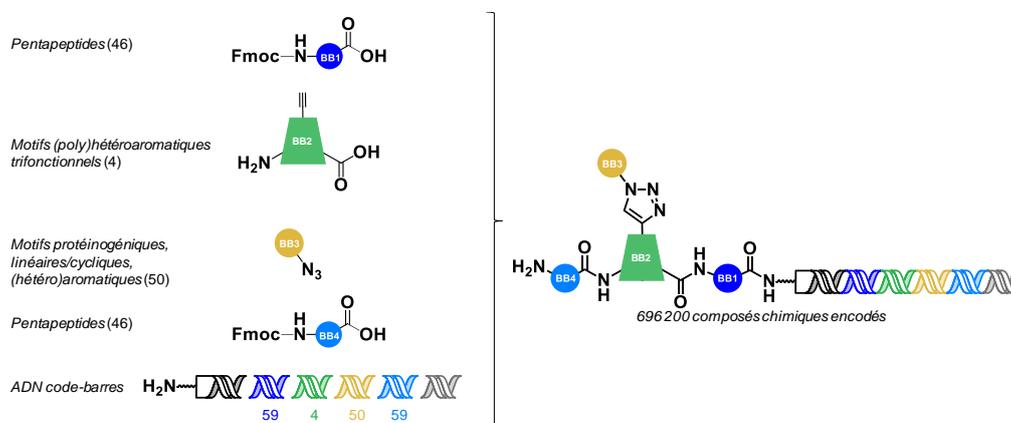


Figure 50 : Mode de construction de la DEL2 – Formation de 696 200 composés uniques.

À partir des librairies DEL1 et DEL2, le projet a consisté à mettre en place des processus de sélection *in vitro* par affinité afin de sélectionner les composés les plus affins et les plus spécifiques des structures G-quadruplex. Parmi les cibles majoritaires, on retrouve les ADN G-quadruplex des séquences promotrices d'oncogènes (Myc, Kit) et des séquences télomériques humaines (wtTel26, 22CTA). La technologie DEL et la mise en place de sélection en parallèle pourraient également permettre de partir à la recherche de ligands présentant une certaine spécificité pour un G-quadruplex en particulier, ce qui consiste un des objectifs d'actualité dans le domaine. Ainsi, en parallèle des sélections par affinités, des expériences de compétitions entre différentes structures G-quadruplex ont également été réalisées.

L'identité des ligands isolés par ces processus de sélections par affinité a ensuite été élucidée par séquençage à haut-débit. Au-delà d'un simple classement, les données considérables obtenues par séquençage nous ont permis de définir des tendances et des profils communs à certaines familles de ligands reconnaissant particulièrement certaines structures G-quadruplex.

Enfin, une fois resynthétisés sans codes-barres ADN, les ligands les plus prometteurs et des molécules contrôles ont pu être évalué en s'appuyant sur des méthodes biophysiques largement décrites dans la littérature afin de décrire l'interaction entre les G-quadruplex et leurs ligands.

Les résultats obtenus au cours de ces différents processus se sont montrés robustes, répétables et permettent de confirmer la validité de l'approche DEL pour le ciblage des G4. Ces travaux sont décrits en détail au sein des trois chapitres qui poursuivent ce manuscrit.

Chapitre 2 — Assemblage de bibliothèques combinatoires encodées par l'ADN ciblant l'ADN G-quadruplex.

Partie 1 : Conception et synthèse de la première génération de ligands encodés par l'ADN.

Préambule.

Au cours de cette thèse, les bibliothèques combinatoires ciblant les G-quadruplex ont été construites suivant la méthode de synthèse par enregistrement, aussi connu sous le nom de *DNA encoded* (voir Chapitre 1, partie 2, section 2.1.1, p65). Pour rappel, cette technique permet la synthèse d'une bibliothèque chimique encodée où la partie chimique est attachée à la partie duplex d'ADN soit par l'intermédiaire d'un brin d'ADN, soit par l'intermédiaire de deux brins d'ADN. L'utilisation d'un ADN double brin au détriment de code-barres ADN simple brin s'explique par la nature plus inerte de l'hélice formée par l'ADN. En effet, les bases engagées par des liaisons hydrogène au sein de l'hélice sont moins susceptibles de subir des lésions potentiellement induites par la synthèse de la bibliothèque de composés chimiques. Il est également possible de postuler que l'ADN double brin sera moins susceptible de jouer un rôle, ou d'interférer, dans la liaison avec la cible lors des étapes de sélections.

Dans le cas des deux bibliothèques combinatoires DEL1 et DEL2 présentées dans ce manuscrit, respectivement pour la bibliothèque de première et de deuxième génération, le précurseur utilisé afin de lier la partie chimique de la partie code-barre ADN est un espaceur bidirectionnel connu sous le nom d'*Headpiece*^{202,284,285}. Ce précurseur est composé de deux simples brins d'ADN attachés de manière covalente par l'intermédiaire d'un espaceur chimique présentant une fonction amine primaire libre (Schéma 51). Les deux séquences simples brins, de 17 et 19 bases respectivement, forment un duplex d'ADN dont l'extrémité 3' du brin supérieur possède deux bases supplémentaires. La fonction amine libre permet d'étendre les diversités chimiques alors que l'extrémité non cohésive de l'ADN permet d'étiqueter chaque diversité *via* un code-barre ADN unique par ligation enzymatique.

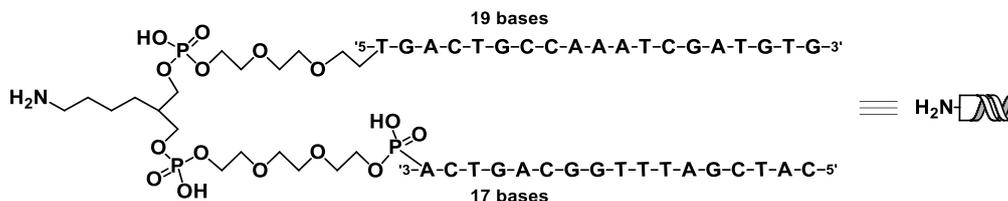


Figure 51 : Structure de l'espaceur bidirectionnel, l'Headpiece.

Non seulement les nucléobases sont protégées des modifications chimiques au sein d'une structure duplex, mais l'apport réel de cet espaceur est de maintenir covalent les deux séquences simples brins. En effet, si les deux brins se dénaturent sous applications de conditions de synthèses dures, chaque brin unique peut se réhybrider avec son brin complémentaire dans des conditions favorables grâce à leur lien covalent. Ceci permet de

limiter des hybridations non contrôlées avec des brins étrangers et d'empêcher la perte de code-barres ADN.

La volonté de former une bibliothèque de molécules encodées est corrélée au fait de pouvoir cribler la librairie contre une cible immobilisée. Il est donc important que l'ADN codant n'interfère pas dans l'interaction avec la cible. De plus, dans l'intérêt de fournir une plus grande distance entre la molécule et le code ADN, l'amine de l'*Headpiece* est étendue par une N-Fmoc-N''-succinyl-4,7,10-trioxa-1,13-tridecanediamine arbitrairement appelée *Extended (Ext)* (Schéma 1).

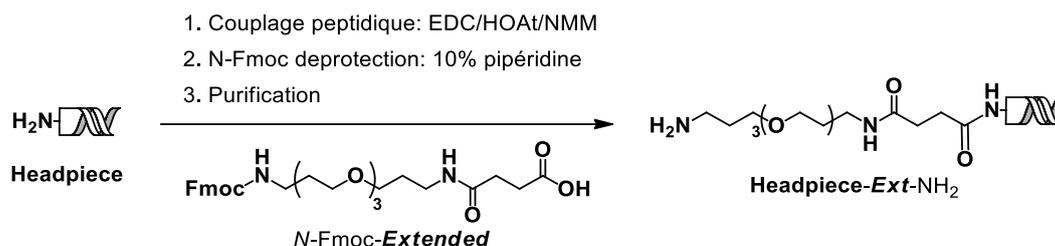


Schéma 1 : Distanciation de la partie chimie de l'ADN codant par couplage d'un espaceur N-Fmoc polyéthylène glycol (PEG) nommé N-Fmoc-Ext.

L'incorporation du *N-Fmoc-Ext* sur l'*Headpiece* a été réalisée par couplage peptidique en présence de l'agent de couplage EDC, de l'agent d'activation HOAt et de la base NMM. S'en suit une étape d'élimination du groupement protecteur Fmoc en milieu aqueux contenant 10% de pipéridine. Après précipitation à l'éthanol et purification sur cartouche de purification en phase inverse (Glen-Pak™), l'*Headpiece-Ext-NH₂* a été aliquoté en grosse quantité et maintenu sec à -20°C.

La stratégie d'encodage utilisée dans la stratégie *DNA recorded* est basée sur la ligation successive de petites séquences d'ADN double brin de 7 paires de bases (Figure 52). Ces séquences code-barres ont la particularité de présenter deux bases supplémentaires du côté 3' de chaque simple brin. Ces extrémités non-cohésives permettent la ligation itérative de code-barres utilisés pour chaque cycle de la librairie. Il est à noter que des exemples d'utilisation d'ADN double brin de longueur supérieure à 7 paires de bases ont également été documentées dans la littérature. L'avantage d'utiliser des séquences plus courtes est de pouvoir plus facilement séparer les code-barres de la librairie lors de processus de purification, par exemple, par HPLC.

De manière générale, les code-barres ADN utilisés à chaque cycle de la librairie ont été achetés sous forme de simple brin chez un revendeur (Integrated DNA Technologies, IDT). Les séquences simples brins obtenues sous forme de solides lyophilisés en plaque 96-puits ont dû être hybridées avec les partenaires afin de former les code-barres ADN doubles brins souhaités. Chaque brin a été dissous à une concentration de 2 mM avant d'être combiné dans un ratio 1:1 au sein d'une plaque 96-puits. Ces plaques sont alors utilisées comme source d'ADN code-barre pour chaque cycle de ligation. Entre ces cycles, les plaques peuvent être stockées à -80°C.

Dans le processus d'encodage, l'ADN code-barre est ajouté à hauteur de 2 équivalents afin de s'assurer d'une ligation complète. Toutefois, l'utilisation de code-barres en excès dans le mélange réactionnel implique la présence d'ADN supplémentaires en fin de cycle. Leurs présences peuvent être problématiques dans la mesure où ils peuvent participer au processus de ligation du cycle suivant. Si cela était le cas, une chaîne ADN dépourvu de diversité chimique grandirait par itération des processus de ligations. Plus la séquence ADN formée est grande, plus il est compliqué de la séparer des composants de la librairie combinatoire encodée d'intérêt. Diverses méthodes de purifications peuvent être appliquées pour pallier ce problème.

Comme indiqué plus haut, la première molécule utilisée dans la formation de librairie encodée est l'*Headpiece* fonctionnalisée par l'extension *Ext* (*Headpiece-Ext-NH₂*). Au-delà de ces particularités de maintien de l'ADN sous forme de duplex, l'*Headpiece* a également été conçue pour être point d'initiation des réactions de PCR, utilisées plus tard afin d'amplifier les touches obtenues de la librairie combinatoire après un processus de sélection. De manière similaire, une séquence terminatrice (en. *closing primer*, CP) commune à chaque membre de la librairie est attachée aux séquences code-barres à la toute fin du dernier cycle itératif de la librairie. Due à ces 27 paires de bases, la séquence terminatrice est utilisée afin d'encoder des informations supplémentaires, ce qui permet notamment de distinguer deux librairies combinatoires distinctes. Ces informations sont retrouvées au sein d'une région variable du CP désignée par X et plus globalement tout le schéma d'encodage des librairies DEL1 et DEL2 présenté en Figure 52.

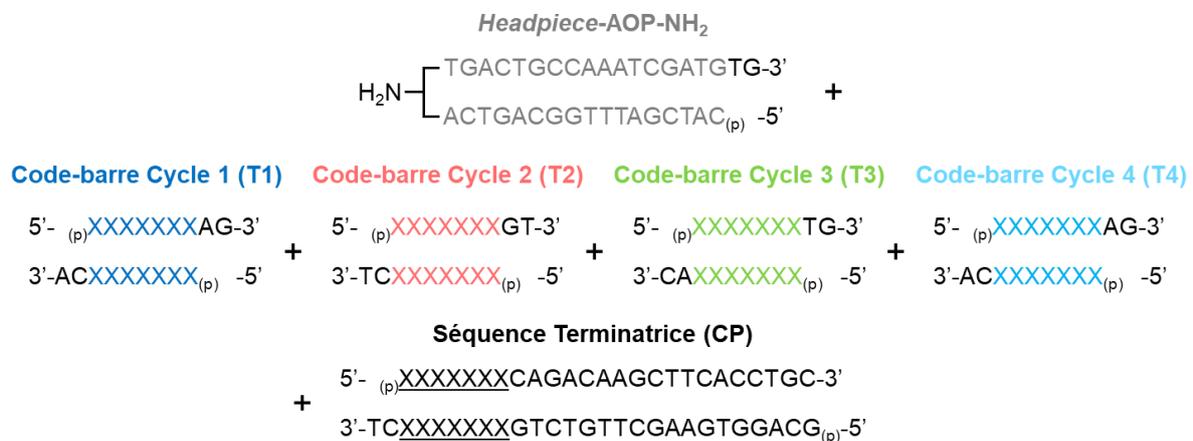


Figure 52 : Stratégie d'encodage appliquée à la DEL1. Chaque segment d'ADN (T1 à T4) va être enzymatiquement lié à l'*Headpiece* de départ par l'intermédiaire des leurs extrémités non-cohésives en 3'. Les extrémités 5', elles, exhibent un groupement phosphate libre. La région X des code-barres T1 à T4 représente des oligonucléotides non définis qui coderont pour les entités chimiques utilisées dans chaque cycle de synthèse. La séquence terminatrice (CP) contient une plus longue séquence parmi laquelle on retrouve les variables identifiables par le symbole X souligné (X).

Même si un nombre important de ligases ont été documentées dans la littérature (Taq DNA ligase, T7 DNA ligase, *E.coli* DNA ligase), celle qui est préférentiellement utilisée par les

acteurs de l'approche DEL est l'ADN T4 ligase²⁰². Cette ligase est référencée pour être particulièrement efficace dans l'incorporation de séquences portant des extrémités non-cohésives²⁸⁶.

Les expériences de ligations enzymatique ont été effectuées en parallèle dans un tube de microcentrifugation (1,5mL) par ajout successif d'*Headpiece*, d'un code-barre ADN, de l'ADN T4 ligase ainsi que du tampon de ligation. Après réaction sur la nuit à 16°C, la ligation enzymatique est généralement quantitative. Toutefois, il est important d'analyser chaque ligation afin d'évaluer son succès. Le meilleur moyen d'analyser une ligation d'ADN est de faire de l'électrophorèse en utilisant soit un gel de polyacrylamide soit un gel d'agarose. Dans le cadre de ce projet, l'analyse des ligations a été effectuée par gel d'agarose. En effet, les gels d'agarose sont globalement plus simples et plus rapides à mettre en œuvre.

Avant d'initier l'assemblage des librairies DEL1 et DEL2, nous avons souhaité vérifier l'efficacité du schéma d'encodage proposé. La Figure 53 ci-dessous montre l'évolution de la taille de l'ADN duplex au fur et à mesure de l'incorporation des code-barres ADN, de l'*Headpiece* jusqu'au *Closing primer*.

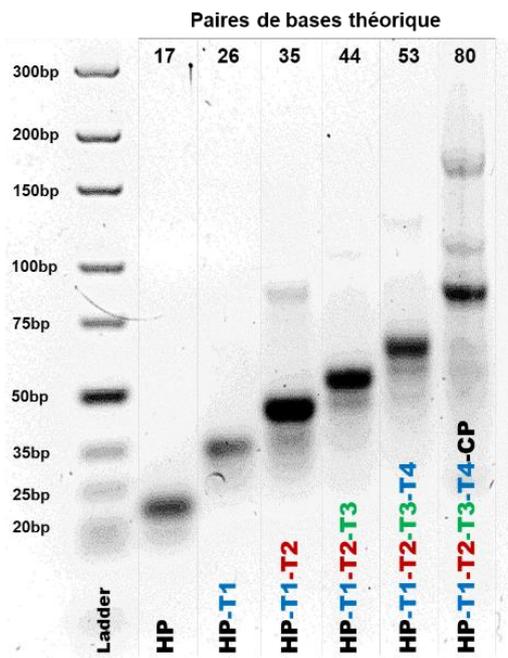


Figure 53 : Assemblage d'un code-barre ADN complet après quatre cycles de ligation (T1 à T4) et ajout du closing primer (CP). Augmentation croissante des masses moléculaires de la librairie après ligation enzymatique. Gel à 5% d'agarose UltraPure (disponible commercialement) précoulé avec du GelRed™.

Migration à 90V pendant 2 h puis exposition à une lampe UV (GelDoc) pour caractérisation.

Des purifications HPLC ou des filtrations sur membrane à exclusion de taille peuvent être utilisées pour éliminer l'excès de code-barres après chaque ligation comme nous verrons pour l'assemblage des DEL1 et DEL2. Cet assemblage de code-barres test a été séquencé avec succès, selon une procédure que nous verrons également plus loin dans ce manuscrit, validant l'approche globale d'encodage.

Après vérification de l'insertion individuelle des code-barres ADN par gel d'agarose à chaque cycle, c'est au tour des entités chimiques d'être insérées. Un des aspects remarquables à l'utilisation de librairie encodée est que la synthèse chimique peut être effectuée à partir d'entités disponibles commercialement. De plus, seuls quelques milligrammes de chaque produit suffisent à étendre la diversité chimique d'une librairie. Bien que le prix et la quantité disponible de toutes les entités chimiques puissent varier considérablement, une collection diversifiée de blocs de construction peut rapidement être collectée en achetant des quantités minimales à divers fournisseurs.

Lors de la mise en place des stratégies de synthèses appliquées à une librairie combinatoire encodée, il est essentiel de considérer des réactions compatibles avec la présence d'ADN. Comme discuté précédemment, cette macromolécule est par nature soluble dans l'eau mais également sujette à dégradation en présence de conditions acides ou par application de températures extrêmes. Pour des réactions chimiques sur molécules d'ADN, il est ainsi nécessaire d'utiliser des conditions adaptées, généralement en milieu aqueux, ou tout au moins au sein de conditions hétérogènes organiques/aqueuses (DMSO, DMF ou DMA/H₂O par exemple).

Lorsque les entités chimiques sont à disposition et que les stratégies de synthèse en présence d'ADN sont en place, il est nécessaire de cribler un à un les blocs de construction afin d'optimiser leurs incorporations dans la librairie. Cette étape va permettre de mesurer individuellement l'efficacité de chaque réaction sur un système modèle. Dans le cas de la DEL1, les entités chimiques ont été testées sur l'espaceur bidirectionnel, l'*Headpiece-Ext-NH₂* ou sur un analogue fonctionnalisé de manière adéquate. L'efficacité des synthèses a ensuite été déterminée par analyse UPLC-MS. Les entités présentant des rendements de synthèse satisfaisants (>60%) peuvent alors être utilisées au sein de la librairie combinatoire de molécules. Ces étapes de validation de blocs de construction sont chronophages mais néanmoins importantes et nécessaires au bon déroulement de la synthèse de la librairie.

Sous-partie 1.1 : Composition, conception et validation des *building blocks* de la DEL1.

Comme indiqué précédemment, l'assemblage de la première librairie combinatoire encodée, DEL1, a été effectué par une stratégie *DNA recorded*. Cet assemblage itératif de code-barres ADN et d'entités chimiques a permis l'obtention d'une librairie de près de 200 000 composés uniques.

La conception rationnelle de molécules capables d'interagir avec les tétrades externes (Figure 54, A), les boucles (Figure 54, B) et les sillons (Figure 54, C) d'une structure G-quadruplex est particulièrement complexe. La plupart des ligands synthétisés jusqu'à présent sont principalement des ligands de tétrades ou des ligands de sillons. La DEL1 ambitionne de proposer des ligands peptido-organiques capables de pouvoir interagir avec les trois composantes des structures G-quadruplex dans le but d'augmenter l'affinité du ligand pour sa cible et pourquoi pas, d'augmenter la spécificité des ligands pour des structures G-quadruplex particulières.

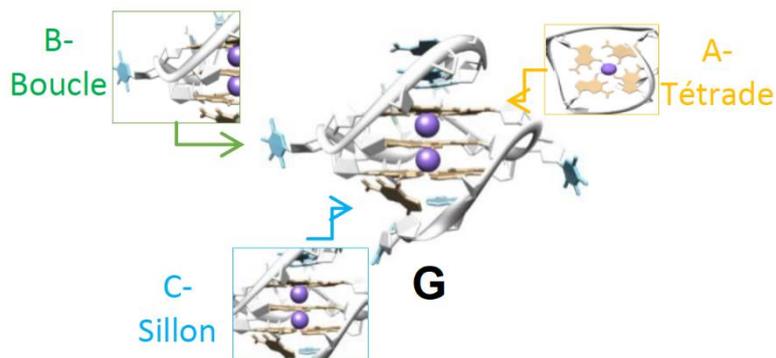


Figure 54 : Représentation schématique des zones distinctes composant les structures G-quadruplex. A- Tétrades de guanines ou G-quartets. B-Boucles et C-Sillons des structures G-quadruplex.

Les ligands décrits dans la cadre de la DEL1 portent en leur centre des motifs trifonctionnels d'origine peptidique (Figure 55, BB2) capable de relier en deux points des éléments de reconnaissances des acides nucléiques, à savoir des séquences peptidiques (Figure 55, BB1 et BB4), ainsi que divers motifs hétérocycliques sélectionnés pour leurs affinités avec les motifs G-quadruplex (Figure 55, BB3).

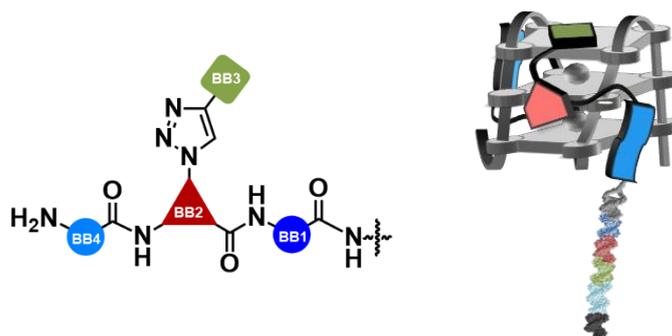


Figure 55 : Structure des ligands de la DEL1 par association de quatre buildings blocks.

Les blocs de construction (en. *building block*, BB) utilisés lors du premier cycle de librairie encodée [Cycle 1] sont des séquences peptidiques de cinq acides aminés (Figure 56) dont la structure générale se présente sous la forme $Fmoc-NH-GXXXG$, ou X correspond à 6 des 20 acides aminés protéinogéniques.

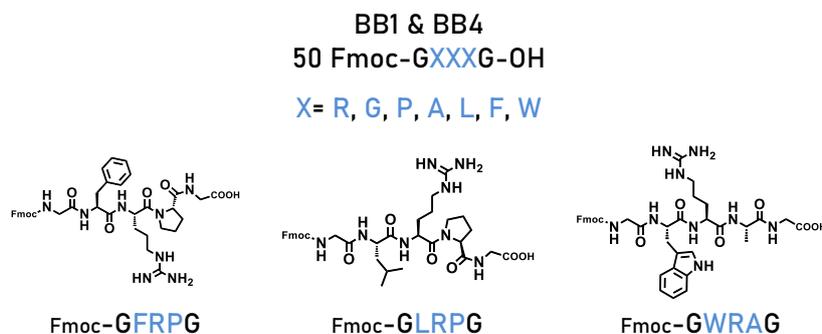


Figure 56 : Structure générale des séquences pentapeptidiques DEL1 — [Cycle 1/4] retenues.

En effet, introduit dans la librairie afin d'interagir au sein des boucles et/ou avec les sillons des motifs G-quadruplex, il est nécessaire de considérer une séquence peptidique suffisamment étendue dans l'espace. L'introduction de glycine (G) de part et d'autre des séquences peptidiques permet à la fois de se prémunir de toutes formes de racémisations lors des couplages peptidiques mais également d'homogénéiser les réactions de couplages puisque celles-ci se feront toutes sur le même acide aminé.

Initialement, le choix des acides aminés présents au sein de la séquence tripeptidique située entre les glycines s'est fait suivant plusieurs rationnels. Le premier a été de sélectionner les acides aminés les plus représentés dans l'interaction protéines/G-quadruplex, à savoir l'Arginine (R), le Tryptophane (W) ou encore la Phénylalanine (F). Dans un deuxième temps il était nécessaire de mettre de côté tous les acides aminés dont les chaînes latérales sont susceptibles d'interférer avec la stratégie de synthèse sans protection choisie, à savoir la Lysine (K), l'acide Aspartique (D), l'acide Glutamique (E), la Sérine (S), la Thréonine (T), l'Asparagine (N), la Glutamine (Q), la Cystéine (C), la Méthionine (M) et enfin la Tyrosine (Y). Un équilibre entre hydrophilie et lipophilie et dans le choix des AA a été considéré. Enfin, afin de limiter les achats et les coûts, une limite de redondance des acides aminés hydrophobes a été considérée et ainsi la Valine (V) et l'Isoleucine (I) ont été exclues. Compte tenu de ces considérations, les séquences pentapeptidiques initialement étudiées étaient alors composées de 7 acides aminés distincts comprenant deux acides aminés chargés positivement, Arginine (R) – Histidine (H), quatre acides aminés hydrophobes, Alanine (A) – Leucine (L) – Phénylalanine (F) – Tryptophane (W) et l'unique acide aminé protéinogène cyclique, la proline (P).

Un travail méthodologique de couplage des BB peptidiques a été réalisé sur l'*Headpiece-Ext-NH₂* par le Dr. Yang SI, postdoctorante au laboratoire (2017-2018) à partir de 114 séquences pentapeptidiques achetées auprès de GenScript et comprenant les 7 acides aminés retenus. La stratégie de couplage utilisée a suivi celle mise au point pour l'élongation de l'*Headpiece* par l'espaceur *N-Fmoc-Ext* à savoir l'utilisation d'EDC en tant qu'agent de couplage, d'HOAt en tant qu'agent activateur et la NMM en tant que base. Les tests de validation sur 2 nmol d'*Headpiece-Ext-NH₂* ont montré que seuls 30% des séquences testées présentaient des rendements supérieurs à 30%. Un changement de base a alors été effectué et l'utilisation de DIPEA a permis de faire passer 44% des séquences à un rendement supérieur à 60%, soit 50 séquences pentapeptidiques distinctes.

Il est intéressant de noter que parmi les 114 pentapeptides testés, 36 d'entre elles contenaient l'acide aminé histidine (H). Lors de l'étape de validation des BB, nous avons pu nous rendre compte que, quelles que soient les conditions de couplage utilisées, l'avancement de la réaction était inférieur à 15% pour ces séquences. Ainsi, parmi les 50 séquences retenues, aucune ne comporte l'acide aminé histidine, faisant ainsi passer à 6 le nombre d'acides aminés protéinogènes distincts au sein des séquences sélectionnées. De plus, une grande majorité des 50 séquences portent au moins une arginine, les peptides possédant le plus d'arginine étant généralement couplés avec les meilleurs rendements, certains à la faveur des meilleures

solubilisations lors des couplages. La liste des séquences sélectionnées ainsi que les rendements d'incorporation sont disponibles dans la section expérimentale (Tableau S4).

Les blocs de construction (BB) utilisés lors du deuxième cycle de librairie encodée [Cycle 2] sont des fragments trifonctionnels de types peptidiques (Figure 57) présentant une fonction amine (—NH-Fmoc), une fonction acide carboxylique (—COOH) ainsi qu'une fonction azoture (—N_3). La fonction acide carboxylique permet le greffage sur les BB1-amine, la fonction azoture permet le greffage des BB3-alcyne et la fonction NH-Fmoc , une fois déprotégée, permet l'incorporation des BB4-COOH. L'incorporation des BB2 a suivi la stratégie de couplage EDC/HOAt/NMM.

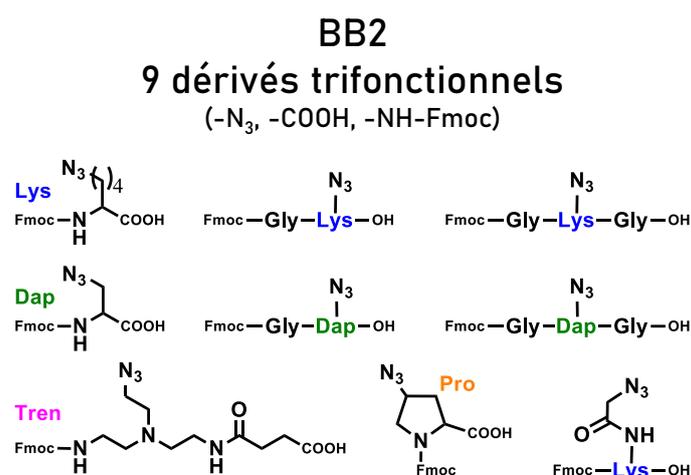


Figure 57 : Structure corps trifonctionnels DEL1 — [Cycle 2].

Au nombre de 9, ces BB2 peuvent être présentés sous forme de trois catégories distinctes (Figure 57). La première catégorie fait intervenir des motifs à longues chaînes latérales. On y retrouve la *N*-Fmoc-azido-norleucine (*i. e.*, une lysine portant un azoture en ϵ de sa chaîne latérale) noté K, mais également des dérivés glycines à savoir la *N*-Fmoc-K-G (KG) ainsi que la *N*-Fmoc-G-K-G (GKG). La deuxième catégorie, opposée à la première, comprend des dérivés trifonctionnels à chaîne courte. L'utilisation de l'acide aminé non-naturel DAP (noté D ; à ne pas confondre avec l'acide Aspartique non présent au sein de la librairie) et de ces dérivés *N*-Fmoc-D-G (DG) ainsi que le *N*-Fmoc-G-D-G (GDG) permettent ainsi de déployer des diversités chimiques (BB3) structurellement plus rapprochées. Enfin, la dernière catégorie est composée de motifs trifonctionnels variés. En effet, l'utilisation du composé Tren (T) permet d'apporter de la flexibilité à l'édifice. De plus, une proline (P) trifonctionnelle a été utilisée dans le but d'induire de la rigidité au système ainsi qu'une structuration particulière au sein de la séquence peptidique et donc au sein du ligand. Enfin, une lysine fonctionnalisée par un acide azidoacétique (Ke) a été utilisée afin d'augmenter la longueur de la chaîne latérale. Il est à noter que la présence de l'acide aminé glycine en position C-term et N-term des dérivés K et D offre divers niveaux de flexibilité aux chaînes peptidiques latérales pour engager des interactions à distance avec les différents sites d'interactions des G-quadruplex. Ces building blocks ont été préparés en parallèle de ces travaux de thèse par C. Van Wesemael (technicienne, 2018).

La validation des BB2 a été effectuée sur 2 nmol d'*Headpiece-Ext-GAWRG* (Figure 58) utilisé comme produit de départ modèle, dans les conditions EDC/HOAt/NMM. L'analyse des composés formés a été effectuée par UPLC-MS et des rendements d'incorporations quantitatifs ont été obtenus. Aucune optimisation supplémentaire n'a alors été réalisée.

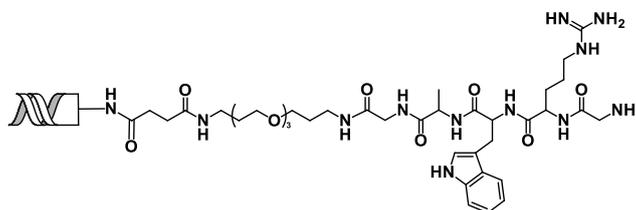


Figure 58 : Structure de l'*HP-Ext-GAWRG* utilisé pour la validation des 9 cœurs trifonctionnels du [Cycle 2].

Les blocs de construction (BB) utilisés lors du troisième cycle de librairie encodée [Cycle 3] sont des fragments (hétéro)aromatiques porteurs de fonctions alcyne. Ces fragments sont couplés par cycloaddition 1,3 — dipolaire catalysée par le cuivre (ou réaction de click- Cu^{II}).

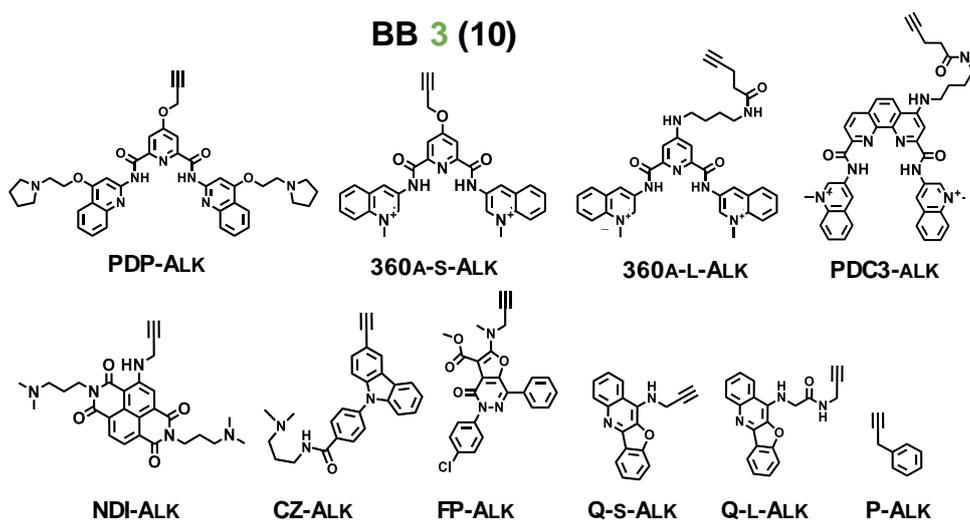


Figure 59 : Représentation exhaustive des fragments hétéroatomatiques alcynes utilisés dans le cadre de la librairie DEL1.

Les 10 fragments utilisés au Cycle 3 de la DEL1 sont pour la plupart des dérivés de molécules connus de la littérature pour leurs reconnaissances des motifs G-quadruplex. De nature hydrophobe, ces ligands fonctionnalisés alcyne atteignent très généralement les tétrades externes des structures G-quadruplex. Les ligands utilisés peuvent être classés en deux catégories (Figure 59) en commençant par celle des systèmes aromatiques fusionnés. Dans cette classe on peut retrouver 6 composés distincts, dont le naphthalène diimide (NDI), les quinolines (Q-s et Q-l), le carbazole (CZ), le furopyridazinone (FP) et la phénanthroline (PDC3). Dans un deuxième temps on peut retrouver la catégorie des systèmes aromatiques non fusionnés faisant intervenir 3 composés, dont la pyridostatine (PDP) et les pyridocarboxamides (360A-s et 360A-l). Une seule molécule n'a pas de précédent dans

l'interaction G-quadruplex, c'est le composé phénylpropyne (P). Cette molécule a été ajoutée à la librairie en guise de contrôle de la réaction de Click-Cu^{II}, également en tant que contrôle négatif à l'interaction avec les G-quadruplex ou encore l'identification d'éventuels ligands exclusivement peptidiques. Certains de ces fragments ont été synthétisés par C. Van Wesemael au cours de son contrat et la majorité a été obtenue par collaboration avec les groupes ayant découvert les composés parents. Leurs synthèses ne sont donc pas détaillées dans ce manuscrit mais plus de détails sont fournis en partie expérimentale.

L'étape de validation des BB3 a été réalisée par réaction de chimie click catalysée par le cuivre sur un composé modèle, l'*Headpiece-Ext-GAWRG-Ke* (Figure 60). Les conditions de cette réaction dérivent de conditions connues et établies au laboratoire afin d'être compatibles avec les ADN (et/ou les ARN). Ainsi, le groupement azoture de la lysine étendue (Ke) a été mis en réaction avec les groupements alcynes de chaque BB3 en présence de sulfate de cuivre (Cu^{II}SO₄, catalyseur de la réaction), d'ascorbate de sodium (agent réducteur du Cu^{II} en Cu^I), de THPTA (stabilisateur du Cu^I) et d'aminoguanidine dans un milieu hétérogène DMF/tampon HEPES. L'aminoguanidine est utilisée pour piéger la forme réactive oxydée de l'ascorbate de sodium afin d'éviter que cette dernière ne réagisse sur les arginines des peptides (Schéma 15, p198). La conversion complète de chacune des 10 réactions de chimie click a pu être confirmée par UPLC-MS dans les conditions employées.

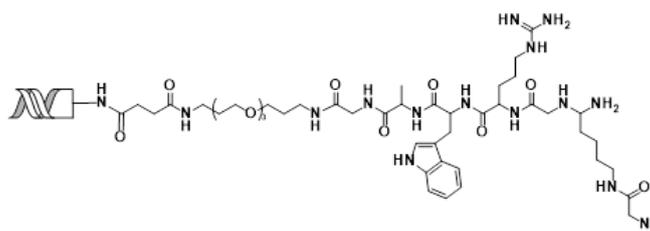


Figure 60 : Synthèse de l'*HP-Ext-GAWRG-Ke* utilisé pour la validation des 10 ligands-alcynes de structures G-quadruplex [Cycle 3].

Les blocs de construction (BB) utilisés lors du quatrième cycle de librairie encodée [Cycle 4] sont les mêmes pentapeptides que ceux utilisés lors du Cycle 1. Cependant, un travail d'optimisation des conditions d'acylations s'est poursuivi en parallèle de la mise au point et de l'assemblage de la librairie DEL1 conduisant à l'identification du DMT-MM comme agent de couplage efficace en milieu hétérogène DMA/borate de sodium. Les optimisations ont montré une incorporation plus efficace pour une majorité de séquences pentapeptidiques. Ainsi, ces conditions ont principalement été utilisées au cycle 4 (voir Tableau S4, p293). Il est à noter que ces nouvelles conditions n'ont toujours pas permis l'incorporation de séquences portant un acide aminé histidine (H).

Sous-partie 1.2 : Assemblage de la DEL1.

Section 1.1.1 : DEL1-Cycle 1.

Dans le cadre du premier cycle de la librairie combinatoire, 50 code-barres ADN T1 ont été individuellement liés à 10 nmol d'*Headpiece-Ext-NH₂*. En présence de l'ADN T4 ligase, de

son buffer, de 2 équivalents de code-barre T1 et de l'*Headpiece* de départ il a été possible de lier enzymatiquement les séquences complémentaires des fragments d'ADN après une nuit à 16°C. Tous les échantillons liés ont ensuite été précipité dans l'éthanol, et la validation des ligations a été effectuée sur gel d'Agarose.



Figure 61 : Augmentation croissante des masses moléculaires de la librairie après ligation enzymatique des code-barres ADN T1 sur l'HP. Exemple de ligation partiellement complète au centre du gel. Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.

Parmi les 50 ligations effectuées, certaines d'entre elles ont dû être relancées, à l'exemple du gel en figure 61. Après validation de toutes les ligations par électrophorèse sur gel d'Agarose (4% dans un tampon TAE), chaque ADN étiqueté a pu être engagé dans la réaction d'acylation avec la séquence pentapeptidique correspondante (Schéma 2).

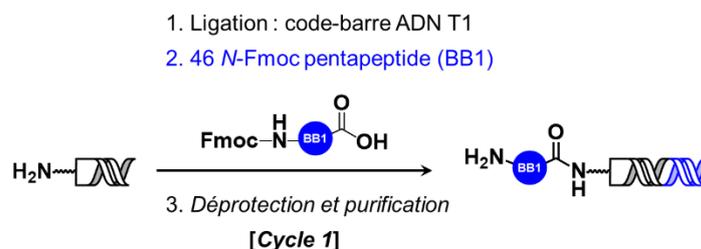


Schéma 2 : DEL1 — [Cycle 1] — Incorporation des motifs pentapeptidiques et de leurs code-barres sur l'*Headpiece*-NH₂. DEL1 — [Cycle 1] = 46 composés.

La réaction d'acylation a été effectuée entre l'amine primaire libre de HPT1-Ext-NH₂ (250 μM ; correspondance code-barres-BB1 : tableau S9, p309) et l'acide carboxylique portée par la séquence pentapeptidique (100mM) en présence d'un mélange EDC/HOAt/NMM (50-10-50 mM) ou DIPEA (30-6-30 mM) dans des conditions hétérogènes DMSO/tampon MOPS. Une étape de précipitation de l'ADN dans l'éthanol a ensuite été effectuée afin de retirer tout excédent chimique au milieu puis chacune des réactions a été analysée en UPLC-MS. A cette étape, le rendement individuel d'acylation a pu être à nouveau déterminé variant de 30 à >95%. En effet, 4 pentapeptides préalablement validés ont été exclus de la librairie pour rendement insuffisant.

Les rendements n'étant pour la plupart pas quantitatifs, nous avons fait face à la présence résiduelle d'HPT1 non fonctionnalisé dans le milieu de synthèse. Afin de s'en affranchir, une étape de purification a été nécessaire.

Afin de purifier un système complexe comme celui-ci, une option est de faire réagir les amines primaires libres de l'HPT1 avec une molécule de biotine activée par un groupement NHS (biotine-NHS). Cette technique dans le cadre de purification de librairie encodée par l'ADN a été décrite par le groupe de D. Neri²⁸⁷.

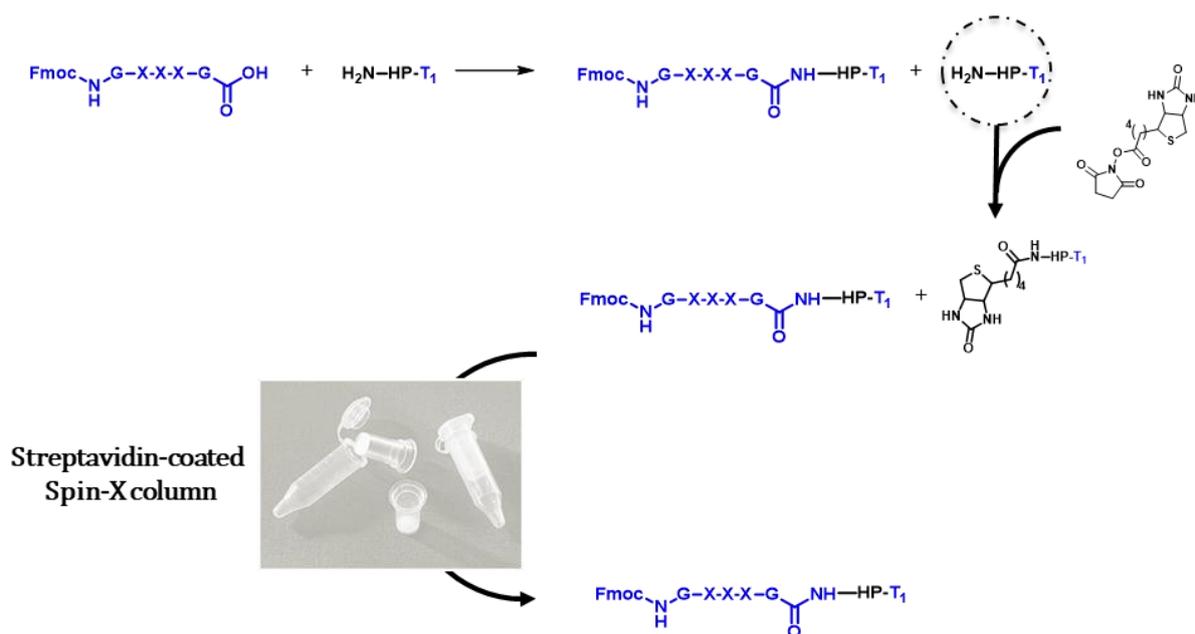


Schéma 3 : Purification DEL1 — [Cycle 1] — Méthode Cap&Catch²⁸⁷. « Cap » des H₂N-HP-T1 non fonctionnalisés par de la biotin-NHS puis « catch » des composés biotinylés sur colonne Streptavidine.

Après avoir fusionné les produits des 46 réactions précipitées, la fonctionnalisation des amines primaires libres avec la biotine-NHS a été réalisée dans un milieu hétérogène DMSO/tampon TEA.HCl sous agitation pendant 4 heures (Schéma 3). L'efficacité de la réaction est analysée par UPLC-MS après une étape de précipitation de l'ADN dans l'éthanol. Une fois les amines primaires libres converties par une fonction biotine, la librairie a été déposée sur une colonne *Spin-X* (tube de microcentrifugation de 1,5mL pourvu d'une membrane à filtration en polypropylène ; CorningTM) contenant une matrice de sépharose fonctionnalisée par de la streptavidine (*Streptavidin SepharoseTM*; GE Healthcare life Sciences). Après incubation de plusieurs heures sous agitation, la librairie a été filtrée puis précipitée dans l'éthanol avant analyse UPLC-MS. Le chromatogramme Figure 62 montre bien la disparition du H₂N-Headpiece-T1 résiduel γ après purification (chromatogramme bleu).

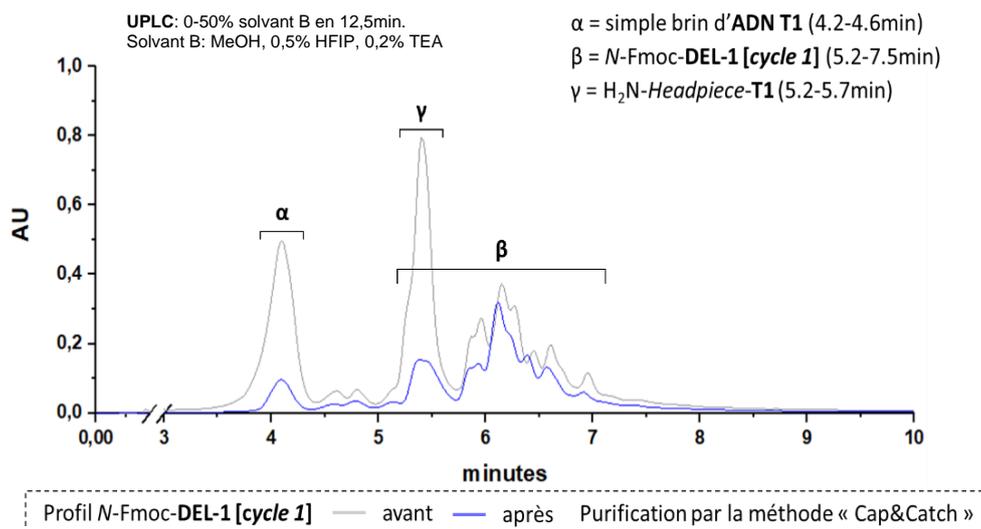


Figure 62 : Chromatogramme UPLC du mélange DEL1 — [Cycle 1] avant (gris) et après (bleu) purification suivant la méthode Cap&Catch. Disparition significative de la fraction γ (H_2N -HP-T1 non fonctionnalisés) après purification.

La quantification de la DEL1 purifiée a été effectuée par dosage spectrométrique. En effet, la radiation UV absorbée par les bases de l'ADN permet de mesurer la concentration de l'ADN double brin à 260nm. Cette concentration molaire peut être calculée à partir de la loi de Beer-Lambert : $A_{260nm} = \epsilon \cdot c \cdot l$, où A représente l'absorbance mesurée, ϵ le coefficient d'extinction molaire du duplex d'ADN considéré en $L \cdot mol^{-1} \cdot cm^{-1}$, c la concentration de l'ADN en solution en $mol \cdot L^{-1}$, et l la longueur du trajet parcouru par le rayonnement dans la solution en cm.

Finalement, 180 nmoles de librairie au Cycle 1 ont pu être obtenues après purification selon la méthode « Cap&Catch », soit un rendement de 39%. L'étape finalisant ce premier cycle est la déprotection des amines primaires portant le groupement protecteur Fmoc présent sur chaque pentapeptide encodé par l'ADN. Cette étape est réalisée rapidement par ajout d'une solution de pipéridine à 10% dans l'eau. Le suivi UPLC-MS de cette étape de déprotection a montré une réaction totale (Figure 63) et quantitative par mesure de l'absorbance. En définitive, 180 nmoles de librairie finale comprenant 46 molécules distinctes ont pu être engagées dans le [Cycle 2] de la DEL1. Il est à noter qu'à ce stade de l'assemblage de la librairie, les analyses de masses permettent facilement de rendre compte du succès des réactions mises en jeu et de la pureté de la librairie en croissance. Ces analyses, associées aux profils UPLC présentés ici, sont présentés en partie expérimentale du manuscrit.

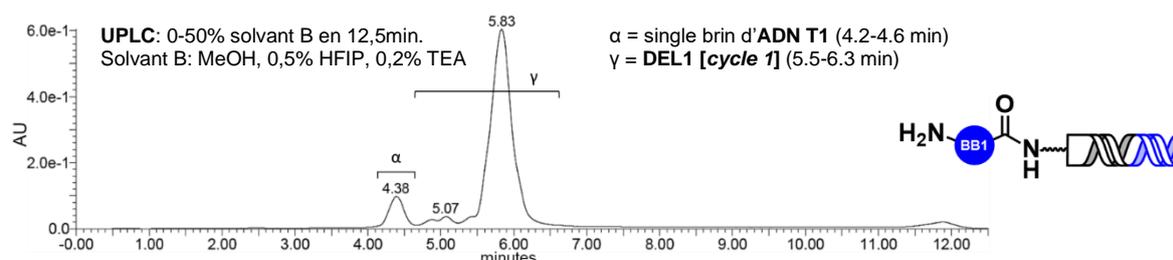


Figure 63 : Chromatogramme UPLC de DEL1 — [Cycle 1] (46 composés en mélange) après purification et déprotection des amines primaires de leur groupement protecteur Fmoc.

Section 1.1.2 : DEL1-Cycle 2.

Le deuxième cycle de librairie combinatoire vise à installer 9 cœurs trifonctionnels au centre de la librairie DEL1. Les 180 nmoles obtenues à l'issue du premier cycle ont été partagées en neuf lots de 20 nmoles afin d'effectuer les ligations des code-barres ADN T2. L'efficacité de ligation des code-barres ADN T2 a été évaluée par électrophorèse sur gel d'agarose. A ce stade de la librairie, le duplex d'ADN à une longueur de 35 paires de bases (17 (*Headpiece*) + 9 (ADN T1) + 9 (ADN T2)). L'étape de couplage a ensuite été conduite individuellement dans des conditions EDC/HOAt/NMM (50-10-50 mM) telles que décrites précédemment (Schéma 4).

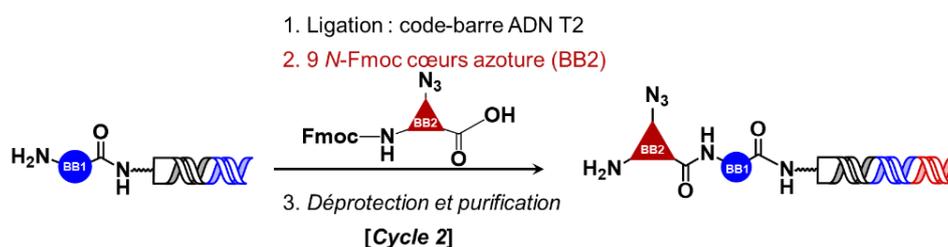


Schéma 4 : DEL1 — [Cycle 2] — Incorporation des cœurs trifonctionnels et de leurs code-barres sur les 46 composés encodés du [Cycle 1]. DEL1 — [Cycle 2] = 414 composés.

Après précipitation dans l'éthanol, les réactions ont ensuite été analysées par UPLC-MS. Comme pour la validation des BB2, les rendements obtenus ont été quantitatifs. En revanche, il a été possible de distinguer trois populations au sein du milieu réactionnel (Figure 64).

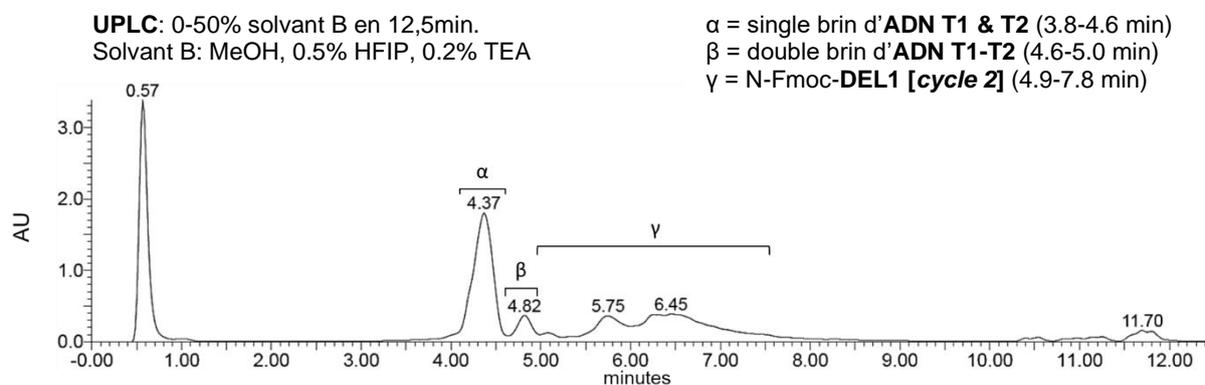


Figure 64 : Chromatogramme UPLC du mélange DEL1 — [Cycle 2]. Présence d'ADN code-barres T1, T2 (α) et T1-T2 (β).

En effet, au-delà de la formation des composés encodés souhaités, nous observons les séquences d'ADN code-barres résiduels (Figure 64, α et β). Ces deux séquences d'ADN sont, respectivement, les ADN code-barres T2 mise en excès lors du processus de ligation enzymatique du [Cycle 2] et de l'ADN code-barres T1 lié à des code-barres T2. Comme le montre la figure 62 représentant l'analyse du milieu réactionnel avant et après purification, il est possible de constater la présence de code-barre T1 résiduel (Figure 64, α). L'ajout en excès des code-barres T2 a donc ainsi pu encoder les code-barres ADN T1 encore présent

dans le mélange. Afin d'éliminer ces séquences ADN parasitant la bibliothèque de composés encodés, le choix a été fait de purifier la librairie par HPLC phase inverse en suivant un gradient linéaire (0 à 80% de solvant B en 32min, débit de 4 mL.min⁻¹) avec solvant A (5% MeCN et 2,5% TEAA dans H₂O, v/v) et solvant B (5% H₂O dans MeCN, v/v). La séparation des code-barres T1 et T2 en excès suivi d'une étape de déprotection des cœurs trifonctionnels des groupements protecteurs Fmoc en milieu basique a permis l'obtention de 65 nmol de DEL1 (Figure 65), soit un rendement cumulé de 36%. A cette étape de la librairie combinatoire, le nombre de molécules présentes atteints 414 composés (46 BB1 x 9 BB2).

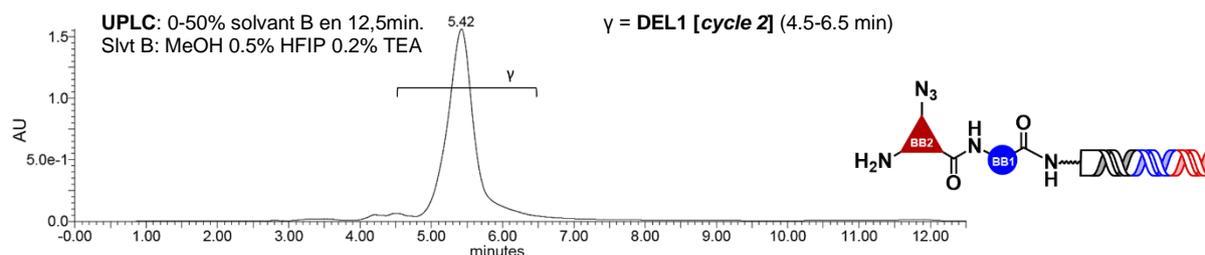


Figure 65 : Chromatogramme UPLC de DEL1 — [Cycle 2] (414 composés en mélange) après purification HPLC et déprotection du groupement protecteur Fmoc.

Section 1.1.3 : DEL1-Cycle 3.

Le troisième cycle de librairie combinatoire permet d'assembler 10 motifs hétéroaromatique pour la plupart intrinsiquement affins des motifs G-quadruplex par cycloaddition 1,3 — dipolaire catalysée par le cuivre. Les 65 nmoles obtenues au [Cycle 2] permettent la ligation enzymatique de chaque code-barre ADN T3 sur des échantillons de 6,5 nmol. La précipitation dans l'éthanol suivi de l'électrophorèse sur gel d'agarose a permis de confirmer l'efficacité de la ligation enzymatique. L'ajout des 9 paires de bases du code-barre T3 fait passer l'ADN duplex à une longueur de 44 paires de bases (Schéma 5).

L'incorporation par chimie click-Cu^{II} des ligands de structures G-quadruplex a été menée conjointement avec le Dr. SI lors de ma première année de doctorat. Pour ce faire, la librairie DEL1 — [Cycle 2] a été préincubée dans une solution hétérogène DMF/tampon HEPES et mise en présence des ligands fonctionnalisés alcynes (13,3mM), de CuSO₄ pentahydraté (6.67mM), d'ascorbate de sodium (20mM) et d'aminoguanidine (66.67mM). Après deux ajouts successifs et une réaction globale de 4 heures, une solution d'EDTA (0,5 M) a été ajoutée au milieu afin de chélater le cuivre résiduel et ainsi de stopper la réaction.

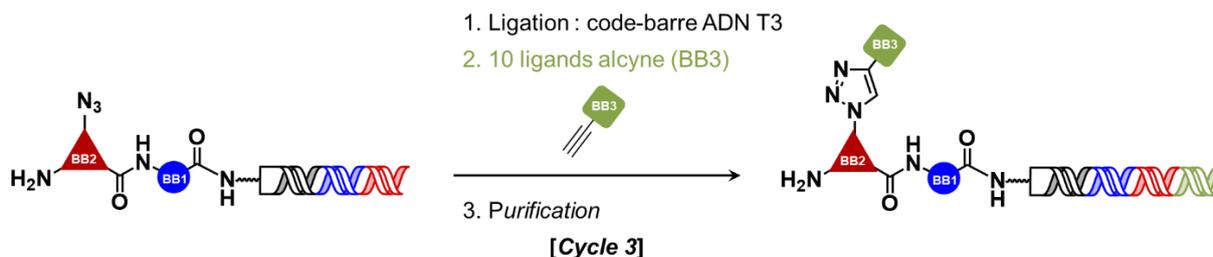


Schéma 5 : DEL1 — [Cycle 3] — Incorporation des ligands G-quadruplex alcynes et de leurs code-barres sur les 414 composés encodés du [Cycle 2]. DEL1 — [Cycle 3] = 4140 composés.

L'analyse UPLC-MS de chaque réaction après précipitation dans l'éthanol n'a pas permis de conclure de manière convaincante sur l'état d'avancement des réactions, la librairie étant à ce stade constituée de 4 140 composés. En effet, la déconvolution UPLC-MS d'un tel mélange est particulièrement difficile. Afin de parer à d'éventuels défauts de conversion, une deuxième réaction de click-Cu^{II} a été réalisée dans les mêmes conditions.

Afin de confirmer l'efficacité des couplages, nous avons souhaité vérifier l'absence de groupement azoture libre dans le milieu réactionnel *via* une technique de biologie moléculaire connue sous le nom de retard sur gel (en. *electrophoretic mobility shift assay*, EMSA). En effet, cette technique permet de détecter par électrophorèse une interaction entre de l'ADN et une protéine. Une bande contenant de l'ADN fixé sur une protéine sera retardée par comparaison à la même bande sans la présence de cette même protéine.

Dans le cas de la DEL1 — [Cycle 3], 2 nmoles d'un mélange de chaque brut réactionnel précipité ont été engagées dans une réaction de chimie click-Cu^{II} (même condition que décrites précédemment) en présence d'une alcyne-PEG₄-biotine. L'intérêt de cette réaction est de faire réagir les éventuels groupements azotures résiduels libres des cœurs trifonctionnels avec un composé biotinylé capable d'être retenu par une protéine (streptavidine) sur un gel d'agarose. En parallèle de cette réaction, l'alcyne-PEG₄-biotine a également été couplée sur une version azoture de l'*Headpiece* (acide acétique azoture couplé à l'*HP-Ext-NH₂*) afin de s'assurer de la bonne incorporation du motif biotinylé. L'analyse UPLC-MS a pu confirmer la conversion totale de l'*HP*-azoture en *HP*-biotine (Tableau 3). La DEL1 — [Cycle 3] biotinylée ainsi que l'*HP*-biotine ont alors été incubés pendant 15 min dans un milieu aqueux contenant la protéine Streptavidine.

Tableau 3 : Tableau associé à l'expérience de gel retard (EMSA) réalisé sur la DEL1 — [Cycle 3] et sur l'*Headpiece*.

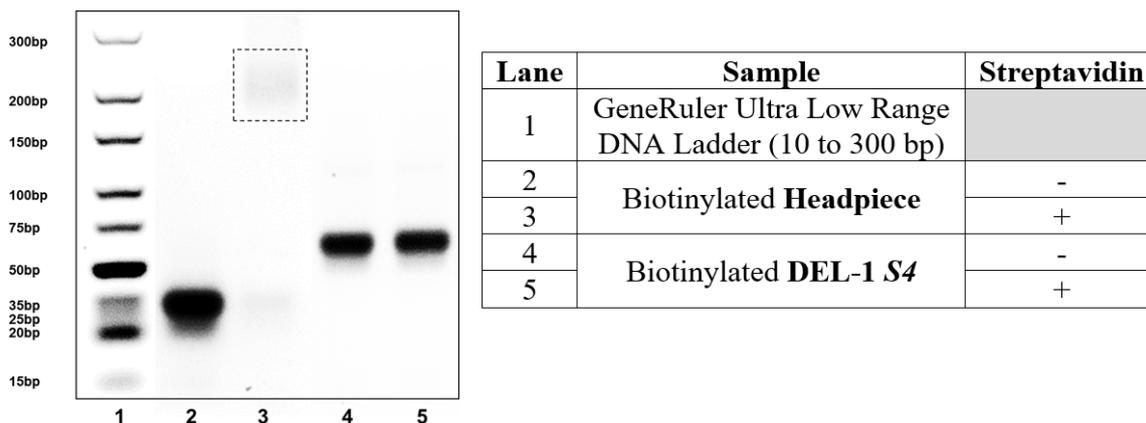


Figure 66 : Gel retard (EMSA) effectué après incorporation d'un alcyne-PEG₄-biotine sur i) l'Headpiece-NH₂ et ii) la DEL1 — [Cycle 3]. L'incubation en présence de streptavidine montre un retard de gel sur le contrôle Headpiece-biotine (ligne 3). Aucun retard observé sur l'échantillon DEL-1 — [Cycle 3] en présence de streptavidine (ligne 5). Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.

La méthode EMSA a été effectuée sur un gel d'Agarose à 4% (Figure 66) et seul le motif HP-biotine a montré une bande supérieure, témoignant de l'effet retard par interaction de la biotine avec la Streptavidine. Cet effet n'a cependant pas été observé sur l'échantillon biotinylé de la DEL1 — [Cycle 3], suggérant ainsi l'absence de fonctions azotures résiduelles au sein de la librairie et donc une conversion totale obtenue lors de la chimie click de la DEL1 en présence des motifs hétéroaromatiques utilisés.

De la même façon que pour les deux cycles précédents, l'utilisation d'un excès de code-barre T3 lors de la ligation enzymatique entraîne la présence résiduelle de ces séquences au sein du milieu réactionnel. A ce stade d'avancement, la différence de masse moléculaire entre la librairie et l'ADN code-barre est marquante puisqu'elle est de 29 488 g/mol et 5596 g/mol respectivement. Il est ainsi possible d'utiliser des membranes de filtration (Amicon® ; Merck-Millipore) à diamètre de pores variables dans le but de purifier et de concentrer la librairie en mélange. Les membranes de filtration à diamètre de pores de 10 kDa ont d'abord été utilisées puisqu'elles sont théoriquement en mesure de laisser passer les structures (globulaires) inférieures à 10 kDa, donc potentiellement les ADN code-barres. Toutefois, l'ADN étant structuré différemment des protéines servant de référence à ce système de filtration, des pores de 10 kDa ne permettent que très peu la séparation des code-barres de la librairie. En revanche l'utilisation de l'Amicon 30 kDa a montré son efficacité pour la séparation des code-barres ADN (Figure 67). Ainsi, après purification sur Amicon 30 kDa, 32 nmoles de la DEL1 — [Cycle 3] ont été obtenues, représentant un mélange de 4 140 composés distincts (46 BB1 x 9 BB2 x 10 BB3).

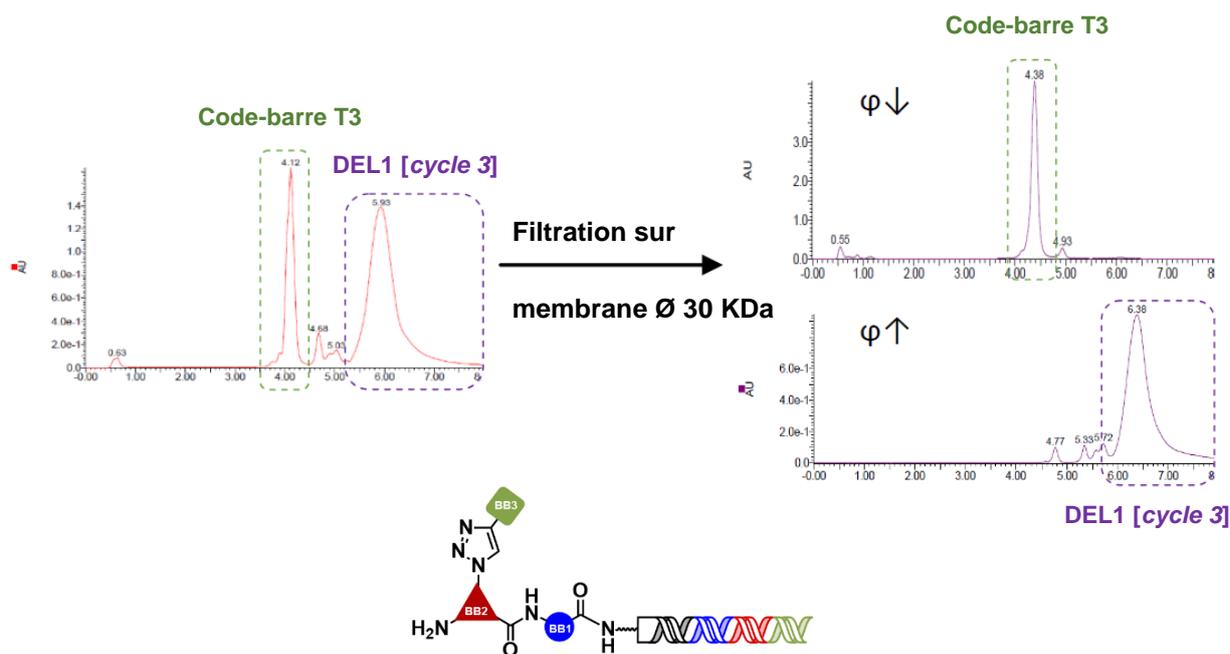


Figure 67 : Filtration sur membranes Ø 30 KDa. Séparation de la DEL1 — [Cycle 3] ($MW_{moy}=29.487,99$ g/mol) des code-barres ADN T3 résiduels.

Section 1.1.4 : DEL1-Cycle 4.

Au même titre que le [Cycle 1], le quatrième et dernier cycle de cette DEL1 a permis l'incorporation de séquences pentapeptidiques susceptibles d'interagir avec les G-quadruplex.

Ainsi, le mélange de 32 nmoles de librairies au Cycle 3 a été fractionné en mélange équimolaire de 0,695 nmoles afin de lier de manière enzymatique les code-barres ADN T4, équivalent aux code-barres T1. Comme explicité lors de la validation des buildings blocks, le couplage via l'activation par le DMT-MM a été principalement utilisé lors de cette étape.

46 pentapeptides ont été insérés dans ces nouvelles conditions, dont 44 peptides communs avec le cycle 1. Travaillant sur petite quantité de librairies, il n'a pas été possible d'analyser individuellement les conversions des couplages peptidiques réalisés en parallèle. En revanche, après avoir rassemblé environ 43 pmoles de chacun des bruts de couplage ($43\text{pmol} \times 46 = 1,98$ nmol), une quantité suffisamment élevée a été obtenue afin d'analyser le mélange réactionnel par UPLC-MS. Malheureusement, le recouvrement des multicharges ainsi que la déconvolution des 190 440 molécules sur un spectre de masse ne permettent pas de conclure de manière exacte sur les rendements de couplage obtenus. Il est à noter que l'analyse de la réaction de couplage peptidique contrôle sur l'*Headpiece-Ext-NH₂* de départ effectué en parallèle des réactions d'acylation du [Cycle 4] ont tous révélés les rendements attendus.

De ce fait, et de manière similaire au [Cycle 1] de la DEL1, il a été postulé que l'amine primaire résiduelle des cœurs trifonctionnels n'ayant pas réagi pourrait être fonctionnalisée par une biotine afin de les éliminer et/ou de quantifier leurs présences. Ainsi, la librairie a été purifiée par la méthode « Cap&Catch » décrite précédemment, en utilisant la biotine-NHS comme

agent de coiffage des amines libres. Après 4 h de réaction à 37°C, la librairie a été précipitée puis purifiée sur matrice de sépharose fonctionnalisée par de la streptavidine. Afin de confirmer l'absence de composés portant une biotine résiduelle, la librairie a été placée sur gel d'électrophorèse, avec le contrôle *HP-biotine*, en présence de streptavidine (Tableau 4, Figure 68).

Une nouvelle fois, le gel a montré que le contrôle positif avec la streptavidine génère un retard du complexe Streptavidine-ADN en haut du gel alors que la librairie, elle, n'est pas retardée. Ceci témoigne de l'absence d'amine biotinylée résiduelle et donc la purification efficace de la DEL1 — [Cycle 4].

Tableau 4 : Tableau associé à l'expérience de gel retard (EMSA) réalisé sur la DEL1 — [Cycle 4] et sur l'Headpiece.

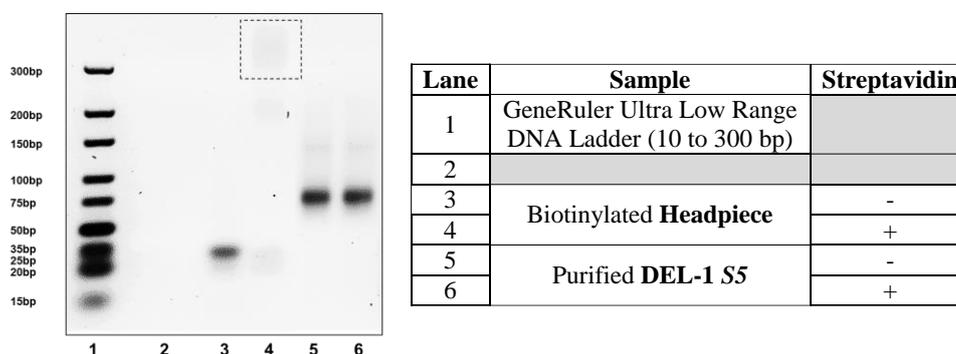


Figure 68 : Gel retard (EMSA) effectué après incorporation d'une biotine-NHS sur i) l'Headpiece-NH₂ et ii) la DEL1 — [Cycle 4]. L'incubation en présence de streptavidine montre un retard de gel sur le contrôle Headpiece-biotine (ligne 3). Aucun retard observé sur l'échantillon DEL-1 — [Cycle 4] en présence de streptavidine (ligne 5).

Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.

L'excédent de code-barres ADN T4 a une nouvelle fois été retiré de la librairie à l'aide de membrane de filtration (Amicon® ; Merck-Millipore) possédant un diamètre de pores de 30 kDa ($MW_{\text{moy}} = 36\,147,51 \text{ g/mol}$). La purification de ce dernier cycle de DEL1 a permis l'obtention de 7,75 nmoles d'un mélange de 190 440 composés (46 BB1 x 9 BB2 x 10 BB3 x 46 BB4).

L'Headpiece sert à la fois de châssis moléculaire mais également de brin matriciel pour l'amplification des étiquettes ADN par PCR. Afin de pouvoir amplifier un signal PCR, il est nécessaire d'avoir, au sein du milieu, deux amorces distinctes capables de reconnaître deux séquences propres, et ceux, dans le but de dédoubler l'ADN à chaque cycle d'amplification (« 2n » copies par cycles, ou n représente le nombre de cycles d'amplification). De ce fait une deuxième séquence d'ADN matricielle, commune à chaque molécule de la librairie, doit être liée de manière enzymatique afin de pouvoir amplifier convenablement la DEL1. Cette dernière étiquette est appelée le *closing primer* (CP). La ligation enzymatique de cette séquence terminatrice sur les 7.75 nmoles obtenues lors du [Cycle 4] de la DEL1 a été réalisée selon la méthodologie décrite précédemment (2 éq. CP, T4 DNA ligase, tampon de ligation). L'excès de *closing primer*, de masse molaire de 16 077 g/mol doit être retiré du milieu réactionnel après précipitation des produits de ligation dans de l'éthanol. La filtration sur membrane à pores de

30 KDa a donc été utilisée. Or, après 3 cycles de filtration, seule la moitié du *closing primer* a pu être séparé de la librairie. Ainsi, l'utilisation d'une membrane à diamètre de pores de 50 kDa a été nécessaire. En effet, seuls 2 cycles supplémentaires ont suffi pour séparer la totalité de la librairie du *closing primer*.

In fine, la quantification par dosage spectrométrique a permis d'obtenir 5,32 nmoles de librairie finale. Cette librairie comporte donc 190 440 molécules, une masse moléculaire moyenne d'environ 52 225 g/mol et une longueur d'ADN duplex de 80 paires de bases (Figure 69).

Tableau 5 : Tableau associé au gel d'électrophorèse final de la DEL1. Caractérisation de tous les intermédiaires encodés par l'ADN dans le cadre de la première librairie combinatoire.

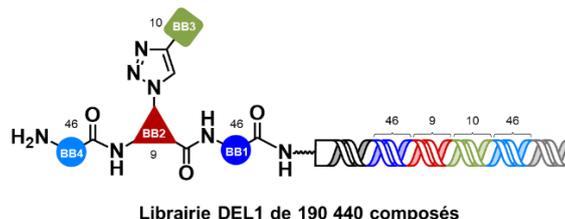
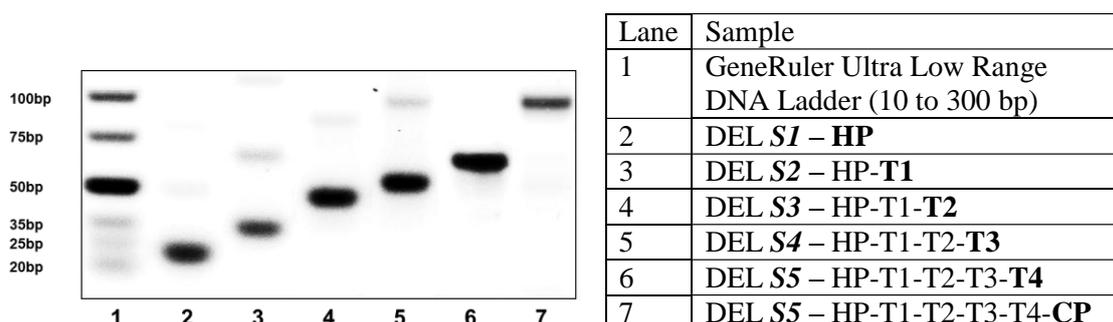


Figure 69 : Gel d'électrophorèse illustrant les 6 tailles d'ADN, de l'Headpiece à la séquence encodée complète (T1 à T4) liée au *closing primer* (CP). Gel 5% (Agarose UltraPure), précoulé au GelRed™ ; migration à 90V/2 h.

Partie 2 : Conception et synthèse de la deuxième génération de ligands encodés par l'ADN.

Sous-partie 2.1 : Composition, conception et validation des buildings blocks de la DEL2.

Au même titre que la DEL1, l'assemblage de la deuxième librairie combinatoire encodée, DEL2, a été effectué suivant la stratégie *DNA recorded*. L'assemblage itératif d'étiquette ADN et de blocs de construction chimique sur plusieurs cycles a permis l'obtention d'une nouvelle librairie de près de 700 000 composés distincts.

La librairie combinatoire DEL2 possède des points communs avec la librairie de première génération. En effet, les molécules chimiques générées lors de la confection de cette librairie encodée ont pour objectif d'interagir le plus vastement possible sur les structures G-quadruplex par l'intermédiaire d'interactions électrostatiques avec les boucles ou les sillons ou par interactions hydrophobes avec les tétrades externes du G-quadruplex. Néanmoins, cette librairie diffère par la nature et le positionnement des éléments hétéroaromatiques et

polycycliques incorporés, ce qui devrait induire un mode de reconnaissance distinct des composés issus de la DEL1.

Une nouvelle fois, fort de notre première expérience en synthèse de DEL, une librairie de composés peptido-organique a été générée à partir de 4 cycles de chimie combinatoire. Le premier et le dernier cycle vont à nouveau permettre l'incorporation de séquences peptidiques visant à générer des interactions électrostatiques avec les boucles ou les sillons des structures quadruplexes (Figure 70, BB1 et BB4). La différence principale entre les deux librairies réside donc dans le positionnement des structures hydrophobes hétéroaromatiques susceptibles d'interagir avec les G-quadruplex. En effet, alors que des fragments aromatiques connus pour interagir avec les G-quadruplex ont été insérés dans la DEL1 au travers des chaînes latérales des cœurs trifonctionnels, ce sont des structures hydrophobes, aliphatiques, aromatiques ou hétéroaromatiques qui sont cette fois-ci au cœur de la DEL2 (Figure 70, BB2). Le positionnement de ces composés au centre de la librairie chimique a donc nécessité la synthèse de cœurs complexes trifonctionnalisés. Leurs synthèses sont un des objectifs de la thèse et un travail collaboratif local. Enfin, l'incorporation de ces cœurs au centre même de la librairie a permis l'incorporation d'un troisième groupe de différenciation. Basé sur des motifs semblables à ceux trouvés sur les chaînes latérales de ligands de G-quadruplex issues de la littérature mais également *via* des motifs moléculaires variés, ce troisième groupe permettra d'établir des interactions supplémentaires avec les tétrades externes, les boucles voire les sillons des différentes structures (Figure 70, BB3).

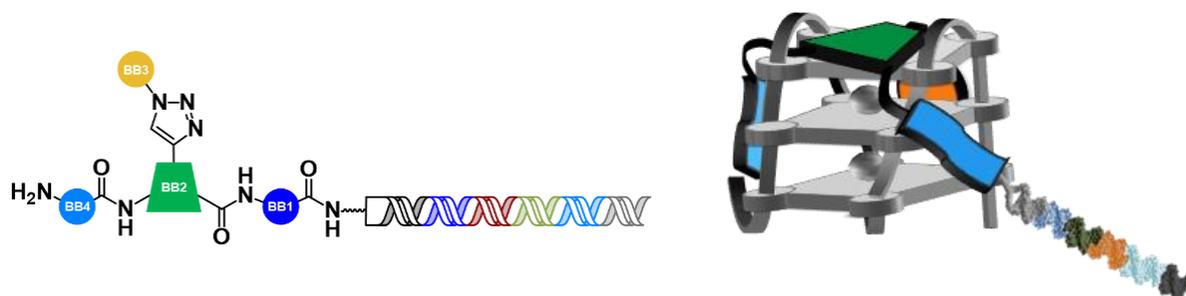


Figure 70 : Structure des ligands de la DEL2 par association de quatre buildings blocks.

L'assemblage itératif de la DEL2 s'effectue selon le même schéma que pour la DEL1.

Les BB utilisés lors du premier cycle de librairie encodée DEL2-[Cycle 1] sont les mêmes blocs de construction que ceux utilisés au sein de la DEL1 (Figure 71). Grâce au protocole de couplage par activation en présence de DMT-MM ce sont 59 pentapeptides qui ont été incorporés au cours du premier cycle.

BB1 & BB4
59 Fmoc-GXXXG-OH

X = R, G, P, A, L, F, W

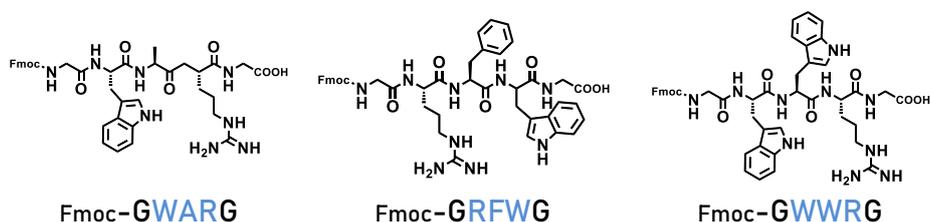


Figure 71 : DEL2 — [Cycle 1] — Structure des 59 motifs pentapeptidiques.

Les BB utilisés lors du deuxième cycle de librairie encodée DEL2-[Cycle 2] sont donc des fragments trifonctionnels associant une fonction amine (NH-Fmoc), une fonction acide carboxylique (—COOH) et enfin un groupement alcyne (—≡). Initialement, 5 fragments ont été envisagés.

Premièrement, on retrouve un composé contrôle de structure aliphatique dérivé du DAP. Ce composé trifonctionnel a été préparé par C. Van Wesemael (technicienne, UGA, Figure 72, A/).

Dans un second temps on retrouve les motifs trioxazole, indolizine et naphthalène diimide connus et choisis pour interagir avec les cibles G-quadruplex avec une affinité de l'ordre du $\mu\text{M}^{288,289}$. Le trioxazole, motif ayant été étudié pour la reconnaissance des G4, a été synthétisé par le Dr. Rémy Lartia au laboratoire (Figure 72, B/). L'indolizine, composé versatile et modulable possédant une structure hétéroaromatique et des propriétés de fluorescence, a été sélectionnée dans le cadre d'une collaboration avec le Département de Pharmacochimie Moléculaire de Grenoble (Dr. Baussanne et Dr. Demeunynck). Il est à noter que deux indolizines respectivement porteuses d'une chaîne latérale courte (Indolizine-s, Figure 72, C/) et d'une chaîne latérale longue (Indolizine-l, Figure 72, D/) ont été synthétisées. Enfin, le dérivé naphthalène diimide (NDI) trifonctionnel préalablement choisi n'a malheureusement pas pu être synthétisé malgré les efforts de nos collaborateurs italiens (Dr. Freccero et Dr. Doria) impliqués sur le projet.

Dans un dernier temps, on retrouve un dérivé pyridostatine (PDS) trifonctionnel. Largement documenté dans la littérature pour ces capacités d'interactions avec les cibles G-quadruplex, ce composé trifonctionnel a été synthétisé avec succès par C. Van Wesemael (Figure 72, E/). En s'appuyant sur les analyses obtenues lors du séquençage et des études d'interactions des ligands issus de la DEL1 (discutées au Chapitre 3 et 4) il a été décidé ne pas inclure ce cœur dans cette deuxième librairie et ce afin de ne pas biaiser les sélections *in vitro* par affinité vers l'enrichissement de ligands portant exclusivement ce cœur particulièrement afin pour les structures G-quadruplex.

Ainsi, seuls les composés trifonctionnels DAP, Trioxazole, Indolizine-s et Indolizine-l ont été retenus en tant que BB de la DEL2-[Cycle 2].

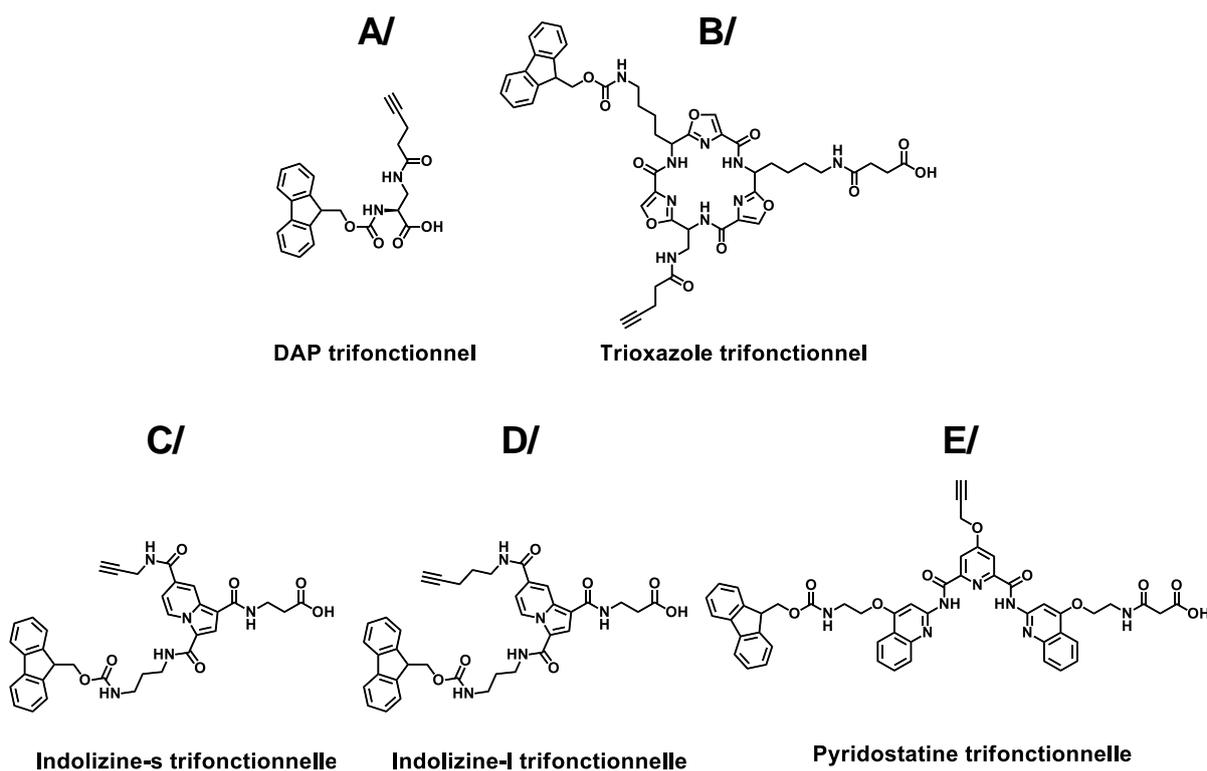


Figure 72 : Série de ligands trifonctionnels (— NH-Fmoc, — COOH, — Alcyne) synthétisés au laboratoire (A/, B/ et E/) et en collaboration (C/ et D/).

Dans le cadre du soutien financier de mon projet de thèse par le Laboratoire d'Excellence (LABEX) Arcane, j'ai donc eu l'occasion de participer à la synthèse d'hétérocycles indolizines en collaboration avec le Département de Pharmacochimie Moléculaire (DPM) de Grenoble. La synthèse d'une indolizine à chaîne alcyne courte (Indolisine-s, Figure 72, C/) ainsi qu'une indolizine à chaîne longue (Indolisine-l, Figure 72, D/) a été effectuée avec et sous la direction d'Isabelle BAUSSANNE (CR, CNRS) et Martine DEMEUNYNCK (DR, CNRS).

L'obtention des deux indolizines repose sur l'assemblage successif de 3 synthons (notés synthon A à C) tel qu'illustré en Figure 73. La synthèse passe par la formation d'un intermédiaire ylure par réaction du N-(3-aminopropyl)-2-bromoacetamide protégé par un groupement Boc (synthon A) avec le N-[alkyn]ylisonicotinamide (synthon B ; n=1 [méthyn] ou n=3 [propyn]). L'ylure formé peut ensuite être cyclisé avec l'acide (propiolamido)propanoïque protégé par le groupement *tert*-butyle (synthon C) afin de former les cycles hétéroaromatiques souhaités.

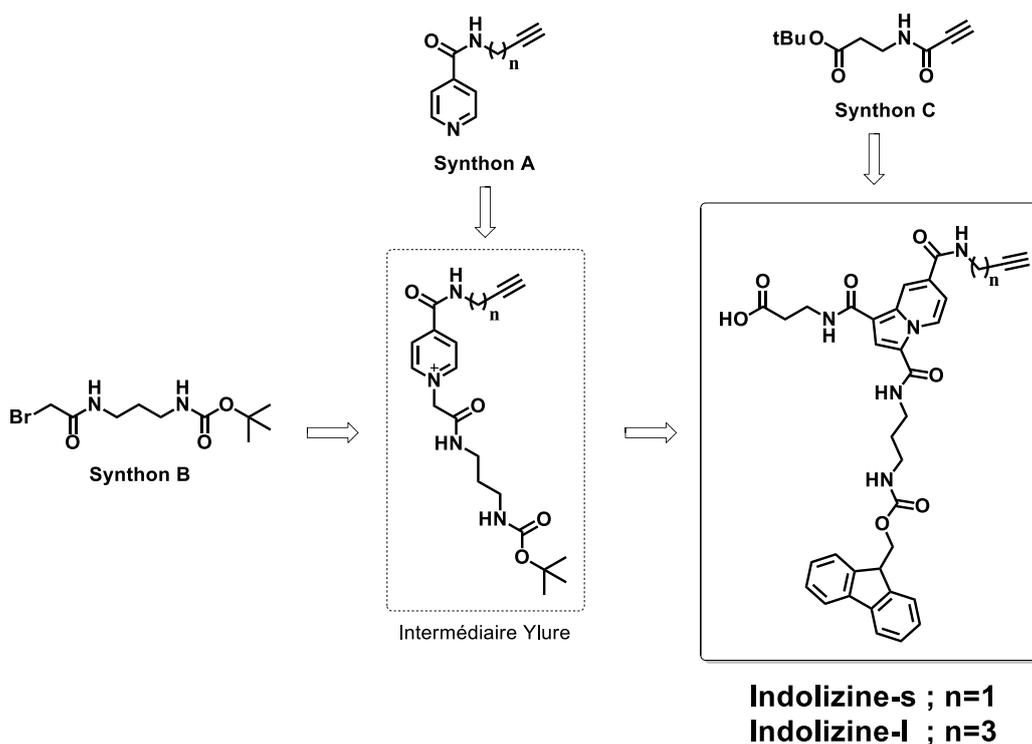


Figure 73 : Synthèses des indolizine-s et -l à partir de trois synthons (A, B et C).

L'évaluation de l'incorporation par couplage peptidique des 5 motifs du [Cycle 2] a été effectuée par réaction des BB2 sur l'*Headpipe-Ext-NH₂*. Les mêmes conditions de couplages utilisées pour le premier cycle de la DEL2 ont été appliquées et la conversion de la réaction a été évaluée par UPLC-MS. Les rendements obtenus ont varié de 5 à 85%, nécessitant ainsi l'optimisation des conditions de couplages. Ces optimisations ont été réalisées en faisant varier les agents de couplage et/ou la température de réalisation de l'expérience. Ainsi l'utilisation des conditions EDC/HOAt/NMM décrites en DEL1, non plus à température ambiante mais à 37°C, a permis d'obtenir des conversions complètes pour l'Indolizine-s et le DAP. Dans ces mêmes conditions, l'indolizine-l a été obtenu avec un rendement satisfaisant de 65%. Le couplage efficace du dérivé trioxazole a été réalisé dans les conditions décrites en 2021 par J.Hunter²⁹⁰ au sein d'une publication dédiée à la formation de lien amide pour la synthèse de librairie encodée. Ainsi l'utilisation à 37°C de l'agent de couplage HATU et de la 2,6-lutidine en tant que base a permis d'observer un rendement de réaction de 85%.

Les building blocks utilisés lors du troisième cycle de librairie encodée [Cycle 3] sont des dérivés azotures de motifs moléculaires variés incluant des cycles (hétéro)aromatiques, des chaînes aliphatiques et protéogéniques sans idées préconçues de leurs possibles interactions avec les cibles G-quadruplex (Figure 74, Tableau exhaustif S8, p306).

Une bibliothèque de 72 composés d'origine commerciale et portant une fonction amine primaire libre a été achetée (Tableau S8, p310). À partir de ces homologues aminés, une réaction de diazo-transfert a été réalisée en présence d'ISAHC (2eq) et de NaHCO₃ (9eq) sous catalyse Cu^{II} afin de convertir les amines libres en fonction azotures.

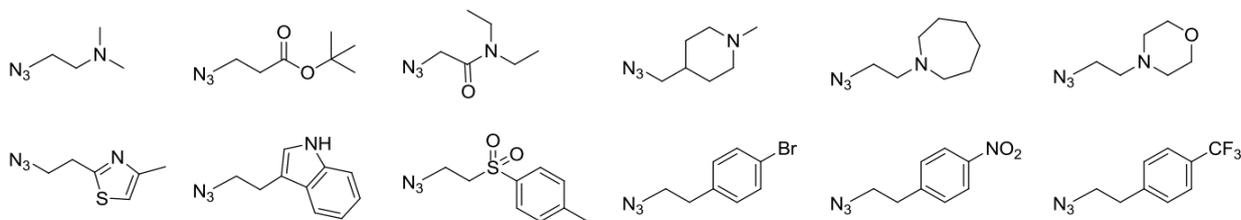
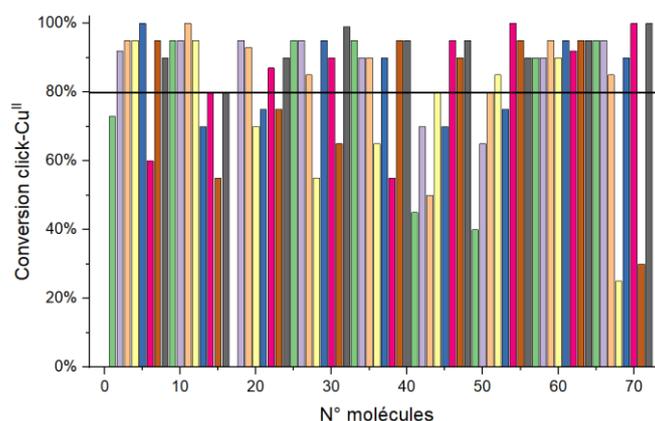


Figure 74 : DEL2 — [Cycle 3] — Structures générales et liste non-exhaustive des azotures (protéogéniques, non-protéogéniques) utilisés dans le cadre du Cycle 3 de la DEL2.

Chacune des réactions de diazotransfert a été analysée par UPLC-MS et les bruts des molécules synthétisées ont été engagés en Click-Cu^{II} sur 2 nmol d'*Headpiece-Ext-NH₂* préalablement couplé avec un acide Pentinoïque. Ces travaux de synthèses organiques et de réactions bio-organiques ont été réalisés par Alicia HOUDANT et Mattéo GARCIA, stagiaires au laboratoire lors de l'été 2019. La caractérisation des réactions de chimie click évalué par UPLC-MS a permis de classer les molécules suivant les avancements de réactions obtenus. Ainsi, parmi les 72 composés, 50 d'entre-eux ont montré des rendements de click supérieurs (ou égaux) à 80% (Graphique 1).



Graphique 1 : Représentation du taux d'incorporation des 72 composés azotures par chimie Click (Cu^{II}) sur un *Headpiece* modifié alcyne. Rendement $\geq 80\%$ observé pour 50 composés distincts.

Les buildings blocks utilisés lors du quatrième cycle de librairie encodée [Cycle 4] sont les mêmes 59 séquences peptidiques que celles utilisées lors du Cycle 1 (Figure 75).

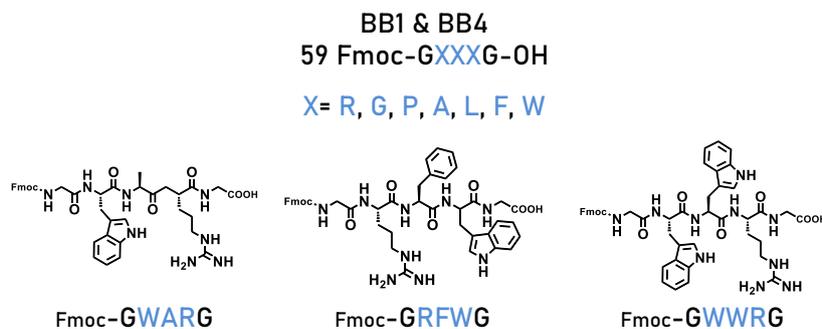


Figure 75 : DEL2 — [Cycle 4] — Structure des 59 motifs pentapeptidiques utilisées lors du Cycle 4 de la DEL2.

Sous-partie 2.2 : Assemblage de la DEL2.

Section 1.2.1 : DEL2-Cycle 1.

Dans le cadre du premier cycle de la deuxième librairie combinatoire, 59 code-barres ADN T1 ont été individuellement engagés avec 10 nmol d'*Headpiece-Ext-NH₂* dans le processus de ligation enzymatique tel que décrit précédemment (ADN T4 ligase et son buffer ainsi que 2 équivalents d'ADN code-barre). Après ligation à 16°C sur la nuit, chacune des réactions a été précipitée dans l'éthanol et analysée pour complétion par électrophorèse sur gel d'agarose.

Les pentapeptides ont été couplés à leur séquence d'ADN étiqueté respectivement en suivant le protocole de couplage optimisé en condition DMT-MM (Schéma 6). En effet, ce couplage s'étant montré suffisamment efficace, et afin de simplifier la réalisation de réactions en parallèle, seul ce protocole a été utilisé au cours de ce premier cycle. Après précipitation dans l'éthanol et analyse UPLC-MS il a été observé des rendements encore supérieurs à ceux obtenus lors des précédentes expériences de validation. Aucune étape de purification n'a donc été nécessaire à ce stade de l'assemblage de la librairie. Les membres de la librairie ont ensuite été assemblés avant d'être soumis à une réaction de déprotection de l'amine primaire de son groupement protecteur Fmoc dans un milieu aqueux contenant 10% de pipéridine.

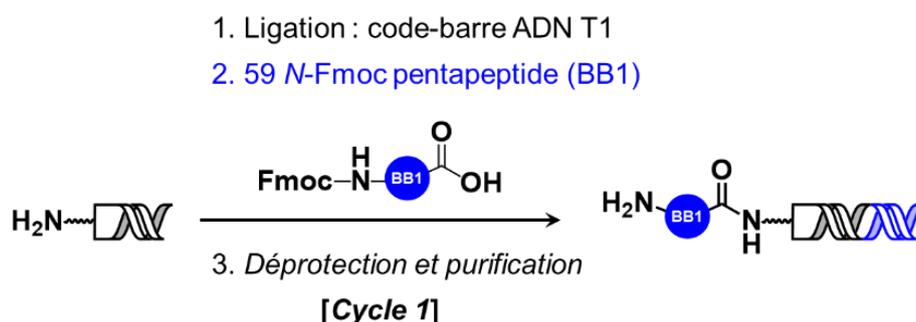


Schéma 6 : DEL2 — [Cycle 1] — Incorporation des motifs pentapeptidiques et de leurs code-barres sur l'*Headpiece-NH₂*. DEL1 — [Cycle 1] = 59 composés.

La quantification spectroscopique du mélange des 59 molécules encodées (Figure 76) a indiqué un rendement final de 69,5%, soit une quantité de DEL2-[Cycle 1] équivalente à 410 nmoles. Le milieu réactionnel a été utilisé en l'état pour le cycle prochain.

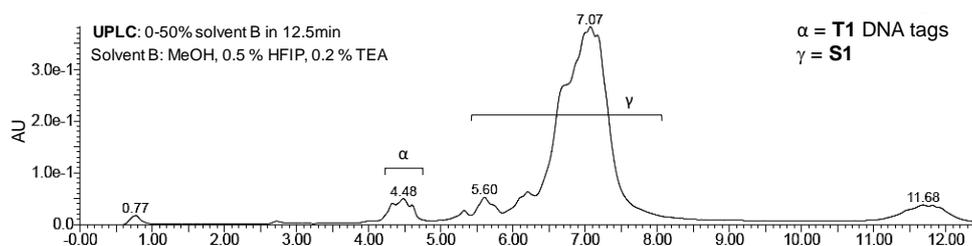


Figure 76 : Chromatogramme UPLC du mélange DEL2 — [Cycle 1] après déprotection du groupement protecteur Fmoc.

Section 1.2.2 : DEL2-Cycle 2.

Le deuxième cycle de la deuxième librairie combinatoire a été effectué d'introduire les 4 cœurs trifonctionnels. À partir des 410 nmoles de DEL1-[Cycle 1], 320 nmoles ont été engagées dans les processus de ligation enzymatique du code-barres ADN T2. Ainsi 80 nmoles ont été utilisées dans chacune des ligations enzymatiques à effectuer. Pour un souci de reproductibilité efficace, les ligations enzymatiques ont été effectuées dans les mêmes conditions que le [Cycle 2] de la DEL1, c'est-à-dire à partir de 20 nmoles de librairie. Chaque portion de 80 nmoles a ainsi été liée à leur code-barre ADN T2 respectif sous forme de 4 réactions de 20 nmoles en présence de l'ADN T4 ligase et de son buffer.

Sans avoir rassemblé les aliquots après ligation enzymatique, les 4 cœurs trifonctionnels ont été couplés à leur cœurs respectif. Les indolizines, le DAP ainsi que le trioxazole ont été couplés, à 37°C, dans un milieu hétérogène DMA/H₂O contenant l'agent de couplage EDC (125eq), l'activateur HOAt (25eq) et la NMM (125eq) pour base. Conjointement avec les optimisations effectuées en amont, le trioxazole a été introduit par couplage peptidique dans un milieu hétérogène DMA/H₂O contenant l'agent de couplage HATU (125eq) et la 2,6 — Lutidine (625eq). Quel que soit les conditions utilisées, les cœurs ont été couplés une première fois pendant 2 heures et une nouvelle solution fraîche de réactifs a été rajoutée aux solutions pour une deuxième activation sur la nuit (Schéma 7).

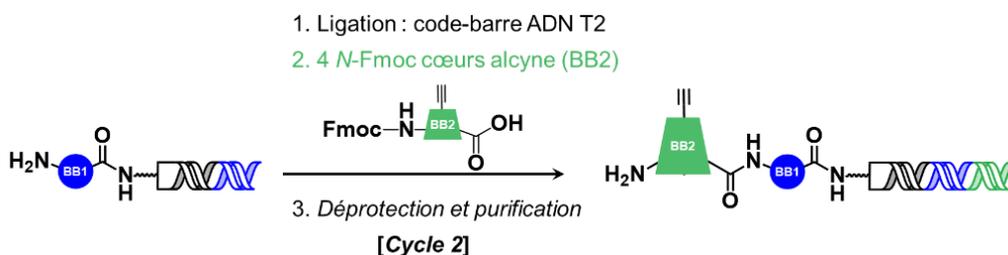


Schéma 7 : DEL2 — [Cycle 2] — Incorporation des cœurs trifonctionnels et de leurs code-barres sur les 59 composés encodés du [Cycle 1].

Le rassemblement des aliquots, la précipitation dans l'éthanol et la purification des échantillons des code-barres ADN T2 en excès par filtration sur membrane (Amicon®, membrane de 30 kDa) ont permis leur analyse par UPLC-MS. L'évaluation de l'incorporation par cette méthode s'est avérée complexe puisque la librairie est très étalée sur le chromatogramme.

Afin d'évaluer les degrés d'incorporation des cœurs, chacun des bruts réactionnels ont été couplés à la biotine-NHS (300eq) afin de quantifier sur gel retard les amines libres résiduelles. Après couplage, les échantillons ont été déposés sur gel d'agarose en présence ou non de Streptavidine afin de percevoir les différences de mobilités sur gel, et donc d'informer sur la présence d'amine libre au sein de l'échantillon de librairie testé (Figure 77).

Tableau 6 : Tableau associé à l'expérience de gel retard (EMSA) réalisée sur les cœurs de la DEL2 — [Cycle 2].

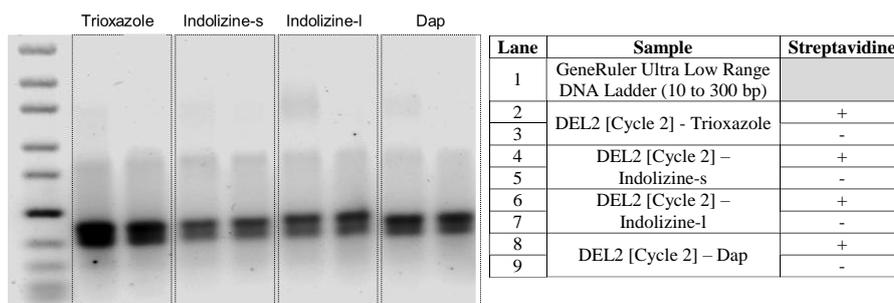


Figure 77 : Gel retard (EMSA) effectué après Cap&Catch²⁸⁷ et réalisé indépendamment sur les cœurs trifonctionnels de la DEL2-[Cycle 2]. Quel que soit le cœur considéré, un retard est observable (lignes 2, 4, 6 et 8). Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.

Le gel retard fait apparaître des bandes en partie supérieures de chaque ligne ayant migré en présence de streptavidine. Cette observation démontre la présence d'amine primaire résiduelle libre au sein de chaque échantillon, permettant ainsi de conclure sur une complétion partielle des couplages des cœurs trifonctionnels, en accord avec les études menées lors de l'évaluation de BB2. Afin de purifier l'ensemble de la librairie, chaque échantillon a été mis en réaction avec la biotine-NHS tel que présenté précédemment. Dans un second temps, les échantillons partiellement fonctionnalisés par de la biotine ont été déposés sur une colonne de séparation contenant une matrice sépharose fonctionnalisée par de la streptavidine. L'incubation de 2 heures avec la matrice a permis d'obtenir environ 60 nmoles de chacun des échantillons purifiés. Toutefois, l'analyse sur gel retard des échantillons purifiés a tout de même montré la présence résiduelle de composés biotinylés dans le mélange réactionnel (Figure 78). Un deuxième tour de purification sur matrice fonctionnalisée streptavidine a permis une purification complète des échantillons de la librairie (Figure 78). En revanche, le dosage spectrométrique a montré une perte importante de quantité de matière par échantillon (rdt=25%-Trioxazole ; 30%-Indolizine-s ; 70%-Indolizine-l et 65%-Dap), indiquant la difficulté de coupler les différents cœurs sur des échantillons de librairie à quantité initiale élevée.

Tableau 7 : Tableau associé à l'expérience de gel retard (EMSA) après deux expériences de catch-streptavidine successives (exemple du cœur trioxazole — DEL2 — [Cycle 2]).

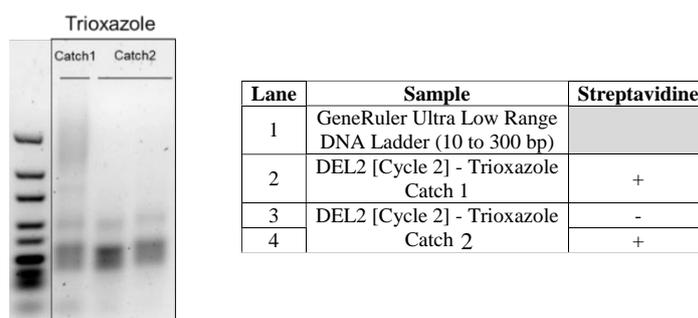


Figure 78 : Gel retard (EMSA) effectué après deuxième expérience de catch (exemple du cœur trifonctionnel trioxazole). Aucun effet retard observé après réalisation de 2 catch successifs (catch 2). Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.

Finalement, les produits réactionnels ainsi purifiés ont été mélangés et les 150 nmoles de DEL2 — [Cycle 2] obtenus ont été soumises à déprotection dans un milieu aqueux contenant 10% de pipéridine afin de déprotéger l'amine primaire portant le groupement protecteur Fmoc. Après précipitation du milieu de déprotection dans l'éthanol et concentration des 236 composés (59 BB1 x 4 BB2) constituant le mélange (Figure 79), le dosage indique que 145 nmoles de mélange peuvent être engagés dans le pénultième cycle de la librairie combinatoire.

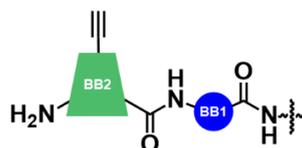


Figure 79 : Structure générale des ligands de la DEL2 en fin de [Cycle 2]. DEL2 — [Cycle 2] = 236 composés.

Section 1.2.3 : DEL2-Cycle 3.

50 motifs moléculaires ont pu être intégrés au sein de la DEL2 par chimie click-Cu^{II}. Les 145 nmoles obtenues au [Cycle 2] permettent la ligation enzymatique de chaque code-barre ADN T3 sur des échantillons de 2,9 nmoles. La visualisation de l'incorporation des code-barres ADN T3 après précipitation dans l'éthanol a permis de confirmer la mauvaise ligation de 2 des 50 code-barres ADN T3. La réalisation d'une seconde ligation et la validation par gel d'agarose de toutes les incorporations ont conduit à des étiquettes ADN d'une longueur de 44 paires de bases (Schéma 8).

L'incorporation par chimie click-Cu^{II} de cycles (hétéro) aromatiques, de chaînes aliphatiques et/ou de motifs protéogéniques a été effectuée dans les conditions classiques de chimie click décrites dans le cadre de la DEL1. En effet, la librairie DEL2 — [Cycle 2] a été préincubée dans une solution hétérogène DMF/tampon HEPES et mise en présence des composés azotures (13,3mM), de CuSO₄ pentahydraté (6.67mM), d'ascorbate de sodium (20mM) et d'aminoguanidine (66.67mM). Après deux ajouts successifs et une réaction globale de 4 heures, une solution d'EDTA (0,5 M) est ajoutée au milieu dans le but de chélater le cuivre résiduel et ainsi stopper la réaction.

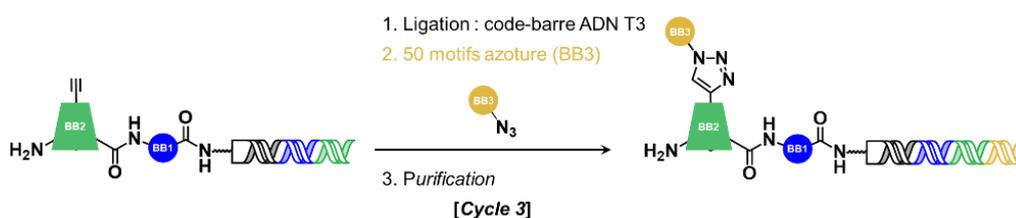


Schéma 8 : DEL2-[Cycle 3] — Incorporation des 50 ligands G-quadruplex azotures et de leurs code-barres ADN.

Tous les milieux réactionnels ont ensuite été rassemblés, précipités dans l'éthanol et filtrés sur membrane d'exclusion de taille (Amicon®, membrane de 30 KDa) afin que la librairie ($MW_{\text{moy}} = 29\,931$ g/mol) soit séparée des ADN code-barres T3 ($MW_{\text{moy}} = 5596$ g/mol) en excès. Le dosage spectrométrique à ce stade d'avancement a permis de quantifier 94 nmoles de librairie.

La DEL2-[Cycle 3] possède 11 800 molécules (59 BB1 x 4 BB2 x 50 BB3) en mélange. L'analyse UPLC-MS de ce mélange ne permet pas de différencier les composés résultant des réactions de cycloaddition des composés de départ. Ainsi, de manière similaire au « Cap&Catch » réalisé lors du [Cycle 2] de la librairie, il est possible d'incorporer un dérivé azoture biotinylé afin de séparer les composés d'intérêts des tronqués de réaction. En effet, les 4 cœurs trifonctionnels possèdent une fonction alcyne qui, lors du [Cycle 3] n'a peut-être pu réagir que partiellement avec les motifs azotures. Le composé biotinylé commercial, le N₃-PEG₃-biotine, a alors été couplé dans les conditions de cycloaddition décrite précédemment afin de fonctionnaliser les dérivés de la librairie chimiquement tronqués. Après biotinylation, la librairie a subi deux processus complets de purification sur matrice fonctionnalisée par de la streptavidine. En effet, il a été possible de visualiser sur gel retard la présence résiduelle de composés biotinylés au sein de la librairie après le premier tour de « Catch » sur résine streptavidine (Figure 80).

Tableau 8 : Tableau associé à l'expérience de gel retard (EMSA) réalisé sur la DEL2 — [Cycle 3].

