

Centre Armand-Frappier Santé Biotechnologie

## **CRÉATION *DE NOVO* DE LIBRAIRIES DÉGÉNÉRÉES DE GRANDE TAILLE**

Par  
Philip Loranger

Mémoire présenté pour l'obtention du grade de  
Maître ès sciences (M.Sc.)  
en microbiologie appliquée

### **Jury d'évaluation**

Président du jury et examinateur interne

Dr Philippe Constant  
Institut National de la Recherche Scientifique

Examineur externe

Dr Martin Bisailon  
Université de Sherbrooke

Directeur de recherche

Dr Jonathan Perreault  
Institut National de la Recherche Scientifique

## REMERCIEMENTS

Tout d'abord, un remerciement est nécessaire pour l'organisme subventionnaire de ma maîtrise, le conseil de recherches en sciences naturelles et en génie du Canada (CRSNG).

Je suis aussi reconnaissant à tous mes collègues de laboratoire m'ayant constamment aidé et supporté lors de mes expériences. Un merci particulier à Samia pour son aide d'analyse en bioinformatique et à Nicolas pour sa main en microscopie.

Un grand merci est de mise pour tous nos collaborateurs de l'INRS ayant fourni les protéines: Pr Calmette (HP0953), Pr Doucet (Gal7), et Pr St-Pierre (Gal1). Un merci additionnel au Pre Jeßberger (toxines) de l'université de Munich Ludwig-Maximilians. Je tiens aussi à remercier le laboratoire de Pr Veyrier (INRS) pour le séquençage Nanopore et le Pr Salim (INRS) pour le temps au microscope inversé.

Je lance aussi un petit mot à mon jury: le Pr Bisailon (USherbrooke) et le Pr Constant (INRS) pour le temps accordé à la lecture de mon mémoire.

Finalement, je tiens à remercier de tout mon cœur mon directeur de recherche, le Pr Jonathan Perreault pour ses excellents conseils et bonnes orientations. Une personne sans qui le projet serait sans doute resté bloqué lors des incessantes impasses.

## RÉSUMÉ

Les aptamères sont des séquences, généralement courtes, d'oligonucléotides se liant avec grande affinité à une molécule cible. Les aptamères sont sélectionnés par un processus appelé *Systematic Evolution of Ligands by Exponential Enrichment* (SELEX), où une grande librairie de séquences d'ADN aléatoires est exposée à un ligand d'intérêt. Cette librairie diversifiée provient typiquement d'un fournisseur commercial qui la synthétise chimiquement. Dans la majorité des cas, la longueur des oligonucléotides ne dépasse pas plus de 100 bases dues à une importante baisse de rendement associé à la synthèse de longs oligonucléotides. Malgré cette contrainte, des librairies contenant de longs ADN pourraient être bénéfiques dû à une plus grande diversité et complexité de structures. De longues librairies ont donc été synthétisées par façon enzymatique, puis ont été testées en parallèle à de courtes librairies lors d'un SELEX. Bien qu'aucun aptamère n'a été trouvé, deux nouvelles méthodes de synthèse de librairies à longue région aléatoire amplifiable par PCR ont été explorées. La première utilise la terminale transférase afin de synthétiser *de novo* une longue séquence dégénérée. La deuxième emploie la T4 ARN ligase 1 afin de joindre des ADN simple-brin dégénérés successivement. Cela ajoute des outils de plus à l'arsenal de la biologie moléculaire.

Mots-clés: terminale transférase (TdT); T4 ARN ligase 1; librairie dégénérée; longue librairie; concatémère; ligation simple-brin; riboqueue; adaptateur; librairie *de novo*; polymérisation sans matrice

## ABSTRACT

Aptamers are short oligonucleotides that bind targets with high affinity. They are selected by Systematic Evolution of Ligands by Exponential Enrichment (SELEX), where sequences interacting with a target are gradually enriched from a large library pool. Libraries are usually ordered from commercial providers and chemically synthesized. Commercial oligonucleotides typically do not reach above 100 bases in length since an increase in size is associated with greatly reduced production yields. However, long random libraries could be beneficial since lengthy sequences have more complex structures. Long libraries were thus enzymatically synthesized to compare them with classical short libraries through SELEX. Even though no aptamers were found, two novel methods have been explored to synthesize PCR amplifiable libraries with long random regions. The first uses terminal transferase to synthesize *de novo* a long random region. The second uses T4 RNA ligase 1 to ligate random single-stranded DNA successively. These new tools could prove useful in the molecular biology arsenal.

Keywords: terminal transferase (TdT); T4 RNA ligase 1; random library; lengthy library; concatamer; single-stranded ligation; ribotail; adaptor; *de novo* library; template-independent polymerization

# TABLES DES MATIÈRES

REMERCIEMENTS .....	II
RÉSUMÉ .....	III
ABSTRACT .....	IV
TABLES DES MATIÈRES .....	V
LISTE DES FIGURES .....	VII
LISTE DES TABLEAUX .....	IX
LISTES DES ABRÉVIATIONS .....	X
<b>1 INTRODUCTION GÉNÉRALE .....</b>	<b>1</b>
1.1 <i>Aptamère</i> .....	1
1.1.1 Structures et interactions .....	2
1.1.2 Applications.....	2
1.1.3 Anatomie d'un aptamère .....	4
1.1.4 Synthèse commerciale d'oligonucléotides .....	5
1.1.5 Longueur des librairies .....	6
1.2 <i>Acides nucléiques catalytiques</i> .....	7
1.3 <i>SELEX</i> .....	7
1.3.1 Étapes principales.....	8
1.3.2 Sélection négative .....	10
1.3.3 Analyses subséquentes .....	10
1.4 <i>Enzymes utilisées</i> .....	12
1.4.1 Terminale transférase .....	12
1.4.2 T4 ARN ligase 1 .....	14
1.4.3 Exonucléase lambda .....	16
1.5 <i>Problématique</i> .....	17
1.6 <i>Hypothèses et objectifs</i> .....	17
<b>2 MATÉRIEL ET MÉTHODES .....</b>	<b>18</b>
2.1 <i>Synthèse des librairies</i> .....	18
2.1.1 Fluorescence .....	18
2.1.2 Échelle de poids moléculaire .....	18
2.1.3 Librairie enzymatique - terminale transférase (N <sub>800</sub> ) .....	19
2.1.4 Librairie enzymatique - T4 ARN ligase 1 (N <sub>200</sub> ) .....	24
2.1.5 Librairies chimiques (N <sub>20</sub> et N <sub>50</sub> ) .....	27
2.2 <i>Ligands</i> .....	27
2.2.1 Choix des cibles .....	27
2.2.2 Fonctionnalisation de billes magnétiques .....	28
2.3 <i>SELEX</i> .....	30
2.3.1 Étapes des cycles .....	30
2.3.1 Cycles négatifs et protéines négatives .....	31
2.3.2 Tableau des cycles SELEX.....	33
2.3.3 Dosage des fractions .....	33
2.3.4 Préparation pour le séquençage .....	34
2.3.5 Analyse des séquences.....	35
2.4 <i>Tests de liaison</i> .....	36
2.4.1 Retard sur gel .....	36
2.4.2 Microscope inversé .....	36
2.4.3 Essai à la goutte.....	37
2.4.4 Thermophorèse à micro-échelle .....	38
<b>3 RÉSULTATS .....</b>	<b>39</b>

3.1	<i>Synthèse des bibliothèques</i> .....	39
3.1.1	Bibliothèque N <sub>800</sub> par terminale transférase .....	39
3.1.2	Bibliothèque N <sub>200</sub> par ligation .....	46
3.1.3	Bibliothèque chimique N <sub>20</sub> et N <sub>50</sub> .....	52
3.1.4	Récapitulatif des bibliothèques et des ligands .....	53
3.2	<i>Cycles SELEX</i> .....	54
3.3	<i>Dosages des fractions</i> .....	56
3.4	<i>Séquençage et analyse</i> .....	58
3.4.1	Préparation des bibliothèques N <sub>20</sub> et N <sub>50</sub> .....	58
3.4.2	Contrôle qualité des bibliothèques N <sub>200</sub> et N <sub>800</sub> après séquençage Nanopore.....	59
3.4.3	Candidats d'aptamères potentiels obtenus .....	60
3.4.4	Contenu nucléotidique de la bibliothèque initiale N <sub>800</sub> .....	61
3.5	<i>Tests de liaison</i> .....	63
3.5.1	Retard sur gel ( <i>gel shift assay</i> ) .....	63
3.5.2	Microscope inversé .....	64
3.5.3	Essai à la goutte ( <i>spotting test</i> ) .....	65
3.5.4	Thermophorèse à micro-échelle .....	66
<b>4</b>	<b>DISCUSSION</b> .....	<b>67</b>
4.1	<i>Synthèse des bibliothèques enzymatiques</i> .....	67
4.1.1	Terminale transférase .....	67
4.1.2	T4 ARN ligase 1 .....	68
4.1.3	Potentielles méthodes additionnelles.....	70
4.2	<i>Cycle SELEX</i> .....	72
4.2.1	Billes magnétiques .....	72
4.2.2	Séquençage des bibliothèques enzymatiques N <sub>200</sub> et N <sub>800</sub> .....	72
4.2.3	Séquençage des bibliothèques chimiques N <sub>20</sub> et N <sub>50</sub> .....	74
4.2.4	Optimisation des PCR .....	74
4.2.5	Erreur humaine .....	76
4.3	<i>Tests de liaison</i> .....	77
4.4	<i>Conclusion</i> .....	78
<b>5</b>	<b>BIBLIOGRAPHIE</b> .....	<b>79</b>
<b>6</b>	<b>ANNEXE I</b> .....	<b>86</b>
<b>7</b>	<b>ANNEXE II</b> .....	<b>87</b>
<b>8</b>	<b>ANNEXE III</b> .....	<b>88</b>
	<b>ANNEXE IV</b> .....	<b>89</b>
<b>9</b>	<b>ANNEXE V</b> .....	<b>90</b>
<b>10</b>	<b>ANNEXE VI</b> .....	<b>91</b>
<b>11</b>	<b>ANNEXE VII</b> .....	<b>92</b>