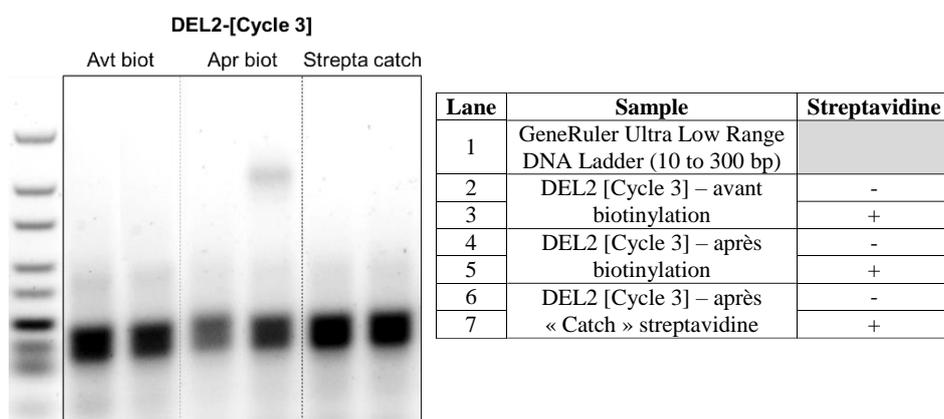


Figure 80 : Gel retard (EMSA) effectué avant biotinylation et après Cap&Catch sur la librairie en mélange DEL2 — [Cycle 3]. L'effet retard observé sur une partie de la DEL2 (ligne 5 ; « apr biot »). Disparition de la bande après Catch en présence de streptavidine. Gel 4% (Agarose), précoulé au GelRed™ ; migration à 90V/2 h.

Après purification du mélange réactionnel et précipitation dans l'éthanol, 47,2 nmoles de librairie combinatoire encodée par l'ADN ont pu être obtenues (Figure 81) et engagées dans le dernier cycle appliqué à la DEL2.

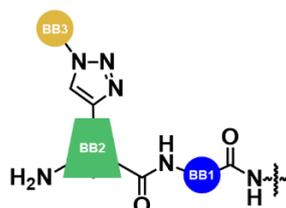


Figure 81 : Structure générale des ligands de la DEL2 en fin de [Cycle 3]. DEL2 — [Cycle 3] = 11 800 composés.

Section 1.2.4 : DEL2-Cycle 4.

De manière analogue à l'incorporation des pentapeptides du [Cycle 1], le [Cycle 4] a été permis l'incorporation de 59 pentapeptides distincts. À partir des 47,2 nmoles de DEL2-[Cycle 3] obtenues, 0,8 nmoles ont été engagées dans des processus de ligation enzymatique avec les code-barres ADN T4 (équivalent à ceux du T1). Les 59 ligations enzymatiques ont été effectuées en présence de 2 eq. de code-barres T4 mais également en présence de l'ADN T4 ligase et de son tampon de travail. Après précipitation dans l'éthanol, chaque pentapeptide (250eq) a été couplé individuellement en présence de son homologue encodée dans des conditions de couplages peptidiques présentant le DMT-MM (250eq) en solution dans un milieu hétérogène DMA/tampon sodium borate. Toutefois, les réactions d'optimisation de couplage ont toutes été réalisées à partir de 2 nmoles d'ADN de départ. Ainsi, pour les 59 réactions de couplages peptidiques, les conditions appliquées ont été celles mimant le couplage sur 2 nmoles. Ainsi, ce ne sont pas réellement 250eq de pentapeptides ou de DMT-MM qui ont été utilisés mais bien 625 eq. de chacun des réactifs.

Après couplage et précipitation, la librairie combinatoire est composée 696 400 membres (59 BB1 x 4 BB2 x 50 BB3 x 59 BB4). Bien évidemment, l'analyse UPLC-MS ne permet pas d'évaluer les rendements de couplages obtenus. En revanche, le dosage spectroscopique de la librairie après filtration sur membrane (Amicon®, membrane de 30 KDa) a permis d'obtenir 16,3 nmoles de DEL2-[Cycle 4]. Une nouvelle fois, afin de s'affranchir des tronqués de réaction, la technique de « Cap&Catch » a été appliquée et réalisée par couplage d'une biotine-NHS suivi d'un passage sur colonne de streptavidine. Cette étape a permis de purifier le mélange réactionnel et d'obtenir 5.05 nmoles d'une librairie finale après déprotection de l'amine primaire de son groupement protecteur Fmoc (solution aqueuse contenant 10% pipéridine, Figure 82).

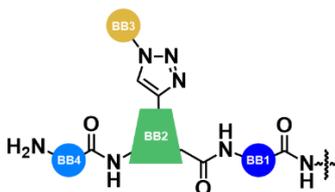
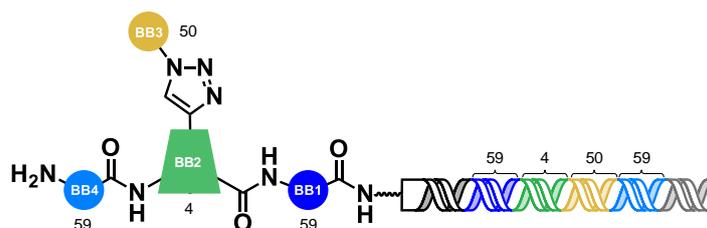


Figure 82 : Structure générale des ligands de la DEL2 en fin de [Cycle 4]. DEL2 — [Cycle 4] = 696 200 composés.

La ligation finale de la séquence terminatrice a été réalisé directement après la fin du [Cycle 4]. Cet ADN code-barre a été lié à la DEL2 par processus classique de ligation enzymatique avec l'ADN T4 ligase en présence de 2 équivalents d'ADN code-barre. L'incorporation du *closing primer* fait ainsi passer l'ADN à une longueur de 80 paires de bases et la librairie à une masse moyenne de 52 067 g/mol. Le *closing primer* présent en large excès a été soustrait de la librairie en mélange en utilisant des filtres sur membranes à diamètre de pores de 50 kDa.

La quantification finale par dosage spectrométrique a ainsi permis d'obtenir 3 nmoles de librairie finale (Figure 83).



Librairie DEL2 de 696 200 composés

Figure 83 : Encodage (Headpiece, code-barres T1 à T4 et Closing primer) et BB utilisé afin de former la DEL2 composée de 696 200 composés en mélange.

Partie 3 : Conclusion des synthèses de bibliothèques chimiques encodées ciblant les G-quadruplex.

L'assemblage d'une première librairie combinatoire de ligands potentiels de structure G-quadruplex a permis l'obtention d'une bibliothèque de 190 440 composés distincts. La ligation enzymatique des code-barres ADN (T1 à T4) ainsi que du *Closing Primer* est un processus simple, robuste et reproductible pour étiqueter les *building blocks* chimiques. Il est à noter que lors des processus de ligation, différentes quantités de départ d'ADN ont été utilisées. Hormis la nécessité ponctuelle de réitérer le processus de ligation, ce processus est applicable sans risque, quelle que soit l'échelle sur laquelle la réaction est initiée. Dans l'exemple de la DEL1, des ligations ont été effectuées sur des échantillons composés de 2 nmoles à 20 nmoles.

Les processus itératifs utilisés au sein de la DEL1 afin d'obtenir une bibliothèque de composés chimiques en mélange sont des réactions essentiellement basées sur des réactions de couplage peptidique mais également sur de la cycloaddition 1,3-dipolaire cuprocatalysée (Schéma 9).

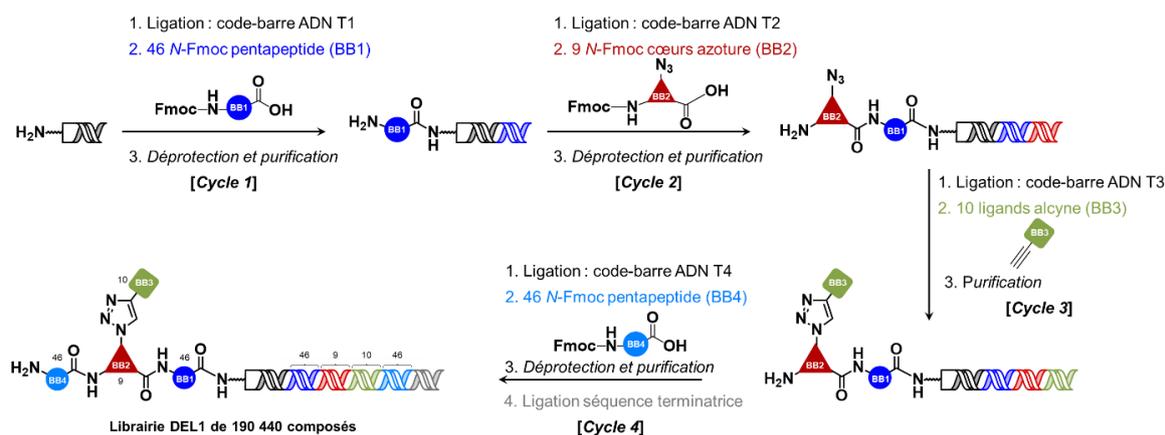


Schéma 9 : Synthèse complète de la première librairie combinatoire encodée DEL1 (190 440 composés). Représentation des cycles 1 à 4 jusqu'à l'incorporation du ClosingPrimer-1. Quantification finale de 5,32 nmoles.

L'incorporation des BB1 et BB2 dans les conditions EDC/HOAt/DIPEA ou NMM ont été optimisées par le Dr. SI. Néanmoins, les réactions n'ont pas permis d'atteindre des rendements complets de conversion pour les BB1. La présence en mélange de 46 composés

en fin du [Cycle 1] et de 414 composés au terme du [Cycle 2] peut encore être analysé afin de distinguer l'avancement des réactions. En revanche dès le [Cycle 3], plus de 4140 molécules en mélange ont été obtenues. Ce nombre important de molécules en mélange ne permet plus à la librairie d'être analysé par UPLC-MS. Des techniques de purification comme la technique de « Cap&Catch » ou d'analyse par électrophorèse retard ont permis d'évaluer les conversions mais également de purifier nos milieux réactionnels. Enfin, les BB4 ont été introduits, après une nouvelle campagne d'optimisation, en utilisant le DMT-MM comme agent de couplage.

Il est important de noter que les expériences de purification de la librairie suivant la méthode générale « (Biotine —) Cap & (Streptavidine —) Catch » se sont avérées particulièrement efficaces pour des systèmes aussi complexes. En parallèle, l'utilisation de l'électrophorèse sur gel d'agarose a permis de comparer les échantillons présentant un effet retard suite à la migration de la librairie en présence, ou non, de streptavidine. En revanche, il n'est nullement possible de s'assurer pleinement que les molécules biotinylées nécessaires à la purification de la librairie ont été introduites sur tous les composés tronqués de la librairie. Il est donc nécessaire de garder à l'esprit que la présence de molécules encodées dépourvues de certains motifs chimiques n'est pas à exclure. Lors de l'évaluation des composés obtenue après les études de sélections *in vitro* par affinité, il sera alors judicieux de considérer la synthèse de dérivés tronqués contrôles des touches obtenues.

La seconde librairie combinatoire de ligands de structure G-quadruplex a permis d'obtenir une librairie encodée de 696 200 molécules en mélange. La ligation enzymatique, au même titre que la DEL1, est une méthode d'étiquetage établie et optimisée permettant d'être réalisée dans des conditions optimales.

Par homologie avec la DEL1, les composés de la DEL2 ont été introduits en utilisant des réactions de couplage peptidique ainsi que des réactions de chimie click cuprocatalysé (Schéma 10).

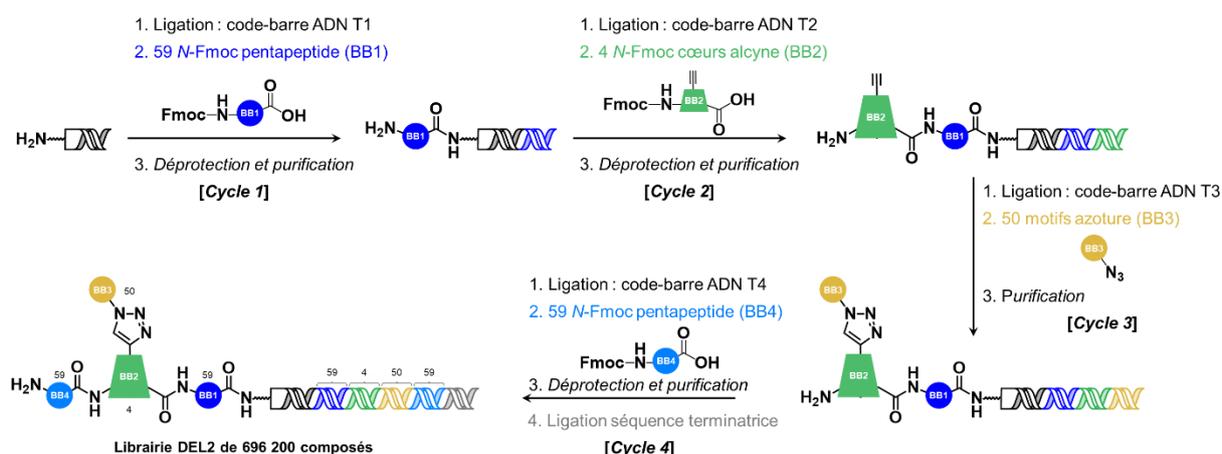


Schéma 10 : Synthèse complète de la deuxième librairie combinatoire encodée DEL2 (696 200 composés). Représentation des cycles 1 à 4 jusqu'à l'incorporation du ClosingPrimer-2. Quantification finale de 3 nmoles.

Contrairement à la première librairie, l'incorporation des *BB1* et *BB4* a été effectuée via l'utilisation de DMT-MM en tant qu'agent de couplage. L'incorporation des cœurs trifonctionnels (*BB2*) a été effectuée après essais de plusieurs méthodes de couplage. Finalement se sont l'EDC/HOAt/NMM qui a été utilisé pour coupler les *BB2* à l'exception du trioxazole qui a été assemblé en utilisant un protocole particulièrement récent, l'HATU ainsi que la 2,6 — Lutidine et qui méritait d'être testé sur tous les substrats amines et acides à notre disposition. Enfin, l'incorporation des motifs azotures des *BB3* a été réalisée par chimie click-Cu^{II}. Une nouvelle fois, je souhaite noter le travail considérable effectué par Alicia HOUDANT et Mattéo GARCIA pour la synthèse et la validation des motifs azotures à partir d'une librairie de composés présentant une fonction amine libre.

L'intérêt de faire la synthèse de touches tronquées s'applique également lors de la resynthèse en grande quantité des ligands identifiés comme prometteurs lors de sélections impliquant la DEL2. En effet, l'unique méthode de purification utilisée lors de l'assemblage itératif de la librairie a été la méthode de « Cap&Catch », laissant possiblement dans le milieu des composés chimiques privés d'un ou l'autre motif chimique.

Chapitre 3 — Sélections et analyses bioinformatiques des bibliothèques encodées par l'ADN ciblant les G-quadruplex.

Partie 1 : Processus de sélection *in vitro* par affinité.

Préambule.

Parmi l'ensemble des méthodes de sélections *in vitro* discutées dans ce manuscrit (Chapitre 1, *Section 2.2.1*, p74) deux principales catégories peuvent être distinguées : i) les sélections au cours desquelles les cibles d'intérêts sont immobilisées sur un support solide et ii) les sélections en solutions.

Dans le cadre de ce projet de thèse, le choix s'est porté vers l'utilisation de cibles immobilisées sur support solide due à la simplicité méthodologique et à la rapidité du processus. Historiquement, les sélections par affinité étaient effectuées en utilisant des billes de sépharose^{205,203,252,207,253}. Les cibles protéiques étaient soit couplées de manière covalente sur des billes de sépharose activé par du bromure de cyanogène (en. *CNBr-activated Sepharose beads*) ou liées à de la sépharose fonctionnalisée par de la streptavidine après biotinylation. Plus tard, des billes magnétiques ont été introduites permettant la transition des sélections réalisées manuellement à des sélections automatiques^{208,218,256,291,292}. Au-delà des fonctionnalisations covalentes ou des immobilisations *via* biotinylation, les billes magnétiques peuvent également utiliser d'autres stratégies de greffage comme des *His-tag*, *Flag-tag*, *strep-tag* par exemple²⁹³⁻²⁹⁶. L'utilisation de billes magnétiques porteuses de streptavidine pour l'immobilisation de séquences cibles fonctionnalisées par une biotine a été retenue dans le cadre du projet de thèse.

Dans les processus de sélections par affinité utilisant des supports solides pour attacher les cibles protéiques ou nucléiques, les billes utilisées disposent d'une capacité de chargement maximale. Cette capacité va déterminer la quantité de cibles maximales pouvant alors être accrochée sur les billes. Les sélections peuvent ainsi être effectuées à deux (ou plusieurs) concentrations de cibles : une première concentration permettant de saturer les billes avec la cible donnée, et une deuxième, bien plus faible, permettant d'appliquer des conditions de sélections plus stringentes. Enfin, la stringence peut être adressée en faisant varier la composition du tampon, le temps d'incubation et le nombre de lavages effectués²⁹⁷ lors des sélections. Enfin, pour récupérer les ligands sélectionnés après lavages, l'application de conditions d'éluions par dénaturation de la cible ou par rupture des interactions ligands/cibles peut être variée. Ces paramètres ont été évalués dans le cadre de cette thèse.

Au cours de cette thèse, les sélections ont été effectuées sur des cibles d'acides nucléiques disponibles commercialement et portant une fonction biotine en 3' de leur séquence (Tableau 9, p129). Corollairement, des billes magnétiques fonctionnalisées par de la streptavidine, et disponibles commercialement, ont donc été utilisées afin d'immobiliser les oligonucléotides biotinylés sur support.

Différents types de billes magnétiques-streptavidine (en. *SA beads*) sont disponibles commercialement chez ThermoFisher Scientific. On y retrouve les Dynabeads M-270 SA, les Dynabeads M-280 SA, les Dynabeads MyOne SA C1 et enfin les Dynabeads MyOne SA T1. Ces 4 types de billes SA diffèrent au niveau du diamètre des billes mais également au niveau de leur composition puisque certaines d'entre elles sont préalablement incubées et bloquées par de l'albumine de sérum bovin (en. *Bovine Serum Albumine*, BSA). Dans la littérature, les Dynabeads SA T1 restent les billes magnétiques les plus documentées pour la sélection de bibliothèques combinatoire^{218,292} et ont été celles utilisées lors de ce projet de thèse. Suivant les groupes et les cibles considérées, les Dynabeads M-270 SA sont également largement documentées^{256,291,292}.

En termes de caractéristiques, les billes Dynabeads MyOne SA T1 possèdent un diamètre de 1,05µm²⁹⁸. Elles sont composées de chaînes de polystyrènes adoptant une forme cylindrique sur laquelle est successivement ajoutée une couche de polymères poreux chargée par de l'oxyde de Fer, donnant les propriétés magnétiques aux matériaux. Enfin, on retrouve une dernière couche externe de polymères électroniquement inerte sur laquelle est ajoutée la fonctionnalisation souhaitée, comme ici, la streptavidine. En terme de chargement, les Dynabeads MyOne SA T1 ont la capacité théorique d'immobiliser 400 pmoles d'oligonucléotides (*i. e* de G-quadruplex) par 1 mg de billes.

Au début de la procédure de sélection, les oligonucléotides biotinylés ont été immobilisés sur des billes magnétiques dans un tampon B/W (en. *Binding and Washing buffer* : 10mM Tris.HCl pH 7.5, 1mM Na₂EDTA et 2M NaCl) tel que préconisé par le fournisseur. Généralement, une expérience de sélection est réalisée à partir de 25 pmoles de G-quadruplex immobilisé (soit 6,25µL de billes). Ainsi, 125 pmoles de G-quadruplex-biotine (5 eq) sont incubées en présence de 6,25µL de billes SA à une concentration finale de 2,5µg billes/µL.

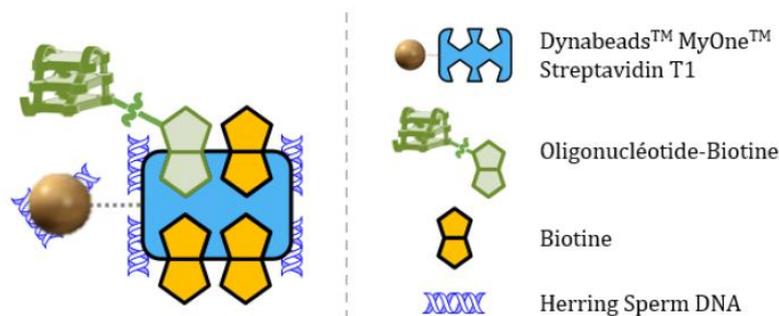


Figure 84 : Représentation schématique du système de billes magnétiques (*Dynabeads™ MyOne™ T1*) streptavidines. Fonctionnalisation des billes par des séquences oligonucléotidiques biotinylés et coiffage des sites de la streptavidine par de la biotine libre en solution. L'ADN génomique est également rajouté au système pour prévenir des interactions non-spécifiques.

La streptavidine est une protéine homotétramérique de 52,8 kDa obtenue par purification de la bactérie *Streptomyces avidinii*. Chaque unité monomérique de cette protéine est composée de 8 feuillets bêta antiparallèles qui se replient afin de former un tonneau bêta antiparallèle dans laquelle la biotine peut se lier avec une constante de dissociation de l'ordre de 10⁻¹⁴

mol/L. Finalement, la biotine est une molécule capable de reconnaître chaque tonneau bêta de la streptavidine permettant ainsi à 4 molécules de biotine d'interagir avec l'homotétramère. Lors de l'immobilisation des cibles biotinylées, distribuées statistiquement sur les billes magnétiques fonctionnalisées SA, de nombreuses unités monomériques seront vacantes. Il est alors important de bloquer ces sites d'interactions. En effet, l'éventualité qu'une librairie combinatoire de plusieurs milliers de composés puisse détenir des molécules capables d'interagir avec les unités monomériques de la streptavidine n'est pas à négliger. Ainsi, après la fonctionnalisation des billes avec les oligonucléotides (5eq), une étape de coiffage des sites SA est réalisée par incubation de la biotine libre en large excès (100 eq) avec les billes magnétiques (25 pmoles G4-immobilisés, 6,25µL de billes, [c]= 2,5µg billes/µL).

Une fois les billes magnétiques fonctionnalisées par les cibles d'intérêts et saturées avec de la biotine en solution (Figure 84), l'incubation de la librairie dans le tampon de sélection peut être réalisée.

Les tampons de sélection utilisés pour les expériences *in vitro* varient en fonction de la cible considérée. Dans le cas du projet de thèse, les cibles considérées sont des cibles nucléiques capables de former des structures G-quadruplex sous conditions stabilisantes. En effet, la plupart des cibles protéiques ou nucléiques peuvent être criblées en présence de tampon PBS voire de tampon HEPES. Dans le cas des cibles quadruplex, le tampon utilisé est le PBS auquel est ajouté 100mM d'ions K⁺ (KCl) afin de stabiliser la structure G-quadruplex. Au sein de ce tampon, il n'est pas rare de retrouver la présence d'un tensioactif non ionique comme le Tween-20 (ou T-20). En effet, l'utilisation de ce détergent permet de réduire l'agglomérat de billes magnétiques ainsi que leur adhésion aux plastiques, des tubes à microcentrifugation jusqu'aux pointes de pipettes. Le T-20 est également utilisé dans le but de diminuer le bruit de fond des résultats faux-positifs, c'est-à-dire diminuer les possibles interactions électrostatiques ou hydrophobes non-spécifiques des ligands sur la cible G-quadruplexe²³⁵ ou les plastiques.

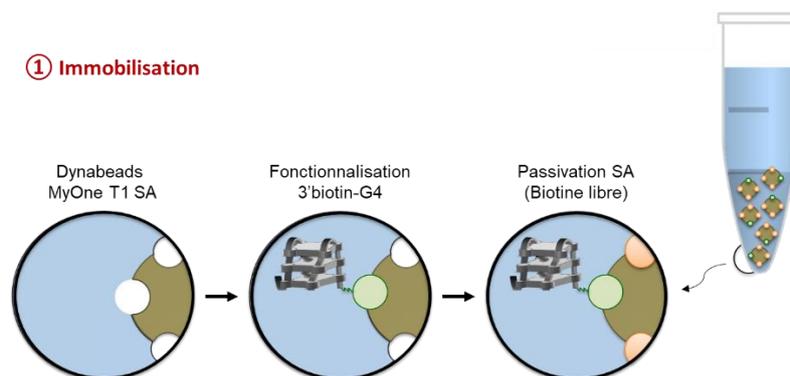
L'incubation de la DEL sur un volume de billes de 6,25µL, soit 25 pmoles de G-quadruplex, a été réalisée principalement avec 5 pmoles de librairie combinatoire. En effet, les conditions standards de travail utilisées rendent compte d'un système où les G4-quadruplexes sont théoriquement dans un excès de cinq par rapport à la librairie en solution, le ratio « G4-immobilisé:DEL » est donc de 5:1. Ces conditions standards ont été choisies au regard des précédents dans la littérature pour la sélection de DEL, et confortées par les résultats empiriques obtenus lors des premières optimisations de sélection. Il sera évidemment discuté plus tard de l'intérêt de travailler en condition plus stringentes par défaut de cible par rapport aux quantités de librairies appliquées. Avant même que la librairie soit mise en présence de la cible, 100 eq. d'un ADN génomique d'hareng (en. *Herring Sperm DNA*) est ajouté au milieu réactionnel. L'intérêt d'ajouter de l'ADN génomique (de tailles moyennes de 200 paires de bases) dès le début de l'incubation est de pouvoir favoriser la diminution des interactions non-spécifiques des composants de la librairie avec la cible d'intérêt, la protéine streptavidine mais également avec les billes magnétiques elles-mêmes. Par ailleurs, ce large excès de double brin d'ADN en solution doit également permettre de piéger en solution les potentiels composés de la librairie combinatoire ayant une affinité exclusive pour l'ADN double brin ou même

partagée avec la cible G-quadruplex. Les incubations standards ont duré 45 min à 25°C dans le tampon de sélection, la concentration finale des billes est alors de 0,85 $\mu\text{g}_{\text{billes}}/\mu\text{L}$, celle en G-quadruplex de 338 nM et celle de la librairie combinatoire encodée de 67,6 nM.

Une fois la période d'incubation terminée, les billes magnétiques sont physiquement séparées du surnageant d'incubation (en. *Flow-through*) à l'aide d'aimants montés sur support. Les ligands présentant une affinité pour la cible sont alors susceptibles d'être retenus sur les billes. Afin de discriminer (séparer) les ligands de haute affinité des ligands de plus modeste affinité, des étapes de lavage peuvent être réalisées. Le nombre de lavages successifs, réalisés avec le tampon de sélection, peut être ajusté suivant la composition de la librairie. Dans un contexte de ligands présentant une haute affinité pour la cible, la stringence des lavages peut être augmentée. Sont compris dans la stringence des lavages la possibilité de jouer sur le nombre de lavages successifs, sur la température de réalisation des lavages mais également sur le temps appliqué aux lavages. A l'opposé, si la librairie n'est composée que de ligands à modeste affinité pour la cible, des conditions de lavages moins stringentes seront appliquées. Le compromis souvent décrit dans la littérature est l'utilisation de cinq étapes de lavages successives avec une durée moyenne de quelques secondes (à minutes) par lavage^{218,256,291,292}.

Une fois sélectionnés après 5 lavages successifs, les composés présentant la plus forte interaction avec la cible seront élués en présence de conditions dénaturantes. Ces conditions peuvent être appliquées par variation de température de solutions aqueuses, tamponnées, voire organiques.

Le schéma ci-dessous récapitule les principales étapes (1 à 5) d'un processus de sélection standard par affinité (Figure 85). Chacune de ces étapes a été étudiée afin de quantifier l'impact d'une ou plusieurs variations des conditions standards.



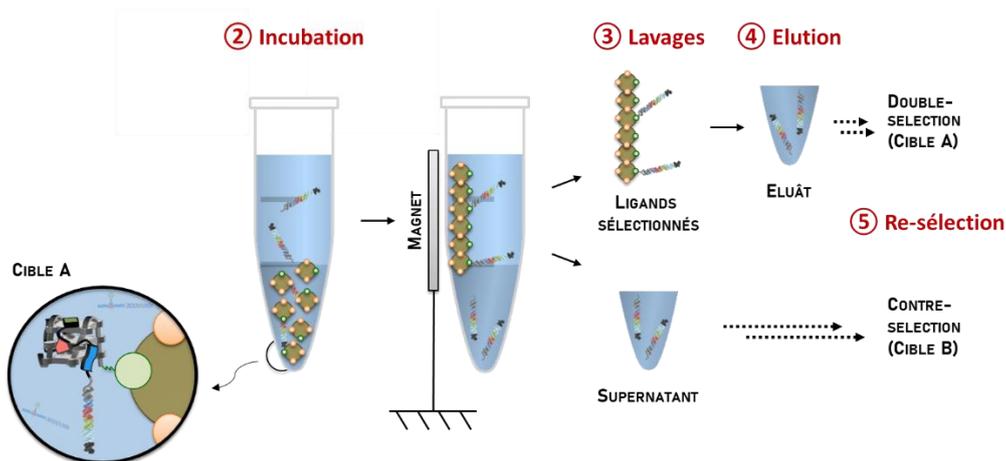


Figure 85 : Expérience de sélection *in vitro* par affinité réalisée de manière séquentielle : 1) immobilisation des cibles, 2) incubation de la DEL, 3) lavages et applications de stringences et 4) élution des composés les plus affins pour la cible. Dans certains cas, une dernière étape de double — ou contre-sélection 5) peut être effectuée.

Sous-partie 1.1 : DEL1 : première campagne de sélections.

Section 1.1.1 : Processus de sélection *in vitro*.

Une première campagne de sélections a été réalisée à partir de la DEL1, vis-à-vis de cibles G-quadruplex adoptant différentes topologies ainsi que différents contrôles. La liste des séquences formant des structures G-quadruplex ainsi que les séquences contrôles est reportée dans la Tableau 9. Pour cette première campagne de sélection nous avons choisi des G-quadruplex cibles d'intérêts thérapeutiques décrits dans le premier chapitre adoptant des topologies variées.

Effet de l'immobilisation des cibles G-quadruplex sur les Dynabeads MyOne T1 SA — ①

Les cibles adoptant une topologie parallèle sont les deux séquences promotrices d'oncogènes, Myc et Kit1. Un variant monomorphique plus stable de la séquence Myc, appelée Myc22, a également été utilisé. La séquence de topologie antiparallèle utilisée est le mutant de la séquence télomérique humaine, 22CTA. Enfin, le modèle polymorphique de la séquence télomérique humaine, wt-Tel26, a été utilisé en tant que G-quadruplex à topologie majoritairement hybride. Au-delà des cibles G-quadruplex, nous avons testé la librairie en présence de différentes cibles contrôles permettant ainsi de repérer les ligands non-spécifique aux structures G-quadruplex. Ainsi, la séquence riche en guanine appelée scramble-26 (scr26) structurellement incapable de former une structure G4 ainsi qu'un ADN double brin en épingle (en. *Hairpin* ou hp1) ont été des contrôles négatifs de choix.

Tableau 9 : Séquences oligonucléotidiques (3' — biotine) utilisées dans le cadre de la première campagne de sélection *in vitro* par affinité à partir de la DEL1.

Nom	Séquence	Description
<i>Myc-3'biot</i>	GGA GGG TGG GGA GGG TGG GGA A/3BioTEG/	G4 (parallèle), promoteur c-myc
<i>Myc22-3'biot</i>	TTG AGG GTG GGT AGG GTG GGT AA/3BioTEG/	G4 (parallèle), promoteur c-myc avec substitution de G en T, PDB: 1XAV
<i>22CTA-3'biot</i>	AGG GCT AGG GCT AGG GCT AGG GTT/3BioTEG/	G4 (antiparallèle), ADN télomérique humain (variant), PDB: 2KM3
<i>Kit1-3'biot</i>	AGG GAG GGC GCT GGG AGG AGG GTT/3BioTEG/	G4 (parallèle avec G isolé), promoteur c-kit1, PDB: 203M
<i>wtTel26-3'biot</i>	TTA GGG TTA GGG TTA GGG TTA GGG TT/3BioTEG/	G4 (hybride), ADN télomérique humain, PDB: 2JPZ
<i>scr26-3'biot</i>	AAG TGT GTG TGT GTG TGT GTG TGA AG/3BioTEG/	Séquence riche en guanines
<i>hp1-3'biot</i>	ATA GCA TAC GTT TTC GTA TGC TAT T/3BioTEG/	ADN en épingle à cheveux de 25 bases

Dans les conditions standards décrites ci-dessus, un ratio oligonucléotide cible/librairie (O/L) de 5:1 a été utilisé pour l'immobilisation des différentes cibles sur les billes magnétiques. Afin d'évaluer l'effet de conditions moins stringentes, un ratio de 10:1 en faveur d'un excès de cible a également été étudié.

A *contrario*, l'application de systèmes O/L avec défaut de cibles a également été réalisé avec des rapports O/L de 0.5:1 et 0.05:1. Dans ce cas, les sélections sont orientées vers des systèmes plus compétitifs ou les ligands présentant les meilleures cinétiques d'association ou constantes d'association sont capables de fixer les cibles G-quadruplex présentes en défaut.

L'évaluation de l'impact du rapport O/L sur le processus de sélection a été déterminée sur la cible Myc. Le paramètre pris en compte a été la quantification de la librairie encodée, théoriquement enrichie en ligands affins, après sélection et amplification en qPCR. Les procédures d'incubation, de lavage et d'éluions utilisées sont les conditions standards telles que décrites en détail plus loin dans le manuscrit et récapitulé dans le tableau 12, p139. La quantification par qPCR réalisée en plaque 96-puits est également détaillée plus loin dans cette section. Elles s'appuient sur les travaux de Dr. Neri et permettent de calculer un nombre de copies obtenues après élution des meilleurs ligands dans une sélection donnée et ainsi de déterminer la stringence des conditions utilisées.

La quantification est ainsi effectuée à partir d'une gamme étalon d'amplification de la librairie à des concentrations connues. En effet, cette gamme varie de 1E+12 à 1E+05 copies (molécules) comme indiquées sur la Figure 86 ci-dessous. Conjointement à la gamme étalon, un contrôle sans matrice (en. *Non-templated control*, NTC) contenant tout le nécessaire à l'amplification (amorces, polymérases, dNTPs, etc...) ainsi qu'un contrôle en présence seule du tampon d'amplification (en. *Buffer*) ont également été effectués.

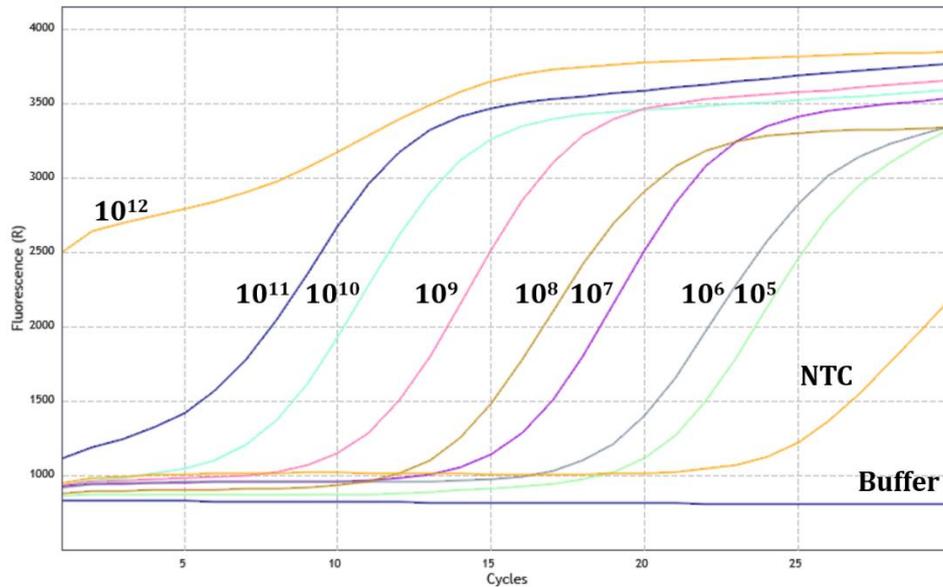


Figure 86 : Courbes d'amplifications obtenues par qPCR. Représentation de l'intensité de fluorescence en fonction du nombre de cycles d'amplifications. Analyse effectuée pour chaque concentration (10^{12} à 10^5 copies) de librairie DEL1 non sélectionnée ainsi que pour les contrôles NTC et buffer.

Initialement 5 pmoles de DEL1 sont incubées à chaque condition de sélection, correspondant à $3E+12$ copies. Le nombre de copies post-sélection reflète donc la stringence du processus de sélection. Celui-ci n'est bien sûr pas le seul indicateur de la pertinence d'une condition de sélection puisqu'à nombre de copies identiques la population des ligands encodés peut largement varier. Néanmoins il est possible de postuler qu'un nombre de copies post-sélection trop élevé témoigne d'une stringence trop faible. Ceci ne permettant pas de discriminer entre les bons ligands et les ligands d'affinité plus modeste. Au contraire, un nombre de copies trop bas doit alerter quant à la présence de biais potentiel lors de l'amplification et donc d'un processus de sélection trop stringent.

Plusieurs observations importantes quant aux profils d'amplification des gammes et des contrôles sont à prendre en considération. Premièrement, la gamme étalon complète est observable en effectuant une amplification de 30 cycles successifs. Notons que pour cette raison, ces conditions seront celles maintenues pour chacun des profils de sélections présentés. Deuxièmement, il est visualisable qu'en présence d'une quantité d'ADN supérieure à $1E+11$ copies, l'intensité de fluorescence initiale est très élevée. Enfin, lorsque le plateau d'amplification des quantités d'ADN inférieures à $1E+06$ copies ont été atteints, la fluorescence du contrôle NTC n'est plus négligeable. En effet, à partir de 21 cycles d'amplifications, une composante NTC ne cesse de croître. La présence unique des amorces de qPCR (adaptateurs *Multiplexing* et *TrueSeq*) au sein du NTC laisse entendre qu'il y a formation de dimères d'amorces (en. *Primer dimers*). Cette hypothèse a d'ailleurs été confirmée sur gel d'agarose. La formation de ces amplicons non pertinents au bout d'un certain nombre de cycles de PCR impose donc un nombre de copies post-sélection suffisamment important pour ne pas avoir à amplifier au point d'amplifier ce bruit de fond.

Quel que soit le système O/L utilisé, 5 pmoles de DEL1 ont été incubées en présence de 50, 25, 2,5 ou 0,25 pmoles de G-quadruplex, représentant respectivement les ratios O/L de 10:1, 5:1, 0,5:1 ou 0,05:1. L'évaluation du nombre de copies après application des différents ratios O/L se retrouve en Tableau 10.

Tableau 10 : Quantification du nombre de copies obtenues après sélection contre la séquence du proto-oncogène Myc en fonction du ratio O/L (oligonucléotide:librairie).

Ratio	Copies	Cible
0,05:1	3.3E+06	G4 (parallèle), promoteur c-myc
0,5:1	2.1E+07	G4 (parallèle), promoteur c-myc
5:1	6.8E+09	G4 (parallèle), promoteur c-myc
10:1	8.8E+09	G4 (parallèle), promoteur c-myc

De manière prévisible, le rapport O/L impact incontestablement le nombre de copies obtenues après élution (en. *output copies*) des meilleurs ligands. En effet, en défaut d'oligonucléotides, les éluats de sélections décroissent proportionnellement au ratio O/L considéré jusqu'à atteindre des valeurs avoisinant 1E +06 copies. À l'inverse, les différents excès d'oligonucléotides n'influencent que très peu les *output copies* et paraissent même plafonner aux alentours de 1E +10 copies.

Alors qu'il est aisé de quantifier de hautes valeurs d'éluats de sélections, il en est expérimentalement plus compliqué de quantifier de faibles valeurs comme celles obtenues en présence d'un ratio 0.05:1. Lors de la quantification du contrôle NTC, une valeur d'amplicon variant de 1E +05 à 1E +06 copies a été obtenue. Autrement dit, à des valeurs suffisamment basses comme la valeur obtenue en présence d'un ratio O/L de 0,05:1, la quantification exacte de l'éluat de sélection est impossible due à ces valeurs de NTC trop proches.

Ainsi pour cette première campagne de sélection, le choix s'est porté vers l'évaluation de sélections à un ratio de 5:1 ainsi qu'à 0.5:1 dans le cadre d'application de stringences.

Incubation de la DEL1 sur les billes post-fonctionnalisées par les cibles d'intérêts - ②

Une fois les séquences G-quadruplex immobilisées sur billes magnétiques et coiffées par de la biotine en solution, les 5 pmoles de la librairie encodée peuvent être incubées en présence de la cible. L'incubation classique a été réalisée dans le tampon d'incubation et en présence d'ADN de sperme de hareng.

Des expériences de sélections ont également été menées en l'absence d'ADN de sperme de hareng afin d'étudier l'évolution du nombre de copies en sortie de sélection et d'informer sur l'éventuelle formation de complexe ligands encodés/ADN duplex. Dans le système standard, la quantité de molécules obtenue après élution demeure sensiblement la même en présence ou en absence d'ADN de sperme de hareng. Ainsi, afin de limiter les faux-positifs, l'ADN génomique a été jugé nécessaire à toutes les expériences et reste constamment présent dans les expériences à une quantité fixée de 500 pmoles (soit 100 équivalents par rapport à la DEL).

À l'exception des sélections réalisées contre l'hairpin (mime de l'ADN double brin) et ceux pour permettre l'identification des « meilleurs » ligands d'hairpin qui serait, en présence d'ADN de sperme de hareng, probablement piégé en solution.

Le tampon de sélection contient 100mM de Chlorure de potassium (KCl), 1X de tampon phosphate salin (en. *Phosphate buffered saline* ou PBS) à pH 7,4 et une concentration variable de Tween-20 (T-20). La concentration en T-20 a également fait l'objet d'expérimentation.

L'impact de la concentration de T-20 dans le processus de sélection a été évalué par application de concentration pouvant varier de 0,1% à 0,01% (du volume de travail). Comme le montre le tableau 11 ci-dessous, l'augmentation de la concentration volumique de Tween-20 n'affecte que très peu le nombre de copies après éluions. Lors de cette première campagne de sélection, nous avons donc retenu des protocoles impliquant 0,1% et 0,01% de T-20. Néanmoins, la concentration de 0,01% étant plus largement utilisée dans des sélections similaires, cette condition a été retenue comme standard et utilisée dans la plupart des sélections.

Tableau 11 : Quantification du nombre de copies obtenues après sélection contre la séquence du proto-oncogène Myc en fonction du pourcentage de détergent (Tween-20) utilisé.

[Tween-20]	Copies	Cible
0,01%	2E+07	G4 (parallèle), promoteur c-myc
0,05%	3.1E+07	G4 (parallèle), promoteur c-myc
0,1%	3.3E+07	G4 (parallèle), promoteur c-myc

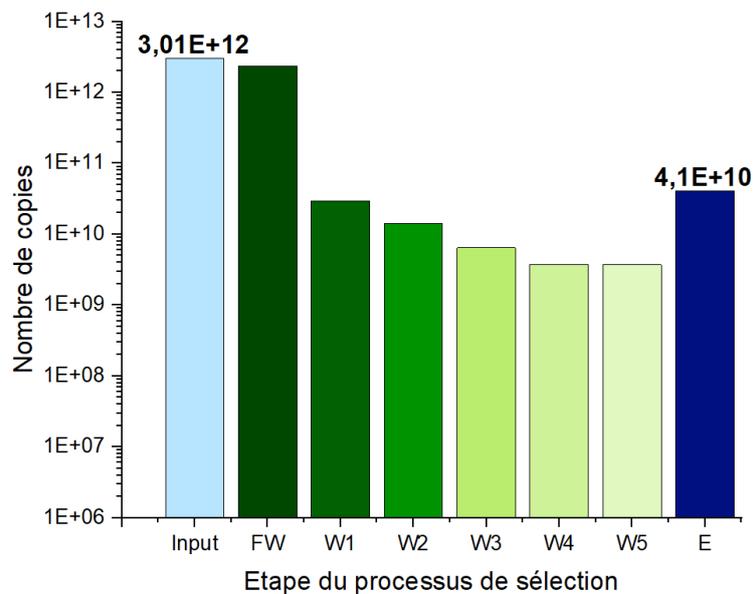
Lavages après incubation de la DEL1 - ③

Après incubation de 5 pmoles de DEL1 (1 h à 25°C) sur une quantité définie d'oligonucléotide immobilisé, le surnageant de la sélection (en. *Flow-through*) est séparé des billes magnétiques à l'aide d'aimants placés sur support. A ce stade d'avancement du processus de sélection, le surnageant est composé des ligands de la librairie non liés à la cible G-quadruplex (en. *non-binders*) en solution dans le tampon de sélection (100 mM KCl, 1X PBS (pH 7.4), 0.01% Tween-20).

Dans l'intérêt d'éliminer les ligands les plus faiblement affins pour la cible en présence (en. *weak-binders*), nous avons utilisé plusieurs lavages successifs. L'effet de ces lavages sur le nombre de copies élué est directement lié aux conditions de lavages. En effet, la stringence peut être dictée par l'application de conditions douces de lavage, comme par exemple l'application de cinq lavages successifs avec le tampon de sélection, ou endurcis par i) des conditions de hautes températures, ii) des conditions de concentration faible en ion K⁺ ou iii) des conditions de concentration en Tween-20 plus élevée.

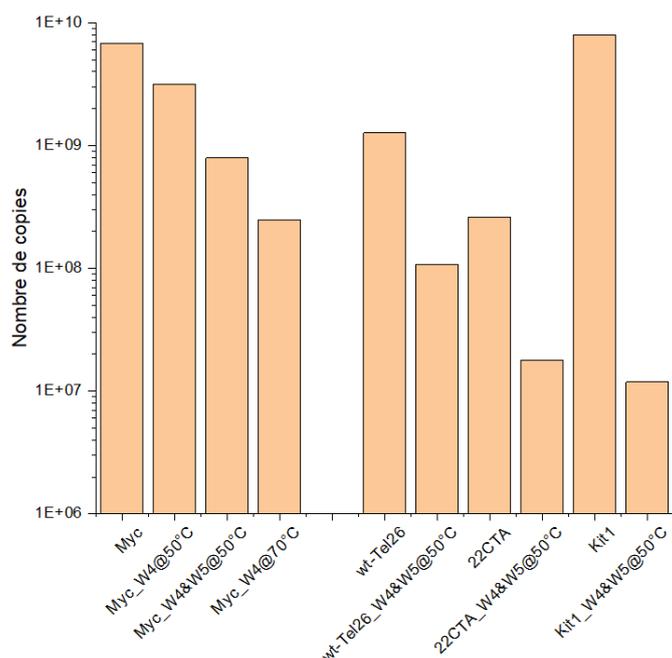
Le graphique 2 montre un exemple de quantification de toutes les solutions obtenues, du *Flow-through* à l'éluion. Dans cet exemple, la DEL1 (5pmol — 3.01E+12copies ; *input* sur le Graphique 2) a été incubée en présence de la cible Myc dans un ratio O/L de 5:1. Tout d'abord,

les non-bindés présents dans *Flow-through* (FW) sont récupérés et présentent un pourcentage de relargage très important estimé à plus de 95%. Les lavages successifs en conditions douces (tampon de sélection : 100mM KCl, 1X PBS pH 7.4, 0.01% Tween-20) relarguent une quantité décroissante de 1% à 1‰ en fonction du cycle de lavage et permet ainsi d'obtenir une quantité d'éluats (E) équivalent à $4.08E+10$ copies soit 1,4% de la quantité initialement incubée (élution dans l'H₂O à 95°C comme explicité plus loin dans le manuscrit, Tableau 12, p139).



Graphique 2 : Quantification du nombre de copies obtenues en fonction des étapes du processus de sélection. 1) Input copies, 2) Flow-through, 3) W1 à W5 : lavages successifs, 4) Output copies.

Comme explicité auparavant, la stringence des sélections peut être appliquée sous différentes formes en commençant par la variation de la température. En effet, au cours des cinq cycles de lavages, nous avons exploré l'utilisation de températures de déstabilisation du complexe ligands/cibles allant de 50 à 80°C (Graphique 3).



Graphique 3 : Quantification du nombre de copies obtenues en fonction de la stringence des lavages (W1 à W5) appliquée et de la cible G-quadruplex considérée. Sans mention contraire, les lavages ont été réalisés à température ambiante (25°C) et sous agitation pendant 15 min.

Comme le montre le graphique ci-dessus, le nombre de copies après élution contre la cible Myc décroît largement en fonction de la température appliquée lors des lavages. En condition douce, c'est-à-dire en appliquant 5 lavages successifs avec le tampon de sélection (« Myc », Graphique 3), l'élution relargue 6.77E+09 copies, soit 2,3‰ de la quantité injectée dans la sélection (en. *input copies*). Lorsque les mêmes conditions sont appliquées mais que le lavage 4 (W4) est chauffé à 50°C, on observe une élution de 3.2E+09 copies, correspondant à 1,1‰. Dans le nouveau cas où le lavage 4 et le lavage 5 (W4+W5) sont effectués à 50°C, seul 0,3‰ des copies sont maintenues, correspondant à un *output* de 7.62E+08 copies. Enfin, lorsque le lavage 4 (W4) est appliqué à une température de 70°C, l'élution permet un relargage de 2.47E+08 copies, soit 0,08‰ de l'*input copies*.

Afin de montrer que cette stringence n'est pas corrélée à une séquence d'oligonucléotide et *in fine* à la topologie adoptée par le G-quadruplex, il a également été montré que le même effet de décroissance du nombre de copies des éluats de sélection est observable sur la séquence polymorphique wtTel26, la séquence de topologie antiparallèle 22CTA ou encore la séquence promotrice d'oncogène kit1 (Graphique 3).

La stringence peut également être contrôlée par un manque d'ions K⁺ en solution puisque l'ion potassium est un stabilisateur de structure G-quadruplex. Ainsi, lorsqu'un des cycles de lavages est appliqué en l'absence de l'ion K⁺, c'est-à-dire dans un tampon contenant uniquement 1X PBS pH 7.4 et 0.01 % T-20, on observe une baisse de copies d'élution passant de 4.08E+10 à 1.46E+10, soit une décroissance de 65%. En revanche, lorsqu'on compare l'effet de la présence de l'ion K⁺ au sein d'une sélection où le quatrième lavage est effectué à 50°C, aucune diminution du nombre de copies n'est observable. Ainsi, l'absence de l'ion K⁺

semble influencer les copies éluées dès lors que les stringences de températures ne sont pas appliquées.

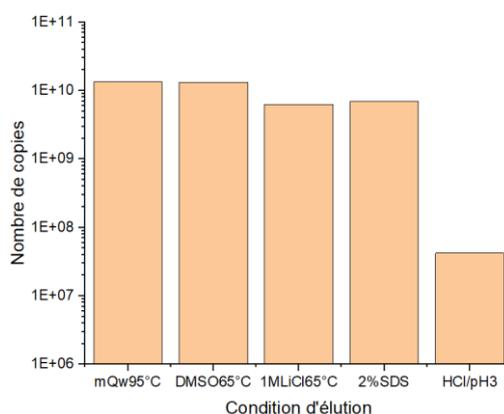
Finalement, malgré une très faible influence de la concentration de Tween-20 dans les expériences réalisées à un ratio de 0.5:1 (cf. section ② Tableau 11, p132), il a été jugé opportun de comparer cette fois-ci la concentration de détergent sur un ratio O/L présentant un excès d'oligonucléotides à hauteur de cinq équivalents (ratio O/L=5:1). Ainsi, l'expérimentation conduite sur le promoteur Myc montre une baisse d'un facteur 5 sur l'*output copies* en comparant cinq lavages successifs en présence de 0,01% de Tween-20 ou en présence de 0,1% de Tween-20.

En conclusion, parmi toutes les pistes explorées afin d'augmenter la stringence des lavages, il apparaît clairement que la réduction de concentration en KCl ou l'augmentation de la concentration de détergent n'influence que très peu les *output copies* (facteur 3 à 5). En revanche, l'effet de la température sur certains cycles de lavages permet d'avoir davantage de disparité puisqu'il est possible de faire diminuer les *output copies* d'un facteur 30 à un facteur 600 en fonction des séquences G-quadruplex considérées.

Conditions d'élution de sélections - ④

Une fois les lavages effectués en présence ou non de conditions stringentes, les meilleurs ligands encodés doivent être élués afin de pouvoir les quantifier et les identifier. Pour ce faire, nous avons exploré différents milieux aussi bien aqueux qu'organique.

Les trois premières conditions à avoir été expérimentées ont été : i) une élution en milieu purement organique (DMSO) à une température de 65°C, puis ii) de l'eau ultrapure chauffée à 95°C et enfin iii) un tampon salin composé de Chlorure de Lithium à 1M chauffé à 65°C (Graphique 4). En effet, ce dernier a pour objectif de déstabiliser les structures quadruplex d'une part grâce à la température mais également par la présence d'une grande concentration d'ions Li⁺. À l'instar du cation potassium, l'ion lithium ne stabilise pas la structure G-quadruplex due à son faible rayon atomique rendant impossible la coordination des oxygènes O₆ des guanines formant les quartets.



Graphique 4 : Quantification du nombre de copies obtenues en fonction des conditions d'élutions utilisées. Exemple sur la cible Myc.

De ces premières conditions, il apparaît que le milieu organique et le milieu aqueux relarguent le même nombre de copies alors que la condition saline ne permet d'éluer que la moitié des ligands. A également noter que l'élution en milieu organique est suivie par un bref lavage en milieu aqueux car certains ligands présents dans la librairie sont des ligands fortement polaires et chargés qui pourraient être relargués en présence de DMSO mais non soluble au milieu organique.

Après ce premier jeu de conditions, d'autres méthodes d'élutions ont été testées. Parmi celles-ci on retrouve une solution aqueuse de dodécylsulfate de sodium (SDS) à 2%, ou encore une solution d'acide chlorhydrique à pH 3 (Graphique 4). En comparaison au milieu aqueux, la solution acide ne permet pas d'éluer de grosse quantité de ligands, ou bien ce milieu conduit à la dégradation du code-barres ADN ne permettant une amplification quantitative. Le milieu détergent est quant à lui capable d'éluer une quantité raisonnable de molécules mais reste néanmoins inférieur à l'élution en milieu aqueux à 95°C.

Pour conclure, la méthode d'élution de choix utilisée pour les tests de sélection a été l'eau ultrapure (*milliQ water*) chauffée à 95°C. Toutefois la condition en milieu organique (DMSO à 65°C) suivi d'un lavage aqueux, a également été utilisée afin de comparer la population des ligands élués.

Double sélection et contre-sélection — ⑤

Après une sélection par affinité vis-à-vis d'une cible G-quadruplex, les ligands sélectionnés peuvent être identifiés grâce à leur code-barre ADN en utilisant une technique de séquençage à haut débit. Néanmoins, dans l'objectif d'identifier les ligands les plus affins ou des ligands spécifiques d'une structure en particulier, il est intéressant d'effectuer plusieurs sélections ou contre-sélections successives.

Afin d'augmenter la stringence par sélections successives, l'incorporation d'éluats de sélections au sein de cycles de sélections supplémentaires nécessite certaines conditions particulières. Dans un premier temps, il est indispensable de considérer le rapport *output/input* copies quantifiable en fin de sélection puisque celui-ci est directement fonction de l'absence ou de la présence de conditions stringentes. Dans notre cas il a été montré (Graphique 3) que ce rapport pouvait varier de 0,25% ($7.5E+9$ copies) en présence de conditions douces à 4E-04% ($1.2E+07$ copies) après application de conditions plus drastiques. Dans un deuxième temps il est nécessaire de se rappeler que la quantification des éluats de sélections n'est possible qu'à des valeurs supérieures à $1E+05$ et $1E+06$ copies, valeur représentant la quantification de l'échantillon NTC (en. *Non-template control*).

Ainsi, la somme des « n » sélections successives doivent être engagées de manière à ce que les *output copies* n'atteignent jamais des valeurs inférieures au seuil minimum mesurable, soit les valeurs de NTC. En ce sens, le choix effectué s'est porté vers l'application de maximum deux tests de sélections successifs en conditions standards. Étant donné que les *inputs copies* d'un test de sélection sont fixées à $3.01E+12$ et qu'il a été montré que le ratio *output/input* en condition standard permettait d'éluer 0,25% de la quantité soumise à sélection, il est

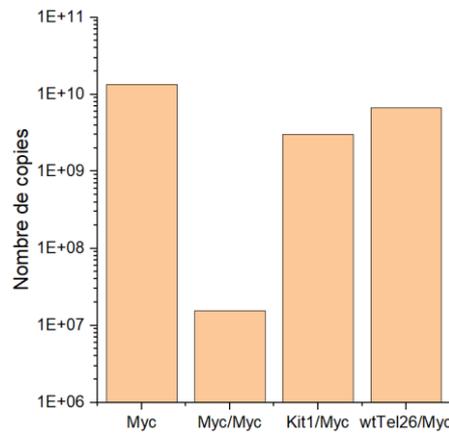
théoriquement possible d'obtenir une valeur d'*output copies* proche de $2E +07$ copies après deux sélections successives. Il est alors évident que l'application d'un troisième test de sélection, même appliqué dans les conditions les plus douces, ne permettrait pas de véritablement quantifier les *outputs copies* ($<10^5$ copies théorique).

Dans l'intérêt d'un gain d'affinité ou de spécificité, ces sélections successives peuvent être de deux types. Premièrement, on peut incuber la librairie sur une cible donnée A, et réinjecter l'éluât de sélection contre cette même cible. L'intérêt de cette double-sélection est d'identifier les composés les plus affins capables de persister après deux tours de sélections. Deuxièmement, il est possible d'incuber la librairie sur une cible A puis d'injecter le surnageant, c'est-à-dire les ligands les moins affins pour la cible A, sur une nouvelle cible B. Cette approche est une des possibilités de « contre-sélection », cette technique permettant dans un premier temps d'éliminer les meilleurs ligands de A, puis, parmi l'ensemble des ligands moins affins pour A, de sélectionner les meilleurs ligands de B (Figure 85).

Une double-sélection a été effectuée en présence de la cible unique Myc dans un ratio de départ O/L de 5:1. L'application des conditions standards aux deux sélections successives ont permis d'obtenir un nombre de copies finales d'environ $1E +07$ copies, en accord avec ce qui a été discuté plus haut dans cette section.

En parallèle, deux contre-sélections distinctes ont été expérimentées en conditions standards. D'abord, les ligands de la librairie combinatoire ont été sélectionnés vis-à-vis du promoteur d'oncogène Kit (Kit1, G4 parallèle) ou de la séquence télomérique humaine (wtTel26, G4 hybride) avec un ratio O/L de 10:1. Il a été pensé que, malgré un très faible écart d'*output copies* entre le ratio 10:1 et 5:1 une plus grande diversité de ligands pourrait être obtenue. Ensuite, les surnageants respectifs ont été réinjectés dans une procédure de sélection contre la séquence oligonucléotidique Myc (G4 parallèle) dans un ratio O/L de 5:1 (Graphique 5).

Le nombre de copies obtenu selon le processus de contre-sélection décrit ci-dessus est élevé. Il témoigne de la grande quantité de ligands encodés présents dans les solutions d'incubation (flow-through) après la première incubation. Il apparaît probable au regard de ces chiffres que cette approche de contre-sélection n'est pas prometteuse, un trop grand nombre de bons ligands potentiels des « non-cibles » persistant dans les flow-through après la première sélection.



Graphique 5 : Quantification du nombre de copies obtenues après sélection classique, double-sélection ou contre-sélection sur la cible principale Myc.

Au regard des optimisations décrites ci-dessus, une liste de premières sélections à séquencer a été déterminée.

Concernant les cibles (①), sept séquences distinctes ont été prises en compte pour la sélection comprenant trois séquences G-quadruplex de topologies parallèles (Myc, Myc22 et Kit1), une séquence de topologie antiparallèle (22CTA) et une séquence de topologie hybride (wt-tel26). A cela ont été ajoutés les contrôles négatifs que sont la séquence riche en guanines scr26 et le duplex d'ADN (hp1). De plus, il est nécessaire de considérer des sélections sur billes magnétiques « nues » c'est-à-dire des sélections en l'absence d'oligonucléotides (ou d'ADN) afin de pouvoir quantifier et identifier les ligands présentant des interactions non-spécifiques avec les billes SA.

L'incubation (②) de la librairie sur les billes magnétiques a également été sujette à expérimentation et le buffer d'incubation en présence de 0,01% de détergent (Tween-20) s'est montré suffisamment efficace pour décrocher les ligands les moins affins. Également discuté en point ③, il est intéressant d'effectuer une manipulation en présence d'une concentration de détergent plus élevée afin de comparer, non pas l'*output copies* puisqu'ils amènent à une quantification similaire, mais les populations de ligands sélections.

Les différents types de lavages (③) permettent d'appliquer une stringence sur les tests de sélections. Il a été montré que des cycles de lavages à différentes températures (température ambiante, 50°C ou 70°C) permettent de faire varier les outputs copies au même titre que la présence ou non d'ions K⁺ au sein du tampon de sélection. Démultipliant les expérimentations, il a été décidé de maintenir stable la concentration en K⁺ tout au long des expérimentations et de ne faire varier que la température dans le but d'imposer des conditions drastiques au processus.

Les éluions (④) expérimentées au cours du travail méthodologique ont été d'origines variées puisqu'il est possible de retrouver des éluions en milieu organique (DMSO @ 65°C) ou aqueux (milliQ water @ 95°C), ainsi que des éluions en milieu détergent, salin ou tamponné à pH acide. L'expérimentation a montré que l'utilisation d'eau ultrapure à haute température permettait d'obtenir le meilleur taux de relargage et est donc la condition principale d'éluion.

Comme discuté auparavant, le milieu organique apparaît également un milieu adéquat pour le relargage de ligand à condition de perpétuer un lavage aqueux postérieurement. Il est donc intéressant de pouvoir différencier les populations de ligands relarguer par des éluants diamétralement opposés.

Enfin, différentes techniques de sélection successives (⑤) ont pu être effectuées afin de gagner en spécificité de ligands. La comparaison de populations issues de double-sélection et celles issues de contre-sélections peuvent apporter de véritables informations sur une éventuelle spécificité structurale des ligands présents au sein de la librairie.

À partir des données acquises lors de l'étude méthodologique des conditions de sélections, 29 sélections distinctes ont été menées et sont listées dans le tableau 12 ci-dessous :

Tableau 12 : Première campagne de sélections réalisées à partir de la DEL1.

Conditions de lavages décrites sous le tableau.

Nom	Duplicate	Cible	ratio O/L	Détails expérimentaux	Output copies (quantif. qPCR)
Sel01	Sel02	Myc	5:1	WASH1	5.91E+09/7.62E+09
Sel03		Myc	5:1	WASH1 @ [T-20]=0,1% v/v	1.29E+09
Sel04		Myc	5:1	WASH2	7.92E+08
Sel05		Myc	5:1	WASH3	3.16E+09
Sel06	Sel07	Myc	5:1	WASH4	1.41E+08/3.53E+08
Sel08		Myc	5:1	WASH4 @ Éluion: DMSO 65°C	6.45E+07
Sel09		Myc	5:1	WASH5	1.42E+8*
Sel10	Sel11	Myc/Myc	5:1	WASH1 @ double sélection	9.18E+06/1.55E+07
Sel12		Myc/Myc	5:1	WASH4 @ double sélection	2.86E+08*
Sel13	Sel14	Myc	0,5:1	WASH1	2.92E+06/3.79E+07
Sel15	Sel16	Streptavidine	5:1	WASH1	1.06E+08/2.44E+08
Sel17		scr26	5:1	WASH1	3.99E+07
Sel18		hp1	5:1	WASH1	2.4E+07
Sel19		hp1	5:1	WASH1 @ sans ADN génomique	6.55E+07
Sel20		Myc22	5:1	WASH4	1.03E+09
Sel21		Kit1	5:1	WASH1	8.01E+09
Sel22		Kit1	5:1	WASH3	1.19E+07
Sel23		22CTA	5:1	WASH1	2.62E+08
Sel24		22CTA	5:1	WASH3	1.78E+07
Sel25		wtTel26	5:1	WASH1	1.27E+09
Sel26		wtTel26	5:1	WASH3	1.08E+08
Sel27		Kit1/Myc	5:1	WASH1	3.05E+09
Sel28		wtTel26/Myc	5:1	WASH1	6.73E+09
Sel29				DEL1 (Librairie naïve) @ 1E+08	

WASH1	WASH2	WASH3	WASH4	WASH5
5 lavages (tampon de sélection : 100mM KCl, 1X PBS pH 7.4, 0.01% v/v Tween-20) @ température ambiante. Élué : 95°C dans l'eau ultrapure.	5 lavages (tampon de sélection) dont le 4 ^{ème} lavage a été effectué à 50°C. Élué : 95°C dans l'eau ultrapure.	5 lavages (tampon de sélection) dont le 3 ^{ème} et le 4 ^{ème} lavage ont été effectués à 50°C. Élué : 95°C dans l'eau ultrapure.	5 lavages (tampon de sélection) dont le 4 ^{ème} lavage a été effectué à 70°C. Élué : 95°C dans l'eau ultrapure.	5 lavages (tampon de sélection) dont le 3 ^{ème} lavage a été effectué à 70°C et le 4 ^{ème} lavage à 80°C. Élué : 95°C dans l'eau ultrapure.

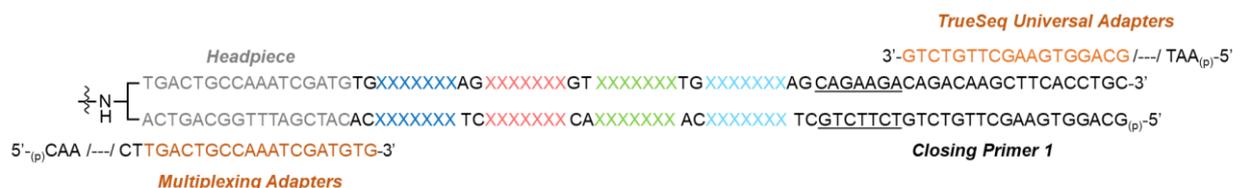
Comme il peut être constaté, l'impact des conditions de sélections a majoritairement été étudié sur la cible parallèle Myc. Cette cible G-quadruplex est un bon modèle dans le cas de la DEL1 puisque dans les conditions standards d'expérimentation le nombre de copies en sortie avoisine les 1E +10 copies.

Section 1.1.2 : Génération d'amplicons par PCR.

Après chaque sélection *in vitro*, les quantités de ligands présents au sein des éluats ont été mesurées grâce à la méthode de polymérisation quantitative en chaîne (en. *Quantitative Polymerase Chain Reactions* ou qPCR) telle que décrite au Chapitre 1 — Section 2.2.2, p77.

La conception de la DEL1 a été rationalisée de telle sorte que chaque ligand encodé de la librairie combinatoire ait la possibilité d'être amplifié simultanément. En effet, chacune des molécules de la librairie possède deux séquences code-barres communes que sont l'*Headpiece* et le *Closing Primer*. La présence de ces séquences communes permet ainsi l'utilisation d'amorce de PCR similaire pour l'amplification de chacune des séquences ADN (Figure 87). Pour rappel, la séquence terminatrice (*CP*) possède une séquence unique parmi laquelle 7 nucléobases (Figure 87, bases soulignées) informent sur l'identité de la librairie combinatoire. Dans le cadre de la DEL1, le *Closing Primer* associé est appelé *Closing Primer 1* ou *CP1*.

Parmi ces amorces, la première est une séquence oligonucléotidique appelée « Multiplexing Adapters » de 83 nucléotides de long. La particularité de cette amorce est qu'elle comporte deux parties distinctes : une partie complémentaire à l'*Headpiece* de chacune des molécules, et une partie comportant une séquence Illumina nécessaire à l'adhésion des amplicons sur la puce Illumina (sujet discuté en partie 2 de ce chapitre). La deuxième amorce utilisée est une amorce de 73 nucléotides appelée « TrueSeq1 Universal Adapter » qui comporte, similairement à la première amorce, une séquence complémentaire au *Closing Primer* de chacune des molécules de la librairie, et une partie adaptateur Illumina.



Multiplexing Adapters

5' CAA GCA GAA GAC GGC ATA CGA GAT **XXX XXX** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC
T TGA CTG CC AA ATC GAT GTG 3'

TrueSeq Universal Adapter

5' A ATG ATA CGG CGA CCA CCG AGA TCT CAC TTT CCC TAC ACG ACG CTC TTC CGA TCT **G CAG GTG**
AAG CTT GTC TG 3'

Figure 87 : Représentation générale de la séquence complète des code-barres ADN des membres de la DEL1 : de l'Headpiece au Closing primer-1. Hybridation des amorces Illumina au niveau du brin sens du CP1 et du brin anti-sens de l'HP.

Alors que l'adaptateur universel *TrueSeq* est utilisé de manière systématique pour l'amplification de chacun des éluats de sélection, l'adaptateur *Multiplexing* lui diffère en fonction des sélections. En effet, parmi les 83 nucléotides présents au sein de la séquence, un index de six nucléotides de long (encadré et grisé en Figure 87) permet de différencier chacune des expérimentations effectuées. Cela permet ainsi de mélanger les amplicons de chaque éluat de sélection afin d'effectuer le séquençage du mélange d'expérimentations.

Ainsi, 29 adaptateurs uniques *Multiplexing* ont été utilisés afin de pouvoir différencier les 29 conditions de sélections effectuées. L'indexage de chacune des expérimentations est représenté dans le tableau 13 ci-dessous.

Tableau 13 : Indexage unique de l'amorce Illumina Multiplexing en fonction de 29 sélections effectuées.

Nom	Indexage	Nom	Indexage	Nom	Indexage	Nom	Indexage
Sel01	CGTGAT	Sel09	TGCCGA	Sel17	ATCGTG	Sel25	GCCATG
Sel02	TACAAG	Sel10	CGTACG	Sel18	TGAGTG	Sel26	ATTCCG
Sel03	TTGACT	Sel11	CCACTC	Sel19	TCGGGA	Sel27	GTCGTC
Sel04	GGAACT	Sel12	CTTCGA	Sel20	GAATGA	Sel28	CGATTA
Sel05	GCGGAC	Sel13	ATCAGT	Sel21	TGTTGG	Sel29	GCTGTA
Sel06	TTTCAC	Sel14	GCTCAT	Sel22	GTATAG		
Sel07	GGCCAG	Sel15	TAGTTG	Sel23	AAAATG		
Sel08	CGAAAC	Sel16	CCGGTG	Sel24	AGCTAG		

La quantification des éluats de sélection a été effectuée séparément au sein d'une plaque 96 puits (en. 96- *well plate*, ou 96-WP). Chaque puits comporte alors une faible quantité d'ADN élue après sélection et tous les composants nécessaires à l'amplification par qPCR.

Les valeurs de quantifications des 28 sélections (ne prenant donc pas en compte la Sel29 qui est l'amplification de la DEL1 naïve, c'est-à-dire, sans sélection préalable) ont ainsi été calculées et sont reportées dans le Tableau 12, p139.

Les profils d'amplification des 28 sélections (28 éluats et la Sel29 amplifiée à 1E +08 copies) sont représentées en Figure 88 (A) et montrent une réelle disparité. En effet, trois populations distinctes peuvent être identifiées :

- une population atteignant leur plateau d'amplification au bout de 20 cycles (Figure 88, B/)

8 sélections / 29 sélections

- une population où l'amplification maximale est atteinte au bout de 23 cycles (Figure 88, C/)

12 sélections / 29 sélections

- et enfin, une population nécessitant 27 cycles d'amplification pour atteindre leur début de plateau (Figure 88, D/).

9 sélections / 29 sélections

À partir de ce constat, chacune des sélections a été réamplifiée du nombre de cycle permettant d'atteindre la phase stationnaire. Aboutir à ce maximum d'amplification garantit une quantité suffisante d'amplicons pour les étapes de purification et de séquençage tout en limitant l'amplification de séquences parasites. Dans l'intérêt de pouvoir séquencer la librairie naïve avec un fort taux de *reads* (voir Chapitre 3 — partie 2, p158), cinq lots de librairie naïve (Sel29) ont été amplifiés en parallèle (soit 5×10^8 copies).

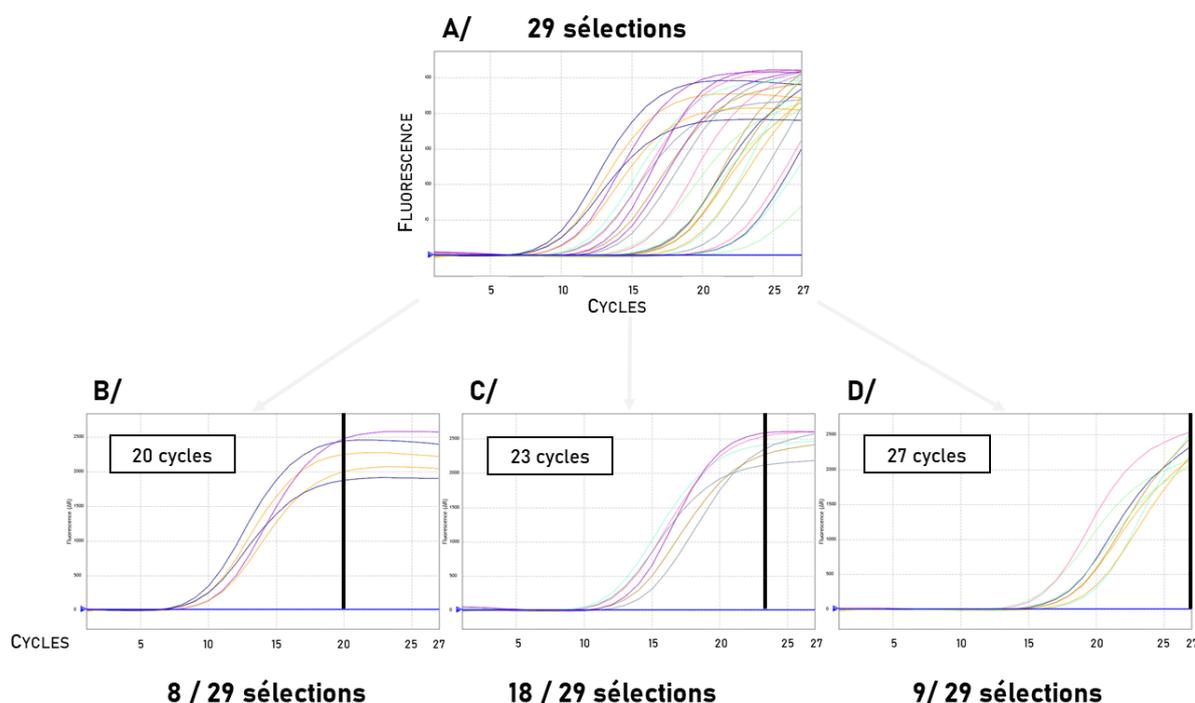


Figure 88 : A/ Amplification des 29 conditions de sélections sur 27 cycles. Classement des sélections en fonction du nombre de cycles nécessaires afin d'atteindre le plateau : B/ 20 cycles, C/ 23 cycles et D/ 27 cycles.

Afin d'obtenir un taux de séquençage similaire pour chacune des sélections, les amplicons de chaque condition de sélection sur cible ont été mélangés en quantité équimolaire. Seule la sélection 29, représentant la DEL1 naïve, a été ajoutée aux autres conditions de sélection en quantité supérieure (plus d'information dans la partie 2, séquençage à haut-débit, p158). Afin

d'effectuer un tel mélange, il est nécessaire de quantifier l'ADN présent après chaque amplification. Cette quantification est effectuée par électrophorèse sur gel d'Agarose.

L'évaluation de la quantité d'ADN présente au sein d'un puits de migration a été réalisée sur gel d'agarose, par comparaison de l'intensité des bandes observées avec celles d'une échelle d'ADN commercial à plage très faible (Figure 89). Également appelée GeneRuler Ultra Low Range (*ThermoFischer*), cette échelle permet la quantification d'ADN contenant entre 10 et 300 paires de bases.

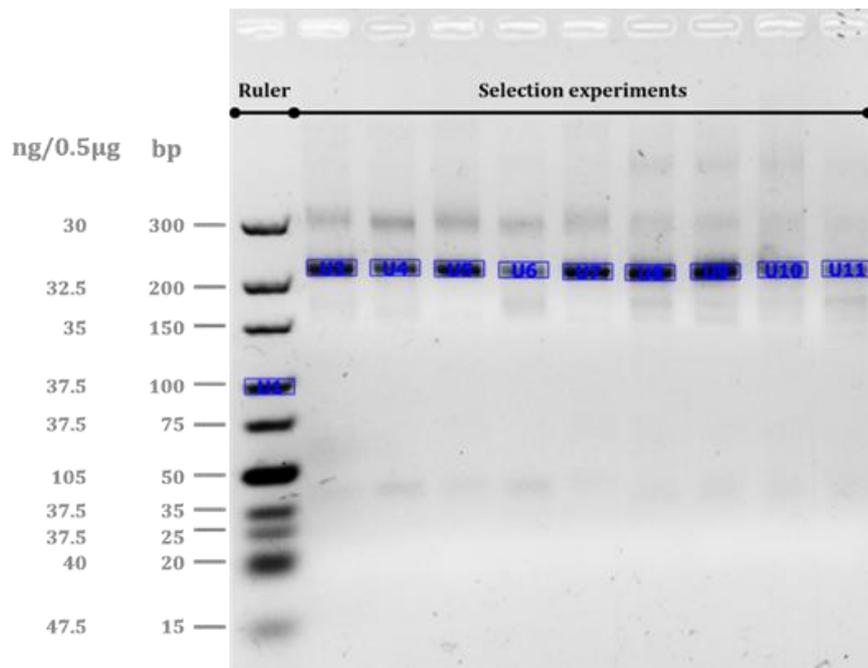


Figure 89 : Quantification massique d'ADN obtenu pour chaque condition de sélection après amplification par qPCR. Quantification réalisée sur gel d'agarose (UltraPure) à 5%. Migration à 90V/2 h.

La quantité massique d'ADN obtenue après amplification de chaque éluat de sélection a varié de 246ng pour la sélection présentant le nombre de copies le plus faible à 730ng pour l'éluat présentant un *output copies* élevé. Afin de rester en quantité équimolaire (ou équimassique), 246ng des 28 sélections (Sel01 à Sel28) ont été rassemblés afin de former un lot unique de 6400ng. Parallèlement, l'amplification 29 (bibliothèque naïve) a également été quantifiée et une masse de 2785ng a été obtenue.

La quantification des bandes par électrophorèse sur gel d'agarose a permis de constater que certaines impuretés subsistent en dessous mais surtout au-dessus des bandes d'intérêts (Figure 90). La pureté des amplicons est primordiale pour une expérience de séquençage à haut débit. Ainsi, afin de s'en affranchir, les amplicons du lot 1 (mélange des sélections 1 à 28) ainsi que les amplicons du lot 2 (amplicons de la sélection 29, bibliothèque naïve) ont été purifiés séparément sur gel d'agarose (Figure 90).

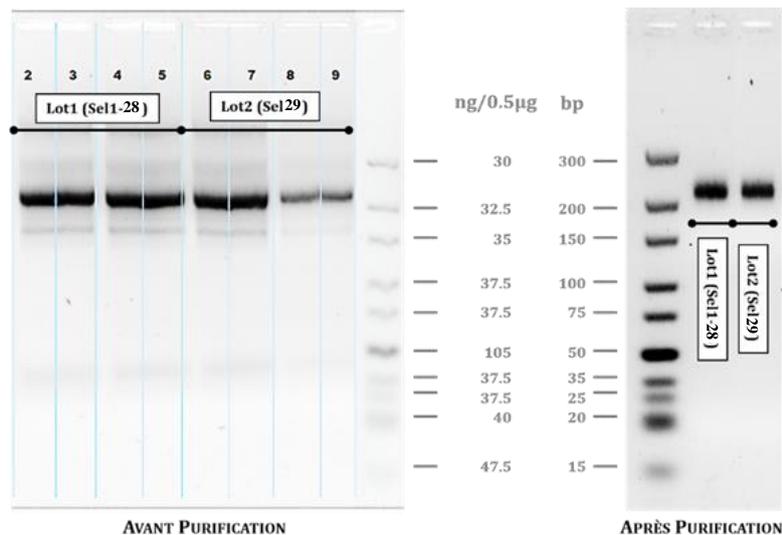


Figure 90 : Purification sur gel d'Agarose 5% (Ultrapure) du mélange équimassique des deux lots de sélections (lot 1 : Sel1 à 28 ; lot 2 : Sel29).

La purification a été réalisée par coupure physique de bandes d'agaroses contenant l'ADN à extraire en utilisant un système d'imagerie *Gel Doc* (Invitrogen). Les bandes d'agaroses ont alors été incubées dans une solution tampon (1M de Tris, 0.5M Na₂EDTA, 25mM NaCl) et mises à l'incubateur 16 h à 30°C afin que l'ADN puisse diffuser au sein du tampon d'incubation. Après filtration et concentration des solutions tampons contenant l'ADN diffusé par ultrafiltration à membrane (Amicon®, MerckMillipore) l'ADN purifié est récupéré en solution dans l'eau.

La quantité d'ADN obtenue pour les deux lots a une nouvelle fois été déterminée sur gel d'agarose afin d'obtenir avec précision la quantité finale d'amplicons pour les lots 1 et 2 (Figure 90). Au total, 1710ng du lot 1 (Sel1-28) et 706ng du lot 2 (Sel29) ont été obtenus. Afin d'obtenir le mélange final nécessaire au séquençage à haut débit, 857ng du lot 1 a été mis en mélange avec 343ng du lot 2 afin d'obtenir une quantité finale de 1200ng. Cette quantité, alors composée à 72% des résultats de sélections et à 28% de la librairie naïve (explication approfondie en partie 2, p159) a ensuite été envoyée pour séquençage à haut-débit en collaboration avec le Dr. Frédéric Ducongé (CEA, Paris).

Sous-partie 1.2 : DEL1 : deuxième campagne de sélections.

Section 1.2.1 : Processus de sélection in vitro.

Une deuxième campagne de sélections à partir de la DEL1 a été réalisée 18 mois après la première. Cette campagne intègre de nouvelles cibles mais avait également pour but de compléter les données obtenues lors des premiers jeux de sélections. Les choix de sélections s'appuient donc sur les résultats de séquençage obtenus dans le cadre de la première campagne de sélection. Néanmoins et pour plus de clarté, cette deuxième campagne de sélection sera décrite ici et l'analyse de toutes les sélections par HTS sera discutée plus loin dans le manuscrit (p162) au sein d'une partie commune.

Parmi les cibles G-quadruplex utilisées on peut retrouver certaines déjà présentes lors de la première campagne de sélection avec la séquence du promoteur d'oncogène Myc, ainsi que son dérivé monomorphique Myc22. Les séquences télomériques 22CTA et wtTel26, respectivement de topologie antiparallèle et hybride, ont également été expérimentées une seconde fois. Les nouvelles cibles G-quadruplex utilisées proviennent notamment de la séquence minisatellite humaine 25CEB, adoptant une topologie parallèle, mais également de la séquence G-quadruplex de topologie hybride retrouvée au sein de la séquence terminale longue répétée (en. *long terminal repeat sequence*, ou LTR) du VIH-1 (Virus de l'immunodéficience humaine)²⁹⁹, au niveau de la région U3 du provirus³⁰⁰.

Toutes les cibles G-quadruplex citées jusqu'à présent sont sous forme ADN. Toutefois, nous avons également utilisé dans cette campagne de sélection un G-quadruplex provenant du transcriptome humain au titre de la séquence répétitive TERRA (en. *telomeric repeat-containing RNA*), ARN non codant transcrit à partir de la région sous-télomérique capable de former une structure G-quadruplex de topologie parallèle⁹⁶.

Outre les structures secondaires capables de former des G-quadruplex, la DEL1 a également été incubée en présence d'autres structures d'acides nucléiques. En effet, des structures tétraplexes de cytosines, également connues sous le nom d'*i-motifs*, ont été utilisées comme cibles. Parmi celles-ci on retrouve la séquence i-motif MycPy27 formée au sein du promoteur d'oncogène Myc, mais également le i-motif de la séquence télomérique humaine hTelo-c. La séquence i-motif de la région promotrice de BCL-2 a également été considérée au même titre que la séquence i-motif DAP issue de la région promotrice du gène DAP (en. *Death-associated protein*). Enfin, un ADN à jonction triple (en. *3-way junctions*) a également été utilisée en tant que cible dans les processus de sélection *in vitro* par affinité. La librairie DEL1 a été principalement conçue pour la reconnaissance des G-quadruplex, les sélections réalisées contre les cibles i-motifs et TWJ ont donc peut de chance d'aboutir. Néanmoins, certaines associations moléculaires non favorables à la reconnaissance des G-quadruplex existent au sein du mélange combinatoire permettant d'envisager la découverte fortuite de pharmacophores affins de ces structures i-motifs et TWJ alternatives.

Tableau 14 : Séquences oligonucléotidiques (3' — biotine) utilisées dans le cadre de la deuxième campagne de sélection *in vitro* par affinité à partir de la DEL1.

Nom	Séquence	Description
Myc-3'biot	GGA GGG TGG GGA GGG TGG GGA A/3BioTEG/	G4 (parallèle), promoteur c-myc
Myc-C18-5'biot	/5Biosg//iSp18/GGA GGG TGG GGA GGG TGG GGA-A	G4 (parallèle), promoteur c-myc
Myc22-3'biot	TTG AGG GTG GGT AGG GTG GGT AA/3BioTEG/	G4 (parallèle), promoteur c-myc avec substitution de G en T, PDB: 1XAV
22CTA-3'biot	AGG GCT AGG GCT AGG GCT AGG GTT/3BioTEG/	G4 (antiparallèle), ADN télomérique humain (variant), PDB: 2KM3
wtTel26-3'biot	TTA GGG TTA GGG TTA GGG TTA GGG TT/3BioTEG/	G4 (hybride), ADN télomérique humain, PDB: 2JPZ

25CEB-3'biot	AAG GGT GGG TGT AAG TGT GGG TGG GTG/3BioTEG/	G4 (parallèle), minisatellite humain, PDB: 2LPW
HIVLTRIII-3'biot	GGG AGG CGT GGC CTG GGC GGG ACT GGG GA/3BioTEG/	G4 (hybride), virus de l'immuno déficience humaine, PDB: 6 h 1 K
TERRA-3'biot	rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrUrA rGrGrG rUrU/3BioTEG/	G4 ARN (parallèle), séquence télomérique répétée, PDB: 2M18
MycPy27-3'biot	CCT TCC CCA CCC TCC CCA CCC TCC CCA/3BioTEG/	i-motif, promoteur c-myc
hTelo c-3'— biot	TAA CCC TAA CCC TAA CCC TAA CCC TAA/3BioTEG/	i-motif, ADN télomérique humaine
BCL-2-3'biot	CAG CCC CGC TCC CGC CCC CTT CCT CCC GCG CCC GCC CCT/3BioTEG/	i-motif, promoteur BCL-2
DAP c-3'-biot	CCC CCG CCC CCG CCC CCG CCC CCG CCC CC/3BioTEG/	i-motif, région promotrice DAP
TWJ1-3'biot	ACT CTT CTC GTT TTT TCG AGA GCG ACT TTT TTG TCG CAG AGT/3BioTEG/	Jonction triple, Three-way junction

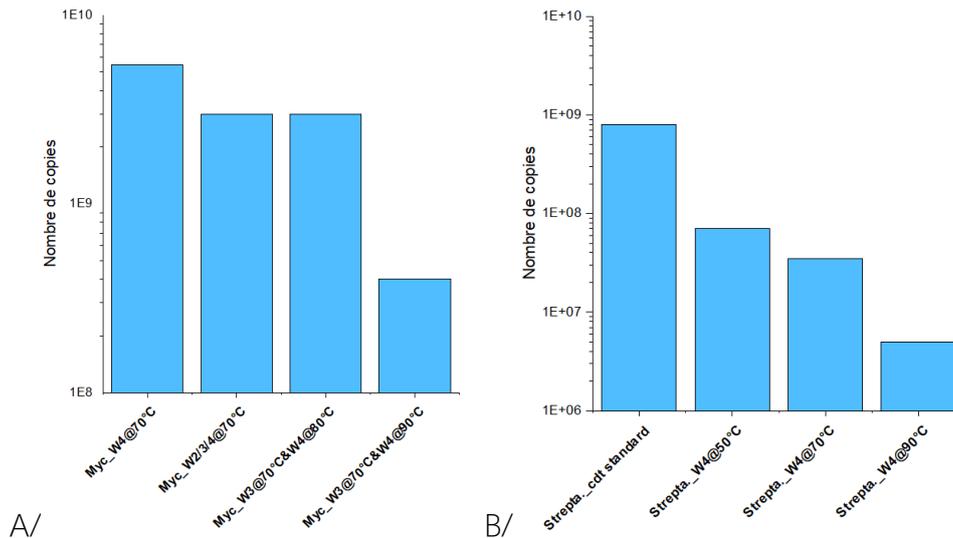
Les conditions standards de sélection *in vitro* décrites précédemment ont été globalement appliquées pour les nouvelles et anciennes cibles G-quadruplex (Tableau 14). Les variations apportées sont reportées ci-dessous. D'autre part, de nouvelles conditions standards ont été optimisées pour l'application de processus de sélection par affinité sur des cibles i-motifs.

[Paragraphe 1.2.1.a\) Nouvelles conditions appliquées au promoteur d'oncogène Myc.](#)

La cible G-quadruplex du promoteur d'oncogène Myc a particulièrement été utilisée afin de faire l'étude méthodologique de nouvelles conditions de stringences. Précédemment, la sélection de 5 pmoles de librairie sur 25 pmoles de la cible Myc immobilisée sur billes magnétiques permettait d'obtenir en sortie de sélection 7E +09 copies en conditions standards et 2E +08 copies en conditions stringentes (5 lavages dont le 4^{ième} appliqué à 70°C).

Lors de l'application de ces mêmes conditions stringentes de sélection (*i. e* lavage 4 @ 70°C) contre la cible Myc pour la nouvelle étude méthodologique, le nombre de copies obtenu était de 5.5E+09. Tous biais lors de la réitération de cette manipulation ont alors été étudiés et des valeurs constantes ont été obtenues en triplicatas. Il a alors été supposé que l'achat de nouvelles Dynabeads MyOne T1 SA était à l'origine de cette augmentation du nombre de copies. En effet, il est possible que le chargement en streptavidine des billes magnétiques utilisée lors de la première campagne de sélection était moindre que le nouveau lot utilisé pour la nouvelle campagne de sélection. Cet accroissement du nombre d'*output copies* après élution a alors été finement utilisé. En effet, le principal objectif lors d'expériences stringentes est de diminuer le nombre de copies d'élution sans jamais descendre en dessous du seuil limite de quantification du qPCR, qui est finalement l'amplification contrôle du NTC, d'environ 5E +06 copies.

Une gamme de stringence plus poussée a alors été étudiée avec des températures de lavages allant de 70 à 90°C (Graphique 6, A/).



Graphique 6 : Quantification du nombre de copies obtenues après renforcement des conditions de stringences. Exemples des cibles A/ Myc et B/ Streptavidine.

L'analyse des données de stringences obtenues montre que l'application d'un ou plusieurs lavages à 70°C ne permet pas de baisser le nombre de copies en sortie. Ajouter une condition de lavage à 80°C permet d'obtenir la même quantité d'*output copies* que dans le processus de lavages avec application de trois lavages à 70°C. Finalement, c'est en combinant un lavage à 70°C et un lavage à 90°C qu'il est possible d'obtenir une quantité de 4E +08 copies.

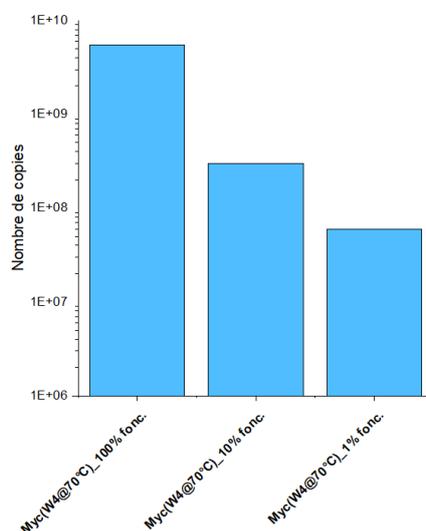
De manière similaire à la cible Myc, des expériences de stringences ont été réalisées sur les billes magnétiques SA dépourvues de séquence oligonucléotidiques afin d'étudier les ligands les plus affins des billes ou de la protéine streptavidine (Graphique 6, B/).

Il est alors possible de constater que plus les lavages appliqués sont effectués à haute température, plus le nombre de copies en sortie décroît. La quantification du nombre de copies obtenue pour la condition la plus stringente est néanmoins compromise puisque celle-ci est proche de la valeur d'amplification du contrôle NTC. Ainsi, le maximum de stringence permis pour une quantification fiable est la condition appliquant un lavage à 70°C. En supplément du contrôle streptavidine effectué à température ambiante lors de la première campagne de sélection, il est intéressant de comparer l'impact d'une condition stringente sur la population de ligands sélectionnés.

Également décrite dans le cadre de la première campagne de sélection, la stringence peut être appliquée par variation du ratio O/L (cf. Tableau 10, p131). Des ratios plus stringents de 0,5:1 et 0,05:1 avaient été appliqués en réduisant la quantité de billes saturées en cibles lors de l'incubation. Cependant cette approche conduisait à la manipulation d'une très faible quantité de billes obtenues par dilutions successives qui pouvaient donner lieu à des biais.

Toutefois, une autre méthode a été envisagée afin de réaliser des ratios de 0,5:1 et 0,05:1. Pour un ratio de 5:1, 6,25µL de billes SA sont fonctionnalisés à leur maximum en suivant le taux de chargement donné par le fournisseur (*ThermoFisher Scientific*, 400pmoles/1mL_{billes}).

Pour un ratio de 0,5:1, nous avons alors utilisé la même quantité de billes, mais seuls 10% des sites SA ont été fonctionnalisés par des oligonucléotides biotinylés. En suivant le même raisonnement, le ratio 0.05:1 a été obtenue en fonctionnalisant 1% des sites dispensés par 6,25 μ L de billes. L'évolution du nombre de copies en fonction du taux de fonctionnalisation des billes magnétiques par la cible G-quadruplex Myc est illustrée sur le Graphique 7.



Graphique 7 : Quantification du nombre de copies obtenues en fonction du taux de fonctionnalisation des billes magnétiques (1, 10 ou 100%).

Le taux de fonctionnalisation des billes magnétiques a en effet permis d'effectuer des expériences de stringences. Il est alors possible d'observer que moins les billes sont fonctionnalisées, plus le nombre de copies en sortie de sélection décroît. Toutefois, il est important de noter que la condition à ratio O/L 0.05:1 présente un nombre de copies bien supérieur à la limite de quantification par qPCR, à *contrario* de l'expérience réalisée lors de la première campagne de sélection.

Lors de toutes les sélections décrites précédemment, la fonctionnalisation des Dynabeads streptavidine a été effectuée par association d'oligonucléotides portant une fonction biotine en 3' de leurs séquences (nom commercial de la fonction : /3BioTEG/ ; Figure 91).

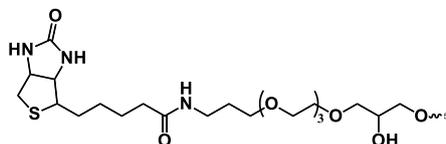


Figure 91 : Fonctionnalisation d'oligonucléotides par une biotine en 3' via l'espaceur triéthylène glycol (/3BioTEG/).

L'effet de la polarité de fixation de la fonction biotine a alors été questionné. Pour ce faire la séquence Myc portant une fonction biotine du côté 5' de l'oligonucléotide a été acquise. Au-delà du lieu de fixation de la biotine, la séquence Myc-5'biotine diffère de son homologue 3' par l'espaceur séparant la biotine de l'oligonucléotide. En effet, alors que l'espaceur du côté 3' porte un motif triéthylèneglycol, la séquence Myc-5'biotine finalement appelée Myc-C18-

5'biotin porte 6 motifs éthylèneglycol (Figure 92). La différence de longueur de chaîne de 11 atomes entre le 3'biotine et le C18-5'biotine a également été étudié lors de cette expérience.

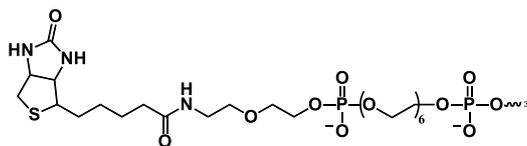


Figure 92 : Fonctionnalisation d'oligonucléotides par une biotine en 5' via l'espaceur hexa-éthylène glycol (/5Biosg/iSp18/).

Les sélections réalisées à partir des deux homologues biotinylés de Myc ont été effectuées dans les conditions standards, en appliquant une stringence lors du 4^{ième} lavage (70°C). Alors que l'éluât de sélection après sélection sur la cible Myc-3'biotine est de 5.4E+09 copies, 3.1E+09 copies ont été obtenues à partir de la cible Myc-C18-5'biotin. En termes d'*output copies*, très peu de différences sont donc à noter. Toutefois, il est intéressant de pouvoir comparer les deux populations de ligands sélectionnés entre-elles afin d'évaluer un éventuel effet de la polarité de la fixation voire de la longueur de l'espaceur.

Paragraphe 1.2.1.b) Expériences de compétitions.

Lors de la deuxième campagne de sélection, des études de compétitions plus approfondies ont également été mises en place afin de pouvoir 1) sélectionner les ligands les plus affins des structures quadruplex vis-à-vis d'ADN duplexes (et/ou du contrôle scr26) et 2) sélectionner d'éventuels ligands capables de différencier des structures G-quadruplex possédant différentes topologies.

Parmi les sélections réalisées lors de la première campagne de sélections, on peut retrouver des sélections faisant intervenir des cibles « contrôles » tels que la séquence riche en guanine incapable de former une structure G-quadruplex, scr26, ainsi que l'hp1 un duplexe d'ADN. L'intérêt d'avoir effectué de telles sélections a été de trouver, parmi les membres de la librairie, des motifs ou des molécules particulièrement affines pour les cibles contrôles.

Dans les conditions standards de sélection, 5 pmoles de DEL1 sont incubées avec la cible quadruplex en présence de 100eq d'ADN génomique. Dans le cadre d'expérience de compétition, il a été postulé qu'ajouter les séquences scr26 et ds26 (duplexe d'ADN ; 100 eq.) dans le milieu d'incubation permettraient de trier les composés présents au sein de la librairie.

Cette expérience de compétition a été réalisée sur la séquence Myc-3'biotine. Le référentiel de l'expérience de sélection stringente sur Myc (W4@70 °C) a permis d'obtenir 5.4E+09 copies après élution. Lorsque cette même expérience a été réalisée en mettant en jeu les séquences scr26 et ds26, 5E +09 copies ont été obtenues. La variation quasiment inexistante entre les deux processus de sélection laisse sous-entendre que les ligands les plus intéressants de la librairie n'ont pas, dans leur majorité, une affinité remarquable pour ces cibles contrôles.

Pour aller plus loin, une autre manière d'isoler une population de ligands ne ciblant que les structures quadruplex est de réaliser un processus séquentiel de sélection et contre-sélection.

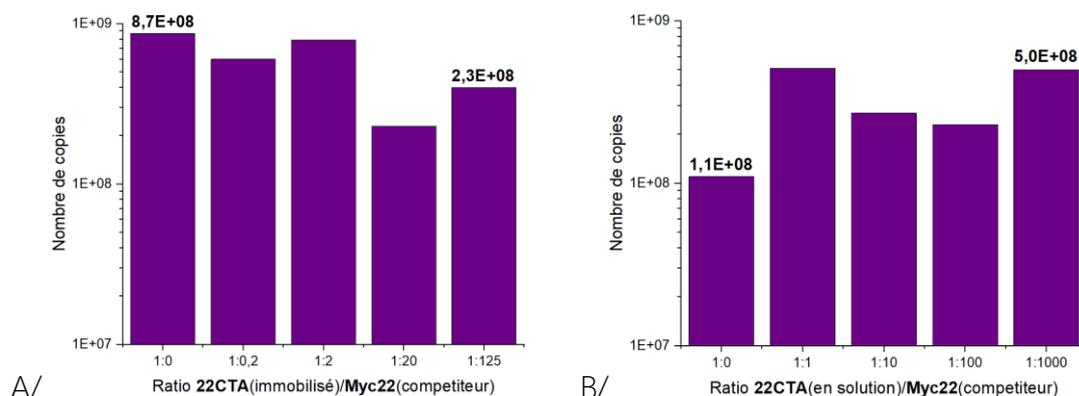
En effet, un premier processus de sélection *in vitro* standard est réalisé sur une cible G-quadruplex. L'éluat de cette première sélection, ne présentant que de meilleurs ligands vis-à-vis de la cible G-quadruplex, est dans un deuxième temps incorporé dans un nouveau cycle de sélection contre, cette fois-ci, une des cibles contrôles. Après incubation du mélange de ligands sur la cible contrôle, le surnageant est récupéré et rend compte des meilleurs ligands de la cible G-quadruplex dépourvus de tous ligands capables d'interagir fortement avec la cible contrôle.

Cette sélection en deux temps a été réalisée dans les conditions classiques de stringence (W4@70 °C) et rend compte d'un nombre de copies de $7E+09$. Une nouvelle fois, en comparaison à la condition stringente sur Myc dont la valeur obtenue est de $5,4E+09$, aucun écart remarquable en termes d'*output copies* n'a pas été observable. Néanmoins, il n'est pas inintéressant de pouvoir comparer les populations sélectionnées par cette méthode de compétition.

Après des études de compétitions de ligands entre des structures G-quadruplex et des contrôles, la sélectivité des ligands de la librairie a été mise à l'épreuve en effectuant des expériences de compétitions sur support solide ou en solution. En effet, la cinétique d'interaction entre un ligand et sa cible peut varier en fonction de la mobilité de la cible.

Parmi les expériences de compétitions réalisées entre différentes topologies de G-quadruplex il est possible de retrouver les couples 22CTA (antiparallèle) /Myc22 (parallèle), TERRA (ARN antiparallèle)/ wtTel26 (antiparallèle) et enfin Myc22 (parallèle)/22CTA (antiparallèle).

Les compétitions sur support solides ont été réalisées similairement aux compétitions avec les contrôles, c'est-à-dire en ajoutant au milieu d'incubation une quantité fixe du G-quadruplex compétiteur. Dans le cas du couple 22CTA/Myc22, une étude sur la quantité de compétiteurs à ajouter a été réalisée (Graphique 8, A). À partir des conditions standards (5 pmoles DEL1, 25 pmoles 22CTA-immobilisées) ont été rajoutés 5 à 2500 pmoles de Myc22, correspondant de 0,2 à 125 équivalents en compétiteur.



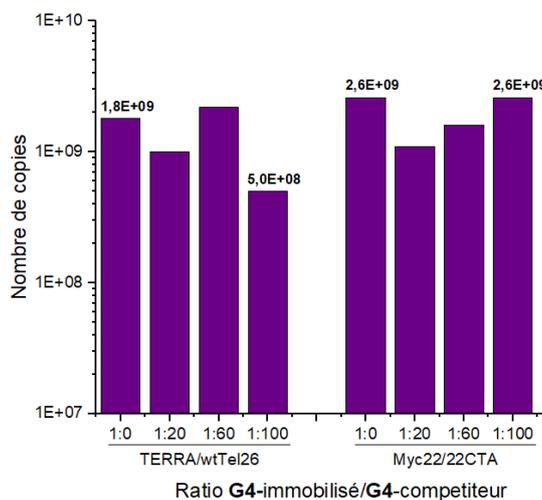
Graphique 8 : Quantification du nombre de copies obtenues en fonction de la quantité de compétiteurs. Expériences réalisées A/ par immobilisation de la cible sur billes magnétiques ou B/ en solution dans des conditions de stringentes (W4@50 °C).

Le nombre d'outputs copies est diminué en présence du compétiteur Myc22 suggérant la fixation d'une partie des ligands potentiels de 22CTA sur cette cible. Néanmoins, la baisse n'est pas directement proportionnelle à la quantité de compétiteurs ajoutée.

Les mêmes expérimentations ont été effectuées à partir de la séquence 22CTA-3'biotine en solution. Dans ces expériences en solution, la cible biotinylée ainsi que le compétiteur non-biotinylé sont incubés en présence de la librairie combinatoire encodée. Une fois l'incubation achevée, le milieu est mis en présence des billes Dynabeads MyOne T1 SA afin d'accrocher sur les billes la cible quadruplex biotinylée. Une fois cette étape réalisée, le processus suit les conditions classiques de lavages et d'éluion.

Dans le cas du mélange 22CTA-3'biotine/Myc22 en solution, une gamme de compétiteurs allant d'un mélange 1:1 à un mélange 1:1000 a été étudiée (Graphique 8, B/). La quantité de DEL1 utilisée par condition de sélection a été de 5 pmoles et la quantité de 22CTA-3'biotine de 10 pmoles. En effet, afin d'effectuer une gamme jusqu'à 1000eq, baisser la quantité de G-quadruplex biotinylée de départ a permis d'utiliser moins de quantité de compétiteurs. Comme le montre l'expérimentation, les variations des *output copies* sont du même ordre de grandeur, entre 1E+08 et 5E+08 copies. Les conditions de compétitions sur support ou en solution entraînent approximativement le même nombre de copies. Il est donc intéressant de vérifier si les populations sélectionnées sont les mêmes par séquençage.

En termes de sélection compétitive, les couples TERRA/wtTel26 et Myc22/22CTA ont également été étudiés (Graphique 9).



Graphique 9 : Quantification du nombre de copies obtenues en fonction de la quantité de compétiteurs ajoutés pour le couple TERRA/wtTel26 (compétiteur) et Myc22/22CTA (compétiteur). Expériences réalisées en conditions stringentes (W4@70 °C).

De manière similaire aux expériences de compétitions décrites précédemment, l'étude du couple TERRA/wtTel26 montre une légère décroissance du nombre de copies en présence de 100 équivalents de wtTel26 dans le milieu d'incubation. La compétition entre Myc22 et

22CTA montre une plus grande stabilité dans le nombre de copies en sortie, même à un nombre d'équivalents très élevé.

Compte tenu des expériences de compétitions réalisées, il n'a pas été possible d'affirmer sur la base des quantifications un quelconque effet remarquable des compétiteurs sur les cibles immobilisées sur support ou les cibles biotinylées en solution. Malgré tout, il est intéressant de pouvoir comparer l'identité des composés sélectionnés après les différentes étapes de sélections en compétitions afin d'évaluer l'enrichissement de certains motifs ou la déplétion d'autres.

Paragraphe 1.2.1.c) Sélection sur i-motif.

Sans en faire le récit, il a été démontré³⁰¹ que des séquences corolaires aux G-quadruplex riches en cytosines étaient capables d'adopter une structure secondaire particulière appelée i-motif (Figure 93, A/). Cette structure tétraplexe possède quatre brins appariés deux à deux en sens opposés. La structure i-motif est formée par intercalation de paires de cytosines appariées grâce à trois liaisons hydrogène (Figure 93, B/). L'une d'entre elle n'est possible qu'après protonation du N(3) en milieu acide. Ainsi, les i-motifs sont des structures favorisées à pH acide.

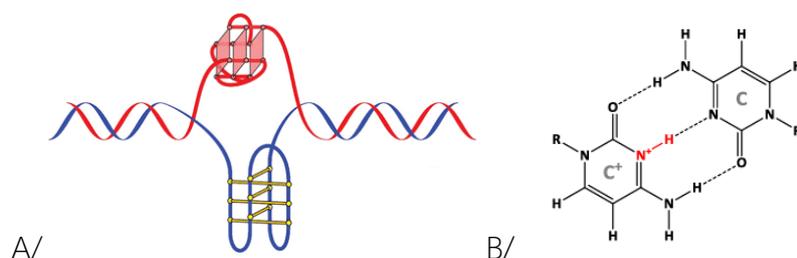


Figure 93 : A/ Représentation schématique des édifices G-quadruplex (rouge) et i-motif (bleu) — Image tirée de la publication de C.E Kaiser³⁰². B/ Formation des structures i-motifs possible en condition acide (dépendant de la séquence considérée).

Compte tenu de la stabilité de la structure tétraplexe i-motif à pH acide, les conditions de sélection standard telles que décrites pour les sélections de ligands de structure G-quadruplex ne sont pas applicable (100mM KCl, 1X PBS pH 7.4, 0.01% Tween-20). De nouvelles conditions pré-établies au laboratoire ont été utilisées. Le nouveau tampon i-motif contient 50mM de Tris base, 50mM de KCl, 35 mM de NaCl et le pH est ajusté avec de l'acide acétique (AcOH). L'étude de la stabilité des séquences i-motifs en fonction du pH du tampon i-motif ainsi qu'en fonction de la température a été effectuée par dichroïsme circulaire (CD ; Figure 94).

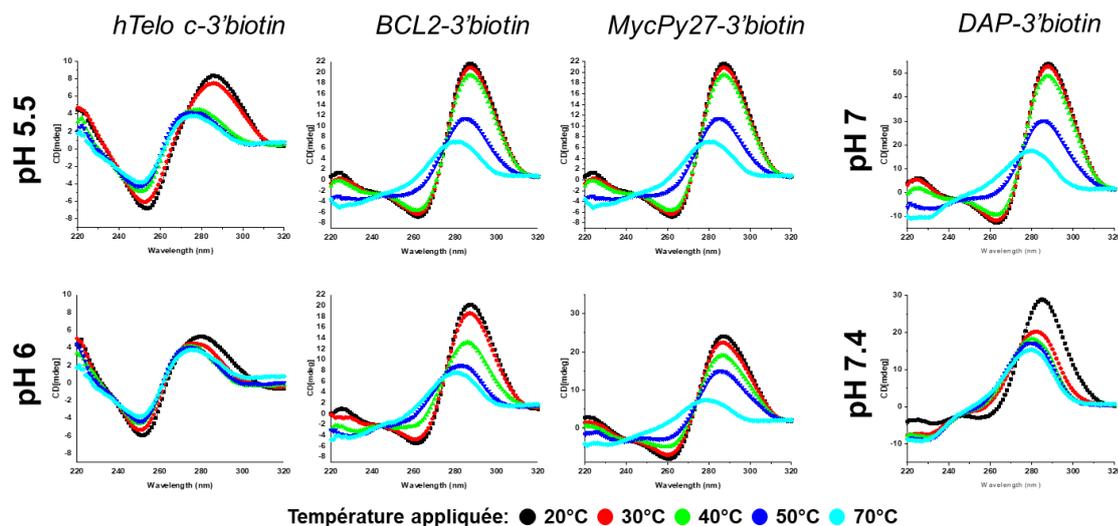


Figure 94 : Etude de la dissociation thermique des structures i-motif par dichroïsme circulaire en fonction de différentes valeurs de pH (5,5 à 7,4).

Le structure i-motif se distingue en CD par un maximum à 288nm et un minimum entre 255 et 260nm³⁰³. Le signal CD des séquences hTelo-c, BCL2 et MycPy27 ont été obtenus à deux valeurs de pH, 5,5 et 6. Dans le cas de la séquence i-motif télomérique un spectre caractéristique est observé à pH 5,5 (maximum à 288nm) alors qu'un maximum à 281nm est observable à pH 6 (Figure 94, courbes noires - 20 °C). Ce nouveau maximum est caractéristique de la formation d'une pelote aléatoire (en. *random coil*) d'une séquence oligonucléotidique³⁰⁴. Au contraire de la séquence i-motif télomérique, les séquences i-motifs BCL2 et MycPy27 présentent des longueurs d'onde maximales et minimales caractéristiques de la formation de la structure i-motif à pH 5.5 et à pH 6.

À l'inverse des autres i-motifs discutés jusqu'à présent, la séquence de la région promotrice du gène DAP est capable de former une séquence i-motif *in vitro* sous conditions physiologique³⁰⁵. Ainsi, la sensibilité du i-motif à la variation de pH a été étudiée entre 7 et 7,4. Dans les conditions à température ambiante, les maximums et minimums observés aux deux pH reflètent la formation de la structure i-motif. Ainsi, parmi toutes les séquences i-motifs considérées, seul le i-motif télomérique ne forme pas la bonne structuration à pH 6 et 20°C.

En effet, la séquence télomérique i-motif supporte des températures jusqu'à 30°C à pH 5.5 alors que les températures supérieures dénaturent la structure secondaire. Les séquences i-motifs BCL2 et MycPy27 ont des comportements stables jusqu'à 40°C quel que soit le pH considéré (Figure 94, courbes noires (20°C), rouges (30°C) et vertes (40°C)). Enfin, la séquence i-motif DAP reste stable à pH 7 jusqu'à 40°C, alors qu'elle subit une déstructuration dès 30°C à pH 7.4.

Tenant compte de toutes les expériences de stabilities des structures i-motifs choisies, les sélections sur ces différentes cibles ont été effectuées à pH 5.5 pour hTelo-c, BCL2 et MycPy27 alors que des sélections à pH 7 ou 7.4 ont été effectuées sur la cible i-motif DAP. L'application de stringences au cours des expériences de sélections a été effectuée par

variation de température lors des lavages. Ces températures ont été choisies suivant la stabilité obtenue dans les expériences de CD.

Au terme de cette deuxième campagne de sélections des structures G-quadruplex, mais également des i-motifs et une jonction triple ont été discutés et utilisés comme cible. Au total, 117 sélections uniques ont été menées. Dans le tableau ci-après est reportée la liste non-exhaustive des principales expériences de sélections menées jusqu'au séquençage (Tableau 15).

Tableau 15 : Deuxième campagne de sélections réalisées à partir de la DEL1.

Conditions de lavages décrites sous le tableau.

Nom	Cible	ratio O/L	Détail expérimentation	Output copies (quantif. qPCR)
Sel01	Streptavidine	5:1	WASH1	4.76E+08
Sel02	Streptavidine	5:1	WASH4	9.92E+06
Sel04	Myc	5:1	WASH4	5.08E+09
Sel05	Myc	5:1	WASH4 (+ Réincubation hp1-3'biotin)	1.33E+10
Sel11	Myc	5:1	WASH4 (ratio O/L: 0.5:1 – loading 10%)	1.34E+08
Sel12	Myc	5:1	WASH4 (ratio O/L: 0.05:1 – loading 1%)	1.71E+07
Sel15	Myc-C18	5:1	WASH4	8.32E+08
Sel22	22CTA	5:1	WASH2 (+ 20eq Myc22 / compétition)	7.31E+07
Sel23	22CTA	5:1	WASH2 (+100eq Myc22 / compétition)	2.71E+08
Sel24	22CTA (en solution)	0.5:1	WASH2 (+1000eq Myc22 / compétition)	2.88E+08
Sel25	22CTA (en solution)	2:1	WASH2 (+100 eq Myc22 / compétition)	5.91E+08
Sel26	CEB25	5:1	WASH4 (dont W3@70 °C)	2.55E+08
Sel27	CEB25	5:1	WASH5	3.35E+10
Sel28	HIV-LTRIII	5:1	WASH4 (dont W3@70 °C)	2.96E+08
Sel29	HIV-LTRIII	5:1	WASH5	1.62E+07
Sel30	TERRA	5:1	WASH4 (dont W3@70 °C)	9.87E+08
Sel31	TERRA	5:1	WASH5	8.12E+09
Sel32	TERRA	5:1	WASH4 (+60 eq wtTel26 / compétition)	9.33E+08
Sel33	TERRA	5:1	WASH4 (+100 eq wtTel26 / compétition)	1.42E+08
Sel34	Myc22	5:1	WASH4	1.03E+09
Sel35	Myc22	5:1	WASH4 (+60 eq 22CTA / compétition)	2.81E+09
Sel36	Myc22	5:1	WASH4 (+100 eq 22CTA / compétition)	2.10E+09
Sel37	TWJ-1	5:1	WASH2	2.79E+08
Sel38	TWJ-1	5:1	WASH4	6.57E+06
Sélections contre i-motifs (tampon i-motif)				
Sel39	hTelo-c	5:1	pH 5.5 _ WASH1	9.27E+06
Sel40	hTelo-c	5:1	pH 5.5 _ WASH2	6.52E+06
Sel41	MycPy27	5:1	pH 5.5 _ WASH1	1.66E+07
Sel42	MycPy27	5:1	pH 5.5 _ WASH2	4.86E+06
Sel43	BCL-2	5:1	pH 5.5 _ WASH1	6.83E+07
Sel44	BCL-2	5:1	pH 5.5 _ WASH2	2.62E+06
Sel45	DAP	5:1	pH 5.5 _ WASH2	1.75E+06
Sel46	DAP	5:1	pH 7 _ WASH2	1.54E+06
Sel47	DAP	5:1	pH 7.4 (tampon sélection G4)_ WASH2	2.84E+08
Sel48	DAP	5:1	pH 7.4 (tampon sélection G4)_ WASH4	3.89E+07

WASH1	WASH2	WASH3	WASH4	WASH5
5 lavages (tampon de sélection : 100mM KCl, 1X PBS pH 7.4, 0.01% v/v Tween-20) @ température ambiante. Élution : 95°C dans l'eau ultrapure.	5 lavages (tampon de sélection) dont le 4 ^{ème} lavage a été effectué à 50°C. Élution : 95°C dans l'eau ultrapure.	5 lavages (tampon de sélection) dont le 3 ^{ème} et le 4 ^{ème} lavage ont été effectués à 50°C. Élution : 95°C dans l'eau ultrapure.	5 lavages (tampon de sélection) dont le 4 ^{ème} lavage a été effectué à 70°C. Élution : 95°C dans l'eau ultrapure.	5 lavages (tampon de sélection) dont le 3 ^{ème} lavage a été effectué à 70°C et le 4 ^{ème} lavage à 80°C. Élution : 95°C dans l'eau ultrapure.

Section 1.2.2 : Génération d'amplicons par PCR.

La quantification et la formation d'amplicons de chacune des sélections ont été effectuées grâce à la méthode de polymérisation quantitative en chaîne (qPCR) similairement aux amplifications réalisées dans le cadre de la première campagne de sélection.

Les adaptateurs *Multiplexing* s'hybridant sur le brin anti-sens du code-barre ADN, au niveau de l'*Headpiece* sont les mêmes adaptateurs Illumina® indexés qu'utilisés lors des premières sélections sur la librairie DEL1. Contenant plus de sélection unique, plusieurs autres séquences indexées ont été utilisées (Tableau 16).

Tableau 16 : Indexage unique de l'amorce Illumina Multiplexing en fonction des sélections de la deuxième campagne.

Nom	Indexage	Nom	Indexage	Nom	Indexage	Nom	Indexage
Sel01	CGTGAT	Sel24	GCTACC	Sel33	CGCCTG	Sel42	CGATTA
Sel02	ACATCG	Sel25	ATCAGT	Sel34	GCCATG	Sel43	GCTGTA
Sel04	TGGTCA	Sel26	GCTCAT	Sel35	AAAATG	Sel44	ATT ATA
Sel05	CAGTGT	Sel27	AGGAAT	Sel36	TGTTGG	Sel45	GAATGA
Sel11	GTAGCC	Sel28	CTTTTG	Sel37	ATTCCG	Sel46	TCGGGA
Sel12	TACAAG	Sel29	TAGTTG	Sel38	AGCTAG	Sel47	CTTCGA
Sel15	TGACAT	Sel30	CCGGTG	Sel39	GTATAG	Sel48	TGCCGA
Sel22	CGTACG	Sel31	ATCGTG	Sel40	TCTGAG		
Sel23	CCACTC	Sel32	TGAGTG	Sel41	GTCGTC		

Comme indiqué dans le Tableau 15, les *output copies* des sélections effectuées varient de 1E+06 à 1E+10 copies. Le nombre de cycles d'amplification appliqué aux sélections uniques est directement dépendant de la quantification de la sélection. Pour un nombre de copies variant entre $1E+06 < X < 8E+07$, 29 cycles d'amplifications ont été appliqués. Lorsque le nombre de copies était supérieur à $8E+07$ et jusqu'à $6E+08$ copies, 26 cycles d'amplifications ont été effectués. Finalement, pour des valeurs supérieures à $6E+08$ copies, 22 cycles d'amplifications ont été réalisés. Le nombre de cycles choisit en fonction du nombre de copies obtenues et permet d'attendre la phase stationnaire lors de l'amplification.

Après amplification, la quantité d'amplicons obtenus a été déterminée par électrophorèse sur gel d'agarose en comparant l'intensité des bandes obtenue avec une échelle commerciale (Figure 95).

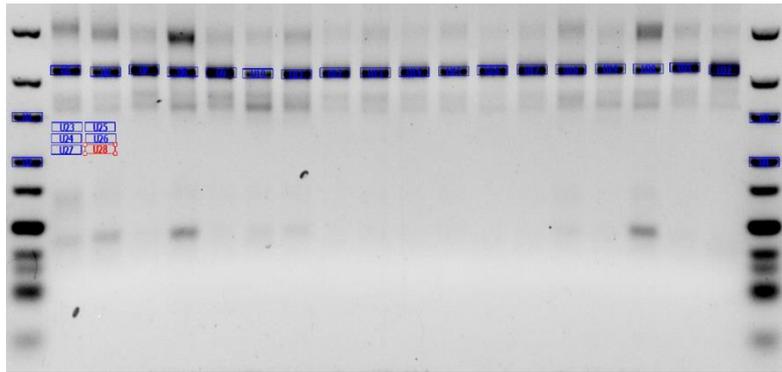


Figure 95 : Quantification massive d'ADN obtenu pour chaque condition de sélection après amplification par qPCR. Quantification réalisée sur gel d'agarose (UltraPure) à 5%. Migration à 90V/2 h.

L'analyse massive de tous les amplicons de sélections obtenus a montré une valeur minimum de 133ng. Ainsi, 133ng de chacune des 48 sélections a été assemblé afin d'obtenir 6384ng d'amplicons. La purification mécanique du mélange d'amplicons a alors été réalisée après migration par découpe des bandes d'intérêts (Figure 96).

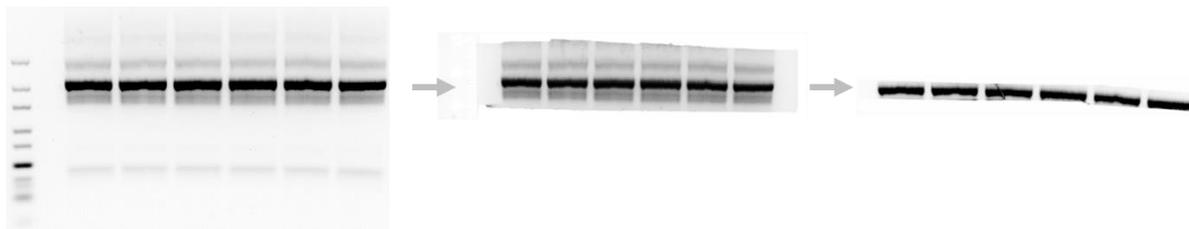


Figure 96 : Migration des quantités équimassiques de sélections (90V/2hrs/5% Agarose UltraPure) suivie de la séparation physique des bandes d'intérêts.

Les amplicons ont été extraits de l'agarose par simple diffusion au sein d'un tampon TE-NaCl (1M de Tris, 0.5M Na₂EDTA, 25mM NaCl) après incubation à 30°C pendant 16 h. L'ADN récupéré a ensuite été filtré sur membrane (Amicon®, MerckMillipore) avant d'être requantifié sur gel d'agarose. Au final, 691ng d'amplicons purifiés ont été obtenus et 300ng ont été envoyés au Dr. Ducongé pour séquençage.

Sous-partie 1.3 : DEL2 : campagne de sélections.

À partir de la librairie combinatoire encodée de deuxième génération et force des résultats des deux campagnes de sélections effectuées avec la DEL1 sur différentes cibles d'acides nucléiques, la DEL2 a été dans un premier temps sélectionnée vis-à-vis des cibles G-quadruplex Myc22 et 22CTA. En effet, ces cibles de topologies variées ont été utilisées comme référence pour des expériences de stringences, de doubles sélections, mais également de compétitions. En parallèle, des études moins poussées ont été réalisées sur les promoteurs d'oncogènes Myc et kit1, la séquence télomérique humaine wtTel26, la séquence minisatellite humaine CEB25 et sur la séquence G-quadruplex du promoteur LTR-III du virus HIV. Enfin, des sélections ont également été réalisées avec la séquence G-quadruplex ARN TERRA et les autres motifs d'acides nucléiques utilisés précédemment à savoir la séquence

i-motif DAP et la jonction triple TWJ1 (voir les séquences tableau 14, p145). Le tableau 17 ci-dessous reprend les sélections qui ont été réalisées.

Tableau 17 : Première campagne de sélections réalisées à partir de la DEL21.

Conditions de lavages décrites sous le tableau.

Nom	Duplicate	Cible	ratio O/L	Détails expérimentaux	
Sel01	Sel02	Streptavidine	5:1	WASH2	
Sel03	Sel04	Hairpin	5:1	WASH2 @ sans ADN génomique	
Sel05		Hairpin	5:1	WASH4	
Sel06		Myc22	5:1	WASH2	
Sel07		Myc22	5:1	WASH4	
Sel08		Myc22	5:1	WASH5	
Sel09		Myc22/Myc22	5:1	WASH1	
Sel10		Myc22	5:1	WASH2 (+1000pmoles 22CTA / <i>compétition</i>)	
Sel11		Myc22	5:1	WASH4 (+1000pmoles 22CTA / <i>compétition</i>)	
Sel12		22CTA	5:1	WASH2	
Sel13		22CTA	5:1	WASH4	
Sel14		22CTA	5:1	WASH5	
Sel15		22CTA/22CTA	5:1	WASH1	
Sel16		22CTA	5:1	WASH2 (+1000pmoles Myc22 / <i>compétition</i>)	
Sel17		22CTA	5:1	WASH4 (+1000pmoles Myc22 / <i>compétition</i>)	
Sel18		Myc	5:1	WASH2	
Sel19		Myc	5:1	WASH4	
Sel20		Kit1	5:1	WASH2	
Sel21		Kit1	5:1	WASH4	
Sel22		CEB25	5:1	WASH2	
Sel23		CEB25	5:1	WASH4	
Sel24		wtTel26	5:1	WASH2	
Sel25		wtTel26	5:1	WASH4	
Sel26		HIV-LTRIII	5:1	WASH2	
Sel27		HIV-LTRIII	5:1	WASH4	
Sel28		TERRA	5:1	WASH2	
Sel29		TERRA	5:1	WASH4	
Sel30		DAP	5:1	WASH2	
Sel31		DAP	5:1	WASH4	
Sel32		TWJ-1	5:1	WASH2	
Sel33		TWJ-1	5:1	WASH4	
Sel34	Sel35	DEL2 (Librairie naïve) @ 1E+08/1E+06			
WASH1 5 lavages (tampon de sélection : 100mM KCl, 1X PBS pH 7.4, 0.01% v/v Tween-20) @ température ambiante. Elution : 95°C dans l'eau ultrapure.		WASH2 5 lavages (tampon de sélection) dont le 4 ^{ème} lavage a été effectué à 50°C. Elution : 95°C dans l'eau ultrapure.		WASH3 5 lavages (tampon de sélection) dont le 3 ^{ème} et le 4 ^{ème} lavage ont été effectués à 50°C. Elution : 95°C dans l'eau ultrapure.	
		WASH4 5 lavages (tampon de sélection) dont le 4 ^{ème} lavage a été effectué à 70°C. Elution : 95°C dans l'eau ultrapure.		WASH5 5 lavages (tampon de sélection) dont le 3 ^{ème} lavage a été effectuée à 70°C et le 4 ^{ème} lavage à 80°C. Elution : 95°C dans l'eau ultrapure.	

Partie 2 : Analyse des sélections réalisées par séquençage à haut-débit.

Alors que les technologies de séquençages à haut-débit se sont particulièrement améliorées ces dernières années, le coût de séquençage de banques d'ADN a lui grandement baissé. Cette baisse a permis à des institutions académiques de pouvoir profiter de nouvelles technologies de pyroséquençage (Roche/454) ou de séquençage par synthèse (Illumina) (voir Chapitre 1, section 2.2.3, p80). Dans le cadre du séquençage de bibliothèques chimiques encodées, le choix de la technologie d'encodage va principalement dépendre de la longueur des fragments d'ADN à séquencer ainsi que du nombre d'évènements uniques à séquencer (on appellera ces évènements des *reads*). La comparaison entre les deux méthodes de séquençage a été évaluée en 2016²³⁵ et la technologie Illumina semble être une méthode universelle de séquençage, quelles que soient les caractéristiques de la bibliothèque (taille de la bibliothèque, longueur des fragments, nombre de *reads* souhaités).

Sans rentrer une nouvelle fois dans les détails de la technologie Illumina, il est bon de rappeler que le séquençage par synthèse est basé sur l'amplification de brins d'ADN groupés sous forme de clusters. La formation de ces clusters individuels est réalisée sur la surface d'une chambre de réaction, c'est-à-dire la cellule d'écoulement (en. *flow cell*) à huit voies. Chaque groupe d'ADN, provenant d'une séquence ADN individuelle, peut être analysé par l'incorporation séquentielle de nucléotides marqués par fluorescence, suivie de procédures de balayage itératives (séquençage d'ADN par synthèse). Comme une bonne résolution de balayage spatial de clusters ne peut pas être obtenue si tous les clusters incorporent la même base à la fois (comme c'est le cas lorsqu'une région constante d'une DEL est séquencée), un ADN génomique aléatoire, à l'exemple du *PhiX Control v3*, doit être ajouté aux amplicons à analyser afin d'obtenir des résultats de séquençage optimaux.

En utilisant le séquenceur Illumina NextSeq 500, 2,1 à 2.5 milliards de *reads* peut être obtenu par *flow cell* en lecture unique. Étant donné qu'une *flow cell* contient huit voies distinctes, une voie permet de lire environ 280 millions de *reads*. En raison de l'ajout de 10% d'ADN génomiques aléatoires *PhiX*, 250 millions de *reads* peut-être alloués, par ligne de *flow cell*, afin de séquencer les amplicons considérés. Tous les séquençages à haut débit effectués dans le cadre de cette thèse ont été réalisés sur une voie de cellule d'écoulement, c'est-à-dire sur un total d'environ 280 millions de *reads*.

Sous-partie 2.1 : Traitement bio-informatique de la première campagne de sélection.

La compilation des données de séquençage a été effectuée par le Dr. Frédéric Ducongé (CEA, DRF, JACOB, MIRCEn, UMR CNRS 9199, Université Paris-Saclay, Fontenay-aux-Roses, France) sur le logiciel PATTERNITY.seq[®]. Ce logiciel, développé par la plateforme MIRCEn de l'institut de biologie François Jacob, a été développé dans le but premier d'améliorer la sélection d'aptamères par séquençage à haut débit. Également parfaitement adapté au séquençage d'amplicons issu de processus de sélections dans le cadre de DEL, le logiciel PATTERNITY.seq[®] peut analyser des millions de séquences afin d'identifier rapidement les composés les plus enrichis.

L'échantillon de la première campagne de sélection envoyé en séquençage contient les amplicons de librairie naïve (DEL1 sans sélection) ainsi que les amplicons de toutes les 28 expériences de sélections réalisées. Toutefois, tel que décrit en conclusion de la partie 1.1 (p144), l'échantillon séquençé a comporté 28% de librairie naïve et 72% d'un mélange équimolaire des 28 sélections. Les pourcentages alloués au séquençage de la librairie naïve ainsi qu'au séquençage des éluats de sélection correspondent respectivement à ~40 et ~100 millions de *reads*.

En effet, de fait que les 190 440 séquences uniques de la DEL1 aient été séquençés pour la première fois, une très grande couverture (en. *coverage*) d'évènements a été souhaitée. L'intérêt d'une telle couverture est de pouvoir déterminer avec un grand indice de confiance la quantité de chacun des ADN uniques (et donc molécules) composant la librairie combinatoire. Ainsi, les 100 millions de *reads* alloués aux amplicons des 28 sélections auraient permis d'obtenir des paquets d'évènements d'environ 3.5-4 millions. En réalité, l'analyse des données de séquençage a montré que le nombre de *reads* alloués à la librairie naïve était de 60 millions et que 150 millions d'évènements ont été recensés pour les amplicons de sélections (soit ~5M *reads*/sélection). Cette évolution du nombre d'évènements attendus dépend de la quantité d'ADN génomiques *PhiX* ajoutés à la banque de données, mais également à la fonctionnalisation de la *flow cell* puisque celle-ci peut avoir une couverture comprise entre 2,1 à 2.5 milliards, soit de ~260 à 312 millions de *reads* par voie de cellule d'écoulement.

Section 2.1.1 : Analyse de la librairie naïve.

L'analyse des 60 millions d'évènements séquençés dans le cadre de la librairie naïve a permis de distinguer deux populations distinctes. La première d'entre elle est une population d'évènements connus (en. *known*) représentant à elle seule 33.15 millions de *reads* soit 55,25% des évènements totaux. Dans un deuxième temps, il a été possible de distinguer une population d'évènements inconnus (en. *unknown*) correspondant à 26.85 millions de *reads* (44,75%).

Les évènements connus sont les séquences d'ADN encodant les molécules chimiques de la librairie combinatoire. *A contrario*, les évènements *unknown* sont des séquences ADN ou l'enchaînement des bases ne correspond pas aux étiquettes des composés chimiques de la librairie. En effet, l'analyse des évènements est effectuée grâce aux 43 nucléotides (composés des code-barres ADN T1, T2, T3 et T4 ainsi que l'index du *Closing Primer*, Figure 97) présents au niveau du brin supérieur des étiquettes d'ADN.



Figure 97 : Structure générale du brin sens des code-barres ADN de la DEL1.

Dès lors qu'un code-barre ADN (T1 à T4) séquencé ne correspond pas à l'étiquette d'un *BB* de la DEL1, l'évènement décrit est considéré comme *unknown*. La possibilité de faire face à la présence de séquences *unknown* lors du séquençage peut être d'origines variées.

Premièrement, il est possible de faire face à des mutations de séquences d'ADN lors des amplifications *post-sélection* due à l'ADN polymérase Taq. Néanmoins, l'ADN polymérase utilisée est reconnue pour avoir un faible taux d'erreurs ($1.5E-04 \pm 0.2E-04^{306}$), n'expliquant ainsi pas le nombre élevé d'évènements inconnus. C'est alors posé la question de l'effet des agents chimiques utilisés lors de l'assemblage de la librairie (EDC, DMT-MM, Cu^I) sur l'amplification des séquences d'ADN en mélange. Afin de répondre à cette question, plusieurs conditions chimiques ont été appliquées sur une séquence nucléotidique commerciale, appelée *HPtoCP* mimant les étiquettes ADN de la DEL1 (Figure 98).

<i>HP</i>	T1-2	T2-1	T3-2	T4-2	<i>CP1</i>
TGACTGCCAAATCGATGTG	GACTAGGAG	GTTGCCTGT	CTCCTGCTG	GACTAGGAC	CAGAAGACAGACAAGCTTCACCTGC
ACTGACGGTTTAGCTACAC	CTGATCCAC	CAACGGACA	GAGGACGAC	CTGATCCTC	GTCTTCTGTCTGTTCGAAGTGGACG

Figure 98 : Séquence « *HPtoCP* » comportant l'*Headpiece* et le *Closing primer-1* ainsi que les code-barres issus de la DEL1 : T1-2, T2-1, T3-1 et T4-2.

Les *HPtoCP* ayant subi les conditions chimiques ont alors été amplifiées et séquencées en parallèle de la DEL1 naïve. Le séquençage a été effectué sur un appareil de routine permettant d'obtenir en peu de temps et sur un faible nombre de *reads* (~150k) l'analyse de tous les évènements. Les résultats de séquençage ont pu montrer que les chimies appliquées dans le cadre de l'assemblage de la DEL1 n'avaient aucun impact sur l'amplification de l'ADN. En effet, le taux de *reads* connu lors du séquençage a dépassé les 92% pour chaque condition appliquée. Par ailleurs, le taux de read connu obtenu lors de séquençage de la librairie naïve sur ce même instrument a été de 80% suggérant une composante instrumentale au fort pourcentage d'*unknown* observé lors de ce premier séquençage illumina.

Parmi les autres évènements identifiés au sein de la librairie naïve, il a été possible de constater la surreprésentation de deux séquences ADN dont le cumul d'évènements uniques dépassait à eux seul 1,5% du total de *reads* (soit environ 900k). Ces deux séquences ADN sont issues d'une librairie modèle développée par le Dr. Yang SI au début du projet de librairie encodée. Cette contamination n'est pas négligeable et illustre l'importance de la rigueur dans le traitement de matériel amplifiable.

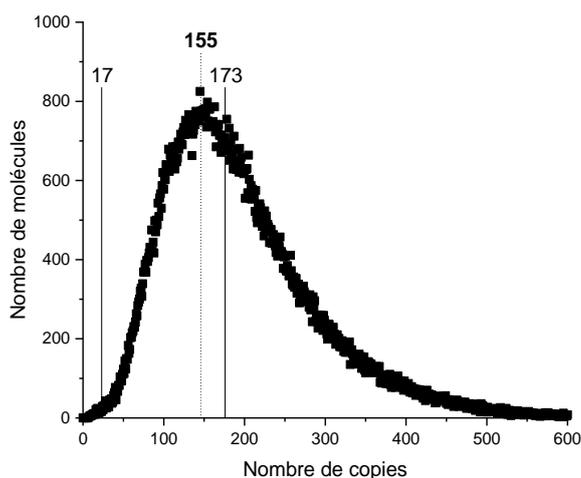
Néanmoins, les 33,2 millions d'évènements connus ont permis d'obtenir un *coverage* largement suffisant pour analyser avec grande confiance toutes les séquences ADN présentes au sein de la DEL1. La confiance des données d'une molécule unique va dépendre de deux paramètres principaux, sa fréquence (calculé par PATTERNITY.seq[®]) et son coefficient de variations (ou CV %) calculé à partir de l'Équation 1 ci-dessous :

$$CV(\%) = \frac{\sigma}{\mu} \times 100$$

Équation 1 : Coefficient de variation calculé à partir de l'écart-type des fréquences (σ) divisé par la moyenne des fréquences (μ).

Dans des conditions idéales de conception de bibliothèques encodées et de séquençage à haute affinité, le nombre de copies (ou la fréquence) de chaque motif chimique devrait être équivalent. À l'exemple de la DEL1 contenant 190 440 composés uniques, chaque composé est théoriquement représenté 173 fois pour un total de 33.2 millions de *reads*, soit une fréquence de 0.0005%.

Lorsqu'on représente la répartition des molécules en fonction de leur nombre de copies il est possible d'apercevoir la formation d'une courbe de Gauss (Graphique 10). Au maximum d'amplitude est représenté le nombre total de molécules ayant une fréquence de 0.00046%, soit 155 copies. La valeur moyenne théorique du nombre de copies/composé est ainsi en accord avec le maximum de la courbe de Gauss.



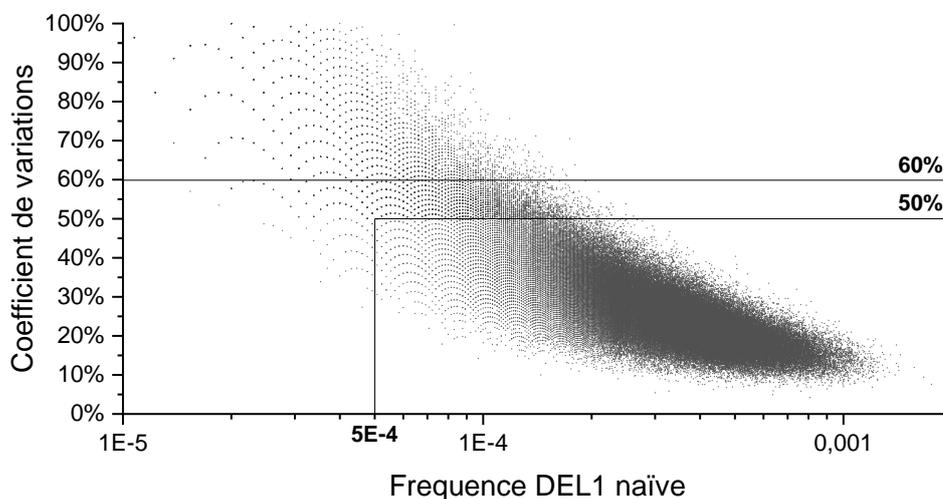
Graphique 10 : Représentation du nombre de molécules en fonction du nombre de copies obtenues pour chaque composé. Valeur moyenne à 173 copies/composés. Amplitude maximale de la sigmoïde à 155 copies/composés.

Parmi l'analyse de tous les événements connus de la bibliothèque naïve, 187 766 séquences ADN différentes ont pu être identifiées sur les 190 440 motifs possibles, soit 98,6% de reconnaissance. Le *coverage* de 60 millions de *reads* (dont 55,25% de séquences *known*) a ainsi permis d'analyser individuellement la fréquence de chacune des 187 766 séquences ADN présentes ainsi que leurs coefficients de corrélations (CV %). La courbe de Gauss présente deux extrêmes dont le minimum atteint un nombre de copies plus de 10 fois inférieur à la valeur théorique moyenne. Il a alors été raisonné que les composés à fréquences très faibles (< 1E-05) ont pu être obtenus dû à des difficultés chimiques. Leur présence au sein de la bibliothèque naïve pourrait avoir un impact lors de l'interprétation des résultats des 28 conditions de sélections car l'obtention de résultats d'enrichissement démesurés pourrait subvenir et

engendrer de faux-positifs. Ainsi, seules les séquences à fréquence $\geq 1E-05$ ont été considérées. D'autre part, l'évaluation des valeurs de CV est également importante car elle reflète la variabilité de la fréquence obtenue pour un motif unique. Pour exemple, si on considère une molécule unique avec une fréquence de 0.0002, cela implique que la molécule est présente dans le milieu à hauteur de 120 copies. Si la valeur du coefficient de variation calculé est de 20%, alors la molécule est présente à 120 ± 24 copies dans le mélange. Il est donc évident qu'il est nécessaire de limiter les hautes variations afin d'avoir un intervalle de confiance très élevé.

Si on représente le coefficient de variation en fonction des fréquences de toutes les séquences ADN retrouvées dans la librairie naïve (Graphique 11), il est possible de remarquer la grande dispersion de tous les membres de la librairie. Néanmoins, plus de 90% des séquences uniques se retrouvent au-dessus d'une fréquence de $1E-04$ et en dessous d'un CV de 50%. Pour l'analyse des relations structures-enrichissements pour chaque fragment (BB) de la librairie, la fréquence a été fixée à des valeurs minimums de $1E-05$ et un CV a 60%. Sous l'application de ces conditions, 180 015 séquences uniques ont été considérées (94,5% des séquences de la DEL1).

Il est également à noter que lors des analyses des composés uniques (en. *single molecule*), pour simplifier le processus individualisé et augmenter l'indice de confiance, seules les séquences présentant des fréquences supérieures à $5E-04$ et des CV inférieurs à 50% ont été considérées.



Graphique 11 : Représentation du Coefficient de variation de chaque composé de la DEL1 en fonction de sa fréquence dans la librairie naïve. Valeurs bornées pour un $CV \leq 50\%$ et une fréquence $\geq 5E-04$ (300 copies).

Section 2.1.2 : Analyse des motifs après sélections.

L'échantillon envoyé en séquençage a contenu 72% d'amplicons issus de processus de sélections contre différentes cibles. Le nombre de *reads* alloué au séquençage des événements des 28 sélections est d'environ 150 millions, permettant ainsi d'obtenir une

couverture d'environ 5 millions de *reads* par sélection unique. En moyenne, 64,6% des évènements constatés pour chaque sélection proviennent de séquences connues au sein de la librairie combinatoire ce qui a permis d'obtenir un *coverage* d'évènements connus d'environ 3.3 millions de *reads* par sélection.

Le Dr. Ducongé a utilisé le logiciel PATTERNITY.seq[®] pour identifier, à partir du processus de séquençage, la fréquence de chaque évènement unique au sein d'une sélection donnée. L'objectif a ensuite été de comparer la fréquence d'un évènement au sein de la librairie naïve à sa nouvelle fréquence après un processus de sélection afin de déterminer son facteur d'enrichissement (en. *Enrichment Fold*, EF). Ceux-ci ont été effectués sur un tableur (Excel, *Microsoft Office*) en collaboration avec le Dr. Frédérique Ducongé.

À partir des fréquences obtenues en fonction de chacune des sélections réalisées (librairie naïve comprise) il a alors été possible de calculer le ratio normalisé (en. *normalized ratio*, RN) relatif à la librairie naïve (R=0) de chaque composé *n* appartenant à la DEL1 (notée B) après un processus de sélection (R=*x*). Ce RN_{*n,x*} est ainsi défini par l'Équation 2.

$$RN_{n,x} = \frac{r_{n,x/0}}{\sum_{n \in B} r_{n,x/0}}$$

Équation 2 : Calcul du Ratio normalisé (RN) relatif à la librairie naïve d'un composé *n* dans une sélection (R= *x*).

Les ratios *r* relatifs à la librairie naïve des *n* composés étant eux calculés à partir des fréquences respectives obtenues avant (R=0) et après (R= *x*) sélection, voir Équation 3.

$$r_{n,x/0} = \frac{f_{n,R=x}}{f_{n,R=0}}$$

Équation 3 : Calcul du ratio (*r*) d'un composé *n* après sélection (R= *x*). *f*_{*n,R=x*} décrivant la fréquence du composé *n* après sélection *x* et *f*_{*n,R=0*} la fréquence du composé *n* au sein de la librairie naïve.

La valeur des RN_{*n,x*} peut d'ores et déjà orienter l'interprétation des résultats. En effet, si le résultat trouvé est supérieur à la valeur de 1, le composé *n* dans la sélection *x* est enrichi. A contrario, si le résultat est inférieur à la valeur de 1, le composé *n* est déplété lors du processus de sélection *x*.

L'enrichissement est caractérisable lorsqu'un évènement est relativement davantage représenté après un processus de sélection que dans la librairie naïve. Au contraire, la déplétion caractérise un évènement proportionnellement moins représenté après un processus de sélection.

Quel que soit la sélection donnée considérée, la somme des RN_{*n,x*} est égal à 100(%). Ainsi, le calcul de l'enrichissement (en. *enrichment*, E, en %) de chaque composé *n* au sein d'une sélection (R=*x*) est calculé à partir du nombre de composés uniques *N* composant la sélection, voir Équation 4.

$$E_{n,x} = \frac{RN_{n,x} - 1/N}{1/N} \times 100$$

Équation 4 : Enrichissement (E) d'un composé n après sélection ($R=x$) calculé à partir du $RN_{n,x}$ et du nombre de molécules N composant la librairie après sélection.

Enfin, le facteur d'enrichissement (EF) de chaque composé n après processus de sélection ($R=x$) peut être calculé suivant l'Équation 5 suivante :

$$EF_{n,x} = \frac{RN_{n,x} \times (E_{n,x} \times RN_{n,x})}{RN_{n,x}}$$

Équation 5 : Facteur d'enrichissement (EF) d'un composé n après sélection ($R=x$) calculé à partir du $RN_{n,x}$ et de $E_{n,x}$.

La série de calcul détaillée précédemment permet de calculer les facteurs d'enrichissements de chacun des événements distincts d'une sélection donnée. De plus, la même méthode de calcul a pu être utilisée afin d'obtenir les données d'enrichissements de chacun des *BB1*, 2, 3 et 4 représentés au sein de la librairie.

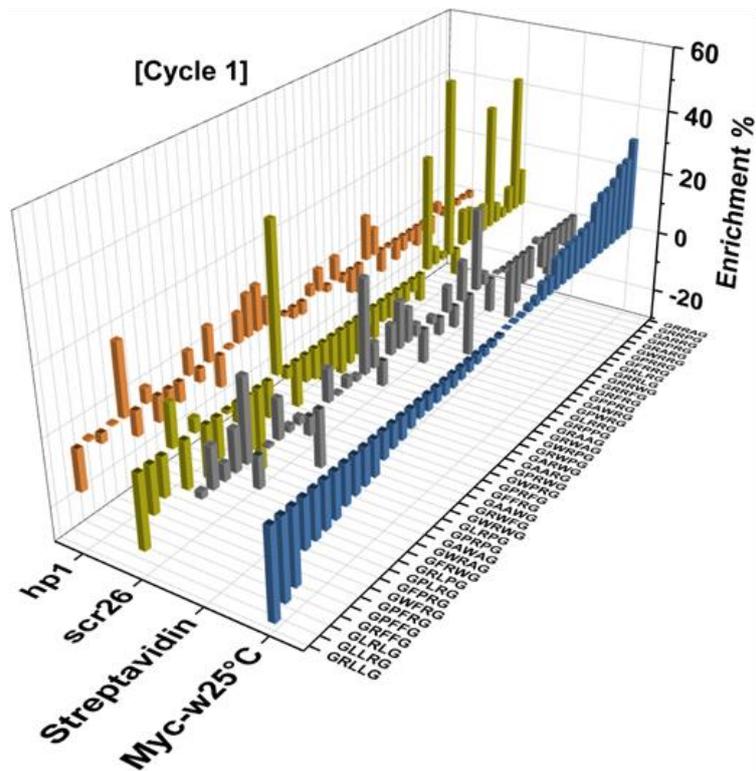
En effet, l'étude des *BB* a permis d'avoir un aperçu général des comportements des différents motifs contre les cibles quadruplex, mais également vis-à-vis des cibles contrôles.

Paragraphe 2.1.2.a) Myc versus contrôles.

L'analyse de la fréquence des blocs de construction du premier cycle de la librairie combinatoire, *BB1*, au sein de la librairie naïve montre des disparités d'incorporation. En effet, des fréquences normalisées de 0,99 à 3,40% ont été obtenues pour une valeur théorique moyenne de 2,17% par *BB1*. L'incorporation des premières briques de construction par couplage peptidique sur l'*HPT1* a été caractérisée par UPLC-MS et a montré des disparités dans les rendements de couplages. De manière similaire, ces mêmes disparités ont pu être retrouvées par séquençage.

L'enrichissement (ou la déplétion) des blocs de construction après un processus de sélection peut être représenté sous la forme d'un histogramme 3D où l'axe des abscisses déploie tous les *BB1*, l'axe des ordonnées les différentes conditions de sélections effectuées, et l'axe des altitudes l'enrichissement calculé sous forme de pourcentage.

Le Graphique 12 ci-dessous permet de comparer l'enrichissement des *BB1* dans les conditions de sélections réalisées contre la cible Myc ainsi que celles réalisées sur les cibles contrôles, à savoir la streptavidine (billes nues), le scr26 ainsi que l'hp1. Il est à noter qu'à l'exception de la condition scr26, les données représentées sont des données moyennées d'un duplicata de sélections.

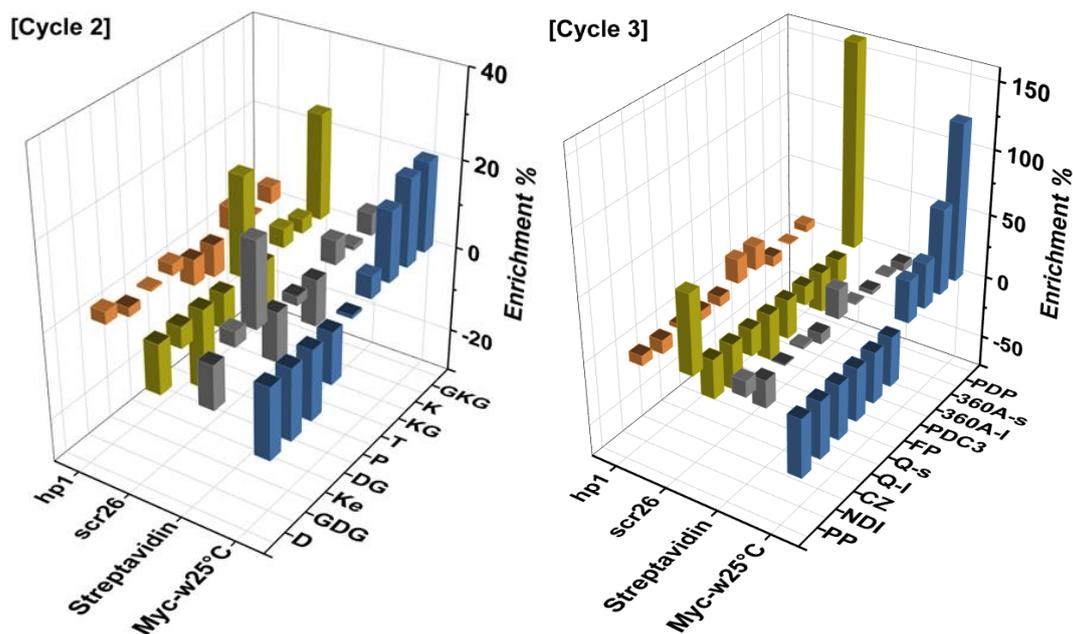


Graphique 12 : Enrichissement (%) des 46 pentapeptides DEL1 — [Cycle 1] en fonction des cibles hp1, scr26, Streptavidine et Myc. Valeurs obtenues en condition de sélection WASH1 (25°C). N=2 (exception de scr26).

Parmi les 46 séquences pentapeptidiques représentant les *BB1*, seules 22 séquences sont enrichies (valeur > 0%) contre la cible de topologie parallèle Myc. Si l'on regarde de plus près les séquences enrichies, il est possible de constater que les motifs doublement arginilés sont majoritairement représentés (20/22 séquences) et que ceux-ci sont jusqu'à 30% plus représentés après sélection. L'arginine est un acide aminé présentant un point isoélectrique de 10,76. Étant donné que les expériences de sélections ont été réalisées dans un tampon à pH de 7,4, les arginines sont alors chargées positivement au sein du mélange. Compte tenu de la nature des liaisons phosphodiester de la séquence oligonucléotidique de Myc, il est alors intelligible que par attraction de charges les séquences les plus arginilées soient celles les plus enrichies dans le mélange.

Lorsque sont comparés les enrichissements des mêmes *BB1* contre les cibles contrôles, il est d'ores et déjà possible de remarquer une réelle disparité dans les enrichissements. Contre les billes magnétiques dépourvues de cible oligonucléotidique, c'est-à-dire exhibant les protéines streptavidines à leurs surfaces, seuls quelques acides aminés ont montré des enrichissements supérieurs à 20% à l'exemple de GPWRG, GRWFG ou encore GPFFG. À l'instar de la cible G-quadruplex Myc, aucune séquence doublement arginilée n'a été enrichie. Contre la cible contrôle scr26, qui est la séquence riche en guanines ne présentant pas la capacité de former une structure G-quadruplex, il est possible de remarquer l'enrichissement de pratiquement toutes les séquences portant deux arginines mais également de quelques séquences hydrophobes ponctuelles comme le GRFFG ou encore le GPRPG enrichi à environ 50%. Enfin, contre la cible duplex hp1, les séquences enrichies sont des séquences essentiellement

hydrophobes (GPFFG, GAAWG, etc) alors que les séquences les plus polaires sont totalement déplétées.



Graphique 13 : Enrichissement (%) des 9 cœurs trifonctionnels DEL1 — [Cycle 2] (gauche) et des 10 ligands de structures G-quadruplex DEL1 — [Cycle 3] (droite) en fonction des cibles hp1, scr26, Streptavidine et Myc. Valeurs obtenues en condition de sélection WASH1 (25°C). N=2 (exception de scr26).

Les blocs de construction du deuxième et troisième cycle, représentent respectivement les cœurs trifonctionnels (BB2) ainsi que les motifs hétéroaromatiques (BB3). La fréquence calculée pour chacun des motifs du [Cycle 2] (Graphique 13, gauche) au sein de la librairie naïve a oscillé entre 10,35% à 12,68% avec une valeur minimum de 8,86% pour le motif D (Fmoc-Dap-OH). En comparaison à la valeur moyenne normalisée de 11,11% par motifs, la plupart des cœurs ont été couplés de manière similaire sur les pentapeptides du [Cycle 1].

Dans le cas des BB3 (Graphique 13, droite), le pourcentage de présence de chaque motif au sein de la librairie naïve devrait être de 10%. Le calcul des fréquences normalisées réelles a montré des variations de fréquences allant de 8,03% pour le ligand Furopyridazinone (FP) à 11,85% pour le Naphtalène diimide (NDI). Malgré des difficultés de validation d'incorporation des BB3 en UPLC-MS, il semblerait que l'incorporation des BB3 par chimie click sur les cœurs du [Cycle 2] aient été réalisées avec des rendements comparables.

L'analyse des données d'enrichissement des BB2 après sélection sur la cible Myc a montré le comportement singulier des trois cœurs lysines (K, KG, GKG). En effet, seuls ces trois BB ont été significativement enrichis à hauteur de 15 à 20%. Comparativement aux autres cœurs trifonctionnels de la librairie, les dérivés Lysines sont les seuls motifs capables de déployer une longue chaîne latérale de 5 atomes. Toutefois, le motif Ke (Lysine étendue par un acide azidoacétique) présente une chaîne latérale de 8 atomes et n'apparaît pourtant pas comme enrichie contre la cible quadruplex. Une hypothèse préliminaire suggérerait que la lysine déploie une longueur de chaîne latérale optimale pour orienter les ligands sur le promoteur

Myc G-quadruplex. La comparaison des enrichissements des *BB2* sur les sélections contrôles ont une nouvelle fois montré des enrichissements ponctuels de certains motifs. Cependant, aucune réelle tendance n'a été notable. Parmi les motifs enrichis dans les conditions de sélection sur les billes streptavidines, seul le cœur Ke a présenté un enrichissement de l'ordre de 20%. Contre la séquence scr26, il est intéressant de noter que tous les dérivés de lysines sont enrichis avec un maximum d'enrichissement pour le dérivé GKG à hauteur de 24%. De plus, le cœur trifonctionnel P (proline) est 22% plus présent après la sélection contre la séquence scr26 que dans la librairie naïve. Cette donnée est intéressante d'autant que la proline est un acide aminé naturel capable d'induire une structuration particulière à la séquence qui la compose. Enfin, quelques motifs uniques sont enrichis contre l'hp1 à l'exemple de KG ou DG mais les taux d'enrichissements ne dépassent guère les 5%.

S'en suit l'analyse des données d'enrichissement des 10 motifs hétérocycliques formant les *BB3*. Pour rappel, les *BB3* peuvent être classés en deux catégories (Figure 99), i) les motifs à systèmes aromatiques fusionnés (NDI, Q-s/Q-l, CZ, FP et PDC3) et ii) les motifs à systèmes aromatiques non fusionnés (PDP, 360A-s/360A-l). Ces motifs possèdent des affinités inhérentes pour les structures G-quadruplex, mais les constantes d'affinités diffèrent. Parmi ces structures, les motifs PDC3, PDP, 360A ont été décrits dans la littérature comme des ligands particulièrement affins des G-quadruplex, PDC3 en particulier. Lorsque l'on étudie le taux d'enrichissement des *BB3* après une sélection contre le promoteur d'oncogène Myc, il est possible d'apercevoir que tous les ligands à systèmes aromatiques non fusionnés sont enrichis, de 35% pour le dérivé pyridocarboxamide à chaîne longue (360A-l) à 122% pour le ligand pyridostatine (PDP). De plus, lorsqu'on regarde l'enrichissement individuel de chacun des ligands catégorisés à systèmes aromatiques fusionnés, tous ont été déplétés à l'exception du ligand phénanthroline (PDC3) enrichi à près de 33%.

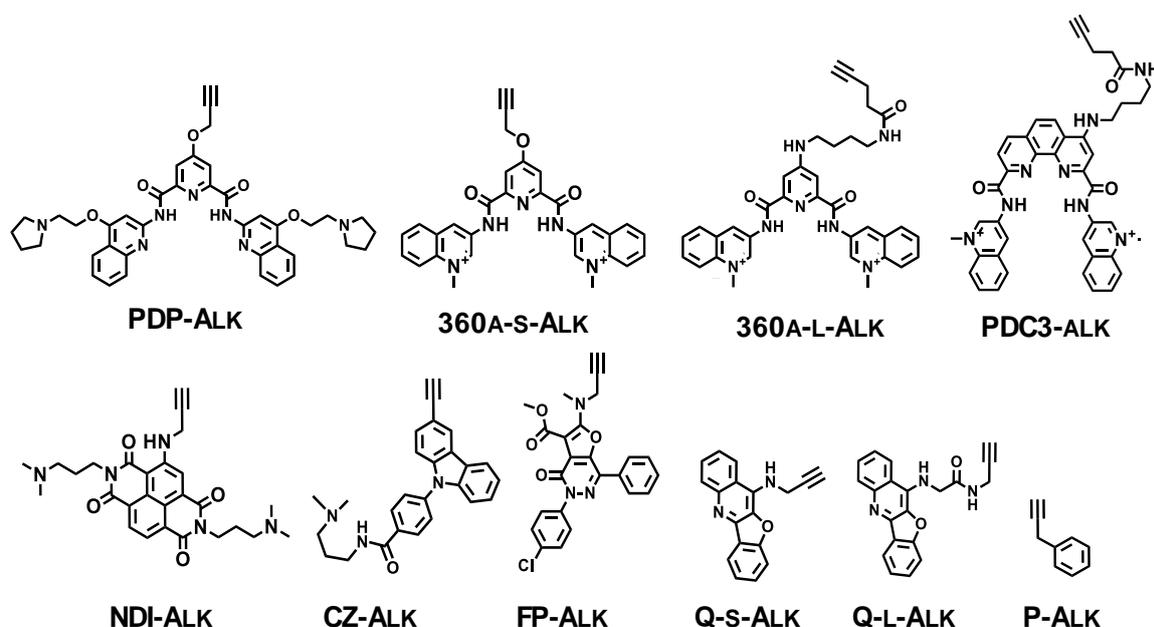


Figure 99 : Représentation exhaustive des 10 ligands formant les Building Blocks de la DEL1 — [Cycle 3].

L'analyse structurale des motifs hétéroaromatiques les plus présents après sélection par affinité sur la cible Myc montre que l'enrichissement semble être dépendant des motifs aromatiques, fusionnés ou non, dérivatisés par un motif quinoline (PDP) ou quinolinium (360A-s/-I, PDC3). En effet, la présence de ces motifs au sein des structures étend l'envergure et l'hydrophobicité du ligand. De plus, les dérivatisations permettent à la molécule d'avoir une flexibilité structurale que les ligands plans ne peuvent pas proposer.

Dans le cadre de la sélection contre la cible streptavidine, aucun réel enrichissement n'est notable. Seuls les ligands Furopyridazinone (FP) et Phénylpropyne (PP, contrôle négatif des BB3) ont été enrichis à hauteur d'environ de 23% et 12% respectivement.

À l'inverse de la cible streptavidine ne présentant pas d'enrichissements significatifs, la sélection contre le contrôle scr26 engendre un enrichissement des ligands Pyridostatine (PDP) et Phénylpropyne (PP) à hauteur de 160% et 62% respectivement. Il est intéressant de constater que le ligand PDP soit fortement enrichi contre une séquence riche en G ne formant pas des G-quadruplex. Le ligand PDS (ou son dérivé pyrrolidine alors nommé PDP) est un ligand connu des structures G-quadruplex. En revanche, l'interaction de ce ligand avec les cibles est toujours controversée car ce ligand est capable de se fixer sur la tétrade externe des G-quadruplexes³⁰⁷ mais également au niveau des sillons selon des stœchiométries de 2:1³⁰⁸. Il est alors envisageable que la séquence scr26 forme une structure favorable pour une interaction avec le ligand PDS.

L'évaluation du taux d'enrichissement des BB3 après sélection sur la séquence hp1 fait apparaître des enrichissements faibles de l'ordre de 19% pour le ligand pyridocarboxamide à chaîne longue (360A-I) ainsi que la phénanthroline (PDC3).

La dernière dérivatisation apportée à la DEL1 (BB4) est l'incorporation de la même diversité insérée lors de la première étape, c'est-à-dire les 46 séquences pentapeptidiques. Ces derniers présentent un taux de présence normalisé au sein de la librairie naïve de 0,8% à 4,4%, avec une valeur moyenne théorique normalisée de 2,17% par composé. Une nouvelle fois, la disparité de taux de présence des motifs est née de la variation de rendements obtenus lors des réactions de couplages peptidiques. L'évaluation des données d'enrichissement des BB4 après sélection contre la cible Myc montre un enrichissement exacerbé pour les pentapeptides composés de 2 acides aminés arginines (Graphique 14). Parmi les 14 séquences enrichies, 9 d'entre elles adressent le motif RR. A *contrario*, les motifs les plus déplétés sont les séquences peptidiques à caractères hydrophobes.

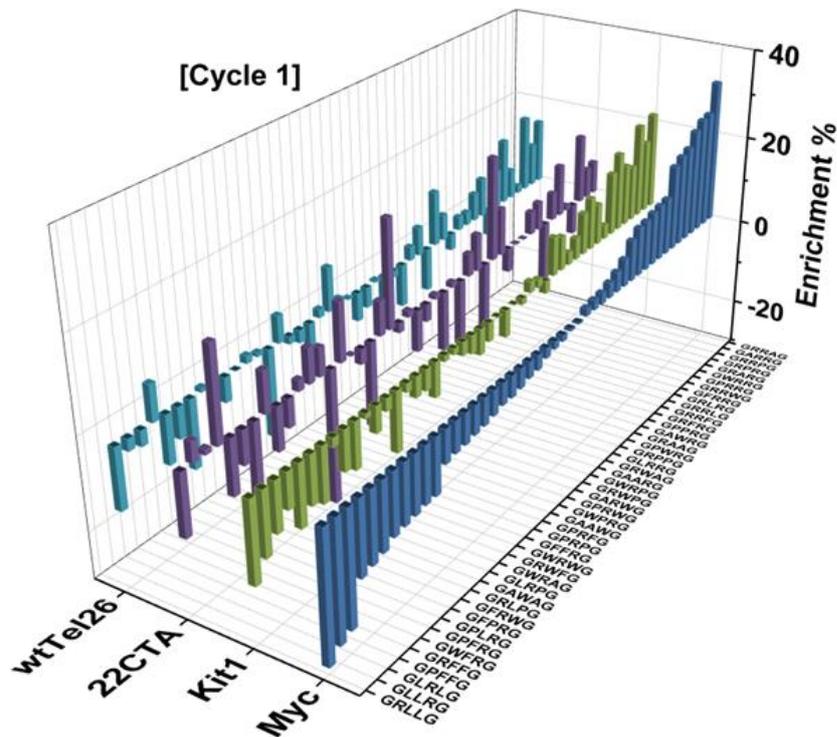
Compte tenu des considérations émises pour l'enrichissement des différents blocs de constructions après sélections sur les cibles contrôles, il sera primordial d'évaluer chaque motif composant un ligand lors de la sélection des meilleures entités chimiques afin de s'affranchir au maximum des effets non-spécifiques envers les contrôles (streptavidine, scr26 et hp1).

Paragraphe 2.1.2.b) Myc versus les autres topologies de structures G-quadruplex.

Après l'analyse des blocs de constructions sur les différentes cibles contrôles, il est intéressant de comparer leur enrichissement individuel en fonction des cibles G-quadruplex adoptant différentes topologies. Pour rappel, les deux promoteurs d'oncogènes Myc et Kit1 adoptent une topologie parallèle, la séquence télomérique 22CTA adopte une topologie antiparallèle alors que la séquence wtTel26 déploie une structure majoritairement hybride.

Les histogrammes 3D ci-dessous permettent ainsi d'illustrer l'enrichissement des *BB* (en %) dans les conditions de sélections réalisées contre les cibles rappelées plus haut. Les conditions de sélections décrites sont équivalentes au WASH1, c'est-à-dire dans des conditions de lavages non-stringentes exclusivement réalisées à température ambiante. Au contraire des données d'enrichissement contre les cibles contrôles et la cible Myc exposée précédemment et issue d'un duplicata (N=2) de sélections, les données obtenues pour les différentes topologies de G-quadruplex, donc la cible Myc, sont issues de sélections réalisées en monocopie (N=1). À ce titre, l'ordre de présentation des enrichissements croissants des *BB* contre la cible Myc peut ponctuellement varier sans altérer les conclusions acquises précédemment.

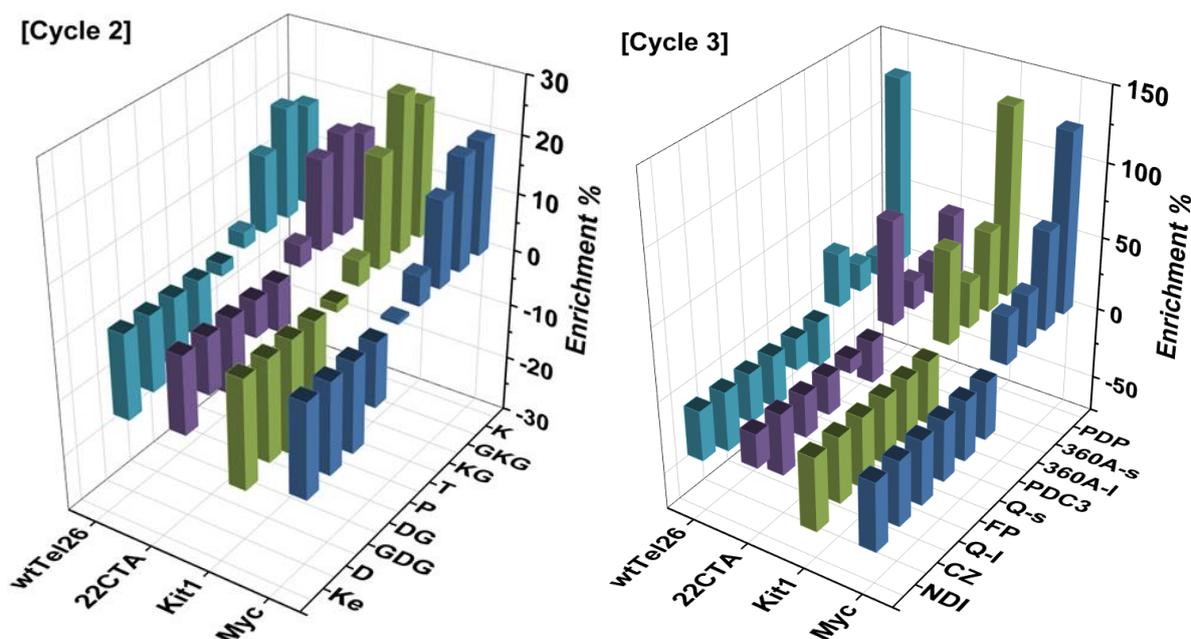
L'étude des données d'accroissements des blocs de constructions commence par l'évaluation des enrichissements des motifs à topologies parallèles, à savoir les promoteurs d'oncogènes Myc et Kit1 (Graphique 15). Conjointement à son homologue Myc, les enrichissements sur Kit1 valorisent les pentapeptides doublement arginilés. Malgré des variations dans les valeurs d'enrichissement avec un maximum aux alentours de 20%, 13 des 18 pentapeptides enrichis portent le motif RR. Ainsi, la topologie de la cible quadruplex semble influencer les motifs enrichis puisque le même groupe de peptides sont retenus par la conformation parallèle.



Graphique 15 : Enrichissement (%) des 46 pentapeptides DEL1 — [Cycle 1] en fonction des cibles wtTel26, 22CTA, Kit1 et Myc. Valeurs obtenues en condition de sélection WASH1 (25°C). N=1.

De façon intéressante, l'enrichissement des *BB1* contre la topologie antiparallèle n'exhibe pas les mêmes groupes de peptides. En effet, ce sont majoritairement les composés hydrophobes qui sont enrichis à l'exemple des séquences GAWRG, GAAWG et GPFFG atteignant des taux de plus de 20%. Certains pentapeptides doublement arginilés sont également enrichis contre la cible 22CTA mais ces enrichissements ne dépassent pas la moyenne de 7%. La cible wtTel26 adopte une structure hybride (3 + 1) au sein du milieu de sélection. Cette topologie présente des caractéristiques communes aux topologies parallèle et antiparallèle et il semble que les enrichissements des *BB1* contre wtTel26 cible affichent également des similitudes avec les données d'enrichissement obtenues pour les deux autres topologies. En effet, tout comme pour les cibles Myc et Kit1, les séquences portant un motif doublement arginilés sont enrichies avec une moyenne de 15%. De plus, des séquences hydrophobes similaires à celles retrouvées contre la cible 22CTA à savoir GAWRG, GAAWG et GPFFG sont enrichies à plus de 13%. Finalement, les caractéristiques des groupes de motifs enrichis au [Cycle 1] de la chimie combinatoire montrent une certaine sélectivité pour les structures quadruplex considérées.

L'évaluation des données d'enrichissements des 9 cœurs trifonctionnels du [Cycle 2] ainsi que des 10 motifs hétéroaromatiques du [Cycle 3] sont représentés sur le Graphique 16 ci-dessous.



Graphique 16 : Enrichissement (%) des 9 cœurs trifonctionnels DEL1 — [Cycle 2] (gauche) et des 10 motifs hétéroaromatiques DEL1 — [Cycle 3] (droite) en fonction des cibles wtTel26, 22CTA, Kit1 et Myc. Valeurs obtenues en condition de sélection WASH1 (25°C). N=1.

Les cœurs trifonctionnels (Graphique 16, gauche) ont montré des enrichissements spécifiques à hauteur de 20% des dérivés de l'acide aminé Lysine (K, KG, GKG) contre la cible parallèle Myc. L'hypothèse formulée plus tôt dans ce rapport, suggérant que la lysine déploie une longueur de chaîne latérale optimale pour orienter les ligands sur le promoteur Myc G-quadruplex, peut être étendue à toutes les cibles quadruplex comparées sur cette figure. En effet, quelle que soit la topologie de la cible considérée, les dérivés de Lysine sont les plus enrichis (enrichissement moyen de 20%) et l'ordre décroissant du podium de tête (GKG > K > KG) reste similaire. En addition aux dérivés lysines, il est également possible de noter l'enrichissement faible (mais constant, quel que soit les conditions d'expériences) d'environ 5% du cœur Tren trifonctionnel. A l'inverse, les dérivés à chaînes courtes (Dap et dérivés) ainsi que l'acide aminé intracyclique (Pro) et même le dérivé Lysine étendue (Ke) ne sont qu'exclusivement déplétés après sélections sur les différentes cibles G-quadruplex.

L'analyse des données d'enrichissements des motifs hétéroaromatiques après sélection sur cibles à topologies variées est représentée sur la Graphique 16, à droite. De manière similaire aux constatations sur la cible Myc, seuls les ligands PDP, 360A-s, 360A-I et PDC3 sont enrichis sur les autres topologies de G-quadruplex. Une nouvelle fois, chacune des molécules à système fusionné, à l'exception de PDC3, a été déplétée.

Malgré l'enrichissement exclusif des 4 ligands cités précédemment, le taux d'enrichissement de chaque motif diffère suivant la cible quadruplex. En effet, vis-à-vis de la cible Myc c'est le ligand Pyridostatine (PDP) qui obtient une valeur d'enrichissement maximum avec plus de 120%. L'enrichissement des autres ligands présente des valeurs inférieures d'un facteur 2 à 5. La deuxième séquence de proto-oncogène utilisée, Kit1, diffère un peu de son homologue

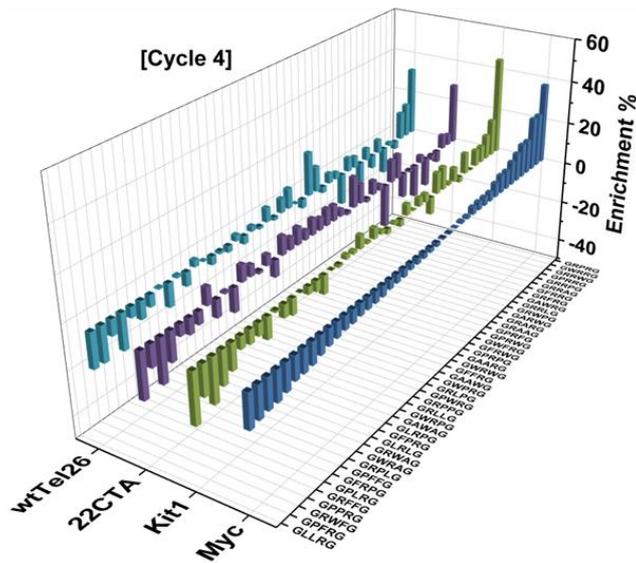
Myc. En effet, malgré un maximum d'enrichissement maintenant pour le motif PDP (~130%) c'est le composé PDC3 qui atteint la deuxième place avec un taux d'environ 50%. L'ordre d'enrichissement des motifs diffère totalement après sélection sur la cible antiparallèle 22CTA. En effet, c'est la phénanthroline (PDC3) qui apparaît comme étant le composé le plus enrichi contre 22CTA avec une valeur de 60%. Le ligand PDP reste néanmoins bien enrichi, atteignant une valeur avoisinant les 50% d'enrichissement. Enfin, la cible hybride de la séquence télomérique wtTel26 présente un enrichissement nettement supérieur (facteur 5) pour la Pyridostatine que pour le deuxième composé le plus enrichi qui est la Phénanthroline.

Tableau 18 : Ordre d'enrichissement des ligands DEL1 — [Cycle 3] en fonction de la cible G-quadruplex.

Ordre enrich. décroissant	Myc (parallèle)	Kit1 (parallèle)	22CTA (antiparallèle)	wtTel26 (Hybride)
1	PDP	PDP	PDC3	PDP
2	360A-s	PDC3	PDP	PDC3
3	360A-l	360A-s	360A-s	360A-l
4	PDC3	360A-l	360A-l	360A-s

Finalement, le ligand Pyridostatine PDP semble être un ligand universel des différentes topologies de G-quadruplex étudiées au sein du processus de sélection effectué. Toutefois, la cible antiparallèle 22CTA présente un meilleur enrichissement pour le ligand Phénanthroline que pour le ligand Pyridostatine. Enfin, parmi les deux dérivés Pyridocarboxamide, le 360A-s à chaîne courte et le 360A-l à chaîne longue, il a été possible de constater que c'est le dérivé à chaîne courte qui est le plus enrichi. En considérant le maximum d'enrichissement des dérivés lysines trifonctionnels (K, KG ou GKG) portant déjà une chaîne latérale longue de 4 carbones, il est plausible d'imaginer que l'ajout d'une autre chaîne longue sur un ligand de G-quadruplex allongerait exagérément la molécule et induirait une moins bonne interaction avec la cible.

L'enrichissement des *BB4*, dont les motifs sont équivalents à ceux utilisés au premier cycle de la bibliothèque combinatoire, a également été étudié en fonction de la topologie des cibles G-quadruplex (Graphique 17).



Graphique 17 : Enrichissement (%) des 46 pentapeptides DEL1 — [Cycle 4] en fonction des cibles wtTel26, 22CTA, Kit1 et Myc. Valeurs obtenues en condition de sélection WASH1 (25°C). N=1.

L'évaluation des données d'enrichissement des *BB4* contre la cible Myc a permis de détecter des enrichissements préférentiels pour les pentapeptides doublement arginilés (Graphique 17) à défaut des séquences peptidiques aux propriétés hydrophobes. Après sélection contre la cible du promoteur d'oncogène Kit1, les tendances retrouvées sont les mêmes que celles obtenues pour le proto-oncogène Myc à savoir un surenrichissement des séquences polaires à double arginine. Il est à noter que la séquence GRPRG a été enrichie avec plus de 45% contre Kit1 alors que la deuxième séquence la plus enrichie, GWRRG, n'a atteint que les 17%. Une analogie dans l'enrichissement des *BB4* a été observée lorsque deux cibles de topologies similaires ont été comparées. *A contrario*, il a été possible de constater une variation des propriétés des *BB4* enrichis dès lors que les autres topologies de G-quadruplex ont été évaluées. En effet, l'étude sur la séquence antiparallèle 22CTA a montré l'enrichissement de quelques séquences doublement arginilés (GRPRG, GWRRG ou encore GRRWG). Néanmoins, l'essentielle des séquences enrichies ont des caractéristiques hydrophobes avec la présence des acides aminés Tryptophane, Phénylalanine, Proline et Alanine. Le même constat a pu être effectué lorsqu'il s'agissait de s'intéresser de l'enrichissement des *BB4* sur la séquence télomérique antiparallèle wtTel26. Toutefois, alors que 21 séquences distinctes présentent des valeurs enrichies contre la cible 22CTA, seules 12 séquences ont été enrichies contre l'hybride télomérique.

L'étude des données d'enrichissement de chaque motif de la librairie combinatoire a montré que les sélections réalisées sur différentes topologies de G-quadruplex exerçaient une influence sur les événements identifiés après séquençage. Les enrichissements des *BB1* et *BB4*, représentant les 46 séquences pentapeptidiques, ont particulièrement été impactés par les différentes topologies étudiées. Il a été possible de constater que les topologies parallèles (Myc et Kit1) renaient préférentiellement les motifs doublement arginilés au détriment des séquences hydrophobes. À l'inverse, le G-quadruplex antiparallèle (22CTA) semble interagir de manière favorisée avec les peptides hydrophobes en diminuant l'interaction avec les

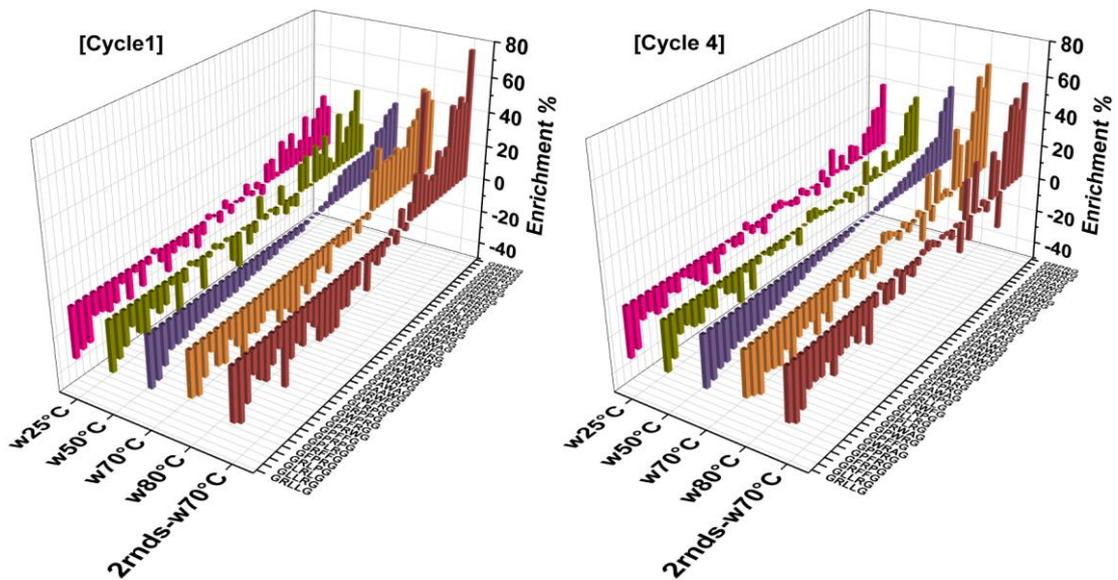
séquences arginilés. Enfin, la séquence G-quadruplex à topologie hybride (wtTel26) semble tirer profit de certains peptides polaires et d'autres apolaires afin d'augmenter les propriétés d'interactions avec la cible. L'interprétation des résultats d'enrichissements des *BB2* a montré que la topologie de la structure quadruplex n'influence pas le cœur trifonctionnel enrichi. Il est alors envisageable que l'enrichissement des motifs du [*Cycle 2*] soit directement dicté par la nature du motif hétéroaromatique (*BB3*). Concernant les *BB3*, seuls les ligands à plateaux étendus larges présentant soit une dérivation quinoline ou quinolinium ont été enrichis (PDP, 360A (— s ou — l) et PDC3). Le ligand Pyridostatine (PDP) présente de fortes affinités pour l'ensemble des ligands des structures quadruplex étudiées alors que le taux d'enrichissement des ligands Phénanthroline (PDC3) et Pyridocarboxamide (360A) influe suivant la topologie adoptée par la structure secondaire.

Paragraphe 2.1.2.c) Variation d'enrichissement par applications de conditions stringentes (cible Myc).

Parmi les 29 conditions de sélections réalisées, 14 sélections ont été réalisées sur la cible du proto-oncogène Myc de topologie parallèle. Au sein de ces 14 sélections, il est possible de retrouver des variations au niveau des conditions de lavages (stringences) ou encore des doubles sélections. L'évaluation des données d'enrichissement obtenues après applications de différentes conditions de stringences est intéressante puisqu'il est alors possible de comparer l'évolution des motifs de la librairie combinatoire en fonction de la dureté des conditions de sélections.

Les histogrammes 3D représentés ci-dessous permettent de comparer l'enrichissement de tous les motifs de la librairie combinatoire dans des processus de sélections à stringences variables contre la cible Myc. Les conditions de sélections comparées sont issues de la Sel01 (Graphique 18, « w25 °C ») ou les cinq lavages ont été effectués à 25°C, la Sel04 (« w50 °C ») ou le 4^{ième} lavage a été effectué à 50°C, la Sel06 (« w70 °C ») ou le 4^{ième} lavage a été effectué à 70°C, la Sel09 (« w80 °C ») ou le 4^{ième} lavage a été effectué à 80°C et enfin la Sel12 (« 2nrds-w70 °C ») ou deux sélections successives ont été réalisées avec l'application de 70°C lors du quatrième lavage.

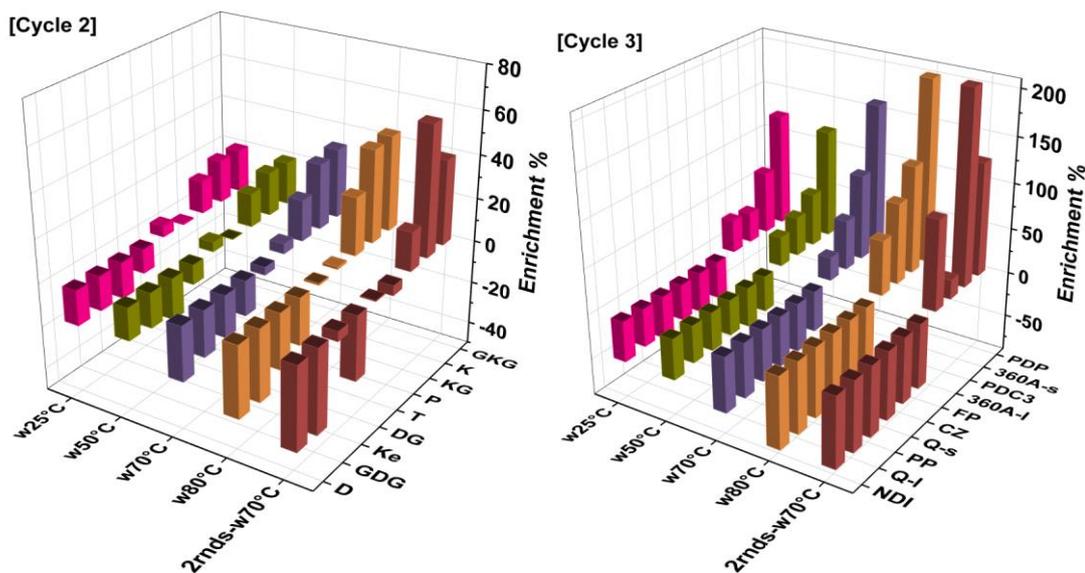
Les données d'enrichissement des pentapeptides utilisés lors du [*Cycle 1*] et [*Cycle 4*] de la librairie combinatoire sont représentées sur la figure ci-dessous. Les motifs enrichis sont classés dans un ordre croissant en fonction de la condition « w70 °C ». De plus, les valeurs d'enrichissements de la condition « w25 °C » sont les celles déjà discutées auparavant lors de la comparaison de la séquence Myc parallèle avec les cibles contrôles ou encore avec les autres topologies de G-quadruplex.



Graphique 18 : Enrichissement (%) des 46 pentapeptides DEL1 — [Cycle 1] (gauche) et DEL1 — [Cycle 4] (droite) en fonction de la stringence appliquée sur la cible Myc. Valeurs obtenues en condition de sélection WASH1, WASH2, WASH4, WASH5 et WASH4/WASH4 (double-sélection). N=1.

Lors de l'identification des tendances d'enrichissement des *BB1* ou *BB4*, il a été possible de distinguer que les tendances générales sont maintenues et que seule la valeur du taux d'enrichissement des motifs individuels a variée. Lors des sélections sans applications stringentes (5 lavages à 25°C) il a été remarqué, et largement discuté précédemment, que les motifs peptidiques enrichis sont principalement polaires puisqu'elles portent majoritairement des motifs doublement arginilés. L'évolution de la stringence des lavages (cf. du 4^{ième} lavage) a permis d'augmenter les valeurs d'enrichissements sans altérer l'identité des motifs enrichis. À l'exemple de la série avec application du lavage à 80°C, les motifs les plus enrichis atteignent plus de 40% alors que les mêmes motifs en conditions classiques (cf. 25°C) n'atteignent pas plus de 25%. De façon intéressante, il est possible de constater que malgré les processus de doubles sélections, l'identité des pentapeptides enrichis ainsi que les valeurs d'enrichissement n'ont pas varié.

Les cœurs trifonctionnels, *BB2*, enrichie sur Myc dans des conditions classiques ont montré un enrichissement préférentiel pour les motifs à chaîne longue tel que les dérivés Lysines (K, KG et GKG). L'évaluation des conditions stringentes (Graphique 19, gauche) a montré une nouvelle fois que l'augmentation des températures de lavages n'a fait qu'amplifier les valeurs d'enrichissement. En effet, les dérivés Lysines restent les motifs les plus enrichis, quelles que soient les conditions appliquées. Après application d'un processus de double sélection seule, les dérivés Lysines apparaissent également toujours enrichis. Dans ce cas de figure, le motif Lysine (K) semble se démarquer de ces homologues KG et GKG puisque ce motif atteint les 60% d'enrichissement.



Graphique 19 : Enrichissement (%) des 9 cœurs trifonctionnels DEL1 — [Cycle 2] (gauche) et des 10 motifs hétéroaromatiques DEL1 — [Cycle 3] (droite) en fonction de la stringence appliquée sur la cible Myc. Valeurs obtenues en condition de sélection WASH1, WASH2, WASH4, WASH5 et WASH4/WASH4 (double-sélection). N=1.

L'analyse de l'enrichissement des motifs de troisième cycle de la librairie combinatoire sur la cible Myc et dans des conditions classiques (« w25 °C) a montré que seuls les motifs hétéroaromatiques à plateau hydrophobe large et à motif quinoline ou quinolinium ont été enrichis, à savoir la Pyridostatine (PDP), les Pyridocarboxamide (360A-s/— I) et la Phénanthroline (PDC3). Il a alors été possible de suivre l'évolution de l'enrichissement de ces ligands après applications de conditions stringentes (Graphique 19, droite). En pratique, et comme pour les motifs *BB1*, *BB2* et *BB4*, l'application de dureté dans les lavages n'a fait qu'amplifier les valeurs d'enrichissement sans pourtant modifier la nature ou l'ordre des ligands enrichis. En effet, les données d'enrichissement du ligand PDP augmentent en fonction de l'augmentation des conditions de stringence. Le deuxième ligand le plus enrichi, la Pyridocarboxamide à chaîne courte, a évolué plus timidement mais reste néanmoins enrichi à plus de 100% dans les conditions de lavage à 80°C.

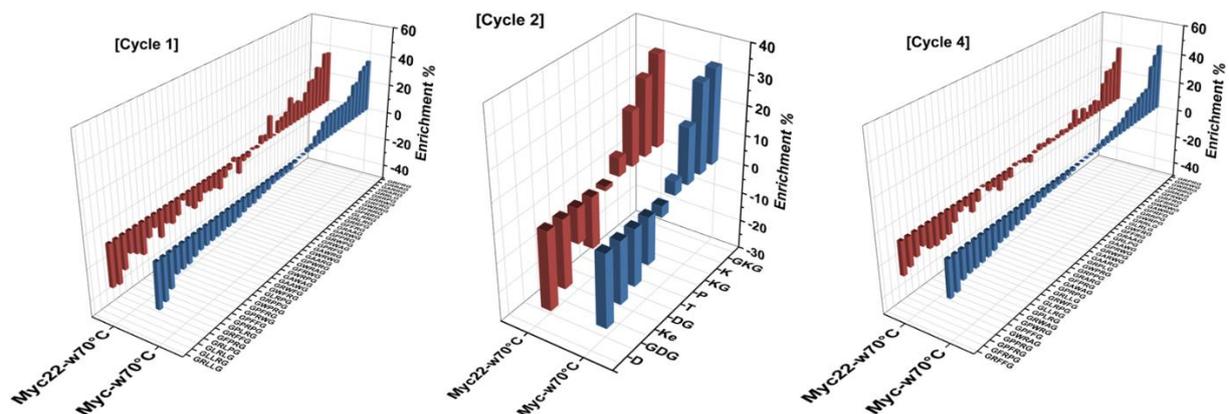
Après application d'un processus de double sélection, une inversion est à noter. Alors que le ligand PDP est enrichi à près de 200% dans les conditions les plus stringentes (80°C), ce même ligand atteint une valeur d'enrichissement divisée par deux après une double sélection (cf. deux sélections successives à 70°C). Toutefois, alors que le 360A-s est enrichi à 100% à 80°C, il est enrichi à près de 200% après une série de deux sélections successives. Ces résultats pourraient s'expliquer par des différences de cinétiques d'association et/ou de dissociation distinctes entre les dérivés PDP et 360A-s, permettant par exemple aux dérivés 360A-s de mieux persister dans un contexte où l'incubation a été prolongée lors du premier cycle, et qu'elle a été réalisée dans des conditions plus diluées en librairie lors du 2^{ème} cycle (le deuxième cycle ayant débuté par l'éluat du 1^{er} à hauteur de 1E +09 copies).

Finalement, l'étude de l'impact des conditions stringentes de sélections sur l'enrichissement des motifs composant la librairie combinatoire a montré que l'identité des *BB1*, *BB2* et *BB4* enrichis a été maintenue et que seuls les taux d'enrichissements des motifs individuels ont varié en fonction de l'intensité de la stringence appliquée. Dans le cadre des ligands de structures G-quadruplex utilisés pour dériver la librairie au [Cycle 3] les mêmes événements ont pu être observés. En effet, les 4 ligands enrichis après sélection en condition classique contre la cible Myc ont été retrouvés enrichis en condition stringentes (50, 70 et 80°C) jusqu'à atteindre des valeurs d'enrichissements de 100% à 200%, suivant le ligand considéré, après application d'un quatrième lavage à 80°C. Toutefois, l'application de plusieurs tours de sélections contre la cible Myc (dans ce cas, deux tours) a généré un événement inattendu. Alors que le ligand PDP a largement dominé les sélections stringentes, il a été constaté que c'est le ligand 360A-s qui a dominé la double sélection. Globalement, ces résultats confortent fortement ceux obtenus dans les conditions initiales et concordent vers l'identification de motifs préférentiels pour la reconnaissance du G-quadruplex Myc.

Paragraphe 2.1.2.d) Myc versus Myc22.

La sélection 20 réalisée contre la cible Myc22, un dérivé de la séquence du proto-oncogène Myc, a été effectuée dans une seule condition, avec application de lavages stringents (cf. 4^{ème} lavage) à 70°C.

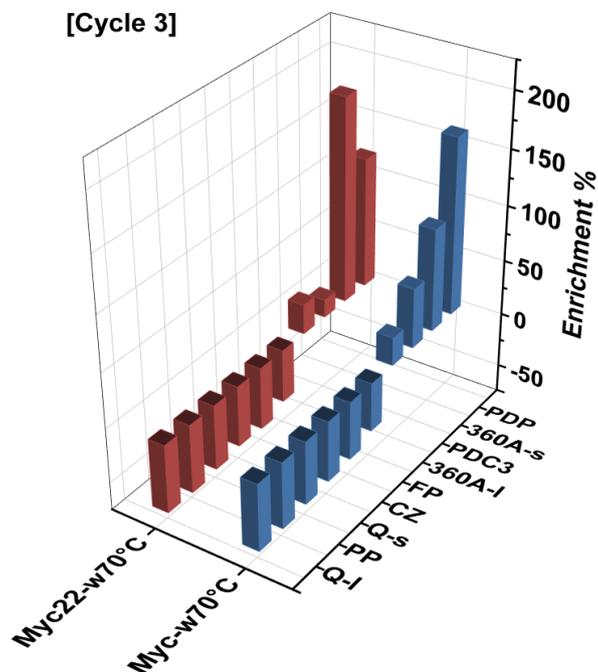
Conjointement aux conclusions préalablement émises, seules les séquences peptidiques polaires (double arginine ; Graphique 20, [Cycle 1] et [Cycle 4]) ainsi que les cœurs trifonctionnels à chaîne longue K, KG et GKG (Graphique 20, [Cycle 2]) ont été enrichis contre la cible Myc22.



Graphique 20 : Enrichissement (%) des 46 pentapeptides DEL1 — [Cycle 1] (gauche), des 9 cœurs trifonctionnels DEL1 — [Cycle 2] (milieu) et des 46 pentapeptides DEL1 — [Cycle 4] (droite) en fonction des cibles Myc22 et Myc. Valeurs obtenues en condition de sélection WASH4 (70°C). N=2.

Alors que les données d'enrichissement obtenues contre les deux cibles semblent similaires sur 3 des 4 cycles de chimie combinatoire, il a été possible de distinguer des enrichissements différents, vis-à-vis des deux cibles, pour certains *BB3*. En effet, contre la cible parallèle Myc dans des conditions de sélections stringentes (70°C), c'est le ligand PDP qui a été le plus

enrichi (150%) suivi par le ligand 360A-s (75%). À l'inverse, et dans les mêmes conditions, le ligand le plus enrichi contre la cible Myc22 a été identifié comme étant le ligand 360A-s avec près de 175% d'enrichissement. Le composé PDP a également été enrichi contre la cible Myc22 mais n'a atteint que la deuxième place en terme d'enrichissement avec une valeur de 110% (Graphique 21).



Graphique 21 : Enrichissement (%) des 10 motifs hétéroaromatiques DEL1 — [Cycle 3] en fonction des cibles Myc22 et Myc. Valeurs obtenues en condition de sélection WASH4 (70°C). N=2.

Comparativement à la double sélection réalisée sur la cible Myc, la cible Myc22 place également le dérivé Pyridocarboxamide à chaîne courte en première place du classement des motifs enrichis. Alors que la Phénanthroline a été enrichie après la double sélection sur Myc, ce même composé ne présente pas d'enrichissement significatif contre la cible Myc22.

En conclusion, si on tient compte de toutes les données d'enrichissement il est possible d'affirmer que les séquences doublement arginilées sont surreprésentées contre les cibles quadruplex, avec un léger enrichissement des séquences hydrophobes contre les cibles à topologies antiparallèle ou hybride. Les expériences de stringences sur les pentapeptides n'ont fait qu'affirmer les données obtenues par augmentation des taux d'enrichissement. Parmi tous les cœurs trifonctionnels utilisés pour dériver la DEL1, seuls les motifs dérivés de la Lysine (K, KG et GKG) sont pertinents et enrichis contre les cibles quadruplex. Enfin, parmi les motifs hétéroaromatiques du cycle 3, le ligand Pyridostaine (PDP) est le ligand le plus représenté, toutes cibles toutes stringences considérées. Néanmoins, le dérivé Phénanthroline (PDC3) reste également intéressant puisque le PDC3 semble être préférentiellement enrichi contre la cible antiparallèle. De plus, la Pyridocarboxamide (à chaîne courte, 360A-s) a montré d'intéressantes propriétés d'affinité après double sélection sur la cible Myc mais également après sélection (stringente) contre la cible Myc22. Les motifs, et leurs caractéristiques associées, pris indépendamment devraient alors pouvoir être associés afin de former des molécules tétra-dérivatisés capable d'interagir avec une haute affinité avec

les cibles quadruplex. Il est dès lors intéressant d'étudier l'enrichissement des molécules uniques de la librairie contre les différentes cibles.

Section 2.1.3 : Analyse des molécules uniques après sélections.

Après l'analyse des motifs de chaque étape de la librairie combinatoire, les molécules uniques ont également été analysées. La méthode d'analyse des données d'enrichissements des molécules uniques a suivi les mêmes procédés que celles appliquées pour les blocs de constructions à savoir le calcul du Ratio Normalisé (R_n) de chaque molécule présente au sein d'une sélection donnée, suivi du Ratio ajusté (R_a) et enfin du facteur d'enrichissement.

Il est tout bonnement impossible de discuter de tous les évènements retrouvés au sein d'une condition de sélection donnée. Toutefois, afin d'illustrer les molécules uniques les plus enrichies, des « top 100 » ont été réalisés. Ces figures illustratives de la variabilité des motifs composants les cent molécules les plus enrichies pour chaque condition ont été réalisées sur le logiciel Cytoscape. À l'exemple de la Figure 100 (gauche) ci-dessous, il est possible de montrer l'occurrence et le poids des motifs issus des 100 molécules les plus enrichies contre le proto-oncogène Myc en condition standard (lavages à 25°C). Les figures sont représentées sous forme d'un quadrilatère dont la légende est illustrée en Figure 100 (droite). Sur chaque partie du quadrilatère sont respectivement symbolisées les blocs de construction trouvés au [Cycle 1-2-3 et 4]. Chaque bloc de construction est symbolisé par un cercle dont la couleur, proportionnelle au volume du cercle, évoque le pourcentage d'enrichissement normalisé de chaque motif. Plus un motif aura un cercle large et plus la couleur associée tend vers le rouge, plus le motif est présent au sein du « top 100 » de la sélection considérée. D'autre part, la corrélation entre les motifs a également été représentée. Cette corrélation représente l'intensité du lien entre deux blocs de construction. Ces intensités de corrélation ont alors été représentées en fonction des liens chimiques retrouvés au sein de la structure des ligands de la DEL1. Ainsi, les motifs du [Cycle 2] représentant les cœurs trifonctionnels centraux de la DEL1, ont alors été corrélés aux pentapeptides [Cycle 1 et 4] mais également aux ligands de structures quadruplex [Cycle 3]. Une nouvelle fois, l'intensité de corrélation entre deux motifs est symbolisée par un trait dont l'épaisseur et la couleur varient en fonction des données. Plus un trait est épais et rouge, plus les motifs sont représentés au sein du « top 100 ».

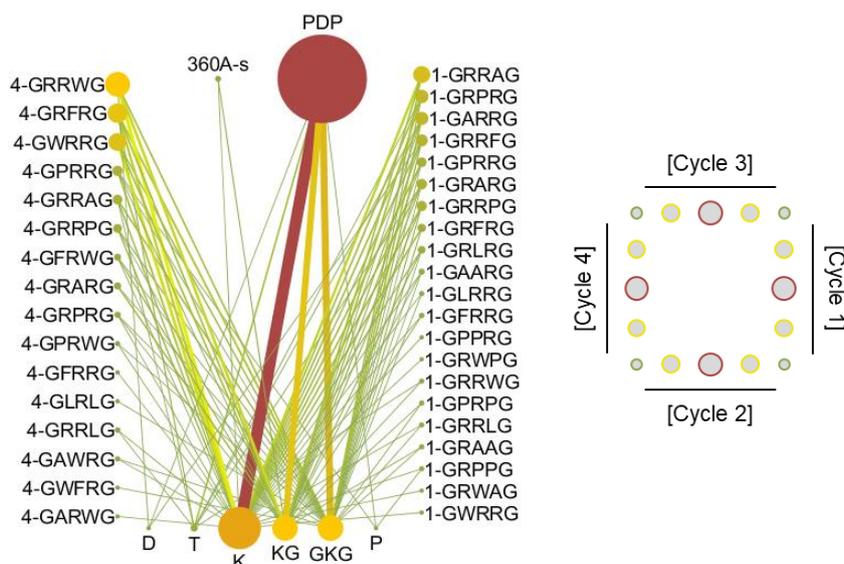


Figure 100 : Top 100 des composés sélectionnés contre la cible Myc — WASH1 (25°C). Représentation i) de l'enrichissement des motifs uniques et ii) leurs corrélations pour chaque molécule du top 100.

Dans le cas de la représentation du « top 100 » contre la cible Myc en condition standard, il est possible de constater que le ligand Pyridostatine est largement majoritaire par rapport au ligand Pyridocarboxamide à chaîne courte. L'intensité de corrélation entre les motifs du [Cycle 2] et le ligand PDP montre que ce sont les dérivés Lysines (K, KG, GKG) qui sont les plus retrouvés en présence de PDP. De plus, parmi ces lysines trifonctionnelles, c'est l'acide aminé K qui a été le plus enrichi.

Après s'être intéressé aux enrichissements des blocs de constructions présents au [Cycle 2] et [Cycle 3] ainsi que le lien entre les deux, il est intéressant de constater que seuls 35 à 45% des motifs du [Cycle 1], et [Cycle 4], sont représentés au sein du « top 100 » et qu'environ 65% de ces séquences sont doublements arginillées. Les séquences portant le motif double arginine apparaissent comme étant les plus enrichies et les liens les plus intenses ont été retrouvés entre les dérivés Lysines ([Cycle 2]) et les pentapeptides arginillés à l'exemple de GRRWG, GRFRG et GWRRG pour le [Cycle 1] et GRRAG, GRPRG et GARRG pour le [Cycle 4].

En résumé, les données d'enrichissement et de corrélation des 100 premières séquences contre la cible Myc (condition standard) montrent un enrichissement préférentiel pour un ligand dont la structure générale est représentée ci-dessous (Figure 101).

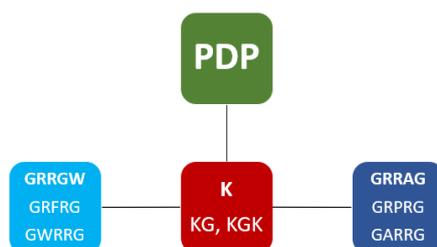


Figure 101 : Composition des motifs les plus représentés au sein du top 100 contre la cible Myc — WASH1 (25°C).

Le même travail a été effectué sur la cible Myc après sélection stringente à 70°C (cf. 4^{ème} lavage à 70°C, Figure 102). En comparaison avec les données obtenues après sélection en condition classique, seul le ligand Pyridosatine est enrichi à 70°C contre la cible Myc. Les cœurs trifonctionnels retrouvés sont une nouvelle fois les dérivés Lysines et le lien connectant les motifs du [Cycle 2] au motif unique du [Cycle 3] apparait le plus intense entre le PDP et l'acide aminé K. Les séquences d'acides aminés issues des [Cycle 1] et [Cycle 4] ont vu leur nombre diminuer après conditions stringentes de sélection. Néanmoins, toutes les séquences présentes au sein du « top 100 » exhibent un motif doublement arginilés. L'application de conditions stringentes lors de processus de sélection relargue les composés mono-arginilés au profit des composés bis arginilés, montrant une nouvelle fois l'importance des deux acides aminés chargés et polaires au sein de la séquence peptidique. Dans ces nouvelles conditions, ce sont les séquences GRRWG, GWRRG et GRFRG qui sont les plus enrichies au [Cycle 4] de la DEL1 alors que les séquences GARRG, GRARG et GRRAG sont-elles plus enrichies au [Cycle 1].

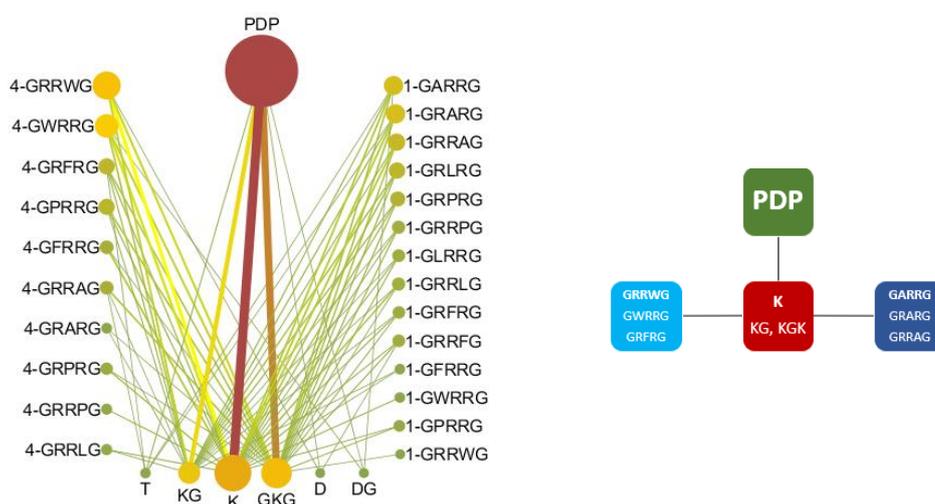


Figure 102 : Gauche. Top 100 des composés sélectionnés contre la cible Myc — WASH4 (70°C). Représentation i) de l'enrichissement des motifs uniques et ii) leurs corrélations pour chaque molécule du top 100. Droite. Composition des motifs les plus représentés au sein du top 100 contre la cible Myc — WASH4 (70°C).

En continuant l'analyse sur les sélections stringentes il est intéressant d'observer l'évolution des molécules issues d'un « top 100 » après un processus de sélections successives. Il a été commenté auparavant que le processus de double sélection permettait au ligand Pyridocarboxamide (360A-s) de surpasser le ligand Pyridostatine (PDP) en terme d'enrichissement. L'illustration de l'enrichissement des 100 premières molécules uniques après une double sélection est représentée sur la Figure 103 ci-dessous. Alors que le ligand PDS dépassait largement les valeurs d'enrichissement des autres ligands du [Cycle 3] dans les conditions de sélections classiques (et stringentes), il est possible de retrouver ce même ligand à la deuxième place après double sélection stringente (cf. 4^{ème} lavage à 70°C). En effet, ce sont les ligands Pyridocarboxamides qui partagent l'identité des événements retrouvés, avec un maximum d'enrichissement pour le dérivé à chaîne courte. De manière intéressante, la quantité de séquences peptidiques exhibée dans ce « top 100 » est nettement supérieure

aux quantités retrouvées après sélections stringentes et même classiques. La majorité des séquences retrouvées avec un fort enrichissement portent le motif « RR » (GWRRG au [Cycle 4] et GLRRG, GRPRG au [Cycle 1]). Les autres séquences présentent, notamment les séquences mono-arginilées, ne semblent être que des évènements à faibles abondances puisque leurs valeurs d'enrichissements sont très faibles. Le cœur trifonctionnel majoritairement enrichi et exhibant des valeurs de corrélations les plus importantes avec les peptides et les ligands de structures quadruplex est l'acide aminé K. Alors que la lysine et ces dérivés étaient les plus présentes au sein des sélections standards ou stringentes, il est possible de constater que K occupe la majorité des cœurs trifonctionnels au sein du « top 100 » de double sélection.

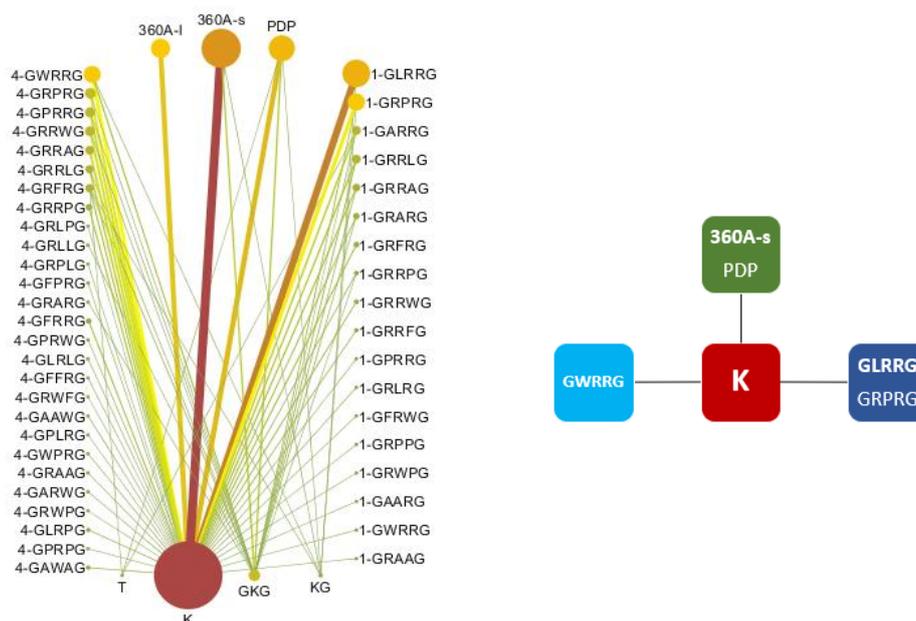


Figure 103 : Gauche. Top 100 des composés sélectionnés après double sélection contre la cible Myc — W4/W4 (70°C). Représentation i) de l'enrichissement des motifs uniques et ii) leurs corrélations pour chaque molécule du top 100. Droite. Composition des motifs les plus représentés au sein du top 100 après double sélection contre Myc.

Alors que la cible Myc et ces conditions stringentes ont été discutées, il est intéressant de s'intéresser aux molécules uniques retrouvées après sélection contre les cibles contrôles, à savoir l'*Hairpin*, hp1, et les billes Streptavidine.

Contre la Streptavidine, Figure 104 (gauche), il est aisé de constater que les motifs enrichis n'ont rien en commun avec ceux discutés jusqu'à présent. En effet, parmi les 10 motifs hétéroaromatiques du [Cycle 3], ce sont les dérivés quinolines, le furopyridazinone et la phénylpropyne qui sont enrichis. Un lien de corrélation intense est retrouvé entre ces motifs et le cœur trifonctionnel Ke (Lysine étendue par un acide acétique azoture) qui s'avère être le motif majoritairement enrichi au [Cycle 2]. De manière intéressante, alors que 36 des 46 séquences de la première étape de la librairie combinatoire sont retrouvées au sein du « top 100 », une séquence pentapeptidique unique a été retrouvée au [Cycle 4], le GWFRG. Il est alors possible d'affirmer qu'au vu des données obtenues, seul le pentapeptide au dernier cycle de la DEL1 interagit avec la cible Streptavidine. En effet, trop de variabilités à faible intensité

ont été obtenues au [Cycle 1], suggérant que l'interaction entre les molécules et la Streptavidine est indépendante de l'identité des motifs du premier cycle.

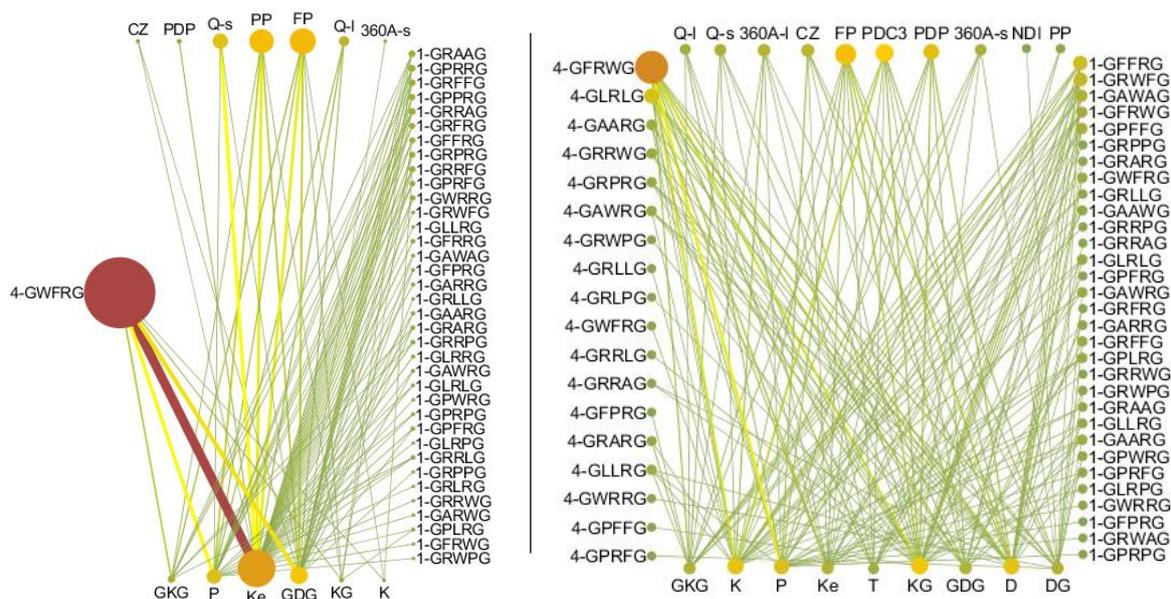


Figure 104 : Top 100 des composés sélectionnés contre la cible Streptavidine (gauche) et la cible hp1 (droite) — WASH1 (25°C). Représentation i) de l'enrichissement des motifs uniques et ii) leurs corrélations pour chaque molécule du top 100.

En s'intéressant aux molécules uniques interagissant avec la cible hp1 (Figure 104, droite) il a été possible de constater que tous les motifs des [Cycles 2 et 3] de la DEL1 étaient représentés. Toutefois, l'intensité relative de ces motifs varie puisque parmi les cœurs trifonctionnels les motifs majoritairement enrichis sont les dérivés Lysines (K, KG), la proline (P) et le Dap (D). De manière similaire, ce sont les motifs hétéroaromatiques à système fusionné (PDP, 360A (— s/— l) ainsi que la Furopyridazinone et le PDC3 qui ont été enrichis. L'enrichissement des motifs pentapeptidiques a également présenté de grandes variations. Néanmoins, ce sont les motifs GFRWG et GLRLG qui gouvernent l'enrichissement des motifs du [Cycle 4] alors que GFFRG, GRWFG et GAWAG n'ont exhibé que des enrichissements faibles au [Cycle 1].

À titre de comparaison avec les données obtenues contre les cibles contrôles et la cible du proto-oncogène Myc, il a été intéressant de voir les molécules les plus enrichies vis-à-vis de la cible de topologie parallèle Myc22 ainsi que la cible de topologie antiparallèle 22CTA. Ces deux cibles se sont démarquées lors de l'étude des blocs de construction puisqu'elles ont présenté des enrichissements pour le ligand 360A-s et PDC3, respectivement.

Comme constaté lors de l'étude des motifs individuels, le « top 100 » contre la cible Myc22 (Figure 105, gauche) montre un enrichissement préférentiel pour le ligand Pyridocarboxamide. La liaison entre ce ligand et les cœurs trifonctionnels est largement favorisée pour les dérivés de Lysines (K, KG, GK), spécialement pour la Lysine (K) qui exhibe la plus haute valeur de présence au sein du « top 100 ». Les pentapeptides retrouvés portent une nouvelle fois en

majorité une séquence doublement arginillée, à l'exemple des séquences GWRRG, GRRAG ou GPRRG au [Cycle 4] et GRPRG, GRRAG et GRARG au [Cycle 1].

L'étude du « top 100 » vis-à-vis de la cible de topologie antiparallèle 22CTA est en accord avec les observations réalisées sur les motifs individuels. En effet, le ligand Phénanthroline gouverne très largement l'identité des 100 premières molécules présentes. Le lien séparant le PDC3 aux cœurs trifonctionnels est majoritaire pour les dérivés Lysines à chaînes longues (K, KG, GKG) avec une légère préférence pour les Lysines fonctionnalisées par une ou deux glycines. Les valeurs de corrélation avec les pentapeptides enrichis sont globalement faibles puisque près des 2/3 des motifs ont été retrouvés dans le « top 100 ». Néanmoins, le motif GFRWG est dominant dans le cas du [Cycle 4].

Les structures générales des molécules uniques contre la cible Myc22 et 22CTA, issues des valeurs d'enrichissement et de corrélation entre les étapes de la librairie combinatoire, sont respectivement retrouvées en bas de la Figure 105.

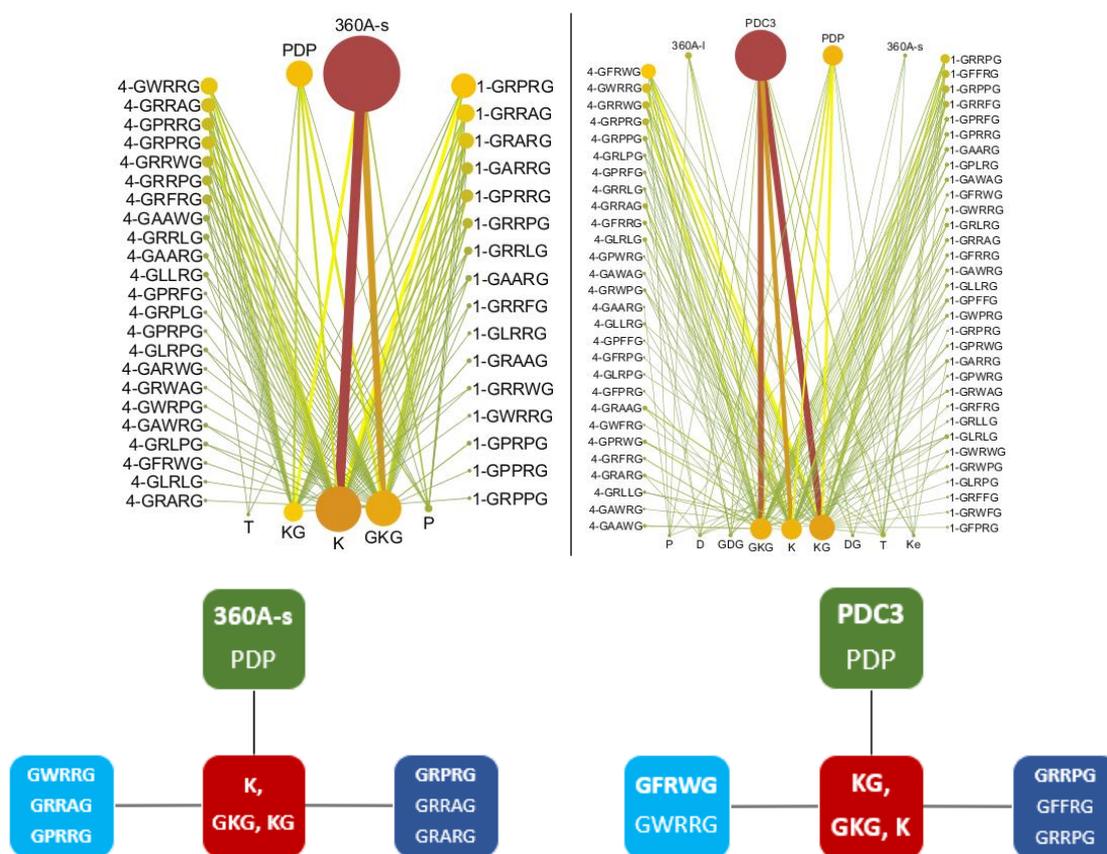


Figure 105 : Top 100 des composés sélectionnés après sélection stringente contre la cible Myc22 — WASH4 (70°C) — ainsi que la composition des motifs les plus représentés (gauche). Top 100 contre la cible 22CTA — WASH2 (50°C) — et que la composition des motifs les plus représentés (droite). Représentation i) de l'enrichissement des motifs uniques et ii) leurs corrélations pour chaque molécule du top 100.

L'analyse des « top 100 » contre les différentes cibles est en corrélation directe avec les données obtenues par analyse des motifs de la librairie combinatoire DEL1. Le ligand Pyridostatine (PDP) est le ligand le plus enrichi, toutes sélections (et conditions) confondues.

Toutefois, les ligands Pyridocarboxamide (360A) et Phenanthroline (PDC3) disputent la première place sous différentes conditions de stringences et contre des cibles de topologies variées. Avec les données de séquençages, il pourrait être pensé que le ligand 360A exhibe une préférence pour les G-quadruplex de topologies parallèles, à l'égard du promoteur d'oncogène Myc (après double sélection stringentes) et de la séquence Myc22 (après sélection stringente). De la même manière, le ligand PDC3 est apparu largement majoritaire dans les 100 premières molécules obtenues après sélection contre la cible G-quadruplex antiparallèle 22CTA.