## LISTE DES FIGURES

Figure 1: Nombre de publications annuelles sur <i>Google Scholar</i> comportant le terme « <i>aptamer</i> » de 1990 à 2021. ....	1
Figure 2: Schématisation d'un ADN simple-brin formant des structures secondaires et tertiaires. ....	2
Figure 3: Schématisation des différentes parties anatomiques d'un aptamère. ....	5
Figure 4: Schématisation des principales étapes d'un cycle SELEX classique. ....	9
Figure 5: Exemple fictif simplifié d'une courbe de liaison montrant une fluorescence d'ADN diminuante dépendamment de la concentration de ligand. ....	11
Figure 6: Schématisation des deux types de réactions de la terminale transférase. ....	13
Figure 7: Schématisation des trois types de réactions et produits par la T4 ARN ligase 1. ....	15
Figure 8: Schématisation de la digestion d'un ADN double-brin phosphorylé par l'exonucléase lambda. ....	16
Figure 9: Limite minimale de détection de la fluoescéine par le Typhoon. ....	18
Figure 10: Schéma des trois PCR successives menant à un amplicon fluorescent. ....	19
Figure 11: Schématisation de l'allongement enzymatique d'une amorce fluorescente à l'aide de la terminale transférase et de dNTP. ....	20
Figure 12: Schématisation de l'anatomie complète de la librairie TdT avant son amplification. ....	21
Figure 13: Schématisation du problème d'amplification de librairies par PCR. ....	23
Figure 14: Schématisation de la création d'une longue région dégénérée par ligation. ....	25
Figure 15: Fonctionnalisation des billes magnétiques avec un peptide fluorescent. ....	29
Figure 16: Photo de la plaque de verre contenant les gouttes des fractions d'un cycle SELEX. ....	34
Figure 17: Migration agarose d'amplicons fluorescents indexés issus du gène ribosomal 16S bactérien. ....	39
Figure 18: Longueur des ADN en fonction du temps d'incubation, du ratio ADN:enzyme, et de la quantité de dNTP. ....	40
Figure 19: Longueur des ADN en fonction du temps d'incubation et du ratio ADN:enzyme. ....	41
Figure 20: Profil de migration de la librairie optimisée finale TdT (région aléatoire, sans adaptateur). ....	42
Figure 21: Preuve de concept démontrant qu'une courte queue d'ARN en 3' de l'ADN accepteur est nécessaire pour liguer des ADN simple-brin. ....	43
Figure 22: PCR finale pour le retour à de l'ADN double-brin et digestion avec une exonucléase. ....	45
Figure 23: Nombre de fragments d'ADN simple-brins ligués en fonction du ratio d'ADN accepteur:donneur. ....	46
Figure 24: Diverses conditions testées ayant pour but d'améliorer le rendement de ligation d'ADN simple-brin. ....	47
Figure 25: Effet de la quantité de $N_{50}$ et d'agents dénaturants sur le nombre de fragments d'ADN simple-brin ligués. ....	48
Figure 26: Effet du nombre de riboadénine (rA) en 3' du donneur et de l'accepteur sur le rendement de ligation d'ADN simple-brin. ....	49
Figure 27: Essai de ligation d'ADN simple-brin de région aléatoire et d'adaptateur en une seule réaction. ....	50
Figure 28: Amplification par RT-PCR de la librairie enzymatique $N_{200}$ . ....	51
Figure 29: Amplification PCR des librairies chimiques $N_{20}$ et $N_{50}$ avec cycle final. ....	52
Figure 30: Importance du chauffage des billes avant la PCR. ....	54
Figure 31: PCR typique (réussie ou avec artefacts) lors des cycles SELEX. ....	55
Figure 32: Dosage sur verre de gouttelettes contenant de l'ADN fluorescent issues des fractions des cycles SELEX. ....	56
Figure 33: Enrichissement des cinq librairies au cours du SELEX. ....	57
Figure 34: Indexation par PCR des librairies en guise de préparation au séquençage à haut débit. ....	58
Figure 35: Exemple typique de contrôle qualité du séquençage Nanopore (NanoPlot Report) pour les librairies enzymatiques $N_{200}$ et $N_{800}$ . ....	59
Figure 36: Retard sur gel natif PAGE comparant 12 candidats d'ADN (Cy5) seuls ou en présence de leur cible (protéine Gal7, toxines, et HP0953). ....	63
Figure 37: Billes magnétiques (fonctionnalisées avec des protéines) incubées avec de l'ADN fluorescent (candidats). Images prises par un microscope à statif inversé. ....	64

Figure 38: Essai à la goutte mesurant l'interaction entre les candidats (ADN fluorescent) et des billes magnétiques fonctionnalisées. Mesure prise par le niveau de déplétion d'ADN du milieu lorsqu'il est pêché par les billes. ....65

Figure 39: Fichier de sortie typique de thermophorèse à micro-échelle présentant une courbe de liaison entre une concentration fixe de candidat (ADN fluorescent) et une concentration grandissante de sa cible respective. ....66

Figure 40: Schéma visuel des jonctions d'ARN entre chaque fragment N<sub>50</sub> de la librairie N<sub>200</sub>. ....69

Figure 41: Schématisation d'une méthode alternative de création d'une large région dégénérée par ligation double-brin avec la T4 ADN ligase. 71

Figure 42: Schématisation d'une méthode alternative de création d'une large région dégénérée par l'ADN polymérase phi29. ....71

## LISTE DES TABLEAUX

Tableau 1: Paramètres des étapes de SELEX à chaque cycle.....	33
Tableau 2: Paires d'amorces employées lors de l'indexation des courtes librairies N <sub>20</sub> et N <sub>50</sub> .....	35
Tableau 3: Récapitulatif des trois ligands et cinq librairies employées dans les cycles SELEX.....	53
Tableau 4: Candidats d'aptamères potentiels obtenus suite à l'analyse des fichiers de sortie de séquençage Miseq et Nanopore pour toutes les librairies et cibles. ....	60
Tableau 5: Proportion nucléotidique (A, T, G, ou C) de la librairie enzymatique N <sub>800</sub> initiale avant les cycles SELEX. ....	61
Tableau 6: Séquence et modification des oligonucléotides commerciaux utilisés.....	86
Tableau 7: Candidats d'aptamères potentiels sélectionnés des fichiers de sorties de séquençage pour toutes les librairies de la protéine galectine-7. ....	89
Tableau 8: Candidats d'aptamères potentiels sélectionnés des fichiers de sorties de séquençage pour toutes les librairies des toxines (NheA et HbIB). ....	90
Tableau 9: Candidats d'aptamères potentiels sélectionnés des fichiers de sorties de séquençage pour toutes les librairies de la protéine HP0953. ....	91
Tableau 10: Séquence complète de trois lectures du séquençage Nanopore de la librairie initiale N <sub>800</sub> . ....	92

## LISTES DES ABRÉVIATIONS

∅:	symbole signifiant « absence de [...] »
ADN:	acide désoxyribonucléique (H en 2')
amol:	attomole ( $10^{-18}$ )
ARN:	acide ribonucléique (OH en 2')
BSA:	<i>bovine serum albumin</i>
Cy5:	cyanine 5 (marqueur fluorescent)
DMSO:	diméthylsulfoxyde (solvant organique dénaturant)
dNTP:	désoxyribonucléotide triphosphate (A, T, G, ou C)
EDTA:	acide éthylènediaminetétraacétique
enz restr:	enzyme de restriction
exo:	exonucléase lambda
fM:	femtomolaire ( $10^{-15}$ )
fmol:	femtomole ( $10^{-15}$ )
kDa:	kilodalton (unité de mesure de la masse d'une molécule)
mA:	milliampère
mL:	millilitre
N <sub>50</sub> :	50 nucléotides inconnus successifs (mélange de A, T, G, ou C)
ng:	nanogramme
nm:	nanomètre
nM:	nanomolaire ( $10^{-9}$ )
nmol:	nanomole ( $10^{-9}$ )
nt:	nucléotide (en contexte d'ADN ou d'ARN simple-brin)
PAGE:	<i>polyacrylamide gel electrophoresis</i>
pb:	paire de bases (en contexte d'ADN double-brin)
PCR:	chaîne de réaction par polymérase (technique d'amplification d'ADN)

pM:	picomolaire ( $10^{-12}$ )
pmol:	picomole ( $10^{-12}$ )
rA:	riboadénine
rATP:	riboadénine triphosphate (aussi appelé ATP)
rNTP:	ribonucléotide triphosphate (rA, U, rG, ou rC)
RT:	<i>reverse transcription</i> (synthèse d'un brin d'ADN à partir d'ARN)
SELEX:	<i>Systematic Evolution of Ligands by Exponential enrichment</i>
TAE:	Tris - <i>Acetic acid</i> - EDTA (tampon pour gel agarose)
TBE:	Tris - <i>Boric acid</i> - EDTA (tampon pour gel PAGE)
TdT:	désoxyribonucléotidyl-transférase terminale (enzyme)
$\mu$ L:	microlitre
$\mu$ m:	micromètre
$\mu$ M:	micromolaire ( $10^{-6}$ )
V:	volt

# 1 INTRODUCTION GÉNÉRALE

---

## 1.1 Aptamère

Les aptamères sont de courtes chaînes d'acides nucléiques capables de se coller fortement à une cible donnée (Zhang *et al.*, 2019). Le mot aptamère dérive du langage grec. *Aptus* signifie attaché, joint, ou ajusté, et *meros* désigne une partie ou plusieurs parties (Gragoudas *et al.*, 2004). Ils ont été décrits pour la première fois simultanément par deux groupes de chercheurs distincts: Tuerk et Gold (1990) et Ellington et Szostak (1990). Initialement, seuls des aptamères d'ARN étaient employés, mais ceux en ADN ont gagné en popularité au fil du temps due à leur meilleure stabilité sans sacrifier leur affinité (McKeague *et al.*, 2015). Les aptamères sont des outils biotechnologiques utilisés en recherche, en industrie alimentaire et en surveillance de l'environnement (McKeague *et al.*, 2015). Ils jouent aussi un rôle prometteur en diagnostic clinique ou comme agent thérapeutique (Zon, 2020). Leur utilisation devient de plus en plus courante et leur nombre de publications gagne en importance au fil des années (McKeague *et al.*, 2015; Zon, 2020).