À partir des molécules composant la DEL1 et des données de séquençage discutées jusqu'alors, il serait tentant de synthétiser un grand nombre de composés d'intérêt et de composés contrôles afin de procéder à des études biophysiques (mesure d'affinité) comparatives permettant de confirmer ou d'invalider les hypothèses formulées. Néanmoins, et pour des raisons évidentes de faisabilité, un choix plus restreint a dû être réalisé dans le cadre de cette thèse sur la base des données d'enrichissements et de spécificités suggérées par séquençage des composés les plus probants. Ainsi, trois ligands potentiels ont été resynthétisés en priorité, dérivant des motifs hétéroaromatiques Pyridostatine (PDP), Phénanthroline (PDC3) et Pyridocarboxamide (360A-s). Ces composés sont représentés ci-dessous (Figure 106) :

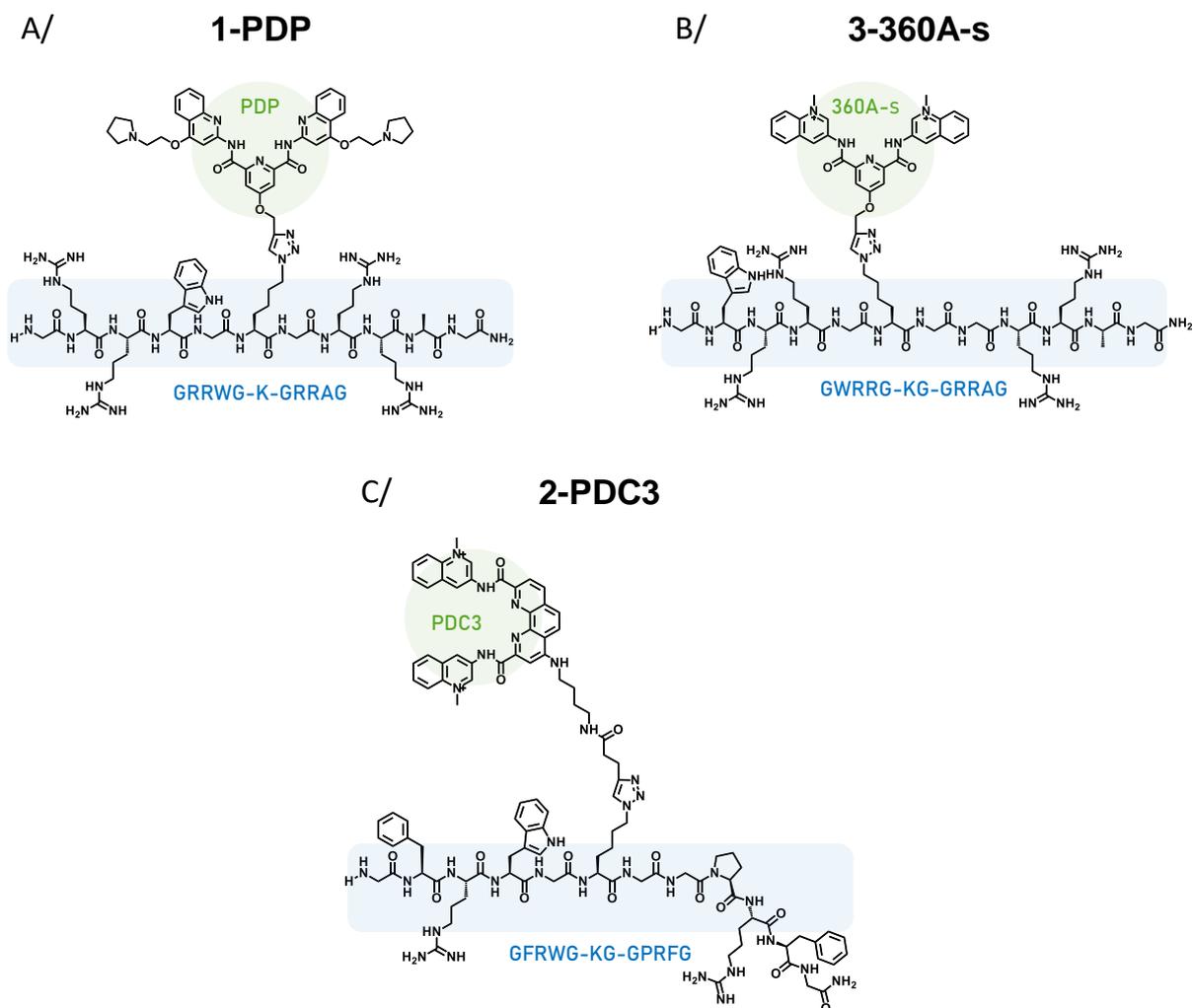


Figure 106 : Composés identifiés pour leurs enrichissements individuels à l'égard de différentes sélections (dont contrôles). A/ Dérivé Pyridostatine 1-PDS. B/ Dérivé Pyridocarboxamide 3-360A-s. C/ Dérivé Phénanthroline 2-PDC3.

Tableau 19 : Facteurs d'enrichissements et classement des trois composés retenus en fonction des différentes cibles.

	Myc @w25 °C	Myc @w70 °C	Myc @double Sel	Myc22 @w70 °C	wtTel26 @w25 °C	22CTA @w25 °C	Kit1 @w25 °C	hp1 @w25 °C	scr26 @w25 °C	Strept. @w25 °C
1- PDP % enrich.	605%	1270%	991%	471%	399%	155%	409%	-3%	396%	-34%
Classement	11	9	104	517	149	2298	113	74 621	1549	108 787
2- PDC3 % enrich.	99%	359%	73%	247%	141%	389%	210%	146%	347%	-4%
Classement	16 401	2706	31 882	5137	6812	9	3427	1043	2211	52 973
3-360A-s % enrich.	351%	438%	826%	600%	67%	86%	263%	17%	-3%	-59%
Classement	482	1671	289	120	20 147	12 657	1494	43 706	57 156	157 700

Le dérivé Pyridostatine, également appelé 1-PDP, possède un cœur Lysine (K) fonctionnalisée par deux séquences pentapeptidiques doublement arginilées, GRRWG et GRRAG. Comme

le montre le tableau 19, cette molécule atteint des valeurs d'enrichissements de plus de 600% contre la cible parallèle Myc en condition classique et double sa valeur après l'application de conditions stringentes (cf. 4^{ème} lavage à 70°C). De plus, cette molécule exhibe des enrichissements décents contre la séquence promotrice d'oncogène Kit1 et contre la cible de topologie hybride wtTel26. D'autre part, 1-PDP possède une bonne sélectivité pour l'*Hairpin 1* (hp1) et pour la streptavidine, avec des valeurs d'enrichissements négatives (déplétions). Il est toutefois à noter qu'une valeur d'enrichissement non négligeable a été trouvée contre la séquence *scramble 26* (scr26) avec un enrichissement de près de 400%. Néanmoins cette dernière expérience n'a pas été réalisée en duplicata.

Le dérivé Phénanthroline, également appelé 2-PDC3, est un composé possédant un KG trifonctionnel en son centre permettant d'être dérivatisé en deux points par une séquence GFRWG et GPRFG. Comparativement au 1-PDP, cette molécule est moins enrichie contre les cibles de topologies parallèle et hybride. En revanche, une valeur d'enrichissement de près de 400% a été observée pour la cible de topologie antiparallèle 22CTA. Enfin, et malgré une déplétion après sélection contre la cible streptavidine, le 2-PDC3 présente une certaine affinité pour les contrôles hp1 et scr26 avec des enrichissements de 150% à 350%, respectivement. Avec un comportement vis-à-vis des contrôles nettement inférieurs au 1-PDP, la synthèse du dérivé Phénanthroline a tout de même été intéressante pour la capacité supposée de la molécule à interagir préférentiellement avec la cible 22CTA, avec un caractère non négligeable pour les autres topologies de G-quadruplex.

Enfin le dérivé Pyridocarboxamide à chaîne courte, également appelé 3-360A-s, est une molécule structurellement proche du dérivé 1-PDP puisqu'elle exhibe de part et d'autre du cœur trifonctionnel des pentapeptides GWRRG et GRRAG. L'association des peptides et du 360A est réalisée par un cœur Lysine-Glycine (KG). Dans le tableau 19, il est possible de remarquer un enrichissement préférentiel du 3-360A-s pour les cibles parallèles, à l'exemple des 600% d'enrichissement contre Myc22. D'autre part, le 3-360A-s a présenté une valeur d'enrichissement de plus de 800% après une double sélection stringente, soit près de 2 fois la valeur obtenue après une sélection stringente unique. Cette valeur doit-être mise en relation à la valeur du 1-PDP obtenue puisque, en effet, l'enrichissement du dérivé Pyridostatine baisse après double sélection. Enfin, les valeurs d'enrichissements du 3-360A-s vis-à-vis des cibles contrôles sont remarquables puisque celles-ci sont extrêmement basses, voire négatives.

Après validation des composés d'intérêts issus de la première campagne de sélection et séquençage, l'objectif a ensuite été de synthétiser le 1-PDP, 2-PDC3 et 3-360A-s en quantité suffisante pour étudier l'affinité et la capacité stabilisatrice de ces molécules contre différentes cibles quadruplex.

Au-delà de ces 3 composés prometteurs, nous avons également sélectionné des composés ayant pour vocation de confirmer les tendances obtenues lors des séquençages. Nous avons par exemple sélectionné des molécules composées des motifs PDP et 360A en association avec des peptides non-enrichis afin de vérifier la corrélation entre enrichissement et affinité réelle. De même nous avons également synthétisé des versions tronquées en Cycle 4 des

composés d'intérêt 1-PDP et 3-360A-s, et ceux afin de confirmer l'importance de chaque bras peptidique. Le descriptif de ces séquences est fourni dans le chapitre suivant.

Au cours de ce chapitre, trois campagnes de sélections ont été mentionnées et seule la première campagne de sélections réalisée à partir de la DEL1 a été entièrement décrite. En effet, alors que la deuxième campagne de sélection a belle et bien été réalisée, certains résultats de séquençage n'ont pas été entièrement traités (TERRA, HIV-LTRIII, TWJ, i-motifs) au jour de la rédaction du manuscrit. Ainsi, l'analyse des données de séquençages ne figurent pas dans le document. De plus, la campagne de sélections réalisée à partir de la deuxième librairie encodée (DEL2) a été réalisée en parallèle de l'écriture de ce manuscrit. Le séquençage à haut-débit est en cours de réalisation.

Chapitre 4 — Synthèse des touches et contrôles identifiés et caractérisations biophysiques.

Partie 1 : Synthèse des composés peptido-organiques.

Les molécules d'intérêts sélectionnées après divers processus de sélections sont des composés peptido-organiques issus de l'assemblage combinatoire d'une librairie de ligands ciblant les G-quadruplex. Confectionnés par chimie combinatoire *Split and mix* au sein de la DEL1, ces ligands peptido-organiques ont été resynthétisés à l'échelle du milligramme afin de pouvoir effectuer plusieurs caractérisations biophysiques de leurs interactions avec les G4.

Les premiers ligands sélectionnés, **1-PDP**, **2-PDC3** et **3-360A-s**, possèdent une structure générale aux caractéristiques peptido-organiques (Figure 107). De fait, la confection de ces ligands a nécessité la synthèse de molécules organiques à système fusionné (Phénanthroline, **PDC3**) ou non-fusionné (Pyridostatine, **PDP** – Pyridocarboxamide, **360A**) mais également la synthèse de séquences peptidiques possédant jusqu'à douze acide aminés.



Figure 107 : Composition organo-peptidique des dérivés Pyridostatine 1-PDP et Pyridocarboxamide 3-360A-s.

Les motifs hétérocycliques fonctionnalisés **PDP**, **360A** et **PDC3** ont été synthétisés et obtenus par collaboration avec différents partenaires. Le dérivé pyrrolidine de la Pyridostatine (**PDP**³⁰⁹) a été synthétisée au laboratoire par C. Van Wesemael à partir de l'acide chelidamique (acide 4-hydroxypyridine-2, 6-dicarboxylique). Le ligand Pyridocarboxamide à chaîne courte³¹⁰ (**360A-s**) a été synthétisé par l'équipe du professeur J.-F. Riou (MNHN – UMR 7196 – Paris) et le dérivé Phénanthroline (**PDC3**³¹⁰) ainsi que le ligand Pyridocarboxamide à chaîne longue³¹¹ (**360A-l**) ont été synthétisés par l'équipe de de M.-P. Teulade-Fichou (Institut Curie — UMR 176 — Orsay).

La synthèse des peptides nécessaire à la formation de composés peptido-organiques a été effectuée au laboratoire suivant la stratégie de synthèse supportée en phase solide (en. *Solid Phase Peptide Synthesis*, SPPS). Cette méthode, dérivée de R. B. Merrifield³¹¹, est basée sur l'addition successive d'acides aminés par formation de liens peptidiques. L'élongation de la chaîne peptidique est réalisée sur billes de résine par cycles itératifs de déprotection/couplage, de la partie Carboxy-terminale (*C-term*) à la partie Amino-terminale (*N-term*, Figure 107).

La stratégie de synthèse supportée suivie est connue sous le nom de stratégie «Fmoc/tBu ». Celle-ci consiste en l'incorporation d'acides aminés dont l'amine en α est protégée par le

groupement protecteur fluorénylméthoxycarbonyle (Fmoc). Lors de cette synthèse, les étapes de couplages sont réalisées en présence d'un agent activateur (HBTU) et d'une base (DIEA). L'HBTU, agent de couplage de la famille des Uroniums, joue un double rôle au sein de la synthèse (robotisée) de peptide. Premièrement, HBTU est utilisé pour ces propriétés limitatrices de racémisation³¹². D'autre part, HBTU est capable d'activer l'acide carboxylique en formant un ester d'hydroxybenzotriazole (HOBt) *in situ* (Schéma 11, 2. Activation). L'ester d'HOBt est un excellent groupe partant capable d'être déplacé par aminolyse afin de former le lien amide souhaité³¹³ (Schéma 11, 3. Couplage). Dans l'intérêt d'incorporer l'acide aminé suivant (AA [R₂]), une étape de déprotection du groupement protecteur Fmoc de l'acide aminé AA [R₁] est nécessaire. Cette déprotection est réalisée en présence de pipéridine (20%) dans la NMP (ou DMF) afin de libérer le dibenzofulvène (DBF) par β-élimination (Schéma 11, 1. Déprotection).

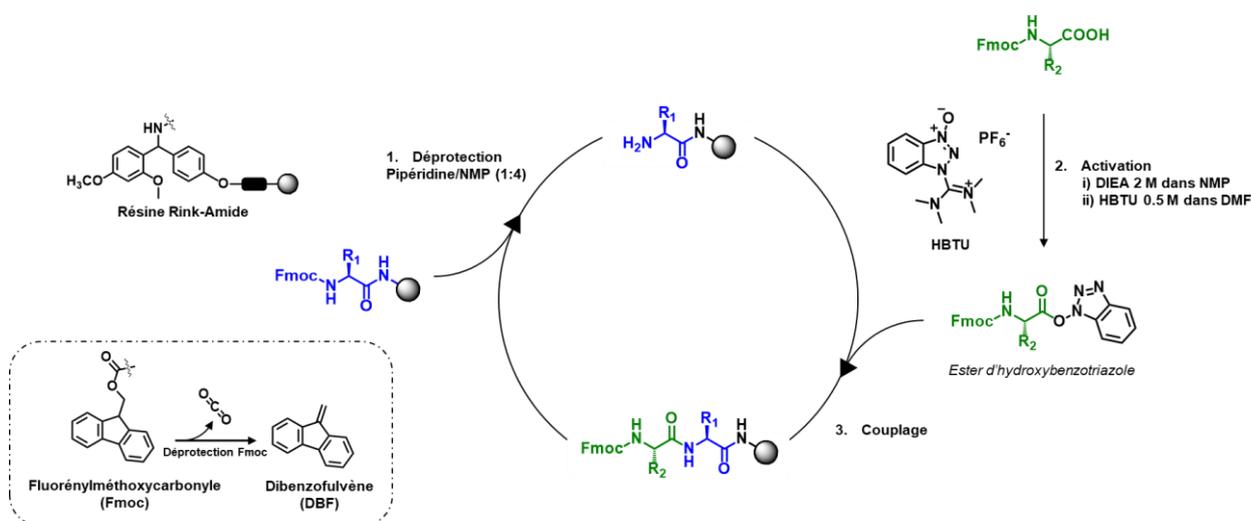


Schéma 11 : Synthèse peptidique supportée sur phase solide. 1/ Déprotection. 2/ Activation. 3/ Couplage. Résine utilisée dans pour la synthèse de peptides : Rink-Amide.

Certains des acides aminés naturels utilisés lors des synthèses possèdent des chaînes latérales capables d'interférer dans le processus d'élongation, à l'exemple du tryptophane (W) ou encore de l'arginine (R). Des acides aminés à chaînes latérales protégées par un groupement acido-labiles ont alors été utilisés (R-N_wPbf et W-N_{in}-Boc) afin de perpétuer l'orthogonalité nécessaire à la stratégie « Fmoc/tBu ».

Comme indiqué précédemment, la synthèse a été effectuée sur des billes de résines préfonctionnalisées par un espaceur chimique (en. linkers). Différents types de linkers existent sur le marché et leur utilisation va dépendre de la stratégie de synthèse utilisé ainsi que de la fonction chimique souhaitée en C-term. Selon la stratégie « Fmoc/tBu », il est possible d'utiliser des résines acido-labiles capables d'être déprotégées en présence d'acide trifluoroacétique (TFA) comme la résine Wang ou encore la 2 — Chlorotriptyl. L'utilisation de ces dernières permettent d'obtenir un peptide qui, après clivage, exhibe une fonction acide carboxylique libre en partie C-term. Au sein de la DEL1, les molécules peptido-organiques possèdent une partie C-term inerte dû à la présence de l'espaceur bifonctionnel, l'Headpiece.

Ainsi, dans le but d'obtenir une fonction chimique inerte en partie *C-term* du peptide, l'utilisation des résines « Amides » a été retenue. Parmi les résines délivrant une fonction amide en partie *C-term*, il est possible de retrouver, par exemple, la résine Rink-Amide ou encore la résine Sieber-Amide. Ces deux résines présentent des différences structurales au niveau de leur *linker* et nécessitent des conditions de clivage distinctes. Alors que la résine Sieber nécessite des conditions douces de décrochage (1% TFA dans le dichlorométhane), la résine Rink-Amide sollicite une solution de TFA à 95% pour décrocher un peptide de la résine. Compte tenu des groupements protecteurs présents sur certains des acides aminés nécessitant également des conditions de déprotection à 95% de TFA (R-N_wPbf) le choix de résine s'est porté sur l'utilisation de la Rink-Amide. Sans rentrer dans d'avantages de considérations, la résine utilisée est disponible commercialement chez Iris Biotech sous le nom de résine *Fmoc*-Rink Amide PEG MBHA (Figure 108). L'intérêt de cette Rink-Amide particulière est de porter au cœur de l'édifice des motifs polyéthylène glycols (PEG) capable à la fois de limiter les effets d'agrégations des peptides en augmentant l'hydrophilicité de l'édifice mais également d'augmenter le gonflement de la résine dans les solvants de synthèse. L'augmentation du gonflement permet ainsi d'augmenter les rendements de couplages des acides aminés puisque les fonctions chimiques sont plus disponibles.

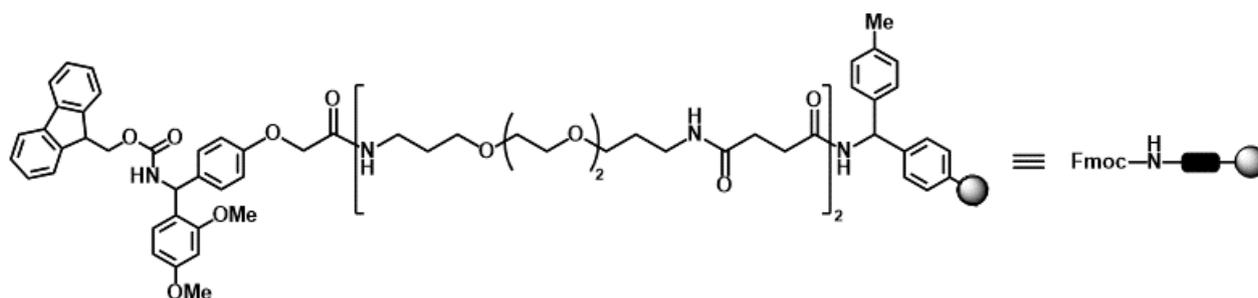


Figure 108 : Résine Rink-Amide Polyéthylène glycol (PEG) 4-mehtylbenzhydrylamine (MBHA) utilisée lors de la synthèse supportée sur phase solide des différentes séquences peptidiques.

Sous-partie 1.1 : Synthèse de séquences peptidiques.

Après traitement bio-informatique, les séquences peptidiques des ligands retenus pour leur enrichissement et/ou spécificité vis-à-vis de la cible quadruplex sont les séquences **1**, **2** et **3** respectives des ligands **1-PDP**, **2-PDC3** et **3-360A-s** :

- **1** : GRRWG-K^{N₃}-GRRAG
- **2** : GFRWG-K^{N₃}G-GPRFG
- **3** : GWRRG-K^{N₃}G-GRRAG

La synthèse des composés peptido-organiques a pour but d'étudier les propriétés physico-chimiques des ligands, en effectuant notamment l'étude des constantes de dissociations (K_d) ainsi que des propriétés stabilisatrices d'édifices G-quadruplex. L'évaluation de propriétés physico-chimiques de ligands nécessite également l'utilisation de contrôles, généralement les

ligands « parents » de structures G-quadruplex, à savoir la Pyridostatine tris-amino (**PDS**), la Phénanthroline (**PDC3**) et le Pyridocarboxamide (**360A**, Figure 109).

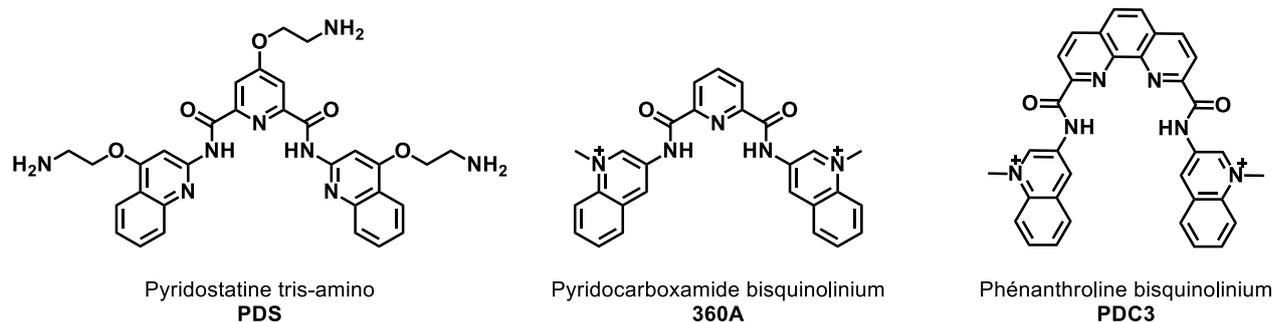


Figure 109 : Structure de ligands parents Pyridostatine, Pyridocarboxamide et Phénanthroline issus de la littérature.

Toutefois, d'autres molécules issues de la DEL1 ont été choisies en tant que contrôles négatifs. En effet, les composés **1-PDP**, **2-PDC3** et **3-360A-s** ont été choisis pour leurs valeurs d'enrichissement importantes vis-à-vis d'une, voire plusieurs cibles quadruplex. Le même travail a alors été effectué afin de trouver des molécules uniques, dérivants de la famille Pyridostatine et Pyridocarboxamide, avec des valeurs d'enrichissements bien inférieures à ceux des ligands sélectionnés. Ce travail a permis d'identifier deux ligands particulièrement déplétés contre les cibles quadruplex, le **4-PDP** et le **5-360A-s**. À titre d'exemple, alors que le ligand **1-PDP** est enrichi à plus de 1200% contre la cible quadruplex Myc en condition stringente (4^{ème} lavage à 70°C), le ligand **4-PDP** présente une déplétion de 41% dans les mêmes conditions. De manière similaire, le dérivé **5-360A-s** est principalement déplété contre les cibles quadruplex et également contre les cibles contrôles.

Les séquences peptidiques des ligands contrôles issus de la DEL1 sont listés ci-dessous :

- **4** : GLLRG-K^{N₃}G-GRFFG
- **5** : GFPRG-K^{N₃}G-GRLLG

Comparativement aux séquences peptidiques **1**, **2** et **3**, les motifs **4** et **5** comportent des pentapeptides du [Cycle 1 et 4] exclusivement mono-arginilées. D'autre part, compte tenu des séquences présentées, il est possible de constater la présence d'un acide aminé non-naturel, la lysine-azoture, également connue sous son nom formel d'azidonorleucine. Cette molécule a été synthétisée au laboratoire à partir de l'acide aminé Fmoc-Lys (N_εBoc) — OH dans les conditions de diazotransfert³¹⁴, après déprotection de l'amine primaire par élimination du groupement protecteur Boc. La réaction de déprotection du groupement protecteur Boc a été réalisée en présence d'acide chlorhydrique puis la réaction de diazotransfert a été réalisée en milieu basique en présence du réactif de diazotransfert (imidazole-1-sulfuryl azoture, ISAHC) et de sulfate de cuivre. La solution de réaction laissée sous agitation à température ambiante est ensuite extraite en milieu acide afin d'obtenir la Fmoc-azidonorleucine avec un rendement usuel de 60 à 80% (Schéma 12).

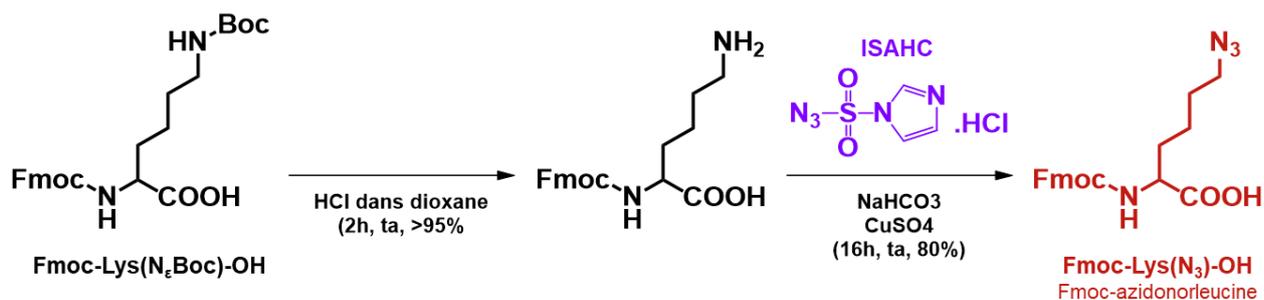


Schéma 12 : Synthèse de la Fmoc-azidonorleucine par réaction de diazotransfert catalysée par le cuivre.

Une fois l'acide aminé non naturel synthétisé, il a été possible d'engager les synthèses robotisées des séquences peptidiques **1**, **2**, **3**, **4** et **5**. Le Schéma 13 détaille la synthèse des différentes séquences peptidiques. Celles-ci ont été réalisées sur la résine Fmoc-Rink Amide PEG MBHA suivant la stratégie « Fmoc/tBu ». L'élongation des chaînes peptidiques ont été opérés sur synthétiseur automatique de peptide tel que présenté précédemment (Schéma 11). Il est à noter que contrairement aux séquences **2** à **5**, seul 11 acides aminés composent la séquence **1**, ainsi cette dernière a fait l'objet d'une synthèse en parallèle.

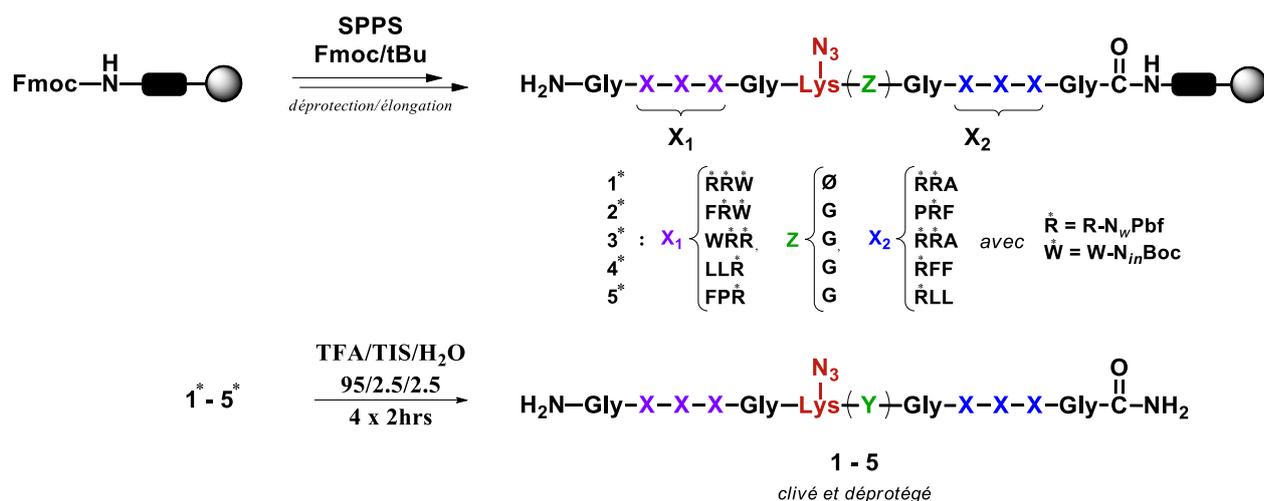


Schéma 13 : Synthèse SPPS des 5 séquences peptidiques.

En fin du processus d'élongation, les peptides linéaires sont clivés de la résine Rink Amide en milieu acide fort, en utilisant une solution de TFA/TIS/H₂O à un ratio de 95/2,5/2,5. Le TIS (Triisopropylsilane) est utilisé dans ce cas en tant piège à carbocation (en. *scavenger*). En effet, le TIS est une molécule capable d'interagir avec les carbocations formés *in situ* après déprotection des fonctions Pbf et Boc, dans l'intérêt de limiter les réactions secondaires non désirables. Les peptides alors clivés et déprotégés ont été obtenus par précipitation dans l'éther, puis purifiés sur colonne greffée C18 semi-préparative en phase inverse avec des rendements variants de 40% à 70%.

La confection de la première librairie combinatoire, DEL1, de molécules a nécessité l'assemblage itératif de 4 motifs uniques afin de former une bibliothèque de près de 200k composés. L'insertion des deux premiers motifs de la DEL1 a pu être caractérisé par UPLC-MS et la caractérisation du troisième cycle par la même méthode d'analyse s'est avéré

complicé dû à la présence, en mélange, de plus de 4k composés. Finalement, l'insertion du dernier motif, le pentapeptide [*Cycle 4*], n'a pas pu être réellement caractérisée. Une purification a été effectuée par la méthode de « Cap&Catch » et une analyse de la pureté a été réalisée par gel retard. Néanmoins, il est possible d'envisager qu'une fraction des ligands, n'ayant pas réagi lors du cap&catch, présente une délétion du [*Cycle 4*]. Si tel était le cas, les ligands sélectionnés par la méthode de sélection par affinité sur billes magnétiques pourraient être des ligands tronqués. Afin de vérifier que les composés peptido-organiques sélectionnés possèdent les motifs des quatre cycles de la chimie combinatoire, il a été proposé de synthétiser les dérivés tronqués du **1-PDP** ainsi que du **3-360A-s** (Figure 110). Si l'étude biophysique des composés tronqués montre une affinité supérieure ou égale pour les ligands présentant une délétion du [*Cycle 4*] que pour les composés originaux de la librairie, il serait possible de conclure sur la présence de composés tronqués au sein de la librairie combinatoire. Les molécules tronquées ont ainsi nécessité la synthèse de nouvelles séquences peptidiques appelées **t1** et **t3**, correspondant aux dérivés tronqués des peptides **1** et **3**. Conjointement au processus décrit précédemment, les peptides **t1** et **t3** de tailles et séquences d'acides aminés différentes ont été synthétisés sur support solide à partir de la Fmoc-Rink Amide PEG MBHA selon la stratégie « Fmoc/tBu ».

Il est également à noter que la synthèse des peptides **t1** et **4** et des dérivés PDP correspondant, à savoir **t1-PDP** et **4-PDP** a fait l'objet d'un stage de L3 et a été réalisée par Samantha Nichillo. De façon similaire, les peptides **t3** et **5**, ainsi que les dérivés **360A-s**, **t3-360A-s** et **5-360A-s** a fait l'objet d'un projet de stage de M1. Ce projet a été réalisé par Rémy Gay et complété par de la caractérisation biophysique qui sera présentée plus tard dans ce chapitre.

Finalement, 7 séquences peptidiques distinctes ont été synthétisées au cours de cette thèse. Parmi celles-ci on retrouve les séquences des composés sélectionnés après séquençage à haut débit (**1**, **2** et **3**), ainsi que celles des contrôles négatifs présentant des enrichissements faibles voire négatifs (**4** et **5**). Enfin, des peptides contrôles de l'incorporation des derniers motifs du [*Cycle 4*] de la chimie combinatoire ont été préparés avec les deux dérivés tronqués (**t1** et **t3**).

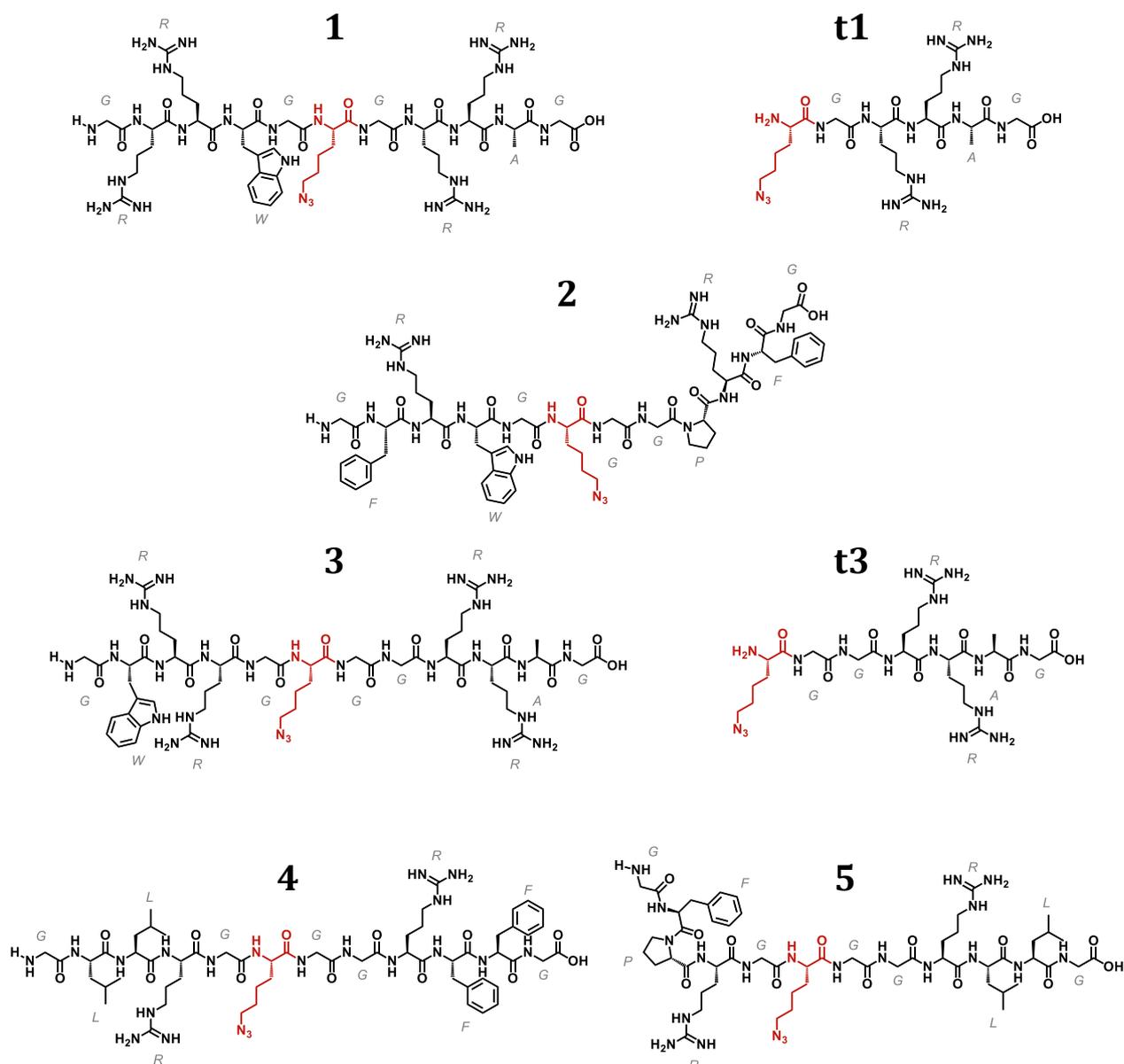


Figure 110 : Structures développées des 5 séquences peptidiques et de leurs tronqués (t1 et t3).

Sous-partie 1.2 : Synthèse des composés peptido-organiques.

Les composés peptido-organiques retenus pour les études biophysiques sont issus de la famille **PDS**, **360A** et **PDC3**. Au même titre que l'incorporation des ligands de structures quadruplex au sein de la chimie combinatoire encodée ([*Cycle 3*]), ces ligands ont été couplés aux séquences peptidiques par ligation chimio-sélection catalysée par le cuivre, la cycloaddition 1,3 — dipolaire. Initialement décrite par Huisgen^{315,316} puis adaptée par Meldal³¹⁷ et Sharpless³¹⁸, la cycloaddition azoture-alcyne cupro-catalysée par le Cuivre (en. *Copper(I)-catalyzed alkyne-azide cycloaddition*, CuAAC) est une ligation de choix grandement utilisée en chimie, biochimie et biologie. Alors qu'aucune certitude dans le mécanisme de la réaction n'ait été acquise jusqu'alors, de nombreux mécanismes ont été proposés dans la littérature. Initialement, les études de la cycloaddition [3 + 2] ont montré la participation unique d'un ion cuivreux. Au début des années 2000, d'autres groupes de recherches ont proposés

l'intervention de deux ions cuivreux dans le mécanisme réactionnel afin d'influencer favorablement le profil d'énergie de la réaction^{319,320}. Toutefois, quel que soit les mécanismes proposés, la CuAAC nécessite l'utilisation d'une source de cuivre (sous forme d'ion cuivreux) afin de conduire exclusivement à la formation d'un triazole 1,4 — disubstituée (Schéma 14).

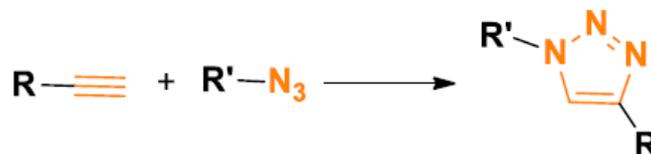


Schéma 14 : Réaction de cycloaddition 1,3 — dipolaire catalysée par Cu (I). Cycloaddition d'un alcyne et d'un azote afin de former le 1,4 — triazole.

Les sources de cuivre peuvent varier suivant les conditions expérimentales utilisées. Néanmoins, dans l'intérêt de limiter la sensibilité de la réaction face à la présence d'oxygène, l'utilisation d'une source d'ions cuivrique est favorisée par rapport à l'utilisation d'une source d'ion cuivreux³²¹.

Les réactions de cycloaddition alors initiées au laboratoire ont été effectuées à partir de sulfate de cuivre (pentahydraté), générant une source d'ions cuivriques. Toutefois, la CuAAC telle que décrite par Huisgen fait intervenir un ion cuivre mono-oxydé. Ainsi, les ions Cu(II) provenant du sulfate de cuivre ont été réduits *in situ* par l'ajout d'un agent réducteur, l'Ascorbate de sodium (NaAsc), capable de maintenir l'état d'oxydation du cuivre tout au long la réaction. Les réactions de click réalisées ont été effectuées au sein d'un mélange hétérogène (eau-solvant organique) pour des questions de solubilité à la fois des peptides mais également des ligands de structures G-quadruplex. Dans ces conditions, un ion cuivrique non stabilisé serait capable d'être oxydé par son environnement. Ainsi, dans l'intérêt de stabiliser le cuivre sous forme Cu(I), un ligand stabilisateur a été utilisé à l'exemple du tris (benzyltriazolylmethyl) amine (TBTA) ou le tris (hydroxypropyltriazolyl methyl) amine (THPTA). L'avantage du ligand THPTA est d'être plus hydrosoluble que son homologue plus hydrophobe, le TBTA.

Une considération importante est à prendre en compte lorsque l'Ascorbate de sodium est utilisé en tant qu'agent réducteur de Cu(II) dans une réaction de chimie click faisant intervenir des partenaires contenant un (ou plusieurs) acide aminé Arginine (R). En effet, l'arginine est un acide aminé portant une chaîne latérale composée d'un motif guanidine capable d'interagir avec la forme oxydée de l'Ascorbate de sodium (Schéma 15). La réaction d'oxydation de l'Ascorbate de sodium conduit à l'acide déhydroxyascorbique (DHA) capable, dans un milieu aqueux, de former l'hémicétal correspondant. La partie γ -lactone de l'hémicétal DHA peut ensuite réagir avec un motif guanidine, la chaîne latérale de l'arginine, afin de former un adduit DHA³²². Afin de limiter l'apparition de ce produit secondaire covalent, l'aminoguanidine est ajoutée en large excès au milieu réactionnel afin de prévenir la formation de l'adduit DHA sur les chaînes latérales de l'arginine.

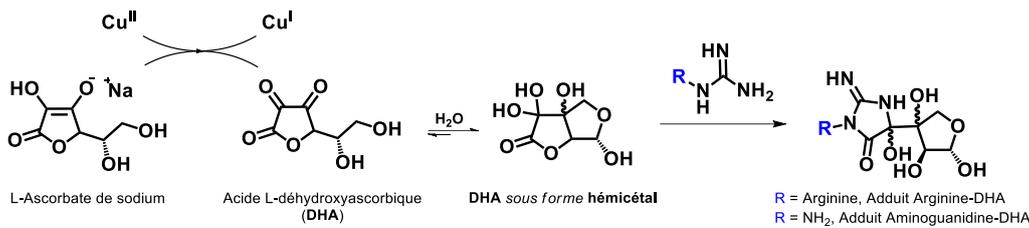


Schéma 15 : Réaction secondaire de la forme oxydée de l'Ascorbate de sodium, la DHA, avec les fonctions guanidines, par exemple, des arginines.

Comme indiqué auparavant, le ligand Pyridostatine (**PDP**), les deux ligands Pyridocarboxamides (chaîne courte — **360A-s** — et chaîne longue — **360A-l**) ainsi que et le ligand Phénanthroline (**PDC3**) ont été obtenus en collaboration. Les réactions de cycloaddition [3 + 2] entre les peptides (10mg) et les ligands de structures G-quadruplex (0.9eq) ont alors été effectuées au laboratoire dans un milieu hétérogène DMSO/H₂O (4 :1), tamponné à pH 7.4 dans l'HEPES, en présence d'Ascorbate de sodium (15 eq.), de Sulfate de cuivre (1.5 eq.), de THPTA (3 eq.) et d'Aminoguanidine (15 eq.). Après analyse de la complétion des réactions par HPLC, une solution saline de Na₂EDTA a été rajouté dans aux milieux réactionnels afin de piéger les ions cuivres résiduels. Les produits bruts ont alors été purifiés sur HPLC en phase inverse et analysés par HR-MS.

Les réactions de click cupro-catalysées ont ainsi permis d'obtenir les dérivés Pyridostatines **1-PDP**, **t1-PDP** et **4-PDP**, le dérivé Phénanthroline **2-PDC3** ainsi que les dérivés Pyridocarboxamides **3-360A-s**, **t3-360A-s** et **5-360A-s** (Figure 111).

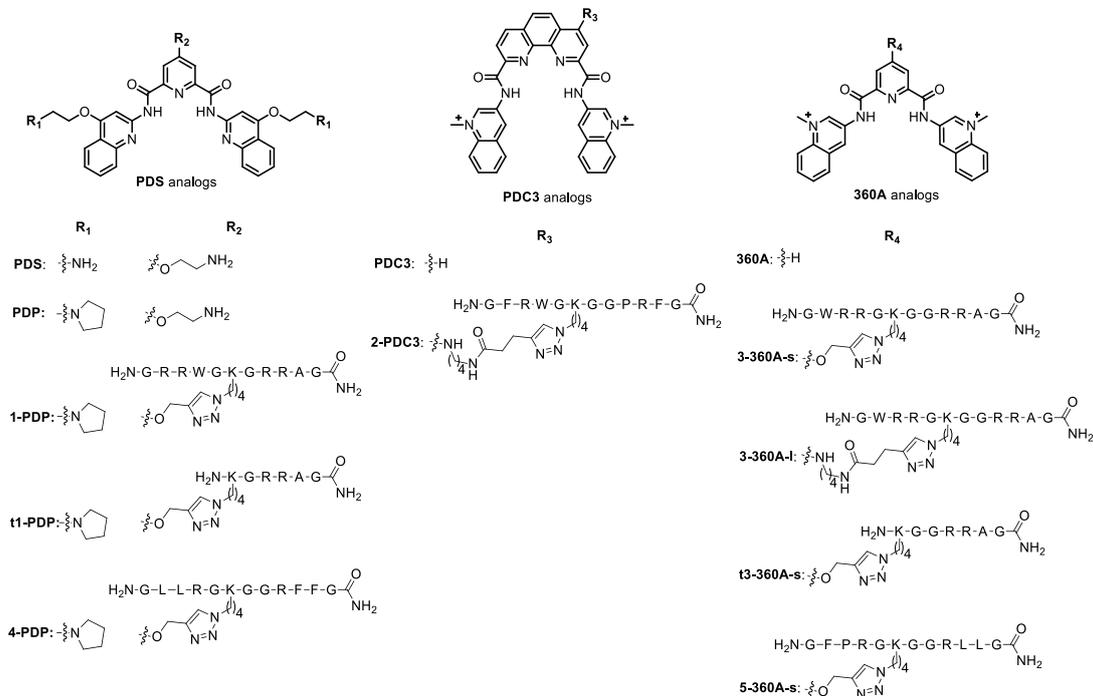


Figure 111 : Représentation exhaustive de tous les composés synthétisés. Dans le cas des familles Pyridostatine et Pyridocarboxamide : synthèse du dérivé enrichi 1-PDP et 3-360A-s respectivement ainsi que leurs tronqués t1-PDP et t3-360A-s. Synthèses des deux dérivés non-enrichis 4-PDP et 5-360A-s. Dans la famille Phénanthroline : synthèse d'un seul composé 2-PDC3.

La Figure 111 décrit une nouvelle molécule appelée **3-360A-I**, composé de la séquence peptidique **3** et du ligand Pyridocarboxamide à chaîne longue, le **360A-I**. En effet, l'analyse des enrichissements des molécules uniques au sein de la librairie combinatoire a montré que la longueur de chaîne différenciant le ligand 360A-s du 360A-I avait une importance dans l'affinité des édifices pour la cible considérée. En comparant les valeurs d'enrichissement (Tableau 20) des deux composés peptido-organiques, il est possible de remarquer que l'augmentation de la longueur de chaîne entraîne une baisse d'enrichissement contre toutes les cibles G-quadruplex utilisées (à l'exception de la cible 22CTA). Le composé **3-360A-I** a donc été synthétisé afin de vérifier, en comparaison avec le **3-360A-s**, si les valeurs d'enrichissement obtenues par séquençage à haut-débit sont des bons indicateurs de la capacité, ou non, d'un ligand à reconnaître voire à stabiliser le G-quadruplex ciblé.

Tableau 20 : Enrichissement du dérivé 3-360A à chaîne courte (-s) et chaîne longue (-I) en fonction des cibles utilisés.

	Myc @w25 °C	Myc @w70 °C	Myc @double Sel	Myc22 @w70 °C	wtTel26 @w25 °C	22CTA @w25 °C	Kit1 @w25 °C	hpl @w25 °C	scr26 @w25 °C	Strept. @w25 °C
3-360A-s % enrich.	351%	438%	826%	600%	67%	86%	263%	17%	-3%	-59%
3-360A-I % enrich.	187%	251%	360%	150%	56%	96%	147%	4%	-50%	-38%

Finalement, 8 composés peptido-organiques ont donc été obtenus après cycloaddition [3 +2] entre l'azidonorleucine des séquences peptidiques et la fonction alcyne portée par les dérivés de **PDC3**, **360A(-s et -I)** et **PDP**. Une fois obtenus, ces ligands ont été soumis à des caractérisations biophysiques afin de déterminer leur constante d'affinité pour les différents G-quadruplex utilisés comme cibles ainsi que leur propriété stabilisatrice de structure G-quadruplex.

Partie 2 : Etudes biophysiques : stabilisation et affinité.

Les propriétés d'interactions entre des cibles G-quadruplex et des ligands peuvent être étudiées par de nombreuses méthodes physico-chimiques afin de déterminer de multiples paramètres tels que, les valeurs thermodynamiques et cinétiques de la formation du complexe, la stoechiométrie d'interactions ou encore les capacités de stabilisation thermique. Parmi les nombreuses approches utilisées de manière régulière pour caractériser les interactions ligands/G-quadruplex nous avons choisi d'explorer les techniques de FRET-Melting, d'extinction de fluorescence par titration et la technique optique de Bio-Layer Interferometry (BLI). Ces méthodes et les résultats obtenus sont décrits dans cette partie.

Sous-partie 2.1 : FRET-Melting.

L'expérience de FRET (en. *Förster Resonance Energy Transfer*) est définie par « un processus physique distance-dépendant par lequel l'énergie non-radiative est transférée d'une molécule fluorescente excitée (le donneur) à un autre composé fluorescente (l'accepteur) au moyen d'une interaction dipôle-dipôle intermoléculaire à longue portée »³²³. Le FRET-Melting associe ainsi le principe de FRET avec une variation progressive de la température permettant

ainsi de mesurer la température de dénaturation d'une structure G-quadruplex et d'évaluer la stabilisation thermique induite par un ligand en interaction avec sa cible. Cette méthode d'analyse, largement décrite et développée par l'équipe de Mergny^{324,325} dans le cadre des G-quadruplexes, fait ainsi intervenir des séquences oligonucléotidiques doublement fonctionnalisées par un fluorophore, dont l'un va servir de donneur de « fluorescence » et l'autre d'accepteur de « fluorescence ». Afin que le phénomène physique de transfert d'énergie entre le donneur et l'accepteur puisse être réalisée, il est nécessaire que le spectre d'émission du donneur puisse recouvrir, partiellement tout au moins, le spectre d'absorption de l'accepteur. Ainsi, les paires donneur-accepteur doivent être choisies avec soin afin de pouvoir visualiser le phénomène de FRET.

Plusieurs couples de fluorophores peuvent être utilisés afin de distinguer ce phénomène (Alexa488/Cy5, Cy3/Cy5). Toutefois, dans le cadre de l'évaluation du processus physique sur les édifices G-quadruplex, c'est la paire fluorescente composée du donneur FAM (carboxyfluorescéine, fonctionnalisation 5') ainsi que de l'accepteur TAMRA (carboxytétraméthylrhodamine, fonctionnalisation 3') qui a été retenue³²⁵ (Figure 112). Le transfert d'énergie de type Förster entre la Fluorescéine (FAM), excité à 488nm, et la Rhodamine (TAMRA), émettant à 588nm, n'est en théorie efficace que si la distance entre les deux fluorophores est suffisamment proche ($1/R^6$)³²⁶. A température ambiante, lorsque la séquence oligonucléotidique forme un G-quadruplex, le donneur et l'accepteur sont suffisamment proches pour visualiser l'effet de FRET (Figure 112). En revanche lorsqu'un édifice G-quadruplex doublement marqué est dénaturé thermiquement, la distance entre les deux fluorophores augmente et l'émission de l'accepteur ne sera plus permise. Dans ce cas, la Fluorescéine alors excitée à 488nm va pouvoir émettre sa propre fluorescence à 515nm (Figure 112).

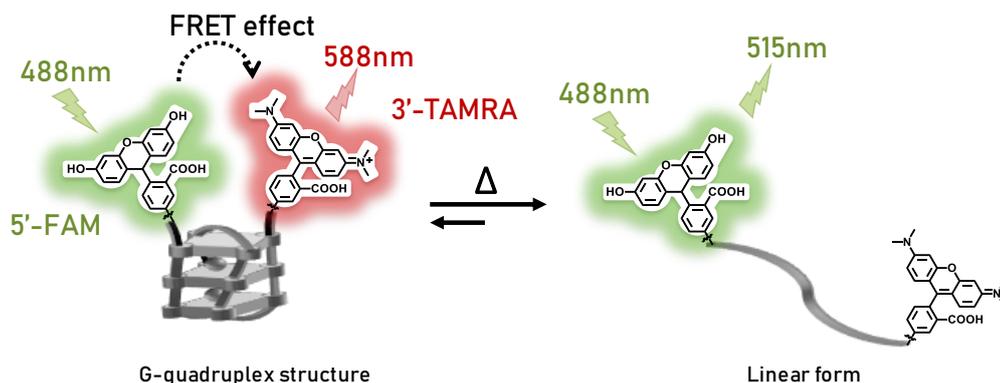
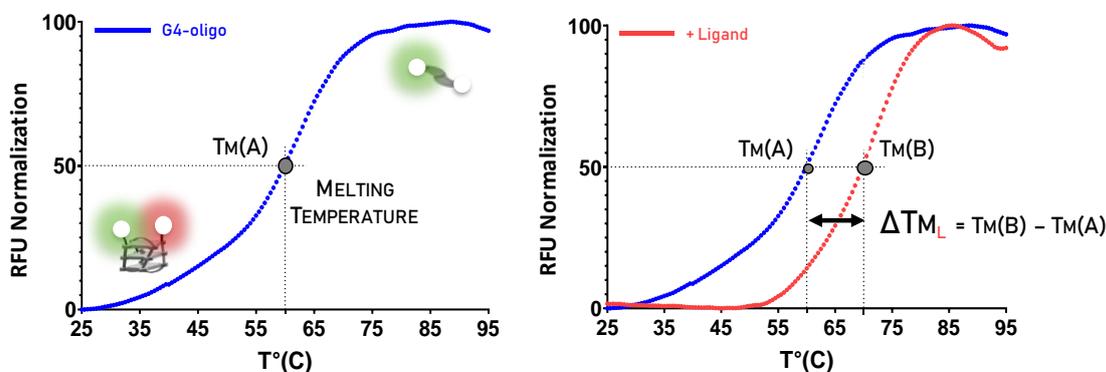


Figure 112 : Phénomène de FRET illustré au sein d'une séquence G-quadruplex doublement marquée par une Fluorescéine (FAM) et une Rhodamine (TAMRA). Dissipation de l'effet FRET par dénaturation thermique.

Dans les expériences de FRET-Melting réalisées au laboratoire les données de fluorescences enregistrées sont celles observées par émission de la fluorescéine (FAM) en 5' des séquences nucléiques. L'augmentation linéaire de la température va ainsi permettre d'augmenter la proportion de séquence simple brin doublement marqué. Le Graphique 22 (gauche) montre l'allure typique des courbes obtenues après expérience de FRET-Melting. Cette courbe

exprime la variation de fluorescence normalisée en fonction de l'augmentation graduelle de la température (25°C à 95°C). La température de fusion (en. *melting temperature*, T_m ou $T_{1/2}$) d'une structure quadruplex correspond à la température pour laquelle la moitié de la séquence doublement marqué est sous forme linéaire. Elle peut être déterminée pour une valeur de 0,5 (ou 50%) après normalisation des valeurs de fluorescences.



Graphique 22 : Courbe de fluorescence normalisée en fonction de la température. Expérience réalisée par qPCR par mesure de la fluorescence d'émission du FAM. Caractérisation de la température de dénaturation (T_m) de l'oligonucléotide en absence (gauche) ou en présence de ligand stabilisateur (droite). Capacité stabilisateur d'un ligand mesuré par calcul du ΔT_m .

Lorsqu'une séquence G-quadruplex forme un complexe avec un ligand stabilisateur, la séquence nécessite une énergie thermique plus importante afin d'atteindre la valeur du T_m . Dans le cas illustratif du Graphique 22 (droite), il est possible de constater que le T_m de la séquence G-quadruplex, noté T_m (A), est bien inférieure au T_m obtenue en présence d'un ligand, T_m (B). L'effet du ligand sur la stabilisation de l'édifice G-quadruplex (ΔT_m) peut alors être calculé en soustrayant les deux valeurs de T_m obtenues tel que $\Delta T_{mL} = T_m$ (B) — T_m (A).

Les expériences de FRET-Melting ont été utilisées en routine au laboratoire afin d'évaluer les capacités stabilisatrices des composés nouvellement synthétisés. Cette méthode rapide et facile à mettre en œuvre a été réalisée sur un appareil de qPCR, permettant de suivre en temps réel l'effet de la dénaturation thermique sur l'intensité de la fluorescence. La réalisation d'expériences de FRET-Melting sur qPCR a été effectuée à haut débit (plaque 96-puits), permettant ainsi de comparer les informations de nombreuses conditions (ligand, oligonucléotide, compétiteur) en parallèles.

Pour une expérience classique de FRET-Melting, la concentration d'oligonucléotide doublement marqué a été fixée à 0,2 μ M dans un tampon Cacodylate de Lithium. Conjointement aux expériences de sélections par affinités, les cibles oligonucléotidiques doublement marqué utilisées sont issues des séquences promotrices d'oncogènes (Myc, Myc22, Kit1), de la séquence télomérique (22CTA, wtTel26) et d'une séquence oligonucléotide contrôle capable de former un Hairpin (duplexe) d'ADN (hp2). Les séquences discutées sont décrites dans le Tableau 21 ci-dessous.

Tableau 21 : Séquences G-quadruplex FAM/TAMRA utilisées lors des expériences de FRET-Melting.

5' – F-oligo-T-3'	Séquences	Description
5' – F-Myc-T-3'	5'-/56-FAM/GGAGGGTGGGGAGGGTGGGGAA/36-TAMRA/-3'	G4 (parallèle), promoteur c-myc
5' – F-Myc22-T-3'	5'-/56-FAM/TTGAGGGTGGGTAGGGTGGGTAA/36-TAMRA/-3'	G4 (parallèle), promoteur c-myc avec substitution de G en T
5' – F-Kit1-T-3'	5'-/56-FAM/AGGGAGGGCGCTGGGAGGAGGGTT/36-TAMRA/-3'	G4 (antiparallèle), ADN télomérique humain (variant)
5' – F-22CTA-T-3'	5'-/56-FAM/AGGGCTAGGGCTAGGGCTAGGGTT/36-TAMRA/-3'	G4 (parallèle avec G isolé), promoteur c-kit1
5' – F-wtTel26-T-3'	5'-/56-FAM/TTAGGGTTAGGGTTAGGGTTAGGGTT/36-TAMRA/-3'	G4 (Hybride), ADN télomérique humain
5' – F-hp2-T-3'	5' — /56-FAM/ATAGCATACGTTTTCGTATGCTATT/36-TAMRA/-3'	Duplexe d'ADN (<i>Hairpin</i>)

Le tampon d'expérimentation (i. e. *tampon FRET*) est composé de Cacodylate de Lithium (LiAsO₂Me₂, 10mM, pH 7.2), de sels de potassium (KCl, X mM) et de sels de lithium (LiCl, 100-X mM)³²⁷. La charge ionique du tampon est fixée à 100 mM et la proportion des deux ions en présence dépend de la stabilité de la structure quadruplex cherchée. En effet, les ions K⁺ sont capables de coordonner les tétrades de guanines et donc de stabiliser l'édifice G-quadruplex. Au contraire, les ions Li⁺ disposent d'un rayon atomique ne permettant pas de stabiliser l'édifice. Ainsi, plus le tampon dispose d'une concentration élevée en potassium, plus le T_m de la structure quadruplex sera élevée. À l'exemple des deux séquences proto-oncogènes Myc et Myc22, l'application d'une concentration en KCl à 1 mM permet d'obtenir des T_m respectifs de 51,1°C et 59°C. Si la concentration d'ions K⁺ atteint 10 mM lors de l'expérimentation, les T_m augmentent jusqu'à 65,5°C et 70,7°C, respectivement. Les conditions de travail, et donc de concentration des deux ions en présence, ont été choisies en fonction de la valeur du T_m obtenue pour chaque séquence oligonucléotidique. En effet, il est important d'avoir des T_m homogènes entre les séquences quadruplex afin de les comparer. De plus, les valeurs ne doivent pas être initialement trop élevées afin de pouvoir mesurer avec confiance les valeurs de T_m obtenues après stabilisation par des ligands. Les ratios de concentrations en KCl et LiCl ainsi que les valeurs de T_m obtenues ont été recensées dans le Tableau 22 ci-dessous.

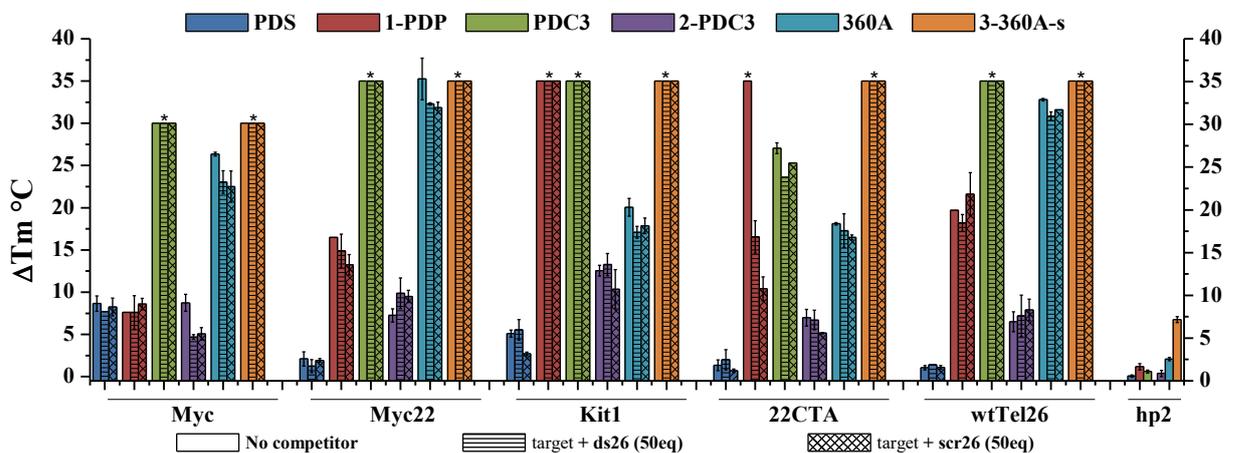
Tableau 22 : Mesure de la température de dénaturation des cibles FAM/TAMRA, en absence ou en présence de compétiteurs, à une concentration saline fixé.

FAM-Oligo-TAMRA	[KCl/LiCl] mM	T _m cible (°C)	T _m cible + scr26 (°C)	T _m cible + ds26 (°C)
F-Myc-T	1/99	57,8	60,8	61,3
F-Myc22-T	1/99	49,2	52,3	51
F-Kit1-T	1/99	50	52,7	50,2
F-22CTA-T	1/99	49,8	50,2	49,7
F-wtTel26-T	100/0	50,5	51,7	51,3
F-hp2-T	1/99	62,6	-	-

Les expériences de FRET-Melting ont également permis d'effectuer des expériences de compétitions afin de déterminer la sélectivité des ligands vis-à-vis de certaines séquences contrôles. Parmi celles-ci, il est possible de retrouver les séquences non-marquées scr26 (5'—AAGTGTGTGTGTGTGTGTGTGAAG-3') et ds26 (5'—CAATCGGATCGAATTCGATCCGATT G-3') duplexe d'ADN. Afin de réaliser ces expériences de compétitions, l'oligonucléotide G-quadruplex (0,2 μ M) a été incubé en présence de 50 eq d'une des séquences contrôles (scr26/ ds26, 10 μ M). L'évaluation des T_m dans ces conditions a montré que l'ajout de compétiteur joue un rôle dans l'analyse des valeurs de stabilisation. Ainsi, le calcul du ΔTm_L des ligands au sein d'expériences de compétitions va devoir prendre en compte les nouvelles valeurs de T_m obtenu pour l'oligonucléotide en présence de l'une ou l'autre séquence compétitrice.

Section 2.2.1 : Effet stabilisant de PDS, PDC3, 360A et leurs dérivés peptido-organiques.

Les expériences de FRET-Melting ont été réalisées au sein de plaques 96-puits. Chaque puits est composé d'une des séquences doublement marqué (5'—F-oligo-T-3', 0,2 μ M dans le tampon FRET), du ligand à étudier (5eq, 1 μ M), et éventuellement du compétiteur (50eq, 10 μ M), pour un volume final de 25 μ L. La mesure de fluorescence lié à la dénaturation thermique de l'oligonucléotide FAM/TAMRA a été effectuée toutes les minutes, après augmentation de 0,5 $^{\circ}$ C, de 25 $^{\circ}$ C à 95 $^{\circ}$ C. Les données de fluorescence acquises ont permis de caractériser le T_m 1) de l'oligonucléotide en absence de ligand, 2) de l'oligonucléotide en interaction avec un ligand et 3) de l'oligonucléotide et d'un compétiteur en interactions avec le ligand. Les valeurs reportées sur l'histogramme (Figure X) représentent les ΔTm_L calculés à partir des données obtenues.



Graphique 23 : Représentation des valeurs de stabilisation (ΔTm $^{\circ}$ C) de chaque ligands en fonction de la cible considérée, en absence ou présence de compétiteurs (ds26 ou scr26).

L'analyse des données de stabilisation obtenues (Graphique 23) permet de comparer l'effet stabilisateur du **1-PDP**, du **2-PDC3** ainsi que du **3-360A-s** en fonction de leur composé parent le **PDS**, le **PDC3** et le **360 A**.

Paragraphe 2.2.1.a) Stabilisation des composés vis-à-vis des proto-oncogènes Myc et Kit.

L'étude de la capacité de stabilisation des 6 ligands contre la cible Myc a montré que le ligand **PDC3** et le ligand **3-360A-s** stabilisaient très fortement le G-quadruplex Myc avec une valeur ΔTm estimée (symbolisée par un astérisque) à plus de 30°C. Pour ces composés, il n'a pas été possible de déterminer précisément une valeur de ΔTm dans la mesure ou le plateau de fluorescence maximum n'a pas été atteint à la température maximum, indiquant ainsi la coexistence de séquences oligonucléotides dénaturés et sous forme G-quadruplex. Les astérisques (*) caractérisent donc les valeurs de stabilisation fixé arbitrairement au maximum mesurable et sont donc sous estimées.

Le dérivé **2-PDC3** ($\Delta Tm_{2-PDC3/Myc} = 9^\circ\text{C}$) issu de la DEL1 apparait nettement moins stabilisant que son dérivé parent, alors que le **360A** ($\Delta Tm_{360A/Myc} = 27^\circ\text{C}$) stabilise moins la séquence Myc que le dérivé **3-360A-s**. Les ligands Pyridostatines **PDS** et **1-PDP** offrent des propriétés stabilisatrices similaires avec des valeurs $\Delta Tm_{PDS|1-PDP/Myc}$ avoisinant les 7°C.

De grandes similitudes ont pu être observées pour la séquence Myc22 et la séquence Kit1. En effet, les ligands **PDC3** et **3-360A-s** stabilise une nouvelle fois extrêmement bien ces deux séquences ($\Delta Tm_{PDC3|3-360A-s/Myc22|Kit1} > 35^\circ\text{C}$), alors que la stabilisation du **2-PDC3** ($\Delta Tm_{2-PDC3/Myc22|Kit1} \sim 13^\circ\text{C}$) reste bien inférieure à son composé parent. Enfin, malgré des valeurs stabilisatrices importantes ($\Delta Tm_{360A/Myc22} = 35^\circ\text{C}$, et $\Delta Tm_{360A/Kit} = 20^\circ\text{C}$) le **360A** reste moins stabilisateur que son dérivé peptidique **3-360A-s**. De manière surprenante, alors que le **PDS** et le 1-PDP présentent les mêmes capacités de stabilisation contre la cible Myc, c'est le **1-PDP** qui a montré la meilleure capacité stabilisatrice contre la cible Myc22 ($\Delta Tm_{1-PDP/Myc22} = 16^\circ\text{C}$) jusqu'à atteindre une valeur estimée de stabilisation de plus de 35°C contre la cible Kit1 ($\Delta Tm_{1-PDP/Kit1} > 35^\circ\text{C}$).

Toutes les expériences décrites jusqu'alors ont également été effectuées en présence des séquences compétitrices non-marquées, le scr26 ou ds26. Dans le cas des composés dont les ΔTm ne sont pas estimés, l'effets des compétiteurs sont faibles mais présents. En effet, contre les séquences des proto-oncogènes les composés les plus affectés par l'effet des compétiteurs sont le **2-PDC3** et le **360A**. Il a été possible de constater que le scr26 et le ds26 engendre des baissent allant de 2 à 4°C en présence de ces composés. Vis-à-vis des ligands Pyridostatines, il est possible de constater c'est en présence de la séquence ds26 que le ΔTm chute le plus.

Paragraphe 2.2.1.b) Stabilisation des composés vis-à-vis des séquences télomériques.

Vis-à-vis de la séquence de topologie antiparallèle 22CTA, seuls les ligands **1-PDP** et **3-360A-s** ont montré des valeurs stabilisatrices estimées à plus de 35°C. En revanche, alors que les séquences compétitrices n'ont pas induites de baisse de stabilisation importante des ligands contre les cibles de topologies parallèles, le **1-PDP** lui subit une chute importante de stabilisation lorsqu'il a été incubé en présence du F-22CTA-T et des compétiteurs, ds26 ou scr26. En effet, en présence de scr26, le $\Delta Tm_{1-PDP/22CTA+ds26} = 17^\circ\text{C}$ alors que la valeur de stabilisation sans présence de compétiteur est sensiblement le double. De façon similaire, le

$\Delta Tm_{1-PDP/22CTA+scr26}$ chute à 12°C. Ces données sont intéressantes car, malgré un léger effet contre les cibles parallèles, les chutes observées en présence du ligand **1-PDP** et des séquences compétitrices n'ont pas été aussi importantes. D'autre part, les composés **PDS** et **360A** stabilisent nettement moins bien la structure 22CTA que leur dérivé peptidique respectifs ($\Delta Tm_{PDS/22CTA} = 3^\circ\text{C}$; $\Delta Tm_{360A/22CTA} = 19^\circ\text{C}$). À l'inverse, c'est le ligand parent **PDC3** ($\Delta Tm_{PDC3/22CTA} = 27^\circ\text{C}$) qui présente la meilleure capacité stabilisatrice en comparaison avec son dérivé **2-PDC3** ($\Delta Tm_{2-PDC3/22CTA} = 8^\circ\text{C}$).

Contre la cible de topologie hybride wtTel26, le **1-PDP** exhibe une capacité de stabilisation de 20°C alors que le **PDS** ne stabilise la structure qu'à 2°C. Une nouvelle fois, les ligands **PDC3** et **3-360A-s** ont atteints des valeurs estimées supérieures à 35°C. Malgré une meilleure stabilisation pour le dérivé peptidique Pyridocarboxamide, le **360A** montre une excellente stabilité pour la séquence wtTel26 avec une valeur approchant les 35°C.

Dans le cas des séquences télomériques 22CTA (antiparallèle) et wtTel26 (hybride) la présence de 50 équivalents des séquences ds26 ou scr26 n'influence que très peu les valeurs de stabilisation obtenues pour ces derniers ligands.

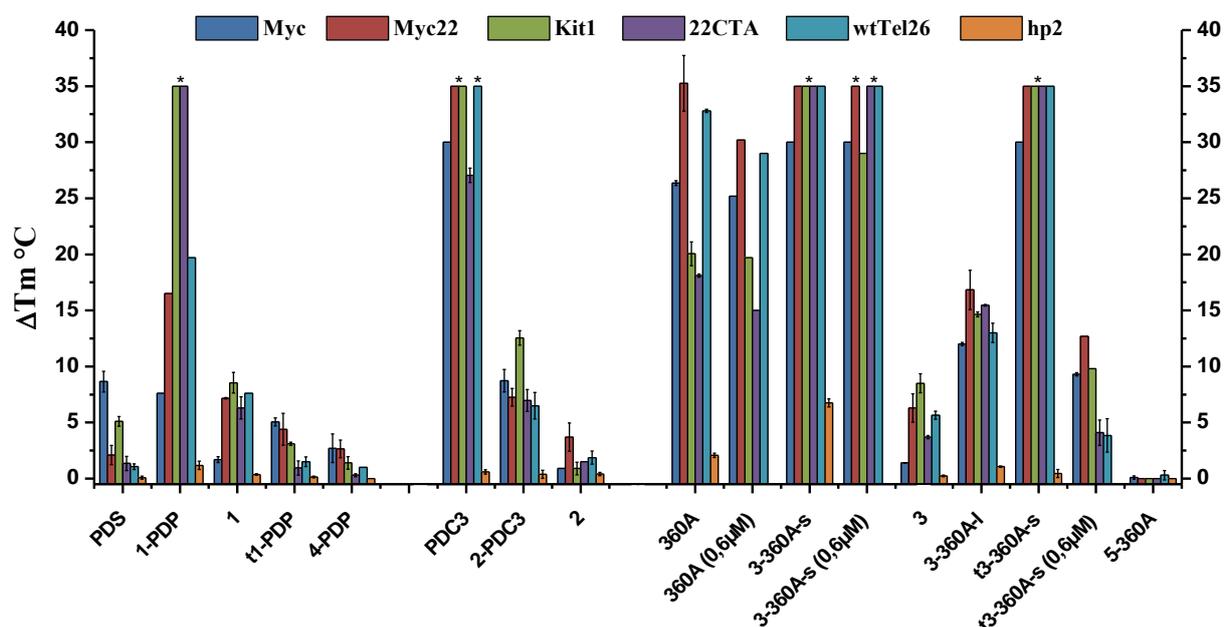
Paragraphe 2.2.1.c) Stabilisation du contrôle *Hairpin* d'ADN.

L'évaluation de la capacité de stabilisation des ligands retenues, et de leur parent, sur la séquence contrôle *hp2* a permis d'étudier la spécificité des molécules vis-à-vis d'un mime de duplex d'ADN. Malgré des valeurs faibles, les composés étudiés présentent tous un certain effet stabilisant sur la séquence *Hairpin hp2*. Alors que le composé **1-PDP** stabilise de 1,5°C le duplex d'ADN, le dérivé **2-PDC3** semble avoir un effet stabilisant moins marqué ($\Delta Tm_{2-PDC3/hp2} \sim 0,5^\circ\text{C}$). Parmi les trois ligands parents étudiés, le ligand présentant la meilleure capacité stabilisatrice de duplex d'ADN est le **360A** ($\Delta Tm_{360A/hp2} = 2^\circ\text{C}$). Enfin, c'est le dérivé **3-360A-s** qui exhibe le meilleur effet stabilisateur de structure *hairpin* ($\Delta Tm_{3-360A-s/hp2} = 7^\circ\text{C}$). Malgré une capacité stabilisatrice très important ($\Delta Tm_{3-360A-s/\text{toutes cibles}} > 35^\circ\text{C}$) vis-à-vis de toutes les cibles utilisées, la stabilisation de l'*hairpin* est un point négatif pour le composé **3-360A-s** dans le contexte d'une utilisation en milieu cellulaire. Le composé **3-360A-s** porte un nombre de charge net total de 7, provenant du motif hétéroaromatique et des bras peptidiques riches en arginine. La stabilisation de l'*hairpin* n'est donc pas inattendu malgré les bons résultats de ce ligand en séquençage vis-à-vis de la cible *hairpin* ou encore de la présence d'ADN double brin compétiteur lors des sélections. D'après les données obtenues il est effectivement possible de constater que plus globalement les ligands présentant deux séquences doublement arginillées (**1-PDP** et **3-360A-s**) présence une stabilisation plus importante de *hp2* que PDS et 360A respectivement. A *contrario*, la séquence **2-PDC3** possédant deux fois moins d'arginines et complémentée par des acides aminés hydrophobes stabilise moins l'*hp2* tout comme **PDC3**.

Section 2.2.2 : Etude de l'effet stabilisant des ligands contrôles (peptides, tronqués, non-enrichis).

L'étude par FRET-Melting des composés Pyridostatine, Phénanthroline, Pyridocarboxamide et leurs dérivés peptidiques, tronqués et non-enrichis a également été

menées sur les cibles quadruplex et hairpin décrites précédemment. Les capacités de stabilisations de chacune des molécules sont illustrées sur l'histogramme en Graphique 24.



Graphique 24 : Représentation des valeurs de stabilisation (ΔT_m °C) de tous les ligands (peptides, composés peptido-organiques et tronqués) pour chacune des cibles utilisées.

Paragraphe 2.2.2.a) Dérivés de la famille Pyridostatine.

Les conclusions exprimées précédemment ont permis de montrer que le ligand **1-PDP** exhibait un effet stabilisant plus important que son dérivé parent **PDS**, notamment contre la cible de topologie parallèle Kit1 et antiparallèle 22CTA. Pour rappel, le composé peptido-organique **1-PDP** a été synthétisé à partir de la séquence peptidique **1**, soumis à réaction de Click cuprocatalysée avec le PDP. La séquence peptidique **1** (GRRWG-K^{N3}-GRRAG) a également été étudiée par FRET-Melting afin de visualiser la contribution de l'effet stabilisateur engendré par la séquence peptidique seule. De manière générale, l'efficacité de stabilisation de **1** est sensiblement la même pour toutes les séquences quadruplex étudiées ($\Delta T_{m1} \sim 7^\circ\text{C}$) à l'exception de la séquence Myc ($\Delta T_{m1/\text{Myc}} = 2^\circ\text{C}$). Le dérivé tronqué **t1-PDP** a également été étudié. De façon rassurante, le composé tronqué exhibe des valeurs de stabilisations nettement inférieures à celles obtenues pour **1-PDP** ($\Delta T_{m1-\text{PDP}i} = 2$ à 5°C). En effet, cette valeur est concordante avec les hypothèses émises, à savoir que le motif du [Cycle 4] a vraisemblablement contribué au bon enrichissement de **1-PDP**. Enfin, le composé **4-PDP**, peu enrichi lors des campagnes de sélections, a été étudié. Les valeurs de stabilisation de **4-PDP** ($\Delta T_{m4-\text{PDP}i} = 1$ à 3°C), en comparaison à celles de **1-PDP**, confirment parfaitement le lien entre l'enrichissement des motifs individuels (et des composés uniques) et la capacité d'interaction des ligands avec les structures G-quadruplex. Même si l'interaction, l'affinité, d'un ligand pour une structure quadruplex ne soit pas systématiquement en lien avec la capacité d'une molécule à stabiliser un édifice G-quadruplex, il semblerait qu'un lien soit présent dans le cadre de l'étude des molécules issues de la DEL1 et présentant des structures comparables.

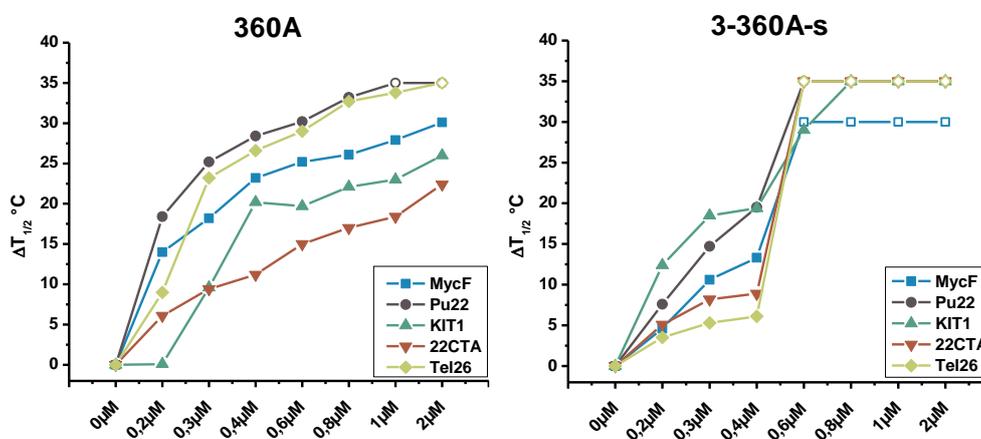
Paragraphe 2.2.2.b) Dérivés de la famille Phénanthroline.

Bien qu'aucun composé tronqué ou non-enrichi aient été synthétisés dans la famille des Phénanthroline, le peptide **2** seul a été étudié. Présentant une stabilisation générale encore inférieure à la séquence **1** ($\Delta T_{m2} = 2$ à 4°C), la séquence peptidique GFRWG-K^{N3}G-GPRFG ne semble pas apporter de contribution particulière à la stabilisation de **2-PDC3** avec les cibles. Ceci implique que la fonctionnalisation alcyne de **PDC3** et/ou l'ajout de la séquence peptidique réduisent la capacité stabilisatrice de la molécule. Il est à garder en tête que la dérivation de la Phénanthroline **PDC3** a été largement étudiée dans la littérature et qu'en général, l'ajout d'un motif supplémentaire à la structure de base engendre une baisse de stabilisation en comparaison avec le **PDC3**³²⁸. Ainsi, il n'est pas improbable que la dérivation même du **PDC3** avec un bras porteur d'une fonction alcyne joue un rôle négatif dans la stabilisation du **2-PDC3** pour les cibles quadruplex.

Paragraphe 2.2.2.c) Dérivés de la famille Pyridocarboxamide.

De façon similaire à la famille pyridostatine, la famille pyridocarboxamide comprend le ligand **360A**, son dérivé peptido-organique **3-360A-s** ainsi que le peptide **3**, le dérivé tronqué **t3-360A-s** et enfin le dérivé non-enrichi **5-360A**. Il est à rappeler que le dérivé à chaîne longue du **3-360A-s**, appelé **3-360A-l**, a également été synthétisé afin d'étudier l'influence de la longueur de chaîne sur la capacité de stabilisation d'une structure quadruplex.

L'étude de stabilisation du dérivé tronqué du **3-360A-s**, le **t3-360A-s**, a montré un comportement intéressant vis-à-vis de toutes les séquences quadruplex. En effet, de manière similaire au **3-360A-s**, le dérivé tronqué stabilise les G-quadruplex à des valeurs dépassant les 35°C . S'est alors posé la question de l'incorporation du motif du [Cycle 4] et de son intérêt dans l'effet stabilisateur des séquences quadruplex. Dans les conditions classiques d'expérimentations, le ligand considéré est présent à hauteur de 5 équivalent ($1\mu\text{M}$) par rapport à la séquence oligonucléotidique ($0,2\mu\text{M}$). Afin d'essayer de comprendre l'effet stabilisateur semblable du **3-360A-s** et de son dérivé tronqué **t3-360A-s**, nous avons évalué l'effet de la concentration du ligand sur la stabilisation des cibles quadruplex.



Graphique 25 : Etude des valeurs de stabilisation (ΔT_m °C) de **360A** (gauche) et **3-360A-s** (droite) à différentes concentrations pour toutes les cibles G-quadruplex.

Sur le Graphique 25 (gauche) il est possible de constater que l'effet stabilisateur de 360A sur les différentes séquences quadruplex est fonction du nombre d'équivalent mis en présence au sein de l'expérimentation (0 éq./0 μM à 10 éq./2 μM). La même étude a alors été effectuée à partir du dérivé **3-360A-s** (Graphique 25, droite). Quel que soit la cible considérée, il est possible d'apercevoir une nouvelle fois que l'augmentation des valeurs de stabilisation est croissante en fonction du nombre d'équivalent, et ceux jusqu'à l'ajout de 2 équivalents de ligand. Néanmoins, à partir de 3 équivalents (0,6 μM) les capacités stabilisatrices des ligands plafonnent (à l'exception de Kit1) à des valeurs de stabilisation estimées à plus de 35°C (30°C). Le ligand **t3-360A-s** a alors été étudié à la concentration de 0,6 μM afin d'être comparé au dérivé non tronqué (Graphique 24).

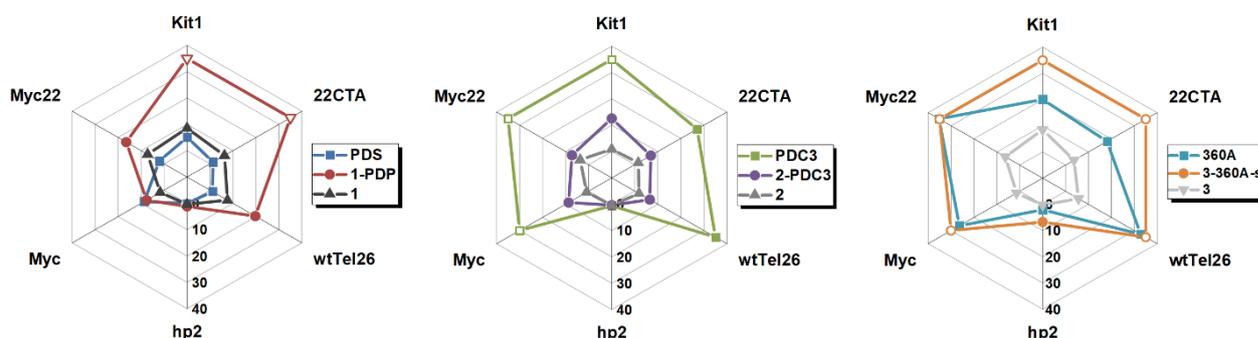
Dans des conditions expérimentales comprenant 3 équivalents de ligand, l'effet stabilisateur de **3-360A-s** reste inchangé puisque des $\Delta Tm_{\text{estimé}}$ supérieurs à 35°C ont été observés contre toutes les cibles quadruplex, à l'exception du proto-oncogène Kit ($\Delta Tm_{3-360A-s/\text{Kit1}} = 28^\circ\text{C}$). Dans les mêmes conditions, les capacités stabilisatrices de **t3-360A-s** ont drastiquement chuté ($\Delta Tm_{t3-360A-s/} = 4$ à 13°C). Toutefois, il est à noter que la baisse de stabilisation est plus importante pour les cibles antiparallèles et hybride que pour les cibles parallèles. La comparaison des deux dérivés à différentes concentrations suggère à nouveau que la présence de la séquence peptidique du [Cycle 4] est favorable à l'augmentation des propriétés stabilisatrices de la molécule.

Pour finir, le **5-360A-s** a été analysés pour sa capacité à stabiliser les structures G-quadruplex. Pour rappel, ce composé a été choisi pour ces valeurs d'enrichissements négatives lors du processus de sélection contre les cibles G-quadruplex. A l'égard des données obtenues après expériences de FRET-Melting ($\Delta Tm_{5-360A-s/} = 0^\circ\text{C}$), il est une nouvelle fois possible d'exprimer une certaine corrélation entre l'enrichissement (reflétant l'affinité de la molécule pour sa cible) et la capacité de stabilisation de la molécule considérée, ici le **5-360A-s**. Ce qui est également intéressant de constater c'est que la séquence peptidique **5**, hydrophobe et présentant deux arginines GFPRG-K^{N3}G-GRLLG, une fois lié au ligand **360A** ne présente aucune affinité pour la séquence duplexe d'ADN *hp2*. Cette constatation renforce le fait qu'il serait alors envisageable de trouver dans la combinatoire de ligand **360A** présent dans la DEL1, un composé présentant des caractéristiques de stabilisation contre les cibles quadruplex et une sélectivité pour l'ADN duplexe.

En résumé

L'étude par FRET-Melting des composés sélectionnés après séquençage à haut débit et de leurs dérivés a pu montrer que deux des dérivés stabilisés plus fortement les séquences quadruplex que leurs composés parents, le **1-PDP** et le **3-360A-s**. Le **1-PDP** stabilise mieux les G-quadruplex à topologies parallèles que le ligand **PDS**, et **1-PDP** stabilise grandement la séquence antiparallèle 22CTA et hybride wtTel26 avec des valeurs de stabilisations estimées à plus de 35°C (Graphique 26). Il a été montré que les études de compétitions n'ont pratiquement pas influencées les valeurs obtenues, à l'exception de **1-PDP** contre la cible

22CTA ou la présence soit de scr26 soit de ds26 a diminué de moitié sa capacité de stabilisation.



Graphique 26 : Représentation araignée des différentes familles des ligands en fonction des différentes cibles.

Le ligand 360A présente déjà des capacités de stabilisation très intéressante vis-à-vis de toutes les séquences quadruplex étudiées, quel que soit leur topologie ($\Delta Tm_{360A} = 20$ à 35°C). Toutefois, **3-360A-s** a montré des caractéristiques supérieures au **360A**. En effet, des valeurs des stabilisations estimées à plus de 35°C ont été observées contre toutes les cibles G-quadruplex (Graphique 23). Cependant, le ligand **3-360A-s** exhibe un certain effet stabilisateur pour le duplexe d'ADN *hp2*. L'étude du dérivé tronqué **t3-360A-s** a également montré que l'ajout d'une séquence chargée positivement ([*Cycle* 4]) sur le dérivé Pyridocarboxamide entraîne un effet positif sur sa stabilisation.

Enfin, le seul dérivé issu de la DEL1 ne présentant pas des propriétés stabilisatrices supérieures à son parent est le dérivé **2-PDC3** (Graphique 24). En effet, alors que **PDC3** stabilise fortement toutes les cibles G-quadruplex à l'instar de **3-360A-s**, le composé **2-PDC3** exhibe des valeurs d'enrichissements modestes. Bien que ce composé figure ait figuré parmi les mieux classés de sa famille lors de l'analyse du séquençage, il est à noter que son enrichissement pour la plupart des cibles est modeste. De plus, comme discuté auparavant, il est compliqué de travailler avec des dérivés de **PDC3** puisque la dérivatisation chimique d'une position du **PDC3** semble engendrer une perte notable de stabilisation. Il serait néanmoins intéressant d'évaluer des conjugués de **PDC3** issus de dérivés dont la chaîne portant la fonction alcyne serait plus courte.

Au-delà de la capacité de stabilisation des ligands et contrôles synthétisés, nous avons également souhaité déterminer les constantes d'affinités de ces différentes molécules. Nous avons donc poursuivi ces travaux avec deux approches, des mesures d'affinités par extinction de fluorescence (en. *Fluorescence Quenching*, FQ) ou par réfraction d'ondes lumineuses. (BLI, en. *BioLayer interferometry*).

[Sous-partie 2.2 : Titration de fluorescence.](#)

La méthode de détermination de K_d^{APP} par titration de fluorescence a été décrite la première fois par le groupe de Balasubramanian³²⁹ en 2015. Dans cette publication, les auteurs ont montré qu'il était possible d'induire une extinction de fluorescence d'une séquence oligonucléotidique G-quadruplex, marquée par un fluorophore, par l'interaction d'un ligand à

proximité du fluorophore (Figure 113). Des études complémentaires ont été réalisées et il a été démontré que la disparition de fluorescence a pu être restaurée par addition de compétiteurs G-quadruplex non-marqués.

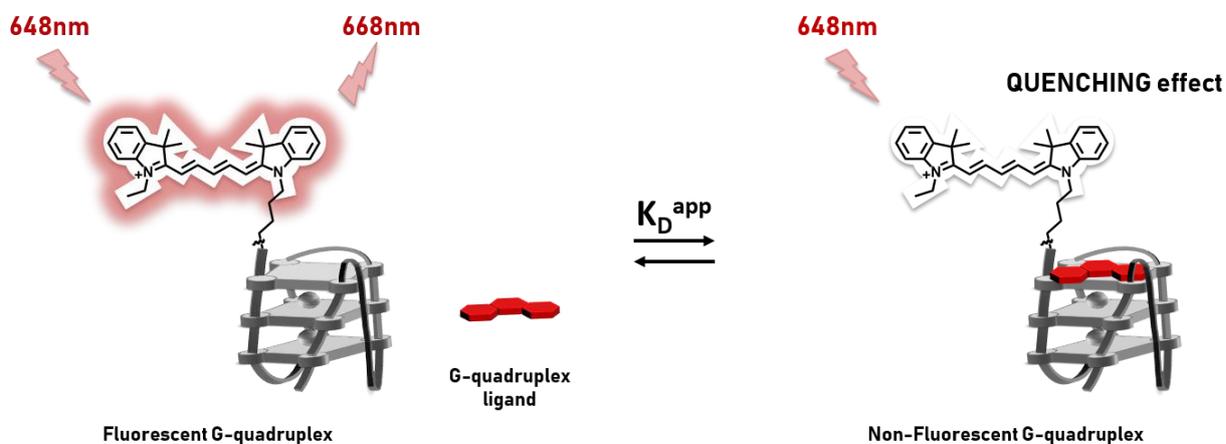
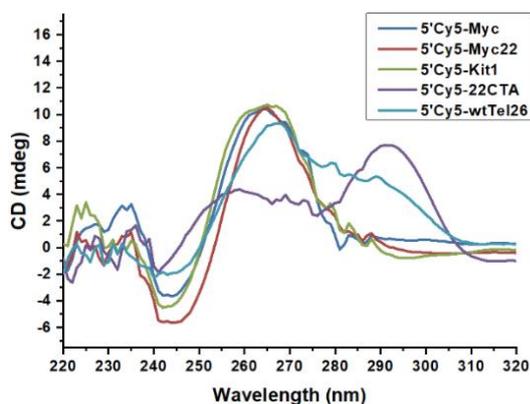
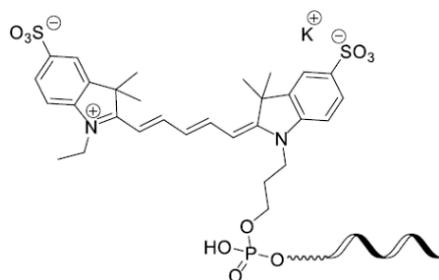


Figure 113 : Expérience de titration de fluorescence d'une séquence G-quadruplex. Phénomène d'extinction de fluorescence d'une Cyanine-5 par ajout croissante d'une concentration de ligands. Figure adaptée de D.D. Le²⁵.

Les travaux effectués dans cette publication ont cherché à valider la méthode de caractérisation d'affinité par extinction de fluorescence en comparant, notamment, les affinités des ligands **PDS** et **PDC3** en interaction avec le G-quadruplex *hTelo* issu de la séquence télomérique humaine (5' – AGGGTTAGGGTTAGGGTTAGGGT-3', dérivé de la séquence wtTel26) ainsi que le proto-oncogène Kit1 (5' – AGGGAGGGCGCTGGGAGGAGGG-3') aux autres méthodes de caractérisations de la littérature (titration UV-Visible, FRET-quenching, SPR, ITC ; informations supplémentaire de la publication³²⁹). De façon intéressante, alors que le ligand **PDS** exhibe la même affinité pour la structure Kit1 que la séquence soit fonctionnalisée par le fluorophore en 5' ou en 3' ($K_d = 640$ nM et 440 nM, respectivement), il a été montré que le ligand **PDC3** induit une interaction dont l'affinité est 46 fois plus faible lorsque la séquence Kit1 est fonctionnalisée par le fluorophore en 5' ($K_d = 7$ nM ; et fluorophore en 3' : $K_d = 330$ nM). Les données obtenues ont alors été comparées à celles de la littérature et la mesure du K_d^{app} tel que décrit par cette méthode permet d'obtenir des valeurs mesurées présentant une bonne corrélation (« Pearson correlation analysis : $r=0.88$ »).

Parallèlement à cette étude, un autre groupe a montré par deux fois qu'il était possible d'utiliser des séquences d'oligonucléotides modifiés, par un 5 – (benzofuran-2-yl) uracil³³⁰ ou une 2'deoxyuridine 5-selenophène³³¹, afin d'estimer les propriétés d'association d'un ligand avec des séquences G-quadruplex en suivant le même principe d'extinction de fluorescence.

Les deux méthodes peuvent ainsi être adaptées pour déterminer les propriétés d'interactions d'un ligand avec sa cible. L'avantage de la méthode utilisant des bases modifiés est qu'elle n'affecte pas la topologie des séquences quadruplex. En effet, lors d'un travail préparatoire consistant à vérifier par CD la formation du G-quadruplex attendu pour les cibles marquées par une Cy5 en 5', il a été possible de constater un changement de conformation d'une séquence oligonucléotidique fonctionnalisée (Graphique 27, gauche).



Graphique 27 : Dichroïsme circulaire des séquences G-quadruplex cyanines. Expérience réalisée au sein du tampon de titration.

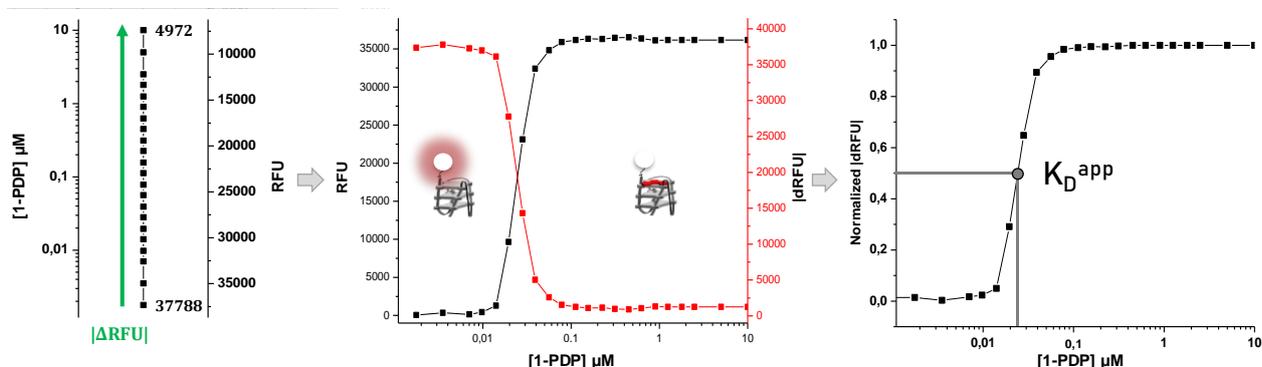
Les 5' — Cy5-oligonucléotides utilisées dans les expériences de titration de fluorescence sont les mêmes que les séquences oligonucléotidiques doublement marqués, utilisées dans les expériences de FRET-Melting. Il est ainsi possible de retrouver les trois séquences parallèles issues des promoteurs d'oncogènes Myc et Kit (Myc, Myc22 et Kit1) ainsi que les deux séquences issues de la séquence télomérique humaine, le 22CTA de topologie antiparallèle et le wtTel26 de topologie hybride.

Sur le Graphique 27 sont représentés les spectres CD des cibles quadruplex cyanines. En comparaison à leurs homologues biotinylés largement décrits dans la littérature, les séquences promotrices d'oncogènes Myc, Myc22 et Kit1 adoptent une topologie formellement parallèle. La cible 5'Cy5-wtTel26, elle, semble effectivement adopter une structure hybride (max 290, épaulement 260-270, min 240) avec une valeur maximale à 260 nm laissant penser à une double conformation parallèle/hybride³²⁹. Enfin, la cible F-22CTA-T adopte une structure particulière. En effet, cette structure présente deux maximums (255 nm et 295 nm) sans minimum réellement caractérisable. Cette dernière topologie semble également être un mixte entre une topologie antiparallèle et une légère composante hybride (épaulement vers 260nm).

Le titrage fluorimétrique isotherme a donc été réalisé au laboratoire à partir des séquences G-quadruplex fonctionnalisées en 5' par une Cyanine5. Effectuée à température ambiante, cette méthode permet de caractériser des ligands dont les valeurs de K_d peuvent atteindre le nM. De plus, la valeur de la constante de dissociation apparente K_d^{App} du complexe G4-ligand est obtenue directement après traitements mathématiques.

Les expériences de titration fluorimétrique ont été réalisées dans un tampon de titration (50mM Tris.HCl, pH 7.2, 150 mM KCl, 0.05 v/v % Triton X-100) composé de l'oligonucléotide marqué ($[c]_{finale} = 2$ nM) et d'une concentration définie du ligand à étudier. En effet, afin de réaliser la courbe de titration d'un ligand donné vis-à-vis d'une cible quadruplex, 22 concentrations différentes de ligand ont été utilisées (1,8 nM à 10 μ M ; Graphique 28, gauche). Afin que les réactions puissent être effectuées en parallèle et à haut-débit, les expérimentations ont été réalisées en plaques 96 puits. Une fois l'oligonucléotide et le ligand mise en mélange dans le tampon de titration, la plaque 96-puits a été laissée pour incubation pendant 2 heures, sous agitation. La mesure de l'intensité de fluorescence pour chaque concentration de ligand a

permis de décrire une courbe de titration (Graphique 28, milieu) dont l'ajustement par un modèle mathématique Dose-Réponse, ou Bidose-Réponse suivant le cas (*Origin 2019*), permet d'obtenir la valeur de K_d^{APP} caractérisant chaque couple ligand/5'Cy5-oligo (Graphique 28, droite).



Graphique 28 : Courbes caractéristiques des expériences de titrations de fluorescence. A gauche : Représentation de la baisse des valeurs de fluorescences (RFU) en fonction de la concentration en ligand (ex : [1-PDP] μ M). Au centre. Représentation de la baisse de fluorescence (rouge) ou l'augmentation de l'extinction de fluorescence (noir) en fonction des différentes concentrations en ligands. A droite : Normalisation et ajustement des valeurs de fluorescence en fonction de la concentration en ligand. La valeur de fluorescence à 0,5 permettant de mesurer le K_D^{app} du ligand considéré pour sa cible.

Les valeurs K_d^{APP} (Tableau X) obtenues sont représentées en Tableau 23.

Tableau 23 : Mesure du K_D^{app} des ligands en fonction des cibles G-quadruplex par extinction de fluorescence. $N=3$. K_D^{app} mesurés suivant un mode d'interaction G4 : ligand 1:1 (a), 2:1 (b) ou valeurs non-déterminés (N.D).

Ligand	Myc	Myc22	Kit1	22CTA	wtTel26
PDS	93.3 \pm 7.4 ^a	84.6 \pm 5.6 ^a	177.5 \pm 5.4 ^a	398.1 \pm 21.0 ^a	316.8 \pm 21.7 ^a
1-PDP	16.8 \pm 0.7 ^a	16.0 \pm 1.1 ^a	25.1 \pm 0.6 ^a	104.6 \pm 7.4 ^a	47.2 \pm 4.3 ^a
t1-PDP	66.7 \pm 2.7 ^a	59.1 \pm 3.1 ^a	129.6 \pm 8.22 ^a	<i>N.D</i>	<i>N.D</i>
4-PDP	83.1 \pm 3.8 ^a	74.9 \pm 6.5 ^a	97.1 \pm 3.2 ^a	500.1 \pm 29.6 ^a	448.0 \pm 30.8 ^a
PDC3	29.6 \pm 0.9 ^b	31.5 \pm 0.8 ^b	41.2 \pm 2.0 ^b	30.9 \pm 1.2 ^b	19.4 \pm 0.8 ^b
2-PDC3	25.8 \pm 2.1 ^a	31.1 \pm 3.2 ^a	125.1 \pm 11.4 ^a	835.2 \pm 56.5 ^a	15.4 \pm 1.3 ^b
360A	90.7 \pm 4.9 ^b	206.4 \pm 16.7 ^a	<i>N.D</i>	55.4 \pm 2.5 ^b	43.6 \pm 2.6 ^b
3-360A-s	12.6 \pm 0.7 ^b	14.5 \pm 0.7 ^b	56.6 \pm 5.2 ^a	10.5 \pm 0.5 ^b	9.5 \pm 0.5 ^b
t3-360A-s	23.6 \pm 1.3 ^a	24.3 \pm 1.0 ^a	76.3 \pm 6.8 ^a	16.2 \pm 0.7 ^b	15.0 \pm 1.0 ^b
3-360A-l	26.2 \pm 1.0 ^b	26.6 \pm 1.3 ^b	92.6 \pm 5.9 ^a	27.3 \pm 1.3 ^b	25.0 \pm 2.0 ^b
5-360A-s	76.9 \pm 2.7 ^a	73.2 \pm 3.5 ^a	390.7 \pm 37.5 ^a	68.8 \pm 2.8 ^b	76.6 \pm 6.6 ^b

Section 2.2.1 : Dérivés de la famille Pyridostatine.

Plusieurs K_d ont été proposés dans la littérature concernant le ligand PDS. Des valeurs comprises entre 200 et 500 nM vis-à-vis des G-quadruplex issus de la séquence télomérique humaine ont par exemple été rapportées^{144,332}. Les expériences de titration de fluorescence

contre les G-quadruplex télomérique ont permis d'obtenir des valeurs de K_d^{App} de 398 et 317nM, respectivement contre la cible de topologie antiparallèle 22CTA et la cible hybride wtTel26. Les valeurs obtenues sont en accord avec les données issues de la littérature. L'affinité de PDS contre les cibles du proto-oncogènes Myc ont montrés des K_d^{App} de l'ordre de 100 nM et de 200nM contre l'autre proto-oncogène, Kit1.

Le dérivé peptidique **1-PDP** a montré des valeurs de K_d^{App} près de 4 à 7 fois inférieures à celles obtenues pour le ligand PDS. En effet, **1-PDP** exhibe des affinités d'environ 20 nM contre les cibles de topologies parallèles (Myc, Myc22, Kit1) et des valeurs de 50 à 100 nM pour les cibles télomériques (wtTel26, 22CTA). En comparaison direct, le **t1-PDP** présente des valeurs d'affinités qui restent intéressantes (67 à 130 nM), mais néanmoins inférieures à celles obtenues pour le **1-PDP**. Ces valeurs sont concordantes avec les données de stabilisation obtenues en FRET-Melting, dans le sens ou la présence de la séquence pentapeptidique du [Cycle 4] permet d'améliorer les caractéristiques intrinsèques d'affinité et de stabilisation du composé **1-PDP**.

Le dérivé « non-enrichi » **4-PDP** présente également de bonnes affinités de l'ordre de 100 nM contre les cibles G-quadruplex parallèles et d'environ 500 nM pour les séquences télomériques. Néanmoins, ces valeurs de K_d^{App} restent près de 4 à 10 fois supérieures aux valeurs obtenues pour le **1-PDP**.

Section 2.2.2 : Dérivés de la famille Phénanthroline.

Le dérivé Phénanthroline **PDC3** est un des ligands les plus affins des structures quadruplex avec un K_d compris entre 2 et 20 nM selon diverses sources^{180,329,333}. Les expériences de titrations de fluorescences ont permis d'obtenir des données similaires à celles de la littérature avec des valeurs de K_d^{App} de 30 à 40 nM contre les cibles de topologies parallèles et de 20 à 30 nM contre les cibles télomériques wtTel26 et 22CTA, respectivement. De manière surprenante, le dérivé **2-PDC3** exhibe des affinités similaires au **PDC3** pour les cibles du promoteur d'oncogène Myc et pour la cible télomérique wtTel26 (K_d^{App} de 15 à 30 nM). Toutefois, des valeurs biens supérieures ont été observées contre les cibles Kit1 et 22CTA, avec une augmentation drastique de 27 fois contre cette dernière. Ces résultats pourraient indiquer que la capacité de stabilisation et l'affinité inhérente ne sont pas systématiquement en lien au sein de notre librairie et semble élucider des résultats honorables d'enrichissement de **2-PDC3**.

Section 2.2.3 : Dérivés de la famille Pyridocarboxamide.

Les expériences de titrations montrent que le ligand **360A** présente une très bonne affinité pour les séquences télomériques (K_d^{App} de 50 nM) et des valeurs correctes pour le promoteur d'oncogène Myc (K_d^{App} de 90 et 206 nM, respectivement pour Myc et Myc22). En cohérence avec les données de stabilisation, le composé peptido-organique **3-360A-s** exhibe des valeurs d'environ 10 nM pour toutes les séquences quadruplex, sauf vis-à-vis du proto-oncogène Kit1 ou le K_d^{App} , qui reste excellent, est de 50 nM. **3-360A-s** apparait ainsi comme le ligand le plus affiné parmi tous les composés étudiés.

Les valeurs de K_d^{App} du dérivé 3-360A-s a été comparé aux valeurs obtenues pour le ligand tronqué, **t3-360A-s**, et le ligand Pyridocarboximide à chaîne longue, **3-360A-I**. Comparativement, les deux ligands précédemment cités présentent les mêmes constantes d'affinités vis-à-vis des cibles quadruplex (K_d^{App} de 15 à 92 nM). Ces valeurs montrent que ces deux ligands contrôles sont globalement deux fois moins affins que le ligand **3-360A-s**, cette différence est modeste, au même titre que celle observée en FRET-Melting suggérant un effet réel mais faible de la nature du linker ainsi que de la présence du peptide en Cycle 4 au sein de cette famille de dérivé.

Enfin, le dérivé « non-enrichi » **5-360A-s** a été synthétisé en raison de ces faibles valeurs d'enrichissements (voire déplétions) contre les cibles quadruplex. Des K_d^{App} de 5 à 8 fois supérieurs à ceux trouvés pour le **3-360A-s** ont été obtenues. En effet, l'affinité mesurée vis-à-vis de toutes les cibles approche les 70 nM, à l'exception de Kit1, où le K_d^{App} est de 390 nM. Ces valeurs sont significativement inférieures à celles obtenues pour le ligand enrichi mais reste néanmoins bonne, soulignant la forte affinité de la famille des dérivés de **360A**.

En résumé

Les cibles oncogéniques Myc et Kit ainsi que des séquences télomériques capables de former des structures G-quadruplex ont été utilisées afin de déterminer par titration de fluorescence les constantes d'affinités apparentes (K_d^{App}) des ligands dérivés de la Pyridostatine (**PDS**, **1-PDP**, **t1-PDP**, **4-PDP**), de la Phénanthroline (**PDC3**, **2-PDC3**) et du Pyridodicarboxamide (**360A**, **3-360A-s**, **t3-360A-s**, **3-360A-I**, **5-360A-s**).

Dans la famille Pyridostatine il a pu être montré que le ligand **1-PDP** présentait des meilleurs K_d^{App} que le **PDS**, avec des affinités variant de 16 à 25 nM vis-à-vis des cibles parallèles, et de 47 à 105 nM pour les séquences télomériques wtTel26 et 22CTA. La conception rationnelle de la librairie combinatoire montre bien qu'il est possible de trouver des dérivés peptidiques de **PDS** pouvant présenter de meilleures caractéristiques en termes d'affinité. D'autre part, par l'étude du dérivé tronqué **t1-PDP** il est possible de suggérer que chaque motif de la librairie combinatoire participe à l'affinité du **1-PDP**. En effet, l'omission volontaire de la dernière séquence peptidique a un effet défavorisant les interactions entre le ligand et sa cible.

Dans la famille des Pyridocarboxamide, il apparaît nettement que le ligand **3-360A-s** exhibe une meilleure affinité que le **360A** vis-à-vis des cibles G-quadruplex. Une nouvelle fois, cette augmentation d'affinité est due à la présence d'une séquence peptidique favorisant des interactions additionnelles avec la structure G-quadruplex. Alors que les différences entre le **1-PDP** et son dérivé tronqué marquaient fortement l'intérêt de la séquence peptidique au [Cycle 4] de la librairie combinatoire, les conclusions sont plus délicates à propos de leurs homologues **360A**. En effet, les constantes d'affinités entre le **3-360A-s** et le **t3-360A-s** sont du même ordre de grandeur, à un facteur 2 près, interrogeant ainsi l'intérêt de considérer une séquence peptidique présentant quatre charges positives au détriment de la version tronquée ne présentant que deux arginines. Le **360A-s** et **360A-I** ne présentent de divergences qu'au niveau de leur bras alcyne, où **360A-I** exhibe 3 atomes de carbones supplémentaires. La variation de cette chaîne latérale a montré qu'elle avait un rôle dans les valeurs

d'enrichissements obtenues puisque les dérivés à chaînes longues étaient globalement moins enrichis que leur homologue **360A-s**. Cette variabilité d'enrichissement est faiblement retrouvée après étude de leurs affinités par titration de fluorescence. En effet, le composé **3-360A-s** et **3-360A-I** présentent le même ordre d'affinité quel que soit les cibles considérées (à un facteur 2 près).

Enfin, parmi les trois familles étudiées, le seul composé ne présentant pas de constantes d'affinités meilleures que son composé parent est le **2-PDC3**. En effet, alors que les K_d^{APP} sont identiques pour les cibles Myc, My22 et wtTel26, une augmentation considérable a été mesurée vis-à-vis de la séquence Kit1 et plus particulièrement pour la séquence 22CTA où le K_d^{APP} passe de 31 nM pour **PDC3** à 835 nM pour **2-PDC3**.

Afin de confirmer ces valeurs de constantes d'affinités pour les tous meilleurs ligands, une deuxième méthode de caractérisation biophysique a été utilisée, la *BioLayer Interferometry* (fr. interférométrie de bicouche, ou BLI). Seuls les ligands les plus affins et leurs dérivés peptidiques ont donc été utilisés, **PDS/1-PDP** & **360A/3-360A-s**, et étudiés contre la seule et unique cible Myc. Ce choix a également été motivé par des considérations de temps et de coût.

Sous-partie 2.3 : Biolayer Interferometry (BLI).

Contrairement à la méthode de titration par fluorescence, le BLI est une technique optique d'analyse d'interaction biomoléculaire ne nécessitant pas la présence de cibles marquées. Décrite et développée par ForteBio en 2005, le BLI repose sur le principe d'interférométrie en lumière blanche (400 à 800 nm). Dans cette technique, une lumière blanche provenant d'une LED (en. *light-emitting diode*) est dirigée dans un biocapteur au travers d'une fibre optique. L'extrémité du biocapteur est composée d'une couche dite biocompatible (Figure 114, en violet), sur laquelle peut être fonctionnalisée de manière covalente ou non des structures protéiques, des ligands ou encore des structures nucléiques comme les G-quadruplex, et une couche de référence interne (Figure 114, cf. *optical layer*, en orange).

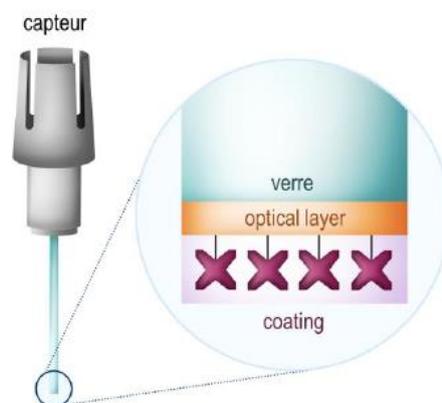


Figure 114 : Biolayer Interferometry (BLI). Utilisation d'un biocapteur fonctionnalisé par la cible d'intérêt de manière covalente ou non. Image tirée de la thèse du Dr. Eugénie Laigre, avec approbation de l'auteure.

Lorsque le spectre de lumière blanche est envoyé à travers la pointe optique du biocapteur, la majorité du faisceau n'est pas réfléchi. L'autre partie des faisceaux est réfléchi sur la couche biocompatible et la couche de référence afin de donner des interférences (constructives ou destructives) pour chacune des longueurs d'ondes du spectre (Figure 115, A/ et B/). Les faisceaux réfléchis sont ainsi enregistrés grâce à un capteur CDD (en. *charge coupled device*) situé en partie haute du biocapteur. Le profil interférométrique peut ainsi être obtenu par représentation des amplitudes des interférences observées en fonction des longueurs d'ondes du spectre de lumière blanche.

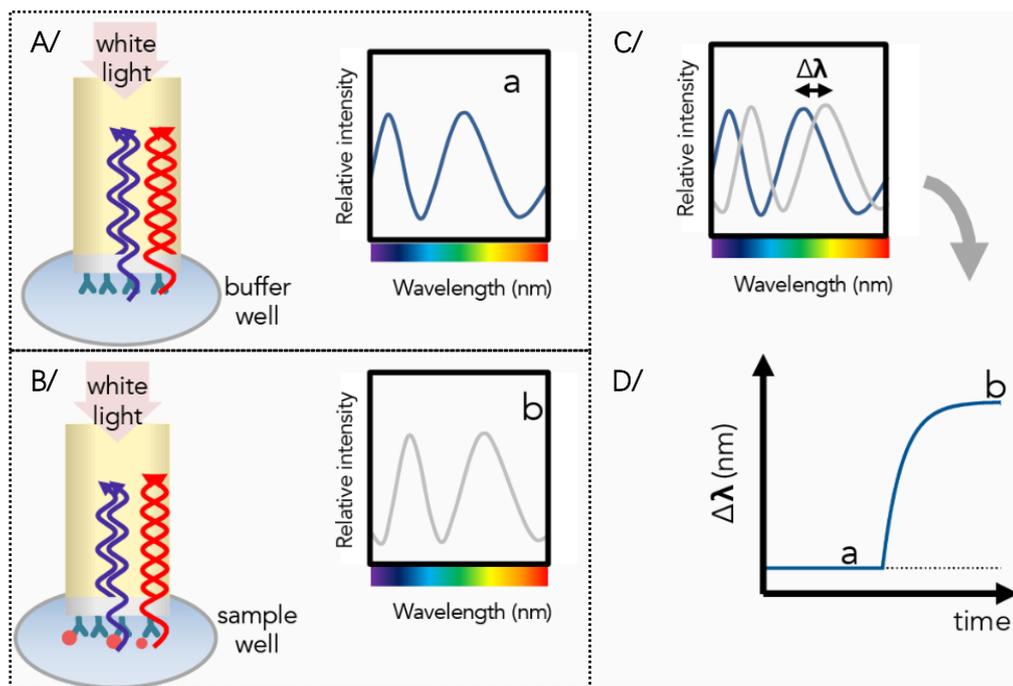


Figure 115 : BLI. Mesure interférométrique afin de caractériser l'association et la dissociation d'une interaction G-quadruplex/ligand. Figure de ForteBio adaptée par le Centre d'Interactions Macromoléculaires (CMI) d'Harvard.

Les mesures d'interférométries vont varier suivant deux cas.

Dans le premier cas, il faut considérer un biocapteur fonctionnalisé par la cible d'intérêt en absence de ligand (Figure 115, A/). Les interférences alors analysées vont être obtenues par i) réflexion de la lumière (toutes les longueurs d'ondes du spectre) sur la couche de référence interne, et par ii) réflexion de la lumière après avoir traversée le milieu biocompatible fonctionnalisé.

Dans le second cas, il est nécessaire de considérer un biocapteur fonctionnalisé en présence du ligand à étudier (Figure 115, B/). Une nouvelle fois, l'une des interférences analysées va être obtenue par réflexion de la lumière sur la couche de référence interne. Cette interférence mesurée est sensiblement la même que celle obtenue dans le premier cas. En revanche, la deuxième interférence analysée va être obtenue par réflexion de la lumière sur le système cible immobilisée-ligand, ou dans notre cas, G-quadruplex-ligand. La longueur du trajet de réflexion parcourue par l'onde lumineuse augmente lorsque le ligand est en interaction avec son ligand. En conséquence, l'interférence réfléchi ne sera plus en phase avec l'onde de

référence. Ainsi, la modification des modèles d'interférences pour chaque longueur d'onde (400nm à 800nm) va engendrer un profil d'interférométrie dont le maximum d'amplitude est décalé d'une valeur de $\Delta\lambda$ par rapport au profil d'interférométrie d'origine (Figure 115, C/).

La construction optique de ces deux systèmes permet ainsi de suivre en temps réel les événements de liaison moléculaire à la surface du biocapteur. De façon générale, plus l'épaisseur de la surface augmente (plus les molécules se fixent), plus la valeur $\Delta\lambda$ entre les deux profils d'interférométrie va être élevée. À l'inverse, la dissociation du complexe ligand/G-quadruplex entraîne une diminution du $\Delta\lambda$, jusqu'à atteindre sa position d'origine ($\Delta\lambda = 0$). Le suivi temporel du décalage des deux spectres interférométriques permet ainsi d'obtenir une courbe d'association et de dissociation d'une interaction G-quadruplex/ligand (Figure 115, D/).

La mise en place d'une expérience de BLI nécessite plusieurs étapes. Premièrement, il est nécessaire de fonctionnaliser le biocapteur (cf. la couche biocompatible du biocapteur) avec la cible d'intérêt. Cette fonctionnalisation peut être réalisée de manière covalente, à l'exemple des biocapteurs AR2G (en. *Amine Reactive 2nd Generation*) fonctionnalisée par des groupements carboxylates permettant ainsi de lier des *Amino*-protéines/oligonucléotides en présence d'EDC/NHS, ou de façon non covalente, à l'exemple des biocapteurs SA (*StreptAvidine*). Les expériences de sélection *in vitro* par affinité réalisées en amont ont nécessitées l'utilisation de cibles biotinylées. Par rappel, les cibles biotinylées ont été fonctionnalisées sur billes magnétiques MyOne T1 préfonctionnalisées par de la streptavidine. Afin de s'approcher au plus proche des conditions de sélections et dans l'intérêt d'utiliser les séquences G-quadruplex biotinylées déjà disponibles au laboratoire, les expériences de BLI ont été réalisées à partir de biocapteurs SA afin d'immobiliser les G-quadruplex (Figure 116, « Immobilization »).

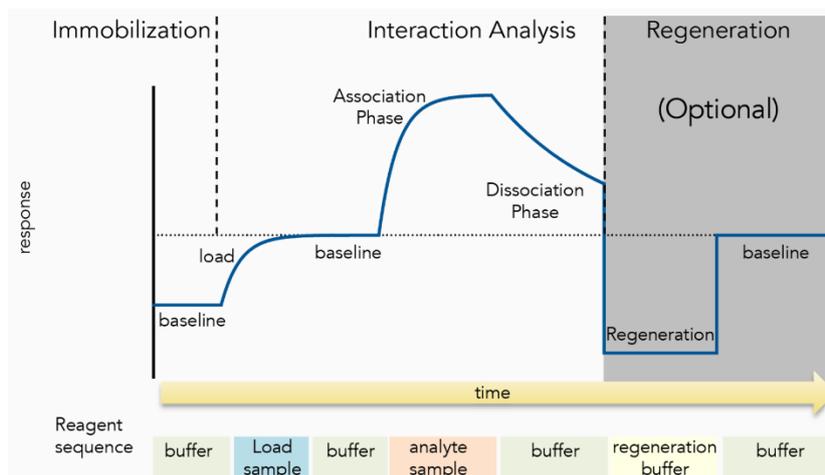


Figure 116 : BLI. Expérience de BLI caractérisée par une étape d'immobilisation, d'interaction et analyses et de régénération. Figure de ForteBio adaptée par le Centre d'Interactions Macromoléculaires (CMI) d'Harvard.

Après avoir équilibré le système en présence des biocapteurs fonctionnalisés les étapes d'association et de dissociation ont pu être effectuées en présence des ligands d'intérêt (Figure 116, « Interaction Analysis »). Enfin, dans l'intérêt de vouloir réutiliser les biocapteurs, une étape de régénération peut être réalisée (Figure 116, « Regeneration »). Cette dernière étape

permet donc de décrocher totalement les ligands toujours associés à sa cible après la phase de dissociation.

Section 2.3.1 : Fonctionnalisation du biocapteur.

Les biocapteurs commerciaux SA (Streptavidine) utilisés lors de expériences de BLI présentent au niveau de la couche biocompatible une protection faite à base de sucrose. En effet, cette zone protectrice des fonctionnalisations streptavidine doit être éliminée avant utilisation. Pour ce faire, les biocapteurs ont été incubés pendant une dizaine de minutes dans le tampon de travail (tampon BLI : 10mM HEPES, 150 mM KCl, 0,05% v/v T-20, pH 7.4). Une fois le sucrose éliminé, les biocapteurs ont été trempée dans une solution d'oligonucléotides biotinylées (100 nM, Figure 116 « immobilisation »). Il est à rappeler que les expériences de BLI n'ont été effectuées que sur le promoteur d'oncogène Myc afin de vérifier les données obtenues par titration de fluorescence. La phase d'immobilisation, d'une durée de 10 min, est suivie par une étape de rinçage des biocapteurs dans le tampon BLI afin d'éliminer l'excédent de Myc-biotine non immobilisé.

Section 2.3.2 : Analyse de l'interaction G-quadruplex/ligand.

Après fonctionnalisation par la cible Myc, les capteurs ont été mis en interactions avec les ligands à étudier, à savoir le **PDS** et son dérivé **1-PDP** ainsi que le **360A** et le composé peptido-organique **3-360A-s**. Les ligands considérés, solubilisés dans le tampon BLI, ont été préparés à 6 concentrations différentes. Ces concentrations (Tableau 24) ont été choisies en fonction des valeurs de K_d^{APP} précédemment obtenues par titration de fluorescence afin d'encercler au mieux la constante d'affinité. Étant donné que l'interaction de 4 ligands a due être caractérisée avec le proto-oncogène Myc, 4 capteurs fonctionnalisés uniques ont été utilisés, 1 biocapteur par ligand.

Tableau 24 : BLI. Gamme de concentration pour la caractérisation des interactions G-quadruplex/ligand.

Ligands	Association time (s)	Dissociation time (s)	Concentrations
PDS	7200	3600	20 – 50 – 80 – 200 - 500 – 1000
1-PDP	7200	3600	20 – 50 – 80 – 200 - 500 – 1000
360A	7200	3600	18.75 – 37.5 – 75 – 150 – 300 – 600
3-360A-s	7200	3600	5 – 10 – 15 – 25 – 50 – 100

Les biocapteurs ont tout d'abord été incubés, sous agitation, dans la solution la moins concentrée pour un temps d'association de 7200 secondes (2 heures) puis dissociés pendant 3600 secondes (1 heure) dans le tampon BLI (Figure 116, « Interaction Analysis »). Les mêmes étapes d'association/dissociation sont ensuite effectuées séquentiellement jusqu'à atteindre la concentration la plus élevée de chacun des ligands.

Section 2.3.3 : Dissociation partielle du ligand.

Suivant l'affinité du ligand pour la cible quadruplex et sa capacité à être dissocié, il est possible que la phase de dissociation dans le tampon BLI ne soit pas suffisamment efficace

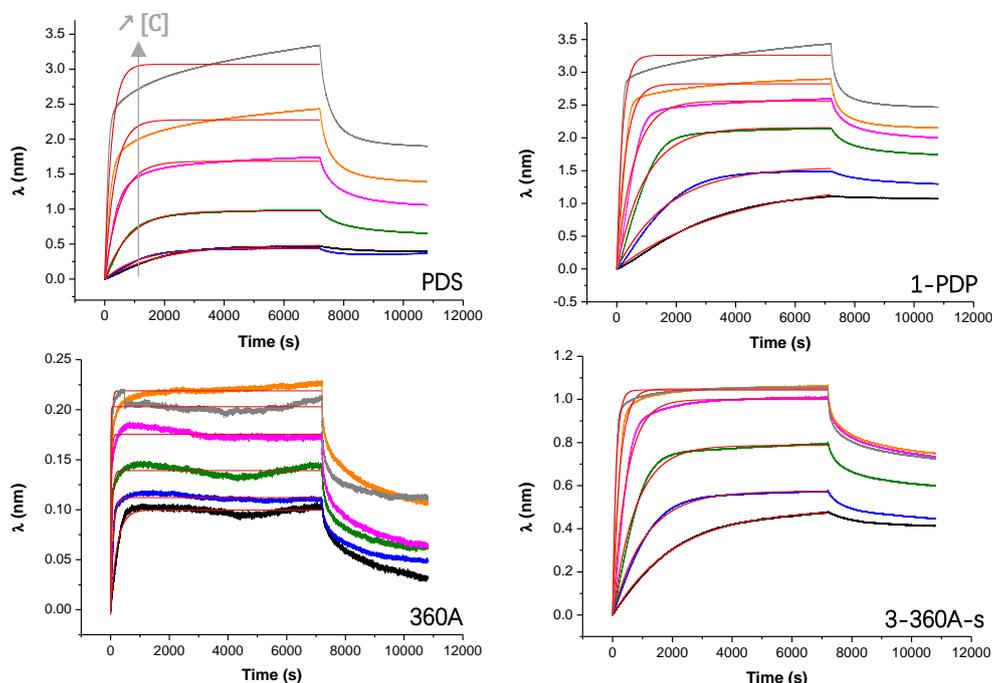
pour relarguer la totalité des ligands en interaction avec la cible. Dans ce cas, deux choix sont possibles. Dans un premier temps, il est possible d'effectuer une étape de régénération du biocapteur entre chaque cycle (Figure 116, « Regeneration »). Toutefois, avec des étapes d'association et de dissociation cumulées de plus de 3 heures, l'ajout de cette étape de régénération engendrerait des expériences longues risquant d'induire des biais à l'analyse (notamment par l'évaporation des solutions de ligands).

La deuxième méthode permettant de palier les dissociations partielles du ligand consiste à utiliser plusieurs biocapteurs pour un ligand donné. En effet, alors qu'un biocapteur unique est utilisé en routine pour analyser l'interaction G-quadruplex/ligand à 6 concentrations différentes, 6 biocapteurs seront utilisés en parallèle dans cette stratégie afin de caractériser l'interaction avec le G-quadruplex pour chaque concentration de ligand.

Les ligands **PDS**, **360A** et leurs dérivés peptidiques ont montré des problèmes de dissociation vis-à-vis de la cible Myc. Afin de limiter les effets de dissociation partielle, la deuxième méthode (6 biocapteurs/6 concentrations) de caractérisation a été utilisée. Les durées d'association et de dissociation alors appliquées (Tableau 24) ont permis d'atteindre des plateaux de saturation et un décrochage maximum de ligands.

Section 2.3.4 : Sensorgrammes des couples ligands/G-quadruplex.

Les sensorgrammes obtenus pour les différents ligands en interaction avec la cible Myc sont représentés sur le Graphique 29 ci-dessous.



Graphique 29 : BLI. Sensorgrammes de l'interaction de PDS/1-PDP et 360/3-360A-s avec la cible Myc.

Les constantes d'affinités thermodynamique et cinétique ont été calculées pour chacun des ligands étudiés et sont reportées dans le Tableau 25. Le ligand **PDS** exhibe un K_d cinétique de 255 nM alors que son homologue peptidique, **1-PDP**, obtient une valeur près de 10 fois

inférieure de 21 nM (Tableau 25). De manière étonnante, alors que la K_d cinétique est en accord avec le K_d thermodynamique pour le composé peptido-organique **1-PDP**, une très grande variation est observée pour le ligand **PDS**. En effet, un K_d cinétique de 36 nM a été mesuré.

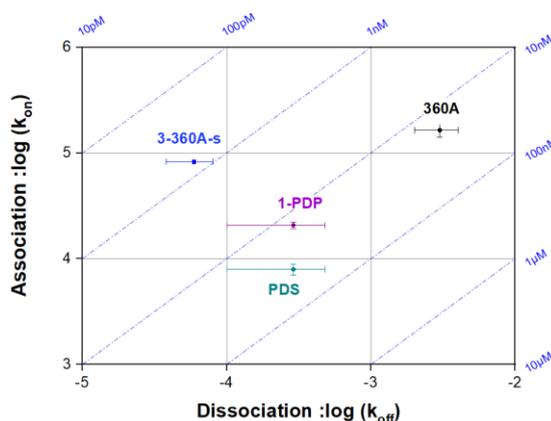
Tableau 25 : BLI. Constantes cinétiques et thermodynamiques obtenues sur la cible Myc. N=2.

Ligands	K_d thermo. (nM)	K_d cin. (nM)	K_{on} ($10^4 M^{-1}.s^{-1}$)	K_{off} (s^{-1})
PDS	255 ± 50	36 ± 25	0.8 ± 0,1	3 10 ⁻⁴ ± 2 10 ⁻⁴
1- PDP	21 ± 4	14 ± 10	2.9 ± 0,2	3 10 ⁻⁴ ± 2 10 ⁻⁴
360A	28 ± 7	18 ± 3	16 ± 2	3.0 10 ⁻³ ± 1 10 ⁻³
3-360A-s	8 ± 3	0.7 ± 0,3	8.3 ± 0,5	6 10 ⁻⁵ ± 2 10 ⁻⁵

Le ligand **360A** ainsi que le dérivé peptidique exhibant le Pyridocarboxamide à chaîne courte, **3-360A-s**, ont également fait l'objet de mesure de constantes d'affinités. Le ligand parent **360A** montre une constante d'affinité thermodynamique et cinétique semblable de 28nm et 18 nM respectivement. En comparant les K_d du **360A** avec celles du **3-360A-s**, il est possible de constater une amélioration d'affinité d'un facteur 3. En effet, le K_d thermodynamique mesuré du **3-360A-s** est de 8 nM alors que le K_d cinétique obtenue est à l'échelle subnanomolaire, de 0,7 nM.

En résumé

Afin de comparer les différences entre les ligands de la même famille, un graphique d'iso-affinité a été tracé. Celui-ci décrit le logarithme de la constante cinétique d'association — $\log(k_{on})$ — en fonction du logarithme de la constante cinétique de dissociation — $\log(k_{off})$. Les valeurs du K_d cinétique sont également représentées en pointillées par les diagonales (Graphique 30).



Graphique 30 : BLI. Iso-affinité des composés caractérisés par BLI sur la cible G-quadruplex Myc.

Le ligand **PDS** est le ligand présentant la moins bonne affinité (cinétique) vis-à-vis de la cible G-quadruplex du promoteur d'oncogène Myc (K_d cinétique = 36 nM). C'est également le ligand qui possède le k_{on} le plus faible ($0.79E+04 M^{-1}.s^{-1}$), c'est-à-dire que parmi tous les ligands, c'est **PDS** qui s'associe le plus lentement. En termes de k_{off} , les ligands **PDS** et **1-PDP**

possèdent une dissociation comparable ($2.9E-04 \text{ s}^{-1}$). Toutefois, le k_{on} du **1-PDP** est supérieur au **PDS**, ce qui explique que **1-PDP** présente une meilleure constante d'affinité (K_d cinétique = 14 nM).

De manière transversale à **PDS**, le ligand **360A** possède le k_{off} le plus élevé ($3.0E-03 \text{ s}^{-1}$) de la série de ligand. Cette information indique que **360A** a la capacité de dissociation la plus rapide parmi tous les ligands considérés. De plus, **360A** possède le k_{on} le plus élevé ($16.4E+04 \text{ M}^{-1}.\text{s}^{-1}$) ce qui permet au ligand d'obtenir une excellente constante d'affinité (K_d cinétique = 18 nM). En comparaison avec **360A**, le **3-360A-s** s'associe deux fois moins vite ($k_{on} = 8.3E+04 \text{ M}^{-1}.\text{s}^{-1}$). En revanche, la dissociation de **3-360A-s** est très lente ($k_{off} = 5.9E-05 \text{ s}^{-1}$), faisant ainsi de **3-360A-s** le composé avec une constante d'affinité au subnanomolaire (K_d cinétique = 0,7 nM). Ce résultat semble cohérent avec le mode de sélection des ligands, à savoir une stringence basée sur la dissociation, favorisant les ligands capables de rester associés dans des conditions défavorables.

Une nouvelle fois, la caractérisation de l'interaction entre les composés **PDS/1-PDP** et **360A/3-360A-s** montrent que les composés issus de la librairie combinatoire présentent de meilleures propriétés d'associations que leurs dérivés parents.

Sous-partie 2.4 : Comparaison des données obtenues.

Les valeurs de stabilisation mesurées par FRET-Melting, ainsi que les constantes d'affinités obtenues par titration de fluorescence et *BioLayer Interferometry* ont été reportées dans la Tableau 26 ci-dessous et permettent de donner une vue d'ensemble à la caractérisation des interactions réalisées au cours de cette thèse. Seuls les composés enrichis de la DEL1 (**1-PDP**, **2-PDC3** et **3-360A-s**) et leurs parents (**PDS**, **PDC3** et **360A**) sont représentés.

Tableau 26 : Résumé des données obtenues pour la caractérisation des composés par FRET-Melting, Titration de fluorescence et BLI sur les cibles de topologies parallèles (Myc, Myc22 et Kit1), antiparallèles (22CTA), hybride (wtTel26) et le contrôle Hairpin (hp1).

Cibles Ligands	Myc			Myc22		Kit1	
	ΔT_m (°C)	K_d (titration)	K_d (BLI, cinétique)	ΔT_m (°C)	K_d (titration)	ΔT_m (°C)	K_d (titration)
PDS	9 ± 1	93 ± 7	36 ± 25	2 $\pm 0,8$	85 ± 6	5 $\pm 0,4$	178 ± 5
1- PDP	8 ± 0	17 ± 1	14 ± 10	17 ± 0	16 ± 1	> 35	25 $\pm 0,6$
PDC3	> 30	30 ± 1	/	> 35	32 ± 1	> 35	41 ± 2
2-PDC3	9 ± 1	26 ± 31	/	7 $\pm 0,8$	31 ± 3	13 $\pm 0,7$	125 ± 11
360A	26 $\pm 0,2$	91 ± 5	18 ± 3	35 ± 2	206 ± 17	20 ± 1	/
3-360A-s	> 30	13 $\pm 0,7$	0,7 $\pm 0,3$	> 35	15 $\pm 0,7$	> 35	57 ± 5

Cibles	22CTA		wtTel26		Hp1	
	ΔT_m (°C)	K_d (titration)	ΔT_m (°C)	K_d (titration)	ΔT_m (°C)	K_d (titration)
PDS	1,4 $\pm 0,6$	398 ± 21	1 $\pm 0,2$	317 ± 22	0,1 $\pm 0,1$	/
1- PDP	> 35	105 ± 7	20 ± 0	47 ± 4	1,2 $\pm 0,4$	/
PDC3	27 $\pm 0,7$	31 ± 1	> 35	20 $\pm 0,8$	0,6 $\pm 0,2$	/
2- PDC3	7 ± 1	835 ± 57	7 ± 1	15 ± 1	0,4 $\pm 0,3$	/
360A	18 $\pm 0,1$	55 ± 3	33 $\pm 0,2$	44 ± 3	2 $\pm 0,2$	/
3-360A-s	> 35	9 $\pm 0,5$	> 35	9 $\pm 0,5$	7 $\pm 0,4$	/

Il est intéressant de constater que les valeurs de stabilisation et d'affinité ne sont pas toujours corrélées. Dans le cas de la Pyridostatine **1-PDP** et du Pyridocarboxamide **3-360A-s**, l'amélioration du ΔT_m par rapport aux composés parents permet l'obtention d'une valeur de K_d encore plus faible. Toutefois, l'exemple de l'étude sur les dérivés Phénanthroline montre que malgré une baisse significative de stabilisation du composé **2-PDC3** sur toutes les cibles quadruplex, les valeurs de constantes d'affinité restent dans la plupart des cas comparable à celles de **PDC3**.

D'après les données obtenues, il ne fait aucun doute que le composé exhibant les meilleures caractéristiques physico-chimiques à l'égard des cibles G-quadruplex est le dérivé peptido-organique **3-360A-s**. Ce composé stabilise ces structures à des valeurs très élevées et présente également des valeurs de stabilisation de 5 à 12°C pour des expériences de FRET-Melting en quantité équimolaire de ligand et de cible. En termes d'affinité, le **3-360A-s** présente des K_d de la dizaine de nanomolaire et des expériences complémentaires en BLI sur la cible Myc ont montré des valeurs cinétiques de K_d au subnanomolaire. En revanche, malgré des propriétés remarquables vis-à-vis des G-quadruplex, il a également été montré que le **3-360A-s** était capable de stabiliser un hairpin d'ADN ($\Delta T_m = 6.75$ °C). Au-delà de la quantité intrinsèque des ligands, ces études montrent fermement que le processus de sélection réalisé et l'approche DEL pour la sélection de ligands de G-quadruplex est viable et prometteuse.

[Sous-partie 2.5 : Modélisation moléculaire.](#)

Afin de rationaliser les résultats obtenus et d'identifier les déterminants des interactions favorables et moins favorables, nous avons réalisés, en collaboration avec le Dr. Hélène Jamet, des études de mécanique moléculaire.

Le G-quadruplex issu du promoteur d'oncogène Myc a été utilisé comme cible afin de déterminer les interactions probables avec les dérivés **3-360A-s**, **3-360A-I** et **5-360A-s**. La structure tridimensionnelle de Myc (id PDB : 2L7V), résolue en 2011 par Résonance Magnétique Nucléaire (J. Dai¹¹⁷ et *al.*), a servi de base pour les études de dynamique moléculaire. Dans cette publication, la cible Myc est en interaction avec 2 composés

« Quindolines », dont l'une est disposée sur la tétrade supérieure et l'autre sur la tétrade inférieure du G-quadruplex. Dans le cadre de la modélisation du dérivé **360A**, les deux composés quindolines ont été retirés de la séquence Myc et remplacés par un seul ligand **360A** en interaction avec la tétrade externe de la cible G-quadruplex. A partir du **360A** docké sur la structure Myc, la séquence peptidique GWRRG-K^{N3}G-GRRAG a été adjointe au Pyridocarboxamide afin d'obtenir la structure finale du **3-360A-s**. Une fois le **360A** docké sur la cible Myc et le peptidique aligné, des études de dynamique moléculaire ont été réalisées dans un champ de force AMBER (en. *Assisted Model Building and Energy Refinement*) spécialement adapté à la modélisation de protéines, d'ADN et d'ARN.

Initialement, la dynamique de ligand **3-360A-s** a été effectuée sur un temps de 500 ns. L'analyse par cluster des conformations ont montré que le ligand **360A** reste fidèlement à son emplacement d'origine, c'est-à-dire en interaction avec la tétrade externe de Myc, et que le dodécapeptide se déploie différemment autour de la séquence quadruplex. Une dizaine de cluster ont été obtenus après dynamique moléculaire pour un total de 4244 frames et un cluster majoritaire, appelé C0, a été déterminé avec 1313 frames.

Au sien du cluster C0, différents types d'interactions entre le ligand **3-360A-s** et la cible Myc ont été retrouvés comme des interactions hydrogènes (contacts polaires), des interactions π (interactions non-covalentes entre deux systèmes π) ainsi que des interactions π -cation. Les différentes interactions recensées entre le **3-360A-s** et la cible Myc sont illustrées sur la Figure 117 ci-après. Les contacts polaires sont exclusivement réalisés au travers des acides aminés Arginines. Au sein du pentapeptide du [Cycle 1] de la librairie combinatoire, c'est l'Arg(R)₃ qui interagit à la fois avec le phosphate de la Thymidine₇ (dT₇) mais également avec le phosphate de la Guanidine₈ (dG₈). D'autre part l'Arg(R)₉, issu du pentapeptide du [Cycle 4], interagit exclusivement avec les groupements phosphates de la Guanosine₁₀ (dG₁₀), de la Thymidine₁₁ (dT₁₁) et de l'Adénosine₁₂ (dA₁₂). Les interactions non-covalente, de nature π -cation et π - π , ont été retrouvés par interaction de l'Arg(R)₄ et du Trp(W)₂ avec le centre Pyrimidine de la Thymidine₇ (dT₇), respectivement.

Après analyse du meilleur composé issu de la librairie combinatoire, le **3-360A-s**, les dérivés **3-360A-I** et **5-360A-s** ont également été analysés. D'après les études de dynamiques moléculaires réalisées jusqu'alors, il a été observé que la longueur de chaîne de la lysine portant la fonction triazole permettait de délivrer efficacement le **360A** sur la tétrade externe ainsi que le peptide autour de la structure quadruplex. Il était alors intéressant d'étudier le comportement du Pyridocarboxamide à chaîne longue (**360A-I**) en présence du peptide dans le but d'observer l'influence de la longueur de chaîne sur l'interaction avec le quadruplex Myc. Dans un deuxième temps, le composé « non-enrichi » **5-360A-s** également été étudié afin d'observer ces interactions avec la cible. Au cours des expériences de sélections, le **5-360A-s** était faiblement enrichi voire déplété vis-à-vis de plusieurs cibles quadruplex donc la cible Myc. L'intérêt de cette étude de dynamique moléculaire sur le **5-360A-s** est de montrer que les interactions de la chaîne peptidique **5** avec la cible sont amoindries vis-à-vis de la cible quadruplex Myc.

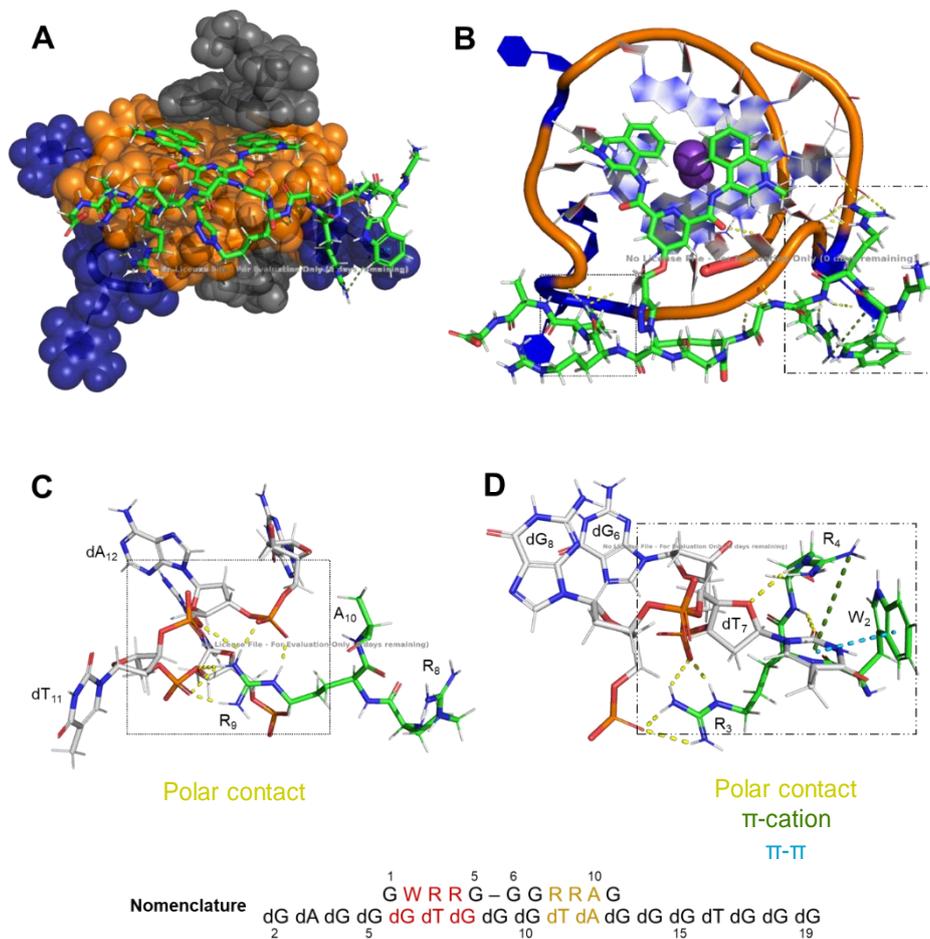


Figure 117 : A/ Modélisation moléculaire du ligand 3-360A-s en interaction avec la cible Myc22. B/ Détermination des interactions de la chaîne peptidique avec le tétraplexe de guanines. C/ et D/ Caractérisation de contacts polaires, hydrophobes et ioniques.

Une nouvelle dynamique a ainsi été réalisée sur les 3 ligands pendant un temps de 100 ns. L'analyse des clusters après chaque dynamique a été réalisée et l'énergie d'interaction du ligand pour la cible Myc a été calculée à partir du cluster le plus stable. Dans le cadre du ligand **3-360A-s**, le cluster le plus stable trouvé est similaire à celui discuté auparavant (Figure 118, **3-360A-s**). Si on regarde le cluster la plus stable retrouvé pour le ligand à chaîne longue, **3-360A-I**, on peut apercevoir que l'orientation générale de la séquence peptidique n'est pas disposée de la même manière que pour le ligand **3-360A-s**. En effet, l'augmentation de la longueur de chaîne du bras du Pyridocarboxamide complique le déploiement du peptide sur le G-quadruplex. En conséquence, les interactions entre la séquence peptidique et l'édifice G-quadruplex sont moindres, ce qui est cohérent avec la baisse (relative) d'affinité et de stabilité retrouvé pour ce ligand contre la cible du proto-oncogène Myc (Figure 118, **3-360A-I**). Enfin, l'évaluation du cluster le plus stable obtenue avec le **5-360A-s** montre que le peptide déploie ces chaînes pentapeptidiques de façon plus aléatoire sur l'édifice. En effet, quasiment aucune interactions n'ont été observées lorsque le peptide oriente ces bras vers la tétrade supérieure et inférieure (Figure 118, **5-360A-s**). Encore une fois, la quasi disparition des interactions

illustrée par la dynamique moléculaire du **5-360A-s** sur la cible Myc permet de rationaliser la baisse d'interaction du dérivé peptido-organique pour sa cible.

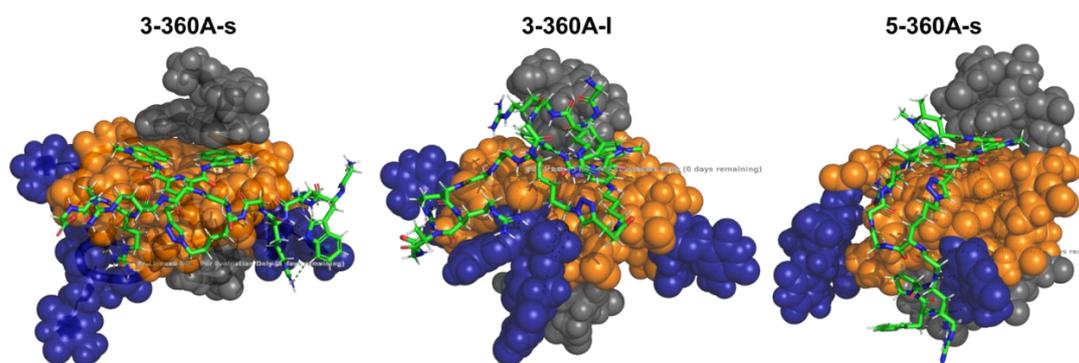


Figure 118 : Comparaison du déploiement tridimensionnel des dérivés Pyridocarboxamides (3-360A-s, 3-360A-l et 5-360A-s) par modélisation moléculaire sur la cible Myc2.

Chapitre 5 — Conclusion et perspectives.

L'objectif de ce projet de thèse a été d'identifier de nouveaux composés affins des structures G-quadruplex en utilisant la chimie combinatoire associée à une technologie d'encodage par l'ADN récemment développée. L'application de cette technologie, qui a conduit à un regain d'intérêt considérable pour les approches combinatoires et qui a permis l'identification de nombreux hits reconnaissant des protéines, au ciblage des acides nucléiques n'en est qu'à son balbutiement. Dans un premier temps, une librairie combinatoire encodée composés de 190 440 molécules uniques (DEL1) a été synthétisée par assemblage itératif de code-barres ADN et de motifs (poly)hétéroaromatiques et peptidiques. Cet assemblage multi-étape a permis l'obtention d'environ 5 nmoles de librairie encodée, une quantité suffisante à la réalisation de centaines d'expériences de sélection par affinité. Au cours de cette synthèse, l'incorporation des code-barres ADN (T1 à T4) et du Closing primer a pu être contrôlée efficacement par électrophorèse sur gel d'agarose. Des étapes de purification par microcentrifugation sur membranes à diamètre de pores variées ou par HPLC ont été réalisées au cours de l'assemblage afin d'éliminer efficacement les excès de code-barres. L'incorporation chimique des Building blocks des [Cycle 1 à 4] a été contrôlée par chromatographie couplée à la spectroscopie de masse (UPLC-MS). Toutefois, à partir du troisième cycle de la DEL1 l'interprétation des analyses UPLC-MS n'a pas permis de différencier les composés résultants des réactions de cycloaddition (DEL1-[Cycle 3]) ou de couplage peptidique (DEL1-[Cycle 4]) des composés de départ. Ainsi, des expériences de Cap&Catch associées à des expériences de gel retard (EMSA) ont permis de maintenir et de confirmer la pureté de la DEL1.

Des expériences de sélection *in vitro* par affinité ont ensuite été effectuées à l'échelle de 5 pmoles de DEL1 sur différentes cibles acides nucléiques. Dans le cadre de la première campagne de sélection, plus d'une centaine de sélections distinctes ont été réalisées sur des cibles G-quadruplex de topologies et d'origines variées. Parmi celles-ci, 28 ont été retenues afin d'être séquencées à haut-débit après amplification des code-barres ADN par PCR. Le traitement des données de séquençages Illumina® a alors permis de dégager des tendances nettes pour la reconnaissance des motifs G-quadruplex et d'identifier des ligands prometteurs tels que **1-PDP**, **2-PDC3** et **3-360A-s**.

Lors du traitement bioinformatique des différentes expériences de sélections réalisées, une tendance générale relative aux motifs (poly)hétéroaromatiques de la DEL1-[Cycle 3] a pu être constatée. En effet, les motifs **PDP**, **PDC3** et **360A** sont systématiquement au cœur (sauf exceptions ponctuelles) des composés les plus enrichis lors des sélections contre les cibles G-quadruplex. Au contraire, les autres motifs du [Cycle 3] (**Q**, **FP**, **NDI**, etc.) ont principalement été déplétés. Ce phénomène s'explique sans doute par la disparité des propriétés d'associations intrinsèque de ces différents fragments qui présentent des K_d de l'ordre de la dizaine à la centaine de nanomolaire pour les composés **PDP**, **PDC3** et **360A** et des K_d à l'échelle du micromolaire pour les dérivés **Q**, **FP** et **NDI**. Ainsi, la composition même de la DEL1 présente un biais contre la sélection de composés à motif hétéroaromatique faiblement affins pour les structure G-quadruplex. Une deuxième tendance majeure concerne la longueur

du lien entre le motif hétéroaromatique [Cycle 3] et le motif peptidique associée [Cycle 2]. Globalement les meilleurs ligands présentent un lien de même longueur, généralement associant motifs hétéroaromatique fonctionnalisés par un alcyne à chaîne courte et fragment peptidique dont l'azoture est portée par un dérivé de Lysine.

Afin de valider les données issues des sélections et du séquençage, des composés contrôles, tronqués ou peu enrichis, ont également été synthétisés portant à onze le nombre de molécules re-synthétisés sans code-barres et soumis à évaluation biophysiques. Leurs constantes d'association avec divers G4 (et contrôles) ont été obtenues (titration de fluorescence, BLI) ainsi que leurs propriétés de stabilisation des G4 (et contrôles) (FRET-Melting).

Les données d'affinités et de stabilisations obtenues ont ainsi pu confirmer que les dérivés **1-PDP** et **3-360A-s** sont plus affins et plus stabilisants que les ligands parents dont ils dérivent. En particulier, le composé **3-360A-s** est capable d'interagir avec les différentes cibles G-quadruplex utilisées avec un K_d d'environ 10 nM et de les stabiliser à plus de 35°C. Toutefois, malgré des valeurs très intéressantes vis-à-vis des structures G-quadruplex, nous avons détecté la propensité de **3-360A-s** à interagir et à stabiliser faiblement un mime d'ADN duplex.

Globalement les résultats obtenus, par l'assemblage et le criblage de cette première librairie encodée DEL1 spécialement dédiée à la reconnaissance des structures ADN G-quadruplex, permettent de confirmer que la technologie DEL est adaptée à la découverte de nouveaux ligands pour le ciblage de structures d'origines nucléiques.

Afin d'aller plus loin dans l'identification de ligands d'acides nucléiques et de structures G-quadruplex, une seconde campagne de sélection a été réalisée à partir de la DEL1. Lors de cette nouvelle campagne, des expériences de compétitions ont été effectuées dans l'intérêt de pouvoir isoler d'éventuels composés spécifiques d'une ou l'autre topologie de structure G-quadruplex. De plus, des cibles ARN G-quadruplex ainsi que d'autres cibles acides nucléiques (jonction triple, i-motifs) ont été utilisées. Plus de 150 expériences de sélections ont été réalisées et 48 ont été choisies afin d'être séquencées à haut-débit. Le séquençage a été réalisé et les données bioinformatiques obtenues. Toutefois, l'interprétation de ces données de séquençages n'a pas été effectuée au terme de cette thèse. Un objectif à court terme du projet sera donc de traiter les données obtenues afin de sélectionner différents ligands aux propriétés intéressantes, d'en faire la synthèse à grande échelle et d'explorer les propriétés biophysiques de ces ligands.

Au cours de cette thèse, une deuxième librairie encodée par l'ADN a également été synthétisée afin de cibler spécifiquement les structures ADN G-quadruplex. L'assemblage itératif de séquences peptidiques, de motifs polyhétéroaromatiques et de diversité chimique variée a ainsi permis d'obtenir une librairie de 696 200 composés encodés distincts. 2 nmoles de DEL2 ont été obtenues avec un rendement global de 3.4%. En parallèle de l'écriture de ce manuscrit, différentes expériences de sélections ont été réalisées, principalement sur les cibles ADN G-quadruplex, et soumises à séquençage. Après retour de séquençage, les données bioinformatiques obtenues devront elles aussi être interprétées afin d'identifier, cette fois-ci, des ligands de structures distinctes de ceux identifiés via la DEL1.

Lors de ma thèse, j'ai également eu l'occasion de travailler sur une troisième librairie, DEL3, visant à assembler des peptoïdes cycliques ciblant les structures G-quadruplex. Ce projet, non décrit dans ma thèse, a été initié au laboratoire et des études d'assemblage préliminaire de 72 building blocks R-NH₂ ont été réalisées. A terme, cette librairie sera constituée de 3 cycles itératifs formant ainsi 373 248 entités peptoïdiques cycliques distinctes. Des conditions chimiques pour la synthèse de la DEL3 ont alors été testées, et des conditions de couplages peptidiques (sur amines primaires et secondaires) et de substitutions nucléophiles (S_N2) ont été mise au point. L'assemblage de cette librairie sera poursuivi au laboratoire sur la base des résultats obtenus.

A partir des deux librairies fonctionnelles, il est possible d'imaginer plusieurs suites à ce projet. En effet, des sélections ont déjà été initiées sur différentes cibles d'acides nucléiques (jonction triple, i-motifs) mais peuvent être encore être élargies à d'autres cibles d'intérêts biologiques. Des conditions de sélections plus exotiques permettant d'appliquer de nouvelles stringences (force ionique, etc.) ou encore permettant de mettre en concurrence les cibles (contre-sélections) pourront être évaluées. Plus tard, des expériences de dynamiques appelées DNA-Encoded Dynamic Chemical library pourrait également être imaginée à partir des DEL1 et DEL2 pour l'identification de ligands spécifiques de différents gènes porteurs de G-quadruplex. En effet, dans cette technique de dynamique combinatoire l'amine primaire présente en fin du [Cycle 4] des deux librairies DEL1 et DEL2 pourrait être utilisée afin d'assembler de manière dynamique de nouveaux éléments de reconnaissance de l'ADN double brin porteur de fonctions aldéhydes. Cette reconnaissance pourrait par exemple être adressée à la séquence ADN duplex présente en amont des séquences riches en guanines capables de former des structures G-quadruplex. Par ailleurs, l'approche DEL dite à « dual pharmacophore » décrite dans l'introduction de ce manuscrit semble également parfaitement adaptée aux cibles G4 et pourrait permettre de cumuler reconnaissance des plateaux et reconnaissance des boucles ou des sillons.

Enfin, les perspectives de ce travail sont également à terme d'évaluer dans des systèmes cellulaires pertinent les propriétés et le devenir des ligands identifiés par la technologie DEL.

Chapitre 6 — Materials and methods.

1. Experimental Section

1.1 General information, material and instrumentation

1.1.1 Materials

Spin-X columns were purchased from Corning-Costar (Acton, MA, USA). Microcentrifugal tubes (Amicon[®] Ultra-0.5mL, 10, 30 and 50 kDa) were purchased from Merck KGaA (Darmstadt, DE). UltraPure[™] Agarose-1000 was purchased from Invitrogen[™] (Carlsbad, USA). qPCR 96-well plates (Skirted, Polypropylene) were purchased from Agilent Technologies (Santa Clara, CA, USA). Fluorometric titration 96-well plates (black, flat-bottom polystyrene non-binding (NBS) microplates) were purchased from Corning[®], Fisher Scientific (Loughborough, UK).

1.1.2 Beads and Resins

Fmoc-Rink-Amide PEG MBHA Resin was purchased from Iris Biotech (Marktredwitz, DE). Streptavidin Sepharose[®] High Performance resin was purchased from GE Healthcare (Uppsala, SE). Streptavidin-coupled magnetic Dynabeads (Dynabeads[™] MyOne[™] Streptavidin T1) was purchased from Invitrogen[™] (Carlsbad, USA).

1.1.3 Chemical compounds

All solvents and standard chemical reagents used were of the highest purity commercially available. Fmoc N-terminal protected pentapeptides (Purity >95%) were purchased from GenScript (Piscataway, NJ, USA) – used in DEL1 & DEL2. Amino building blocks (Purity > 95%) – engaged in diazo-transfer reaction – DEL2 [Cycle 3] – were purchased from Enamine (Latvia, Lettonie).

1.1.4 Oligonucleotides and Enzymes

T4 DNA Ligase and buffers (T4 DNA ligase buffer) were purchased from New England Biolabs (Ipswich, MA, USA). PowerUp[™] SYBR[™] Green Master Mix 2X was purchased from Applied Biosystems[™] (Waltham, USA). All oligonucleotides (DNA tags, biotin and fluorophore modified oligonucleotides) were purchased HPLC purified from Integrated DNA Technologies (Leuven, BE).

1.1.5 Mass Spectrometry and NMR

ESI-MS spectra were performed on an Esquire 3000 spectrometer from Bruker (Ettlingen, Germany) or on an Acquity UPLC/MS system from Waters Corporation (Milford, USA) equipped with a SQ Detector 2 within a column heater set at 40°C for peptide analysis or 60°C for oligonucleotide analysis. Peptides were analyzed in positive mode and oligonucleotides and conjugates in negative mode. The mass spectra display either the relative abundance of

ion signals (total ion counts) against the m/z ratios or the total ion counts against the m/z ratios. All solvents and reagents used were of highest purity commercially available. NMR spectra were obtained with either a 400 or 500 MHz Avance III Bruker spectrometer (Ettlingen, Germany).

1.1.6 Ethanol Precipitation of DNA

Ethanol precipitation of DNA was performed by adding 10% (v/v) of 5 M NaCl buffer and 220% (v/v) of ethanol relative to the volume of the DNA sample. After 2 h incubation at -80°C , DNA samples were centrifuged in a table-top centrifuge for 20 min (15,000 rpm) at 4°C . After removal of the supernatant, pellets were washed with 500 μL of EtOH 70% (v/v) and centrifuged for 20 min (15,000 rpm) at 4°C . Pellets were dried under reduced pressure and dissolved in H₂O or an adequate buffer.

1.1.7 Agarose gel electrophoresis

Agarose gel electrophoresis was performed using 4% or 5% UltraPure™ Agarose (Invitrogen) in TBE buffer. Electrophoresis gels were precast using GelRed® Nucleic Acid Gel Stain (10,000X; Biotium) and 90 V / 300 mA was applied for 120 min. Samples (2 μL) were mixed with 2 μL DNA Gel Loading Dye (6X; Thermo Scientific) and H₂O (or additives if required) were added to reach a final volume of 12 μL before loading. Afterwards, gels stained DNA were detected by UV transillumination using a Gel Doc XR+ System (Bio-Rad).

1.2 Synthesis of DEL-1 and DEL-2 building blocks

DEL1-[Cycle 2]

Fmoc-Dap(N₃)-OH (building block ID: D)³³⁴, Fmoc-Lys(N₃)-OH (building block ID: K)³³⁵, azido lysine (building block ID: Ke)³³⁶, the tris(aminoethyl)amine (TREN) derivative (Building block ID: T), the Fmoc-Dap(N₃)-Gly-OH (DG) derivative (Building block ID: DG), the Fmoc-Gly-Dap(N₃)-Gly-OH derivative (Building block ID: GDG), the Fmoc-Lys(N₃)-Gly-OH derivative (Building block ID: KG) and the Fmoc-Gly-Lys(N₃)-Gly-OH derivative (Building block ID: GKG) were synthesized in the lab by C. Van Wesemal (DCM, I2BM).

DEL1-[Cycle 3]

The alkyne modified pyridostatin analog (building block ID: PDP-alk)¹⁸¹, and the Benzofuro[3,2-b]quinolone derivatives (building block ID: Q-s-alk and Q-l-alk)³³⁷, were synthesized by C. Van Wesemael (DCM, I2BM) as previously described.

The alkyne modified phenanthroline analog (building block ID: PDC3-alk)³¹⁰ and the alkyne modified pyridocarboxamide derivatives (building block ID: 360A-l-alk)³³⁸ were synthesized as previously described by M.-P. Teulade-Fichou's group (Paris-Sud, Institut Curie).

The alkyne modified pyridocarboxamide derivatives (building block ID: 360A-s-alk) was synthesized as previously described by J.-F. Riou's group (Paris, MNHN).

The alkyne furopyridazinone analog (building block ID: FP-alk)¹³⁹ was synthesized as previously described by M. Giustiniano (Naples – Italy, Department of Pharmacy).

The alkyne carbazole analog (building block ID: CZ-alk)³³⁹ was synthesized as previously described by J. Dash's group (Kolkata – India, IACS).

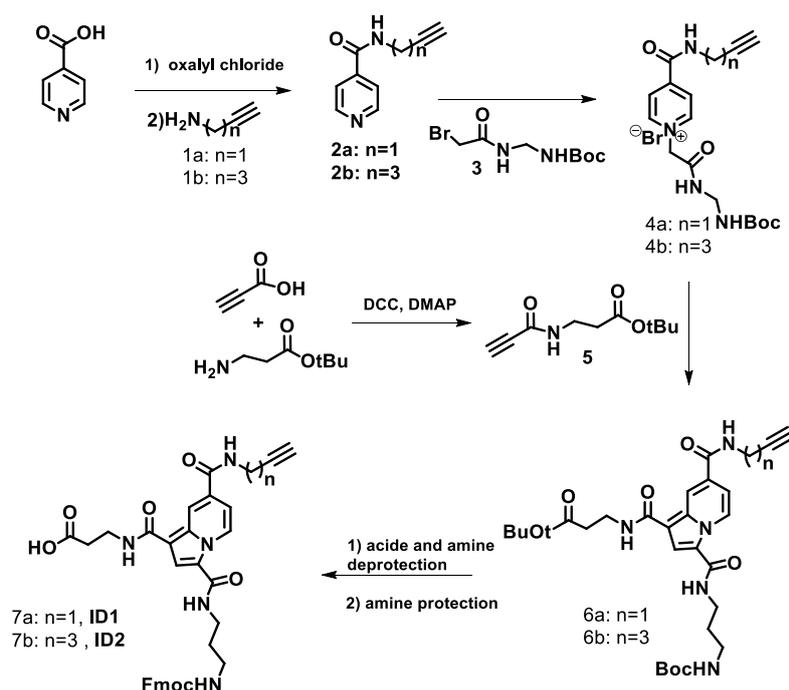
The alkyne naphthalene diimide analog (building block ID: NDI-alk)³⁴⁰ was synthesized as previously described by F. Doria's group (Pavia – Italy, Department of Chemistry).

DEL2-[Cycle 2]

The Fmoc-Dap(alkyne)-OH derivative (Building block ID: DE) was synthesized in the lab by C. Van Wesemal (DCM, I2BM).

The trifunctional trioxazole derivative (Building block ID: TO) was synthesized by adapting previously described procedures³⁴¹ and using Boc-Lys(Tfa)-OH, Boc-Lys(N₃)-OH and Boc-DAP(pent-4-ynamide)-OH synthons. This analog was synthesized by R. Lartia in the lab (DCM, I2BM).

The indolizine derivatives was synthesized (in collaboration with I. Baussanne and M. Demeunynck –DPM, Grenoble) according to Scheme S1.



Scheme S1 : Synthesis of building block ID: ID1 and ID2.

N-Prop-2-ynyl-isonicotinamide 2a³⁴²: To a solution containing isonicotinic acid (2.46 g, 20.0 mmol) and DMF (0.025 mL, 0.32 mmol) dissolved in dry dichloromethane (50 mL) was added a solution of oxalyl chloride (1.86 mL, 22 mmol) dissolved in dry dichloromethane (5 mL) dropwise at 0 °C. After stirring the reaction mixture at r.t. for 3 h, the solution was concentrated under reduced pressure to give the corresponding acyl chloride. To a solution composed of propargylamine 1a (1.4 mL, 22 mmol) and triethylamine (4.2 mL, 30 mmol) dissolved in dry

dichloromethane (50 mL) was added dropwise at 0 °C a solution of the corresponding acyl chloride (1.55 g, 10.9 mmol) previously dissolved in dry dichloromethane (15 mL). After stirring at r.t. for 3 h, the mixture was washed with 50 mL of HCl 1 M followed by an extraction with dichloromethane (3 x 50 mL). The combined organic layers were then dried over anhydrous sodium sulfate and then concentrated under reduced pressure to afford the crude product 2a. The residue was subjected to a silica gel column chromatography with cyclohexane/ethyl acetate: 1/9 as eluant. After evaporation of the solvent, the desired compound 2a was obtained as a white solid (0.866 g, 50 % yield). The analytical results of compound 2a were identical with the ones described in the literature.

N-Pent-2-ynyl-isonicotinamide 2b: To a solution containing isonicotinic acid (0.370 g, 3 mmol) and DMF (3 drops) dissolved in dry dichloromethane (5 mL) was added a solution of oxalyl chloride (0.280 mL, 3.3 mmol) dissolved in dry dichloromethane (2 mL) dropwise at 0 °C. After stirring the reaction mixture at r.t. for 3 h, the solution was concentrated under reduced pressure to give the corresponding acyl chloride. To a solution composed of 1-amino-4-pentyne³⁴³ 1b (0.221 g, 2.6 mmol) and triethylamine (0.700 mL, 5 mmol) dissolved in dry dichloromethane (5 mL) was added dropwise at 0 °C a solution of the corresponding acyl chloride (1.55 g, 10.9 mmol) previously dissolved in dichloromethane (5 mL). After stirring at r.t. for 3 h, the mixture was washed with 50 mL of HCl 1 M followed by an extraction with dichloromethane (3 x 50 mL). The combined organic layers were then dried over anhydrous sodium sulfate and then concentrated under reduced pressure to afford the crude product 2b. The residue was subjected to a silica gel column chromatography with a step gradient of cyclohexane/ethyl acetate: 2/9 to 0/10. After evaporation of the solvent, the desired compound 2b was obtained as a colorless oil (0.278 g, 57 % yield).

¹H-NMR (400 MHz, CDCl₃): δ 8.79 (d, 2H, J = 5.9 Hz ; 2CH pyr), 7.70 (d, 2H, J = 5.8 Hz ; 2CH pyr), 6.71 (br s, 1H ; NH), 3.67 (td, 2H, J = 6.8, 6.0 Hz ; CH₂-NH), 2.39 (td, 2H, J = 6.8, 2.8 Hz ; CH₂-Cq), 2.09 (t, 1H, J = 2.8 Hz ; CH alkyne), 1.93 (tt, 2H, J = 6.8 Hz : CH₂-CH₂-CH₂).
¹³C NMR (100 MHz, CDCl₃): δ 165.48 (CO), 150.28 (2CH pyr), 142.01 (Cq pyr), 121.03 (2CH pyr), 83.55 (Cq alkyne), 69.61 (CH alkyne), 39.65 (CH₂-NH), 27.75 (CH₂-Cq), 16.38 (CH₂-CH₂-CH₂). ESI MS(+) Calcd. for C₁₁H₁₂N₂O: 188.09 Found 188.97 [M+H]⁺.

Pyridinium 4a : A mixture of isonicotinamide 2a (0.208 g, 1.3 mmol) and tert-butyl 3-(2-bromoacetamido)propylcarbamate³⁴⁴ 3 (0.192 g, 0.65 mmol) in ethyl acetate (5 mL) was stirred 6 h at 70 °C then 12 h at r.t.. The precipitate was filtered off, washed with ethyl acetate, diethyl ether and dried. The pyridinium salt 4a was thus isolated as a hygroscopic white solid (0.239 g, 80 % yield).

¹H-NMR (400 MHz, CD₃OD): δ 9.12 (d, 2H, J = 6.8 Hz ; 2CH pyr), 8.47 (d, 2H, J = 6.8 Hz ; 2CH pyr), 6.65 (br s, 1H ; NH), 5.57 (s, 2H ; N-CH₂-CO), 4.27 (d, 2H, J = 2.4 Hz ; CH₂-NH), 3.34 (m; CH₂-NH + MeOH), 3.14 (m, 2H ; CH₂-NH), 2.74 (t, 1H, J = 2.4 Hz ; CH alkyne), 1.74 (tt, 2H, J = 6.8 Hz : CH₂-CH₂-CH₂), 1.45 (s, 9H ; 3CH₃).

¹³C NMR (100 MHz, CD₃OD): δ 165.57 (CO), 163.73 (CO), 158.62 (Cq pyr), 150.74 (CO), 148.55 (2CH pyr), 126.83 (2CH pyr), 80.09 (Cq alkyne), 73.08 (CH alkyne), 62.90 (N-CH₂-

CO), 38.55 (CH₂-NH), 38.31 (CH₂-NH), 30.62 (NH-CH₂-CH₂-CH₂-NH), 30.51 (CH₂-NH), 28.79 (3CH₃).

Pyridinium 4b : A mixture of isonicotinamide 2b (0.935 g, 4.97er mmol) and tert-butyl 3-(2-bromoacetamido)propylcarbamate³⁴⁴ 3 (1.47 g, 4.97 mmol) in ethyl acetate (50 mL) was stirred 12 h at 70 °C then 12 h at r.t.. The precipitate was filtered off, washed with ethyl acetate, diethyl ether and dried. The pyridinium salt 4b was thus isolated as a hygroscopic white solid (1.58 g, 66 % yield).

¹H-NMR (400 MHz, CDCl₃): δ 9.01 (d, 2H, J = 6.8 Hz ; 2CH pyr), 8.37 (d, 2H, J = 6.8 Hz ; 2CH pyr), 6.55 (br s, 1H ; NH), 5.47 (s, 2H ; N-CH₂-CO), 3.50 (t, 2H, J = 7.0 Hz ; CH₂-NH), 3.25 (m; CH₂-NH + MeOH), 3.04 (m, 2H ; CH₂-NH), 2.25 (td, 2H, J = 7.0, 2.6 Hz ; CH₂-Cq), 2.23 (m, 1H ; CH alkyne), 1.81 (tt, 2H, J = 6.8 Hz : CH₂-CH₂-CH₂), 1.65 (tt, 2H, J = 6.8 Hz : CH₂-CH₂-CH₂), 1.36 (s, 9H ; 3CH₃).

¹³C NMR (100 MHz, CD₃OD): δ 165.60 (CO), 164.16 (CO), 158.63 (Cq pyr), 151.33 (CO), 148.43 (2CH pyr), 126.74 (2CH pyr), 84.10 (Cq alkyne), 70.28 (CH alkyne), 61.65 (N-CH₂-CO), 40.76 (CH₂-NH), 38.53 (CH₂-NH), 38.28 (CH₂-NH), 30.62 (NH-CH₂-CH₂-CH₂-NH), 29.05 (CH₂-Cq), 28.78 (3CH₃), 16.81 (CH₂-CH₂-CH₂).

ESI MS(+) Calcd. for C₂₁H₃₁N₄O₄⁺: 403.23 Found 403.21 [M]⁺.

Propiolamide 5 : To a stirred solution containing protected β-alanine ter-butyl ester hydrochloride (0.726 g, 4 mmol), DMAP (24 mg, 0.2 mmol) and propiolic acid (0.271 mL, 4.4 mmol) in dichloromethane (10 mL) at 0°C was added slowly a solution of DCC (0.908 g, 4.4 mmol) dissolved in dichloromethane (10 mL). The reaction mixture was stirred at r.t. for 24 h. After the precipitated urea filtration, the solution was washed with water, NaHCO₃ aq. sat. solution and brine before being dried over anhydrous magnesium sulfate and then concentrated under reduced pressure to afford the crude product. The residue was subjected to a silica gel column chromatography with cyclohexane/ethyl acetate : 8/2 as eluant. After evaporation of the solvent, the desired compound 5 was obtained as a yellow oil (0.497 g, 63 % yield).

¹H-NMR (400 MHz, CDCl₃): δ 6.46 (br s, 1H ; NH), 4.03 (d, 2H, J = 5.2 Hz ; CH₂), 2.87 (s, 1H ; CH). 1.53 (s, 9H ; 3CH₃).

¹³C NMR (100 MHz, CDCl₃): δ 168.6 (CO), 151.86 (CO), 82.96 (Cq alkyne), 73.88 (CH alkyne), 42.21 (CH₂), 28.04 (3CH₃).

Indolizine 6a : A mixture of pyridinium 4a (0.248 g, 0.54 mmol), propiolamide 5 (0.107 g, 0.54 mmol) and NEt₃ (0.148 mL, 1.1 mmol) in acetonitrile (8 mL) was stirred 6 h at 50 °C then 12 h at r.t.. The solution was concentrated under reduce pressure. The residue was subjected to a silica gel column chromatography with a step gradient of cyclohexane / ethyl acetate : 3/7 to 2/8. After evaporation of the solvent, the desired compound 6a was obtained as a yellow oil (0.075 g, 24 % yield).

¹H-NMR (400 MHz, CDCl₃): δ 9.34 (t, 1H, J = 6.8 Hz ; CH indo), 8.51 (br s, 1H ; CH indo), 7.54 (br s, 1H ; NH), 7.49 (s, 1H ; CH indo), 7.38 (br s, 1H ; NH), 7.15 (m, 2H ; NH + CH indo), 5.26 (br s, 1H ; NH), 4.27 (dd, 2H, J = 5.2, 2.4 Hz ; CH₂ allyl), 3.69 (m, 2H ; CH₂NH), 3.50 (m, 2H ; CH₂NH), 3.29 (m, 2H ; CH₂NH), 2.64 (t, 2H, J = 6 Hz ; CH₂CO), 2.34 (t, 1H, J = 2.4 Hz ; CH

allyl), 1.78 (m, 2H ; CH₂-CH₂-CH₂), 1.48 (s, 9H ; 3CH₃), 1.47 (s, 9H ; 3CH₃). ¹³C NMR (100 MHz, CD₃OD): δ 172.43 (CO), 165.67 (CO), 164.34 (CO), 161.50 (CO), 157.02 (CO), 134.90 (Cq), 128.14 (Cq), 126.92 (CH indo), 118.36 (Cq), 117.66 (CH indo), 115.83 (CH indo), 111.70 (CH indo), 110.11 (Cq), 81.23 (Cq alkyne), 79.65 (CH alkyne), 37.35 (CH₂-NH), 35.75 (CH₂-NH), 35.37 (CH₂-NH), 35.29 (CH₂-CO), 30.17 (CH₂-NH allyl), 29.76 (CH₂-CH₂-CH₂), 28.47 (3CH₃), 28.13 (3CH₃).

Indolizine 6b : A mixture of pyridinium 4b (0.490 g, 1.03 mmol), propiolamide 5 (0.209 g, 1.06 mmol) and K₂CO₃ (0.209 mL, 1.51 mmol) in acetonitrile (15 mL) was stirred 6 h at 50 °C then 12 h at r.t.. The solution was filtrated before being concentrated under reduce pressure. The residue was dissolved with ethyl acetate, the solution washed twice with water, brine before being dried over anhydrous magnesium sulfate and then concentrated under reduced pressure to afford the pure indolizine 6b as a yellow oil (0.580 g, 94 % yield). ¹H-NMR (400 MHz, CD₃OD): δ 9.56 (t, 1H, J = 7.6 Hz ; CH indo), 8.80 (m, 1H ; CH indo), 7.81 (s, 1H ; CH indo), 7.32 (dd, 1H, J = 7.5, 2.0 Hz ; CH indo), 3.68 (t, 2H, J = 6.8 Hz ; CH₂NH), 3.55 (t, 2H, J = 7.0 Hz ; CH₂NH), 3.47 (m, 2H ; CH₂NH), 3.20 (m, 2H ; CH₂NH), 2.64 (t, 2H, J = 6.8 Hz ; CH₂CO), 2.49 (t, 2H, J = 6.8 Hz ; CH₂ allyl), 2.35 (m, 1H ; CH allyl), 1.89 (m, 2H ; CH₂-CH₂-CH₂), 1.81 (m, 2H ; CH₂-CH₂-CH₂), 1.49 (s, 9H ; 3CH₃), 1.47 (s, 9H ; 3CH₃).

Indolizine 7a – ID1 : To a solution of indolizine 6a (0.075 g, 0.13 mmol) in dichloromethane (3 mL) at 0°C was added TFA (1,5 mL). After stirring at r.t. for 4 h, the solution was concentrated under reduce pressure and coevaporated with methanol. The deprotected indolizine was obtained in a quantitative yield enough pure to be used without purification. ¹H-NMR (400 MHz, CD₃OD): δ 9.60 (dd, 1H, J = 7.4, 1.0 Hz ; CH indo), 8.80 (m, 1H ; CH indo), 7.95 (m, 1H ; CH indo), 7.33 (dd, 1H, J = 7.6, 2.0 Hz ; CH indo), 4.22 (d, 2H, J = 2.4 Hz ; CH₂ allyl), 3.70 (t, 2H, J = 6.8 Hz ; CH₂NH), 3.56 (t, 2H, J = 6.8 Hz ; CH₂NH), 3.07 (t, 2H, J = 7.2 Hz ; CH₂NH), 2.70 (t, 2H, J = 6.8 Hz ; CH₂CO), 2.67 (t, 1H, J = 2.4 Hz ; CH allyl), 2.01 (m, 2H, J = 6.8 Hz ; CH₂-CH₂-CH₂).

The deprotected indolizine was then solubilized in 50 % aq. acetonitrile, FmocOSuc (0.050 g, 0.15 mmol) then K₂CO₃ (0.062 g, 0.45 mmol) were added. After 4 h stirring at r.t., water and ethyl acetate were added and the organic phase separated. The aqueous phase was acidified with HCl 1M and extracted twice with ethyl acetate. The combined organic phases were washed with HCl 1M, water and brine before being dried over anhydrous magnesium sulfate and then concentrated under reduced pressure. The crude product was subjected to a silica gel column chromatography with a step gradient of dichloromethane/methanol: 96/4 to 90/10 with 0.5 % formic acid. After evaporation of the solvent, the desired compound 7a was obtained as a slight yellow solid (0.030 g, 32 % yield).

¹H-NMR (400 MHz, CD₃OD): δ 9.55 (d, 1H, J = 7.2 Hz ; CH indo), 8.80 (s, 1H ; CH indo), 7.81 (m, 3H ; CH indo + 2CH fmoc), 7.66 (d, 2H, J = 7.2 Hz ; 2CH fmoc), 7.40 (dd, 2H, J = 7.4 Hz ; 2CH fmoc), 7.31 (m, 3H ; CH indo + 2CH fmoc), 4.40 (d, 2H, J = 6.8 Hz ; CH₂ fmoc), 4.21 (m, 3H ; CH₂ allyl + CH fmoc), 3.67 (t, 2H, J = 6.8 Hz ; CH₂NH), 3.44 (t, 2H, J = 6.6 Hz ; CH₂NH), 3.27 (t, 2H, J = 6.6 Hz ; CH₂NH), 2.68 (m, 3H ; CH₂CO + CH allyl), 1.82 (m, 2H ; CH₂-CH₂-CH₂). ¹³C NMR (100 MHz, CD₃OD): δ 175.57 (CO), 168.60 (Cq), 167.78 (Cq), 166.56 (CO),

163.41 (CO), 145.30 (CO), 142.63 (CO), 136.89 (Cq), 129.73 (Cq), 128.77 (2CH fmoc), 128.31 (CH indo), 128.12 (2CH fmoc), 126.13 (2CH fmoc), 120.93 (2CH fmoc), 120.22 (CH indo), 119.91 (Cq), 118.09 (CH indo), 112.38 (CH indo), 111.50 (Cq), 80.60 (Cq alkyne), 72.25 (CH alkyne), 67.68 (CH₂ fmoc), 46.13 (CH fmoc), 39.23 (CH₂-NH), 37.63 (CH₂-NH), 36.55 (CH₂-NH), 34.95 (CH₂-CO), 30.85 (CH₂-NH allyl), 30.10 (CH₂-CH₂-CH₂). ESI MS(+) Calcd. for C₃₅H₃₃N₅O₇: 635.24 Found 636.23 [M+H]⁺.

HRMS (ESI) m/z: calcd. for C₃₅H₃₃N₅O₇ [M-H]⁻ 634.23072, obsd 634.22884

Indolizine 7b – ID2 : To a solution of indolizine 6b (0.580 g, 0.98 mmol) in dichloromethane (8 mL) at 0°C was added TFA (4 mL). After stirring at r.t. for 4 h, the solution was concentrated under reduce pressure and coevaporated with methanol. The deprotected indolizine was obtained in a quantitative yield enough pure to be used without purification. ¹H-NMR (400 MHz, CD₃OD): δ 9.63 (dd, 1H, J = 7.5, 1.0 Hz ; CH indo), 8.80 (m, 1H ; CH indo), 7.84 (s, 1H ; CH indo), 7.33 (dd, 1H, J = 7.5, 2.0 Hz ; CH indo), 3.71 (m, 4H ; 2CH₂NH), 3.57 (m, 2H ; CH₂NH), 3.07 (t, 2H, J = 7.0 Hz ; CH₂NH), 2.71 (t, 2H, J = 6.8 Hz ; CH₂CO), 2.59 (t, 2H, J = 6.8 Hz ; CH₂ allyl), 2.35 (m, 1H ; CH allyl), 2.01 (m, 2H ; CH₂-CH₂-CH₂), 1.90 (m, 2H ; CH₂-CH₂-CH₂).

The deprotected indolizine (0.350 g, 0.63 mmol) was then solubilized in a mixture of dioxane/NaHCO₃ aq.sat. (30 mL, v/v: 1/1) at 0 °C. Then FmocCl (0.260 g, 1 mmol) in dioxane (5 mL) was added dropwise to the solution. After 20 h stirring at r.t., water and ethyl acetate were added and the organic phase separated. The aqueous phase was acidified with HCl 1M and extracted twice with ethyl acetate. The combined organic phases were washed with HCl 1M, water and brine before being dried over anhydrous magnesium sulfate and then concentrated under reduced pressure. The crude product was subjected to a silica gel column chromatography with a step gradient of dichloromethane/methanol: 98/2 to 96/4 with 0.5 % formic acid. After evaporation of the solvent, the desired compound 7b was obtained as a slight yellow solid (0.135 g, 20 % yield).

¹H-NMR (400 MHz, CD₃OD): δ 9.46 (d, 1H, J = 7.4 Hz ; CH indo), 8.70 (s, 1H ; CH indo), 7.75 (m, 3H, CH indo + 2CH fmoc), 7.62 (d, 2H, J = 7.2 Hz ; 2CH fmoc), 7.36 (dd, 2H, J = 7.3 Hz ; 2CH fmoc), 7.28 (dd, 2H, J = 7.3 Hz ; 2CH fmoc), 7.23 (d, 1H, J = 7.4 Hz ; CH indo), 4.37 (d, 2H, J = 6.8 Hz ; CH₂ fmoc), 4.16 (t, 1H, J = 6.7 Hz ; CH fmoc), 3.66 (t, 2H, J = 6.6 Hz ; CH₂NH), 3.52 (m, 2H ; CH₂NH), 3.42 (m, 2H ; CH₂NH), 3.25 (t, 2H, J = 6.4 Hz ; CH₂NH), 2.68 (t, 2H, J = 6.7 Hz ; CH₂CO), 2.32 (m, 3H, CH₂ allyl + CH allyl), 1.87 (m, 2H ; CH₂-CH₂-CH₂), 1.80 (m, 2H ; CH₂-CH₂-CH₂).

¹³C NMR (100 MHz, CD₃OD): δ 174.15 (CO), 168.21 (CO), 166.55 (CO), 163.35 (CO), 159.15 (CO), 145.29 (Cq), 142.60 (Cq), 136.89 (Cq), 130.30 (Cq), 128.77 (2CH fmoc), 128.23 (CH indo), 128.12 (2CH fmoc), 126.13 (2CH fmoc), 120.95 (2CH fmoc), 119.85 (CH indo), 119.70 (Cq), 118.04 (CH indo), 112.45 (Cq), 111.23 (CH indo), 84.37 (Cq alkyne), 70.23 (CH alkyne), 68.16 (CH₂ fmoc), 48.61 (CH fmoc), 40.43 (CH₂-NH), 39.23 (CH₂-NH), 37.59 (CH₂-NH), 36.55 (CH₂-NH), 34.99 (CH₂-CO), 30.88 (CH₂-NH allyl), 29.47 (CH₂-CH₂-CH₂). ESI MS(+) Calcd. for C₃₇H₃₇N₅O₇: 663.27 Found 664.29 [M+H]⁺.

HRMS (ESI) m/z: calcd. for C₃₇H₃₆N₅O₇ [M-H]⁻ 662.26202, obsd 662.26067

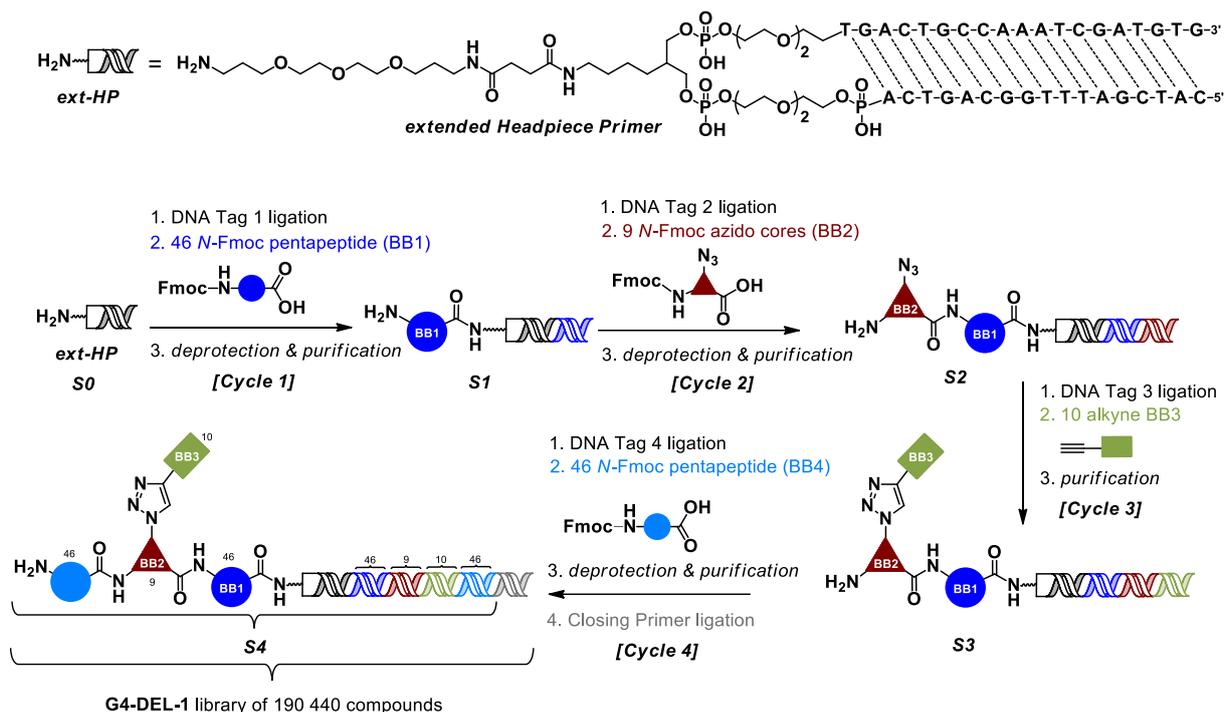
DEL2-[Cycle 3]

General procedure for diazo-transfer reaction: To a solution of NH₂-derivative (1 eq) in 400 μL MeOH/H₂O (3/1, v/v) was added NaHCO₃ (3 eq.) until reaching pH 7-8. Imidazole-1-sulfonyl azide hydrochloride (ISAHC, 3 eq.), aqueous CuSO₄ solution (0.04 eq., 50 μL) and 100 μL of MeOH were added and the mixture was stirred at 25°C for 12h. After completion, determined by UPLC-MS, crude material was concentrated to dry and extracted with brine and EtOAc. The combined organic layers were washed with brine, dried over anhydrous magnesium sulfate, filtrated and concentrated under reduce pressure to afford the corresponding N₃-derivative. The azido-compound was resuspended in DMSO at 133.3 mM and used without further purifications.

1.3 Preparation of the 190.440 G4-DEL-1 library

1.3.1 General synthetic sequence for G4-DEL-1 and barcoding scheme

DNA encoded library **DEL-1** was assembled in four cycles as depicted in Scheme S2. First, an amino PEG3 spacer was installed onto the previously reported extended headpiece primer (ext-HP, Figure S2). Cycle 1 consisted in the parallel ligation of 46 DNA tags 1 (Table S9) followed by the parallel acylation with 46 N-fmoc pentapeptides (Table S4), the N-fmoc deprotection of the pooled reaction mixtures and purification. Cycle 2 consisted in the parallel ligation of 9 DNA tags 2 (Table S10) followed by the parallel acylation with 9 azide-modified N-fmoc aminoacid and peptides (Table S5), the N-fmoc deprotection of the pooled reaction mixtures and purification. Cycle 3 consisted in the parallel ligation of 10 DNA tags 3 (Table S12) followed by the parallel Copper-Catalyzed Azide Alkyne “click” of 10 alkyne containing (hetero)aromatic compounds (Table S7) and purification. Cycle 4 consisted in the parallel ligation of the 46 DNA tags 4 (Table S9), the parallel acylation with 46 N-fmoc pentapeptides (Table S4) followed by the N-fmoc deprotection of the pooled reaction mixtures and purification. The pooled members of the resulting DEL were ligated with a closing primer containing a library tag.



Scheme S2 : Assembly scheme for the four-step split-and-pool synthesis of the DNA-encoded chemical library **DEL-1**.

The **DEL-1** barcoding scheme is depicted in Figure S1.

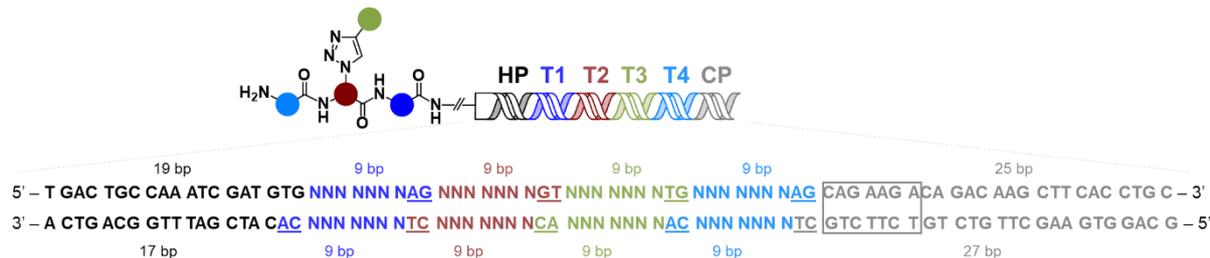


Figure S1 : Barcoding scheme for **DEL-1**. The headpiece (HP) containing the forward primer sequence is in black, DNA Tag 1 in blue, DNA Tag 2 in red, DNA Tag 3 in green, DNA Tag 4 in light blue, Closing Primer in gray, ligation overhangs are underlined, the library barcode is boxed.

1.3.2 General procedures for building block couplings in G4-DEL-1

On-DNA amide coupling protocol 1: To a solution of Fmoc-peptide in DMA (100 μ L, 250 eq., 50 mM) was added an aqueous solution of DMTMM (100 μ L, 250 eq., 50mM) and 200 μ L DMA. The mixture was allowed to incubate at 25°C for 15 min. An aliquot of the growing DEL library (20 nmol, 100 μ M) was then diluted in sodium borate-buffer (150 mM, pH 9.2) and the solution of activated peptide was added to the mixture. The reaction was allowed to proceed for 2 h at 25 °C under orbital agitation. After 2h, a fresh solution of activated peptide consisting of Fmoc-peptide in DMA (100 μ L, 250 eq., 50 mM) and aqueous DMT-MM (100 μ L, 250 eq., 50mM) incubated for 15 min at 25 °C was added to the reaction mixture. The reaction was allowed to proceed overnight at 25°C under orbital agitation. The on-growing DEL-library was

precipitated in EtOH according to the general procedure and the outcome of the reaction was analyzed by UPLC-MS unless specified otherwise.

On-DNA amide coupling protocol 2: To a solution of Fmoc-peptide in DMSO (20 μ L, 50 eq., 50 mM) was added solutions of EDC (22 eq., 30 mM in DMSO), HOAt (4.4 eq., 6 mM in DMSO) and DIPEA (22 eq., 30 mM in DMSO). The mixture was allowed to incubate at 25°C for 15 min. An aliquot of the growing DEL library (20 nmol, 250 μ M) was then diluted in MOPS-buffer (50 mM, pH 8, 5 mM NaCl) and the activated peptide solution was added to the mixture. The reaction was allowed to proceed for 2 h at 25°C under orbital agitation. After 2, 4 and 6h, fresh solutions of activated peptide (prepared as described above) were added to the reaction mixture. 2h after the last addition of activated peptide, the on-growing DEL-library was precipitated in EtOH according to the general procedure and the outcome of the reaction was analyzed by UPLC-MS unless specified otherwise.

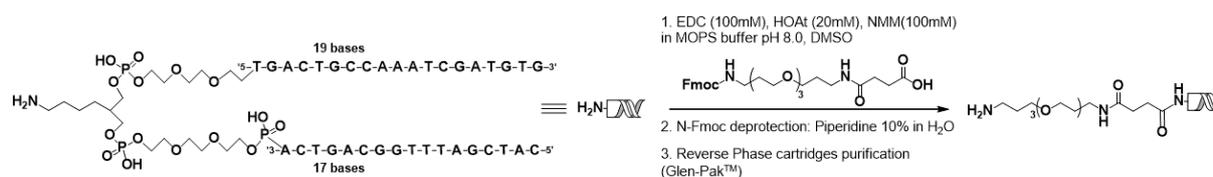
On-DNA amide coupling protocol 3: To a solution of Fmoc-peptide in DMSO (50 μ L, 250 eq., 100 mM) was added a premix solution of EDC/HOAt/NMM in DMSO (50 μ L, 50/10/50 mM respectively) and 100 μ L DMSO. The resulting mixture was allowed to incubate at 30°C for 15 min. An aliquot of the growing DEL library (20 nmol, 200 μ M) was then diluted in MOPS-buffer (50 mM, pH 8.0, 5 mM NaCl) and the activated peptide solution was added to the resulting mixture. The reaction was allowed to proceed for 2 h at 30°C under orbital agitation. After 2h, a fresh solution of activated peptide consisting of Fmoc-peptide in DMSO (500 μ L, 250 eq., 100 mM) and EDC/HOAt/NMM premix solution in DMSO (50 μ L, 50/10/50 mM) incubated for 15 min at 30 °C was added to the reaction mixture. The reaction was allowed to proceed overnight at 30°C under orbital agitation. The on-growing DEL-library was precipitated in EtOH according to the general procedure and the outcome of the reaction was analyzed by UPLC-MS unless specified otherwise.

On-DNA CuAAC reaction: To a solution of on-growing DEL in degassed H₂O (6.5 nmol, 250 μ M) was successively added an alkyne functionalized(hetero)aromatic building block in degassed DMF (40 eq., 13.33 mM), a THPTA solution in degassed H₂O (10 eq., 33.33 mM), aminoguanidine-HCl in degassed H₂O (100 eq., 66.67 mM) and CuSO₄.H₂O in degassed H₂O (10 eq., 6.67 mM). Degassed HEPES-buffer (10X) was then added at 10% of the final volume ($V_f = 182 \mu$ L) and degassed DMF was added to reach 40% v/v. Lastly, sodium ascorbate in degassed H₂O (30 eq., 20 mM) was added to initiate the reaction. The mixture was incubated at 25°C for 2 h under orbital agitation. Identical solutions of THPTA, aminoguanidine, CuSO₄, HEPES, DMF and sodium ascorbate were then added to the mixture and the reaction was allowed to proceed for 2 h at 25 °C under orbital agitation. After 2 h, an aqueous sodium EDTA solution (0.5 M) was added to the solution, the on-growing DEL-library was precipitated in EtOH according to the general procedure and the outcome of the reaction was analyzed by UPLC-MS unless specified otherwise.

1.3.3 Building Blocks Validation

Cycle 1-4 building blocks used to assemble the G4-DEL-1 library were validated on model substrates as described below. The list of all selected building blocks and corresponding validation yields can be found in Table S4.

Cycle 1 & Cycle 4: A total of 78 Fmoc protected pentapeptides (putative cycle 1 & 4 building blocks) sharing the common GNNNG pattern, where N is either A, F, L, P, R, W, were investigated. The coupling efficiency of each Fmoc-protected pentapeptide was assessed using the extended headpiece (Scheme S3) as the starting material, 2 nmoles per reaction.



Scheme S3 : Synthesis of the extended headpiece (ext-HP) from the commercially available Headpiece (HP, IDT code: /5Phos/CAT CGA TTT GGC AGT CA/iSp9//iUniAmM//iSp9/T GAC TGC CAA ATC GAT GTG).

On-DNA amide coupling reactions were investigated with three protocols (vide supra). 46 Fmoc-protected pentapeptides that were found to be coupled with yields above 60% (based UPLC-MS analysis) were deemed suitable for the G4-DEL-1 assembly (Table S9).

Cycle 2: 9 azide-modified N-fmoc amino acid and peptides (putative cycle 2 building blocks) were investigated. The coupling efficiency of each of the 9 putative building blocks was assessed using the extended headpiece grafted with either a GRRAG or a GAWRG pentapeptide (Figure S2) as the starting material at 2 nmole scale. Amide coupling reactions were performed according to general protocol 3 and all 9 building blocks were found to be incorporated with near quantitative yields (based UPLC-MS analysis) and therefore deemed suitable for G4-DEL-1 assembly (Table S5).

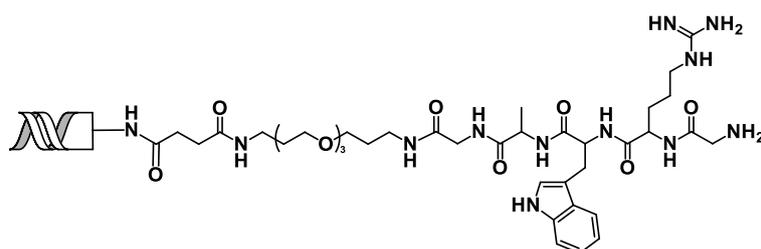


Figure S2. Representative starting material for cycle 2 building block validation

Cycle 3: 10 alkyne containing (hetero)aromatic compounds (putative cycle 3 building blocks) were investigated. The coupling efficiency of each of the 10 putative building blocks was assessed using the extended headpiece grafted with a GAWRG pentapeptide and an azide modified lysine (Figure S3) as the starting material in 2 nmoles scale. CuAAC reactions were performed with the “On-DNA CuAAC reaction”_ protocol (vide supra) and all 10 building blocks were found to be incorporated with quantitative yields (based UPLC-MS analysis) and therefore deemed suitable for G4-DEL-1 assembly (Table S7).

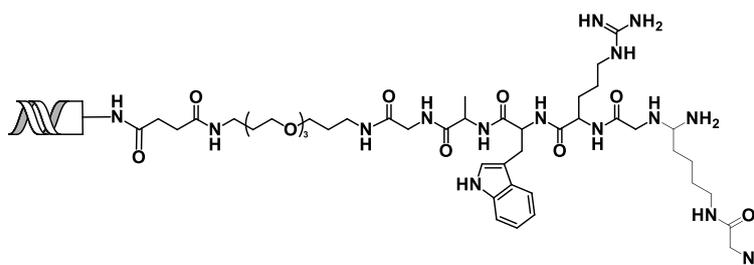


Figure S3 : Representative starting material for cycle 3 building block validation

1.3.4. Assembly of DEL-1:

G4-DEL-1 was assembled mostly according to previously reported procedures.⁷

Procedure for Cycle 1-DEL1

460 nmoles of H₂N-ext-HP **S0** (Scheme S2) were partitioned into 46 wells (10 nmol, 25 μ M in T4 DNA Ligase Reaction Buffer from NEB). In each well distinct Tag 1 duplexes (T1-X, 2 eq, table S9) were ligated using T4 DNA ligase (50 μ M, NEB) for 12h at 16°C. Wells were subjected to ethanol precipitation using the general procedure and ligation successes were assessed using UPLC-MS (MW of 17655 for all ligated products) and agarose gel electrophoresis (according to the general procedure). Next, the selected 46 *N*-Fmoc pentapeptides (BB1) were reacted in parallel onto the T1 encoded amino headpieces (H₂N-ext-HP-T1-X) according to table S4 and using their appropriate on-DNA amide coupling protocol (see table S4 and general coupling procedures). After ethanol precipitations, the coupling efficiencies were confirmed using UPLC-MS and all 46 encoded Fmoc-protected pentapeptides were pooled to provide the crude **S1** library (Figure S4).

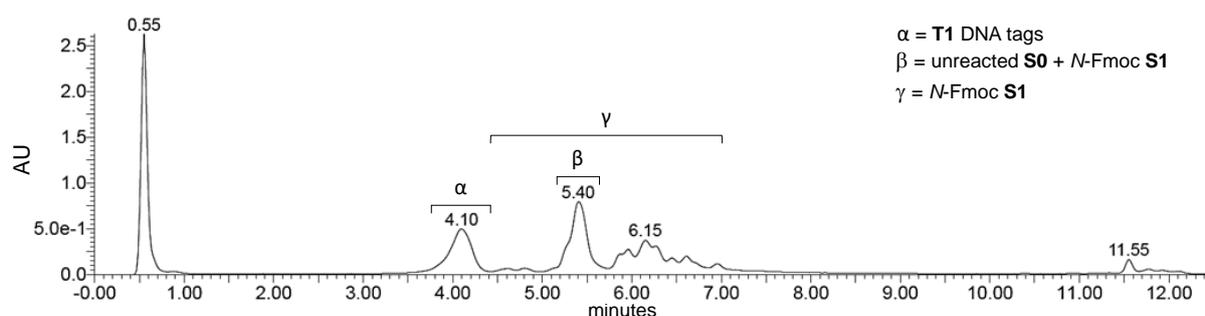


Figure S4 : UPLC chromatogram of crude *N*-Fmoc **S1** library

After an additional ethanol precipitation, the crude **S1** library containing unreacted encoded **S0** was subjected to the cap-and-catch purification procedure described by Neri and colleagues²⁸⁷. The crude **S1** (400 nmol in 700 μ L of TEA.HCl buffer at pH 10.0) was incubated with biotin-NHS (130 eq, 50 mM) in DMSO for 4h at 37°C. After quenching (Tris.HCl, 0.5M, pH 8.1) for 10 min, an EtOH precipitation was performed and the resulting biotin-capped mixture (1 mL in PBS 50 mM NaPi, 100 mM NaCl, pH 7.4) was subjected to streptavidin coated sepharose according to the following procedure. Streptavidin-coated sepharose (1.34 μ L, Streptavidin sepharose high performance, GE Healthcare, 17-5113-01) was transferred into a

spin column (Costar Spin-X centrifuge tube filter, Corning Inc., 8163). The liquid was removed by centrifugation (1 min at 13.2 rpm). The sepharose was rinsed with water (2 × 400 μL) and PBS (50 mM NaPi, 100 mM NaCl, pH 7.4; 2 × 400 μL). The biotin-capped mixture in PBS (1 mL) was added to the streptavidin-coated sepharose and incubated for 10 min with continuous horizontal shaking (150 rpm). The purified **S1** solution (Figure S5) was collected by centrifugation (1 min at 13.2 rpm) into an Eppendorf centrifuge tube.

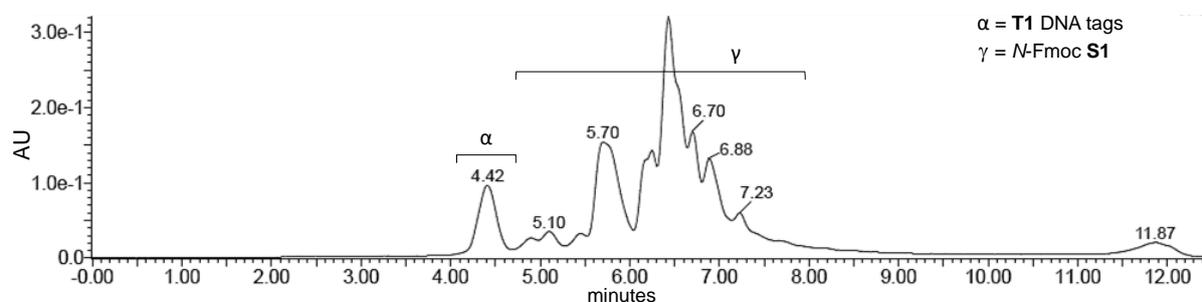


Figure S5 : UPLC chromatogram of purified N-Fmoc **S1** library.

After ethanol precipitation, the purified N-Fmoc **S1** mixture was solubilized in an aqueous solution of 10% piperidine (1 mL, v/v) and incubated at 25°C for 45min. The resulting solution was subjected to EtOH precipitation to afford the purified **S1** library (180 nmoles, 39% yield ; Figure S6).

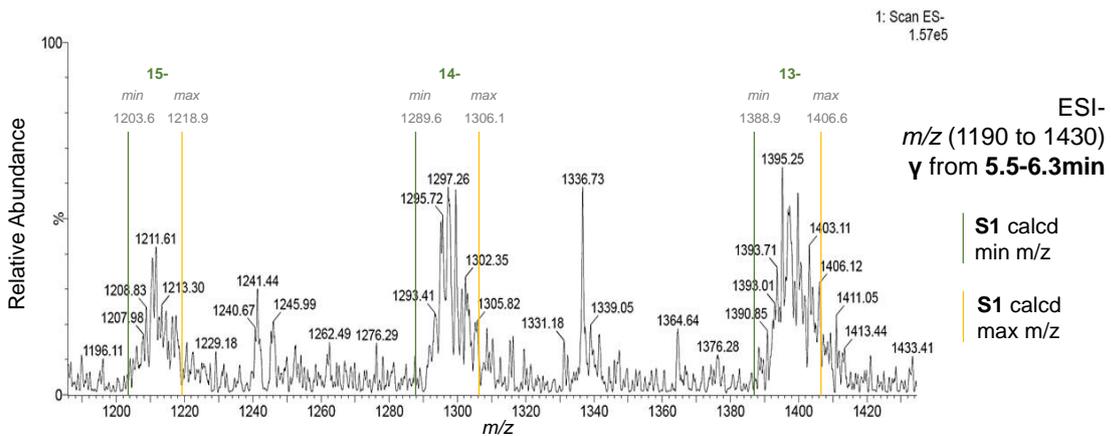
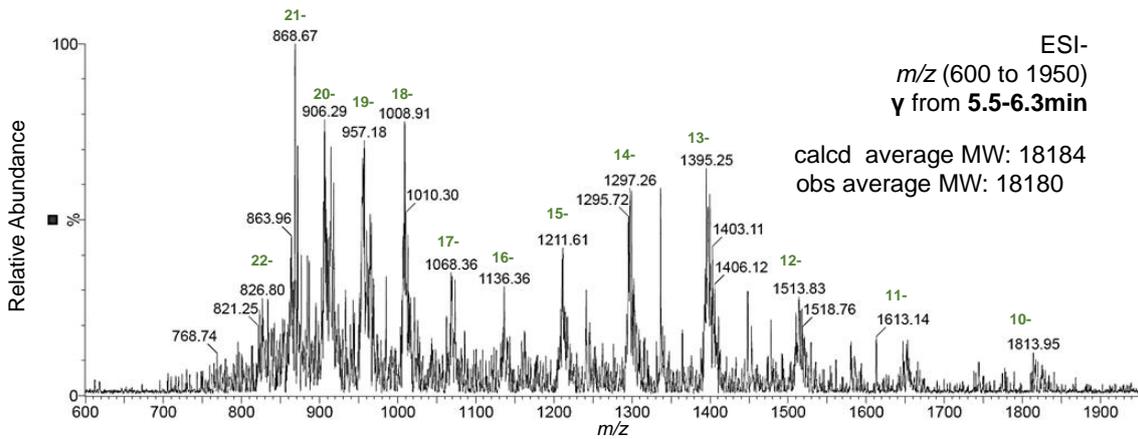
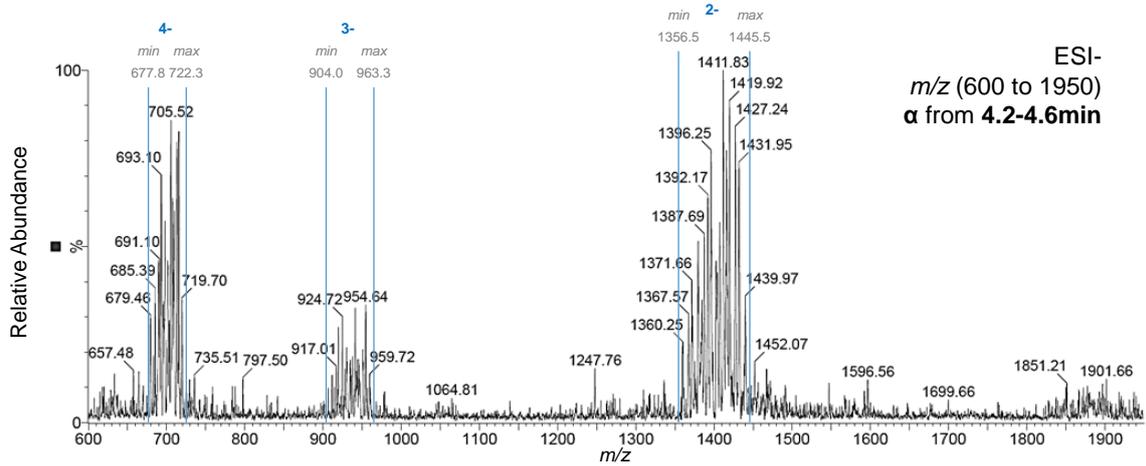
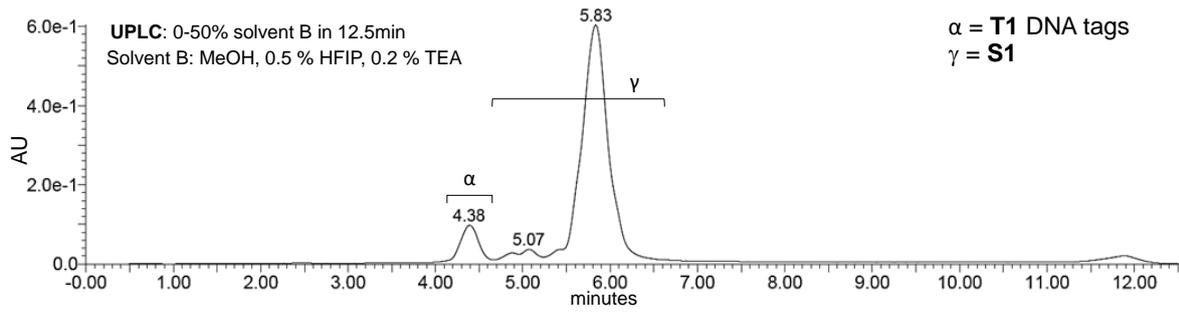


Figure S6 : UPLC chromatogram and ESI mass spectrum of purified S1 library.

Procedure for Cycle 2-DEL1

180 nmoles of **S1** library were partitioned into 9 wells (20 nmoles, 25 μ M in T4 DNA Ligase Reaction Buffer from NEB). In each well distinct Tag 2 duplexes (T2-X, 2 eq, table S10) were ligated using T4 DNA ligase (50 μ M, NEB) for 12h at 16°C. Wells were subjected to ethanol precipitation using the general procedure and ligation successes were assessed using UPLC-MS (average MW of 23959 for ligated products) and agarose gel electrophoresis (according to the general procedure). Next, the selected 9 *N*-Fmoc azido cores (BB2) were reacted onto their corresponding T2 encoded **S1** library members (H₂N-BB1-ext-HP-T1s-T2-X) according to table S5 and using their appropriate on-DNA amide coupling protocol (see table S5 and general coupling procedures). After ethanol precipitations, the coupling efficiencies were confirmed using UPLC-MS and all wells were pooled to provide the crude *N*-Fmoc **S2** library (Figure S7).

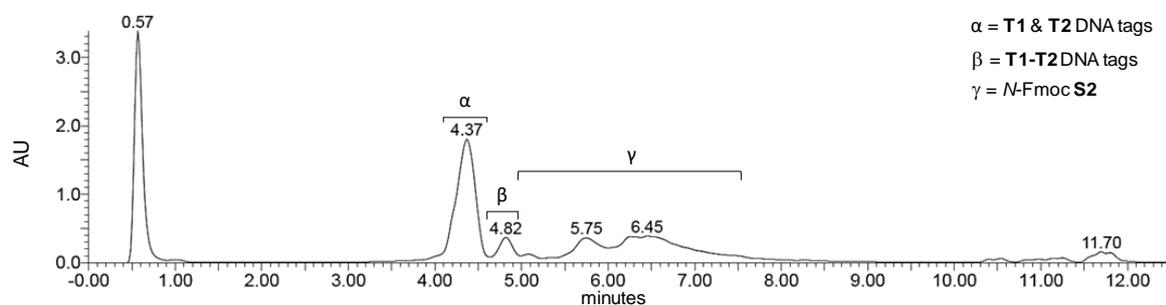
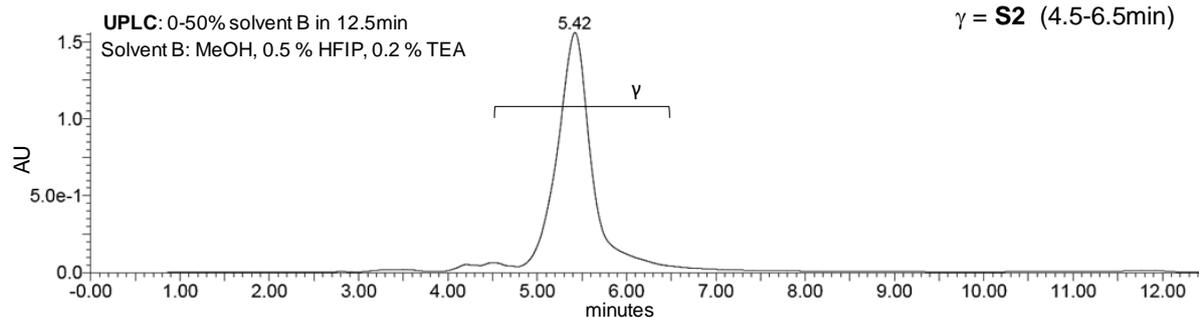


Figure S7 : UPLC chromatogram of crude *N*-Fmoc **S2** library.

The crude *N*-Fmoc **S2** library was purified by reverse-phase HPLC (Gilson – C18 Nucleosil column (Macherey-Nagel, 100 Å, 250 x 10 mm, 7 μ m) with UV-monitoring at 260 nm and 280 nm. A 4 mL/min flow linear gradient was applied from 0 to 80% B solvent for 20min. Solvent A (50 mM triethylammonium acetate buffer with 5% acetonitrile) and solvent B (acetonitrile with 5% water) were used) and the purified *N*-Fmoc **S2** library was solubilized in an aqueous solution of 10% piperidine (1 mL, v/v) and incubated at 25°C for 45min. The resulting solution was subjected to EtOH precipitation then filtered twice using size-exclusion microcentrifuge tubes (Amicon[®]Ultra-0.5mL, 10kDa) to afford the purified **S2** library (65 nmoles, 36% yield ; Figure S8).



$\gamma = \mathbf{S2}$ (4.5-6.5min)

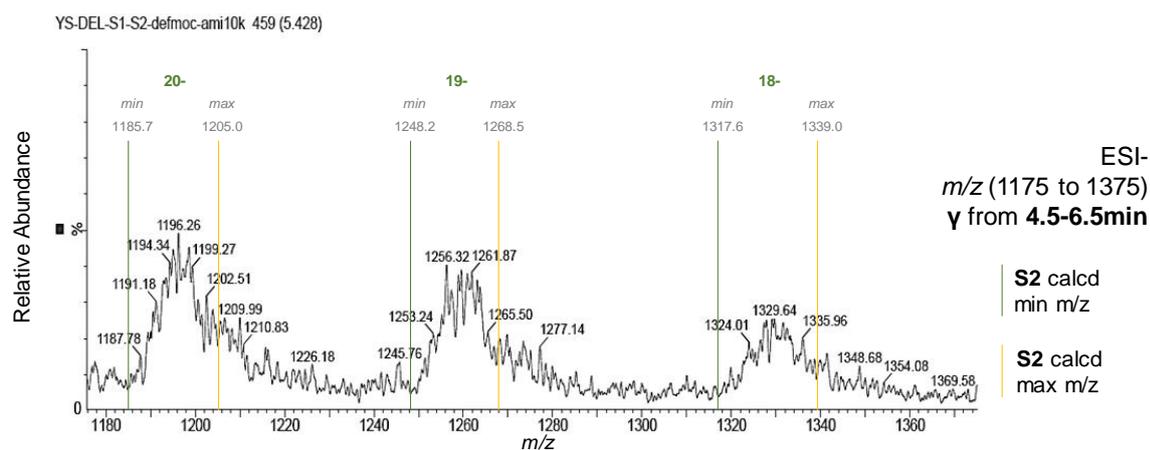
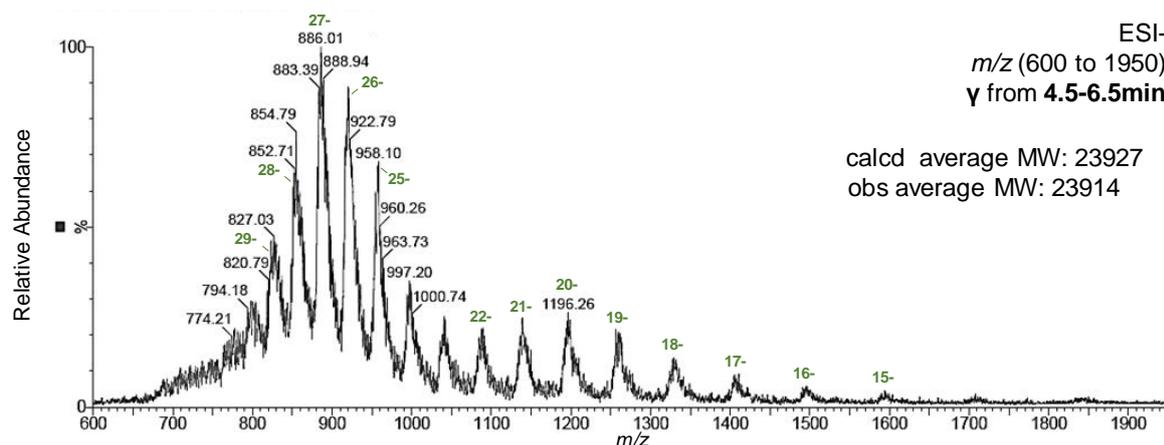


Figure S8 : UPLC chromatogram and ESI mass spectrum of purified **S2** library.

Procedure for Cycle 3-DEL1

65 nmoles of **S2** library were partitioned into 10 wells (6.5 nmoles, 25 μ M in T4 DNA Ligase Reaction Buffer from NEB). In each well distinct Tag 3 duplexes (T3-X, 2 eq, table S12) were ligated using T4 DNA ligase (50 μ M, NEB) for 12h at 16°C. Wells were subjected to ethanol precipitation using the general procedure and ligation successes were assessed using UPLC-MS (average MW of 29710 for ligated products) and agarose gel electrophoresis (according to the general procedure). Next, the selected 10 alkyne containing (hetero)aromatic compounds (BB3) were reacted onto their corresponding T3 encoded **S2** library members (N₃-BB2-BB1-ext-HP-T1s-T2s-T3-X) according to table S7 and using the On-DNA CuAAC reaction protocol. After ethanol precipitations, the coupling efficiencies could not be confirmed using

UPLC-MS due to mass redundancies and were assessed via streptavidin induced electrophoretic mobility shift assay (EMSA) using the following procedure. All wells were pooled to provide the **S3** library which was filtrated twice using size-exclusion microcentrifuge tubes (Amicon® Ultra-0.5mL, 10kDa ; Figure S9).