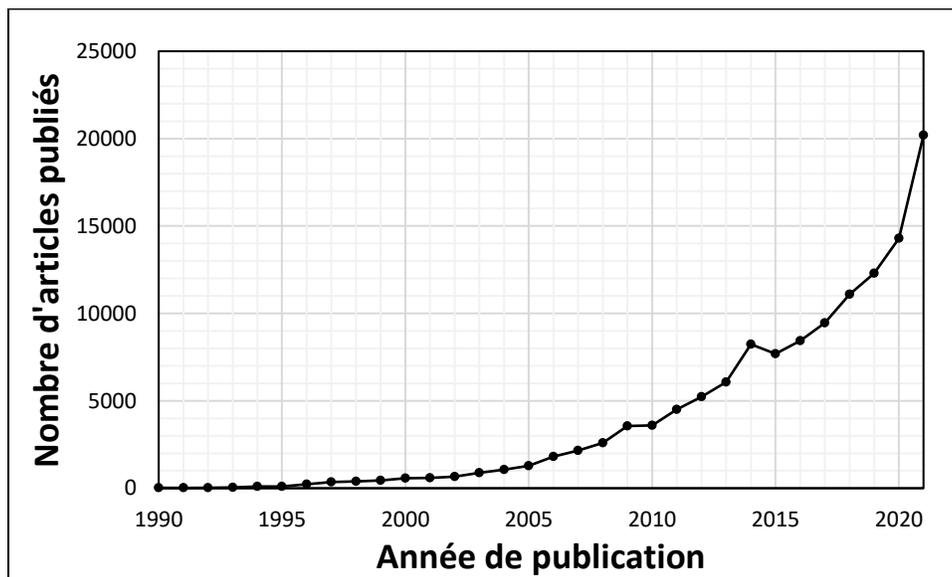


Figure 1: Nombre de publications annuelles sur *Google Scholar* comportant le terme « *aptamer* » de 1990 à 2021.

Le terme « *aptamer* » a été inséré sur le site [WWW.scholar.google.ca](http://WWW.scholar.google.ca), et le nombre d'articles publiés a été compilé pour chaque année (1990-1990, 1991-1991... jusqu'à 2021 inclusivement). Les points de la courbe ne sont pas cumulés, mais correspondent plutôt au nombre d'articles publiés à cette année seulement.

### 1.1.1 Structures et interactions

Étant des ADN ou ARN simple-brin, les aptamères se replient sur eux-mêmes et adoptent des structures secondaires et tertiaires (Zhang *et al.*, 2019). Ces structures leur permettent de se lier avec haute affinité avec leur ligand (aussi appelé analyte ou cible) (Zhang *et al.*, 2019). Les aptamères sont capables de lier un large éventail de cible, variant d'ions à des protéines à haut poids moléculaire (Zhou et Rossi, 2017). Ils sont aussi capables de lier un épitope spécifique de plus grandes entités, telle qu'une cellule entière (Zhou et Rossi, 2017). Généralement, le type d'interaction pour les petites cibles (ex: ion ou acide aminé) est par empilement (*pi stacking* ou *stacking interactions* en anglais), par complémentarité électrostatique, ou par des ponts hydrogènes (Kohlberger et Gadermaier, 2021). Pour les plus grandes cibles telles que les protéines, les interactions ont tendance à être plus diversifiées et souvent combinatoires dépendamment de l'aptamère utilisé (Kohlberger et Gadermaier, 2021).

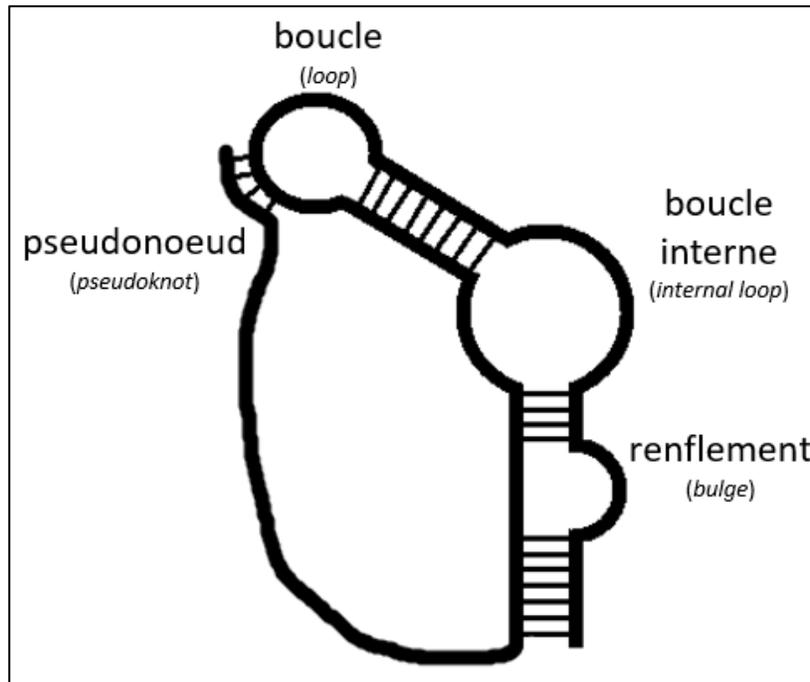


Figure 2: Schématisation d'un ADN simple-brin formant des structures secondaires et tertiaires.

### 1.1.2 Applications

Un médicament à base d'aptamère (appelé *Pegaptanib*) a été approuvé par la FDA en 2004 comme traitement efficace contre la dégénérescence maculaire de la rétine (Sundaram *et al.*, 2013). Cet aptamère d'ARN a été fortement modifié chimiquement à plusieurs niveaux pour pouvoir contrer les ribonucléases du corps (Sundaram *et al.*, 2013). Un grand nombre d'articles traitent du potentiel thérapeutique des

aptamères, surtout au niveau des cancers (Chen *et al.*, 2017; Han *et al.*, 2020; Zon, 2020), mais à ce jour, très peu semblent avoir dépassé les études précliniques (Zon, 2020). Ce manque de succès semble attribué à une spécificité et affinité insuffisante pour les standards d'un remède pour humain (Zon, 2020). Ni *et al.* (2020) suggèrent que ces lacunes *in vivo* sont attribuées à plusieurs facteurs, entres-autres: élimination rapide du corps, vulnérabilité aux nucléases, ainsi que la réponse immunitaire non-spécifique. Contrairement à l'application concrète du *Pegaptanib*, la plupart des nouveaux aptamères découverts ne finissent pas par être utilisés dans un contexte autre qu'académique. Plusieurs preuves de concept sont établies en laboratoire, mais semble rarement utilisées comme outils dans des sphères plus applicables. Ci-bas se trouvent plusieurs exemples de preuves de concepts trouvées dans la littérature.

**Diagnostic:** Une équipe de recherche a développé un dispositif microfluidique se servant d'aptamère capable de capter diverses sortes de cellules cancéreuses dans des échantillons de sang humain (Xu *et al.*, 2009). Cet article a été cité plus de 200 fois.

**Imagerie:** L'équipe de Shen *et al.* (2019) a développé un aptamère fluorescent capable de se coller aux nucléoprotéines de cellules mortes, tout en étant non-toxique et non-spécifique aux cellules vivantes environnantes.

**Polluants:** Weerathunge *et al.* (2019) ont développé un biosenseur en jumelant un nanozyme (des nanoparticules d'argent avec une activité peroxydase) et un aptamère se collant à des pesticides organophosphorés. La détection du polluant se fait par réaction colorimétrique à partir d'échantillon d'eau de rivière.

**Alimentation:** Un autre laboratoire a développé un aptamère jumelé à des nanoparticules d'or capable de détecter *Cronobacter sakazakii*, une dangereuse bactérie pathogène présente dans le lait en poudre pour les nouveau-nés (Kim *et al.*, 2017). La présence de bactéries dans un échantillon mène à l'agglutination des aptamères-nanoparticules et change la couleur du milieu de rouge à mauve.

**Outil de recherche:** En 1994, Jenison *et al.* ont développé un aptamère contre la théophylline, une petite molécule présente dans la feuille de plusieurs plantes. En présence de ce ligand, cet aptamère change sa conformation. Ce phénomène a été utilisé pour permettre de séquestrer, ou rendre disponible, un site de liaison de ribosomes en aval de l'aptamère (un tel aptamère se nomme *riboswitch*). Cet aptamère est depuis couramment utilisé en recherche comme outil d'expression de gènes, avec l'article ayant plus de 1300 citations.

### 1.1.3 Anatomie d'un aptamère

Pour qu'un oligonucléotide soit qualifié comme étant un aptamère, il doit être capable d'interactions avec un ligand. Pour obtenir un aptamère, il faut débiter avec une large librairie initiale comportant des séquences aléatoires (aussi appelé région dégénérée ou variable), et identifier les quelques séquences capables de liaison avec la cible (Sabeti *et al.*, 1997). Une librairie se réfère à un amalgame d'oligonucléotides ayant plusieurs séquences nucléotidiques. Ainsi, dans une librairie, une diversité élevée de molécules est attendue (Sabeti *et al.*, 1997).

Le SELEX est une méthode permettant de ressortir des candidats d'aptamères d'une librairie (Sabeti *et al.*, 1997). Les détails de cette technique seront couverts plus bas, mais il est important de mentionner que toutes les molécules d'une librairie doivent être amplifiables par PCR. Pour que ceci soit possible, chacun des oligonucléotides doit posséder une région conservée à l'extrémité 5' et une autre région conservée en 3' (Sabeti *et al.*, 1997). Une région conservée consiste en une séquence connue et bien définie. Ces régions agissent comme assises d'amorces lors des PCR. Par contre, la région centrale des oligonucléotides d'une librairie possède une séquence inconnue comportant plusieurs « N » successifs, c'est-à-dire un A, T, G, ou C. Ces régions dégénérées sont ce qui donne aux librairies leur diversité (Sabeti *et al.*, 1997). Le concept d'une librairie dans le cadre d'un SELEX est que de toutes les séquences comprises dans leur répertoire, seulement quelques-unes peuvent interagir avec la cible d'intérêt. Une librairie N<sub>50</sub> se réfère à des oligonucléotides ayant 50 nucléotides dégénérés flanqués de part et d'autre par ~20 nucléotides connus, pour un total de 90 nucléotides. Autrefois, des librairies en ARN étaient favorisées aux dépens de celle en ADN, car elles étaient crues à tort de posséder une meilleure variété de motifs, lorsque repliés en structure secondaire (Wang *et al.*, 2019). Depuis, plusieurs ouvrages ont réfuté cette hypothèse et démontrent même de clairs avantages à travailler avec de l'ADN: meilleure stabilité et plus efficace en temps et argent (Wang *et al.*, 2019). Ceci rend la sélection d'aptamères et leur application plus facile, et ainsi, plus populaire dans les ouvrages récents (Wang *et al.*, 2019).