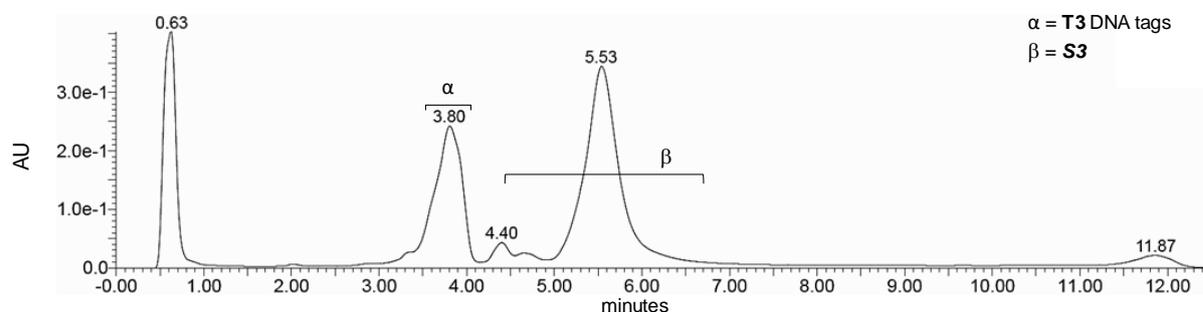
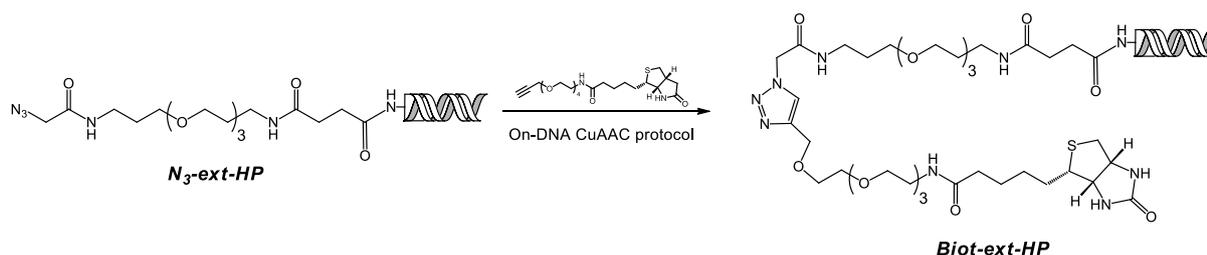


Figure S9 : UPLC chromatogram of **S3** library.

An aliquot of the **S3** library (2 nmoles) was reacted with alkyne-PEG₄-biotin according to the On-DNA CuAAC reaction protocol. In parallel an aliquot of the extended headpiece **S0** functionalized with an azido group (2 nmoles) was also reacted with alkyne-PEG₄-biotin according to the On-DNA CuAAC reaction protocol (Scheme S4).



Scheme S4 : Synthesis of the biotin functionalized extended-headpiece.

Both crude reactions were subjected to ethanol precipitation and a streptavidin gel-shift assay was performed by loading both reactions on agarose gel (according to the general procedure) in the absence or presence (5 μg) of streptavidin (Figure S10).

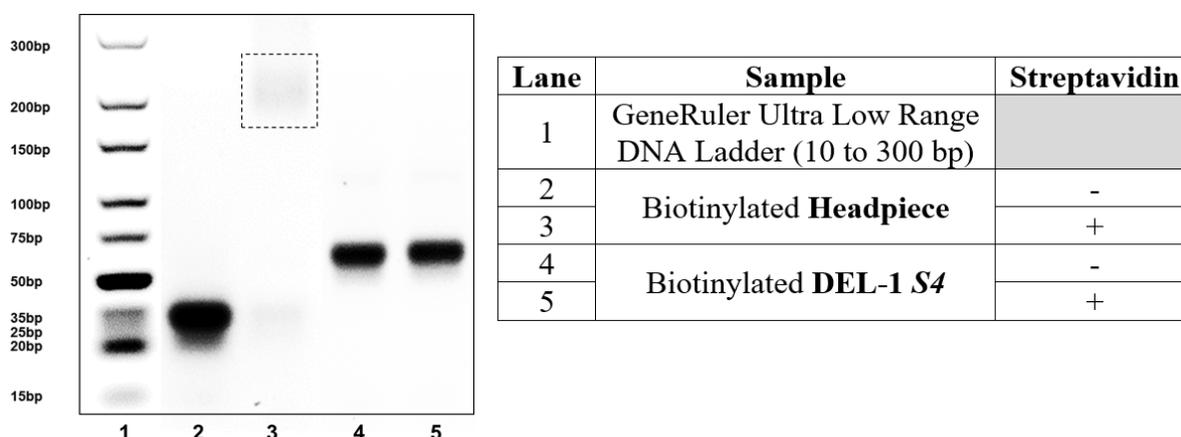


Figure S10 : **S3** coupling efficiencies assessment via streptavidin electrophoretic mobility shift assay.

The **S3** library was further purified by performing four consecutive size-exclusion filtration using microcentrifuge tubes (Amicon[®]Ultra-0.5mL, 30kDa) to afford the purified **S3** library (31.7 nmoles, 48.7% yield ; Figure S11).

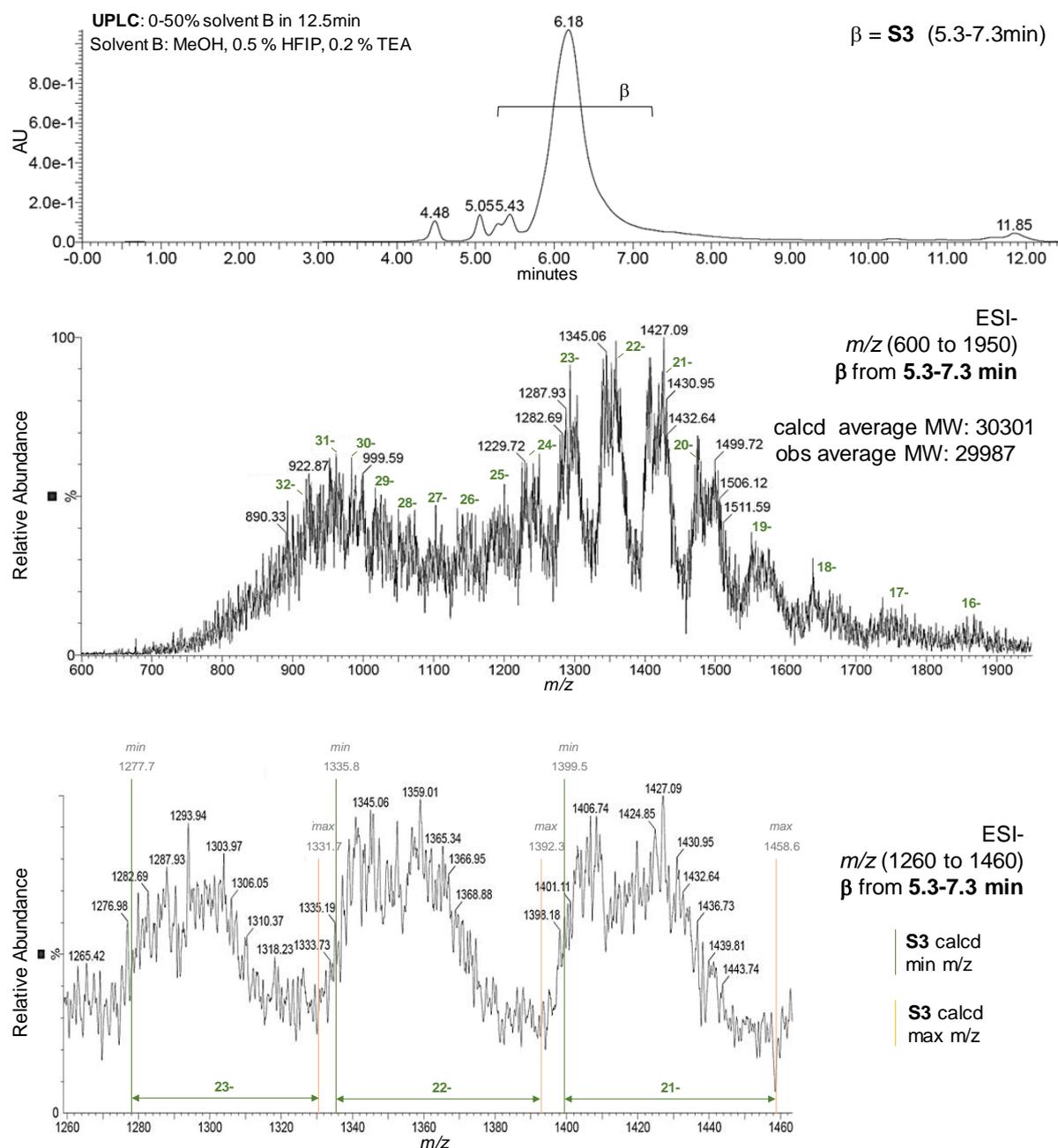


Figure S11 : UPLC chromatogram and ESI mass spectrum of purified **S3** library.

Procedure for Cycle 4-DEL1

31 nmoles of **S3** library were partitioned into 46 wells (0.67 nmole, 25 μ M in T4 DNA Ligase Reaction Buffer from NEB). In each well distinct Tag 4 duplexes (T4-X, 2 eq, table S9) were ligated using T4 DNA ligase (50 μ M, NEB) for 12h at 16°C. Wells were subjected to ethanol precipitation using the general procedure and ligation successes were assessed using agarose

gel electrophoresis (according to the general procedure). Next, the selected 46 *N*-Fmoc pentapeptides (BB4) were reacted in parallel onto the T4 encoded **S3** library members (NH₂-(BB3)-BB2-BB1-ext-HP-T1s-T2s-T3s-T4-X) according to table S4 and using their appropriate on-DNA amide coupling protocol (see table S4 and general coupling procedures). The coupling efficiencies could not be confirmed using UPLC-MS due to quantity limitations and mass redundancies. After ethanol precipitations, all wells were pooled to provide the crude *N*-Fmoc **S4** library (Figure S12).

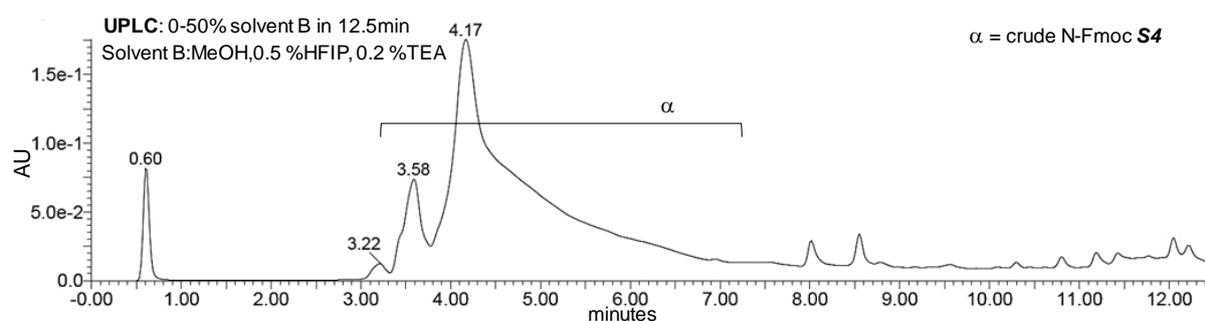


Figure S12 : UPLC chromatogram of crude *N*-Fmoc **S4** library.

The crude *N*-Fmoc **S4** library containing unreacted **T4** encoded **S3** library members was subjected to the biotin/streptavidin cap-and-catch purification procedure described in Cycle 1. To confirm the success of the cap and catch procedure streptavidin induced EMSA was performed as described in Cycle 3. The crude reaction was subjected to ethanol precipitation and a streptavidin gel-shift assay was performed by loading both reactions on agarose gel (according to the general procedure) in the absence or presence (5 μ g) of streptavidin (Figure S13).

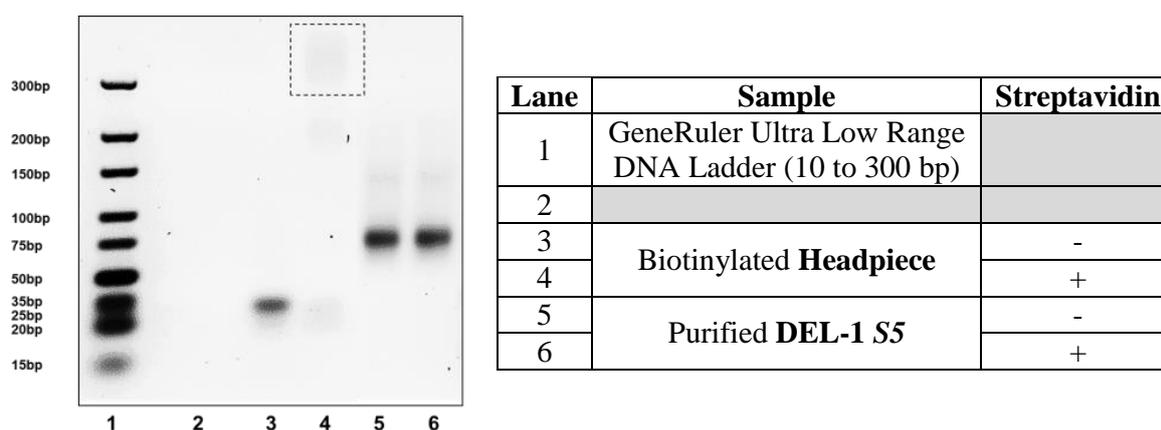


Figure S13 : *N*-Fmoc **S4** purity assessment via streptavidin electrophoretic mobility shift assay.

The purified *N*-Fmoc **S4** was solubilized in an aqueous solution of 10% piperidine (1 mL, v/v) and incubated at 25°C for 45min. The resulting solution was subjected to EtOH precipitation then filtered four times using size-exclusion microcentrifuge tubes (Amicon® Ultra-0.5mL, 30kDa) to afford the purified **S4** library (7.75 nmoles, 24,5% yield).

Cycle 4 was completed by the ligation of a common closing primer (CP1) containing a 7 bp library barcode according to standard procedures. After EtOH precipitation, the fully assembled **G4-DEL1** library was filtered four times using size-exclusion microcentrifuge tubes (Amicon[®]Ultra-0.5mL, 50 kDa) to afford the purified **G4-DEL1** library (5.47 nmoles, overall yield of 1.2 %).

Figure SX recapitulate the agarose gel electrophoresis analysis of the pooled **S0**, **S1**, **S2**, **S3**, **S4** and **G4-DEL1** library members.

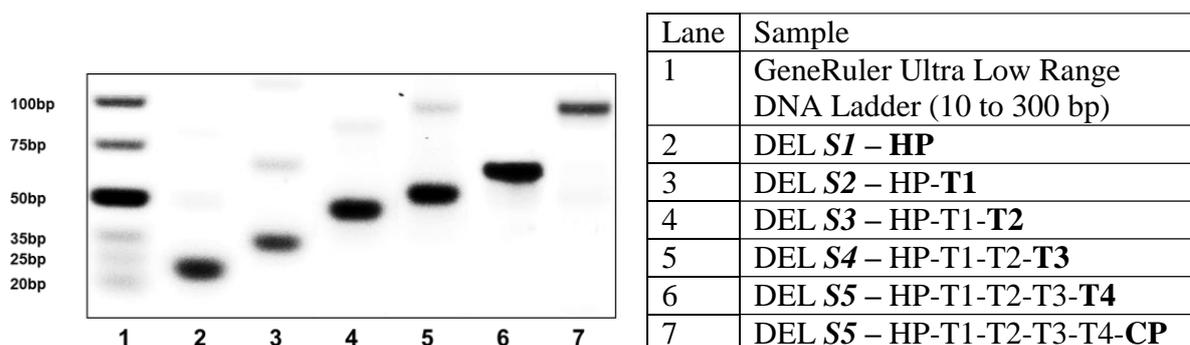
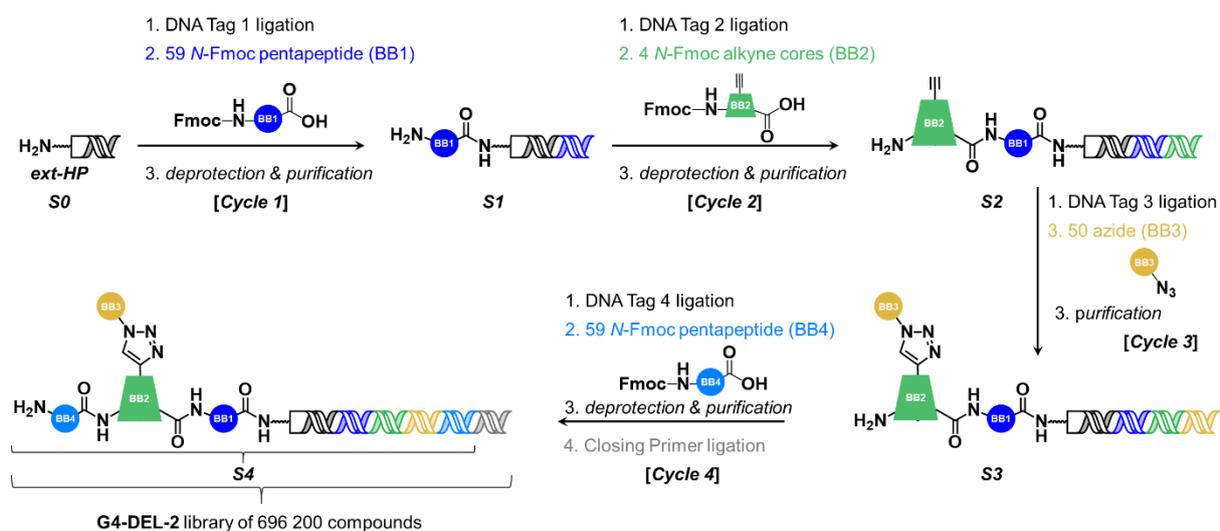


Figure S14 : Agarose gel electrophoresis analysis of cycle 1 to cycle 4 **G4-DEL1** library.

1.4 Preparation of the 696.200 G4-DEL-2 library

1.4.1 General synthetic sequence for G4-DEL-2 and barcoding scheme.

DNA encoded library G4-DEL-2 was assembled in four cycles as depicted in Scheme S5. First, an amino PEG3 spacer was installed onto the previously reported extended headpiece primer (ext-HP, Figure S5). Cycle 1 consisted in the parallel ligation of 59 DNA tags 1 (Table S9) followed by the parallel acylation with 59 N-fmoc pentapeptides (Table S4), the N-fmoc deprotection of the pooled reaction mixtures and purification. Cycle 2 consisted in the parallel ligation of 4 DNA tags 2 (Table S11) followed by the parallel acylation with 4 trifunctional (azide/carboxylic acid/N-fmoc) heteroaromatic cores (Table S6), the N-fmoc deprotection of the pooled reaction mixtures and purification. Cycle 3 consisted in the parallel ligation of 50 DNA tags 3 (Table S13) followed by the parallel Copper-Catalyzed Azide Alkyne “click” of 50 azide containing non- and proteinogenic side-chains (Table S8) and purification. Cycle 4 consisted in the parallel ligation of the 59 DNA tags 4 (Table S9), the parallel acylation with 59 N-fmoc pentapeptides (Table S4) followed by the N-fmoc deprotection of the pooled reaction mixtures and purification. The pooled members of the resulting DEL were ligated with a closing primer containing a library tag.



Scheme S5 : Assembly scheme for the four-step split-and-pool synthesis of the DNA-encoded chemical library **G4-DEL-2**.

The **DEL-2** barcoding scheme is depicted in figure S15.

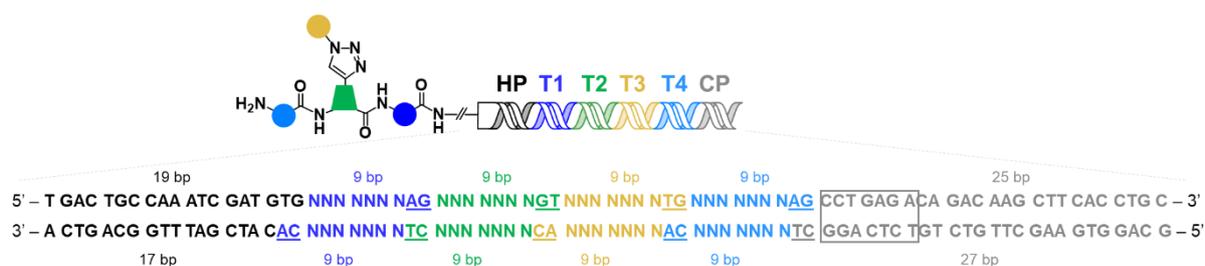


Figure S15 : Barcoding scheme for **DEL-2**. The headpiece (HP) containing the forward primer sequence is in black, DNA Tag 1 in blue, DNA Tag 2 in green, DNA Tag 3 in orange, DNA Tag 4 in light blue, Closing Primer in gray, ligation overhangs are underlined, the library barcode is boxed.

1.4.2 General procedures for building block couplings in G4-DEL-2.

On-DNA amide coupling protocol 1, 2 and 3 were used as described above (p239-240).

On-DNA amide coupling protocol 4: To a solution of Fmoc-trifonctionnal cores in DMA (50 μ L, 250 eq., 100 mM) was added an organic solution of HATU (50 μ L, 125 eq., 50 mM) and 2,6-Lutidine (50 μ L, 625 eq., 10 mM) in DMA. The mixture was allowed to incubate at 30°C for 15 min. An aliquot of the growing DEL library (20 nmol, 100 μ M) was then diluted in sodium borate-buffer (150 mM, pH 9.2) and the solution of activated cores was added to the mixture. The reaction was allowed to proceed for 2 h at 25 °C under orbital agitation. After 2h, a fresh solution of activated cores consisting of Fmoc-trifunctional cores in DMA (100 μ L, 250 eq., 50 mM) and HATU (50 μ L, 125 eq., 50 mM)/2,6-Lutidine (50 μ L, 625 eq., 10 mM) incubated for 15 min at 30°C was added to the reaction mixture. The reaction was allowed to proceed overnight at 30°C under orbital agitation. The on-growing DEL-library was precipitated in EtOH

according to the general procedure and the outcome of the reaction was analyzed by UPLC-MS unless specified otherwise.

On-DNA CuAAC reaction was used as described above (p240).

1.4.3 Building Blocks Validation.

Cycle 1-4 building blocks used to assemble the G4-DEL-2 library were validated on model substrates as described below. The list of all selected building blocks and corresponding validation yields can be found in Table S4.

Cycle 1 & Cycle 4: A total of 78 Fmoc protected pentapeptides (putative cycle 1 & 4 building blocks) sharing the common GNNNG pattern, where N is either A, F, L, P, R, W, were investigated. The coupling efficiency of each Fmoc-protected pentapeptide was assessed using the extended headpiece (Scheme S5) as the starting material, 2 nmoles per reaction. On-DNA amide coupling reactions were investigated with amide coupling protocol 1 (DMT-MM). 59 Fmoc-protected pentapeptides that were found to be coupled with yields above 70% (based UPLC-MS analysis) were deemed suitable for the G4-DEL-1 assembly (Table S4).

Cycle 2: 4 alkyne/carboxylic acid/N-Fmoc trifunctional cores (putative cycle 2 building blocks) were investigated. The coupling efficiency of each of the 4 putative building blocks was assessed using the *ext*-HP as starting material at 2 nmole scale. Amide coupling reactions were performed according to general protocol 3 (and 4 for the trioxazole derivative). All 4 building blocks were found to be incorporated from 65 to quantitative yields (based UPLC-MS analysis) and therefore deemed suitable for G4-DEL-2 assembly (Table S6).

Cycle 3: 50 azide-bearing compounds (putative cycle 3 building blocks) were investigated. The coupling efficiency of each of the 50 putative building blocks was assessed using the extended headpiece grafted with a pentinoic acid (Figure S16, left) as the starting material in 2 nmoles scale. The coupling of some of them were also assessed using the extended headpiece grafted with the trifunctional indolizine-2 and the *ext*-DAP (Figure S16, right). CuAAC reactions were performed with the “On-DNA CuAAC reaction” protocol (p240) and all 50 building blocks were found to be incorporated with yields above 80% (based UPLC-MS analysis) and therefore deemed suitable for G4-DEL-2 assembly (Table S8).

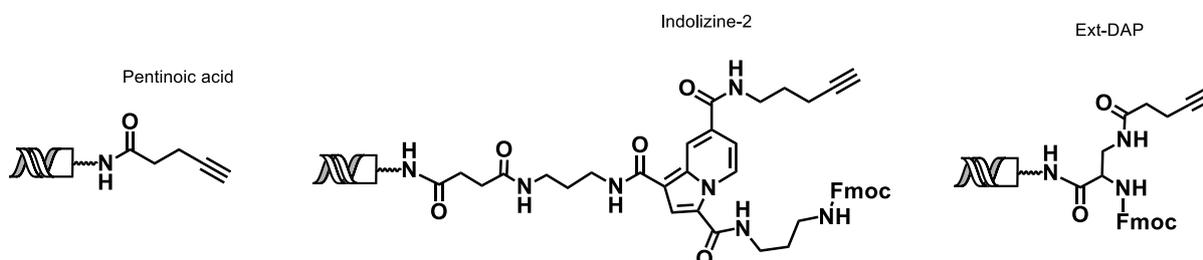


Figure S16 : Representative starting materials for cycle 3 building block validation

1.4.4 Assembly of DEL-2

DEL-2 was assembled mostly according to previously reported procedures.

Procedure for Cycle 1-DEL2

590 nmoles of H₂N-ext-HP **S0** (Scheme S5) were partitioned into 59 wells (10 nmol, 25 μM in T4 DNA Ligase Reaction Buffer from NEB). In each well distinct Tag 1 duplexes (T1-X, 2 eq, table S9) were ligated using T4 DNA ligase (50 μM, NEB) for 12h at 16°C. Wells were subjected to ethanol precipitation using the general procedure and ligation successes were assessed using UPLC-MS (MW of 17655 for all ligated products) and agarose gel electrophoresis (according to the general procedure). Next, the selected 59 N-Fmoc pentapeptides (BB1) were reacted in parallel onto the T1 encoded amino headpieces (H₂N-ext-HP-T1-X) according to table S4 and using their appropriate on-DNA amide coupling protocol (see table S4 and general coupling procedures). After ethanol precipitations, the coupling efficiencies were confirmed using UPLC-MS and all 59 encoded Fmoc-protected pentapeptides were pooled to provide the crude **S1** library.

After an additional ethanol precipitation, the crude **S1** library was filtrated twice using size-exclusion microcentrifuge tubes (Amicon® Ultra-0.5mL, 10kDa) to provide the purified S1 mixture into an Eppendorf centrifuge tube.

After ethanol precipitation, the purified N-Fmoc **S1** mixture was solubilized in an aqueous solution of 10% piperidine (1 mL, v/v) and incubated at 25°C for 45min. The resulting solution was subjected to EtOH precipitation to afford the purified **S1** library (430 nmoles, 73% yield ; Figure S17).

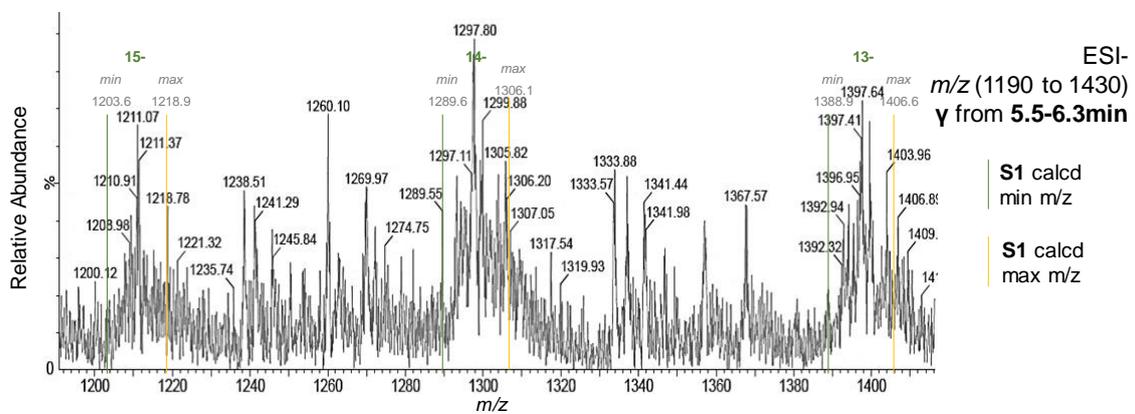
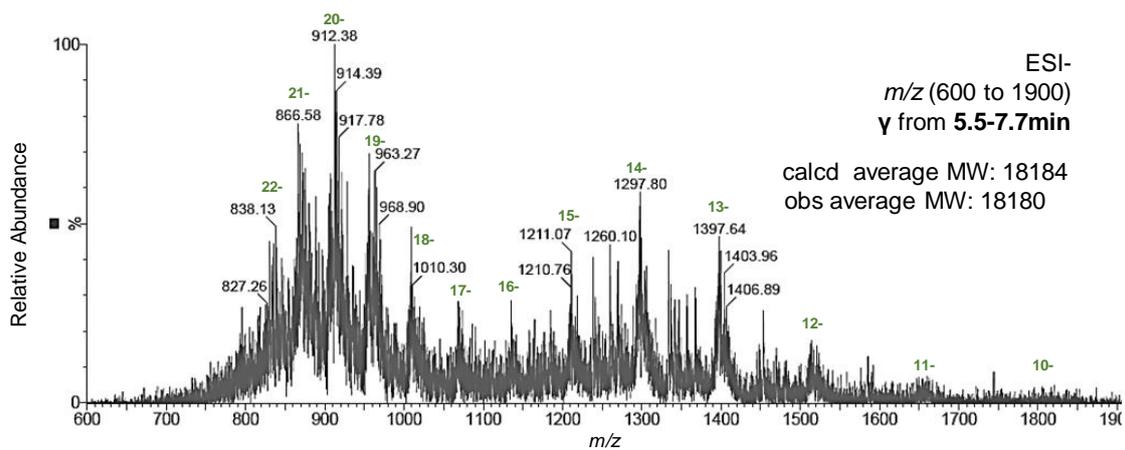
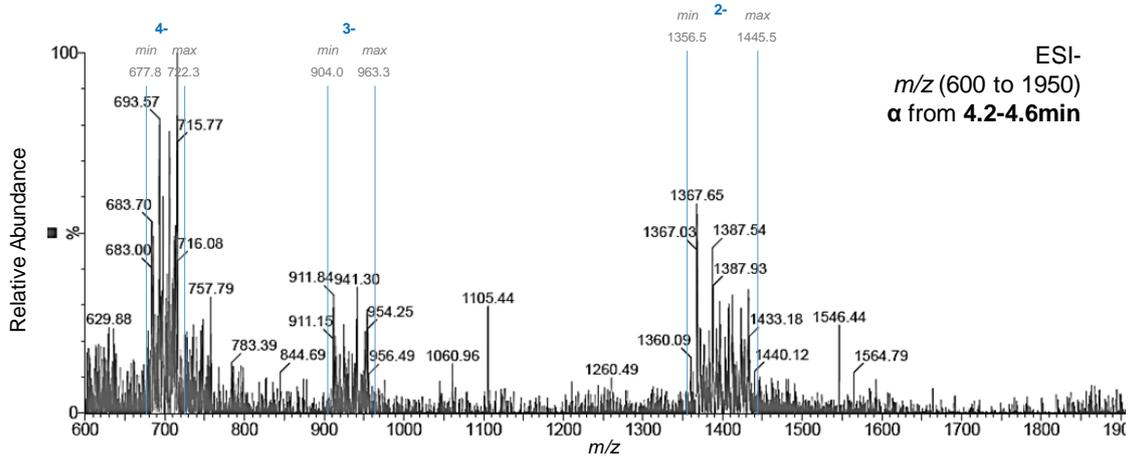
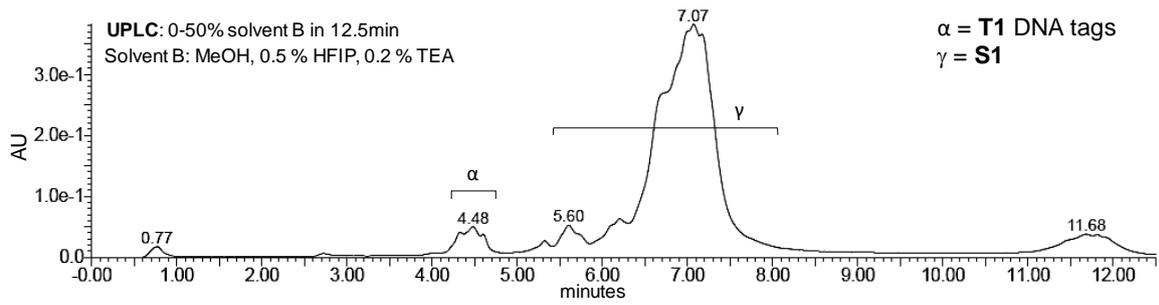


Figure S17 : UPLC chromatogram and ESI mass spectrum of purified S1 library.

Procedure for Cycle 2-DEL2

320 nmoles of **S1** library were partitioned into 4 wells (80 nmoles, 25 μ M in T4 DNA Ligase Reaction Buffer from NEB). In each 80 nmoles wells (subdivided in 4 x 20 nmol), distinct Tag 2 duplexes (T2-X, 2eq., table S11) were ligated using T4 DNA ligase (50 μ M, NEB) for 12h at 16°C. Wells were subjected to EtOH precipitation according to general procedure. Ligation successes were assessed using UPLC-MS (average MW 23737 for ligated products) and agarose gel electrophoresis (according to the general procedure). Next, the selected 4 trifunctional cores (BB2) were reacted onto their corresponding T2 encoded **S1** library members (H₂N-BB1-ext-HP-T1s-T2-X) according to table S6 and using their appropriate on-DNA amide coupling protocol (see table S6 and general coupling procedures). After ethanol precipitations, the coupling efficiencies were confirmed using UPLC-MS.

After filtration using size-exclusion microcentrifuge tubes (Amicon® Ultra-0.5mL, 10kDa), each well (80nmoles – 4x20 nmoles aliquot) was subjected to the cap-and-catch purification procedure described by Neri and colleagues²⁸⁷. The crude wells (80 nmol in 320 μ L of TEA.HCl buffer at pH 10.0) was incubated with biotin-NHS (300 eq, 50 mM) in DMSO for 4h at 37°C. After quenching (Tris.HCl, 0.5M, pH 8.1) for 10 min, an EtOH precipitation was performed and the resulting biotin-capped mixture (500 μ L in PBS 50 mM NaPi, 100 mM NaCl, pH 7.4) was subjected to streptavidin coated sepharose according to the following procedure. Streptavidin-coated sepharose (534 μ L, Streptavidin sepharose high performance, GE Healthcare, 17-5113-01) was transferred into a spin column (Costar Spin-X centrifuge tube filter, Corning Inc., 8163). The liquid was removed by centrifugation (1 min at 13.2 rpm). The sepharose was rinsed with water (8 x 650 μ L) and PBS (50 mM NaPi, 100 mM NaCl, pH 7.4; 8 x 650 μ L). The biotin-capped mixture in PBS (500 μ L) was added to the streptavidin-coated sepharose and incubated for 10 min with continuous horizontal shaking (150 rpm). The purified solutions (Figure S18) were collected by centrifugation (1 min at 13.2 rpm) and filtrated using size-exclusion microcentrifuge tubes (Amicon® Ultra-0.5mL, 10kDa). Streptavidin gel-shift assay showed residual biotinylated compounds on agarose gel electrophoresis (according to the general procedure). The wells were independently engaged in a second cap-and-catch procedure following the same protocol. The purified solutions were collected, and a streptavidin gel-shift assay was performed by loading the mixture on agarose gel (according to the general procedure) in the absence or presence (5 μ g) of streptavidin (Figure 78, p116). The purified *N*-Fmoc S2 solutions were pooled to provide the crude *N*-Fmoc **S2** library (Figure S18).

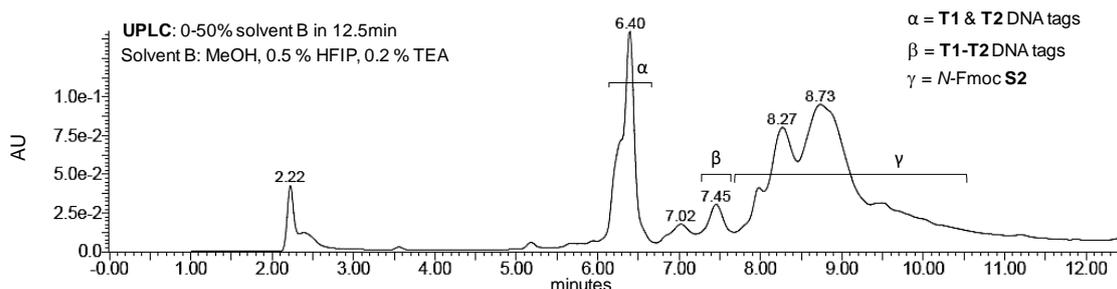


Figure S18 : UPLC chromatogram of purified *N*-Fmoc **S2** library.

After ethanol precipitation, the purified *N*-Fmoc **S2** library was solubilized in an aqueous solution of 10% piperidine (1 mL, v/v) and incubated at 25°C for 45min. The resulting solution was subjected to EtOH precipitation to afford the purified **S2** library (150 nmoles, 47% yield ; Figure S19).

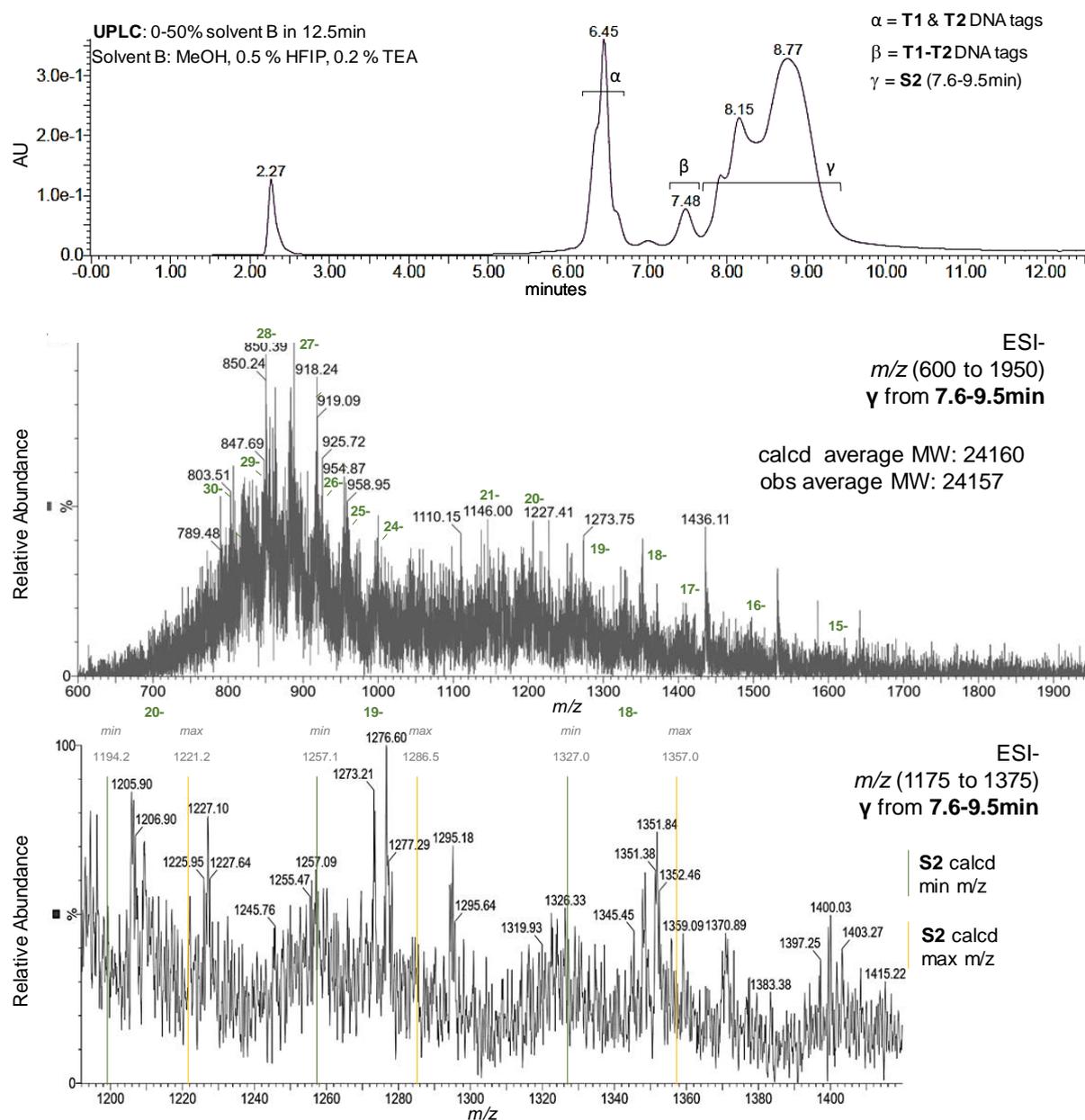


Figure S19 : UPLC chromatogram and ESI mass spectrum of purified **S2** library.

Procedure for cycle 3-DEL2

150 nmoles of **S2** library were partitioned into 50 wells (3 nmoles, 25 μ M in T4 DNA Ligase Reaction Buffer from NEB). In each well, distinct Tag 3 duplexes (T3-X, 2eq., table S13) were ligated using T4 DNA ligase (50 μ M, NEB) for 12h at 16°C. Wells were subjected to ethanol precipitation using the general procedure and ligation successes were assessed using UPLC-MS (average MW of 29722 for ligated products) and agarose gel electrophoresis (according to the general procedure). Next, the selected 50 azide containing proteinogenic or non-

proteinogenic side chains (BB3) were reacted onto their corresponding T3 encoded **S2** library members (Alk-BB2-BB1-ext-HP-T1s-T2s-T3-X) according to table S8 and using the On-DNA CuAAC reaction protocol. After ethanol precipitations, the coupling could not be confirmed using UPLC-MS due to mass redundancies and were assessed via streptavidin induced electrophoretic mobility shift assay (EMSA) using the following procedure. All wells were pooled to provide the crude **S3** library which was filtrated twice using size-exclusion microcentrifuge tubes (Amicon[®] Ultra-0.5mL, 10kDa ; Figure S20).

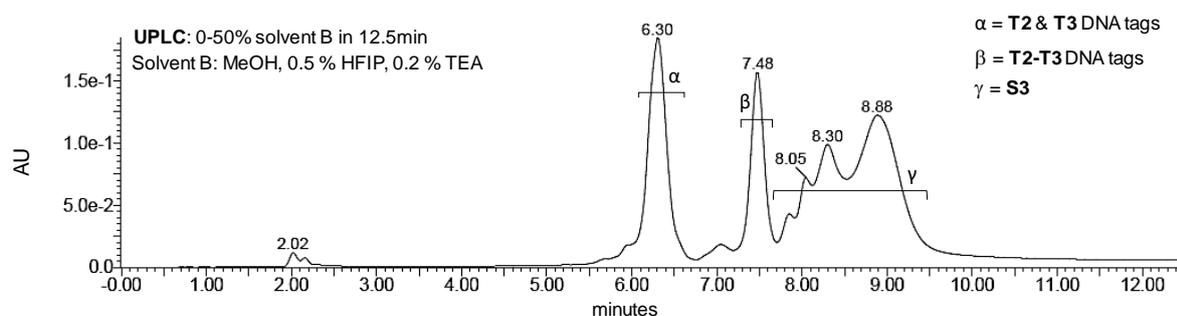
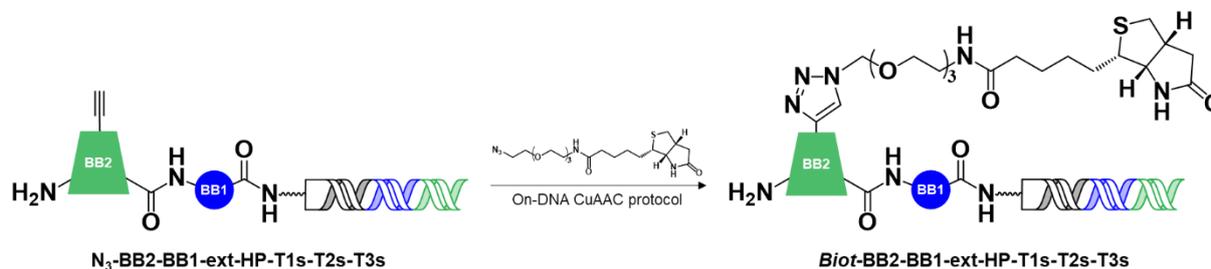


Figure S20 : UPLC chromatogram of crude **S3** library.

The crude **S3** library was reacted with azide-PEG₃-biotin according to the On-DNA CuAAC reaction protocol (Scheme S6). The mixture was subjected to the biotin/streptavidin catch purification procedure as described in Cycle 2. Two rounds of biotin-catch were sequentially performed to fully purify the **S3** library.



Scheme S6 : Functionalization of unreacted N_3 -BB2-BB1-ext-HP-T1s-T2s-T3s with biotin.

After filtration using size-exclusion microcentrifuge tubes (Amicon[®] Ultra-0.5mL, 30kDa), the success of the cap-and-catch procedure was assessed by EMSA (Figure 80, p118) on agarose gel as described in Cycle 2. The **S3** library was subjected to ethanol precipitation to afford the purified **S3** library (49 nmoles, 33% yield ; Figure S21).

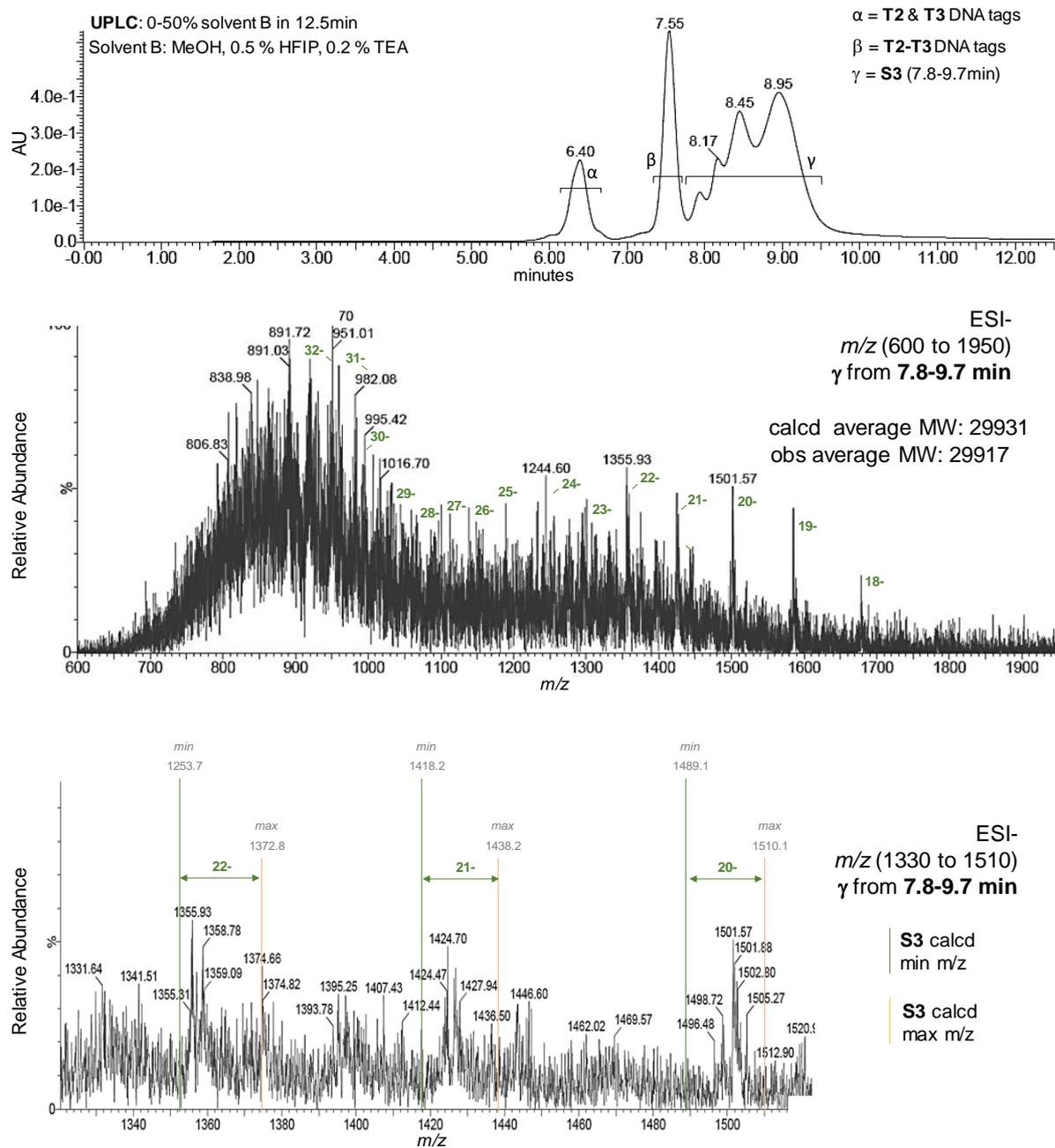


Figure S21 : UPLC chromatogram and ESI mass spectrum of purified S3 library.

Procedure for Cycle 4-DEL2

47.2 nmoles of S3 library were partitioned into 59 wells (0.8 nmole, 25 μ M in T4 DNA Ligase Reaction Buffer from NEB). In each well distinct Tag 4 duplexes (T4-X), 2 eq., table S9) were ligated using T4 DNA ligase (50 μ M, NEB) for 12h at 16°C. Wells were subjected to ethanol precipitation using the general procedure and ligation successes were assessed using agarose gel electrophoresis (according to the general procedure). Next, the selected 59 N-Fmoc pentapeptides (BB4) were reacted in parallel onto the T4 encoded S3 library members (NH2-(BB3)-BB2-BB1-ext-HP-T1s-T2s-T3s-T4-X) according to table S4 and using their appropriate on-DNA amide coupling protocol (see table S4 and general coupling procedures). The coupling

efficiencies could not be confirmed using UPLC-MS due to quantity limitations and mass redundancies. After ethanol precipitations, all wells were pooled to provide the crude *N*-Fmoc **S4** library (Figure S22).

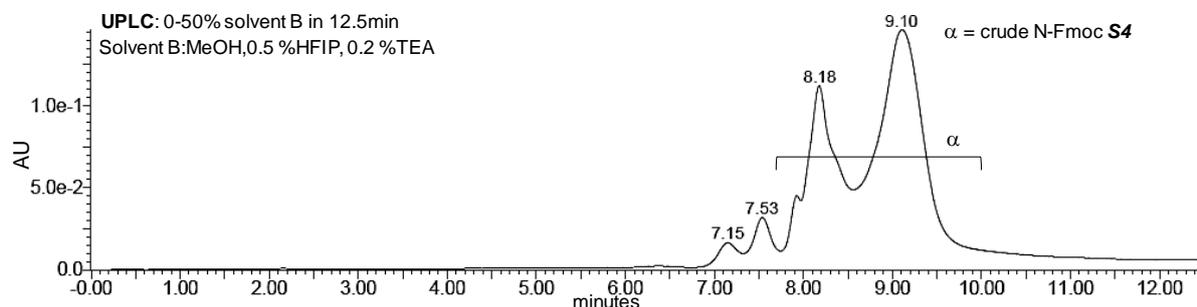


Figure S22 : UPLC chromatogram of purified *N*-Fmoc **S4** library.

The crude *N*-Fmoc **S4** library containing unreacted **T4** encoded **S3** library members was subjected to the biotin/streptavidin cap-and-catch purification procedure described in Cycle 2. The purified *N*-Fmoc **S4** was solubilized in an aqueous solution of 10% piperidine (1 mL, v/v) and incubated at 25°C for 45min. The resulting solution was subjected to EtOH precipitation then filtered four times using size-exclusion microcentrifuge tubes (Amicon®Ultra-0.5mL, 30kDa) to afford the purified **S4** library (5.04 nmoles, 10.7% yield).

Cycle 4 was completed by the ligation of a common closing primer (CP2) containing a 7 bp library barcode according to standard procedures. After EtOH precipitation, the fully assembled G4-DEL1 library was filtered six times using size-exclusion microcentrifuge tubes (Amicon®Ultra-0.5mL, 50 kDa) to afford the purified G4-DEL1 library (3 nmoles, overall yield of 0.51 %).

1.5 Affinity selection of G4-DEL-1 against streptavidin bound G4 structures and controls

1.5.1. Affinity selection targets

All affinity selections were performed with biotin oligonucleotides immobilized onto magnetic Dynabeads® MyOne™ Streptavidin T1. The full list of oligonucleotides used as targets can be found on Table 9, p129, and Table 14, p145.

Circular dichroism studies were performed with the biotinylated oligonucleotide to confirm the expected folding. Annealing/folding were performed by heating each oligonucleotide (1.67 μM) at 80°C for 5 min in selection buffer (100mM KCl, 0.01% Tween-20, 1X PBS pH 7.4) and cooling to room temperature over 2 h. CD spectra were recorded at 25°C with wavelength range of 220 to 320 nm on a Jasco J-810 spectropolarimeter using 1 cm length quartz cuvette. Each spectrum was an average of three scans with a 0.5 s response time, a 1 nm data pitch, a 4 nm bandwidth and a 200 nm.min⁻¹ scanning speed.

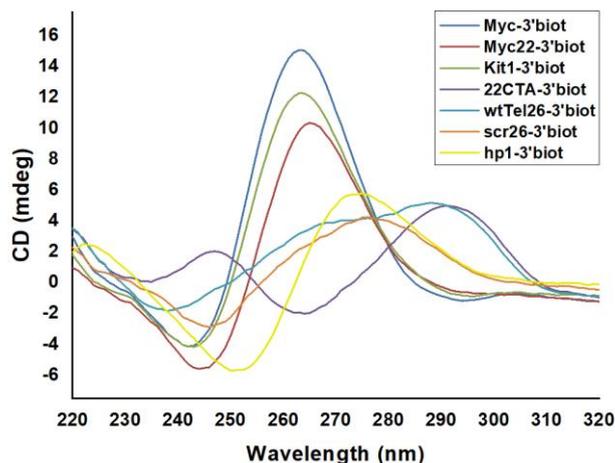


Figure S23 : CD analyses (1.67 μ M in 100 mM KCl, 0.01% Tween-20, 1X PBS pH 7.4) of biotin-oligonucleotide targets.

1.5.2. General procedure for affinity selections

All affinity selections were performed according to the following general procedure.

1/ Target loading

Dynabeads® MyOne™ Streptavidin T1 (80 μ L, 10 mg/mL, ss oligonucleotides binding ~400 pmol/mg) were transferred to a microcentrifugal tube (0.5 mL) and placed on a magnet for 2 min. After removal of the supernatant by pipetting, the beads were washed three times with 160 μ L of 1X **Binding and Washing buffer** (1X **B/W buffer** = 5 mM Tris-HCl, 0.5 mM EDTA, 1 M NaCl, pH 7.5, filtered with 0.45 μ m filter). The magnetic beads were resuspended in 160 μ L (5 mg/mL, 320 pmol) of 2X B/W buffer and 160 μ L of biotinylated oligonucleotide target (5 eq., 10 pmol/ μ L in H₂O) was added. The heterogeneous mixture was incubated for 15 min at room temperature using gentle rotation of 500 rpm. Tubes were then placed on a magnet for 2 min, the supernatant was removed by pipetting and the oligonucleotide coated magnetic beads were washed twice with 160 μ L 1X B/W buffer.

2/ Beads capping

The oligonucleotide coated magnetic beads were resuspended in 160 μ L (5 mg/mL, 320 pmol) of 2X B/W buffer and 160 μ L of biotin (100 eq., 200 pmol/ μ L in H₂O) was added. The heterogeneous mixture was incubated for 15 min at room temperature using gentle rotation of 500 rpm. Tubes were then placed on a magnet for 2 min, the supernatant was removed by pipetting and the oligonucleotide coated magnetic beads were washed twice with 160 μ L 1X B/W buffer.

3/ Beads splitting and target folding

The oligonucleotide coated magnetic beads were resuspended in 80 μL (10 mg/mL, 320 pmol) of 1X B/W buffer and washed 3 times with 160 μL of **incubation buffer (i-buffer: 100 mM KCl, 0.01% Tween-20 (unless otherwise specified), 1X PBS pH 7.4)**. The oligonucleotide coated magnetic beads were then suspended in 80 μL of i.buffer and splitted in 12 microcentrifuge tubes (6.25 μL , 25 pmol,). Each batch of beads (6.25 μL , 25 pmol) were diluted up to 62.5 μL with i.buffer (final concentration of 1 mg/mL), and subjected to annealing/folding by heating at 80°C for 5 min and slowly cooling (2h) to room temperature. 6.5 μL of genomic DNA (10X Herring Sperm DNA, 100 eq., 10 $\mu\text{g}/\mu\text{L}$) were added in each aliquot (unless otherwise specified) and the heterogeneous mixture was incubated at room temperature for 15 min at 500 rpm. Each aliquot can be used for one selection experiment.

4/ Incubation, washes and elution

A selection was performed by adding 5 μL of **G4-DEL-1** (1 μM , 5 pmoles) to one aliquot of oligonucleotide coated magnetic beads (67.57 nM final **G4-DEL-1** concentration) and incubating for 1h at room temperature at 500 rpm. After 1h selection tubes were placed on a magnet for 2 min, the supernatant (flow-through of selection) was removed by pipetting and the ligands/oligonucleotide coated magnetic beads were subjected to various wash procedures listed below depending on the selection stringencies.

WASH1: 5x i.buffer (100 μL) at 25°C

WASH2: 3x i.buffer (100 μL) at 25°C then 1x i.buffer (100 μL) at 50°C then 1x i.buffer (100 μL) at 25°C

WASH3: 3x i.buffer (100 μL) at 25°C then 2x i.buffer (100 μL) at 50°C then 1x i.buffer (100 μL) at 25°C

WASH4: 3x i.buffer (100 μL) at 25°C then 1x i.buffer (100 μL) at 70°C then 1x i.buffer (100 μL) at 25°C

WASH5: 2x i.buffer (100 μL) at 25°C then 1x i.buffer (100 μL) at 70°C then 1x i.buffer (100 μL) at 80°C then 1x i.buffer (100 μL) at 25°C.

The washes incubation time was 10 min and pre-heated i-buffer were used for hot washes. Elutions of the selected ligands were performed by incubating 100 μL milliQ water (unless otherwise specified) pre-heated at 95°C for 5 min. The tubes were then placed on a magnet for 2 min and the supernatant (selection elution) were collected using a pipette.

1.5.3. Post-selection PCR amplification of the DNA barcodes for illumina HTS sequencing

The elution solutions of each selection experiments were EtOH precipitated according to the general procedure and solubilized in 10 μL of milliQ H₂O. 2 μL of each elution solutions were used to quantify the amount of DNA recovered after selections using quantitative PCR (qPCR) according to reported procedures³⁴⁵. qPCR was carried out by using a AriaMx Real-Time PCR System (Agilent). In a 96-well reaction plate (Agilent Technologies, Skirted, Polypropylene) were combined to a total volume of 20 μL : selection eluate (2 μL), forward primer (2 μL , 5 μM), reverse primer (2 μL , 5 μM), PowerUp SYBR Green Master Mix (10 μL , Applied Biosystems) and 4 μL of water. All qPCR experiments were conducted at 95°C for 10 min followed by 20 to 30 cycles (depending on the initial quantity of DNA) of 95°C for 30 sec (denature), 68°C for 3

min (anneal/extend). The specificity of the reactions was verified by melting curve analysis and gel electrophoresis.

Each selection elutes were then amplified by PCR for high throughput DNA sequencing using primers that are compatible with Illumina sequencing instruments:

Forward primer with illumina multiplexing adapter. CAA GCA GAA GAC GGC ATA CGA GAT **NNN NNN** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T TGA CTG CCA AAT CGA TGT G. The underlined sequence is complementary to the eluate DNA sequences, the bold sequence marks an index unique for each eluate amplification.

Reverse primer = TruSeq Universal Adapter. A ATG ATA CGG CGA CCA CCG AGA TCT CAC TTT CCC TAC ACG ACG CTC TTC CGA TCT G CAG GTG AAG CTT GTC TG. The underlined sequence is complementary to the eluate DNA sequences.

After PCR amplification, the DNA amplicons of each selection were purified using gel electrophoresis. Amplicons were loaded on UltraPure Agarose 5% gels in TBE buffer and the gel were run at 90 V, 300 mA for 2 h. The DNA bands of interest were identified using a Gel Doc XR+ System (Bio-Rad), cut-off the gel and incubated in extraction buffer (1 M Tris-HCl, 0.5 M Na₂EDTA, 25 mM NaCl) at 30°C for 16 h. The extracted DNA was concentrated, filtered twice using size-exclusion microcentrifuge tubes (Amicon® Ultra-0.5mL, 10kDa) and quantified using agarose gel electrophoresis (GeneRuler™ Ultra Low Range DNA Ladder, Thermo Scientific). The DNA amplicons of all selections were pooled in equimolar ratios and submitted to HTS. Deeper sequencing was also performed for PCR-amplified samples of the strating naïve (unselected) library.

1.5.4. List and details of reported selection experiments

The details of reported affinity selection experiments (target, buffer, wash procedure, elution procedure and selection efficiency) are listed on Table S1.

Table S1 : Characteristics of affinity selection experiments

Target	Selection buffer	Additive	Washes	Elution	Output quantity in copies and % recovery ¹
Myc / parallel					
Myc-biot (5eq)	i.buffer	Herring Sperm DNA	WASH1	95°C mQw	o-Sel01 5.91E+09 1.96‰
Myc-biot (5eq)	i.buffer	Herring Sperm DNA	WASH1	95°C mQw	o-Sel02 7.62E+09 2.53‰
Myc-biot (5eq)	i.buffer	Herring Sperm DNA	WASH2	95°C mQw	o-Sel04 7.92E+08 0.26‰
Myc-biot (5eq)	i.buffer	Herring Sperm DNA	WASH4	95°C mQw	o-Sel07 3.53E+08 0.12‰

Myc-biot (5eq)	i.buffer	Herring Sperm DNA	WASH4	95°C mQw	n-Sel05&14 1.41E+08 0.047‰
Myc-biot (5eq)	i.buffer	Herring Sperm DNA	WASH5	95°C mQw	n-Sel06 1.42E+8 0.047‰
Myc-biot (5eq)	i.buffer	Herring Sperm DNA	WASH4/WASH4	95°C mQw	n-Sel07 2.86E+08 0.095‰
Controls					
Biotin coated Streptavidin	i.buffer	Herring Sperm DNA	WASH1	95°C mQw	o-Sel15 1.06E+08 0.035‰
Biotin coated Streptavidin	i.buffer	Herring Sperm DNA	WASH1	95°C mQw	o-Sel16 2.44E+08 0.081‰
hp1 (5eq)	i.buffer	/	WASH1	95°C mQw	n-Sel16 6.55E+07 0.022‰
hp1 (5eq)	i.buffer	/	WASH1	95°C mQw	n-Sel17 1.36E+08 0.045‰
Other G4 targets					
Kit1-biot (5eq)	i.buffer	Herring Sperm DNA	WASH1	95°C mQw	o-Sel21 8.01E+09 2.66‰
Myc22-biot (5eq)	i.buffer	Herring Sperm DNA	WASH4	95°C mQw	n-Sel34 1.03E+09 0.34‰
Myc22-biot (5eq)	i.buffer	Herring Sperm DNA	WASH4	95°C mQw	n-Sel35 2.81E+09 0.94‰
22CTA-biot (5eq)	i.buffer	Herring Sperm DNA	WASH1	95°C mQw	o-Sel20 2.62E+08 0.087‰
wtTel26-biot (5eq)	i.buffer	Herring Sperm DNA	WASH1	95°C mQw	o-Sel19 1.27E+09 0.42‰
Additional selection experiments					
Myc-biot (5eq)	i.buffer 0.1% T-20	Herring Sperm DNA	WASH1	95°C mQw	o-Sel03 1.29E+09 0.43‰
Myc-biot (5eq)	i.buffer	Herring Sperm DNA	WASH4	65°C DMSO	o-Sel08 6.45E+07 0.022‰
Myc-biot (0.5eq)	i.buffer	Herring Sperm DNA	WASH1	95°C mQw	Ave. o-Sel11/12 2.04E+07 0.0068‰
Kit1-biot (5eq)	i.buffer	Herring Sperm DNA	WASH3	95°C mQw	o-Sel24 1.19E+07 0.004‰
22CTA-biot (5eq)	i.buffer	Herring Sperm DNA	WASH3	95°C mQw	o-Sel23 1.78E+07 0.006‰
wtTel26-biot (5eq)	i.buffer	Herring Sperm DNA	WASH3	95°C mQw	o-Sel22 1.08E+08 0.035‰

1.6 High Throughput Sequencing

High Throughput Sequencing was performed on an Illumina NextSeq 500 instrument using High Output Kit v2 (75 cycles). The pool of selection and naïve library amplicons were mixed with 10 % PhiX. Deep sequencing data were demultiplexed and recovered in FASTQ format using bcl2fastq Conversion Software (Illumina). The resulting raw data files (FASTQ files) were processed using a software suite called PATTERNITY-SEQ, developed from the aptamer platform in MIRGen (<https://jacob.cea.fr/drf/ifrancoisjacob/english/Pages/Departments/MIRGen/Platforms.aspx?Type=Chapitre&numero=6>). The different steps of the sequencing analysis are recapitulated below. First, adapter and primer sequences were removed from each sequence, keeping only variable regions. Then, sequences that contained at least one base with a quality score (Q) below 30 were removed before being saved in a FASTA format. This quality score can be converted to a probability of error (P) using the formula $P = 10(-Q/10)$. Thus, the recovered sequences contain bases with a potential probability of error below 0.001 (1 in 1000). Only sequences with a frequency >0.0001% in at least one selection were recovered to decrease the time of analysis and to remove sequences of poor interest. Sequence data were then converted back into encoded chemical information computationally. The frequency of each compound for every selection were output into an excel file and the frequency of each individual building block was calculated for every selection. For each compound n, a normalized ratio (RN) relative to unselected naïve library was calculated for each selection x as described by the formula:

$$RN_{n,x} = \frac{r_{n,x/0}}{\sum_{n \in B} r_{n,x/0}}$$

where B = all library members and

$$r_{n,x/0} = \frac{f_{n,R=x}}{f_{n,R=0}}$$

where $f_{n,R=x}$ is the frequency of the compound n in the library after the selection x and $f_{n,R=0}$ is the frequency of the compound n in the starting library. Accordingly, the sum of RNs for a selection x is equal to 1 and the enrichment (E) in % of each compound n after selection x can be calculated by the formula:

$$E_{n,x} = \frac{RN_{n,x} - 1/N}{1/N} \times 100$$

where N is the number of compounds for which a RN could be calculated for the selection x. Accordingly, the sum of Es for a selection x is equal to 0%. Finally, an Enrichment Fold (EF) can be calculated by the formula:

$$EF_{n,x} = \frac{RN_{n,x} \times (E_{n,x} \times RN_{n,x})}{RN_{n,x}}$$

The same calculation was also performed to calculate RN, E and EF for each building block.

In total, about 60 million reads were determined from the no-target (naïve) selection and about 5 million reads were determined from each of the reported selections (Table S1). To estimate a coefficient of variation of the measured frequencies of compounds inside 5 million reads, the 60 millions of sequencing reads from the naive library were randomly distributed into 12 files of 5 million reads, the mean frequency and its standard deviation were calculated for each compounds in order to calculate for each frequency a coefficient of variation, which corresponds to the ratio of standard deviation to the mean (Figure S24). Of 190,440 sequences, only those sequences for which the CV was below 60% (180 014 sequences) were considered for analysis (Figure S25). Only sequences for which CV was below 50% and frequency within the naïve library was above $5 \cdot 10^{-4}$ (175 242 sequences) were considered for the selection of enriched library members and controls to be resynthesized (Figure S25).

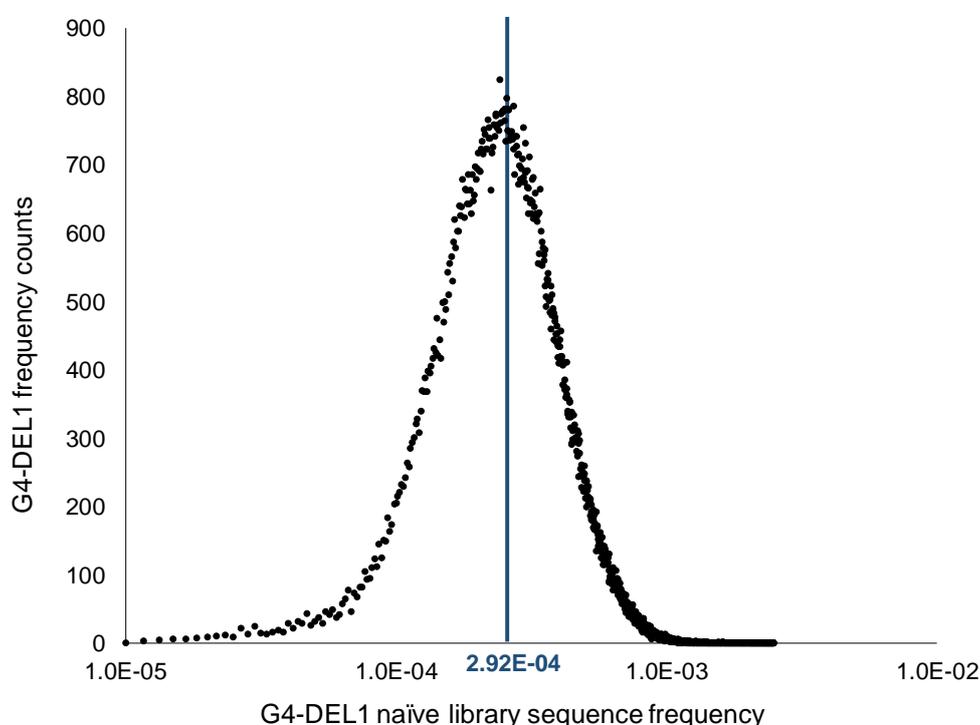


Figure S24 : Analysis of high-throughput sequencing results. Actual (dots) vs expected (vertical line) frequency for all library members obtained from naïve library sequencing. This plot shows the distribution of sequence frequencies within the 190 440 sequences of the naïve G4-DEL1 library (x axis) versus the abundance of each frequency in sequence counts (y axis). The blue vertical indicates the theoretical frequency for each sequence assuming equal abundance of each of the 190 400 sequences analyzed by 60 million sequence reads.

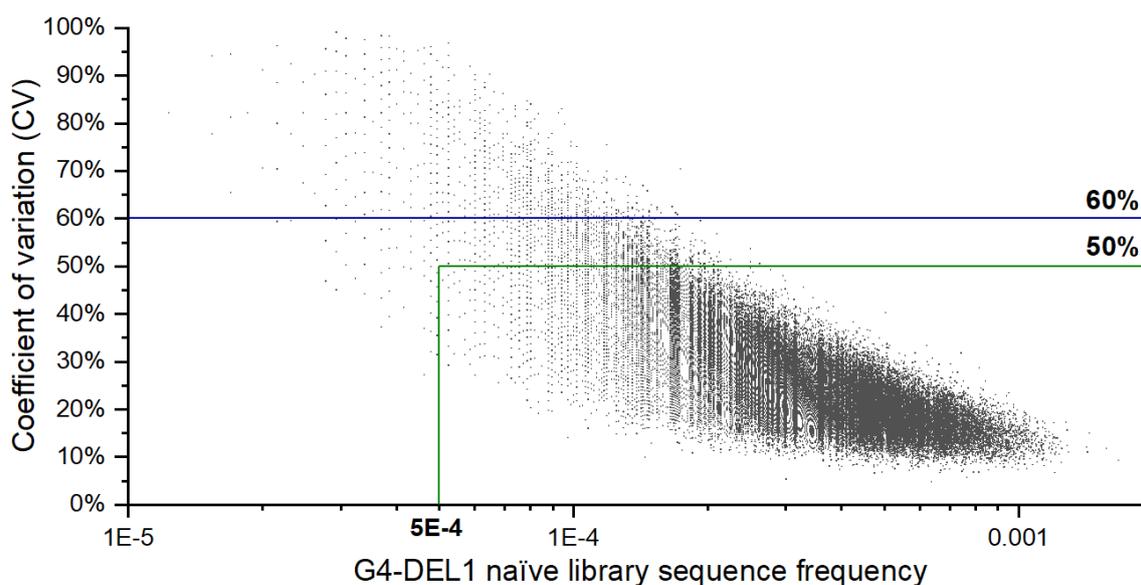


Figure S25 : Graphical representation of the correlation between the frequency of the 190 440 sequences in the naïve G4-DEL1 library (x axis) and their coefficient of variation (y axis). To determine the coefficient of variation for each sequence, 60 million of the reads were randomly distributed into clusters of 5 million reads and the ratio of standard deviation to the mean was calculated for each sequence from those clusters. Only sequences for which CV was below 60% (180 014 sequences) were considered for statistical analysis of the selections (below the blue horizontal). Only sequences for which CV was below 50% and frequency within the naïve library was above $5 \cdot 10^{-4}$ (175 242 sequences) were considered for the selection of enriched library members and controls to be resynthesized (within the green box).

Individual building blocks enrichment were calculated for each DEL cycle and each selection experiments as depicted in Figure S26, Figure S27, Figure S28. Abundance and correlation were schematically represented against Myc G4 and 22CTA G4 in Figure S29.

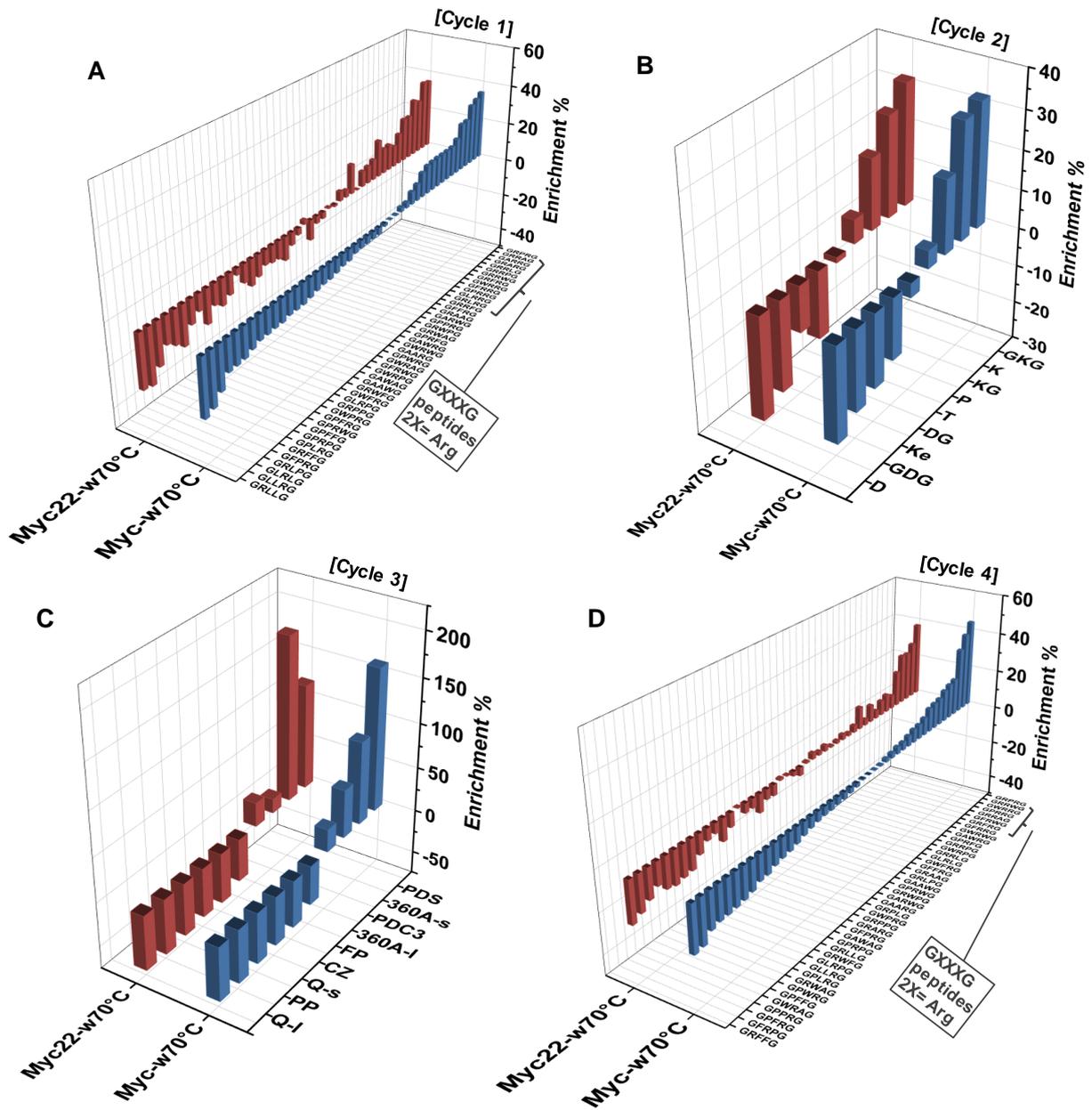
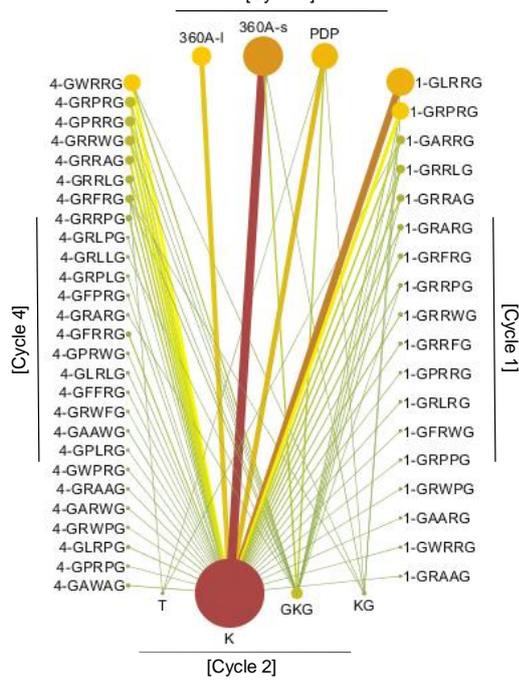


Figure S28 : Schematic representation of the Enrichment (E) in % of each **DEL1** building blocks after selections against **Myc** and **Myc22** G4. Harsh (70°C) wash procedures were used in both selections.

A 2rnds-Myc-w70°C



B 22CTA-w50°C

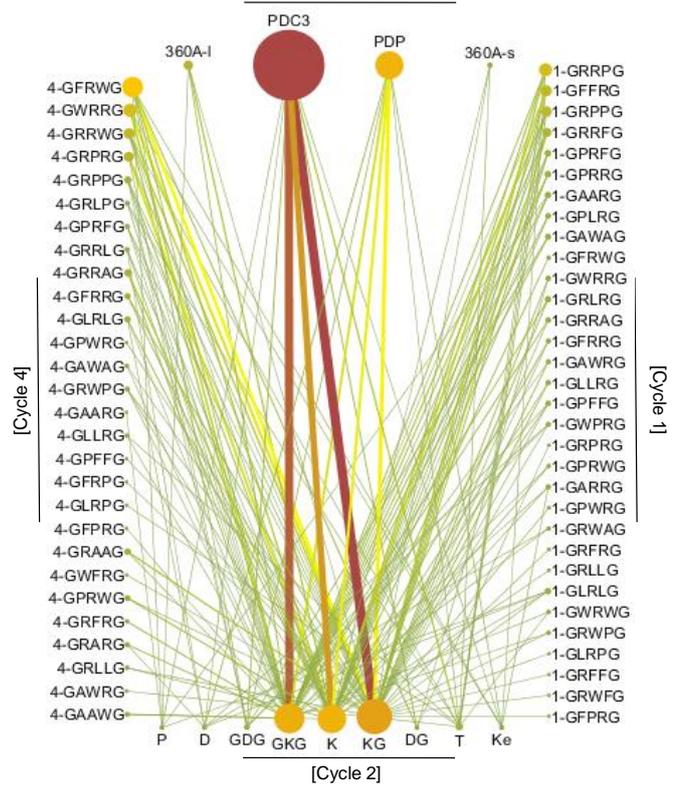


Figure S29 : Schematic representation of DEL-1 building blocks abundance and correlation within the 100 most enriched compounds identified after two successive rounds (2rnds) of selections against **Myc** G4 (A) and selection with **22CTA** G4 (B). A mild wash procedure (50°C) was implemented in the selection displayed in B whereas a harsh wash procedure (70°C) was implemented in the selection displayed in A (Table S1). Variation of node size and color (gradient from green to red) represent the abundance of individual building blocks across selections. Variation of building block connecting line size and colors represent the abundance of building block correlations across selections. Created with Cytoscape.

Table S2 : HTS normalized reads pre- and post-selections for selected compounds and controls and corresponding enrichment values.

Target	Compound	Normalized reads pre-selection	Normalized reads post-selection	Wash protocol	Enrichment	
MYC	1-PDP	221	1907	WASH1	8.6	
			2401	WASH3	10.9	
			3021	WASH4	13.7	
			4757	WASH5	21.5	
			2273	WASH4/WASH4	10.3	
			Strept.	181	WASH1	0.8
			hp1	436	WASH1	2
			Myc22	1964	WASH4	8.9
			Kit1	1381	WASH1	6.3
			22CTA	736	WASH1	3.3
wtTel26	1450	WASH1	6.6			
MYC	2-PDC3	36	88	WASH1	2.4	
			111	WASH3	3.1	
			158	WASH4	4.4	
			183	WASH5	5.1	
			59	WASH4/WASH4	1.6	
			Strept.	43	WASH1	1.2
			hp1	34	WASH1	0.9
			Myc22	194	WASH4	5.4
			Kit1	137	WASH1	3.8
			22CTA	230	WASH1	6.4
wtTel26	114	WASH1	3.2			
MYC	3-360A-l	222	778	WASH1	3.5	
			918	WASH3	4.1	
			1028	WASH4	4.6	
			1327	WASH5	6.0	
			843	WASH4/WASH4	3.8	
			Strept.	170	WASH1	0.8
			hp1	56	WASH1	0.3
			Myc22	604	WASH4	2.7
			Kit1	675	WASH1	3.0
			22CTA	567	WASH1	2.6
wtTel26	456	WASH1	2.1			
MYC	3-360A-s	208	1148	WASH1	5.5	
			959	WASH3	4.6	
			1360	WASH4	6.5	
			1879	WASH5	9.0	
			1816	WASH4/WASH4	8.7	
			Strept.	106	WASH1	0.5
			hp1	224	WASH1	1.1
			Myc22	2266	WASH4	10.9
			Kit1	928	WASH1	4.5
			22CTA	506	WASH1	2.4
wtTel26	456	WASH1	2.2			

MYC	4-PDP	281	224	WASH1	0.8	
			111	WASH3	0.4	
			221	WASH4	0.8	
			229	WASH5	0.8	
			105	WASH4/WASH4	0.4	
			Strept.	160	WASH1	0.6
			hp1	201	WASH1	0.7
			Myc22	194	WASH4	0.7
			Kit1	137	WASH1	0.5
			22CTA	92	WASH1	0.3
wtTel26	114	WASH1	0.4			
MYC.	5-360A-s	187	156	WASH1	0.8	
			161	WASH3	0.9	
			411	WASH4	2.2	
			366	WASH5	2.0	
			293	WASH4/WASH4	1.6	
			Strept.	149	WASH1	0.8
			hp1	22	WASH1	0.1
			Myc22	453	WASH4	2.4
			Kit1	232	WASH1	1.2
			22CTA	138	WASH1	0.7
wtTel26	228	WASH1	1.2			

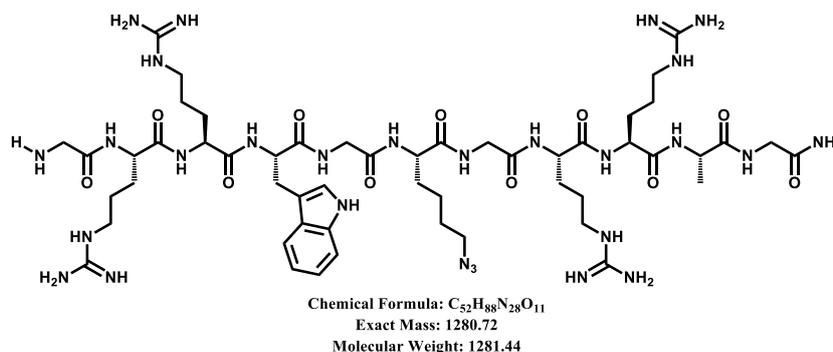
1.7 Off –DNA re-synthesis of hits and controls

Synthesis of azido peptides 1, t1, 2, 3, t3, 4 and 5

General details.

Peptide synthesis were performed on a Syro II synthesizer using the Fmoc/tBu strategy on a Rink-Amide PEG MBHA resin. The courses of reactions were monitored by using a Waters UPLC-MS system including reverse phase chromatography using a Peptide XB-C18 column (100 Å, 2,1 × 50 mm, 1,7µm) and with detection by UV spectrometry at 214 nm and 250 nm and combined electron spray ionization mass spectrometry. RP-HPLC purifications were performed on a Gilson system with a Nucleosil C18 column (100Å, 250 ×21 mm, 7µm) with UV monitoring at 214 nm and 250 nm. A 20 mL min⁻¹ flow linear gradient was applied from 95% solvent A1 (0.1% trifluoroacetic acid in water) and 5% solvent B1 (0.1% trifluoroacetic acid in acetonitrile) to 60% B1 for 20 min.

Peptide 1. H_2N -GRRWG-K(N₃)-GRRAG-CONH₂



Peptide 1 was synthesized using Fmoc-Rink-Amide PEG MBHA (loading of 0.41 mmol/g) commercial resin. The elongation was performed on a Syro II using the Fmoc/tBu strategy on the rink resin using HBTU as activator. Fmoc-Ala-OH, Fmoc-Gly-OH, Fmoc-Arg(Pbf)-OH and Fmoc-Trp(Boc)-OH were commercially available. Fmoc-azidonorleucine-OH (Fmoc-Lys(N₃)-OH) was synthesized according to reported protocols. The elongated peptide was then cleaved from the resin and deprotected using TFA/TIS/H₂O solution (95/2.5/2.5, v/v/v). The resulting solution was evaporated under vacuum and the peptide was precipitated in ether to obtain a white powder. A portion of the crude product was purified by RP-HPLC and freeze-dried to obtain peptide 1 as a white powder (15.3 µmol, 19.6 mg).

HRMS (+): m/z calcd for C₅₂H₉₁N₂₆O₁₁: 427.91352 [M+3H]³⁺, m/z found: 427.91267.

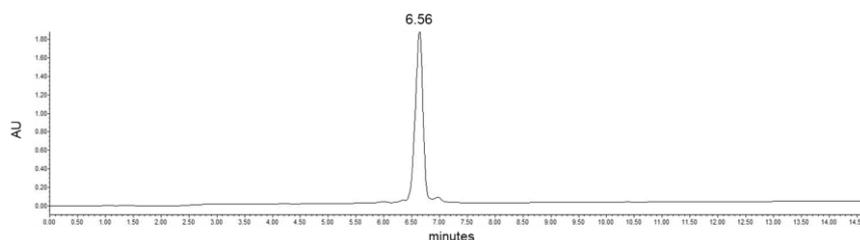
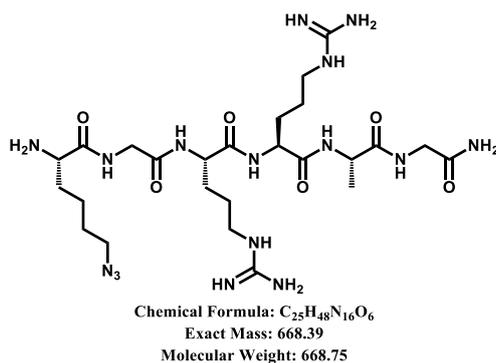


Figure S30 : HPLC chromatogram of pure peptide 1.

Peptide t1. H_2N -K(N₃)-GRRAG-CONH₂



Peptide **t1** was synthesized as described above. The elongated peptide was then cleaved from the resin and deprotected by using TFA/TIS/H₂O solution (95/2.5/2.5, v/v/v). The resulting solution was evaporated under vacuum and the peptide was precipitated in ether to obtain a white powder. A portion of the crude product was purified by RP-HPLC and freeze-dried to obtain peptide **t1** as a white powder (41.0 μmol, 27.4 mg).

ESI-MS (+): m/z calcd for C₂₅H₅₀N₁₆O₆: 335.1947 [M+2H]⁺, m/z found: 335.1952.

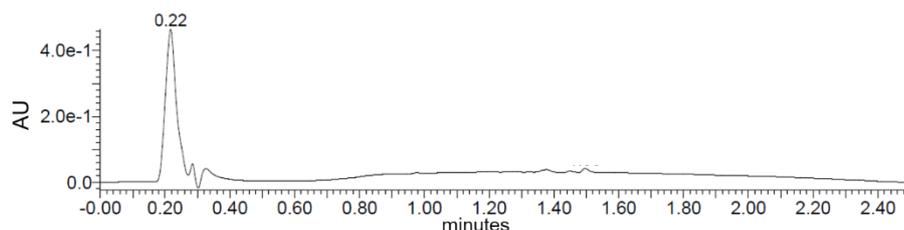
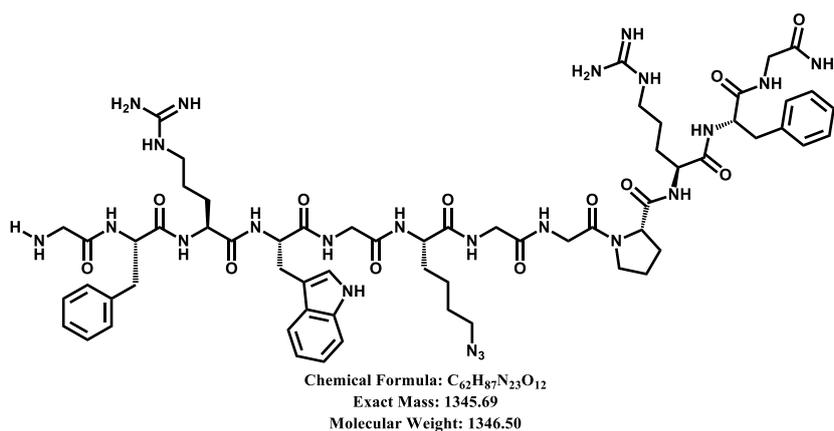


Figure S31 : UPLC chromatogram of pure peptide **t1**.

Peptide 2. H_2N -GFRWG-K(N₃)G-GPRFG-CONH₂



Peptide **2** was synthesized as described above using the following protected amino acids Fmoc-Gly-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Trp(Boc)-OH and Fmoc-Lys(N₃)-OH. The elongated peptide was then cleaved from the resin and deprotected by using TFA/TIS/H₂O solution (95/2.5/2.5, v/v/v). The resulting solution was evaporated under vacuum and the peptide was precipitated in ether to obtain a white powder. A portion of the

crude product was purified by RP-HPLC and freeze-dried to obtain peptide **2** as a white powder (22.28 μmol , 30 mg).

HRMS (+): m/z calcd for $\text{C}_{62}\text{H}_{89}\text{O}_{12}\text{N}_{23}$: 673.8525 $[\text{M}+2\text{H}]^+$, m/z found: 673.85215.

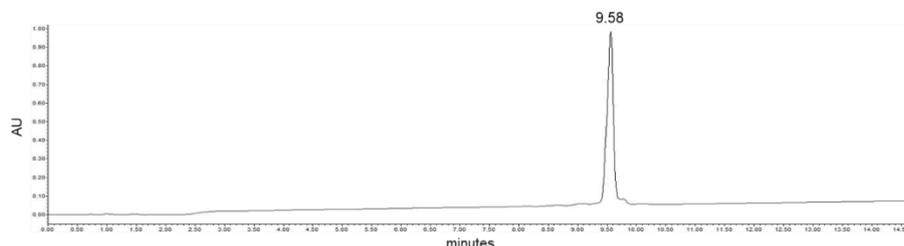
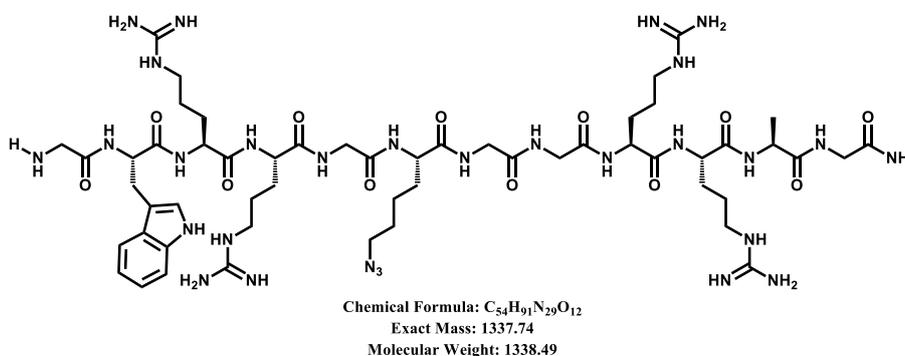


Figure S32 : HPLC chromatogram of pure peptide **2**.

Peptide 3. $\text{H}_2\text{N-GWRRG-K(N}_3\text{)G-GRRAG-CONH}_2$



Peptide **3** was synthesized as described above using the following protected amino acids Fmoc-Ala-OH, Fmoc-Gly-OH, Fmoc-Arg(Pbf)-OH and Fmoc-Trp(Boc)-OH and Fmoc-Lys(N_3)-OH. The elongated peptide was then cleaved from the resin and deprotected by using TFA/TIS/ H_2O solution (95/2.5/2.5, v/v/v). The resulting solution was evaporated under vacuum and the peptide was precipitated in ether to obtain a white powder. A portion of the crude product was purified by RP-HPLC and freeze-dried to obtain peptide **3** as a white powder (15.1 μmol , 20.2 mg).

HRMS (+): m/z calcd for $\text{C}_{54}\text{H}_{93}\text{N}_{29}\text{O}_{12}$: 669.87738 $[\text{M}+2\text{H}]^+$, m/z found: 669.87695.

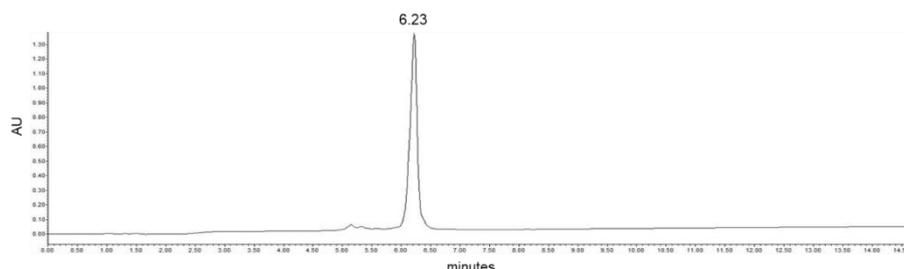
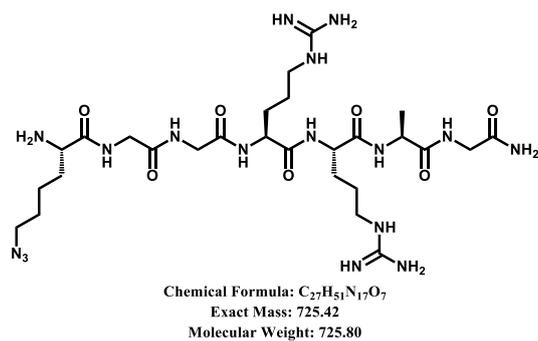


Figure S33 : HPLC chromatogram of pure peptide **3**.

Peptide t3. $H_2N-K(N_3)G-GRRAG-CONH_2$



Peptide **t3** was synthesized as described above using the following protected amino acids Fmoc-Ala-OH, Fmoc-Gly-OH, Fmoc-Arg(Pbf)-OH and Fmoc-Lys(N_3)-OH. The elongated peptide was then cleaved from the resin and deprotected by using TFA/TIS/ H_2O solution (95/2.5/2.5, v/v/v). The resulting solution was evaporated under vacuum and the peptide was precipitated in ether to obtain a white powder. A portion of the crude product was purified by RP-HPLC and freeze-dried to obtain peptide **t3** as a white powder (32.3 μ mol, 23.4 mg).

ESI-MS (+): m/z calcd for $C_{27}H_{53}N_{17}O_7$: 363.7089 $[M+2H]^+$, m/z found: 363.7105.

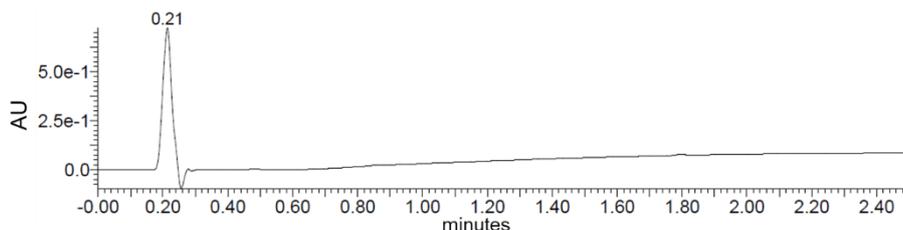
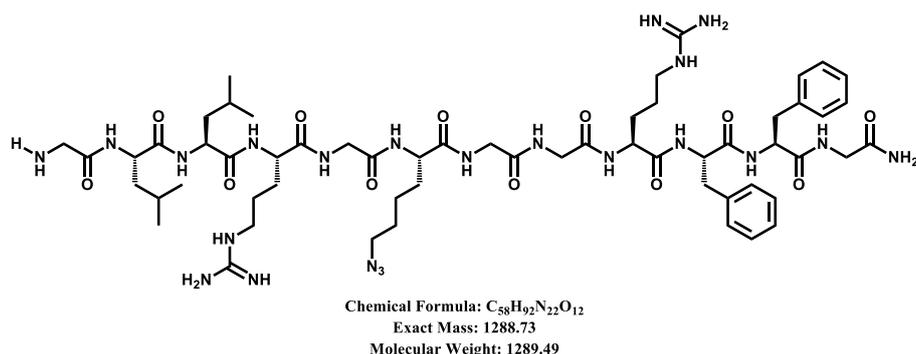


Figure S34 : UPLC chromatogram of pure peptide **t3**.

Peptide 4. $H_2N-GLLRG-K(N_3)G-GRFFG-CONH_2$



Peptide **4** was synthesized as described above using the following protected amino acids Fmoc-Gly-OH, Fmoc-Phe-OH, Fmoc-Leu-OH, Fmoc-Arg(Pbf)-OH and Fmoc-Lys(N_3)-OH. The elongated peptide was then cleaved from the resin and deprotected by using TFA/TIS/ H_2O solution (95/2.5/2.5, v/v/v). The resulting solution was evaporated under vacuum and the peptide was precipitated in ether to obtain a white powder. A portion of the crude product was

purified by RP-HPLC and freeze-dried to obtain peptide **2** as a white powder (19.62 μmol , 25.3 mg).

HRMS (+): m/z calcd for $\text{C}_{58}\text{H}_{94}\text{O}_{12}\text{N}_{22}$: 645.37053 $[\text{M}+2\text{H}]^+$, m/z found: 645.36978.

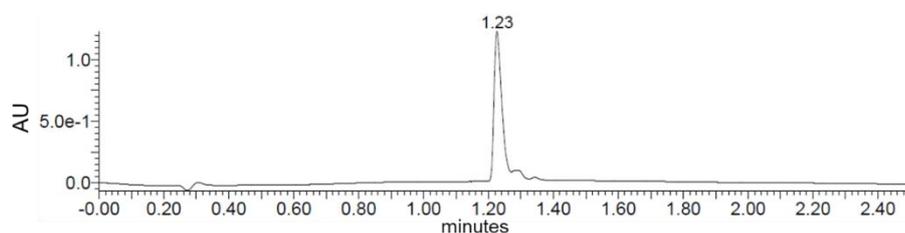
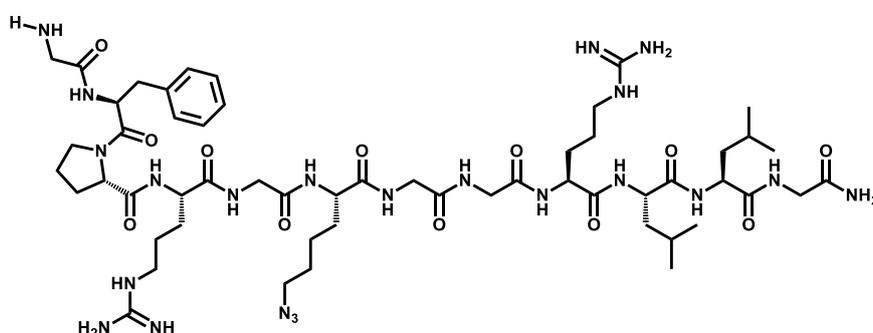


Figure S35 : UPLC chromatogram of pure peptide **4**.

Peptide 5. H_2N -GFPRG-K(N₃)G-GRLLG-CONH₂



Chemical Formula: $\text{C}_{54}\text{H}_{90}\text{N}_{22}\text{O}_{12}$
Exact Mass: 1238.71
Molecular Weight: 1239.43

Peptide **5** was synthesized as described above using the following protected amino acids moc-Gly-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Leu-OH, Fmoc-Arg(Pbf)-OH and Fmoc-Lys(N₃)-OH. The elongated peptide was then cleaved from the resin and deprotected by using TFA/TIS/H₂O solution (95/2.5/2.5, v/v/v). The resulting solution was evaporated under vacuum and the peptide was precipitated in ether to obtain a white powder. A portion of the crude product was purified by RP-HPLC and freeze-dried to obtain peptide **5** as a white powder (23.2 μmol , 28.8 mg).

ESI-MS (+): m/z calcd for $\text{C}_{54}\text{H}_{92}\text{O}_{12}\text{N}_{22}$: 620.3576 $[\text{M}+2\text{H}]^+$, m/z found: 653.3543.

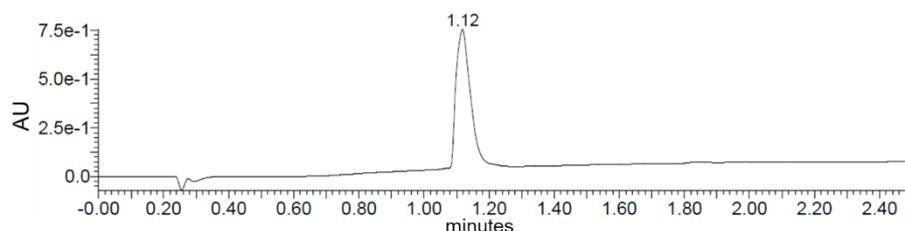
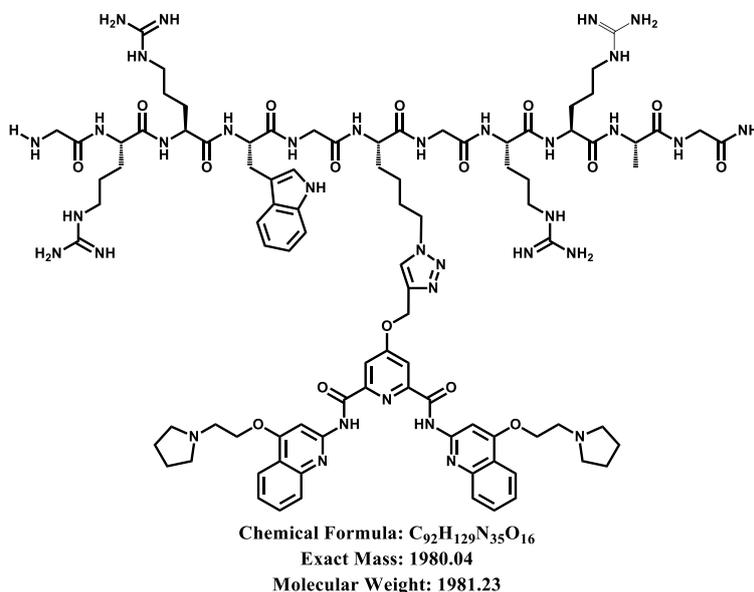


Figure S36 : UPLC chromatogram of pure peptide **5**.

Synthesis of compounds 1-PDP, t1-PDP, 2-PDC3, 3-360A-s, 3-360A-l, t3-360A-s, 4-PDP, 5-360A-s.

General protocol for Off-DNA CuAAC reaction: To a solution of peptide (10 mg, 100 μL , 1 eq.) was successively added a solution of the alkyne functionalized(hetero)aromatic derivative (0.9 eq., 100 μL in degassed DMSO/ H_2O solution (4:1)), THPTA (3 eq., 80 μL in degassed DMSO/ H_2O solution (4:1)), aminoguanidine-HCl (15 eq., 80 μL in degassed DMSO/ H_2O solution (4:1)) and $\text{CuSO}_4 \cdot \text{H}_2\text{O}$ (1.5 eq., 40 μL in degassed DMSO/ H_2O solution (4:1)). Degassed HEPES-buffer (10X, 65 μL) was then added at 10% of the final volume ($V_f = 650 \mu\text{L}$) and degassed DMSO was added (65 μL). Lastly, sodium ascorbate in degassed H_2O (15 eq., 120 μL) was added to initiate the reaction. The mixture was incubated at 25°C for up to 6 h. The progress of the reaction was monitored every hour using UPLC-MS. After completion of the reaction, an aqueous sodium EDTA solution (0.5 M, 130 $\mu\text{L} = 20\%$ of V_f) was added to the mixture. The crude product was purified by semi-preparative RP-HPLC (Gilson system – Nucleosil C18 column 100\AA , $250 \times 21 \text{ mm}$, $7\mu\text{m}$) with UV monitoring at 214 nm and 250 nm. A 20 mL min^{-1} flow linear gradient was applied from 95% solvent A1 (0.1% trifluoroacetic acid in water) and 5% solvent B1 (0.1% trifluoroacetic acid in acetonitrile) to 60% B1 for 20 min to afford the desired compound as a solid powder after lyophilization. The purity and identity of the product was confirmed by UPLC-MS and HRMS according to the corresponding general procedures.

Synthesis of **1-PDP**.



1-PDP was synthesis from peptide **1** (5.44 μmol) and **PDP-alk** building blocks using the Off-DNA CuAAC reaction protocol. The crude product was purified by RP-HPLC and freeze-dried (1.76 μmol , 36% yield).

HRMS (+): m/z calcd for $\text{C}_{92}\text{H}_{134}\text{N}_{35}\text{O}_{16}$: 397.0078 $[\text{M}+5\text{H}]^{5+}$, m/z found: 397.0081.

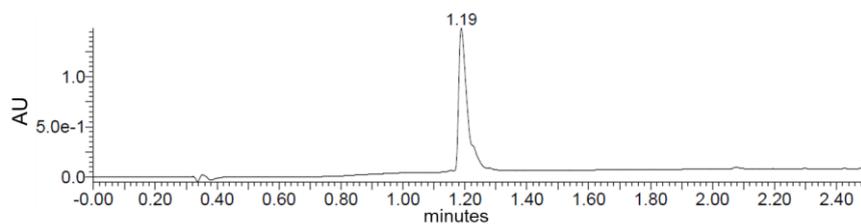
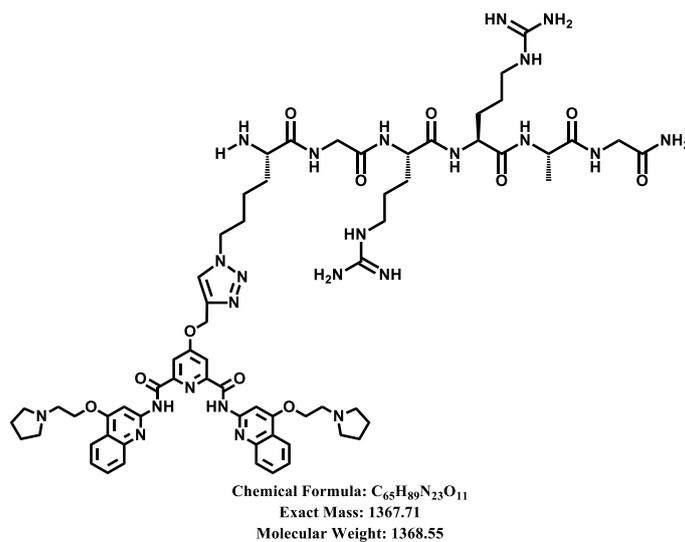


Figure S37 : UPLC chromatogram of pure **1-PDP**.

Synthesis of **t1-PDP**.



t1-PDP was synthesis from peptide **t1** (6.40 μ mol) and **PDP-alk** building blocks using the Off-DNA CuAAC reaction protocol. The crude product was purified by RP-HPLC and freeze-dried (0.79 μ mol, 14% yield).

HRMS (+): m/z calcd for $C_{65}H_{91}N_{23}O_{11}$: 684.86287 $[M+2H]^{2+}$, m/z found: 684.86204.

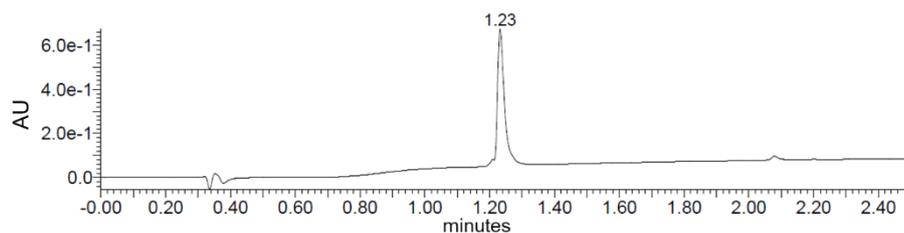
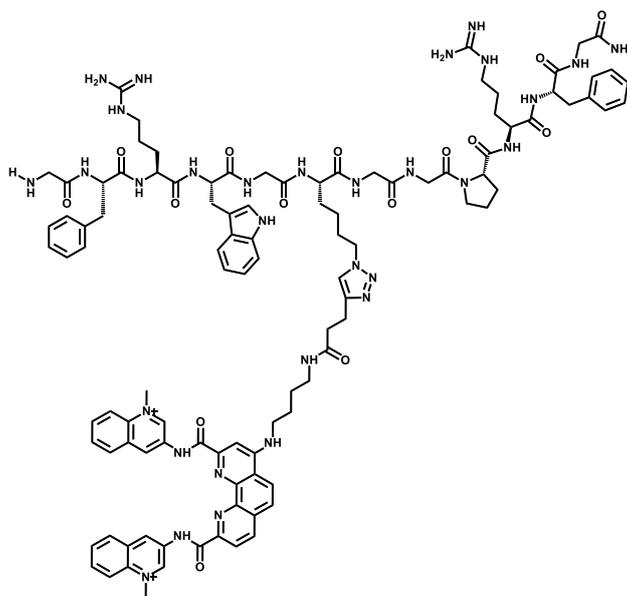


Figure S38 : UPLC chromatogram of pure **t1-PDP**.

Synthesis of **2-PDC3**.



Chemical Formula: $C_{105}H_{127}N_{31}O_{15}^{2+}$
Exact Mass: 2062.01
Molecular Weight: 2063.33

2-PDC3 was synthesized from peptide **2** (5.55 μmol) and **PDC3-alk** building blocks using the Off-DNA CuAAC reaction protocol. The crude product was purified by RP-HPLC and freeze-dried (2.37 μmol , 47% yield).

HRMS (+): m/z calcd for $C_{105}H_{129}N_{31}O_{15}^{2+}$: 516.00656 $[M+2H]^{4+}$, m/z found: 516.00555.