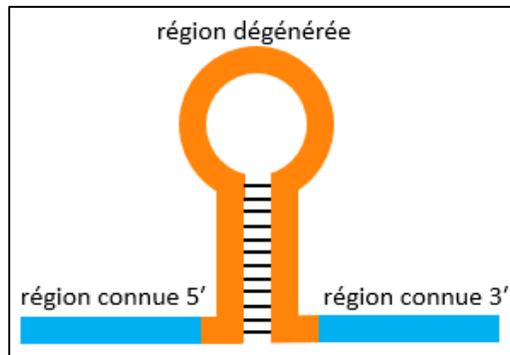


Figure 3: Schématisation des différentes parties anatomiques d'un aptamère.

#### 1.1.4 Synthèse commerciale d'oligonucléotides

La synthèse *de novo* d'oligonucléotides a été développée dans les années 1950 (Michelson et Todd, 1955), alors que leur automatisation et commercialisation date des années 1980 (Kosuri et Church, 2014). Malgré 40 ans d'optimisation, il demeure aujourd'hui encore difficile de produire de longs acides nucléiques de plus de 200 nucléotides (Kosuri et Church, 2014). La méthode commerciale la plus répandue opère par chimie de phosphoramidite (Kosuri et Church, 2014). Une synthèse de ce type se fait par l'ajout successif de bases, une à la fois, sur une chaîne grandissante immobilisée sur un support solide (Kosuri et Church, 2014). Chaque base est bloquée en 5' par un groupement de diméthoxytrityle (celui-ci empêche plusieurs bases de se coller successivement) qui est éventuellement enlevé par une solution acide (Kosuri et Church, 2014). L'acidité libère la base en enlevant le diméthoxytrityle, ce qui permet l'ajout d'une autre base en 5' (Kosuri et Church, 2014). Il est donc question de plusieurs incubations successives, ce qui allonge la chaîne de l'oligonucléotide en direction 3' vers 5' (Kosuri et Church, 2014), l'inverse des polymérase d'ADN. Ce procédé chimique est aujourd'hui très optimisé et rapide, mais ne demeure pas sans contrainte. Même avec un rendement de 99% à chaque ajout de base, seul un rendement final de 13% est possible pour la synthèse d'un oligonucléotide à 200 bases (Kosuri et Church, 2014). En plus de cette contrainte, un clivage d'ADN est aussi associé au retrait du groupement diméthoxytrityle, ce qui diminue la quantité finale de produits (Kosuri et Church, 2014). Or, il est difficile d'atteindre des longueurs allant au-delà de 200 bases en raison de faible rendement et coût élevé, ce qui n'est pas profitable aux yeux des fournisseurs.

Une méthode additionnelle existe permettant de construire des fragments allant jusqu'à 5000 pb (Kosuri et Church, 2014). Celle-ci opère en assemblant 5 à 50 oligonucléotides synthétiques partiellement

complémentaires en leur extrémité (Kosuri et Church, 2014). Une fois appariés, ces oligonucléotides sont joints enzymatiquement à l'aide de ligases, puis amplifiés par PCR (Czar *et al.*, 2009). La méthode d'assemblage en série (*poly chain assembly* ou PCA en anglais) est une alternative à la ligation (Czar *et al.*, 2009). Elle consiste en l'amplification successive de plusieurs complexes ADN-ADN partiellement double-brin en leur centre, avec des extrémités simple-brin en 5' (Czar *et al.*, 2009). La polymérase rend double-brin ces extrémités, ce qui allonge le brin au fil des cycles (Czar *et al.*, 2009). Ces cycles ne sont pas à amplification exponentielle puisque des oligonucléotides équimolaires de départ sont utilisés (Czar *et al.*, 2009). Cependant, suite à 55 cycles de ce type, une amplification finale exponentielle de l'assemblage complet est faite en utilisant des amorces englobant le fragment complet (Czar *et al.*, 2009). Des optimisations récentes ont rendu ce processus rapide et fiable, mais elle demeure coûteuse et ne permet pas de créer n'importe quelle séquence (Kosuri et Church, 2014). Notamment, les séquences riches en GC ou en structure secondaire (comme pour le cas de bibliothèques dégénérées) rendent plus difficile l'appariement ordonné des oligonucléotides (Kosuri et Church, 2014). Il faut aussi garder en tête que le principe repose sur l'assemblage par complémentarité de bases, certaines séquences peuvent donc aussi se positionner au mauvais endroit (Kosuri et Church, 2014). Par exemple, les séquences répétées posent un défi majeur à ce genre de technique puisqu'elles peuvent faire un grand nombre de blocs répétés plutôt que le nombre exact attendu (Kosuri et Church, 2014). Ce type de technologie est bien adapté pour la fabrication synthétique de gènes à des fins de production de protéine (Kosuri et Church, 2014), mais non-applicable pour la création de bibliothèques dégénérées.

En somme, il n'est pas possible de se procurer des oligonucléotides avec une longue séquence dégénérée (>200 nt), issus d'une synthèse commerciale.

### **1.1.5 Longueur des bibliothèques**

La longueur de la région variable des bibliothèques employées pour développer des aptamères semble choisie de façon arbitraire (Velez *et al.*, 2012). En règle générale, les aptamères varient entre 22 à 100 nucléotides en longueur totale (McKeague *et al.*, 2015; Sun *et al.*, 2015; Wang *et al.*, 2019). Très peu d'ouvrages scientifiques ont exploré des bibliothèques de longue taille. Dans la littérature, la bibliothèque la plus longue ayant mené à la découverte d'un aptamère était de N<sub>200</sub> (McKeague *et al.*, 2015) ou N<sub>220</sub> (Kohlberger et Gadermaier, 2021). En la date d'écriture de ce mémoire, seuls quelques articles ont comparé la longueur des régions aléatoires de différentes bibliothèques, mais ces bibliothèques semblent toujours s'en tenir à des tailles sous 100 nucléotides (McKeague *et al.*, 2015). Ce fait semble attribué à la difficulté de se procurer des bibliothèques de longues tailles dans le passé (Sun *et al.*, 2015), comme discuté à la section

précédente. De plus, même si un groupe de recherche avait recours à une librairie avec une région aléatoire supérieure à  $N_{25}$ , cela n'augmente pas le nombre de séquences uniques qui s'y trouve (Hall *et al.*, 2009). Par exemple, une librairie  $N_{50}$  possède plus de  $10^{30}$  possibilités de séquences ( $4^{50}$ ). Une librairie de cette envergure avec toutes les séquences possibles présentes aurait une masse d'environ 92000 tonnes. Par contre, mettant de côté la couverture des séquences possibles, une longue séquence possède certains avantages (Sun *et al.*, 2015). Quelques ouvrages suggèrent que des librairies plus longues peuvent mener à de meilleures affinités dues à des structures plus complexes qui augmenteraient le nombre de sites de liaison potentiels avec leur cible (Blind et Blank, 2015; Kohlberger et Gadermaier, 2021; Komarova et Kuznetsov, 2019; Qi *et al.*, 2022; Sun *et al.*, 2015). De plus, une librairie possédant une courte région aléatoire mène souvent à des séquences ayant des motifs similaires, alors qu'une bonne diversité de conformation mène généralement à de meilleurs aptamères (Kohlberger et Gadermaier, 2021).

## 1.2 Acides nucléiques catalytiques

Dans un contexte autre que celui de sélection d'aptamères, des librairies dégénérées sont aussi employées pour découvrir des acides nucléiques fonctionnels capables d'activité catalytique (Velez *et al.*, 2012). Certains ARN catalytiques (communément appelé ribozyme) sont capables de divers types d'activité, notamment: clivage d'ARN, ligation, ou formation de lien peptidique (Velez *et al.*, 2012). Ce type d'activité est aussi possible pour des ADN catalytiques, communément appelés ADNzyme ou déoxyribozyme (Velez *et al.*, 2012). Similairement aux aptamères, ces acides nucléiques catalytiques sont aussi sélectionnés par évolution dirigée, où une librairie dégénérée diversifiée subit plusieurs rondes d'éliminations successives (Silverman, 2016). Typiquement, la longueur de la région aléatoire pour ce type de sélection est de  $N_{40}$  (Silverman, 2016), mais la plus longue utilisée a été de  $N_{228}$  (Velez *et al.*, 2012). Une longue région dégénérée peut engendrer certaines difficultés, telles qu'un mauvais repliement menant à une séquestration potentielle du site actif (Velez *et al.*, 2012). Par contre, dans un contexte d'activité catalytique, il est préférable d'utiliser des structures secondaires et tertiaires complexes pour augmenter les chances d'obtenir un site actif (Velez *et al.*, 2012). Il serait donc potentiellement avantageux d'employer des librairies de longues tailles afin de chercher de nouveaux (deoxy)ribozymes.

## 1.3 SELEX

Un SELEX est une technique de laboratoire permettant d'identifier d'une librairie diversifiée les séquences capables d'interactions avec un ligand (Kohlberger et Gadermaier, 2021). SELEX est un acronyme signifiant *Systematic Evolution of Ligands by Exponential enrichment* (Kohlberger et

Gadermaier, 2021). Tout comme le premier aptamère, cette technique a été décrite pour la première fois en 1990 par Tuerk et Gold (1990), et Ellington et Szostak (1990). En bref, cette méthode se résume à faire des sélections et amplifications successives pour repêcher les séquences qui mènent à une interaction entre un acide nucléique et une cible afin de ressortir les séquences s'attachant le plus fortement à cette cible (Kohlberger et Gadermaier, 2021). Cette technique est aussi parfois appelée évolution *in vitro* ou dirigée, où il y a un enrichissement des séquences interagissantes aux dépens des séquences n'interagissant pas (Kohlberger et Gadermaier, 2021). Il est donc question d'employer des stratégies permettant une compétition intermoléculaire entre les bonnes et les mauvaises séquences. Un SELEX comprend plusieurs étapes et plusieurs cycles. Au fil des cycles, les populations de séquences enrichies arriveront à un point où les séquences capables d'interagir avec la cible seraient dominantes en termes d'abondance (Kohlberger et Gadermaier, 2021). En bref, le SELEX est donc une façon accélérée de sélectionner des séquences démontrant une affinité pour le ligand et éliminer les séquences non-spécifiques (Kohlberger et Gadermaier, 2021).

### 1.3.1 Étapes principales

Voici les principales étapes d'un SELEX classique: la première étape est l'incubation entre la librairie et la cible d'intérêt (Zhuo *et al.*, 2017). Généralement, un minimum de  $10^{15}$  molécules est utilisé comme librairie de départ (Ni *et al.*, 2020; Stoltenburg *et al.*, 2007). À la suite de cette première incubation, une portion des séquences interagiront avec la cible afin de former un complexe cible-séquence, mais la grande majorité des séquences de la librairie est incapable d'interaction et restera séparée de la cible (Zhuo *et al.*, 2017). La deuxième étape consiste à éliminer ces séquences non-interagissantes (Zhuo *et al.*, 2017). Cette étape se fait souvent sous forme d'un simple lavage lorsque la cible est attachée à un support solide (Zhuo *et al.*, 2017). La troisième étape consiste à récupérer les séquences interagissantes par élution, afin de les amplifier par PCR (Zhuo *et al.*, 2017). La PCR permet d'obtenir une multitude de copies de chaque séquence s'étant collées sur la cible (Zhuo *et al.*, 2017). Ceci entraîne un enrichissement de la librairie envers la cible d'intérêt (Zhuo *et al.*, 2017). Puisqu'une PCR produit de l'ADN double-brin, la dernière étape consiste à retourner à de l'ADN simple-brin afin de permettre à ces séquences de se replier en leur forme native (Svobodová *et al.*, 2012). Rappelons que les structures secondaires et tertiaires sont ce qui permet l'interaction entre un oligonucléotide et la cible (Zhang *et al.*, 2019). Le passage de double-brin à simple-brin est donc primordial. Le premier cycle SELEX maintenant complet, il est maintenant question de répéter ces étapes à plusieurs reprises (Zhuo *et al.*, 2017). Généralement, 5 à 20 cycles de SELEX sont effectués en total (Wu et Kwon, 2016). Au cours de ces cycles, il est possible d'employer des

stratégies permettant une augmentation de la stringence de la sélection, c'est-à-dire augmenter la difficulté avec laquelle un oligonucléotide se lie à sa cible. Ceci permet de favoriser les séquences se liant fortement aux cibles aux dépens des séquences se liant faiblement (Komarova et Kuznetsov, 2019). Cette compétition intermoléculaire est nécessaire, car elle permet l'enrichissement des séquences démontrant la plus forte affinité pour leur ligand aux fils des cycles (Komarova et Kuznetsov, 2019). Voici quelques exemples permettant une augmentation de la stringence: diminution de la quantité de cibles disponibles, diminution du temps d'incubation entre la cible et les séquences, des lavages plus fréquents ou plus longs (Komarova et Kuznetsov, 2019). Il est important de mentionner que la technique SELEX classique, datant de 32 ans, a reçu une grande quantité d'altérations et variations au fil des années. En effet, plus de 30 différentes versions de SELEX ont été recensées en date de 2015 (Acquah *et al.*, 2018). Ces changements permettent, entre autres, à amener des ajouts bénéfiques au protocole original, comme une diminution du temps requis (*one-step* SELEX), immobiliser la librairie plutôt que le ligand (*capture*-SELEX), ou encore le type de ligand employé (*cell*-SELEX). Ces SELEX modifiés ne seront pas couverts en détail dans le cadre du mémoire présent, mais cela souligne la plasticité de la méthode, capable d'être facilement changée afin de contourner une lacune propre à une équipe de recherche ou une cible donnée.

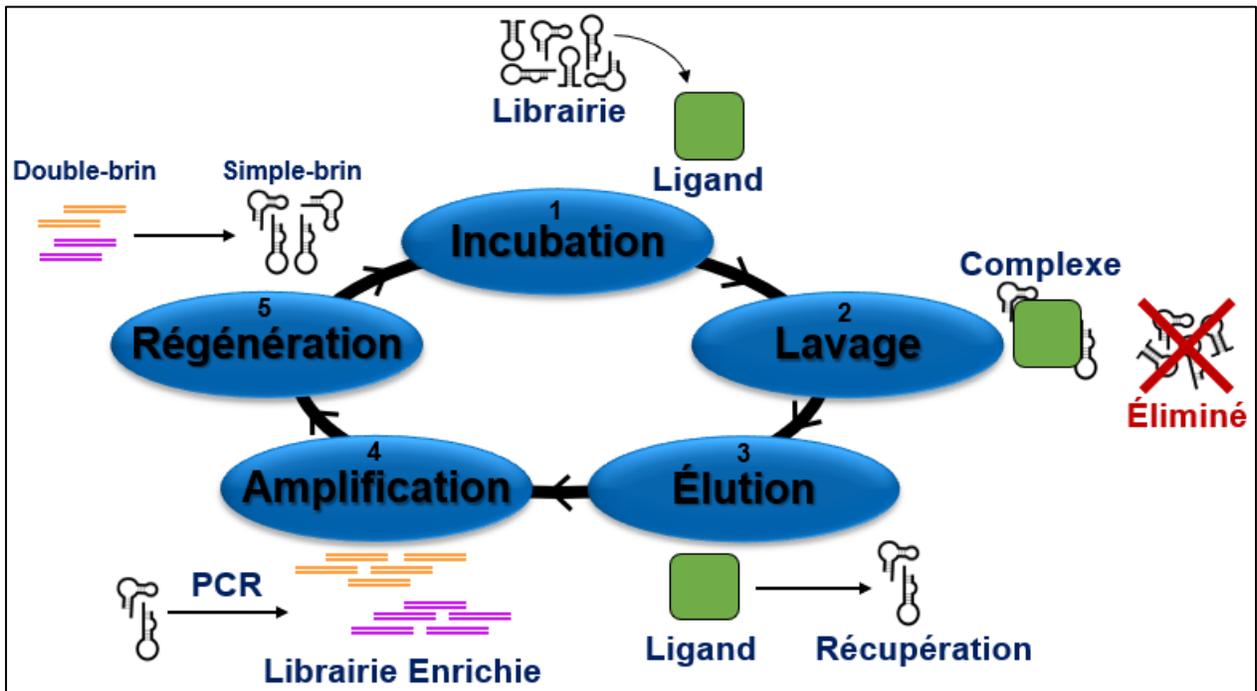


Figure 4: Schématisation des principales étapes d'un cycle SELEX classique.

### 1.3.2 Sélection négative

Une difficulté rencontrée lorsque l'on travaille avec des agents capables de liaison, comme les anticorps et les aptamères, est la liaison non-spécifique (Komarova et Kuznetsov, 2019). Il est possible d'établir qu'un anticorps ou un aptamère est capable de liaison avec une cible, mais il n'y a aucune garantie que ces agents ne lient pas d'autres cibles non-attendues (Zhuo *et al.*, 2017). Ceci peut être particulièrement dommageable dans un contexte de biosenseur ou de diagnostic, qui entraînerait de faux positifs (Zhuo *et al.*, 2017). Un des avantages des aptamères comparativement aux anticorps est que l'on peut faire des sélections négatives (Zhuo *et al.*, 2017). Tout comme un SELEX permet de sélectionner des séquences s'attachant à une cible, il est tout autant possible de sélectionner des séquences ne s'attachant pas à une autre cible (Zhuo *et al.*, 2017). C'est-à-dire, faire une étape de liaison entre les bibliothèques d'oligonucléotides et une cible non-désirée, mais cette fois, plutôt récupérer la fraction comportant les séquences n'interagissant pas avec ces cibles négatives (Zhuo *et al.*, 2017). Une autre stratégie employée est d'immobiliser les cibles désirées sur un support immobile, et de mettre les cibles non-désirées en solution (Zhuo *et al.*, 2017). Ainsi, lorsque l'incubation sera effectuée, les oligonucléotides interagissant avec les cibles négatives dans le surnageant seront éliminés alors que ceux collés sur les cibles positives sur les billes magnétiques seront conservés (Zhuo *et al.*, 2017). Il est aussi possible d'employer plusieurs cibles négatives en même temps (ou en alternance) si l'aptamère est utilisé dans un environnement hétérogène (Komarova et Kuznetsov, 2019). Il est aussi possible de faire des cycles négatifs sur des supports immobiles, comme des billes magnétiques sans ligand attaché, afin d'éliminer les oligonucléotides interagissant directement sur les billes plutôt que sur les cibles d'intérêts collées sur les billes (Zhuo *et al.*, 2017). Faire plusieurs cycles négatifs à travers les cycles normaux de SELEX permet aux aptamères de conserver une spécificité (Zhuo *et al.*, 2017). Ainsi, ceci permet aux aptamères de discriminer entre des cibles très similaires (Komarova et Kuznetsov, 2019).