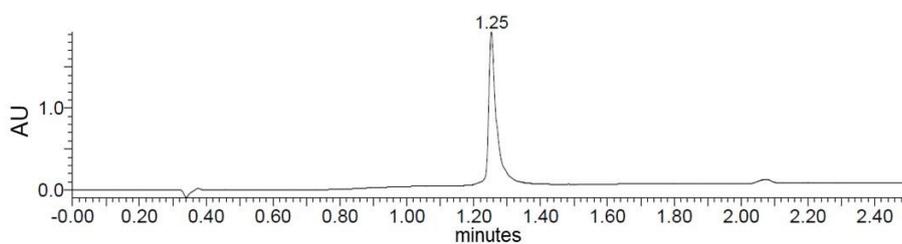
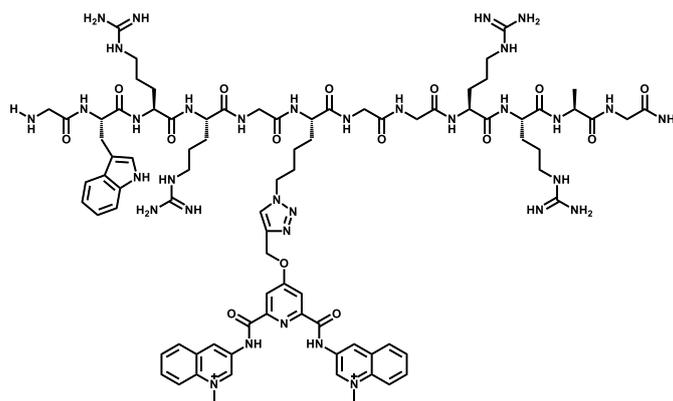


Figure S39 : UPLC chromatogram of pure **2-PDC3**.

Synthesis of **3-360A-s**.



Chemical Formula: $C_{84}H_{116}N_{34}O_{15}^{2+}$
Exact Mass: 1840.93
Molecular Weight: 1842.04

3-360A-s was synthesis from peptide **3** (4.95 μ mol) and **360A-s-alk** building blocks using the Off-DNA CuAAC reaction protocol. The crude product was purified by RP-HPLC and freeze-dried (2.62 μ mol, 59% yield).

HRMS (+): m/z calcd for $C_{84}H_{118}N_{34}O_{15}^{2+}$: 460.73735 $[M+2H]^{4+}$, m/z found: 460.73666.

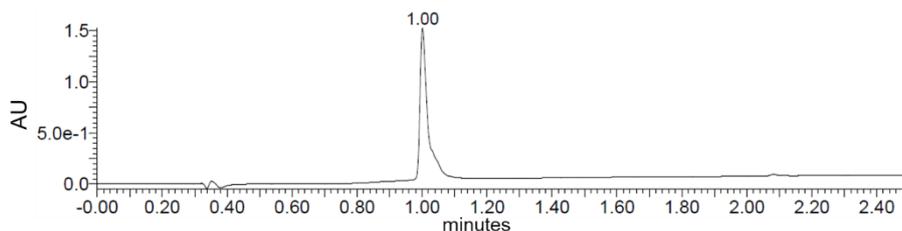
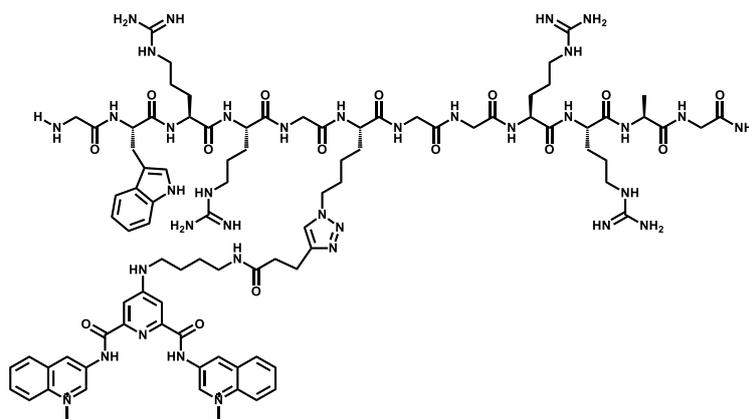


Figure S40 : UPLC chromatogram of pure **3-360A-s**.

Synthesis of **3-360A-I**.



Chemical Formula: $C_{90}H_{128}N_{36}O_{15}^{2+}$
Exact Mass: 1953.03
Molecular Weight: 1954.21

3-360A-I was synthesis from peptide **3** (1.74 μ mol) and **360A-I-alk** building blocks using the Off-DNA CuAAC reaction protocol. The crude product was purified by RP-HPLC and freeze-dried (0.37 μ mol, 24% yield).

HRMS (+): m/z calcd for $C_{90}H_{130}N_{36}O_{15}^{2+}$: 489.2575 $[M+4H]^{6+}$, m/z found: 489.2578

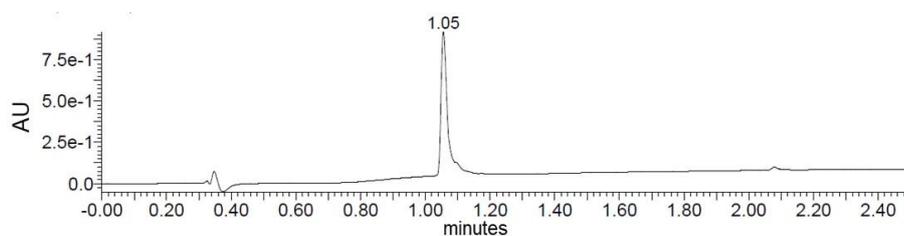
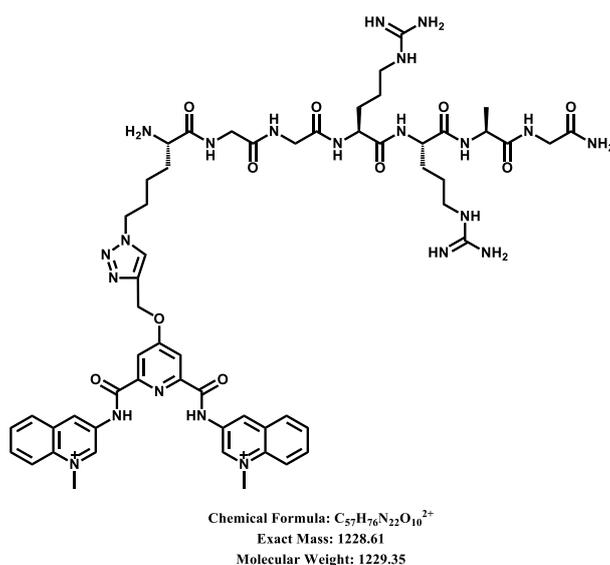


Figure S41 : UPLC chromatogram of pure **3-360A-I**.

Synthesis of **t3-360A-s**.



t3-360A-s was synthesis from peptide **t3** (9.36 μ mol) and **360A-s-alk** building blocks using the Off-DNA CuAAC reaction protocol. The crude product was purified by RP-HPLC and freeze-dried (5.13 μ mol, 61% yield).

HRMS (+): m/z calcd for $C_{57}H_{79}N_{22}O_{10}^{2+}$: 409.87255 $[M+3H]^{5+}$, m/z found: 409.87191

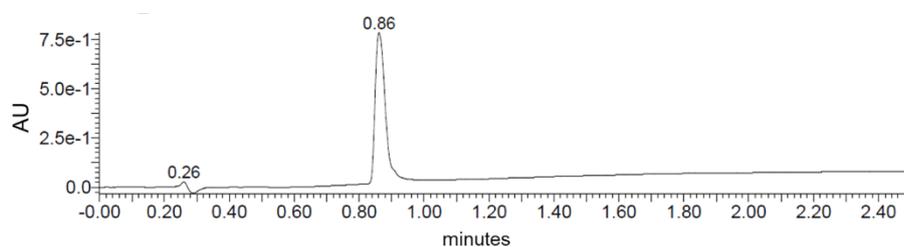
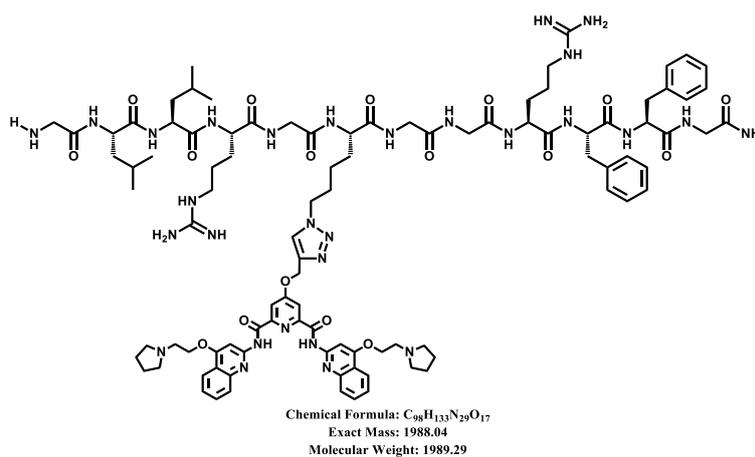


Figure S42 : UPLC chromatogram of pure **t3-360A-s**.

Synthesis of **4-PDP**.



4-PDP was synthesis from peptide **4** (6.13 μmol) and **PDP-alk** building blocks using the Off-DNA CuAAC reaction protocol. The crude product was purified by RP-HPLC and freeze-dried (1.58 μmol , 29% yield).

HRMS (+): m/z calcd for $C_{98}H_{135}N_{29}O_{17}$: 995.02899 $[M+2H]^{2+}$, m/z found: 995.02824

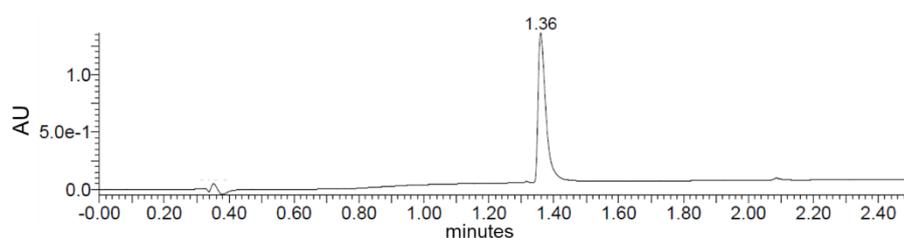
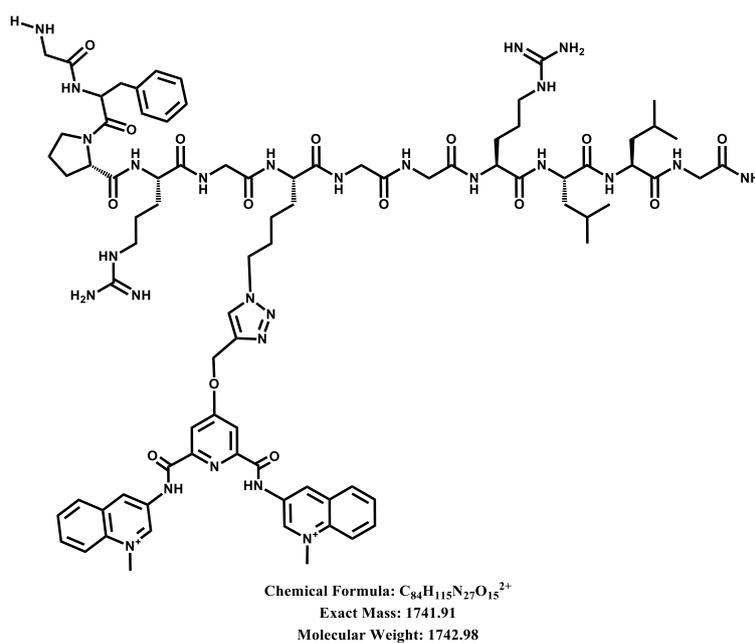


Figure S43 : UPLC chromatogram of pure **4-PDP**.

Synthesis of **5-360A-s**.



5-360A-s was synthesis from peptide **5** (6.33 μmol) and **360A-s-alk** building blocks using the Off-DNA CuAAC reaction protocol. The crude product was purified by RP-HPLC and freeze-dried (2.55 μmol , 44% yield).

HRMS (+): m/z calcd for $\text{C}_{84}\text{H}_{118}\text{N}_{27}\text{O}_{15}^{2+}$: 580.97093 $[\text{M}+3\text{H}]^{5+}$, m/z found: 580.96969

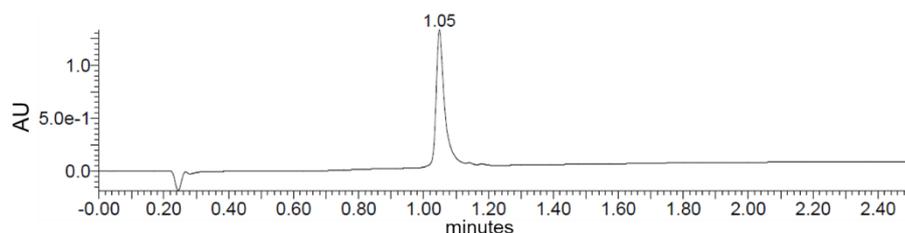


Figure S44 : UPLC chromatogram of pure **5-360A-s**.

1.8 Biophysical evaluation of hits and controls

1.8.1 Fluorescence melting experiments (FRET-melting)

Fluorescence melting experiments were performed using an Agilent real-time qPCR machine according to reported procedures. The assay was performed with double-labelled oligonucleotides (5'-FAM, 3'-TAMRA, Tableau 21) and the fluorescence intensity was monitored in the FAM channel (462/516nm, respectively for the excitation and the detection wavelength). Thermal denaturation runs were performed in 96-well plates, starting with an initial incubation at 25°C for 10 min. Then, a heating ramp of 0.5 °C per minute was applied from 25 to 95 °C. Each well contained 0.2 μM of double-labelled G4, 1 μM of tested compound (unless otherwise specified), in a final volume of 25 μL of buffer (10mM lithium cacodylate (pH 7.2), 100 mM ionic strength with KCl and LiCl as shown in Tableau 22). In competition experiments, 10 μM (50 eq) of competitor (scr26 or hp1) was present in the mixture. Each experiments were run, at least, in triplicates on, at least, three separate plates. The denaturation temperature (T_m) were determined from normalized curves (OriginPro) of FAM emission intensity vs. temperature. Ligand-induced T_m shifts ΔT_m are reported as a difference of mid-transition temperature in the presence and in the absence of ligands. Tableau 22 shows the melting temperature of the oligo targets alone or in presence of competitors.

1.8.2 Fluorimetric titration

Fluorimetric titrations were performed as previously reported. The 5'-Cy5-labelled oligonucleotides listed in table S3 were used for the titrations.

Table S3 : DNA sequences used as targets for fluorimetric titrations.

Target	Sequence (5'Cy5→3')	Structure	Source
5'Cy5-Myc	/5Cy5/TGA GGG TGG GGA GGG TGG GGA A	G4 (parallel)	IDT
5'Cy5-Myc22	/5Cy5/TTG AGG GTG GGT AGG GTG GGT AA	G4 (parallel)	IDT
5'Cy5-Kit1	/5Cy5/AGG GAG GGC GCT GGG AGG AGG GTT	G4 (parallel, isolated G)	IDT
5'Cy5-22CTA	/5Cy5/AGG GCT AGG GCT AGG GCT AGG GTT	G4 (hybrid)	IDT
5'Cy5-wtTel26	/5Cy5/TTA GGG TTA GGG TTA GGG TTA GGG TT	G4 (hybrid)	IDT

Circular dichroism studies were first performed with the 5'Cy5 oligonucleotide to confirm the expected folding (Figure S45). Annealing/folding were performed by heating each oligonucleotide (1,6 μM) at 80°C for 5 min in titration buffer (50 mM Tris.HCl, 150 mM KCl, 0.05%, v/v, Triton X-100) and cooling to room temperature over 2 h. CD spectra were recorded at 25°C with wavelength range of 220 to 320 nm on a Jasco J-810 spectropolarimeter using 1 cm length quartz cuvette. Each spectrum was an average of three scans with a 0.5 s response time, a 1 nm data pitch, a 4 nm bandwidth and a 200 nm.min⁻¹ scanning speed. CD analysis suggest that 5'Cy5-22CTA adopts a hybrid type G4 conformation unlike 22CTA-biot and F-22CTA-T which were shown to fold into and antiparallel G4 by CD.

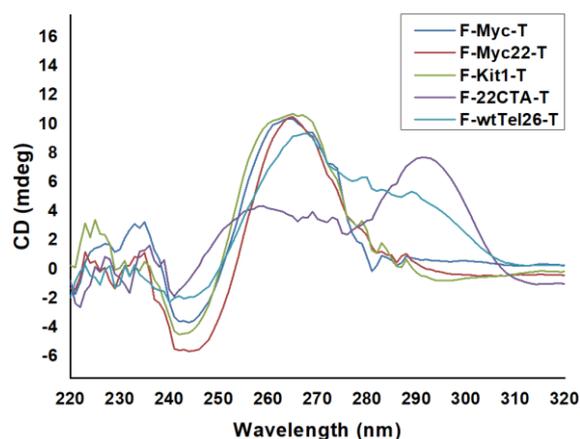
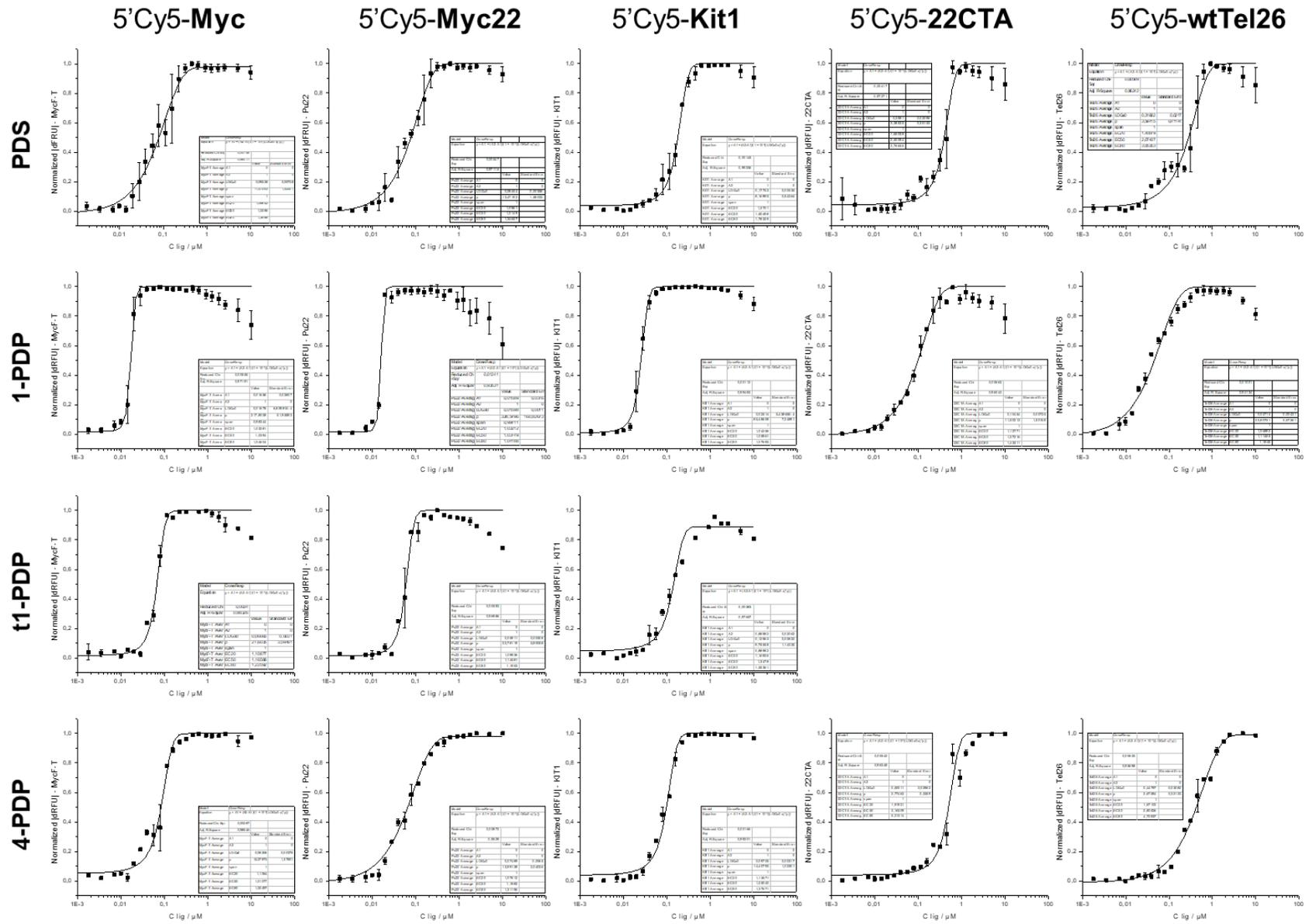
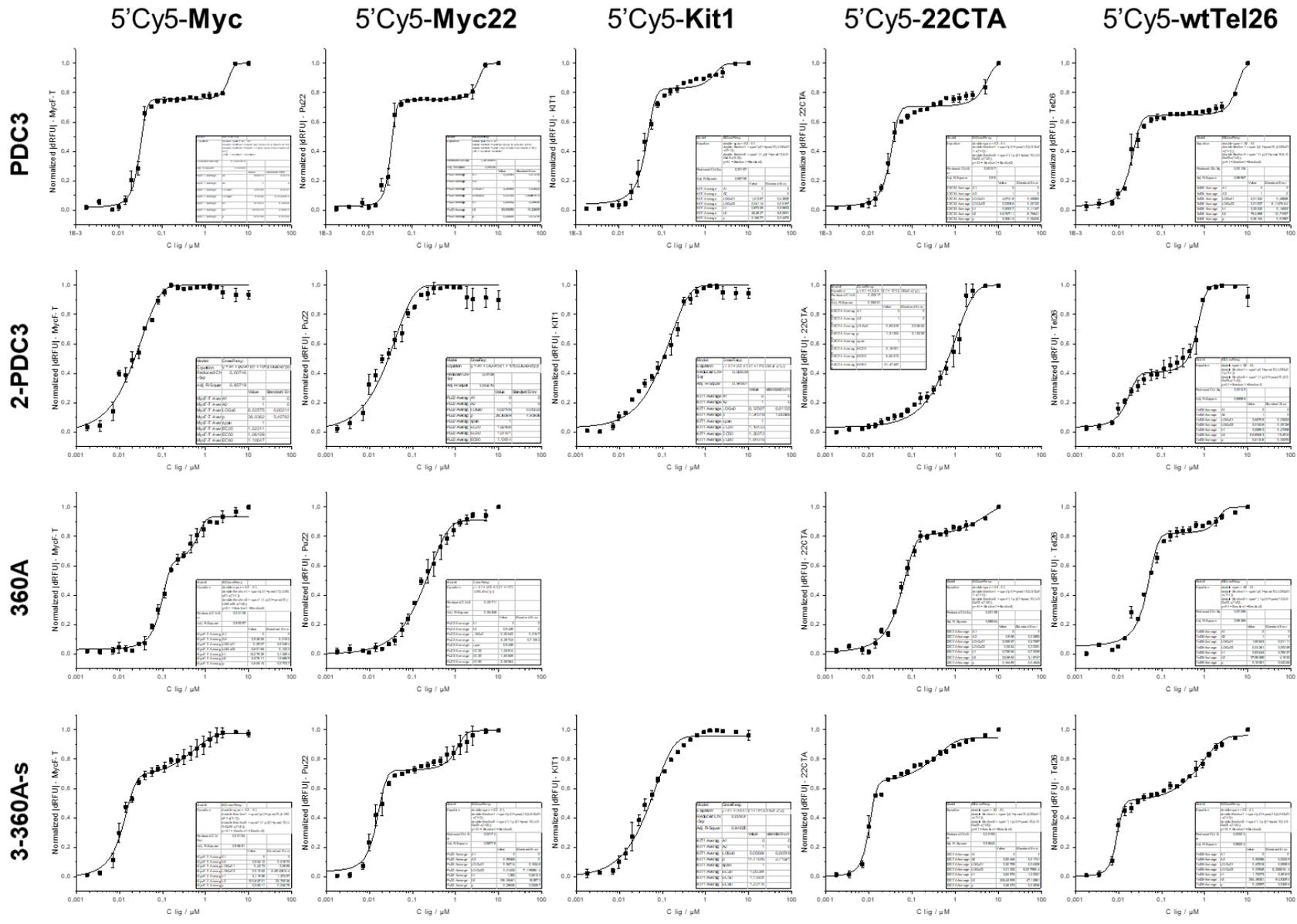


Figure S45: CD analyses (1.6 μM in 50 mM Tris.HCl, 150 mM KCl, 0.05%, v/v, Triton X-100) of 5'Cy5 oligonucleotide targets.

For fluorimetric titrations, 5'-Cy5-labelled oligonucleotides were dissolved at a concentration of 100 μM in water. Then the solutions were further diluted with the titration buffer (50mM TrisHCl, pH 7.2, 150 mM KCl, 0.05 v/v % Triton X-100) to a final concentration of 2.22 nM. Diluted solutions were annealed at 95°C for 5 min and slowly cooled to room temperature. During the cooling step, ten serial dilutions (1:1) of ligand in 0.1 v/v % DMSO and water was prepared from 100 μM and 18 μM respectively resulting in a total of 20 ligands solutions with different ligand concentrations (100 μM , 50 μM , 25 μM , 18 μM , 12.5 μM , 9 μM , 6.25 μM , 4.5 μM , 3.12 μM , 2.25 μM , 1.56 μM , 1.12 μM , 781 nM, 563 nM, 391 nM, 281 nM, 195 nM, 141 nM, 98 nM, 70 nM, 35 nM, 18 nM) and two no-compound controls. Then, 10 μL per solution of the dilution series was transferred to a 96-well plate (black, flat-bottom polystyrene non-binding (NBS) microplates – Corning). This was repeated five times to create five replicate rows containing the same dilution series. Annealed 5'-Cy5-labelled oligonucleotide solutions (90 μL) were added to the wells to reach a final oligonucleotide solution of 2 nM and a final ligand concentration from 10 μM to 1.8nM. After incubation for 2 hr at room temperature, fluorescence intensity was measured on a CLARIOstar Plusmicroplate reader (BMG) using 610BP30 (excitation), LP638 (dichroic) and 675BP50 (emission) filters and an integration time of 0.5 s per well. The experiments were performed in triplicate. The difference in fluorescence intensity (RFU) were converted to absolute values relative to the no-compound control (buffer control). The range of observed absolute RFU change was used to normalize the measurements (Figure S46). Normalized values were used as inputs in a one- or two-site binding model (OriginPro) for curve fitting (Figure S46).





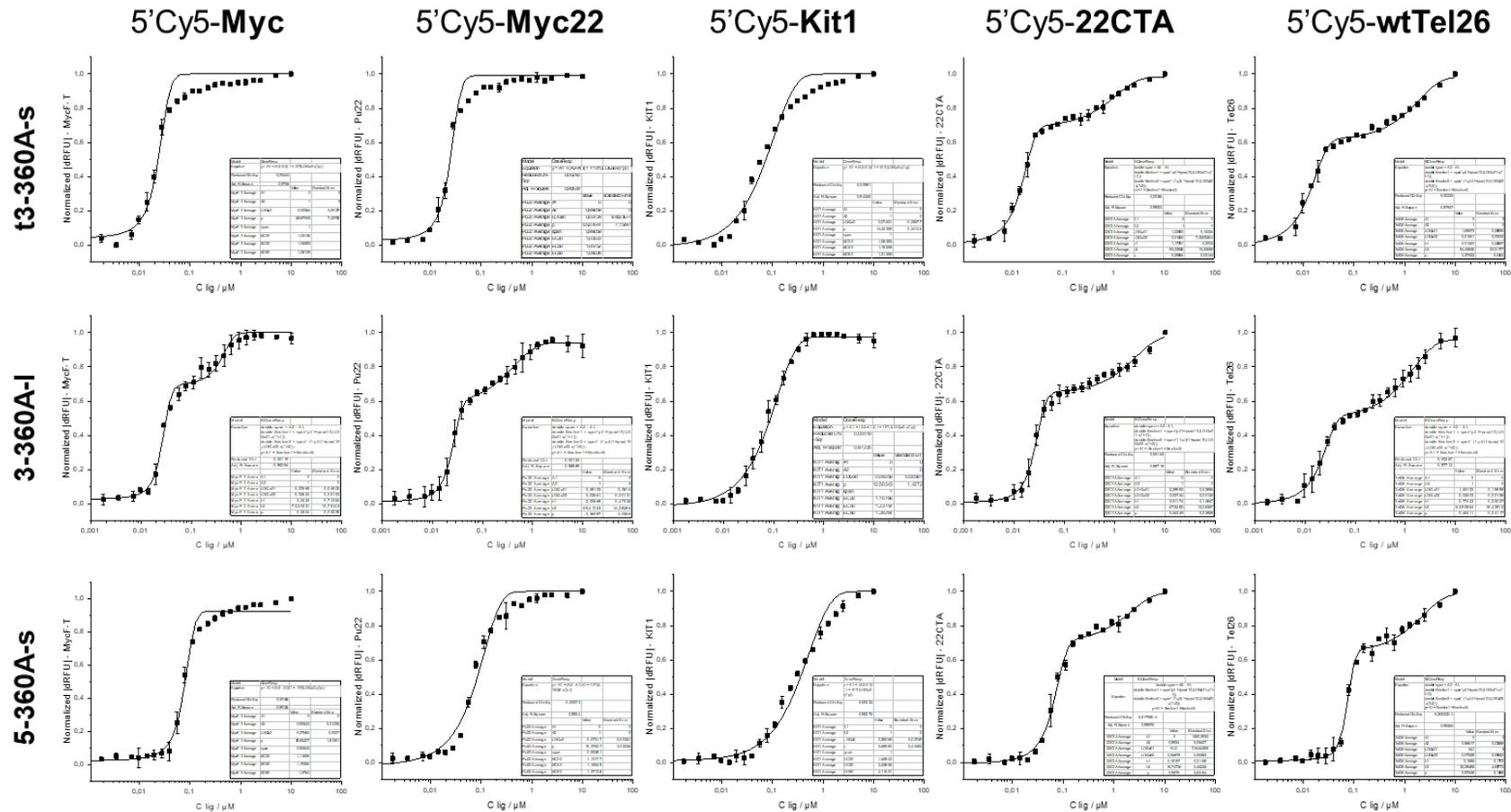


Figure S46 : Normalized of fluorescence intensity ($\lambda_{ex} = 610\text{nm}$, $\lambda_{em}=675\text{nm}$) upon titrations of ligands (final concentration: 1.8nM to 10 μM) to 5'-Cy5-labelled oligonucleotides ($[c]=2\text{nM}$ in 50mM Tris.HCl, 150mM KCl, 0.05 v/v % Triton X-100). Black lines represent fits of experimental data to a 1:1 or 2:1 binding model. Data are means \pm s.d. from three experiments (N=3).

1.8.3 Bio Layer Interferometry (BLI)

BLI sensors coated with streptavidin (SA sensors) were purchased from Forte Bio (PALL). They were immersed 10 minutes in buffer solution (10 mM HEPES, 150 mM KCl, 0.05% v/v P20, pH 7.4) before functionalization to remove the protective sucrose layer.

Then, the sensors were dipped for 15 minutes in biotinylated Myc G4 (100 nM in buffer solution, Myc-biot, Tableau 9, p129) and rinsed in buffer solution for 10 minutes. The functionalized sensors were dipped in G4 ligand solutions at different concentrations in buffer solution (Tableau 24, p218) and, after each association step, rinsed in the same buffer. The experimental parameters are reported in Tableau 24. Reference sensors not functionalized with biotinylated oligonucleotide were used as blank to correct for non-specific adsorption to the raw data. The sensorgrams are shown in Graphique 29, p219.

The association steps were fitted using a 1:1 model to obtain the k_{obs} value³⁴⁶. The k_{obs} parameter depends on the kinetic association rate k_{on} and dissociation rate k_{off} : $k_{obs} = k_{on} * C + k_{off}$, where C is the concentration of the analyte in the microplate. The plot of k_{obs} versus C allows the determination of the two kinetic constants (Figure S47): k_{on} the slope of the curve and k_{off} the intercept. From the kinetic constants, the equilibrium K_D^{Kin} was calculated from k_{off}/k_{on} . The equilibrium constant, K_D^{eq} , was also obtained by fitting the equilibrium responses in R_{eq} , as a function of analyte concentration from the Langmuir equation, assuming a 1:1 binding stoichiometry (Figure S48). The reported values are obtained from the average of representative independent experiments, and the errors provided are standard deviations from the mean. Each experiment was repeated at least two times.

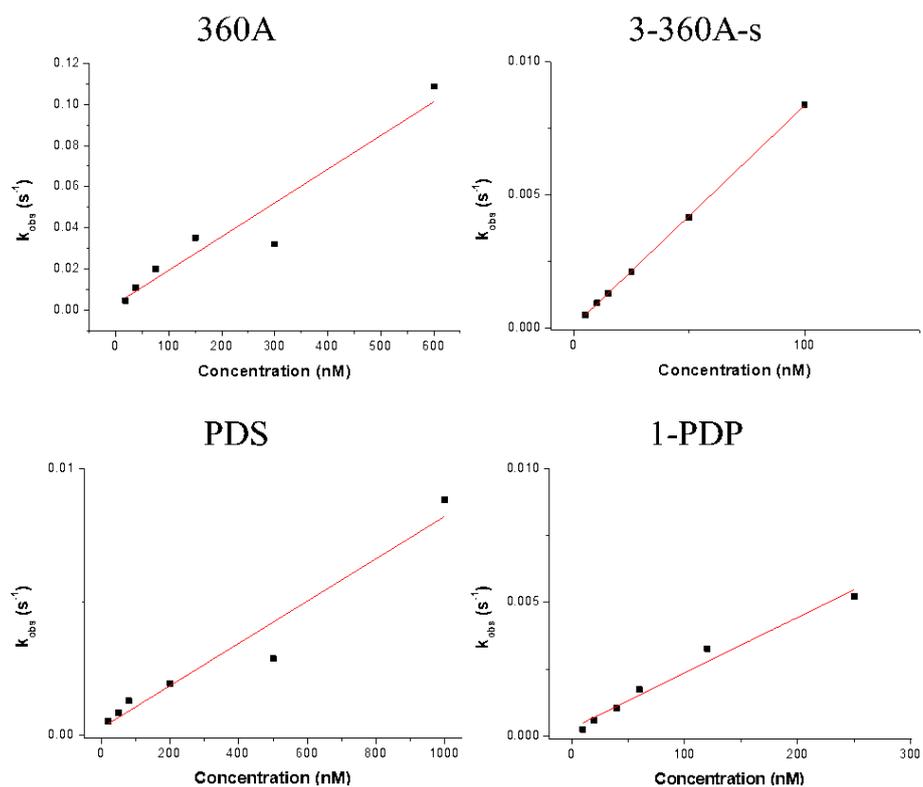


Figure S47 : Plots of observed rate constants (k_{obs}) versus ligand concentrations for ligands **1-PDP** and **3-360A-s** and reference ligands **PDS** and **360A** in interaction with Myc G4 (Myc-biot).

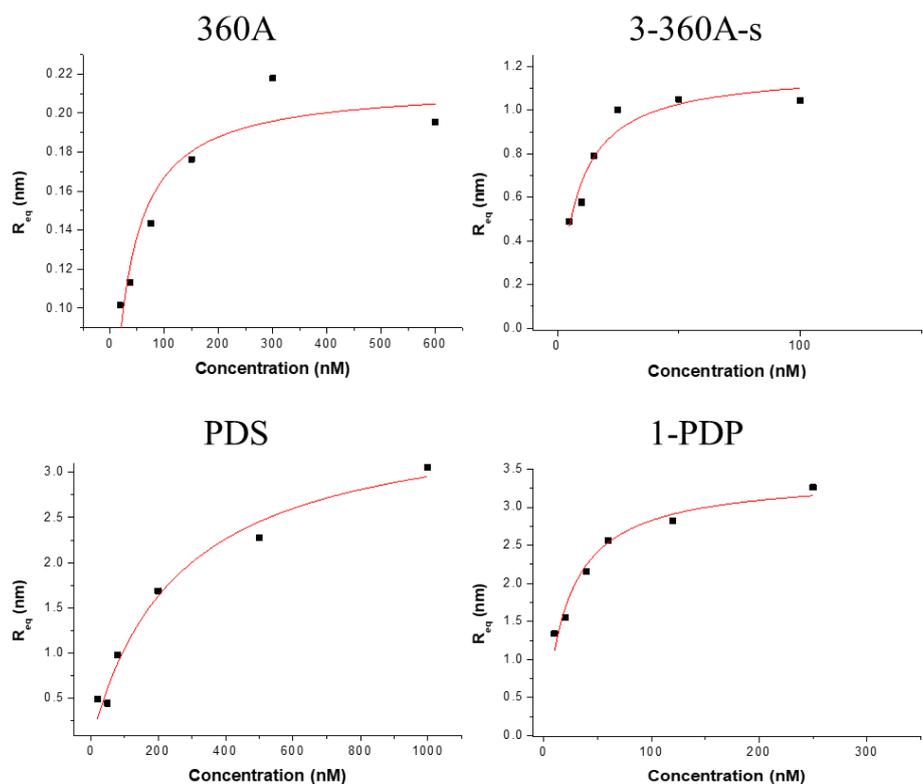


Figure S48 : Adsorption isotherm (square) and fitting curve (line) using a 1:1 Langmuir interaction model for ligands **1-PDP** and **3-360A-s** and reference ligands **PDS** and **360A** in interaction with Myc G4 (Myc-biot).

1.9 Computational Analysis

Using structure in PDB 2L7V of the Myc22, we deleted the experimentally solved ligand from the complex. Next, we docked 360A into the G4 using the software Autodock 4.2³⁴⁷. Similar protocol described already by Harikrishna and al. was used³⁴⁸. The geometry of 360A was optimized at HF/6-31g* level using Gaussian 16^{Gaussian 349}. The input files for docking calculations were prepared with ADT Tools using Gasteiger partial charges. The docking area was centered on the Cartesian coordinates at the center mass of the G4 structure. The grid points of 60*60*60 with 0.375 spacing were calculated for all the ligand atom types using AutoGrid 4.2. A conformational search was performed using a genetic algorithm with an initial population of 300 randomly placed individuals, a maximum number of 2.5×10^5 energy evaluations, a maximum number of 2.7×10^4 generations, a mutation rate of 0.02, and a crossover rate of 0.8. One hundred different conformers of the ligand were generated and the energy values of conformers were reported as a histogram. The best rank position was taken as the docked position.

Ligands 3-360A-s, 3-360A-l and 5-360A were constructed using the software Sybyl³⁵⁰. Then we aligned their 360A part with the docked position of 360A obtained in Myc22 using the “pair fitting” tool of Pymol (ref). Molecular mechanic (MM) dynamics were performed on these positions using the software Amber18³⁵¹. Specific parameters were constructed for the 360A part of the ligands with the generalized Amber force field (Gaff) and the antechamber RESP fitting procedure for the charges. ParmBSC1 force field were used for the G4 DNA and for the peptide linker of the ligands. Structures were immersed in a 10 Å rectangular box of TIP3P water molecules. Potassium cations were added and long-range electrostatic interactions were computed using the particle-mesh Ewald method with a cutoff value of 10 angstroms. The time step was set at 1fs and 500 ns of simulations were done in the NPT ensemble using the CUDA version of PMED in a GPU accelerated version of AMBER 18.

Simulations were performed for the three ligands 3-360A-s, 3-360A-l and 5-360A in interaction with Myc22. Then waters molecules and potassium cations except cations in the channel of the G4 were removed from the MD trajectories to perform a cluster analysis using CPPTRAJ module in AmberTools 18. For each system, the cluster analysis was carried using the density_based clustering algorithm, with the coordinate RMSD as distance matrix and in generating the pair_wise distance matrix by using 10 frames for the initial clustering.

Geometry of the representative structure of the most populated cluster, structure which has the minimal sum of the squared displacements with other structures in the cluster and itself, was optimized at crude level using the semi-empirical density functional tight binding GFN2-xTB³⁵². Solvation is included with the analytical linearized Poisson Boltzmann (ALPB) model.

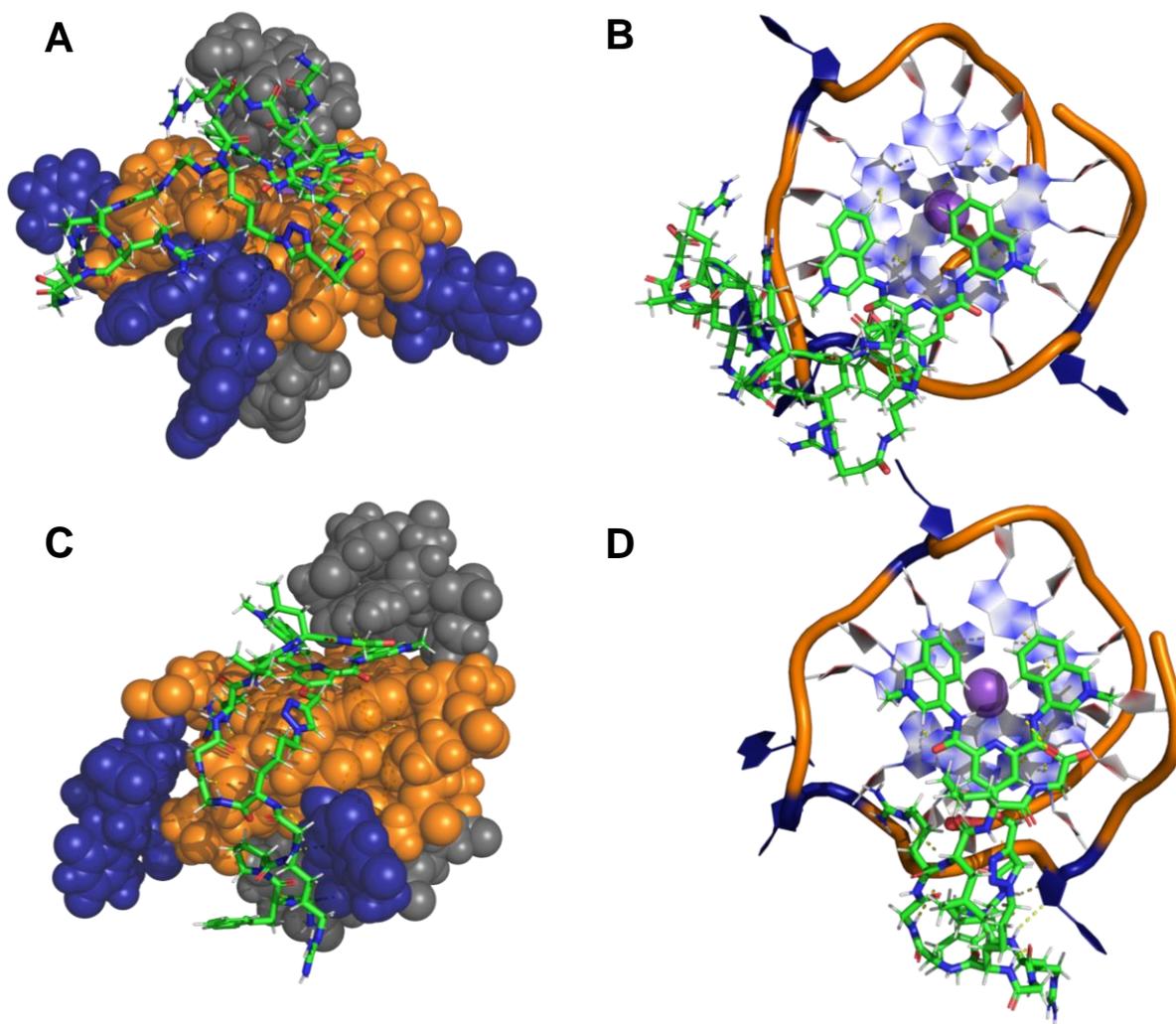


Figure S49 : Best pose obtained from the molecular dynamic simulation of the interaction of **3-360A-I** (A: side view, B: top view) and **5-360A** (C: side view, D: top view) with Myc22 (PDB:2L7V). Ligands in side views are shown as sticks and Myc22 is shown as spheres with G-tetrads in orange, loop residues in blue and flanking residues in dark grey. Ligands in top views are shown as sticks and Myc22 is displayed as residue-based cartoons and ribbon, G4 loop residues are colored in blue, the 5' flanking sequence dTdGdA has been removed for clarity.

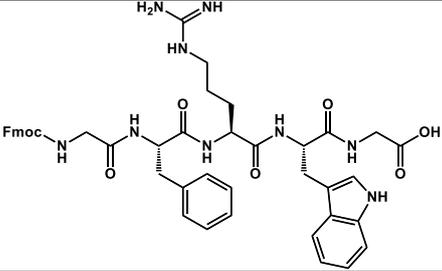
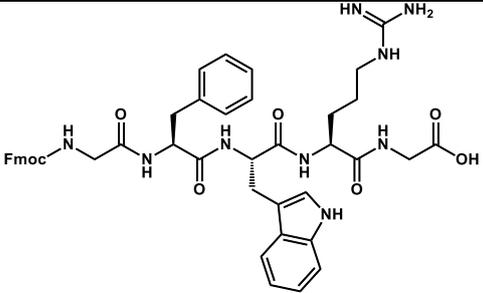
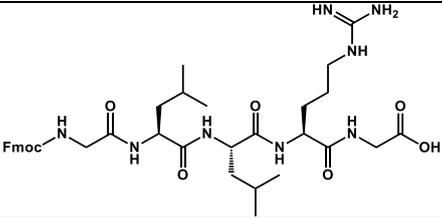
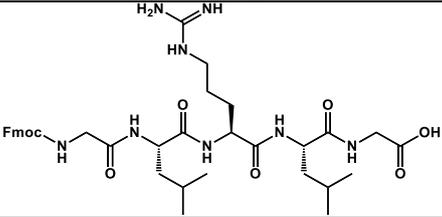
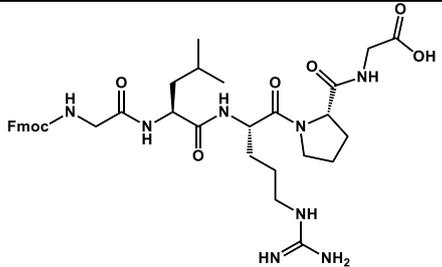
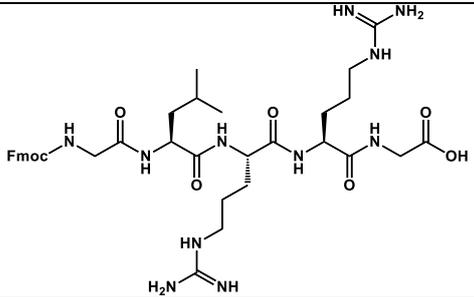
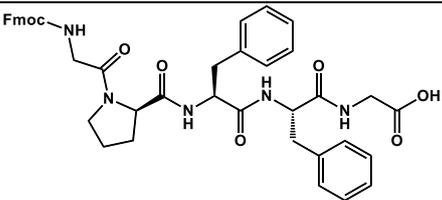
2. Supplementary tables

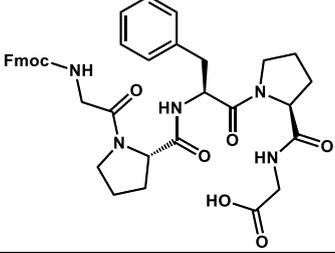
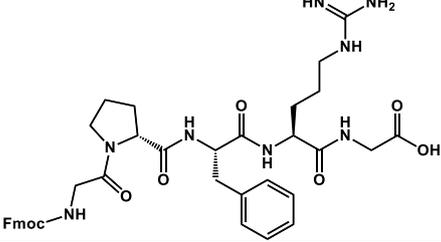
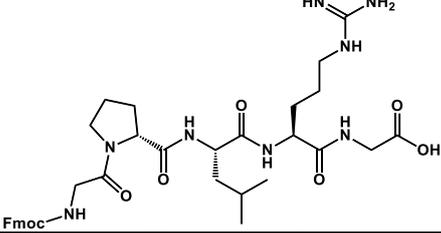
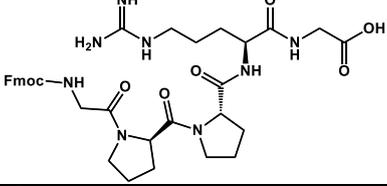
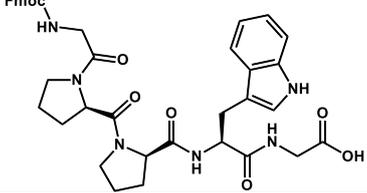
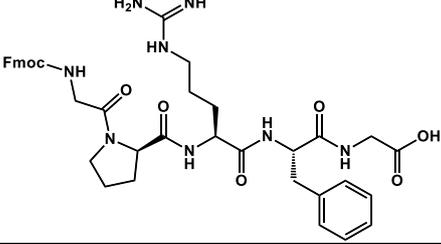
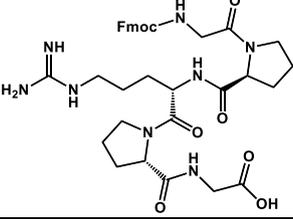
Chemical structure, validation and source of building blocks.

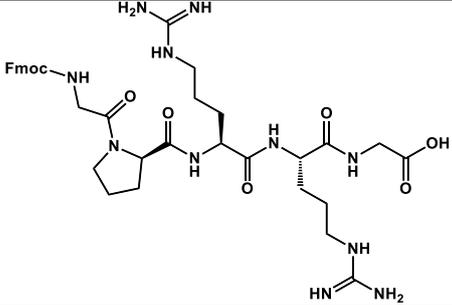
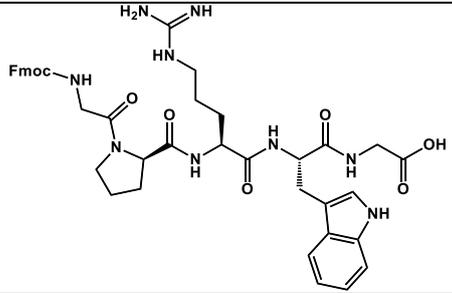
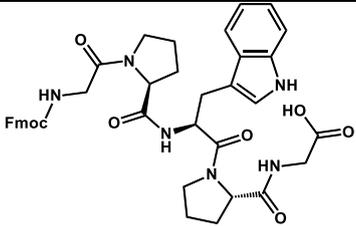
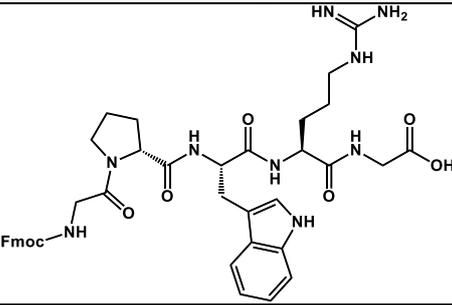
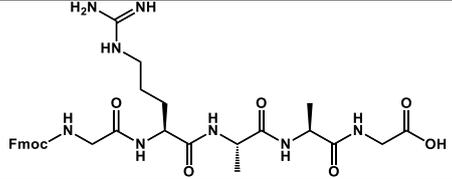
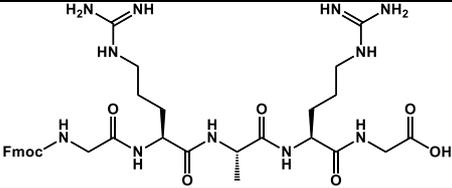
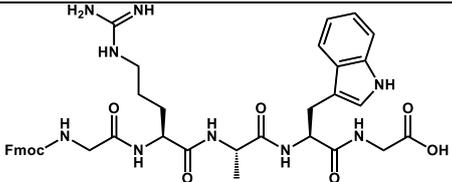
Building blocks [Cycle 1 & 4] – DEL1 & DEL2

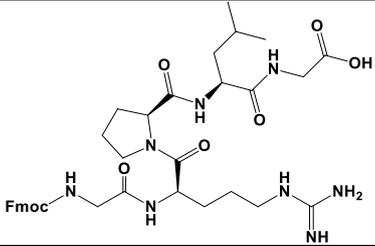
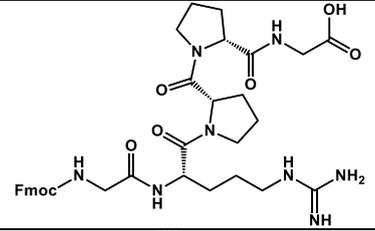
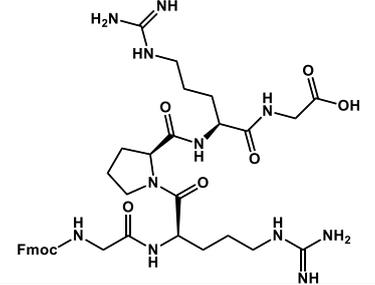
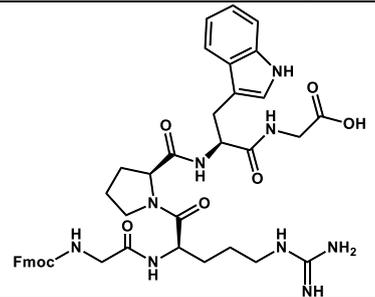
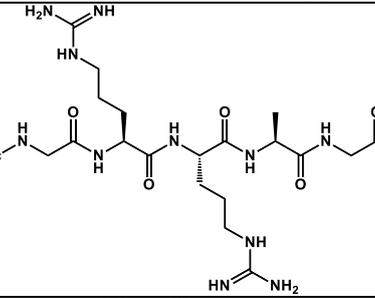
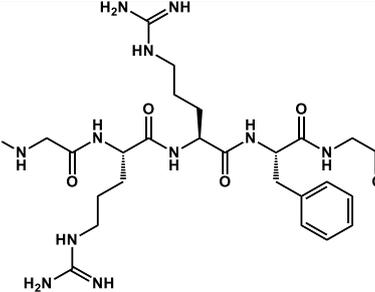
Table S4 : BB1 & BB4 – DEL1 & DEL2.

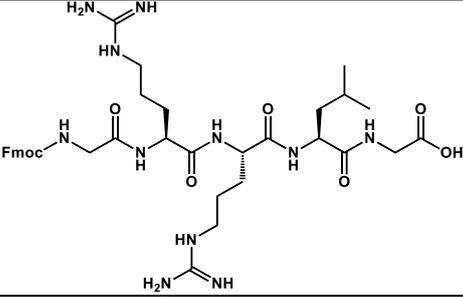
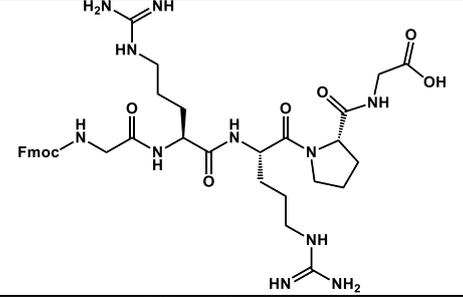
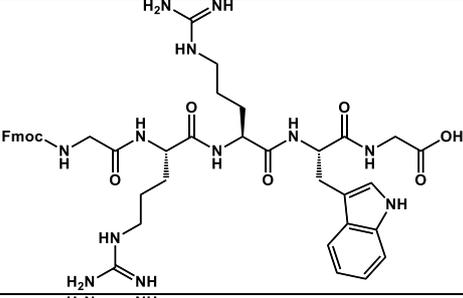
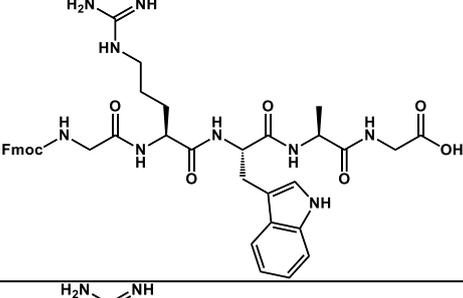
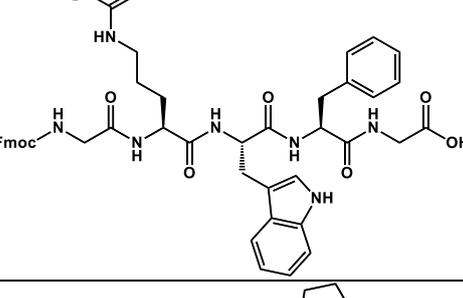
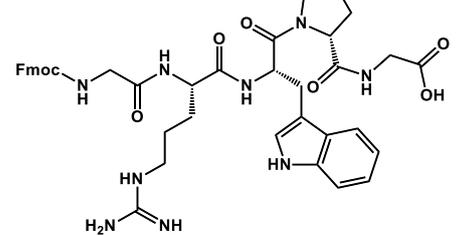
Building block ID	Structure	Acylation yield & Protocol	Source
GAARG DEL1 DEL2		73% Protocol 2	Genscript
GAAWG DEL1 DEL2		73% Protocol 2	Genscript
GARAG DEL2		75% Protocol 1	Genscript
GARRG DEL1		60% Protocol 2	Genscript
GARWG DEL1 DEL2		67% Protocol 2	Genscript
GAWAG DEL1 DEL2		67% Protocol 2	Genscript

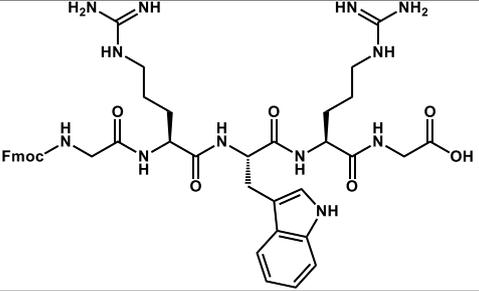
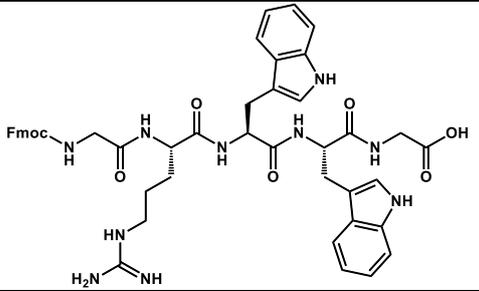
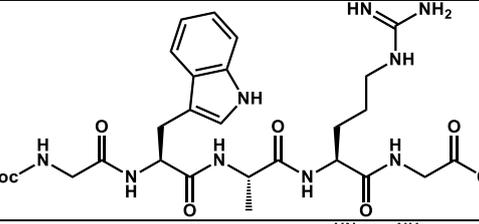
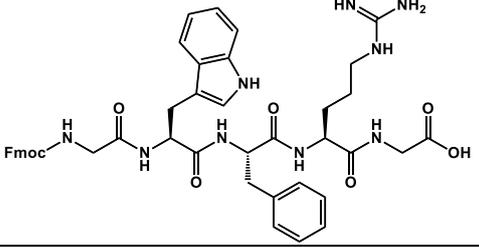
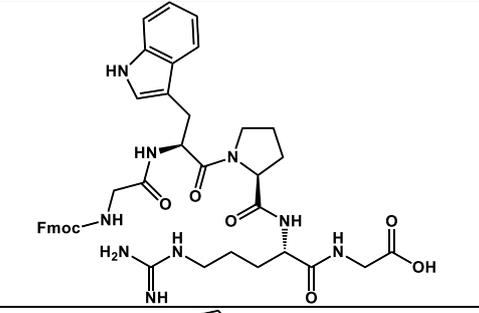
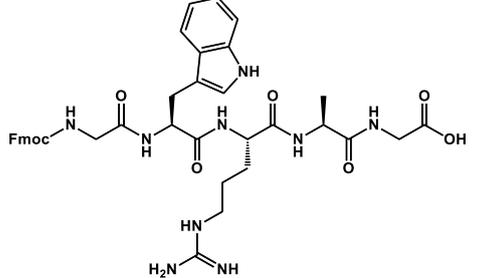
<p>GFRWG DEL1 DEL2</p>		<p>95% Protocol 3</p>	<p>Genscript</p>
<p>GFWRG DEL1 DEL2</p>		<p>60% Protocol 1</p>	<p>Genscript</p>
<p>GLLRG DEL1 DEL2</p>		<p>80% Protocol 1</p>	<p>Genscript</p>
<p>GLRLG DEL1 DEL2</p>		<p>95% Protocol 3</p>	<p>Genscript</p>
<p>GLRPG DEL1 DEL2</p>		<p>77% Protocol 2</p>	<p>Genscript</p>
<p>GLRRG DEL1 DEL2</p>		<p>100% Protocol 3</p>	<p>Genscript</p>
<p>GPFFG DEL1 DEL2</p>		<p>80% Protocol 1</p>	<p>Genscript</p>

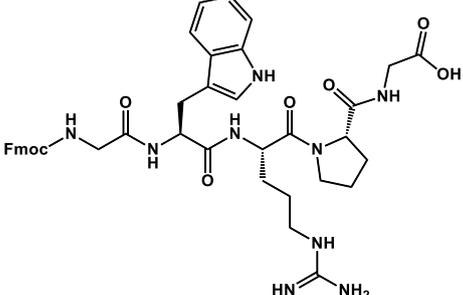
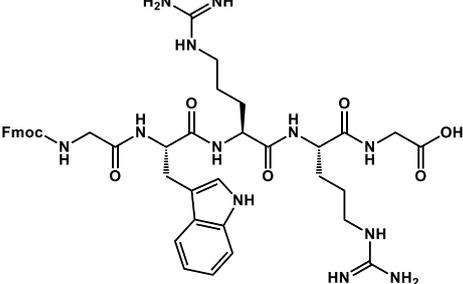
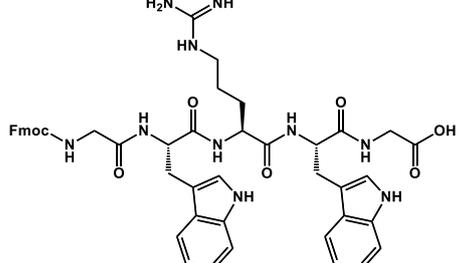
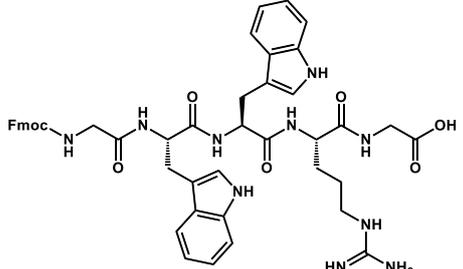
<p>GPFPG DEL2</p>		<p>50% Protocol 1</p>	<p>Genscript</p>
<p>GPRFG DEL1 DEL2</p>		<p>80% Protocol 2</p>	<p>Genscript</p>
<p>GPLRG DEL1 DEL2</p>		<p>70% Protocol 1</p>	<p>Genscript</p>
<p>GPPRG DEL1 DEL2</p>		<p>78% Protocol 2</p>	<p>Genscript</p>
<p>GPPWG DEL2</p>		<p>75% Protocol 1</p>	<p>Genscript</p>
<p>GPRFG DEL1 DEL2</p>		<p>65% Protocol 1</p>	<p>Genscript</p>
<p>GPRPG DEL1 DEL2</p>		<p>90% Protocol 1</p>	<p>Genscript</p>

GPRRG DEL1 DEL2		80% Protocol 1	Genscript
GPRWG DEL1 DEL2		70% Protocol 1	Genscript
GPWPG DEL2		65% Protocol 1	Genscript
GPWRG DEL1 DEL2		70% Protocol 1	Genscript
GRAAG DEL1 DEL2		92% Protocol 2	Genscript
GRARG DEL1 DEL2		95% Protocol 1	Genscript
GRAWG DEL2		70% Protocol 1	Genscript

<p>GRPLG DEL1 DEL2</p>		<p>65% Protocol 1</p>	<p>Genscript</p>
<p>GRPPG DEL1 DEL2</p>		<p>95% Protocol 2</p>	<p>Genscript</p>
<p>GRPRG DEL1 DEL2</p>		<p>95% Protocol 3</p>	<p>Genscript</p>
<p>GRPWG DEL2</p>		<p>80% Protocol 1</p>	<p>Genscript</p>
<p>GRRAG DEL1 DEL2</p>		<p>82% Protocol 2</p>	<p>Genscript</p>
<p>GRRFG DEL1</p>		<p>78% Protocol 2</p>	<p>Genscript</p>

<p>GRRLG DEL1 DEL2</p>		<p>>95% Protocol 3</p>	<p>Genscript</p>
<p>GRRPG DEL1 DEL2</p>		<p>70% Protocol 2</p>	<p>Genscript</p>
<p>GRRWG DEL1 DEL2</p>		<p>90% Protocol 1</p>	<p>Genscript</p>
<p>GRWAG DEL1 DEL2</p>		<p>80% Protocol 1</p>	<p>Genscript</p>
<p>GRWFG DEL1 DEL2</p>		<p>90% Protocol 1</p>	<p>Genscript</p>
<p>GRWPG DEL1 DEL2</p>		<p>80% Protocol 1</p>	<p>Genscript</p>

<p>GRWRG DEL2</p>		<p>85% Protocol 1</p>	<p>Genscript</p>
<p>GRWWG DEL2</p>		<p>65% Protocol 1</p>	<p>Genscript</p>
<p>GWARG DEL2</p>		<p>55% Protocol 1</p>	<p>Genscript</p>
<p>GWFRG DEL1 DEL2</p>		<p>80% Protocol 1</p>	<p>Genscript</p>
<p>GWPRG DEL1 DEL2</p>		<p>80% Protocol 1</p>	<p>Genscript</p>
<p>GWRRG DEL1 DEL2</p>		<p>70% Protocol 1</p>	<p>Genscript</p>

<p>GWRPG DEL1 DEL2</p>		<p>76% Protocol 2</p>	<p>Genscript</p>
<p>GWRRG DEL1 DEL2</p>		<p>80% Protocol 3</p>	<p>Genscript</p>
<p>GWRWG DEL1 DEL2</p>		<p>70% Protocol 1</p>	<p>Genscript</p>
<p>GWWRG DEL2</p>		<p>75% Protocol 1</p>	<p>Genscript</p>

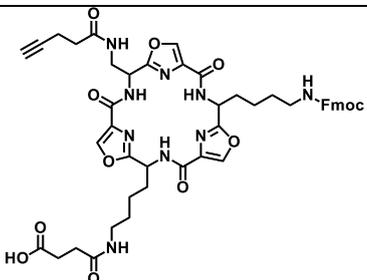
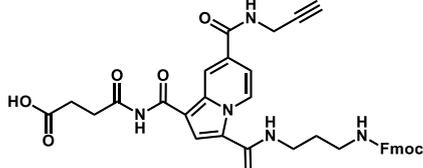
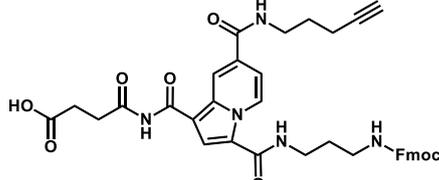
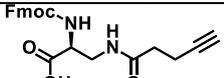
Building blocks [Cycle 2] – DEL1

Table S5 : BB2 – DEL1.

Building block ID	Structure	Acylation yield & reagent	Source
K		>95% Protocol 3	In house (C. Van Wesemael)
KG		>95% Protocol 3	In house (C. Van Wesemael)
GKG		>95% Protocol 3	In house (C. Van Wesemael)
D		>95% Protocol 3	Activotec
DG		>95% Protocol 3	In house (C. Van Wesemael)
GDG		>95% Protocol 3	In house (C. Van Wesemael)
T		>95% Protocol 3	In house (C. Van Wesemael)
P		>95% Protocol 3	Activotec
Ke		>95% Protocol 3	In house (C. Van Wesemael)

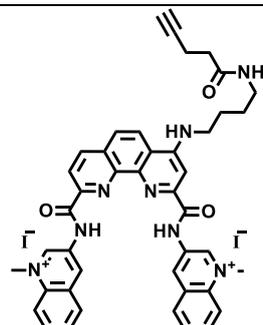
Building blocks [Cycle 2] – DEL2

Table S6 : BB2 – DEL2.

Building block ID	Structure	Acylation yield & reagent	Source
T0		85% Protocol 4	In house (R. Lartia)
ID1		>95% Protocol 3	In house (I. Baussanne)
ID2		65% Protocol 3	In house (I. Baussanne)
ED		>95% Protocol 3	Activotec

Building blocks [Cycle 3] – DEL1

Table S7 : BB3 – DEL1.

Building block ID	Structure	CuAAC yield	Source
PDC3-alk		>95%	In house (M.-P. Teulade-Fichou)

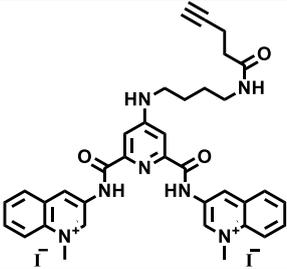
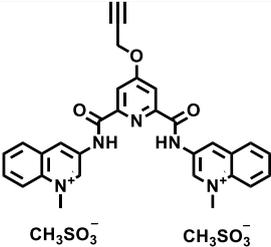
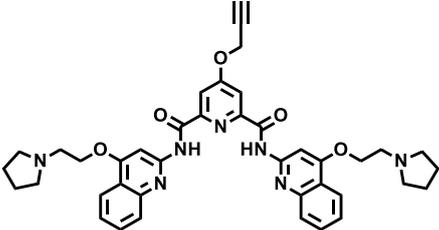
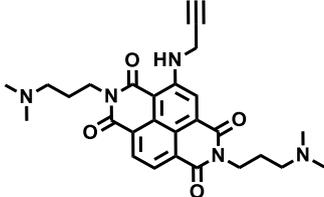
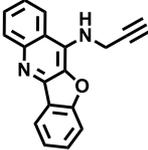
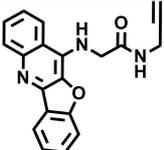
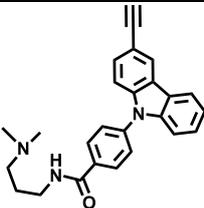
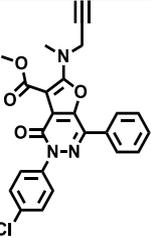
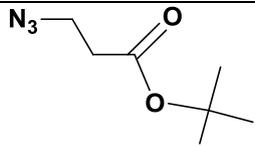
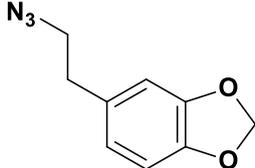
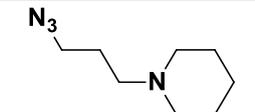
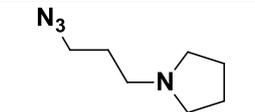
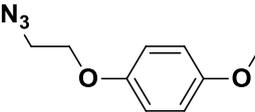
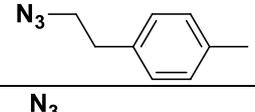
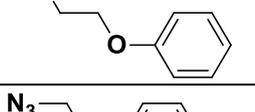
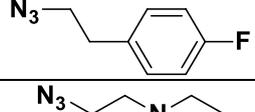
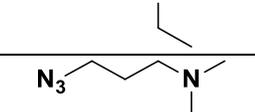
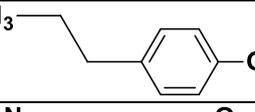
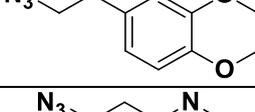
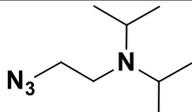
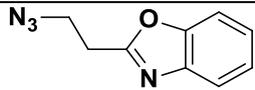
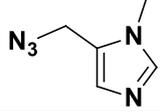
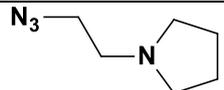
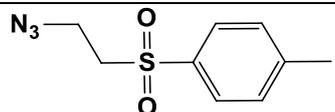
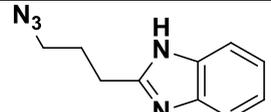
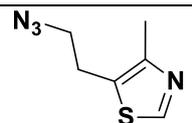
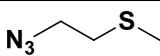
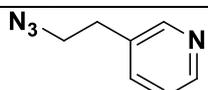
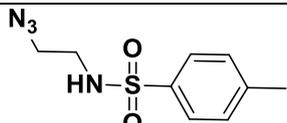
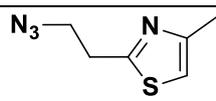
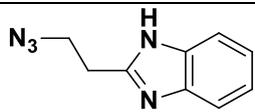
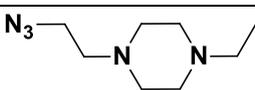
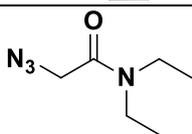
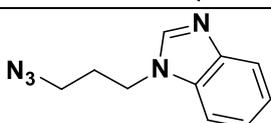
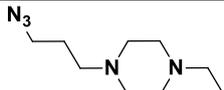
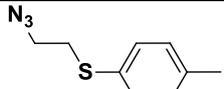
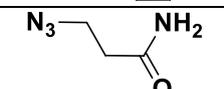
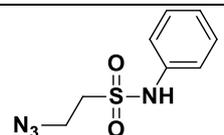
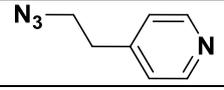
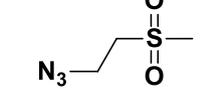
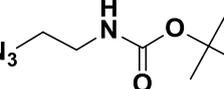
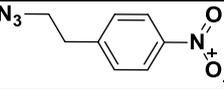
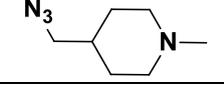
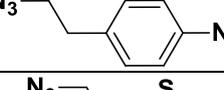
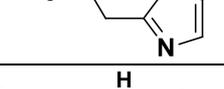
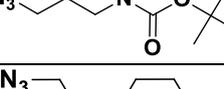
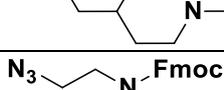
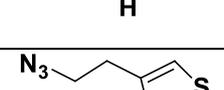
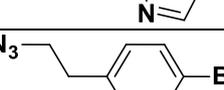
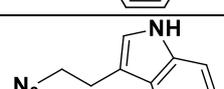
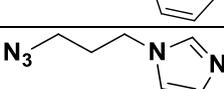
360A-l-alk		>95%	In house (M.-P. Teulade-Fichou)
360A-s-alk		>95%	In house (J.-F. Riou)
PDP-alk		>95%	In house (C. Van Wesemael)
NDI-alk		>95%	In house (F. Doria)
Q-s-alk		>95%	In house (C. Van Wesemael)
Q-l-alk		>95%	In house (C. Van Wesemael)
CZ-alk		>95%	In house (J. Dash)
FP-alk		>95%	In house (M. Giustiniano)
P-alk		>95%	Sigma Aldrich

Table S8 : BB3 – DEL2.

Building block ID	Structure	CuAAC yield	Source
CLK1		92%	In house From H ₂ N-cpd
CLK2		95%	In house From H ₂ N-cpd
CLK3		95%	In house From H ₂ N-cpd
CLK4		>95%	In house From H ₂ N-cpd
CLK5		95%	In house From H ₂ N-cpd
CLK6		90%	In house From H ₂ N-cpd
CLK7		95%	In house From H ₂ N-cpd
CLK8		95%	In house From H ₂ N-cpd
CLK9		>95%	In house From H ₂ N-cpd
CLK10		95%	In house From H ₂ N-cpd
CLK11		80%	In house From H ₂ N-cpd
CLK12		80%	In house From H ₂ N-cpd
CLK13		95%	In house From H ₂ N-cpd
CLK14		93%	In house From H ₂ N-cpd
CLK15		87%	In house From H ₂ N-cpd

CLK16		90%	In house From H ₂ N-cpd
CLK17		95%	In house From H ₂ N-cpd
CLK18		95%	In house From H ₂ N-cpd
CLK19		55%	In house From H ₂ N-cpd
CLK20		95%	In house From H ₂ N-cpd
CLK21		90%	In house From H ₂ N-cpd
CLK22		>95%	In house From H ₂ N-cpd
CLK23		95%	In house From H ₂ N-cpd
CLK24		90%	In house From H ₂ N-cpd
CLK25		90%	In house From H ₂ N-cpd
CLK26		90%	In house From H ₂ N-cpd
CLK27		95%	In house From H ₂ N-cpd
CLK28		95%	In house From H ₂ N-cpd
CLK29		80%	In house From H ₂ N-cpd
CLK30		85%	In house From H ₂ N-cpd
CLK31		90%	In house From H ₂ N-cpd

CLK32		95%	In house From H ₂ N-cpd
CLK33		95%	In house From H ₂ N-cpd
CLK34		90%	In house From H ₂ N-cpd
CLK35		>95%	In house From H ₂ N-cpd
CLK36		95%	In house From H ₂ N-cpd
CLK37		90%	In house From H ₂ N-cpd
CLK38		90%	In house From H ₂ N-cpd
CLK39		90%	In house From H ₂ N-cpd
CLK40		95%	In house From H ₂ N-cpd
CLK41		90%	In house From H ₂ N-cpd
CLK42		95%	In house From H ₂ N-cpd
CLK43		92%	In house From H ₂ N-cpd
CLK44		95%	In house From H ₂ N-cpd
CLK45		95%	In house From H ₂ N-cpd
CLK46		95%	In house From H ₂ N-cpd
CLK47		95%	In house From H ₂ N-cpd
CLK48		90%	In house From H ₂ N-cpd
CLK49		>95%	In house From H ₂ N-cpd
CLK50		>95%	In house From H ₂ N-cpd

Sequence of DNA tags and corresponding building block ID.

DNA Tags [Cycle 1 & 4] – DEL1 & DEL2

Table S9 : DNA Tag sequences of BB1 & BB4 – DEL1 & DEL2.

Tag Sequence (5'→3' bottom strand)		Building block ID	Source
DEL1	T1-14 // 5'-TTCGGCTCA	GAARG	IDT
DEL2	T4-12 // 5'-GTTAGCGCA		
DEL1	T1-14 // 5'-TTCGGCTCA		
DEL2	T4-12 // 5' – GTTAGCGCA		
DEL1	T1-15 // 5'-CTACCGACA	GAAWG	IDT
DEL2	T4-13 // 5'-GTCGGCTCA		
DEL1	T1-15 // 5'-CTACCGACA		
DEL2	T4-13 // 5'-GTCGGCTCA		
DEL1	---	GARAG	IDT
DEL2	---		
DEL1	T1-56 // 5'-GCTTCGGCA		
DEL2	T4-52 // 5'-GACTGGACA		
DEL1	T1-34 // 5'-CGCCTAACA	GARRG	IDT
DEL2	---		
DEL1	---		
DEL2	---		
DEL1	T1-26 // 5'-GGTGTCCA	GARWG	IDT
DEL2	T4-24 // 5'-ACCTGACCA		
DEL1	T1-26 // 5'-GGTGTCCA		
DEL2	T4-24 // 5'-ACCTGACCA		
DEL1	T1-25 // 5'-AACAGGCCA	GAWAG	IDT
DEL2	T4-23 // 5'-GCTCGCACA		
DEL1	T1-25 // 5'-AACAGGCCA		
DEL2	T4-23 // 5'-GCTCGCACA		
DEL1	T1-19 // 5'-CCTCGCACA	GAWRG	IDT
DEL2	T4-17 // 5'-ACGCACTCA		
DEL1	T1-19 // 5'-CCTCGCACA		
DEL2	T4-17 // 5'-ACGCACTCA		
DEL1	T1-39 // 5'-CGGCTACCA	GFFRG	IDT
DEL2	T4-35 // 5'-CACAGGCCA		
DEL1	T1-39 // 5'-CGGCTACCA		
DEL2	T4-35 // 5'-CACAGGCCA		
DEL1	T1-48 // 5'-CCAGAGTCA	GFPRG	IDT
DEL2	T4-44 // 5'-TGCCGTCCA		
DEL1	T1-48 // 5'-CCAGAGTCA		
DEL2	T4-44 // 5'-TGCCGTCCA		
DEL1	---	GFRFG	IDT
DEL2	T4-9 // 5'-GCTTCAGCA		
DEL1	T1-11 // 5'-CGCCAAACA		
DEL2	T4-9 // 5'-GCTTCAGCA		
DEL1	---	GFRPG	IDT
DEL2	T4-36 // 5'-ACAGGAGCA		
DEL1	T1-40 // 5'-TACAGGCCA		
DEL2	T4-36 // 5'-ACAGGAGCA		
DEL1	T1-33 // 5'-GGTCCACA	GFRRG	IDT
DEL2	T4-31 // 5'-CGTCCACA		
DEL1	T1-33 // 5'-GGTCCACA		
DEL2	T4-31 // 5'-CGTCCACA		

DEL1	T1-5 // 5'-GCTAGTCCA	GFRWG	IDT
DEL2	T4-4 // 5'-CCTTCAGCA		
DEL1	T1-5 // 5'-GCTAGTCCA	GFRWG	IDT
DEL2	T4-4 // 5'-CCTTCAGCA		
DEL1	---	GFWRG	IDT
DEL2	---		
DEL1	T1-63 // 5'-CCTTTCGCA	GFWRG	IDT
DEL2	T4-59 // 5'-CCGATAGCA		
DEL1	T1-49 // 5'-TTCCAGCCA	GLLRG	IDT
DEL2	T4-45 // 5'-GCGAAAGCA		
DEL1	T1-49 // 5'-TTCCAGCCA	GLLRG	IDT
DEL2	T4-45 // 5'-GCGAAAGCA		
DEL1	T1-6 // 5'-CTTAGCGCA	GLRLG	IDT
DEL2	T4-5 // 5'-GCTAGTCCA		
DEL1	T1-6 // 5'-CTTAGCGCA	GLRLG	IDT
DEL2	T4-5 // 5'-GCTAGTCCA		
DEL1	T1-12 // 5'-GTTAGCGCA	GLRPG	IDT
DEL2	T4-10 // 5'-ACTCGCACA		
DEL1	T1-12 // 5'-GTTAGCGCA	GLRPG	IDT
DEL2	T4-10 // 5'-ACTCGCACA		
DEL1	T1-2 // 5'-CCTAGTCCA	GLRRG	IDT
DEL2	T4-2 // 5'-CCTAGTCCA		
DEL1	T1-2 // 5'-CCTAGTCCA	GLRRG	IDT
DEL2	T4-60 // 5'-CCTTTCGCA		
DEL1	T1-24 // 5'-ACCTGACCA	GPFFG	IDT
DEL2	T4-22 // 5'-ACCGGAACA		
DEL1	T1-24 // 5'-ACCTGACCA	GPFFG	IDT
DEL2	T4-22 // 5'-ACCGGAACA		
DEL1	---	GPFFG	IDT
DEL2	---		
DEL1	T1-61 // 5'-CGCCAGACA	GPFFG	IDT
DEL2	T4-57 // 5'-CGGAACACA		
DEL1	T1-32 // 5'-TCGCACTCA	GPFRG	IDT
DEL2	T4-30 // 5'-CGGACAACA		
DEL1	T1-32 // 5'-TCGCACTCA	GPFRG	IDT
DEL2	T4-30 // 5'-CGGACAACA		
DEL1	T1-31 // 5'-CGTTCCACA	GPLRG	IDT
DEL2	T4-29 // 5'-CCGCACTCA		
DEL1	T1-31 // 5'-CGTTCCACA	GPLRG	IDT
DEL2	T4-29 // 5'-CCGCACTCA		
DEL1	T1-30 // 5'-CGGACAACA	GPPRG	IDT
DEL2	T4-28 // 5'-TCTCGCACA		
DEL1	T1-30 // 5'-CGGACAACA	GPPRG	IDT
DEL2	T4-28 // 5'-TCTCGCACA		
DEL1	---	GPPWG	IDT
DEL2	---		
DEL1	T1-59 // 5'-CCGATAGCA	GPPWG	IDT
DEL2	T4-55 // 5'-CCTTCGGCA		
DEL1	T1-38 // 5'-TCCTGACCA	GPRFG	IDT
DEL2	T4-34 // 5'-CGCCTAACA		
DEL1	T1-38 // 5'-TCCTGACCA	GPRFG	IDT
DEL2	T4-34 // 5'-CGCCTAACA		
DEL1	T1-17 // 5'-ACGCACTCA	GPRPG	IDT
DEL2	T4-15 // 5'-CTACCGACA		
DEL1	T1-17 // 5'-ACGCACTCA	GPRPG	IDT
DEL2	T4-15 // 5'-CTACCGACA		

		T4-15 // 5'-CTACCGACA		
DEL1		T1-22 // 5'-ACCGGAACA	GPRRG	IDT
DEL2		T4-20 // 5'-CGTGTTC		
DEL1		T1-22 // 5'-ACCGGAACA	GPRWG	IDT
DEL2		T4-20 // 5'-CGTGTTC		
DEL1		T1-27 // 5'-GGACAAGCA	GPRWG	IDT
DEL2		T4-25 // 5'-AACAGGCCA		
DEL1		T1-27 // 5'-GGACAAGCA	GPWPG	IDT
DEL2		T4-25 // 5'-AACAGGCCA		
DEL1		---	GPWPG	IDT
DEL2		---		
DEL1		T1-58 // 5'-CGCCAGACA	GPWRG	IDT
DEL2		T4-54 // 5'-TCGCTCTCA		
DEL1		T1-37 // 5'-GCCTGACCA	GPWRG	IDT
DEL2		T4-33 // 5'-GGTTCCACA		
DEL1		T1-37 // 5'-GCCTGACCA	GRAAG	IDT
DEL2		T4-33 // 5'-GGTTCCACA		
DEL1		T1-7 // 5'-CTCGGCTCA	GRAAG	IDT
DEL2		T4-6 // 5'-CTTAGCGCA		
DEL1		T1-7 // 5'-CTCGGCTCA	GRARG	IDT
DEL2		T4-6 // 5'-CTTAGCGCA		
DEL1		T1-4 // 5'-CCTTCAGCA	GRARG	IDT
DEL2		T4-3 // 5'-CATACGGCA		
DEL1		T1-4 // 5'-CCTTCAGCA	GRAWG	IDT
DEL2		T4-3 // 5'-CATACGGCA		
DEL1		---	GRAWG	IDT
DEL2		---		
DEL1		T1-53 // 5'-CCGCTCTCA	GRFFG	IDT
DEL2		T4-49 // 5'-TTCCAGCCA		
DEL1		T1-29 // 5'-CCGCACTCA	GRFFG	IDT
DEL2		T4-27 // 5'-GGACAAGCA		
DEL1		T1-29 // 5'-CCGCACTCA	GRFPG	IDT
DEL2		T4-27 // 5'-GGACAAGCA		
DEL1		---	GRFPG	IDT
DEL2		---		
DEL1		T1-60 // 5'-CCTTTCGCA	GRFRG	IDT
DEL2		T4-56 // 5'-GCTTCGGCA		
DEL1		T1-16 // 5'-CGGCAACCA	GRFRG	IDT
DEL2		T4-14 // 5'-TTCGGCTCA		
DEL1		T1-16 // 5'-CGGCAACCA	GRFWG	IDT
DEL2		T4-14 // 5'-TTCGGCTCA		
DEL1		---	GRFWG	IDT
DEL2		---		
DEL1		T1-62 // 5'-CCGATAGCA	GRLLG	IDT
DEL2		T4-58 // 5'-CGCCAGACA		
DEL1		T1-9 // 5'-GCTTCAGCA	GRLLG	IDT
DEL2		T4-8 // 5'-GATACGGCA		
DEL1		T1-9 // 5'-GCTTCAGCA	GRLPG	IDT
DEL2		T4-8 // 5'-GATACGGCA		
DEL1		T1-42 // 5'-CCTTACGCA	GRLPG	IDT
DEL2		T4-38 // 5'-TCCTGACCA		
DEL1		T1-42 // 5'-CCTTACGCA	GRLRG	IDT
DEL2		T4-38 // 5'-TCCTGACCA		
DEL1		T1-3 // 5'-CATACGGCA	GRLRG	IDT
DEL2		---		

DEL2	---		
DEL1	---		
DEL2	T4-40 // 5'-TACAGGCCA	GRPLG	IDT
DEL1	T1-44 // 5'-TGCCGTCCA		
DEL2	T4-40 // 5'-TACAGGCCA		
DEL1	T1-21 // 5'-CGACAAGCA	GRPPG	IDT
DEL2	T4-19 // 5'-CCTCGCACA		
DEL1	T1-21 // 5'-CGACAAGCA		
DEL2	T4-19 // 5'-CCTCGCACA		
DEL1	T1-13 // 5'-GTCGGCTCA	GRPRG	IDT
DEL2	T4-11 // 5'-CGCCAAACA		
DEL1	T1-13 // 5'-GTCGGCTCA		
DEL2	T4-61 // 5'-CGCCAGACA		
DEL1	---		
DEL2	---		
DEL1	T1-57 // 5'-CGGAACACA	GRPWG	IDT
DEL2	T4-53 // 5'-CCGCTCTCA		
DEL1	T1-8 // 5'-GATACGGCA	GRRAG	IDT
DEL2	T4-7 // 5'-CTCGGCTCA		
DEL1	T1-8 // 5'-GATACGGCA		
DEL2	T4-7 // 5'-CTCGGCTCA		
DEL1	T1-10 // 5'-ACTCGCACA	GRRFG	IDT
DEL2	---		
DEL1	---		
DEL2	---		
DEL1	T1-1 // 5'-GAGTGACCA	GRRLG	IDT
DEL2	T4-1 // 5'-GAGTGACCA		
DEL1	T1-1 // 5'-GAGTGACCA		
DEL2	T4-1 // 5'-GAGTGACCA		
DEL1	T1-23 // 5'-GCTCGCACA	GRRPG	IDT
DEL2	T4-21 // 5'-CGACAAGCA		
DEL1	T1-23 // 5'-GCTCGCACA		
DEL2	T4-62 // 5'-CCGATAGCA		
DEL1	T1-43 // 5'-ACGCTCTCA	GRRWG	IDT
DEL2	T4-39 // 5'-CGGCTACCA		
DEL1	T1-43 // 5'-ACGCTCTCA		
DEL2	T4-39 // 5'-CGGCTACCA		
DEL1	T1-47 // 5'-CCATAGCCA	GRWAG	IDT
DEL2	T4-43 // 5'-ACGCTCTCA		
DEL1	T1-47 // 5'-CCATAGCCA		
DEL2	T4-43 // 5'-ACGCTCTCA		
DEL1	T1-50 // 5'-ACTTCGGCA	GRWFG	IDT
DEL2	T4-46 // 5'-GCTTACGCA		
DEL1	T1-50 // 5'-ACTTCGGCA		
DEL2	T4-46 // 5'-GCTTACGCA		
DEL1	T1-18 // 5'-TGGCAACCA	GRWPG	IDT
DEL2	T4-16 // 5'-CGGCAACCA		
DEL1	T1-18 // 5'-TGGCAACCA		
DEL2	T4-16 // 5'-CGGCAACCA		
DEL1	---		
DEL2	---		
DEL1	T1-51 // 5'-TCAGGAGC	GRWRG	IDT
DEL2	T4-47 // 5'-CCATAGCCA		
DEL1	---		

DEL2	T1-52 // 5'-GACTGGAC	GRWWG	IDT
	T4-48 // 5'-CCAGAGTCA		
DEL1	---		

DEL2	T1-54 // 5'-TCGCTCTC	GWARG	IDT
	T4-50 // 5'-ACTTCGGCA		
DEL1	T1-20 // 5'-CGTGTCCA	GWFRG	IDT
	T4-18 // 5'-TGGCAACCA		
DEL2	T1-20 // 5'-CGTGTCCA	GWPRG	IDT
	T4-18 // 5'-TGGCAACCA		
DEL1	T1-46 // 5'-GCTTACGCA	GWRAG	IDT
	T4-42 // 5'-CCTTACGCA		
DEL2	T1-46 // 5'-GCTTACGCA	GWRAG	IDT
	T4-42 // 5'-CCTTACGCA		
DEL1	T1-45 // 5'-GCGAAAGCA	GWRAG	IDT
	T4-41 // 5'-CCGAAAGCA		
DEL2	T1-45 // 5'-GCGAAAGCA	GWRAG	IDT
	T4-41 // 5'-CCGAAAGCA		
DEL1	T1-36 // 5'-ACAGGAGCA	GWRPG	IDT
	T4-32 // 5'-TCGCACTCA		
DEL2	T1-36 // 5'-ACAGGAGCA	GWRPG	IDT
	T4-32 // 5'-TCGCACTCA		
DEL1	T1-28 // 5'-TCTCGCACA	GWRRG	IDT
	T4-26 // 5'-GGTGTCCA		
DEL2	T1-28 // 5'-TCTCGCACA	GWRRG	IDT
	T4-26 // 5'-GGTGTCCA		
DEL1	T1-41 // 5'-CCGAAAGCA	GWRWG	IDT
	T4-37 // 5'-GCCTGACCA		
DEL2	T1-41 // 5'-CCGAAAGCA	GWRWG	IDT
	T4-37 // 5'-GCCTGACCA		
DEL1	---	GWWRG	IDT

DEL2	T1-55 // 5'-CCTTCGGC	GWWRG	IDT
	T4-51 // 5'-TCAGGAGCA		

DNA Tags [Cycle 2] – DEL1

Table S10 : DNA Tag sequences of BB2 – DEL1.

Tag Sequence (5'→3' bottom strand)	Building block ID	Source
T2-1 // 5'- AGGCAACCT	K	IDT
T2-2 // 5'- CGTCCTGCT	KG	IDT
T2-3 // 5'- CACGTCTCT	GKG	IDT
T2-4 // 5'- GGTCCTGCT	D	IDT
T2-5 // 5'- TGTCCTGCT	DG	IDT
T2-6 // 5'- CAGAGTGCT	GDG	IDT
T2-7 // 5'- AGCCGTCCT	T	IDT
T2-8 // 5'- GAGAGTGCT	P	IDT
T2-9 // 5'- AGGCTACCT	Ke	IDT

DNA Tags [Cycle 2] – DEL2

Table S11 : DNA Tag sequences of BB2 – DEL2.

Tag Sequence (5'→3' bottom strand)	Building block ID	Source
T2-38 // 5'- GCGAAGGCT	DE	IDT
T2-34 // 5'- AGGCCATCT	I1	IDT
T2-35 // 5'- AGACTGGCT	I2	IDT
T2-30 // 5'- ACCTGGCCT	TO	IDT

DNA Tags [Cycle 3] – DEL1

Table S12 : DNA Tag sequences of BB3 – DEL1.

Tag Sequence (5'→3' bottom strand)	Building block ID	Source
T3-2 // 5'- GCAGGAGAC	360A-l	IDT
T3-10 // 5'- GTAGGCAAC	360A-s	IDT
T3-8 // 5'- GGTTTCAGAC	CZ	IDT
T3-9 // 5'- GTCCAGAAC	FP	IDT
T3-4 // 5'- GTTCCTGAC	NDI	IDT
T3-1 // 5'- GTAGCTGAC	PDC3	IDT
T3-3 // 5'- AGGCAGCAC	PDP	IDT
T3-5 // 5'- GCGTGTGAC	PP	IDT
T3-7 // 5'- CGTTTCAGAC	Q-l	IDT
T3-6 // 5' – AGGCTGCAC	Q-s	IDT

DNA Tags [Cycle 3] – DEL2

Table S13 : DNA Tag sequences of BB3 – DEL2.

Tag Sequence (5'→3' bottom strand)	Building block ID	Source
T3-1 // 5'- GTAGCTGAC	CLK1	IDT
T3-2 // 5'- GCAGGAGAC	CLK2	IDT
T3-3 // 5'- AGGCAGCAC	CLK3	IDT
T3-4 // 5'- GTTCCTGAC	CLK4	IDT
T3-5 // 5'- GCGTGTGAC	CLK5	IDT
T3-6 // 5'- AGGCTGCAC	CLK6	IDT
T3-7 // 5'- CGTTTCAGAC	CLK7	IDT
T3-8 // 5'- GGTTTCAGAC	CLK8	IDT
T3-9 // 5'- GTCCAGAAC	CLK9	IDT
T3-10 // 5'- GTAGGCAAC	CLK10	IDT
T3-11 // 5'- GTATGCCAC	CLK11	IDT
T3-12 // 5'- GTCACCGAC	CLK12	IDT
T3-13 // 5'- GTCGTTGAC	CLK13	IDT
T3-14 // 5'- AGAGGAGAC	CLK14	IDT
T3-15 // 5'- GTCCTGAAC	CLK15	IDT
T3-16 // 5'- CGCCTTTAC	CLK16	IDT
T3-17 // 5'- CGAGGAGAC	CLK17	IDT
T3-18 // 5'- GCATCTGAC	CLK18	IDT
T3-19 // 5'- TTGCTGCAC	CLK19	IDT
T3-20 // 5'- ACTCCACAC	CLK20	IDT

T3-21 // 5'- CTA CTGGAC	CLK21	IDT
T3-22 // 5'- GTGCCATAC	CLK22	IDT
T3-23 // 5'- ACAGTCCAC	CLK23	IDT
T3-24 // 5'- AGTTCGGAC	CLK24	IDT
T3-25 // 5'- TGAGGAGAC	CLK25	IDT
T3-26 // 5'- AGCAGTGAC	CLK26	IDT
T3-27 // 5'- GCCTGCTAC	CLK27	IDT
T3-28 // 5'- TCCTGCTAC	CLK28	IDT
T3-29 // 5'- TGGCTCTAC	CLK29	IDT
T3-30 // 5'- CAACGACAC	CLK30	IDT
T3-31 // 5'- CGTTCGGAC	CLK31	IDT
T3-32 // 5'- CGCAGTGAC	CLK32	IDT
T3-33 // 5'- GCTCCACAC	CLK33	IDT
T3-34 // 5'- TCTCCACAC	CLK34	IDT
T3-35 // 5'- GCAGTCCAC	CLK35	IDT
T3-36 // 5'- GGTTTCGGAC	CLK36	IDT
T3-37 // 5'- GGCAGTGAC	CLK37	IDT
T3-38 // 5'- CGTTTCGAC	CLK38	IDT
T3-39 // 5'- TCAGTCCAC	CLK39	IDT
T3-40 // 5'- TGTTTCGGAC	CLK40	IDT
T3-41 // 5'- GGTTTCGAC	CLK41	IDT
T3-42 // 5'- AAGCCAGAC	CLK42	IDT
T3-43 // 5'- AGGTGTGAC	CLK43	IDT
T3-44 // 5'- GGTCGTTAC	CLK44	IDT
T3-45 // 5'- GCTGGATAC	CLK45	IDT
T3-46 // 5'- CTTCGCAAC	CLK46	IDT
T3-47 // 5'- GTTCGCAAC	CLK47	IDT
T3-48 // 5'- CAGCCAGAC	CLK48	IDT
T3-49 // 5'- CGGTGTGAC	CLK49	IDT
T3-50 // 5'- CTGCACTAC	CLK50	IDT

Bibliographie.

- (1) Dahm, R. Friedrich Miescher and the Discovery of DNA. *Dev. Biol.* **2005**, 278 (2), 274–288. <https://doi.org/10.1016/j.ydbio.2004.11.028>.
- (2) Levene, P. A.; London, E. S. THE STRUCTURE OF THYMONUCLEIC ACID. *J. Biol. Chem.* **1929**, 83 (3), 793–802. [https://doi.org/10.1016/S0021-9258\(18\)77108-1](https://doi.org/10.1016/S0021-9258(18)77108-1).
- (3) Levene, P. A.; Mori, T. RIBODESOSE AND XYLODESOSE AND THEIR BEARING ON THE STRUCTURE OF THYMINOSE. *J. Biol. Chem.* **1929**, 83 (3), 803–816. [https://doi.org/10.1016/S0021-9258\(18\)77109-3](https://doi.org/10.1016/S0021-9258(18)77109-3).
- (4) Chargaff, E. Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymatic Degradation. *Experientia* **1950**, 6 (6), 201–209. <https://doi.org/10.1007/BF02173653>.
- (5) Zamenhof, S.; Brawerman, G.; Chargaff, E. On the Desoxypentose Nucleic Acids from Several Microorganisms. *Biochim. Biophys. Acta* **1952**, 9, 402–405. [https://doi.org/10.1016/0006-3002\(52\)90184-4](https://doi.org/10.1016/0006-3002(52)90184-4).
- (6) Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature* **1953**, 171 (4356), 738–740. <https://doi.org/10.1038/171738a0>.
- (7) Franklin, R. E.; Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **1953**, 171 (4356), 740–741. <https://doi.org/10.1038/171740a0>.
- (8) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, 171 (4356), 737–738. <https://doi.org/10.1038/171737a0>.
- (9) Wood, B. R. The Importance of Hydration and DNA Conformation in Interpreting Infrared Spectra of Cells and Tissues. *Chem. Soc. Rev.* **2016**, 45 (7), 1980–1998. <https://doi.org/10.1039/C5CS00511F>.
- (10) Macgregor, R. B.; Poon, G. M. K. The DNA Double Helix Fifty Years On. *Comput. Biol. Chem.* **2003**, 27 (4–5), 461–467. <https://doi.org/10.1016/j.compbiolchem.2003.08.001>.
- (11) Wang, A. H.-J.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. Molecular Structure of a Left-Handed Double Helical DNA Fragment at Atomic Resolution. *Nature* **1979**, 282 (5740), 680–686. <https://doi.org/10.1038/282680a0>.
- (12) Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Crystal Structure Analysis of a Complete Turn of B-DNA. *Nature* **1980**, 287 (5784), 755–758. <https://doi.org/10.1038/287755a0>.
- (13) Takezawa, Y.; Shionoya, M. Supramolecular DNA Three-Way Junction Motifs With a Bridging Metal Center. *Front. Chem.* **2020**, 7, 925. <https://doi.org/10.3389/fchem.2019.00925>.
- (14) Gellert, M.; Lipsett, M. N.; Davies, D. R. Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci.* **1962**, 48 (12), 2013–2018.
- (15) Miescher, F. Hoppe-Seyler, S. *Med-Chem. Untersuch.* **1871**, 441, 502.
- (16) Hammarsten, O. Zur Kenntniss Der Nucleoproteide. *Z. Phys. Chem.* **1895**, 19, 19–37.
- (17) Levene, P. A.; Jacobs, W. A. Über Guanylsäure. *Ber. Dtsch. Chem. Ges.* **1909**, 42, 2469–2473. <https://doi.org/10.1002/Cber.190904202147>.
- (18) Bang, I. Untersuchungen Über Die Guanylsäure, *Biochem. Z.* **1910**, 26, 293–311.
- (19) 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.
- (20) Davis, J. T. G-Quartets 40 Years Later: From 5'-GMP to Molecular Biology and Supramolecular Chemistry. *Angew. Chem. Int. Ed.* **2004**, 43 (6), 668–698. <https://doi.org/10.1002/anie.200300589>.
- (21) Pinnavaia, T. J.; Miles, H. T.; Becker, E. D. Self-Assembled 5'-Guanosine Monophosphate_nuclear Magnetic Resonance Evidence Ofr a Regular Ordered Structure and Slow Chemical Exchange. *J. Am. Chem. Soc.* **1975**, 97 (24), 7198–7200.
- (22) Pinnavaia, T. J.; Marshall, C. L.; Mettler, C. M.; Fisk, C. L.; Miles, H. T.; Becker, E. D. Alkali Metal Ion Specificity in the Solution Ordering of a Nucleotide, 5'-Guanosine Monophosphate. *J. Am. Chem. Soc.* **1978**, 100 (11), 3625–3627. <https://doi.org/10.1021/ja00479a070>.
- (23) Hardin, C. C.; Watson, T.; Corregan, M.; Bailey, C. Cation-Dependent Transition between the Quadruplex and Watson-Crick Hairpin Forms of d(CGCG3GCG). *Biochemistry* **1992**, 31 (3), 833–841. <https://doi.org/10.1021/bi00118a028>.
- (24) Karsisiotis, A. I.; Hessari, N. M.; Novellino, E.; Spada, G. P.; Randazzo, A.; Webba da Silva, M. Topological Characterization of Nucleic Acid G-Quadruplexes by UV Absorption and Circular Dichroism. *Angew. Chem. Int. Ed.* **2011**, 50 (45), 10645–10648. <https://doi.org/10.1002/anie.201105193>.
- (25) Webba da Silva, M.; Trajkovski, M.; Sannohe, Y.; Ma'ani Hessari, N.; Sugiyama, H.; Plavec, J. Design of a G-Quadruplex Topology through Glycosidic Bond Angles. *Angew. Chem. Int. Ed.* **2009**, 48 (48), 9167–9170. <https://doi.org/10.1002/anie.200902454>.
- (26) Guédin, A.; De Cian, A.; Gros, J.; Lacroix, L.; Mergny, J.-L. Sequence Effects in Single-Base Loops for Quadruplexes. *Biochimie* **2008**, 90 (5), 686–696. <https://doi.org/10.1016/j.biochi.2008.01.009>.
- (27) Wang, K.-B.; Dickerhoff, J.; Wu, G.; Yang, D. PDGFR- β Promoter Forms a Vacancy G-Quadruplex That Can Be Filled in by DGMP: Solution Structure and Molecular Recognition of Guanine Metabolites and Drugs. *J. Am. Chem. Soc.* **2020**, 142 (11), 5204–5211. <https://doi.org/10.1021/jacs.9b12770>.
- (28) Mukundan, V. T.; Phan, A. T. Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **2013**, 135 (13), 5017–5028. <https://doi.org/10.1021/ja310251r>.
- (29) Ngoc Nguyen, T. Q.; Lim, K. W.; Phan, A. T. Duplex Formation in a G-Quadruplex Bulge. *Nucleic Acids Res.* **2020**, 48 (18), 10567–10575. <https://doi.org/10.1093/nar/gkaa738>.
- (30) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, 5 (3), 182–186. <https://doi.org/10.1038/nchem.1548>.
- (31) Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and Selective Chemical Targeting of RNA G-Quadruplex Structures in the Cytoplasm of Human Cells. *Nat. Chem.* **2014**, 6 (1), 75–80. <https://doi.org/10.1038/nchem.1805>.
- (32) Qiu, J.; Chen, S.; Su, L.; Liu, J.; Xiao, N.; Ou, T.-M.; Tan, J.-H.; Gu, L.-Q.; Huang, Z.-S.; Li, D. Cellular Nucleic Acid Binding Protein Suppresses Tumor Cell Metastasis and Induces Tumor Cell Death by Downregulating Heterogeneous Ribonucleoprotein K in Fibrosarcoma

- Cells. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2014**, *1840* (7), 2244–2252. <https://doi.org/10.1016/j.bbagen.2014.02.025>.
- (33) Huppert, J. L. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Res.* **2005**, *33* (9), 2908–2916. <https://doi.org/10.1093/nar/gki609>.
- (34) Eddy, J.; Maizels, N. Gene Function Correlates with Potential for G4 DNA Formation in the Human Genome. *Nucleic Acids Res.* **2006**, *34* (14), 3887–3896. <https://doi.org/10.1093/nar/gkl529>.
- (35) Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-Evaluation of G-Quadruplex Propensity with G4Hunter. *Nucleic Acids Res.* **2016**, *44* (4), 1746–1759. <https://doi.org/10.1093/nar/gkw006>.
- (36) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proc. Natl. Acad. Sci.* **2002**, *99* (18), 11593–11598. <https://doi.org/10.1073/pnas.182256799>.
- (37) Lee, S. C.; Zhang, J.; Strom, J.; Yang, D.; Dinh, T. N.; Kappeler, K.; Chen, Q. M. G-Quadruplex in the NRF2 mRNA 5' Untranslated Region Regulates *De Novo* NRF2 Protein Translation under Oxidative Stress. *Mol. Cell. Biol.* **2017**, *37* (1). <https://doi.org/10.1128/MCB.00122-16>.
- (38) Endoh, T.; Kawasaki, Y.; Sugimoto, N. Stability of RNA Quadruplex in Open Reading Frame Determines Proteolysis of Human Estrogen Receptor α . *Nucleic Acids Res.* **2013**, *41* (12), 6222–6231. <https://doi.org/10.1093/nar/gkt286>.
- (39) 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* (1), 1796. <https://doi.org/10.1038/ncomms2792>.
- (40) Dickerhoff, J.; Onel, B.; Chen, L.; Chen, Y.; Yang, D. Solution Structure of a MYC Promoter G-Quadruplex with 1:6:1 Loop Length. *ACS Omega* **2019**, *4* (2), 2533–2539. <https://doi.org/10.1021/acsomega.8b03580>.
- (41) Bertin, A.; Mangenot, S. Structure et dynamique de la particule cœur de nucléosome. *médecine/sciences* **2008**, *24* (8–9), 715–719. <https://doi.org/10.1051/medsci/20082489715>.
- (42) Lomberk, G.; Wallrath, L.; Urrutia, R. The Heterochromatin Protein I Family. *Genome Biol.* **2006**, *7* (7), 228. <https://doi.org/10.1186/gb-2006-7-7-228>.
- (43) Peterson, C. L. The SMC Family: Novel Motor Proteins for Chromosome Condensation? *Cell* **1994**, *79* (3), 389–392. [https://doi.org/10.1016/0092-8674\(94\)90247-X](https://doi.org/10.1016/0092-8674(94)90247-X).
- (44) McClintock, B. The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis. *Proc Natl Acad Sci U S A.* **1939**, *25* (8), 405–416. <https://doi.org/10.1073/Pnas.25.8.405>.
- (45) Blackburn - 2001 - Switching and Signaling at the Telomere.Pdf.
- (46) Kar, A.; Jones, N.; Arat, N. Ö.; Fishel, R.; Griffith, J. D. Long Repeating (TTAGGG)_n Single-Stranded DNA Self-Condenses into Compact Beaded Filaments Stabilized by G-Quadruplex Formation. *J. Biol. Chem.* **2018**, *293* (24), 9473–9485. <https://doi.org/10.1074/jbc.RA118.002158>.
- (47) Kim, N.; Piatyszek, M.; Prowse, K.; Harley, C.; West, M.; Ho, P.; Coviello, G.; Wright, W.; Weinrich, S.; Shay, J. Specific Association of Human Telomerase Activity with Immortal Cells and Cancer. *Science* **1994**, *266* (5193), 2011–2015. <https://doi.org/10.1126/science.7605428>.
- (48) Shay, J. W.; Bacchetti, S. A Survey of Telomerase Activity in Human Cancer. *Eur. J. Cancer* **1997**, *33* (5), 787–791. [https://doi.org/10.1016/S0959-8049\(97\)00062-2](https://doi.org/10.1016/S0959-8049(97)00062-2).
- (49) Wright, W. E.; Piatyszek, M. A.; Rainey, W. E.; Byrd, W.; Shay, J. W. Telomerase Activity in Human Germline and Embryonic Tissues and Cells. *1996* *18* (173–179), 7.
- (50) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. **2006**, 14.
- (51) Neidle, S.; Parkinson, G. N. The Structure of Telomeric DNA. *Curr. Opin. Struct. Biol.* **2003**, *13* (3), 275–283. [https://doi.org/10.1016/S0959-440X\(03\)00072-1](https://doi.org/10.1016/S0959-440X(03)00072-1).
- (52) Wang, Y.; Patel, D. J. Guanine Residues in d(T2AG3) and d(T2G4) Form Parallel-Stranded Potassium Cation Stabilized G-Quadruplexes with Anti Glycosidic Torsion Angles in Solution. *Biochemistry* **1992**, *31* (35), 8112–8119. <https://doi.org/10.1021/bi00150a002>.
- (53) Müller, S.; Rodriguez, R. G-Quadruplex Interacting Small Molecules and Drugs: From Bench toward Bedside. *Expert Rev. Clin. Pharmacol.* **2014**, *7* (5), 663–679. <https://doi.org/10.1586/17512433.2014.945909>.
- (54) Cuesta, J.; Read, M.; Neidle, S. The Design of G-Quadruplex Ligands as Telomerase Inhibitors. *Mini-Rev. Med. Chem.* **2003**, *3* (1), 11–21. <https://doi.org/10.2174/1389557033405502>.
- (55) Gomez, D.; Mergny, J.-L.; Riou, J.-F. Detection of Telomerase Inhibitors Based on G-Quadruplex Ligands by a Modified Telomeric Repeat Amplification Protocol Assay. *Nature* **1991**, *350*, 718–720.
- (56) Zakian, V. A. Telomeres: Beginning to Understand the End. *Science* **1995**, *270* (5242), 1601–1607. <https://doi.org/10.1126/science.270.5242.1601>.
- (57) Azzalin, C. M.; Reichenbach, P.; Khoraiuli, L.; Giulotto, E.; Lingner, J. Telomeric Repeat Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science* **2007**, *318* (5851), 798–801. <https://doi.org/10.1126/science.1147182>.
- (58) Schoeftner, S.; Blasco, M. A. Developmentally Regulated Transcription of Mammalian Telomeres by DNA-Dependent RNA Polymerase II. *Nat. Cell Biol.* **2008**, *10* (2), 228–236. <https://doi.org/10.1038/ncb1685>.
- (59) Cifuentes-Rojas, C.; Shippen, D. E. Telomerase Regulation. *Mutat. Res. Mol. Mech. Mutagen.* **2012**, *730* (1–2), 20–27. <https://doi.org/10.1016/j.mrfmmm.2011.10.003>.
- (60) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy? *Nat. Rev. Drug Discov.* **2011**, *10* (4), 261–275. <https://doi.org/10.1038/nrd3428>.
- (61) Seenisamy, J.; Rezler, E. M.; Powell, T. J.; Tye, D.; Gokhale, V.; Joshi, C. S.; Siddiqui-Jain, A.; Hurley, L. H. The Dynamic Character of the G-Quadruplex Element in the c-MYC Promoter and Modification by TMPyP4. *J. Am. Chem. Soc.* **2004**, *126* (28), 8702–8709. <https://doi.org/10.1021/ja040022b>.
- (62) Fernando, H.; Reszka, A. P.; Huppert, J.; Ladame, S.; Rankin, S.; Venkitaraman, A. R.; Neidle, S.; Balasubramanian, S. A Conserved Quadruplex Motif Located in a Transcription Activation Site of the Human C-Kit Oncogene. *Biochemistry* **2006**, *45* (25), 7854–7860. <https://doi.org/10.1021/bi0601510>.
- (63) Cogo, S.; Paramasivam, M.; Spolaore, B.; Xodo, L. E. Structural Polymorphism within a Regulatory Element of the Human KRAS Promoter: Formation of G4-DNA Recognized by Nuclear Proteins. *Nucleic Acids Res.*

- 2008**, 36 (11), 3765–3780. <https://doi.org/10.1093/nar/gkn120>.
- (64) Cogoi, S.; Xodo, L. E. G-Quadruplex Formation within the Promoter of the KRAS Proto-Oncogene and Its Effect on Transcription. *Nucleic Acids Res.* **2006**, 34 (9), 2536–2549. <https://doi.org/10.1093/nar/gkl286>.
- (65) Paramasivam, M.; Membrino, A.; Cogoi, S.; Fukuda, H.; Nakagama, H.; Xodo, L. E. Protein HnRNP A1 and Its Derivative Up1 Unfold Quadruplex DNA in the Human KRAS Promoter: Implications for Transcription. *Nucleic Acids Res.* **2009**, 37 (9), 2841–2853. <https://doi.org/10.1093/nar/gkp138>.
- (66) 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>.
- (67) 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 Res.* **2005**, 33 (18), 6070–6080. <https://doi.org/10.1093/nar/gki917>.
- (68) De Armond, R.; Wood, S.; Sun, D.; Hurley, L. H.; Ebbinghaus, S. W. Evidence for the Presence of a Guanine Quadruplex Forming Region within a Polypurine Tract of the Hypoxia Inducible Factor 1 α Promoter †. *Biochemistry* **2005**, 44 (49), 16341–16350. <https://doi.org/10.1021/bi051618u>.
- (69) Palumbo, S. L.; Memmott, R. M.; Uribe, D. J.; Krotovakhan, Y.; Hurley, L. H.; Ebbinghaus, S. W. A Novel G-Quadruplex-Forming GGA Repeat Region in the c-Myb Promoter Is a Critical Regulator of Promoter Activity. *Nucleic Acids Res.* **2008**, 36 (6), 1755–1769. <https://doi.org/10.1093/nar/gkm1069>.
- (70) Qin, Y.; Rezler, E. M.; Gokhale, V.; Sun, D.; Hurley, L. H. Characterization of the G-Quadruplexes in the Duplex Nuclease Hypersensitive Element of the PDGF-A Promoter and Modulation of PDGF-A Promoter Activity by TMPyP4. *Nucleic Acids Res.* **2007**, 35 (22), 7698–7713. <https://doi.org/10.1093/nar/gkm538>.
- (71) Qin, Y.; Fortin, J. S.; Tye, D.; Gleason-Guzman, M.; Brooks, T. A.; Hurley, L. H. Molecular Cloning of the Human Platelet-Derived Growth Factor Receptor β (PDGFR- β) Promoter and Drug Targeting of the G-Quadruplex-Forming Region To Repress PDGFR- β Expression. *Biochemistry* **2010**, 49 (19), 4208–4219. <https://doi.org/10.1021/bi100330w>.
- (72) Xu, Y. Formation of the G-Quadruplex and i-Motif Structures in Retinoblastoma Susceptibility Genes (Rb). *Nucleic Acids Res.* **2006**, 34 (3), 949–954. <https://doi.org/10.1093/nar/gkj485>.
- (73) Palumbo, S. L.; Ebbinghaus, S. W.; Hurley, L. H. Formation of a Unique End-to-End Stacked Pair of G-Quadruplexes in the HTERT Core Promoter with Implications for Inhibition of Telomerase by G-Quadruplex-Interactive Ligands. *J. Am. Chem. Soc.* **2009**, 131 (31), 10878–10891. <https://doi.org/10.1021/ja902281d>.
- (74) 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>.
- (75) 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>.
- (76) Hershman, S. G.; Chen, Q.; Lee, J. Y.; Kozak, M. L.; Yue, P.; Wang, L.-S.; Johnson, F. B. Genomic Distribution and Functional Analyses of Potential G-Quadruplex-Forming Sequences in *Saccharomyces Cerevisiae*. *Nucleic Acids Res.* **2008**, 36 (1), 144–156. <https://doi.org/10.1093/nar/gkm986>.
- (77) Datta, A.; Pollock, K. J.; Kormuth, K. A.; Brosh Jr, R. M. G-Quadruplex Assembly by Ribosomal DNA: Emerging Roles in Disease Pathogenesis and Cancer Biology. *Cytogenet. Genome Res.* **2021**, 161 (6–7), 285–296. <https://doi.org/10.1159/000516394>.
- (78) Falabella, M.; Kolesar, J. E.; Wallace, C.; de Jesus, D.; Sun, L.; Taguchi, Y. V.; Wang, C.; Wang, T.; Xiang, I. M.; Alder, J. K.; Maheshan, R.; Horne, W.; Turek-Herman, J.; Pagano, P. J.; St. Croix, C. M.; Sondheimer, N.; Yatsunyk, L. A.; Johnson, F. B.; Kaufman, B. A. G-Quadruplex Dynamics Contribute to Regulation of Mitochondrial Gene Expression. *Sci. Rep.* **2019**, 9 (1), 5605. <https://doi.org/10.1038/s41598-019-41464-y>.
- (79) Falabella, M.; Fernandez, R. J.; Johnson, F. B.; Kaufman, B. A. Potential Roles for G-Quadruplexes in Mitochondria. *Curr. Med. Chem.* **2019**, 26 (16), 2918–2932. <https://doi.org/10.2174/0929867325666180228165527>.
- (80) Kwok, C. K.; Marsico, G.; Balasubramanian, S. Detecting RNA G-Quadruplexes (RG4s) in the Transcriptome. *Cold Spring Harb. Perspect. Biol.* **2018**, 10 (7), a032284. <https://doi.org/10.1101/cshperspect.a032284>.
- (81) Fay, M. M.; Lyons, S. M.; Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* **2017**, 429 (14), 2127–2147. <https://doi.org/10.1016/j.jmb.2017.05.017>.
- (82) Endoh, T.; Sugimoto, N. Unusual -1 Ribosomal Frameshift Caused by Stable RNA G-Quadruplex in Open Reading Frame. *Anal. Chem.* **2013**, 85 (23), 11435–11439. <https://doi.org/10.1021/acs.402497x>.
- (83) Liu, H.; Zheng, K.; He, Y.; Chen, Q.; Hao, Y.; Tan, Z. RNA G-Quadruplex Formation in Defined Sequence in Living Cells Detected by Bimolecular Fluorescence Complementation. *Chem. Sci.* **2016**, 7 (7), 4573–4581. <https://doi.org/10.1039/C5SC03946K>.
- (84) Endoh, T.; Sugimoto, N. Mechanical Insights into Ribosomal Progression Overcoming RNA G-Quadruplex from Periodical Translation Suppression in Cells. *Sci. Rep.* **2016**, 6 (1), 22719. <https://doi.org/10.1038/srep22719>.
- (85) Beaudoin, J.-D.; Perreault, J.-P. 5'-UTR G-Quadruplex Structures Acting as Translational Repressors. *Nucleic Acids Res.* **2010**, 38 (20), 7022–7036. <https://doi.org/10.1093/nar/gkq557>.
- (86) Bugaut, A.; Balasubramanian, S. 5'-UTR RNA G-Quadruplexes: Translation Regulation and Targeting. *Nucleic Acids Res.* **2012**, 40 (11), 4727–4741. <https://doi.org/10.1093/nar/gks068>.
- (87) Beaudoin, J.-D.; Perreault, J.-P. Exploring mRNA 3'-UTR G-Quadruplexes: Evidence of Roles in Both Alternative Polyadenylation and mRNA Shortening. *Nucleic Acids Res.* **2013**, 41 (11), 5898–5911. <https://doi.org/10.1093/nar/gkt265>.
- (88) Rouleau, S.; Glouzon, J.-P. S.; Brumwell, A.; Bisailon, M.; Perreault, J.-P. 3' UTR G-Quadruplexes Regulate miRNA Binding. *RNA* **2017**, 23 (8), 1172–1179. <https://doi.org/10.1261/rna.060962.117>.