### 1.3.3 Analyses subséquentes

Une fois tous les cycles de SELEX finalisés, il est maintenant question de séquencer les séquences sélectionnées lors du SELEX (Stoltenburg *et al.*, 2007). Jadis, des clonages suivis d'un séquençage Sanger étaient nécessaires pour obtenir des candidats d'aptamères (Komarova et Kuznetsov, 2019). Les séquences les plus abondantes étaient retenues puis considérées comme des candidats d'aptamères potentiels (Komarova et Kuznetsov, 2019). De nos jours, il est possible de faire un séquençage à haut débit (NGS ou *Next Generation Sequencing* en anglais) sur la bibliothèque entière afin d'avoir un portrait global complet sur les populations des séquences ainsi que leur abondance (Komarova et Kuznetsov, 2019). Suite

à une analyse bioinformatique, les séquences ayant une abondance élevée en lectures sont généralement les candidats d'aptamère retenus, et seront subséquemment testées en laboratoire (Komarova et Kuznetsov, 2019). Un test de liaison (*binding test*) permet de confirmer s'il y a interaction entre les candidats et la cible d'intérêt en plus de calculer leur affinité (Pollard, 2010). Selon cette méthode, l'affinité de chaque candidat est décrite par leur constante de dissociation ( $K_D$ ). Cette constante désigne la concentration requise du ligand afin d'atteindre un équilibre 1:1 entre les complexes aptamère-ligand et les molécules libres; plus un  $K_D$  est bas, plus l'affinité entre un aptamère et son ligand est élevée. Il existe diverses façons expérimentales de mesurer un  $K_D$ , mais en règle générale, il est question de mesurer une différence de signal entre des molécules libres et complexées (Jerabek-Willemsen *et al.*, 2014). Par exemple, un ADN fluorescent pourrait perdre sa fluorescence sous forme de complexe ADN-ligand. Dans cette situation, il est possible d'observer une perte de signal de fluorescence en fonction de la quantité de complexes formés. Ces signaux quantifiables peuvent ensuite être mis sur un graphique afin de tracer une courbe et déterminer le point milieu ( $K_D$ ) entre le point minimal et maximal.

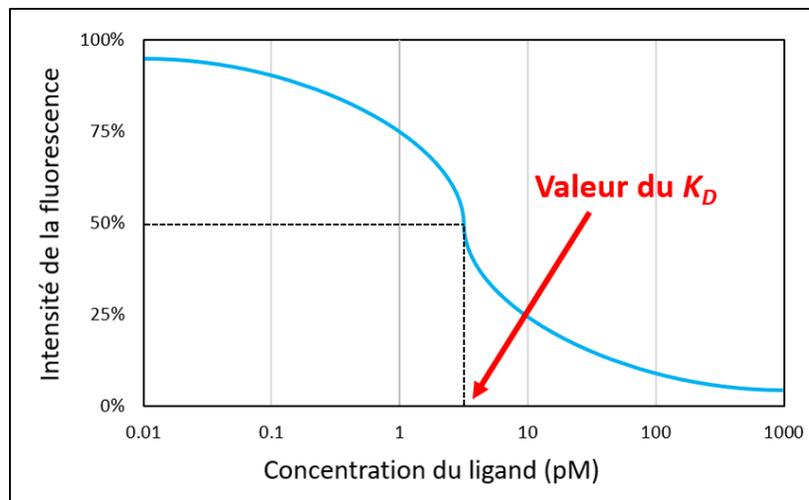


Figure 5: Exemple fictif simplifié d'une courbe de liaison montrant une fluorescence d'ADN diminuante dépendamment de la concentration de ligand.

## 1.4 Enzymes utilisées

L'optique de la recherche présente est d'explorer les bibliothèques de longues tailles comparativement à de courtes tailles. Puisqu'il n'est pas possible de synthétiser chimiquement de longues bibliothèques dégénérées, une option est d'avoir recours à des méthodes enzymatiques. Deux enzymes ont été utilisées afin de construire de longues régions aléatoires: la terminale transférase et la T4 ARN ligase 1. La lambda exonucléase a aussi été utilisée afin de retourner à de l'ADN simple-brin à partir d'ADN double-brin. Ces trois enzymes seront décrites brièvement dans leur mode d'action et utilités.

### 1.4.1 Terminale transférase

La désoxyribonucléotidyl-transférase terminale (communément appelé TdT ou *terminal transferase* en anglais) est une enzyme du groupe des polymérases d'ADN (Fowler et Suo, 2006). Les polymérases sont des enzymes ayant une multitude de rôles biologiques, substrats, localisations cellulaires et mécanismes (Fowler et Suo, 2006; Sarac et Hollenstein, 2019). Les polymérases d'ADN sont divisées en six familles (A, B, C, D, X et Y) selon leur rapprochement phylogénétique et homologie de séquence (Fowler et Suo, 2006). En général, les polymérases des familles B, C et D sont majoritairement impliquées dans la réplication de l'ADN, alors que la famille A touche à la fois la réplication et la réparation de l'ADN (Fowler et Suo, 2006). En revanche, les familles X et Y s'occupe généralement que de la réparation et modification de l'ADN (Sarac et Hollenstein, 2019). La terminale transférase appartient à la famille X (Sarac et Hollenstein, 2019). Plus spécifiquement, cette enzyme est impliquée dans la modification d'ADN chez les lymphocytes B et T, lors de la recombinaison V(D)J (Ashley *et al.*, 2021; Fowler et Suo, 2006; Sarac et Hollenstein, 2019). Lors de la genèse des lymphocytes T et des anticorps (issu des lymphocytes B), les gènes codant pour leurs récepteurs antigéniques subissent l'addition de quelques nucléotides aléatoires entre les segments variable (V) et de diversité (D) (Sarac et Hollenstein, 2019). Cette addition catalysée par la terminale transférase est une stratégie employée par le système immunitaire adaptatif permettant d'augmenter la diversité immunitaire, ce qui augmente la capacité de l'arsenal contre les agents pathogènes (Sarac et Hollenstein, 2019).

Contrairement aux polymérases communes comme la Taq, la terminale transférase ne requiert pas de brin matrice (Sarac et Hollenstein, 2019; Tang *et al.*, 2017). L'enzyme s'accroche à l'extrémité 3' d'un oligonucléotide faisant un minimum de quatre nucléotides de longueur, puis commence son élongation (Sarac et Hollenstein, 2019; Tang *et al.*, 2017). Ces nucléotides additionnels en 3' sont ajoutés de nature aléatoire et ne correspondent pas à une copie conforme complémentaire d'un brin matrice, il

est donc question d'une synthèse *de novo* d'ADN (Sarac et Hollenstein, 2019). La terminale transférase a une préférence naturelle pour les guanines et les cytosines (dGTP et dCTP), mais il est possible de détourner ce biais (Sarac et Hollenstein, 2019). Similairement aux autres polymérase, la TdT a aussi besoin de cofacteur pour son activité catalytique (Sarac et Hollenstein, 2019). Plusieurs cations métalliques divalents peuvent accomplir cette tâche, comme le  $Mg^{2+}$ ,  $Zn^{2+}$ ,  $Co^{2+}$  et le  $Mn^{2+}$  (Sarac et Hollenstein, 2019). Par contre, la particularité de la terminale transférase est que dépendamment du cation employé, sa préférence de nucléotide change (Sarac et Hollenstein, 2019). Par exemple, utiliser le cobalt permet une augmentation d'un facteur de dix l'incorporation de pyrimidines (C et T) plutôt que celle des purines (A et G) (Sarac et Hollenstein, 2019). Le cobalt augmente aussi significativement le rendement des réactions, et est ainsi un cofacteur souvent utilisé dans les trousse commerciales (Sarac et Hollenstein, 2019).

Dans son contexte biologique, la terminale transférase n'ajoute que quelques nucléotides en 3', mais dans des conditions optimales *in vitro*, elle peut ajouter des centaines, voire des milliers de nucléotides d'ADN dépendamment des conditions de réaction (Sarac et Hollenstein, 2019). En effet, Tang *et al.* (2017) ont explorés plusieurs différentes conditions de réaction afin d'observer leur effet sur l'allongement du brin initial. Le ratio enzyme:ADN, la quantité de dNTP et le temps d'incubation semblent tous avoir un impact important sur la longueur du produit créé (Tang *et al.*, 2017). Une autre particularité importante de la TdT est le fait qu'elle peut aussi ajouter des nucléotides en ARN (ou ribonucléotide) (Sarac et Hollenstein, 2019). Cependant, seuls un à quatre ribonucléotides peuvent être ajoutés en 3' plutôt que plusieurs centaines (Miura *et al.*, 2019; Sarac et Hollenstein, 2019). Cette limite de longueur est attribuée au 2'OH des ARN qui amène à un changement de conformation et freine l'enzyme (Sarac et Hollenstein, 2019). Typiquement, la TdT est un outil employé en biologie moléculaire comme agent de marquage, de blocage, ou pour ajouter des queues de monomères (polyA) (Sarac et Hollenstein, 2019).

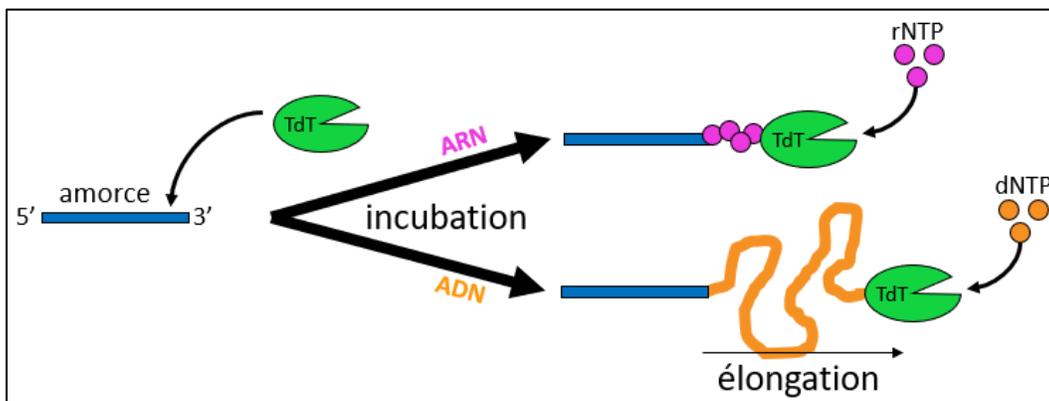


Figure 6: Schématisation des deux types de réactions de la terminale transférase.

### 1.4.2 T4 ARN ligase 1

Les ligases sont des enzymes catalysant la formation de liens phosphodiester entre deux brins d'acide nucléique (Eun, 1996). Cette réaction nécessite typiquement un groupement hydroxyle (OH) à l'extrémité 3' d'un acide nucléique accepteur, et un groupement phosphate au 5' d'un donneur (Eun, 1996). En plus de ces deux conditions, la réaction exige un cofacteur énergétique, souvent l'ATP pour les eucaryotes et le NAD<sup>+</sup> pour les bactéries (Eun, 1996; Shuman, 2009). Les ligases d'ADN sont présentes dans pratiquement tous les organismes vivants (Eun, 1996). Leur rôle biologique se centre autour de la réparation d'ADN, la recombinaison, mais principalement de la réplication de l'ADN (Eun, 1996). En contraste, l'enzyme T4 ARN ligase 1 (aussi appelé Rnl1) provient des virus bactériophages T4 et permet d'évader les défenses de leurs hôtes (Wang *et al.*, 2006; Ho *et al.*, 2004). Cette enzyme a été décrite pour la première fois en 1972 (Silber *et al.*, 1972) et son activité catalytique est similaire aux ligases d'ADN.

La réaction formant le lien phosphodiester se fait en trois étapes (Shuman, 2009). La première est l'attaque nucléophile de l'ATP (triphosphate) environnant, ce qui le transforme en AMP (monophosphate) et relâche une molécule de pyrophosphate inorganique (PP<sub>i</sub>) dans le milieu. Cette réaction lie le phosphore restant de l'AMP à un azote d'une lysine de l'enzyme. La deuxième étape est le transfert de l'AMP à un phosphate 5' de l'acide nucléique donneur. Ce transfert coupe la liaison AMP-enzyme et résulte en un acide nucléique adénylé, possédant à la fois un phosphate et un AMP en 5'. La troisième et dernière étape est l'attaque du phosphate du donneur adénylé par le groupement 3'OH de l'acide nucléique accepteur. La formation de ce lien relâche l'AMP dans le milieu et lie les deux acides nucléiques (accepteur et donneur) ensemble, avec un groupement phosphate entre eux. Ces informations sont plutôt pertinentes à garder en tête, car certaines ligases commerciales peuvent contourner certaines de ces étapes intermédiaires en ayant recours à des réactifs modifiés. Par exemple, la T4 ARN ligase 2 ne nécessite pas d'ATP pour faire la réaction, mais doit utiliser des acides nucléiques donneurs pré-adénylés (Wang *et al.*, 2006).

Les ligases d'ADN sont des outils de recherche polyvalents en biologie moléculaire (Eun, 1996). Généralement, leurs utilisations est associée à la ligation d'ADN double-brin (Eun, 1996). En revanche, la T4 ARN ligase 1 est employée afin de liguer à un fragment d'ARN simple-brin un adaptateur 3' en ADN ou ARN (Wang *et al.*, 2006; Zhuang *et al.*, 2012). Un autre type de ligation de ces ligases est la circularisation (Wang *et al.*, 2006). Lorsqu'un ARN possède à la fois un phosphate en 5' et un 3'-OH non-bloqué, l'enzyme peut joindre les deux extrémités et former un anneau à partir de la même molécule (Eun, 1996). Ce phénomène se produit lorsque le brin d'ARN fait un minimum de 8 nt en longueur (Eun, 1996). Une

particularité des ligases d'ARN est que l'accepteur (extrémité 3'OH) doit à tout prix être en ARN (ribonucléotide), alors que le donneur (extrémité 5' phosphate) peut être en ADN ou ARN (Miura *et al.*, 2019). Certains ouvrages ont rapporté des ligations ADN-ADN simple-brin, mais à des rendements très faibles (Miura *et al.*, 2019).

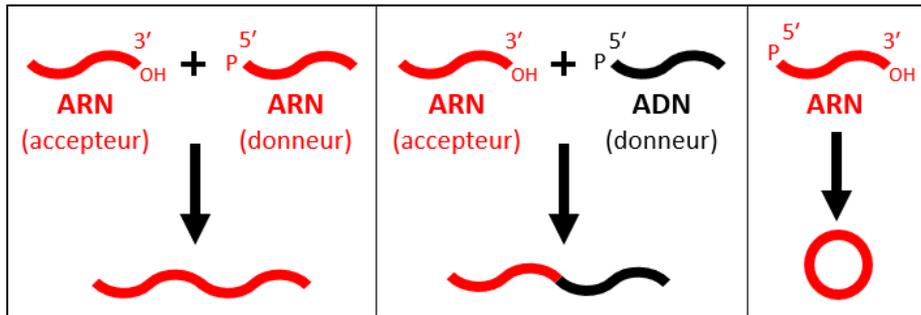


Figure 7: Schématisation des trois types de réactions et produits par la T4 ARN ligase 1.

### 1.4.3 Exonucléase lambda

Les nucléases sont des enzymes catabolisant la dégradation d'acide nucléique (Rittié et Perbal, 2008). Plus spécifiquement, elles clivent la liaison phosphodiester entre les groupements de sucres des bases (Rittié et Perbal, 2008). Les exonucléases sont actives aux extrémités (5' ou 3') des molécules d'acides nucléiques, alors que les endonucléases opèrent à l'intérieur même du brin (Rittié et Perbal, 2008). L'enzyme exonucléase lambda ( $\lambda$ ) provient du virus bactériophage lambda (Sriprakash *et al.*, 1975). Cette nucléase d'ADN est impliquée dans la recombinaison génétique de l'ADN lors de l'infection virale (Marimuthu *et al.*, 2012; Sriprakash *et al.*, 1975). L'enzyme reconnaît les extrémités 5' phosphorylées d'ADN double-brin et digère ce brin en direction 5'  $\rightarrow$  3' (Marimuthu *et al.*, 2012). Elle préfère, par un facteur de 20, les extrémités phosphorylées comparativement aux non-phosphorylées, et obtient un rendement beaucoup plus bas pour les ADN simple-brin phosphorylés (Marimuthu *et al.*, 2012).

En biologie moléculaire, cette digestion à conditions bien précises rend cette enzyme idéale pour la régénération d'amplicons PCR en ADN simple-brin lors des cycles SELEX (Marimuthu *et al.*, 2012). Ce retour au simple-brin permet ainsi aux ADN de retrouver leurs structures secondaires et tertiaires, ce qui leur permet à nouveau d'interagir avec leur ligand (Marimuthu *et al.*, 2012). Ceci est fait en utilisant une amorce antisens de PCR avec un groupement phosphate en 5'. Ainsi, seul le brin matrice possède un phosphate en 5' et seul celui-ci est ciblé par l'exonucléase lambda.

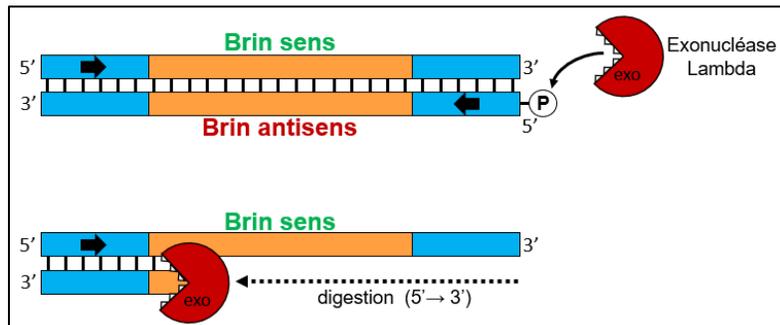


Figure 8: Schématisation de la digestion d'un ADN double-brin phosphorylé par l'exonucléase lambda.

## **1.5 Problématique**

En date d'écriture de ce mémoire, seuls quelques aptamères de longue taille ont été découverts. Ceci est en partie dû à la limite en longueur lors de synthèse des librairies dégénérées chez les fournisseurs commerciaux. En effet, la synthèse chimique d'oligonucléotides aléatoire limite le nombre de nucléotides à cause d'une importante diminution de rendement associée à une longue synthèse. Cependant, il pourrait être avantageux d'employer des librairies ayant une région dégénérée beaucoup plus longue. Celles-ci mèneraient potentiellement à des complexités de structures plus élevées. Il est donc question d'explorer une nouvelle approche permettant de synthétiser, de façon efficace, de longs oligonucléotides dégénérés pouvant agir comme des librairies dans un contexte de génération d'aptamères.

## **1.6 Hypothèses et objectifs**

L'hypothèse de l'ouvrage présent est, dans un premier temps, qu'il est possible de synthétiser enzymatiquement de longues librairies et, dans un deuxième temps, que celles-ci permettront de sélectionner des aptamères plus efficaces en termes de force de liaison ( $K_D$ ).

Trois objectifs ont été posés. Le premier était d'optimiser la synthèse de librairies de longue taille amplifiables par PCR. Le deuxième était d'effectuer des cycles SELEX sur ces librairies en parallèle avec des librairies classiques de courte taille. Un séquençage suivi d'une analyse bioinformatique sommaire ont aussi été faits pour ressortir les potentiels candidats issus de ces SELEX. Le dernier objectif était de comparer les candidats des différentes librairies afin de vérifier si la longueur des régions dégénérées joue un rôle sur l'affinité de liaison entre les librairies et leur cible.

## 2 MATÉRIEL ET MÉTHODES

---

### 2.1 Synthèse des bibliothèques

#### 2.1.1 Fluorescence

La plupart des colorants communs en biologie moléculaire s'accrochent beaucoup plus efficacement à l'ADN double-brin que simple-brin. Puisque la majorité des manipulations utilise de l'ADN simple-brin, il était important d'employer une méthode sensible et efficace afin de pouvoir mesurer la longueur des ADN simple-brin. Pour ce faire, des amorces à séquence connue avec une modification de fluorescéine en 5' ont été utilisées. La fluorescéine est un fluorophore ayant une excitation à 492 nm et une émission à 516 nm (Christopoulos et Diamandis, 1996). Le Typhoon™ FLA 9500 a été utilisé pour détecter cette fluorescence. L'excitation était à 473 nm et le filtre d'émission était le BPB1 (520-540 nm). La quantité minimale d'ADN marqué détectable par cette machine était d'environ 50 fmol lorsque l'échantillon est sous forme de gouttelettes sur plaque de verre, et de 10 fmol lorsque migré dans un gel de polyacrylamide dénaturant.