- (89) 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>.
- (90) Kumari, S.; Bugaut, A.; Balasubramanian, S. Position and Stability Are Determining Factors for Translation Repression by an RNA G-Quadruplex-Forming Sequence within the 5' UTR of the NRAS Proto-Oncogene. *Biochemistry* **2008**, *47* (48), 12664–12669. <https://doi.org/10.1021/bi8010797>.
- (91) Tsai, Z. T.-Y.; Chu, W.-Y.; Cheng, J.-H.; Tsai, H.-K. Associations between Intronic Non-B DNA Structures and Exon Skipping. *Nucleic Acids Res.* **2014**, *42* (2), 739–747. <https://doi.org/10.1093/nar/gkt939>.
- (92) Gomez, D. Telomerase Downregulation Induced by the G-Quadruplex Ligand 12459 in A549 Cells Is Mediated by HTERT RNA Alternative Splicing. *Nucleic Acids Res.* **2004**, *32* (1), 371–379. <https://doi.org/10.1093/nar/gkh181>.
- (93) Didiot, M.-C.; Tian, Z.; Schaeffer, C.; Subramanian, M.; Mandel, J.-L.; Moine, H. The G-Quartet Containing FMRP Binding Site in FMR1 mRNA Is a Potent Exonic Splicing Enhancer. *Nucleic Acids Res.* **2008**, *36* (15), 4902–4912. <https://doi.org/10.1093/nar/gkn472>.
- (94) Neidle, S. Human Telomeric G-Quadruplex: The Current Status of Telomeric G-Quadruplexes as Therapeutic Targets in Human Cancer: G-Quadruplexes as Cancer Drug Targets. *FEBS J.* **2010**, *277* (5), 1118–1125. <https://doi.org/10.1111/j.1742-4658.2009.07463.x>.
- (95) Xu, Y.; Suzuki, Y.; Ito, K.; Komiyama, M. Telomeric Repeat-Containing RNA Structure in Living Cells. *Proc. Natl. Acad. Sci.* **2010**, *107* (33), 14579–14584. <https://doi.org/10.1073/pnas.1001177107>.
- (96) Martadinata, H.; Phan, A. T. Structure of Human Telomeric RNA (TERRA): Stacking of Two G-Quadruplex Blocks in K⁺ Solution. *Biochemistry* **2013**, *52* (13), 2176–2183. <https://doi.org/10.1021/bi301606u>.
- (97) Biffi, G.; Tannahill, D.; Balasubramanian, S. An Intramolecular G-Quadruplex Structure Is Required for Binding of Telomeric Repeat-Containing RNA to the Telomeric Protein TRF2. *J. Am. Chem. Soc.* **2012**, *134* (29), 11974–11976. <https://doi.org/10.1021/ja305734x>.
- (98) Gros, J.; Guédin, A.; Mergny, J.-L.; Lacroix, L. G-Quadruplex Formation Interferes with P1 Helix Formation in the RNA Component of Telomerase HTERC. *ChemBioChem* **2008**, *9* (13), 2075–2079. <https://doi.org/10.1002/cbic.200800300>.
- (99) Lattmann, S.; Stadler, M. B.; Vaughn, J. P.; Akman, S. A.; Nagamine, Y. The DEAH-Box RNA Helicase RHAU Binds an Intramolecular RNA G-quadruplex in TERC and Associates with Telomerase Holoenzyme. *Nucleic Acids Res.* **2011**, *39* (21), 9390–9404. <https://doi.org/10.1093/nar/gkr630>.
- (100) Kharel, P.; Becker, G.; Tsvetkov, V.; Ivanov, P. Properties and Biological Impact of RNA G-Quadruplexes: From Order to Turmoil and Back. *Nucleic Acids Res.* **2020**, *48* (22), 12534–12555. <https://doi.org/10.1093/nar/gkaa1126>.
- (101) Bhartiya, D.; Chawla, V.; Ghosh, S.; Shankar, R.; Kumar, N. Genome-Wide Regulatory Dynamics of G-Quadruplexes in Human Malaria Parasite *Plasmodium Falciparum*. *Genomics* **2016**, *108* (5–6), 224–231. <https://doi.org/10.1016/j.ygeno.2016.10.004>.
- (102) Dumetz, F.; Merrick, C. Parasitic Protozoa: Unusual Roles for G-Quadruplexes in Early-Diverging Eukaryotes. *Molecules* **2019**, *24* (7), 1339. <https://doi.org/10.3390/molecules24071339>.
- (103) Bartas, M.; Čutová, M.; Brázda, V.; Kaura, P.; Šťastný, J.; Kolomazník, J.; Coufal, J.; Goswami, P.; Červeň, J.; Pečinka, P. The Presence and Localization of G-Quadruplex Forming Sequences in the Domain of Bacteria. *Molecules* **2019**, *24* (9), 1711. <https://doi.org/10.3390/molecules24091711>.
- (104) Holder, I. T.; Hartig, J. S. A Matter of Location: Influence of G-Quadruplexes on Escherichia Coli Gene Expression. *Chem. Biol.* **2014**, *21* (11), 1511–1521. <https://doi.org/10.1016/j.chembiol.2014.09.014>.
- (105) Belmonte-Reche, E.; Martínez-García, M.; Guédin, A.; Zuffo, M.; Arévalo-Ruiz, M.; Doria, F.; Campos-Salinas, J.; Maynadier, M.; López-Rubio, J. J.; Freccero, M.; Mergny, J.-L.; Pérez-Victoria, J. M.; Morales, J. C. G-Quadruplex Identification in the Genome of Protozoan Parasites Points to Naphthalene Diimide Ligands as New Antiparasitic Agents. *J. Med. Chem.* **2018**, *61* (3), 1231–1240. <https://doi.org/10.1021/acs.jmedchem.7b01672>.
- (106) Artusi, S.; Perrone, R.; Lago, S.; Raffa, P.; Di Iorio, E.; Palù, G.; Richter, S. N. Visualization of DNA G-Quadruplexes in Herpes Simplex Virus 1-Infected Cells. *Nucleic Acids Res.* **2016**, gkw968. <https://doi.org/10.1093/nar/gkw968>.
- (107) Tlučková, K.; Marušič, M.; Tóthová, P.; Bauer, L.; Šket, P.; Plavec, J.; Viglasky, V. Human Papillomavirus G-Quadruplexes. *Biochemistry* **2013**, *52* (41), 7207–7216. <https://doi.org/10.1021/bi400897g>.
- (108) Fleming, A. M.; Ding, Y.; Alenko, A.; Burrows, C. J. Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. *ACS Infect. Dis.* **2016**, *2* (10), 674–681. <https://doi.org/10.1021/acsinfecdis.6b00109>.
- (109) Zhao, C.; Qin, G.; Niu, J.; Wang, Z.; Wang, C.; Ren, J.; Qu, X. Targeting RNA G-Quadruplex in SARS-CoV-2: A Promising Therapeutic Target for COVID-19? *Angew. Chem. Int. Ed.* **2021**, *60* (1), 432–438. <https://doi.org/10.1002/anie.202011419>.
- (110) Perrone, R.; Nadai, M.; Poe, J. A.; Frasson, I.; Palumbo, M.; Palù, G.; Smithgall, T. E.; Richter, S. N. Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 Nef Coding Region: Implications for Antiviral Activity. *PLoS ONE* **2013**, *8* (8), e73121. <https://doi.org/10.1371/journal.pone.0073121>.
- (111) Perrone, R.; Nadai, M.; Frasson, I.; Poe, J. A.; Butovskaya, E.; Smithgall, T. E.; Palumbo, M.; Palù, G.; Richter, S. N. A Dynamic G-Quadruplex Region Regulates the HIV-1 Long Terminal Repeat Promoter. *J. Med. Chem.* **2013**, *56* (16), 6521–6530. <https://doi.org/10.1021/jm400914r>.
- (112) Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. Small-Molecule-Induced DNA Damage Identifies Alternative DNA Structures in Human Genes. *Nat. Chem. Biol.* **2012**, *8* (3), 301–310. <https://doi.org/10.1038/nchembio.780>.
- (113) 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 Res.* **2021**, gkab952. <https://doi.org/10.1093/nar/gkab952>.
- (114) Santos, T.; Salgado, G. F.; Cabrita, E. J.; Cruz, C. G-Quadruplexes and Their Ligands: Biophysical Methods

- to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals* **2021**, *14* (8), 769. <https://doi.org/10.3390/ph14080769>.
- (115) Phan, A. T. Human Telomeric G-Quadruplex: Structures of DNA and RNA Sequences: Human Telomeric G-Quadruplex Structures. *FEBS J.* **2010**, *277* (5), 1107–1117. <https://doi.org/10.1111/j.1742-4658.2009.07464.x>.
- (116) Arola, A.; Vilar, R. Stabilisation of G-Quadruplex DNA by Small Molecules. *Curr. Top. Med. Chem.* **2008**, *8* (15), 1405–1415. <https://doi.org/10.2174/156802608786141106>.
- (117) Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. Solution Structure of a 2:1 Quindoline–c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J. Am. Chem. Soc.* **2011**, *133* (44), 17673–17680. <https://doi.org/10.1021/ja205646q>.
- (118) Yan, Y.; Tan, J.; Ou, T.; Huang, Z.; Gu, L. DNA G-Quadruplex Binders: A Patent Review. *Expert Opin. Ther. Pat.* **2013**, *23* (11), 1495–1509. <https://doi.org/10.1517/13543776.2013.833187>.
- (119) Sun, Z.-Y.; Wang, X.-N.; Cheng, S.-Q.; Su, X.-X.; Ou, T.-M. Developing Novel G-Quadruplex Ligands: From Interaction with Nucleic Acids to Interfering with Nucleic Acid–Protein Interaction. *Molecules* **2019**, *24* (3), 396. <https://doi.org/10.3390/molecules24030396>.
- (120) Read, M.; Harrison, R. J.; Romagnoli, B.; Tanious, 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.* **2001**, *98* (9), 4844–4849. <https://doi.org/10.1073/pnas.081560598>.
- (121) Martino, L.; Pagano, B.; Fotticchia, I.; Neidle, S.; Giancola, C. Shedding Light on the Interaction between TMPyP4 and Human Telomeric Quadruplexes. *J. Phys. Chem. B* **2009**, *113* (44), 14779–14786. <https://doi.org/10.1021/jp9066394>.
- (122) Zuffo, M.; Guédin, A.; Leriche, E.-D.; Doria, F.; Pirola, V.; Gabelica, V.; Mergny, J.-L.; Freccero, M. More Is Not Always Better: Finding the Right Trade-off between Affinity and Selectivity of a G-Quadruplex Ligand. *Nucleic Acids Res.* **2018**, *46* (19), e115–e115. <https://doi.org/10.1093/nar/gky607>.
- (123) Asamitsu, S.; Bando, T.; Sugiyama, H. Ligand Design to Acquire Specificity to Intended G-Quadruplex Structures. *Chem. - Eur. J.* **2019**, *25* (2), 417–430. <https://doi.org/10.1002/chem.201802691>.
- (124) Duan, W.; Rangan, A.; Vankayalapati, H.; Kim, M.-Y.; Zeng, Q.; Sun, D.; Han, H.; Fedoroff, O. Y.; Nishioka, D.; Rha, S. Y.; Izbicka, E.; Hoff, D. D. V.; Hurley, L. H. Design and Synthesis of Fluoroquinophenoxazines That Interact with Human Telomeric G-Quadruplexes and Their Biological Effects. **2001**, *1*, 103–120.
- (125) Cylene Pharmaceuticals. *A Phase 2, Multi-Center, Open Label Study Evaluating the Efficacy and Safety of Quarfloxin in Patients With Low to Intermediate Grade Neuroendocrine Carcinoma*; Clinical trial registration NCT00780663; clinicaltrials.gov, 2011.
- (126) Guyen, B.; Schultes, C. M.; Hazel, P.; Mann, J.; Neidle, S. Synthesis and Evaluation of Analogues of 10H-Indolo[3,2-b]Quinoline as G-Quadruplex Stabilising Ligands and Potential Inhibitors of the Enzyme Telomerase. *Org. Biomol. Chem.* **2004**, *2* (7), 981. <https://doi.org/10.1039/b316055f>.
- (127) Ou, T.-M.; Lu, Y.-J.; Zhang, C.; Huang, Z.-S.; Wang, X.-D.; Tan, J.-H.; Chen, Y.; Ma, D.-L.; Wong, K.-Y.; Tang, J. C.-O.; Chan, A. S.-C.; Gu, L.-Q. Stabilization of G-Quadruplex DNA and Down-Regulation of Oncogene *c-Myc* by Quindoline Derivatives. *J. Med. Chem.* **2007**, *50* (7), 1465–1474. <https://doi.org/10.1021/jm0610088>.
- (128) Lavrado, J.; Reszka, A. P.; Moreira, R.; Neidle, S.; Paulo, A. C-11 Diamino Cryptolepine Derivatives NSC748392, NSC748393, and NSC748394: Anticancer Profile and G-Quadruplex Stabilization. *Bioorg. Med. Chem. Lett.* **2010**, *20* (23), 7042–7045. <https://doi.org/10.1016/j.bmcl.2010.09.110>.
- (129) Funke, A.; Dickerhoff, J.; Weisz, K. Towards the Development of Structure-Selective G-Quadruplex-Binding Indolo[3,2-*b*]Quinolines. *Chem. - Eur. J.* **2016**, *22* (9), 3170–3181. <https://doi.org/10.1002/chem.201504416>.
- (130) Harrison, R. J.; Cuesta, J.; Chessari, G.; Read, M. A.; Basra, S. K.; Reszka, A. P.; Morrell, J.; Gowan, S. M.; Incles, C. M.; Tanious, 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>.
- (131) Burger, A. M.; Dai, F.; Schultes, C. M.; Reszka, A. P.; Moore, M. J.; Double, J. A.; Neidle, S. The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Res.* **2005**, *65* (4), 1489–1496. <https://doi.org/10.1158/0008-5472.CAN-04-2910>.
- (132) Guo, Q.-L.; Su, H.-F.; Wang, N.; Liao, S.-R.; Lu, Y.-T.; Ou, T.-M.; Tan, J.-H.; Li, D.; Huang, Z.-S. Synthesis and Evaluation of 7-Substituted-5,6-Dihydrobenzo[*c*]Acridine Derivatives as New *c-KIT* Promoter G-Quadruplex Binding Ligands. *Eur. J. Med. Chem.* **2017**, *130*, 458–471. <https://doi.org/10.1016/j.ejmech.2017.02.051>.
- (133) Gunaratnam, M.; Swank, S.; Haider, S. M.; Galesa, K.; Reszka, A. P.; Beltran, M.; Cuenca, F.; Fletcher, J. A.; Neidle, S. Targeting Human Gastrointestinal Stromal Tumor Cells with a Quadruplex-Binding Small Molecule. *J. Med. Chem.* **2009**, *52* (12), 3774–3783. <https://doi.org/10.1021/jm900424a>.
- (134) Marchetti, C.; Zyner, K. G.; Ohnmacht, S. A.; Robson, M.; Haider, S. M.; Morton, J. P.; Marsico, G.; Vo, T.; Laughlin-Toth, S.; Ahmed, A. A.; Di Vita, G.; Pazitna, I.; Gunaratnam, M.; Besser, R. J.; Andrade, A. C. G.; Diocou, S.; Pike, J. A.; Tannahill, D.; Pedley, R. B.; Evans, T. R. J.; Wilson, W. D.; Balasubramanian, S.; Neidle, S. Targeting Multiple Effector Pathways in Pancreatic Ductal Adenocarcinoma with a G-Quadruplex-Binding Small Molecule. *J. Med. Chem.* **2018**, *61* (6), 2500–2517. <https://doi.org/10.1021/acs.jmedchem.7b01781>.
- (135) Chang, C.-C.; Wu, J.-Y.; Chien, C.-W.; Wu, W.-S.; Liu, H.; Kang, C.-C.; Yu, L.-J.; Chang, T.-C. A Fluorescent Carbazole Derivative: High Sensitivity for Quadruplex DNA. *Anal. Chem.* **2003**, *75* (22), 6177–6183. <https://doi.org/10.1021/ac034789i>.
- (136) Chang, C.-C.; Wu, J.-Y.; Chang, T.-C. A Carbazole Derivative Synthesis for Stabilizing the Quadruplex Structure. *J. Chin. Chem. Soc.* **2003**, *50* (2), 185–188. <https://doi.org/10.1002/jccs.200300026>.
- (137) Giustiniano, M.; Meneghetti, F.; Mercalli, V.; Varese, M.; Giustiniano, F.; Novellino, E.; Tron, G. C. Synthesis of Aminocarbonyl *N*-Acylhydrazones by a Three-Component Reaction of Isocyanides, Hydrasonoyl Chlorides, and Carboxylic Acids. *Org. Lett.* **2014**, *16* (20), 5332–5335. <https://doi.org/10.1021/ol502515b>.

- (138) Giustiniano, M.; Mercalli, V.; Amato, J.; Novellino, E.; Tron, G. C. Exploiting the Electrophilic and Nucleophilic Dual Role of Nitrile Imines: One-Pot, Three-Component Synthesis of Furo[2,3-*d*]Pyridazin-4(5*H*)-Ones. *Org. Lett.* **2015**, *17* (16), 3964–3967. <https://doi.org/10.1021/acs.orglett.5b01798>.
- (139) Amato, J.; Pagano, A.; Capasso, D.; Di Gaetano, S.; Giustiniano, M.; Novellino, E.; Randazzo, A.; Pagano, B. Targeting the *BCL2* Gene Promoter G-Quadruplex with a New Class of Furopyridazinone-Based Molecules. *ChemMedChem* **2018**, *13* (5), 406–410. <https://doi.org/10.1002/cmdc.201700749>.
- (140) 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>.
- (141) Dhamodharan, V.; Harikrishna, S.; Bhasikuttan, A. C.; Pradeepkumar, P. I. Topology Specific Stabilization of Promoter over Telomeric G-Quadruplex DNAs by Bisbenzimidazole Carboxamide Derivatives. *ACS Chem. Biol.* **2015**, *10* (3), 821–833. <https://doi.org/10.1021/cb5008597>.
- (142) Medeiros-Silva, J.; Guédin, A.; Salgado, G. F.; Mergny, J.-L.; Queiroz, J. A.; Cabrita, E. J.; Cruz, C. Phenanthroline-Bis-Oxazole Ligands for Binding and Stabilization of G-Quadruplexes. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2017**, *1861* (5), 1281–1292. <https://doi.org/10.1016/j.bbagen.2016.11.024>.
- (143) Pennarun, G.; Granotier, C.; Gauthier, L. R.; Gomez, D.; Hoffschir, F.; Mandine, E.; Riou, J.-F.; Mergny, J.-L.; Mailliet, P.; Boussin, F. D. Apoptosis Related to Telomere Instability and Cell Cycle Alterations in Human Glioma Cells Treated by New Highly Selective G-Quadruplex Ligands. *Oncogene* **2005**, *24* (18), 2917–2928. <https://doi.org/10.1038/sj.onc.1208468>.
- (144) Rodriguez, R.; Müller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, J.-F.; Balasubramanian, S. A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. *J. Am. Chem. Soc.* **2008**, *130* (47), 15758–15759. <https://doi.org/10.1021/ja805615w>.
- (145) Hamon, F.; Largy, E.; Guédin-Beaurepaire, A.; Rouchon-Dagois, M.; Sidibe, A.; Monchaud, D.; Mergny, J.-L.; Riou, J.-F.; Nguyen, C.-H.; Teulade-Fichou, M.-P. An Acyclic Oligoheteroaryle That Discriminates Strongly between Diverse G-Quadruplex Topologies. *Angew. Chem. Int. Ed.* **2011**, *50* (37), 8745–8749. <https://doi.org/10.1002/anie.201103422>.
- (146) Rizeq, N.; Georgiades, S. Investigation of 'Head-to-Tail'-Connected Oligoaryl N,O-Ligands as Recognition Motifs for Cancer-Relevant G-Quadruplexes. *Molecules* **2017**, *22* (12), 2160. <https://doi.org/10.3390/molecules22122160>.
- (147) Mendes, E.; Cadoni, E.; Carneiro, F.; Afonso, M. B.; Brito, H.; Lavrado, J.; dos Santos, D. J. V. A.; Vitor, J. B.; Neidle, S.; Rodrigues, C. M. P.; Paulo, A. Combining 1,3-Ditriazolylbenzene and Quinoline to Discover a New G-Quadruplex-Interactive Small Molecule Active against Cancer Stem-Like Cells. *ChemMedChem* **2019**, *14* (14), 1325–1328. <https://doi.org/10.1002/cmdc.201900243>.
- (148) Moorhouse, A. D.; Haider, S.; Gunaratnam, M.; Munnur, D.; Neidle, S.; Moses, J. E. Targeting Telomerase and Telomeres: A Click Chemistry Approach towards Highly Selective G-Quadruplex Ligands. *Mol. Biosyst.* **2008**, *4* (6), 629. <https://doi.org/10.1039/b801822g>.
- (149) Duarte, A. R.; Cadoni, E.; Ressurreição, A. S.; Moreira, R.; Paulo, A. Design of Modular G-Quadruplex Ligands. *ChemMedChem* **2018**, *13* (9), 869–893. <https://doi.org/10.1002/cmdc.201700747>.
- (150) Tang Y, Xu X, Li Q, et al. Novel Use of Chuanbeinone Compounds. CN101352445A; 2009.
- (151) Minhas, G. S.; Pilch, D. S.; Kerrigan, J. E.; LaVoie, E. J.; Rice, J. E. Synthesis and G-Quadruplex Stabilizing Properties of a Series of Oxazole-Containing Macrocycles. *Bioorg. Med. Chem. Lett.* **2006**, *16* (15), 3891–3895. <https://doi.org/10.1016/j.bmcl.2006.05.038>.
- (152) Shin-ya, K.; Wierzba, K.; Matsuo, K.; Ohtani, T.; Yamada, Y.; Furihata, K.; Hayakawa, Y.; Seto, H. Telomestatin, a Novel Telomerase Inhibitor from *Streptomyces Anulatus*. *J. Am. Chem. Soc.* **2001**, *123* (6), 1262–1263. <https://doi.org/10.1021/ja005780q>.
- (153) 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>.
- (154) Tera, M.; Ishizuka, H.; Takagi, M.; Suganuma, M.; Shin-ya, K.; Nagasawa, K. Macrocyclic Hexaoxazoles as Sequence- and Mode-Selective G-Quadruplex Binders. *Angew. Chem. Int. Ed.* **2008**, *47* (30), 5557–5560. <https://doi.org/10.1002/anie.200801235>.
- (155) Tera, M.; Iida, K.; Ishizuka, H.; Takagi, M.; Suganuma, M.; Doi, T.; Shin-ya, K.; Nagasawa, K. Synthesis of a Potent G-Quadruplex-Binding Macrocyclic Heptaoxazole. *ChemBioChem* **2009**, *10* (3), 431–435. <https://doi.org/10.1002/cbic.200800563>.
- (156) Guo, K.; Pourpak, A.; Beetz-Rogers, K.; Gokhale, V.; Sun, D.; Hurley, L. H. Formation of Pseudosymmetrical G-Quadruplex and i-Motif Structures in the Proximal Promoter Region of the *RET* Oncogene. *J. Am. Chem. Soc.* **2007**, *129* (33), 10220–10228. <https://doi.org/10.1021/ja072185g>.
- (157) Izbicka, E.; Wheelhouse, R. T.; Raymond, E.; Davidson, K. K.; Lawrence, R. A.; Sun, D.; Windle, B. E.; Hurley, L. H.; Hoff, D. D. V. Effects of Cationic Porphyrins as G-Quadruplex Interactive Agents in Human Tumor Cells. *Cancer Res.* **1999**, *59*, 639–644.
- (158) Gonçalves, D. P. N.; Rodriguez, R.; Balasubramanian, S.; Sanders, J. K. M. Tetramethylpyridiniumporphyrazines—a New Class of G-Quadruplex Inducing and Stabilising Ligands. *Chem Commun* **2006**, No. 45, 4685–4687. <https://doi.org/10.1039/B611731G>.
- (159) Puckett, C. A.; Ernst, R. J.; Barton, J. K. Exploring the Cellular Accumulation of Metal Complexes. *Dalton Trans* **2010**, 39 (5), 1159–1170. <https://doi.org/10.1039/B922209J>.
- (160) Dixon, I. M.; Lopez, F.; Tejera, A. M.; Estève, J.-P.; Blasco, M. A.; Pratiel, G.; Meunier, B. A G-Quadruplex Ligand with 10000-Fold Selectivity over Duplex DNA. *J. Am. Chem. Soc.* **2007**, *129* (6), 1502–1503. <https://doi.org/10.1021/ja065591t>.
- (161) Pipier, A.; De Rache, A.; Modeste, C.; Amrane, S.; Mothes-Martin, E.; Stigliani, J.-L.; Calsou, P.; Mergny, J.-L.; Pratiel, G.; Gomez, D. G-Quadruplex Binding Optimization by Gold(III) Insertion into the Center of a Porphyrin. *Dalton Trans.* **2019**, 48 (18), 6091–6099. <https://doi.org/10.1039/C8DT04703K>.
- (162) Łęczkowska, A.; Gonzalez-Garcia, J.; Perez-Arnaiz, C.; Garcia, B.; White, A. J. P.; Vilar, R. Binding Studies of Metal-Salphen and Metal-Bipyridine Complexes

- towards G-Quadruplex DNA. *Chem. – Eur. J.* **2018**, *24* (45), 11785–11794. <https://doi.org/10.1002/chem.201802248>.
- (163) Banerjee, N.; Panda, S.; Chatterjee, S. *G-Quadruplex Targeting Ligands: A Hope and a New Horizon in Cancer Therapeutics*; preprint; CHEMISTRY, 2018. <https://doi.org/10.20944/preprints201811.0409.v1>.
- (164) Ladame, S.; Schouten, J. A.; Stuart, J.; Roldan, J.; Neidle, S.; Balasubramanian, S. Tetrapeptides Induce Selective Recognition for G-Quadruplexes When Conjugated to a DNA-Binding Platform. *Org. Biomol. Chem.* **2004**, *2* (20), 2925. <https://doi.org/10.1039/b409698c>.
- (165) Redman, J. E.; Granadino-Roldán, J. M.; Schouten, J. A.; Ladame, S.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. Recognition and Discrimination of DNA Quadruplexes by Acridine-Peptide Conjugates. *Org. Biomol. Chem.* **2009**, *7* (1), 76–84. <https://doi.org/10.1039/B814682A>.
- (166) Schouten, J. A.; Ladame, S.; Mason, S. J.; Cooper, M. A.; Balasubramanian, S. G-Quadruplex-Specific Peptide-Hemicyanine Ligands by Partial Combinatorial Selection. *J. Am. Chem. Soc.* **2003**, *125* (19), 5594–5595. <https://doi.org/10.1021/ja029356w>.
- (167) Casals, J.; Debéthune, L.; Alvarez, K.; Risitano, A.; Fox, K. R.; Grandas, A.; Pedroso, E. Directing Quadruplex-Stabilizing Drugs to the Telomere: Synthesis and Properties of Acridine-Oligonucleotide Conjugates. *Bioconjug. Chem.* **2006**, *17* (5), 1351–1359. <https://doi.org/10.1021/bc060194t>.
- (168) Kaiser, M.; De Cian, A.; Sainlos, M.; Renner, C.; Mergny, J.-L.; Teulade-Fichou, M.-P. Neomycin-Capped Aromatic Platforms: Quadruplex DNA Recognition and Telomerase Inhibition. *Org. Biomol. Chem.* **2006**, *4* (6), 1049. <https://doi.org/10.1039/b516378a>.
- (169) Paul, A.; Sengupta, P.; Krishnan, Y.; Ladame, S. Combining G-Quadruplex Targeting Motifs on a Single Peptide Nucleic Acid Scaffold: A Hybrid (3+1) PNA-DNA Bimolecular Quadruplex. *Chem. – Eur. J.* **2008**, *14* (28), 8682–8689. <https://doi.org/10.1002/chem.200800605>.
- (170) Wang, K.; Flaherty, D. P.; Chen, L.; Yang, D. High-Throughput Screening of G-Quadruplex Ligands by FRET Assay. In *G-Quadruplex Nucleic Acids*; Yang, D., Lin, C., Eds.; Methods in Molecular Biology; Springer New York: New York, NY, 2019; Vol. 2035, pp 323–331. https://doi.org/10.1007/978-1-4939-9666-7_19.
- (171) Scalabrin, M.; Nadai, M.; Tassinari, M.; Lago, S.; Doria, F.; Frasson, I.; Freccero, M.; Richter, S. N. Selective Recognition of a Single HIV-1 G-Quadruplex by Ultrafast Small-Molecule Screening. *Anal. Chem.* **2021**, *acs.analchem.0c04106*. <https://doi.org/10.1021/acs.analchem.0c04106>.
- (172) Ma, D.-L.; Lai, T.-S.; Chan, F.-Y.; Chung, W.-H.; Abagyan, R.; Leung, Y.-C.; Wong, K.-Y. Discovery of a Drug-Like G-Quadruplex Binding Ligand by High-Throughput Docking. *ChemMedChem* **2008**, *3* (6), 881–884. <https://doi.org/10.1002/cmdc.200700342>.
- (173) Chen, S.-B.; Tan, J.-H.; Ou, T.-M.; Huang, S.-L.; An, L.-K.; Luo, H.-B.; Li, D.; Gu, L.-Q.; Huang, Z.-S. Pharmacophore-Based Discovery of Triaryl-Substituted Imidazole as New Telomeric G-Quadruplex Ligand. *Bioorg. Med. Chem. Lett.* **2011**, *21* (3), 1004–1009. <https://doi.org/10.1016/j.bmcl.2010.12.019>.
- (174) Lam, K. S.; Lebl, M.; Krchňák, V. The “One-Bead-One-Compound” Combinatorial Library Method. *Chem. Rev.* **1997**, *97* (2), 411–448. <https://doi.org/10.1021/cr9600114>.
- (175) Furka, F. History of Combinatorial Chemistry. *Drug Dev. Res.* **1995**, *36* (1), 1–12. <https://doi.org/10.1002/ddr.430360102>.
- (176) Redman, J. E.; Ladame, S.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. Discovery of G-Quadruplex Stabilizing Ligands through Direct ELISA of a One-Bead-One-Compound Library. *Org. Biomol. Chem.* **2006**, *4* (23), 4364. <https://doi.org/10.1039/b611716c>.
- (177) Nasiri, H. R.; Bell, N. M.; McLuckie, K. I. E.; Husby, J.; Abell, C.; Neidle, S.; Balasubramanian, S. Targeting a C-MYC G-Quadruplex DNA with a Fragment Library. *Chem Commun* **2014**, *50* (14), 1704–1707. <https://doi.org/10.1039/C3CC48390H>.
- (178) Tassinari, M.; Lena, A.; Butovskaya, E.; Pirota, V.; Nadai, M.; Freccero, M.; Doria, F.; Richter, S. A Fragment-Based Approach for the Development of G-Quadruplex Ligands: Role of the Amidoxime Moiety. *Molecules* **2018**, *23* (8), 1874. <https://doi.org/10.3390/molecules23081874>.
- (179) Bugaut, A.; Jantos, K.; Wietor, J.-L.; Rodriguez, R.; Sanders, J. K. M.; Balasubramanian, S. Exploring the Differential Recognition of DNA G-Quadruplex Targets by Small Molecules Using Dynamic Combinatorial Chemistry. *Angew. Chem. Int. Ed.* **2008**, *47* (14), 2677–2680. <https://doi.org/10.1002/anie.200705589>.
- (180) Reznichenko, O.; Cucchiarini, A.; Gabelica, V.; Granzhan, A. Quadruplex DNA-Guided Ligand Selection from Dynamic Combinatorial Libraries of Acylhydrazones. *Org. Biomol. Chem.* **2021**, *19* (2), 379–386. <https://doi.org/10.1039/D0OB01908A>.
- (181) Di Antonio, M.; Biffi, G.; Mariani, A.; Raiber, E.-A.; Rodriguez, R.; Balasubramanian, S. Selective RNA Versus DNA G-Quadruplex Targeting by In Situ Click Chemistry. *Angew. Chem. Int. Ed.* **2012**, *51* (44), 11073–11078. <https://doi.org/10.1002/anie.201206281>.
- (182) Keefe, A. D.; Pai, S.; Ellington, A. Aptamers as Therapeutics. *Nat. Rev. Drug Discov.* **2010**, *9* (7), 537–550. <https://doi.org/10.1038/nrd3141>.
- (183) Yang, M.; Sunderland, K.; Mao, C. Virus-Derived Peptides for Clinical Applications. *Chem. Rev.* **2017**, *117* (15), 10377–10402. <https://doi.org/10.1021/acs.chemrev.7b00100>.
- (184) Huang, Y.; Wiedmann, M. M.; Suga, H. RNA Display Methods for the Discovery of Bioactive Macrocycles. *Chem. Rev.* **2019**, *119* (17), 10360–10391. <https://doi.org/10.1021/acs.chemrev.8b00430>.
- (185) Schofield, D. J.; Pope, A. R.; Clementel, V.; Buckell, J.; Chapple, S. D.; Clarke, K. F.; Conquer, J. S.; Crofts, A. M.; Crowther, S. R.; Dyson, M. R.; Flack, G.; Griffin, G. J.; Hooks, Y.; Howat, W. J.; Kolb-Kokocinski, A.; Kunze, S.; Martin, C. D.; Maslen, G. L.; Mitchell, J. N.; O’Sullivan, M.; Perera, R. L.; Roake, W.; Shadbolt, S. P.; Vincent, K. J.; Warford, A.; Wilson, W. E.; Xie, J.; Young, J. L.; McCafferty, J. Application of Phage Display to High Throughput Antibody Generation and Characterization. *Genome Biol.* **2007**, *8* (11), R254. <https://doi.org/10.1186/gb-2007-8-11-r254>.
- (186) 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 Chem. Biol.* **2016**, *23* (10), 1261–1270. <https://doi.org/10.1016/j.chembiol.2016.08.013>.
- (187) Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A. In Vitro Generated Antibodies

- Specific for Telomeric Guanine-Quadruplex DNA React with Stylonychia Lemnae Macronuclei. *Proc. Natl. Acad. Sci.* **2001**, *98* (15), 8572–8577. (202) <https://doi.org/10.1073/pnas.141229498>.
- (188) Litovchick, A.; Tian, X.; Monteiro, M. I.; Kennedy, K. M.; Guié, M.-A.; Centrella, P.; Zhang, Y.; Clark, M. A.; Keefe, A. D. Novel Nucleic Acid Binding Small Molecules Discovered Using DNA-Encoded Chemistry. *Molecules* **2019**, *24* (10), 2026. <https://doi.org/10.3390/molecules24102026>.
- (189) Nielsen, T. E.; Schreiber, S. L. Towards the Optimal Screening Collection: A Synthesis Strategy. *Angew. Chem. Int. Ed.* **2008**, *47* (1), 48–56. <https://doi.org/10.1002/anie.200703073>.
- (190) Halai, R.; Cooper, M. A. Using Label-Free Screening Technology to Improve Efficiency in Drug Discovery. *Expert Opin. Drug Discov.* **2012**, *7* (2), 123–131. <https://doi.org/10.1517/17460441.2012.651121>.
- (191) Folmer, R. H. A. Integrating Biophysics with HTS-Driven Drug Discovery Projects. *Drug Discov. Today* **2016**, *21* (3), 491–498. (204) <https://doi.org/10.1016/j.drudis.2016.01.011>.
- (192) Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty Years on: The Impact of Fragments on Drug Discovery. *Nat. Rev. Drug Discov.* **2016**, *15* (9), 605–619. <https://doi.org/10.1038/nrd.2016.109>.
- (193) Stark, J. L.; Powers, R. Application of NMR and Molecular Docking in Structure-Based Drug Discovery. In *NMR of Proteins and Small Biomolecules*; Zhu, G., Ed.; Topics in Current Chemistry; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; Vol. 326, pp 1–34. https://doi.org/10.1007/128_2011_213.
- (194) Goodnow, R. A. The Changing Feasibility and Economics of Chemical Diversity Exploration with DNA-Encoded Combinatorial Approaches. In *A Handbook for DNA-Encoded Chemistry*; Goodnow, R. A., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2014; pp 417–426. <https://doi.org/10.1002/9781118832738.ch18>.
- (195) Neri, D.; Lerner, R. A. DNA-Encoded Chemical Libraries: A Selection System Based on Endowing Organic Compounds with Amplifiable Information. *Annu. Rev. Biochem.* **2018**, *87* (1), 479–502. <https://doi.org/10.1146/annurev-biochem-062917-012550>.
- (196) Brenner, S.; Lerner, R. A. Encoded Combinatorial Chemistry. *Proc. Natl. Acad. Sci.* **1992**, *89* (12), 5381–5383. <https://doi.org/10.1073/pnas.89.12.5381>.
- (197) Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. Generation and Screening of an Oligonucleotide-Encoded Synthetic Peptide Library. *Proc. Natl. Acad. Sci.* **1993**, *90* (22), 10700–10704. <https://doi.org/10.1073/pnas.90.22.10700>.
- (198) Nielsen, J.; Brenner, S.; Janda, K. D. Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry. *J. Am. Chem. Soc.* **1993**, *115* (21), 9812–9813. <https://doi.org/10.1021/ja00074a063>.
- (199) Gartner, Z. J. DNA-Templated Organic Synthesis and Selection of a Library of Macrocycles. *Science* **2004**, *305* (5690), 1601–1605. (211) <https://doi.org/10.1126/science.1102629>.
- (200) Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. Encoded Self-Assembling Chemical Libraries. *Nat. Biotechnol.* **2004**, *22* (5), 568–574. <https://doi.org/10.1038/nbt961>.
- (201) Halpin, D. R.; Harbury, P. B. DNA Display I. Sequence-Encoded Routing of DNA Populations. *PLoS Biol.* **2004**, *2* (7), e173. <https://doi.org/10.1371/journal.pbio.0020173>.
- Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuzzo, J. W.; Davie, C. P.; Ding, Y.; Franklin, G. J.; Franzen, K. D.; Gefter, M. L.; Hale, S. P.; Hansen, N. J. V.; Israel, D. I.; Jiang, J.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'Keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L.; van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A. Design, Synthesis and Selection of DNA-Encoded Small-Molecule Libraries. *Nat. Chem. Biol.* **2009**, *5* (9), 647–654. <https://doi.org/10.1038/nchembio.211>.
- (203) Buller, F.; Mannocci, L.; Zhang, Y.; Dumelin, C. E.; Scheuermann, J.; Neri, D. Design and Synthesis of a Novel DNA-Encoded Chemical Library Using Diels-Alder Cycloadditions. *Bioorg. Med. Chem. Lett.* **2008**, *18* (22), 5926–5931. <https://doi.org/10.1016/j.bmcl.2008.07.038>.
- (204) Mannocci, L.; Zhang, Y.; Scheuermann, J.; Leimbacher, M.; De Bellis, G.; Rizzi, E.; Dumelin, C.; Melkko, S.; Neri, D. High-Throughput Sequencing Allows the Identification of Binding Molecules Isolated from DNA-Encoded Chemical Libraries. *Proc. Natl. Acad. Sci.* **2008**, *105* (46), 17670–17675. <https://doi.org/10.1073/pnas.0805130105>.
- (205) Buller, F.; Zhang, Y.; Scheuermann, J.; Schäfer, J.; Bühlmann, P.; Neri, D. Discovery of TNF Inhibitors from a DNA-Encoded Chemical Library Based on Diels-Alder Cycloaddition. *Chem. Biol.* **2009**, *16* (10), 1075–1086. <https://doi.org/10.1016/j.chembiol.2009.09.011>.
- (206) Mannocci, L.; Melkko, S.; Buller, F.; Molnár, I.; Gapián Bianké, J.-P.; Dumelin, C. E.; Scheuermann, J.; Neri, D. Isolation of Potent and Specific Trypsin Inhibitors from a DNA-Encoded Chemical Library. *Bioconjug. Chem.* **2010**, *21* (10), 1836–1841. <https://doi.org/10.1021/bc100198x>.
- (207) Buller, F.; Steiner, M.; Frey, K.; Mircsof, D.; Scheuermann, J.; Kalisch, M.; Bühlmann, P.; Supuran, C. T.; Neri, D. Selection of Carbonic Anhydrase IX Inhibitors from One Million DNA-Encoded Compounds. *ACS Chem. Biol.* **2011**, *6* (4), 336–344. <https://doi.org/10.1021/cb1003477>.
- (208) Leimbacher, M.; Zhang, Y.; Mannocci, L.; Stravs, M.; Geppert, T.; Scheuermann, J.; Schneider, G.; Neri, D. Discovery of Small-Molecule Interleukin-2 Inhibitors from a DNA-Encoded Chemical Library. *Chem. - Eur. J.* **2012**, *18* (25), 7729–7737. <https://doi.org/10.1002/chem.201200952>.
- (209) Gartner, Z. J.; Liu, D. R. The Generality of DNA-Templated Synthesis as a Basis for Evolving Non-Natural Small Molecules. *J. Am. Chem. Soc.* **2001**, *123* (28), 6961–6963. <https://doi.org/10.1021/ja015873n>.
- (210) Li, X.; Liu, D. R. DNA-Templated Organic Synthesis: Nature's Strategy for Controlling Chemical Reactivity Applied to Synthetic Molecules. *Angew. Chem. Int. Ed.* **2004**, *43* (37), 4848–4870. <https://doi.org/10.1002/anie.200400656>.
- (211) Li, X.; Gartner, Z. J.; Tse, B. N.; Liu, D. R. Translation of DNA into Synthetic N-Acyloxazolidines. *3*.
- (212) Li, Y.; Zhao, P.; Zhang, M.; Zhao, X.; Li, X. Multistep DNA-Templated Synthesis Using a Universal Template. *J. Am. Chem. Soc.* **2013**, *135* (47), 17727–17730. <https://doi.org/10.1021/ja409936r>.
- (213) Hansen, M. H.; Blakskjær, P.; Petersen, L. K.; Hansen, T. H.; Højfeldt, J. W.; Gothelf, K. V.; Hansen, N. J. V. A

- Yoctoliter-Scale DNA Reactor for Small-Molecule Evolution. *J. Am. Chem. Soc.* **2009**, *131* (3), 1322–1327. <https://doi.org/10.1021/ja088558a>.
- (214) Liu, Y.; West, S. C. Happy Hollidays: 40th Anniversary of the Holliday Junction. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (11), 937–944. <https://doi.org/10.1038/nrm1502>.
- (215) Halpin, D. R.; Harbury, P. B. DNA Display II. Genetic Manipulation of Combinatorial Chemistry Libraries for Small-Molecule Evolution. *PLoS Biol.* **2004**, *2* (7), e174. <https://doi.org/10.1371/journal.pbio.0020174>.
- (216) Halpin, D. R.; Lee, J. A.; Wrenn, S. J.; Harbury, P. B. DNA Display III. Solid-Phase Organic Synthesis on Unprotected DNA. *PLoS Biol.* **2004**, *2* (7), e175. <https://doi.org/10.1371/journal.pbio.0020175>.
- (217) Wrenn, S. J.; Weisinger, R. M.; Halpin, D. R.; Harbury, P. B. Synthetic Ligands Discovered by in Vitro Selection. *J. Am. Chem. Soc.* **2007**, *129* (43), 13137–13143. <https://doi.org/10.1021/ja073993a>.
- (218) Zambaldo, C.; Barluenga, S.; Winssinger, N. PNA-Encoded Chemical Libraries. *Curr. Opin. Chem. Biol.* **2015**, *26*, 8–15. <https://doi.org/10.1016/j.cbpa.2015.01.005>.
- (219) Satz, A. L.; Cai, J.; Chen, Y.; Goodnow, R.; Gruber, F.; Kowalczyk, A.; Petersen, A.; Naderi-Oboodi, G.; Orzechowski, L.; Strelbel, Q. DNA Compatible Multistep Synthesis and Applications to DNA Encoded Libraries. *Bioconjug. Chem.* **2015**, *26* (8), 1623–1632. <https://doi.org/10.1021/acs.bioconjugchem.5b00239>.
- (220) Fitzgerald, P. R.; Paegel, B. M. DNA-Encoded Chemistry: Drug Discovery from a Few Good Reactions. *Chem. Rev.* **2021**, *121* (12), 7155–7177. <https://doi.org/10.1021/acs.chemrev.0c00789>.
- (221) Götte, K.; Chines, S.; Brunschweiler, A. Reaction Development for DNA-Encoded Library Technology: From Evolution to Revolution? *Tetrahedron Lett.* **2020**, *61* (22), 151889. <https://doi.org/10.1016/j.tetlet.2020.151889>.
- (222) McCafferty, J.; Griffiths, A. D.; Winter, G.; Chiswell, D. J. Phage Antibodies: Filamentous Phage Displaying Antibody Variable Domains. *Nature* **1990**, *348* (6301), 552–554. <https://doi.org/10.1038/348552a0>.
- (223) Huse, W.; Sastry, L.; Iverson, S.; Kang, A.; Altling-Mees, M.; Burton, D.; Benkovic, S.; Lerner, R. Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire in Phage Lambda. *Science* **1989**, *246* (4935), 1275–1281. <https://doi.org/10.1126/science.2531466>.
- (224) Marks, J. D.; Hoogenboom, H. R.; Bonnert, T. P.; McCafferty, J.; Griffiths, A. D.; Winter, G. By-Passing Immunization. *J. Mol. Biol.* **1991**, *222* (3), 581–597. [https://doi.org/10.1016/0022-2836\(91\)90498-U](https://doi.org/10.1016/0022-2836(91)90498-U).
- (225) Kang, A. S.; Barbas, C. F.; Janda, K. D.; Benkovic, S. J.; Lerner, R. A. Linkage of Recognition and Replication Functions by Assembling Combinatorial Antibody Fab Libraries along Phage Surfaces. *Proc. Natl. Acad. Sci.* **1991**, *88* (10), 4363–4366. <https://doi.org/10.1073/pnas.88.10.4363>.
- (226) Griffiths, A. D.; Williams, S. C.; Hartley, O.; Tomlinson, I. M.; Waterhouse, P.; Crosby, W. L.; Kontermann, R. E.; Jones, P. T.; Low, N. M.; Allison, T. J. Isolation of High Affinity Human Antibodies Directly from Large Synthetic Repertoires. *EMBO J.* **1994**, *13* (14), 3245–3260. <https://doi.org/10.1002/j.1460-2075.1994.tb06626.x>.
- (227) Plückthun, A. Ribosome Display: A Perspective. In *Ribosome Display and Related Technologies*; Douthwaite, J. A., Jackson, R. H., Eds.; Methods in Molecular Biology; Springer New York: New York, NY, 2012; Vol. 805, pp 3–28. https://doi.org/10.1007/978-1-61779-379-0_1.
- (228) Gironde-Martínez, A.; Donckele, E. J.; Samain, F.; Neri, D. DNA-Encoded Chemical Libraries: A Comprehensive Review with Successful Stories and Future Challenges. *ACS Pharmacol. Transl. Sci.* **2021**, *4* (4), 1265–1279. <https://doi.org/10.1021/acspstci.1c00118>.
- (229) Arnau, J.; Lauritzen, C.; Petersen, G. E.; Pedersen, J. Current Strategies for the Use of Affinity Tags and Tag Removal for the Purification of Recombinant Proteins. *Protein Expr. Purif.* **2006**, *48* (1), 1–13. <https://doi.org/10.1016/j.pep.2005.12.002>.
- (230) Kimple, M. E.; Brill, A. L.; Pasker, R. L. Overview of Affinity Tags for Protein Purification. *Curr. Protoc. Protein Sci.* **2013**, *73* (1). <https://doi.org/10.1002/0471140864.ps0909s73>.
- (231) Chapman-Smith, A.; Cronan, J. E. The Enzymatic Biotinylation of Proteins: A Post-Translational Modification of Exceptional Specificity. *Trends Biochem. Sci.* **1999**, *24* (9), 359–363. [https://doi.org/10.1016/S0968-0004\(99\)01438-3](https://doi.org/10.1016/S0968-0004(99)01438-3).
- (232) Cull, M. G.; Schatz, P. J. [26] Biotinylation of Proteins in Vivo and in Vitro Using Small Peptide Tags. In *Methods in Enzymology*; Elsevier, 2000; Vol. 326, pp 430–440. [https://doi.org/10.1016/S0076-6879\(00\)26068-0](https://doi.org/10.1016/S0076-6879(00)26068-0).
- (233) Rusmini, F.; Zhong, Z.; Feijen, J. Protein Immobilization Strategies for Protein Biochips. *Biomacromolecules* **2007**, *8* (6), 1775–1789. <https://doi.org/10.1021/bm061197b>.
- (234) Wong, L. S.; Khan, F.; Micklefield, J. Selective Covalent Protein Immobilization: Strategies and Applications. *Chem. Rev.* **2009**, *109* (9), 4025–4053. <https://doi.org/10.1021/cr8004668>.
- (235) Decurtins, W.; Wichert, M.; Franzini, R. M.; Buller, F.; Stravs, M. A.; Zhang, Y.; Neri, D.; Scheuermann, J. Automated Screening for Small Organic Ligands Using DNA-Encoded Chemical Libraries. *Nat. Protoc.* **2016**, *11* (4), 764–780. <https://doi.org/10.1038/nprot.2016.039>.
- (236) Dumelin, C. E.; Scheuermann, J.; Melkko, S.; Neri, D. Selection of Streptavidin Binders from a DNA-Encoded Chemical Library. *Bioconjug. Chem.* **2006**, *17* (2), 366–370. <https://doi.org/10.1021/bc050282y>.
- (237) McGregor, L. M.; Gorin, D. J.; Dumelin, C. E.; Liu, D. R. Interaction-Dependent PCR: Identification of Ligand–Target Pairs from Libraries of Ligands and Libraries of Targets in a Single Solution-Phase Experiment. *J. Am. Chem. Soc.* **2010**, *132* (44), 15522–15524. <https://doi.org/10.1021/ja107677q>.
- (238) McGregor, L. M.; Jain, T.; Liu, D. R. Identification of Ligand–Target Pairs from Combined Libraries of Small Molecules and Unpurified Protein Targets in Cell Lysates. *J. Am. Chem. Soc.* **2014**, *136* (8), 3264–3270. <https://doi.org/10.1021/ja412934t>.
- (239) Chan, A. I.; McGregor, L. M.; Jain, T.; Liu, D. R. Discovery of a Covalent Kinase Inhibitor from a DNA-Encoded Small-Molecule Library × Protein Library Selection. *J. Am. Chem. Soc.* **2017**, *139* (30), 10192–10195. <https://doi.org/10.1021/jacs.7b04880>.
- (240) Li, G.; Liu, Y.; Liu, Y.; Chen, L.; Wu, S.; Liu, Y.; Li, X. Photoaffinity Labeling of Small-Molecule-Binding Proteins by DNA-Templated Chemistry. *Angew. Chem. Int. Ed.* **2013**, *52* (36), 9544–9549. <https://doi.org/10.1002/anie.201302161>.
- (241) Zhao, P.; Chen, Z.; Li, Y.; Sun, D.; Gao, Y.; Huang, Y.; Li, X. Selection of DNA-Encoded Small Molecule Libraries Against Unmodified and Non-Immobilized

- Protein Targets. *Angew. Chem. Int. Ed.* **2014**, *53* (38), 10056–10059. <https://doi.org/10.1002/anie.201404830>.
- (242) Li, G.; Liu, Y.; Yu, X.; Li, X. Multivalent Photoaffinity Probe for Labeling Small Molecule Binding Proteins. *Bioconjug. Chem.* **2014**, *25* (6), 1172–1180. <https://doi.org/10.1021/bc500195w>.
- (243) Wu, Z.; Graybill, T. L.; Zeng, X.; Platchek, M.; Zhang, J.; Bodmer, V. Q.; Wisnoski, D. D.; Deng, J.; Coppo, F. T.; Yao, G.; Tamburino, A.; Scavello, G.; Franklin, G. J.; Mataruse, S.; Bedard, K. L.; Ding, Y.; Chai, J.; Summerfield, J.; Centrella, P. A.; Messer, J. A.; Pope, A. J.; Israel, D. I. Cell-Based Selection Expands the Utility of DNA-Encoded Small-Molecule Library Technology to Cell Surface Drug Targets: Identification of Novel Antagonists of the NK3 Tachykinin Receptor. *ACS Comb. Sci.* **2015**, *17* (12), 722–731. <https://doi.org/10.1021/acscombsci.5b00124>.
- (244) Cai, B.; Kim, D.; Akhand, S.; Sun, Y.; Cassell, R. J.; Alpsy, A.; Dykhuizen, E. C.; Van Rijn, R. M.; Wendt, M. K.; Krusemark, C. J. Selection of DNA-Encoded Libraries to Protein Targets within and on Living Cells. *J. Am. Chem. Soc.* **2019**, *141* (43), 17057–17061. <https://doi.org/10.1021/jacs.9b08085>.
- (245) Qian, Z.; Martyna, A.; Hard, R. L.; Wang, J.; Appiah-Kubi, G.; Coss, C.; Phelps, M. A.; Rossman, J. S.; Pei, D. Discovery and Mechanism of Highly Efficient Cyclic Cell-Penetrating Peptides. *Biochemistry* **2016**, *55* (18), 2601–2612. <https://doi.org/10.1021/acs.biochem.6b00226>.
- (246) Petersen, L. K.; Christensen, A. B.; Andersen, J.; Folkesson, C. G.; Kristensen, O.; Andersen, C.; Alzu, A.; Sløk, F. A.; Blakskjær, P.; Madsen, D.; Azevedo, C.; Micco, I.; Hansen, N. J. V. Screening of DNA-Encoded Small Molecule Libraries inside a Living Cell. *J. Am. Chem. Soc.* **2021**, *143* (7), 2751–2756. <https://doi.org/10.1021/jacs.0c09213>.
- (247) Dumelin, C. E.; Trüssel, S.; Buller, F.; Trachsel, E.; Bootz, F.; Zhang, Y.; Mannocci, L.; Beck, S. C.; Drumea-Mirancea, M.; Seeliger, M. W.; Baltes, C.; Müggler, T.; Kranz, F.; Rudin, M.; Melkko, S.; Scheuermann, J.; Neri, D. A Portable Albumin Binder from a DNA-Encoded Chemical Library. *Angew. Chem. Int. Ed.* **2008**, *47* (17), 3196–3201. <https://doi.org/10.1002/anie.200704936>.
- (248) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bembem, L. A.; Berka, J.; Braverman, M. S.; Chen, Y.-J.; Chen, Z.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Godwin, B. C.; He, W.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L. I.; Jarvie, T. P.; Jirage, K. B.; Kim, J.-B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S. M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.; Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. F.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P.; Begley, R. F.; Rothberg, J. M. Genome Sequencing in Microfabricated High-Density Picolitre Reactors. *Nature* **2005**, *437* (7057), 376–380. <https://doi.org/10.1038/nature03959>.
- (249) Pettersson, E.; Lundeberg, J.; Ahmadian, A. Generations of Sequencing Technologies. *Genomics* **2009**, *93* (2), 105–111. <https://doi.org/10.1016/j.ygeno.2008.10.003>.
- (250) Mardis, E. R. Next-Generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet.* **2008**, *9* (1), 387–402. <https://doi.org/10.1146/annurev.genom.9.081307.164359>.
- (251) Lu, Y.; Shen, Y.; Warren, W.; Walter, R. *Next Generation Sequencing in Aquatic Models*; IntechOpen, 2016. <https://doi.org/10.5772/61657>.
- (252) Buller, F.; Steiner, M.; Scheuermann, J.; Mannocci, L.; Nissen, I.; Kohler, M.; Beisel, C.; Neri, D. High-Throughput Sequencing for the Identification of Binding Molecules from DNA-Encoded Chemical Libraries. *Bioorg. Med. Chem. Lett.* **2010**, *20* (14), 4188–4192. <https://doi.org/10.1016/j.bmcl.2010.05.053>.
- (253) Wichert, M.; Krall, N.; Decurtins, W.; Franzini, R. M.; Pretto, F.; Schneider, P.; Neri, D.; Scheuermann, J. Dual-Display of Small Molecules Enables the Discovery of Ligand Pairs and Facilitates Affinity Maturation. *Nat. Chem.* **2015**, *7* (3), 241–249. <https://doi.org/10.1038/nchem.2158>.
- (254) Zimmermann, G.; Li, Y.; Rieder, U.; Mattarella, M.; Neri, D.; Scheuermann, J. Hit-Validation Methodologies for Ligands Isolated from DNA-Encoded Chemical Libraries. *ChemBioChem* **2017**, *18* (9), 853–857. <https://doi.org/10.1002/cbic.201600637>.
- (255) Franzini, R. M.; Randolph, C. Chemical Space of DNA-Encoded Libraries: Miniperspective. *J. Med. Chem.* **2016**, *59* (14), 6629–6644. <https://doi.org/10.1021/acs.jmedchem.5b01874>.
- (256) Samain, F.; Ekblad, T.; Mikutis, G.; Zhong, N.; Zimmermann, M.; Nauer, A.; Bajic, D.; Decurtins, W.; Scheuermann, J.; Brown, P. J.; Hall, J.; Gräslund, S.; Schüler, H.; Neri, D.; Franzini, R. M. Tankyrase 1 Inhibitors with Drug-like Properties Identified by Screening a DNA-Encoded Chemical Library. *J. Med. Chem.* **2015**, *58* (12), 5143–5149. <https://doi.org/10.1021/acs.jmedchem.5b00432>.
- (257) Ding, Y.; O’Keefe, H.; DeLorey, J. L.; Israel, D. I.; Messer, J. A.; Chiu, C. H.; Skinner, S. R.; Matico, R. E.; Murray-Thompson, M. F.; Li, F.; Clark, M. A.; Cuozzo, J. W.; Arico-Muendel, C.; Morgan, B. A. Discovery of Potent and Selective Inhibitors for ADAMTS-4 through DNA-Encoded Library Technology (ELT). *ACS Med. Chem. Lett.* **2015**, *6* (8), 888–893. <https://doi.org/10.1021/acsmedchemlett.5b00138>.
- (258) Petersen, L. K.; Blakskjær, P.; Chaikuad, A.; Christensen, A. B.; Dietvorst, J.; Holmkvist, J.; Knapp, S.; Kořínek, M.; Larsen, L. K.; Pedersen, A. E.; Röhm, S.; Sløk, F. A.; Hansen, N. J. V. Novel P38 α MAP Kinase Inhibitors Identified from YoctoReactor DNA-Encoded Small Molecule Library. *MedChemComm* **2016**, *7* (7), 1332–1339. <https://doi.org/10.1039/C6MD00241B>.
- (259) Usanov, D. L.; Chan, A. I.; Maianti, J. P.; Liu, D. R. Second-Generation DNA-Templated Macrocyclic Libraries for the Discovery of Bioactive Small Molecules. *Nat. Chem.* **2018**, *10* (7), 704–714. <https://doi.org/10.1038/s41557-018-0033-8>.
- (260) Wood, E. R.; Bledsoe, R.; Chai, J.; Daka, P.; Deng, H.; Ding, Y.; Harris-Gurley, S.; Krynn, L. H.; Nartey, E.; Nichols, J.; Nolte, R. T.; Prabhu, N.; Rise, C.; Sheahan, T.; Shotwell, J. B.; Smith, D.; Tai, V.; Taylor, J. D.; Tomberlin, G.; Wang, L.; Wisely, B.; You, S.; Xia, B.; Dickson, H. The Role of Phosphodiesterase 12 (PDE12) as a Negative Regulator of the Innate Immune Response and the Discovery of Antiviral Inhibitors. *J. Biol. Chem.* **2015**, *290* (32), 19681–19696. <https://doi.org/10.1074/jbc.M115.653113>.
- (261) Gilmartin, A. G.; Falt, T. H.; Richter, M.; Groy, A.; Seefeld, M. A.; Darcy, M. G.; Peng, X.; Federowicz, K.; Yang, J.; Zhang, S.-Y.; Minthorn, E.; Jaworski, J.-P.;

- Schaber, M.; Martens, S.; McNulty, D. E.; Sinnamon, R. H.; Zhang, H.; Kirkpatrick, R. B.; Nevins, N.; Cui, G.; Pietrak, B.; Diaz, E.; Jones, A.; Brandt, M.; Schwartz, B.; Heering, D. A.; Kumar, R. Allosteric Wip1 Phosphatase Inhibition through Flap-Subdomain Interaction. *Nat. Chem. Biol.* **2014**, *10* (3), 181–187. <https://doi.org/10.1038/nchembio.1427>.
- (262) Taylor, D. M.; Anglin, J.; Park, S.; Ucisik, M. N.; Faver, J. C.; Simmons, N.; Jin, Z.; Palaniappan, M.; Nyshadham, P.; Li, F.; Campbell, J.; Hu, L.; Sankaran, B.; Prasad, B. V. V.; Huang, H.; Matzuk, M. M.; Palzkill, T. Identifying Oxacillinase-48 Carbapenemase Inhibitors Using DNA-Encoded Chemical Libraries. *ACS Infect. Dis.* **2020**, *6* (5), 1214–1227. <https://doi.org/10.1021/acsinfecdis.0c00015>.
- (263) Ahn, S.; Pani, B.; Kahsai, A. W.; Olsen, E. K.; Husemoen, G.; Vestergaard, M.; Jin, L.; Zhao, S.; Wingler, L. M.; Rambarat, P. K.; Simhal, R. K.; Xu, T. T.; Sun, L. D.; Shim, P. J.; Staus, D. P.; Huang, L.-Y.; Franch, T.; Chen, X.; Lefkowitz, R. J. Small-Molecule Positive Allosteric Modulators of the β_2 -Adrenoceptor Isolated from DNA-Encoded Libraries. *Mol. Pharmacol.* **2018**, *94* (2), 850–861. <https://doi.org/10.1124/mol.118.111948>.
- (264) Dawadi, S.; Simmons, N.; Miklossy, G.; Bohren, K. M.; Faver, J. C.; Ucisik, M. N.; Nyshadham, P.; Yu, Z.; Matzuk, M. M. Discovery of Potent Thrombin Inhibitors from a Protease-Focused DNA-Encoded Chemical Library. *Proc. Natl. Acad. Sci.* **2020**, *117* (29), 16782–16789. <https://doi.org/10.1073/pnas.2005447117>.
- (265) Kunig, V. B. K.; Potowski, M.; Akbarzadeh, M.; Klika Škopić, M.; Santos Smith, D.; Arendt, L.; Dormuth, I.; Adihou, H.; Andlovic, B.; Karatas, H.; Shaabani, S.; Zarganes-Tzitzikas, T.; Neochoritis, C. G.; Zhang, R.; Groves, M.; Guéret, S. M.; Ottmann, C.; Rahnenführer, J.; Fried, R.; Dömling, A.; Brunschweiler, A. TEAD–YAP Interaction Inhibitors and MDM2 Binders from DNA-Encoded Indole-Focused Ugi Peptidomimetics. *Angew. Chem. Int. Ed.* **2020**, *59* (46), 20338–20342. <https://doi.org/10.1002/anie.202006280>.
- (266) Shin, M. H.; Lee, K. J.; Lim, H.-S. DNA-Encoded Combinatorial Library of Macrocyclic Peptoids. *Bioconjug. Chem.* **2019**, *30* (11), 2931–2938. <https://doi.org/10.1021/acs.bioconjchem.9b00628>.
- (267) Bassi, G.; Favalli, N.; Vuk, M.; Catalano, M.; Martinelli, A.; Trenner, A.; Porro, A.; Yang, S.; Tham, C. L.; Moroglu, M.; Yue, W. W.; Conway, S. J.; Vogt, P. K.; Sartori, A. A.; Scheuermann, J.; Neri, D. A Single-Stranded DNA-Encoded Chemical Library Based on a Stereoisomeric Scaffold Enables Ligand Discovery by Modular Assembly of Building Blocks. *Adv. Sci.* **2020**, *7* (22), 2001970. <https://doi.org/10.1002/advs.202001970>.
- (268) Stress, C. J.; Sauter, B.; Schneider, L. A.; Sharpe, T.; Gillingham, D. A DNA-Encoded Chemical Library Incorporating Elements of Natural Macrocycles. *Angew. Chem. Int. Ed.* **2019**, *58* (28), 9570–9574. <https://doi.org/10.1002/anie.201902513>.
- (269) Koesema, E.; Roy, A.; Paciaroni, N. G.; Kodadek, T. Synthesis and Screening of A DNA-Encoded Library of Non-Peptidic Macrocycles. **8**.
- (270) Rizvi, N. F.; Smith, G. F. RNA as a Small Molecule Druggable Target. *Bioorg. Med. Chem. Lett.* **2017**, *27* (23), 5083–5088. <https://doi.org/10.1016/j.bmcl.2017.10.052>.
- (271) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **1997**, *23* (1–3), 3–25. [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1).
- (272) Noe, M. C.; Peakman, M.-C. Drug Discovery Technologies: Current and Future Trends. In *Comprehensive Medicinal Chemistry III*; Elsevier, 2017; pp 1–32. <https://doi.org/10.1016/B978-0-12-409547-2.12312-1>.
- (273) Harris, P. A.; Berger, S. B.; Jeong, J. U.; Nagilla, R.; Bandyopadhyay, D.; Campobasso, N.; Capriotti, C. A.; Cox, J. A.; Dare, L.; Dong, X.; Eidam, P. M.; Finger, J. N.; Hoffman, S. J.; Kang, J.; Kasparcova, V.; King, B. W.; Lehr, R.; Lan, Y.; Leister, L. K.; Lich, J. D.; MacDonald, T. T.; Miller, N. A.; Ouellette, M. T.; Pao, C. S.; Rahman, A.; Reilly, M. A.; Rendina, A. R.; Rivera, E. J.; Schaeffer, M. C.; Sehon, C. A.; Singhaus, R. R.; Sun, H. H.; Swift, B. A.; Totoritis, R. D.; Vossenkämper, A.; Ward, P.; Wisnoski, D. D.; Zhang, D.; Marquis, R. W.; Gough, P. J.; Bertin, J. Discovery of a First-in-Class Receptor Interacting Protein 1 (RIP1) Kinase Specific Clinical Candidate (GSK2982772) for the Treatment of Inflammatory Diseases. *J. Med. Chem.* **2017**, *60* (4), 1247–1261. <https://doi.org/10.1021/acs.jmedchem.6b01751>.
- (274) Madsen, D.; Azevedo, C.; Micco, I.; Petersen, L. K.; Hansen, N. J. V. An Overview of DNA-Encoded Libraries: A Versatile Tool for Drug Discovery. In *Progress in Medicinal Chemistry*; Elsevier, 2020; Vol. 59, pp 181–249. <https://doi.org/10.1016/bs.pmch.2020.03.001>.
- (275) Zhu, H.; Flanagan, M. E.; Stanton, R. V. Designing DNA Encoded Libraries of Diverse Products in a Focused Property Space. *J. Chem. Inf. Model.* **2019**, *59* (11), 4645–4653. <https://doi.org/10.1021/acs.jcim.9b00729>.
- (276) Onda, Y.; Bassi, G.; Elsayed, A.; Ulrich, F.; Oehler, S.; Plais, L.; Scheuermann, J.; Neri, D. A DNA-Encoded Chemical Library Based on Peptide Macrocycles. *Chem. – Eur. J.* **2021**, *27* (24), 7160–7167. <https://doi.org/10.1002/chem.202005423>.
- (277) Cuozzo, J. W.; Centrella, P. A.; Gikunju, D.; Habeshian, S.; Hupp, C. D.; Keefe, A. D.; Sigel, E. A.; Soutter, H. H.; Thomson, H. A.; Zhang, Y.; Clark, M. A. Discovery of a Potent BTK Inhibitor with a Novel Binding Mode by Using Parallel Selections with a DNA-Encoded Chemical Library. *ChemBioChem* **2017**, *18* (9), 864–871. <https://doi.org/10.1002/cbic.201600573>.
- (278) Deng, H.; Zhou, J.; Sandersingh, F.; Messer, J. A.; Somers, D. O.; Ajakane, M.; Arico-Muendel, C. C.; Beljean, A.; Belyanskaya, S. L.; Bingham, R.; Blazensky, E.; Boullay, A.-B.; Boursier, E.; Chai, J.; Carter, P.; Chung, C.-W.; Daugan, A.; Ding, Y.; Herry, K.; Hobbs, C.; Humphries, E.; Kollmann, C.; Nguyen, V. L.; Nicodeme, E.; Smith, S. E.; Dodic, N.; Ancellin, N. Discovery and Optimization of Potent, Selective, and *in Vivo* Efficacious 2-Aryl Benzimidazole BCATm Inhibitors. *ACS Med. Chem. Lett.* **2016**, *7* (4), 379–384. <https://doi.org/10.1021/acsmedchemlett.5b00389>.
- (279) Ahn, S.; Kahsai, A. W.; Pani, B.; Wang, Q.-T.; Zhao, S.; Wall, A. L.; Strachan, R. T.; Staus, D. P.; Wingler, L. M.; Sun, L. D.; Sinnaeve, J.; Choi, M.; Cho, T.; Xu, T. T.; Hansen, G. M.; Burnett, M. B.; Lamerdin, J. E.; Bassoni, D. L.; Gavino, B. J.; Husemoen, G.; Olsen, E. K.; Franch, T.; Costanzi, S.; Chen, X.; Lefkowitz, R. J. Allosteric “Beta-Blocker” Isolated from a DNA-Encoded Small Molecule Library. *Proc. Natl. Acad. Sci.* **2017**, *114* (7), 1708–1713. <https://doi.org/10.1073/pnas.1620645114>.

- (280) Cheng, R. K. Y.; Fiez-Vandal, C.; Schlenker, O.; Edman, K.; Aggeler, B.; Brown, D. G.; Brown, G. A.; Cooke, R. M.; Dumelin, C. E.; Doré, A. S.; Geschwindner, S.; Grebner, C.; Hermansson, N.-O.; Jazayeri, A.; Johansson, P.; Leong, L.; Prihandoko, R.; Rappas, M.; Soutter, H.; Snijder, A.; Sundström, L.; Tehan, B.; Thornton, P.; Troast, D.; Wiggin, G.; Zhukov, A.; Marshall, F. H.; Dekker, N. Structural Insight into Allosteric Modulation of Protease-Activated Receptor 2. *Nature* **2017**, *545* (7652), 112–115. <https://doi.org/10.1038/nature22309>.
- (281) Brown, D. G.; Brown, G. A.; Centrella, P.; Certel, K.; Cooke, R. M.; Cuozzo, J. W.; Dekker, N.; Dumelin, C. E.; Ferguson, A.; Fiez-Vandal, C.; Geschwindner, S.; Guié, M.-A.; Habeshian, S.; Keefe, A. D.; Schlenker, O.; Sigel, E. A.; Snijder, A.; Soutter, H. T.; Sundström, L.; Troast, D. M.; Wiggin, G.; Zhang, J.; Zhang, Y.; Clark, M. A. Agonists and Antagonists of Protease-Activated Receptor 2 Discovered within a DNA-Encoded Chemical Library Using Mutational Stabilization of the Target. *SLAS Discov. Adv. Sci. Drug Discov.* **2018**, *23* (5), 429–436. <https://doi.org/10.1177/2472555217749847>.
- (282) Seigal, B. A.; Connors, W. H.; Fraley, A.; Borzilleri, R. M.; Carter, P. H.; Emanuel, S. L.; Fargnoli, J.; Kim, K.; Lei, M.; Naglich, J. G.; Pokross, M. E.; Posy, S. L.; Shen, H.; Surti, N.; Talbott, R.; Zhang, Y.; Terrett, N. K. The Discovery of Macrocyclic XIAP Antagonists from a DNA-Programmed Chemistry Library, and Their Optimization To Give Lead Compounds with in Vivo Antitumor Activity. *J. Med. Chem.* **2015**, *58* (6), 2855–2861. <https://doi.org/10.1021/jm501892g>.
- (283) Fernández-Montalván, A. E.; Berger, M.; Kuroppa, B.; Koo, S. J.; Badock, V.; Weiske, J.; Puetter, V.; Holton, S. J.; Stöckigt, D.; ter Laak, A.; Centrella, P. A.; Clark, M. A.; Dumelin, C. E.; Sigel, E. A.; Soutter, H. H.; Troast, D. M.; Zhang, Y.; Cuozzo, J. W.; Keefe, A. D.; Roche, D.; Rodeschini, V.; Chaikuad, A.; Díaz-Sáez, L.; Bennett, J. M.; Fedorov, O.; Huber, K. V. M.; Hübner, J.; Weinmann, H.; Hartung, I. V.; Gorjánác, M. Isoform-Selective ATAD2 Chemical Probe with Novel Chemical Structure and Unusual Mode of Action. *ACS Chem. Biol.* **2017**, *12* (11), 2730–2736. <https://doi.org/10.1021/acscchembio.7b00708>.
- (284) Creaser, S. P.; Acharya, R. A. Exercises in the Synthesis of DNA-Encoded Libraries. In *A Handbook for DNA-Encoded Chemistry*; Goodnow, R. A., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2014; pp 123–151. <https://doi.org/10.1002/9781118832738.ch6>.
- (285) Dickson, P.; Kodadek, T. Chemical Composition of DNA-Encoded Libraries, Past Present and Future. *Org. Biomol. Chem.* **2019**, *17* (19), 4676–4688. <https://doi.org/10.1039/C9OB00581A>.
- (286) Method for the Synthesis of a Bifunctional Complex.
- (287) Franzini, R. M.; Biendl, S.; Mikutis, G.; Samain, F.; Scheuermann, J.; Neri, D. “Cap-and-Catch” Purification for Enhancing the Quality of Libraries of DNA Conjugates. *ACS Comb. Sci.* **2015**, *17* (7), 393–398. <https://doi.org/10.1021/acscombsci.5b00072>.
- (288) Chakraborty, T. K.; Arora, A.; Roy, S.; Kumar, N.; Maiti, S. Furan Based Cyclic Oligopeptides Selectively Target G-Quadruplex. *J. Med. Chem.* **2007**, *50* (23), 5539–5542. <https://doi.org/10.1021/jm070619c>.
- (289) Shirude, P. S.; Gillies, E. R.; Ladame, S.; Godde, F.; Shin-ya, K.; Huc, I.; Balasubramanian, S. Macrocyclic and Helical Oligoamides as a New Class of G-Quadruplex Ligands. *J. Am. Chem. Soc.* **2007**, *129* (39), 11890–11891. <https://doi.org/10.1021/ja073775h>.
- (290) Hunter, J. H.; Anderson, M. J.; Castan, I. F. S. F.; Graham, J. S.; Salvini, C. L. A.; Stanway-Gordon, H. A.; Crawford, J. J.; Madin, A.; Pairedeau, G.; Waring, M. J. Highly Efficient On-DNA Amide Couplings Promoted by Micelle Forming Surfactants for the Synthesis of DNA Encoded Libraries. *Chem. Sci.* **2021**, *12* (27), 9475–9484. <https://doi.org/10.1039/D1SC03007H>.
- (291) Franzini, R. M.; Nauer, A.; Scheuermann, J.; Neri, D. Interrogating Target-Specificity by Parallel Screening of a DNA-Encoded Chemical Library against Closely Related Proteins. *Chem. Commun.* **2015**, *51* (38), 8014–8016. <https://doi.org/10.1039/C5CC01230A>.
- (292) Franzini, R. M.; Ekblad, T.; Zhong, N.; Wichert, M.; Decurtins, W.; Nauer, A.; Zimmermann, M.; Samain, F.; Scheuermann, J.; Brown, P. J.; Hall, J.; Gräslund, S.; Schüler, H.; Neri, D. Identification of Structure-Activity Relationships from Screening a Structurally Compact DNA-Encoded Chemical Library. *Angew. Chem. Int. Ed.* **2015**, *54* (13), 3927–3931. <https://doi.org/10.1002/anie.201410736>.
- (293) Bornhorst, J. A.; Falke, J. J. [16] Purification of Proteins Using Polyhistidine Affinity Tags. In *Methods in Enzymology*; Elsevier, 2000; Vol. 326, pp 245–254. [https://doi.org/10.1016/S0076-6879\(00\)26058-8](https://doi.org/10.1016/S0076-6879(00)26058-8).
- (294) Terpe, K. Overview of Tag Protein Fusions: From Molecular and Biochemical Fundamentals to Commercial Systems. *Appl. Microbiol. Biotechnol.* **2003**, *60* (5), 523–533. <https://doi.org/10.1007/s00253-002-1158-6>.
- (295) Waugh, D. S. Making the Most of Affinity Tags. *Trends Biotechnol.* **2005**, *23* (6), 316–320. <https://doi.org/10.1016/j.tibtech.2005.03.012>.
- (296) Köhn, M. Immobilization Strategies for Small Molecule, Peptide and Protein Microarrays. *J. Pept. Sci.* **2009**, *15* (6), 393–397. <https://doi.org/10.1002/psc.1130>.
- (297) Satz, A. L. DNA Encoded Library Selections and Insights Provided by Computational Simulations. *ACS Chem. Biol.* **2015**, *10* (10), 2237–2245. <https://doi.org/10.1021/acscchembio.5b00378>.
- (298) Fonnum, G.; Johansson, C.; Molteberg, A.; Mørup, S.; Aksnes, E. Characterisation of Dynabeads® by Magnetization Measurements and Mössbauer Spectroscopy. *J. Magn. Magn. Mater.* **2005**, *293* (1), 41–47. <https://doi.org/10.1016/j.jmmm.2005.01.041>.
- (299) 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>.
- (300) Piekna-Przybylska, D.; Sullivan, M. A.; Sharma, G.; Bambara, R. A. U3 Region in the HIV-1 Genome Adopts a G-Quadruplex Structure in Its RNA and DNA Sequence. *Biochemistry* **2014**, *53* (16), 2581–2593. <https://doi.org/10.1021/bi4016692>.
- (301) 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>.
- (302) Kaiser, C. E.; Gokhale, V.; Yang, D.; Hurley, L. H. Gaining Insights into the Small Molecule Targeting of the G-Quadruplex in the c-MYC Promoter Using NMR and an Allele-Specific Transcriptional Assay. In *Quadruplex Nucleic Acids*; Chaires, J. B., Graves, D., Eds.; Topics in Current Chemistry; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; Vol. 330, pp 1–21. https://doi.org/10.1007/128_2012_333.

- (303) Lannes, L.; Halder, S.; Krishnan, Y.; Schwalbe, H. Tuning the PH Response of I-Motif DNA Oligonucleotides. *ChemBioChem* **2015**, *16* (11), 1647–1656. <https://doi.org/10.1002/cbic.201500182>.
- (304) Brahms, J.; Mommaerts, W. F. H. M. A Study of Conformation of Nucleic Acids in Solution by Means of Circular Dichroism. *J. Mol. Biol.* **1964**, *10* (1), 73–88. [https://doi.org/10.1016/S0022-2836\(64\)80029-2](https://doi.org/10.1016/S0022-2836(64)80029-2).
- (305) Dzatko, S.; Krafčíková, M.; Hänsel-Hertsch, R.; Fessler, T.; Fiala, R.; Loja, T.; Krafčík, D.; Mergny, J.-L.; Foldynova-Trantírková, S.; Trantírek, L. Evaluation of the Stability of DNA I-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chem. Int. Ed.* **2018**, *57* (8), 2165–2169. <https://doi.org/10.1002/anie.201712284>.
- (306) Potapov, V.; Ong, J. L. Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLOS ONE* **2017**, *12* (1), e0169774. <https://doi.org/10.1371/journal.pone.0169774>.
- (307) Santos, T.; Miranda, A.; Campello, M. P. C.; Paulo, A.; Salgado, G.; Cabrita, E. J.; Cruz, C. Recognition of Nucleolin through Interaction with RNA G-Quadruplex. *Biochem. Pharmacol.* **2021**, *189*, 114208. <https://doi.org/10.1016/j.bcp.2020.114208>.
- (308) Carvalho, J.; Santos, T.; Carrilho, R.; Sousa, F.; Salgado, G. F.; Queiroz, J. A.; Cruz, C. Ligand Screening to Pre-MiRNA 149 G-Quadruplex Investigated by Molecular Dynamics. *J. Biomol. Struct. Dyn.* **2020**, *38* (8), 2276–2286. <https://doi.org/10.1080/07391102.2019.1632743>.
- (309) Müller, S.; Pantoş, G. D.; Rodriguez, R.; Balasubramanian, S. Controlled-Folding of a Small Molecule Modulates DNA G-Quadruplex Recognition. *Chem Commun* **2009**, No. 1, 80–82. <https://doi.org/10.1039/B816861J>.
- (310) Lefebvre, J.; Guetta, C.; Poyer, F.; Mahuteau-Betzer, F.; Teulade-Fichou, M.-P. Copper-Alkyne Complexation Responsible for the Nucleolar Localization of Quadruplex Nucleic Acid Drugs Labeled by Click Reactions. *Angew. Chem. Int. Ed.* **2017**, *56* (38), 11365–11369. <https://doi.org/10.1002/anie.201703783>.
- (311) Merrifield, R. B. **Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide.** *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154. <https://doi.org/10.1021/ja00897a025>.
- (312) Knorr, R.; Trzeciak, A.; Bannwarth, W.; Gillissen, D. New Coupling Reagents in Peptide Chemistry. *Tetrahedron Lett.* **1989**, *30* (15), 1927–1930. [https://doi.org/10.1016/S0040-4039\(00\)99616-3](https://doi.org/10.1016/S0040-4039(00)99616-3).
- (313) Valeur, E.; Bradley, M. Amide Bond Formation: Beyond the Myth of Coupling Reagents. *Chem Soc Rev* **2009**, *38* (2), 606–631. <https://doi.org/10.1039/B701677H>.
- (314) Goddard-Borger, E. D.; Stick, R. V. An Efficient, Inexpensive, and Shelf-Stable Diazotransfer Reagent: Imidazole-1-Sulfonyl Azide Hydrochloride. *Org. Lett.* **2007**, *9* (19), 3797–3800. <https://doi.org/10.1021/ol701581g>.
- (315) Huisgen, R.; Szeimies, G.; Möbius, L. 1,3-Dipolare Cycloadditionen, XXXII. Kinetik der Additionen organischer Azide an CC-Mehrfachbindungen. *Chem. Ber.* **1967**, *100* (8), 2494–2507. <https://doi.org/10.1002/cber.19671000806>.
- (316) Huisgen, R. Kinetics and Reaction Mechanisms: Selected Examples from the Experience of Forty Years. *Pure Appl. Chem.* **1989**, *61* (4), 613–628. <https://doi.org/10.1351/pac198961040613>.
- (317) Tornøe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* **2002**, *67* (9), 3057–3064. <https://doi.org/10.1021/jo011148j>.
- (318) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective “Ligation” of Azides and Terminal Alkynes. *Angew. Chem. Int. Ed.* **2002**, *41* (14), 2596–2599. [https://doi.org/10.1002/1521-3773\(20020715\)41:14<2596::AID-ANIE2596>3.0.CO;2-4](https://doi.org/10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4).
- (319) Rosner, T.; Le Bars, J.; Pfaltz, A.; Blackmond, D. G. Kinetic Studies of Heck Coupling Reactions Using Palladacycle Catalysts: Experimental and Kinetic Modeling of the Role of Dimer Species. *J. Am. Chem. Soc.* **2001**, *123* (9), 1848–1855. <https://doi.org/10.1021/ja003191e>.
- (320) Straub, B. F. M-Acetylide and μ -Alkenylidene Ligands in “Click” Triazole Syntheses. *Chem. Commun.* **2007**, No. 37, 3868. <https://doi.org/10.1039/b706926j>.
- (321) Meldal, M.; Tornøe, C. W. Cu-Catalyzed Azide-Alkyne Cycloaddition. *Chem. Rev.* **2008**, *108* (8), 2952–3015. <https://doi.org/10.1021/cr0783479>.
- (322) Conibear, A. C.; Farbiarz, K.; Mayer, R. L.; Matveenko, M.; Kählig, H.; Becker, C. F. W. Arginine Side-Chain Modification That Occurs during Copper-Catalyzed Azide-Alkyne Click Reactions Resembles an Advanced Glycation End Product. *Org. Biomol. Chem.* **2016**, *14* (26), 6205–6211. <https://doi.org/10.1039/C6OB00932H>.
- (323) Sekar, R. B.; Periasamy, A. Fluorescence Resonance Energy Transfer (FRET) Microscopy Imaging of Live Cell Protein Localizations. *J. Cell Biol.* **2003**, *160* (5), 629–633. <https://doi.org/10.1083/jcb.200210140>.
- (324) Mergny, J.-L.; Maurizot, J.-C. Fluorescence Resonance Energy Transfer as a Probe for G-Quartet Formation by a Telomeric Repeat. *ChemBioChem* **2001**, *2* (2), 124–132. [https://doi.org/10.1002/1439-7633\(20010202\)2:2<124::AID-CBIC124>3.0.CO;2-L](https://doi.org/10.1002/1439-7633(20010202)2:2<124::AID-CBIC124>3.0.CO;2-L).
- (325) Mergny, J.-L.; Lacroix, L.; Teulade-Fichou, M.-P.; Hounsou, C.; Guittat, L.; Hoarau, M.; Arimondo, P. B.; Vigneron, J.-P.; Lehn, J.-M.; Riou, J.-F.; Garestier, T.; Helene, C. Telomerase Inhibitors Based on Quadruplex Ligands Selected by a Fluorescence Assay. *Proc. Natl. Acad. Sci.* **2001**, *98* (6), 3062–3067. <https://doi.org/10.1073/pnas.051620698>.
- (326) Saini, S.; Singh, H.; Bagchi, B. Fluorescence Resonance Energy Transfer (FRET) in Chemistry and Biology: Non-Förster Distance Dependence of the FRET Rate. *J. Chem. Sci.* **2006**, *118* (1), 23–35. <https://doi.org/10.1007/BF02708762>.
- (327) Mergny, J.; Lacroix, L. UV Melting of G-Quadruplexes. *Curr. Protoc. Nucleic Acid Chem.* **2009**, *37* (1). <https://doi.org/10.1002/0471142700.nc1701s37>.
- (328) Lefebvre, J. Outils moléculaires pour l'étude des G-quadruplex au sein du génome. 215.
- (329) Le, D. D.; Di Antonio, M.; Chan, L. K. M.; Balasubramanian, S. G-Quadruplex Ligands Exhibit Differential G-Tetrad Selectivity. *Chem. Commun.* **2015**, *51* (38), 8048–8050. <https://doi.org/10.1039/C5CC02252E>.
- (330) Tanpure, A. A.; Srivatsan, S. G. Conformation-Sensitive Nucleoside Analogues as Topology-Specific Fluorescence Turn-on Probes for DNA and RNA G-Quadruplexes. *Nucleic Acids Res.* **2015**, *43* (22), e149–e149. <https://doi.org/10.1093/nar/gkv743>.
- (331) Sontakke, V. A.; Srivatsan, S. G. A Dual-App Nucleoside Probe Reports G-Quadruplex Formation and Ligand

- Binding in the Long Terminal Repeat of HIV-1 Proviral Genome. *Bioorg. Med. Chem. Lett.* **2020**, *30* (16), 127345. <https://doi.org/10.1016/j.bmcl.2020.127345>.
- (332) Koirala, D.; Dhakal, S.; Ashbridge, B.; Sannohe, Y.; Rodriguez, R.; Sugiyama, H.; Balasubramanian, S.; Mao, H. A Single-Molecule Platform for Investigation of Interactions between G-Quadruplexes and Small-Molecule Ligands. *Nat. Chem.* **2011**, *3* (10), 782–787. <https://doi.org/10.1038/nchem.1126>.
- (333) Bonnat, L.; Bar, L.; Génaro, B.; Bonnet, H.; Jarjayes, O.; Thomas, F.; Dejeu, J.; Defrancq, E.; Lavergne, T. Template-Mediated Stabilization of a DNA G-Quadruplex Formed in the HIV-1 Promoter and Comparative Binding Studies. *Chem. – Eur. J.* **2017**, *23* (23), 5602–5613. <https://doi.org/10.1002/chem.201700417>.
- (334) Lau, Y.; Spring, D. Efficient Synthesis of Fmoc-Protected Azido Amino Acids. *Synlett* **2011**, *2011* (13), 1917–1919. <https://doi.org/10.1055/s-0030-1260950>.
- (335) Katayama, H.; Hojo, H.; Ohira, T.; Nakahara, Y. An Efficient Peptide Ligation Using Azido-Protected Peptides via the Thioester Method. *Tetrahedron Lett.* **2008**, *49* (38), 5492–5494. <https://doi.org/10.1016/j.tetlet.2008.07.037>.
- (336) Fujiwara, Y.; Yamanashi, Y.; Fujimura, A.; Sato, Y.; Kujirai, T.; Kurumizaka, H.; Kimura, H.; Yamatsugu, K.; Kawashima, S. A.; Kanai, M. Live-Cell Epigenome Manipulation by Synthetic Histone Acetylation Catalyst System. *Proc. Natl. Acad. Sci.* **2021**, *118* (4), e2019554118. <https://doi.org/10.1073/pnas.2019554118>.
- (337) Zeng, D.-Y.; Kuang, G.-T.; Wang, S.-K.; Peng, W.; Lin, S.-L.; Zhang, Q.; Su, X.-X.; Hu, M.-H.; Wang, H.; Tan, J.-H.; Huang, Z.-S.; Gu, L.-Q.; Ou, T.-M. Discovery of Novel 11-Triazole Substituted Benzofuro[3,2-*b*]Quinolone Derivatives as *c-Myc* G-Quadruplex Specific Stabilizers via Click Chemistry. *J. Med. Chem.* **2017**, *60* (13), 5407–5423. <https://doi.org/10.1021/acs.jmedchem.7b00016>.
- (338) Thibaut Masson; 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**. <https://doi.org/10.1093/nar/gkab1166>.
- (339) Panda, D.; Saha, P.; Das, T.; Dash, J. Target Guided Synthesis Using DNA Nano-Templates for Selectively Assembling a G-Quadruplex Binding *c-MYC* Inhibitor. *Nat. Commun.* **2017**, *8* (1), 16103. <https://doi.org/10.1038/ncomms16103>.
- (340) Arévalo-Ruiz, M.; Doria, F.; Belmonte-Reche, E.; De Rache, A.; Campos-Salinas, J.; Lucas, R.; Falomir, E.; Carda, M.; Pérez-Victoria, J. M.; Mergny, J.-L.; Freccero, M.; Morales, J. C. Synthesis, Binding Properties, and Differences in Cell Uptake of G-Quadruplex Ligands Based on Carbohydrate Naphthalene Diimide Conjugates. *Chem. - Eur. J.* **2017**, *23* (9), 2157–2164. <https://doi.org/10.1002/chem.201604886>.
- (341) Jantos, K.; Rodriguez, R.; Ladame, S.; Shirude, P. S.; Balasubramanian, S. Oxazole-Based Peptide Macrocycles: A New Class of G-Quadruplex Binding Ligands. *J. Am. Chem. Soc.* **2006**, *128* (42), 13662–13663. <https://doi.org/10.1021/ja064713e>.
- (342) Tian, H.; Xu, Y.; Liu, S.; Jin, D.; Zhang, J.; Duan, L.; Tan, W. Synthesis of Gibberellic Acid Derivatives and Their Effects on Plant Growth. *Molecules* **2017**, *22* (5), 694. <https://doi.org/10.3390/molecules22050694>.
- (343) Yu, T.-B.; Bai, J. Z.; Guan, Z. Cycloaddition-Promoted Self-Assembly of a Polymer into Well-Defined β Sheets and Hierarchical Nanofibrils. *Angew. Chem. Int. Ed.* **2009**, *48* (6), 1097–1101. <https://doi.org/10.1002/anie.200805009>.
- (344) Kwie, F. H. A.; Briet, M.; Soupaya, D.; Hoffmann, P.; Maturano, M.; Rodriguez, F.; Blonski, C.; Lherbet, C.; Baudoin-Dehoux, C. New Potent Bisubstrate Inhibitors of Histone Acetyltransferase P300: Design, Synthesis and Biological Evaluation: New Bisubstrate Inhibitors of P300. *Chem. Biol. Drug Des.* **2011**, *77* (1), 86–92. <https://doi.org/10.1111/j.1747-0285.2010.01056.x>.
- (345) Li, Y.; Zimmermann, G.; Scheuermann, J.; Neri, D. Quantitative PCR Is a Valuable Tool to Monitor the Performance of DNA-Encoded Chemical Library Selections. *ChemBioChem* **2017**, *18* (9), 848–852. <https://doi.org/10.1002/cbic.201600626>.
- (346) Sanchez Perez, E.; Toor, R.; Bruyat, P.; Cepeda, C.; Degardin, M.; Dejeu, J.; Boturny, D.; Coche-Guérente, L. Impact of Multimeric Ferrocene-containing Cyclodecapeptide Scaffold on Host-Guest Interactions at a B-Cyclodextrin Covered Surface. *ChemPhysChem* **2021**, *22* (21), 2231–2239. <https://doi.org/10.1002/cphc.202100469>.
- (347) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2009**, NA-NA. <https://doi.org/10.1002/jcc.21334>.
- (348) Harikrishna, S.; Kotaru, S.; Pradeepkumar, P. I. Ligand-Induced Conformational Preorganization of Loops of *c-MYC* G-Quadruplex DNA and Its Implications in Structure-Specific Drug Design. *Mol. Biosyst.* **2017**, *13* (8), 1458–1468. <https://doi.org/10.1039/C7MB00175D>.
- (349) Gaussian 16, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, J. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.
- (350) Sybyl-X Molecular Modeling Software Packages, Version 2.1. 1 TRIPOS Associates, Inc. - 2012.
- (351) D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M.

(352) York and P.A. Kollman (2018), AMBER 2018, University of California, San Francisco.
Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-XTB—An Accurate and Broadly Parametrized Self-Consistent

Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15* (3), 1652–1671. <https://doi.org/10.1021/acs.jctc.8b01176>.

Annexes.

1. Publication.

(1) Rieber, P.; Si, Y.; Van Wesemal, C.; Gay, R.; They, O.; Bonnet, H.; Lartia, R.; Dejeu, J.; Dash, J.; Giustiniano, M.; Doria, F.; Teulade Fichou, M.-P.; Riou, J.-F.; Jamet, H.; Ducongé, F.; Lavergne, T. Selection of DNA-encoded chemical libraries against G-Quadruplex Nucleic Acid Structures. *Manuscript in preparation for JACS*.

2. Communications.

a) Posters

(1) Journée Scientifique ARCANE — Juin 2019

Development of encoded combinatorial chemistry to identify specific ligands of G-quadruplex DNA structures.

Pierre Rieber, Yang Si, Camille Van Wesemael, Rémy Lartia, Isabelle Baussane, Martine Demeunynck, Thomas Lavergne.

(2) Journée du Printemps de la SCF — Rhône-Alpes — Juin 2019

Development of combinatorial encoded chemistry for the identification of specific ligands of G-quadruplex DNA structures.

Pierre Rieber, Yang Si, Camille Van Wesemael, Rémy Lartia, Isabelle Baussane, Martine Demeunynck, Thomas Lavergne.

(3) DNA-Encoded chemical library – 9th International Symposium – Septembre 2019

Assembly of a combinatorial encoded library for the identification, by selection, of specific ligands of G-quadruplex DNA structures.

Pierre Rieber, Yang Si, Camille Van Wesemael, Rémy Lartia, Isabelle Baussane, Martine Demeunynck, Thomas Lavergne.

b) Communications orales

(1) ACS Spring 2021 – Macromolecular Chemistry: The second century – Avril 2021

DNA-encoded chemical library: drug discovery process applied to G-quadruplex DNA structures.
Pierre Rieber, Yang Si, Camille Van Wesemael, Rémy Lartia, Isabelle Baussane, Martine Demeunynck, Thomas Lavergne.

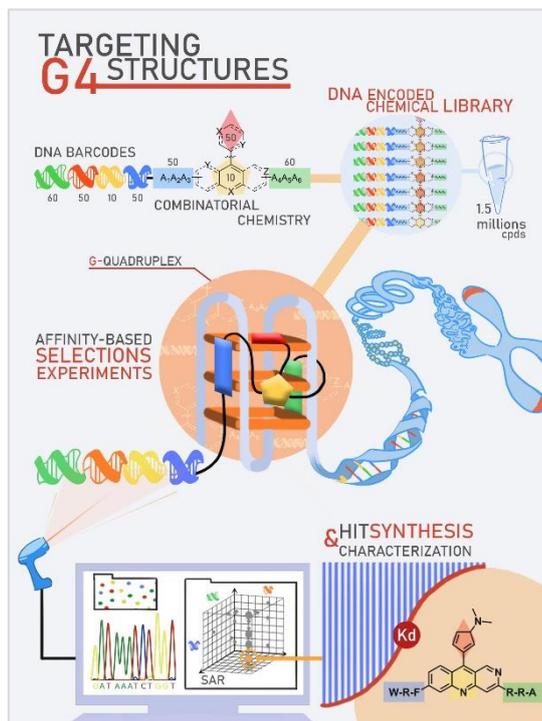
(2) Labex ARCANE Webinar – Avril 2021

Assembly of a combinatorial encoded library for the identification, by selection, of specific ligands of non-usual nucleic acid conformations.

Pierre Rieber, Yang Si, Camille Van Wesemael, Rémy Lartia, Isabelle Baussane, Martine Demeunynck, Thomas Lavergne.

c) Evènements

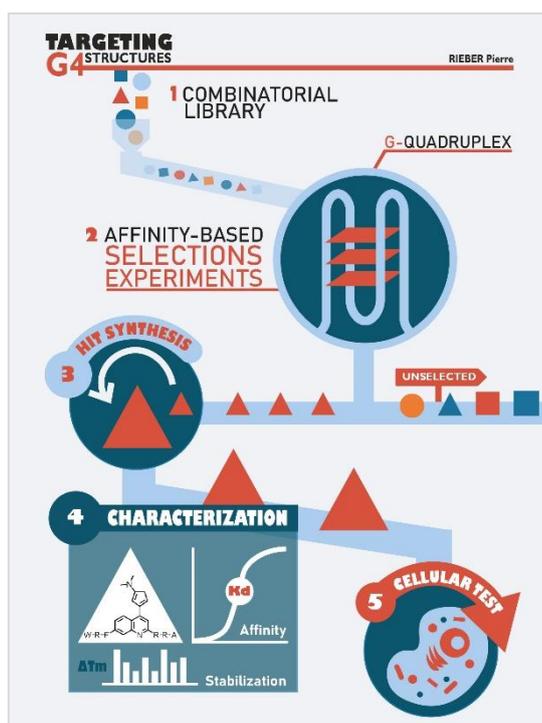
(1) Participation au concours « Graphical Abstract » / Labex ARCANE — Février 2020



(2) Participation au concours « Graphical Abstract » / Labex ARCANE — Février 2021

Prix du meilleur abstract graphique : « Draw your thesis ».

Récompense : 1k€ participation à un congrès international et vidéo promotionnelle du projet de thèse (Youtube, Université Grenoble Alpes: <https://www.youtube.com/watch?v=FIQ-ka43w6Q>).



Développement d'une chimie combinatoire encodée pour l'identification de ligands spécifiques d'ADN G-quadruplex.

Au-delà des structures duplex, il existe des preuves convaincantes qui suggèrent que les acides nucléiques peuvent former des structures secondaires telles que la structure G-quadruplex (G4). Le G-quadruplex est formé par des interactions non canoniques de nucléobases au sein de séquences d'ADN et d'ARN riches en G. Ces structures, formées dans des régions régulatrices clés des génomes et des transcriptomes, sont impliquées dans de nombreux processus biologiques mais aussi au sein de mécanismes moléculaires de plusieurs pathologies. Afin d'identifier des sondes moléculaires et des candidats médicaments capables de lier spécifiquement de telles structures, nous avons construit des bibliothèques combinatoires de ligands potentiels des structures G4. Ces bibliothèques sont assemblées à partir d'éléments de reconnaissance des structures G4 tels que des noyaux hydrophobes hétéroaromatiques, capables d'interagir avec des tétrades externes, des polypeptides chargés positivement et/ou hydrophobes, capables de former une interaction supplémentaire avec des sillons et/ou des boucles chargées négativement, et un troisième groupe de différenciation composé de divers motifs moléculaires. Les bibliothèques ont été soumises à des procédures de sélection *in vitro* par affinité contre plusieurs cibles G4 pour isoler les ligands les plus spécifiques. Afin d'identifier les composés sélectionnés et d'obtenir des données de RSA, nous utilisons une technologie d'encodage permettant d'encoder chaque ligand potentiel avec sa séquence d'ADN unique (DECL). Grâce à cette approche, nous avons identifié des composés prometteurs et des modèles de liaisons en utilisant le séquençage à haut débit. Les composés les plus prometteurs ont été resynthétisés et leurs propriétés de liaison ont été caractérisées par des approches biophysiques. Au cours de ce travail, nous avons trouvé des ligands de G4 très affins validant ainsi pleinement l'utilisation de la technologie DEL pour l'identification de ligands spécifiques de G4.

Mots-clés : bibliothèque encodée par l'ADN, chimie combinatoire, biologie moléculaire, G-quadruplex, structure peptido-organique

Development of combinatorial encoded chemistry for the identification of specific ligands of G-quadruplex DNA structure.

Beyond the canonical duplex, there are now compelling evidences that suggest that nucleic acids can form secondary structures such as G-quadruplex (G4). G4 can arise through non-canonical nucleobase interactions within G-rich DNA and RNA sequences. Those structures, formed in key regulatory regions of genomes and transcriptomes, are involved in many biological processes but also in the molecular mechanisms of several diseases. In order to identify molecular probes and drug candidates capable of specifically binding such structures, we have built combinatorial libraries of potentials ligands of G4 structures. Those libraries are assembled from recognition elements of G4 nucleic acids such as heteroaromatic hydrophobic cores, able to act with external G-quartets, positively charged and/or hydrophobic polypeptides, able to form additional interaction with negatively charged grooves and/or loops, and a third differentiation group composed of various molecular patterns. The libraries were subjected to affinity-based *in vitro* selection procedures against several G4 targets to isolate the most specific ligands. In order to identify the selected compounds and obtain SAR data, we are making use of an encoding technology to barcode each potential ligands with its unique DNA sequence (DECL). Thanks to this approach we have identified promising compounds and binding patterns using high-throughput sequencing. The most promising compounds were re-synthesized and their binding properties were characterized by biophysical approaches. In the course of this work, we found highly potent G4 ligands fully validating the use of DEL technology to identify G4-specific binders.

Key-words: DNA-Encoded chemical library, Combinatorial chemistry, Molecular Biology, G-quadruplex, peptide-organic scaffold