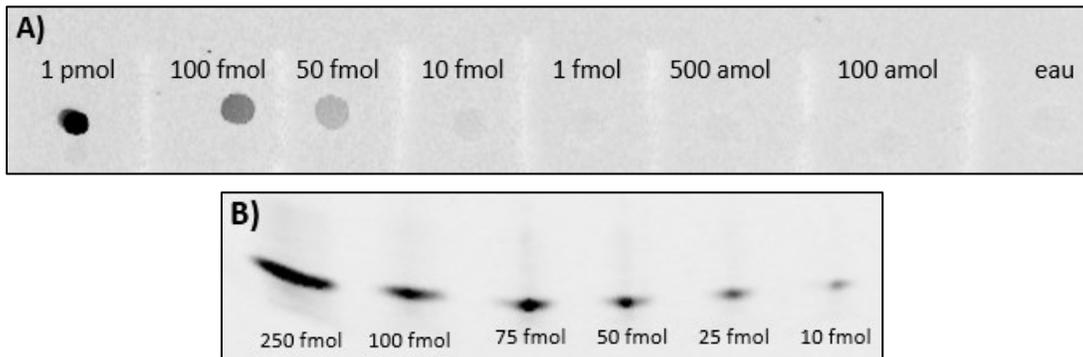


Figure 9: Limite minimale de détection de la fluorescéine par le Typhoon.

A) ADN contenu dans des gouttelettes placées sur verre B) ADN migré dans un gel dénaturant urée-PAGE 6%.

#### 2.1.2 Échelle de poids moléculaire

Les échelles de poids moléculaire des gels agaroses ne sont pas fluorescentes, il est donc important de trouver une autre façon de mesurer la longueur des ADN. Le gène ribosomal 16S bactérien de 1550 paires de bases (pb) a servi à faire les trois PCR successives suivantes: 1) PCR 16S classique sur des bactéries *Escherichia coli* DH5 $\alpha$  entière (amorce 8F avec 518R, 805R, ou 1092R), 2) PCR indexée utilisant la PCR précédente comme matrice, et une amorce sens 8F ayant une protubérance en 5' (amorce index 8F avec

518R, 805R, ou 1092R), 3) PCR classique utilisant la PCR précédente comme matrice, et l'amorce TdT 1 fluorescente (amorce TdT 1 avec 518R, 805R, ou 1092R). Toutes les séquences des amorces utilisées sont disponibles à l'Annexe I. Suite à ces PCR successives, trois fragments de 528, 815, et 1110 pb ont été obtenus comme point de repère pour mesurer la longueur des librairies.

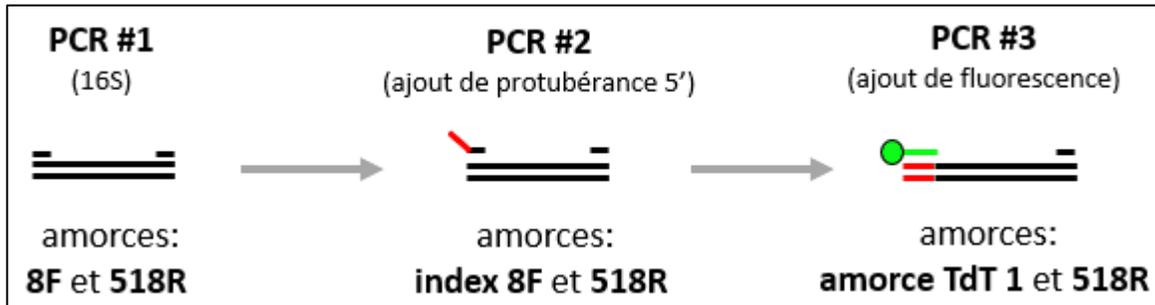


Figure 10: Schéma des trois PCR successives menant à un amplicon fluorescent.

### 2.1.3 Librairie enzymatique - terminale transférase ( $N_{800}$ )

#### 2.1.3.1 Synthèse de la région aléatoire

La polymérase terminale transférase a été utilisée pour synthétiser la région aléatoire à partir d'un oligonucléotide fluorescent de 18 nucléotides (amorce TdT 1). Voici les concentrations finales des réactifs de la réaction enzymatique TdT: tampon 1X (50 mM acétate de potassium, 20 mM Tris-acétate, 10 mM acétate de magnésium, pH 7,9 à 25°C), 0,25 mM  $CoCl_2$ , 2 mM dNTP, 120 unités de terminale transférase (fournisseur: *New England BioLabs Inc.*, #cat: M0315S, 20 unités  $\approx$  8,16 pmol), 0,8  $\mu$ M d'amorce TdT 1 fluorescente, volume final ajusté à 60  $\mu$ L avec de l'eau Milli-Q. Voici les conditions de réaction enzymatique TdT: une fois les réactifs ajoutés, le mélange a bien été brassé par aspiration avec une pipette, puis a été incubé à 37°C pendant 2 heures suivi d'une inactivation à 80°C pendant 20 minutes (par thermocycleur). Le mélange a ensuite été mélangé avec 1 volume de tampon de chargement (15 mM EDTA dans du formamide 100%, sans colorant ajouté, car la couleur interfère avec la fluorescence). L'échantillon a ensuite été réparti dans six puits d'un gel dénaturant urée-PAGE 6%. Les trois marqueurs de poids moléculaire (528, 815, et 1110 nucléotides en longueur) ont été migrés à côté des échantillons enzymatiques pour une comparaison facile. Dimension des vitres du gel PAGE: 24,5 cm x 32,5 cm (horizontal x vertical). Dimension des puits du gel PAGE: 0,4 cm x 2,0 cm (largeur x profondeur). La migration a été faite à 500 V (32 mA et 16 W) pendant 75 minutes à température pièce. Une des plaques de vitre a ensuite été retirée sans glisser des bulles sous le gel, puis a été scannée au Typhoon™ FLA 9500

(version du logiciel: 1.0). Voici les paramètres utilisés du Typhoon: force du laser à 1000 V, résolution de 200  $\mu\text{m}$ , et filtres 473LPR (473 nm) et BPB1 (520-540 nm). Les bandes majeures du mélange (s'étalant de  $\sim 500$  à 1100 nucléotides en longueur) ont ensuite été extraites du gel avec une lame de rasoir. L'ADN a ensuite été élué des bouts de gel par électroélution, puis l'ADN a été précipité par éthanol. Le protocole détaillé de l'électroélution est disponible à l'Annexe II et celui de la précipitation à l'éthanol est à l'Annexe III.

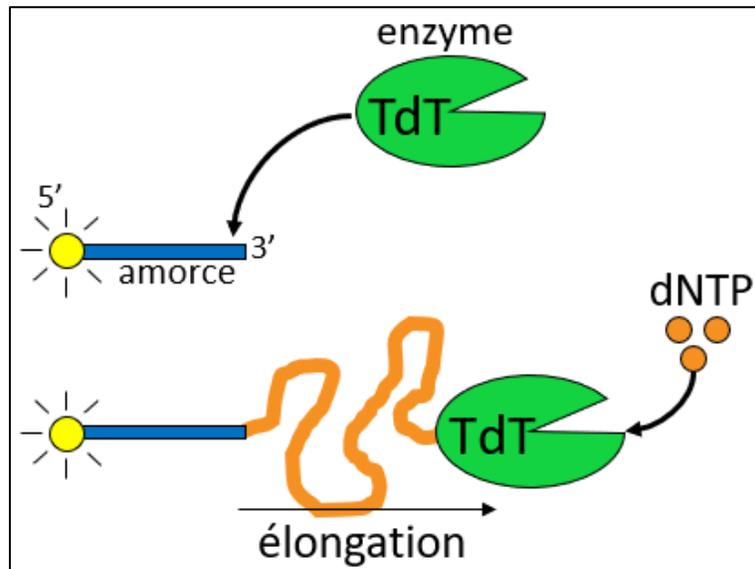


Figure 11: Schématisation de l'allongement enzymatique d'une amorce fluorescente à l'aide de la terminale transférase et de dNTP.

### 2.1.3.2 Ajout d'une courte queue d'ARN (riboqueue)

L'ADN purifié consiste maintenant en une courte séquence connue de 18 nucléotides en 5' suivie d'une longue queue d'ADN dégénéré. Pour rendre ces ADN amplifiables par PCR, il est nécessaire d'ajouter une deuxième séquence connue en 3' qui agira comme assise d'amorce antisens. La T4 ARN ligase 1 a été employée pour catalyser la ligation de cette assise en 3' des fragments d'ADN simple-brin. Cette ligase se distingue de la T4 ADN ligase en agissant sur des acides nucléiques simple-brin plutôt que double-brin. Une importante limite de cette enzyme est qu'elle nécessite un accepteur (3'OH) en ARN. Une équipe de chercheur japonais ont contourné cette limitation en utilise la terminale transférase, mais en employant du rATP plutôt que du dNTP (Miura *et al.*, 2019). Comme mentionné dans l'introduction (section enzyme), la terminale transférase n'ajoute que <5 ribonucléotides lorsqu'il s'agit d'ARN (Miura *et al.*, 2019). Ces quelques ajouts sont suffisants pour « tricher » la T4 ARN ligase 1 à penser que l'accepteur est en ARN

alors que ce n'est le cas que pour les quelques nucléotides en 3' (Miura *et al.*, 2019). Voici les concentrations finales des réactifs de la réaction enzymatique de *ribotailing*: tampon 1X (50 mM acétate de potassium, 20 mM Tris-acétate, 10 mM acétate de magnésium, pH 7,9 à 25°C), 0,25 mM CoCl<sub>2</sub>, 0,4 mM ATP, 20 unités de terminale transférase (fournisseur: *New England BioLabs Inc.*, #cat: M0315S), volume final ajusté à 10 µL avec l'ADN purifié resuspendu dans de l'eau Milli-Q. Voici les conditions de réaction enzymatique de *ribotailing*: une fois les réactifs ajoutés, le mélange a bien été brassé par aspiration avec une pipette, puis a été incubé à 37°C pendant 1 heure suivi d'une inactivation à 80°C pendant 20 minutes (par thermocycleur).

### 2.1.3.3 Ligation d'adaptateur en 3'

L'ADN possède maintenant une courte queue d'ARN et est prêt à recevoir une séquence connue d'ADN simple-brin (aussi appelé adaptateur) en 3' par ligation. Il est recommandé d'utiliser un adaptateur possédant une modification phosphate en 5' et en 3'. Le phosphate en 5' est nécessaire à la réaction enzymatique de ligation, alors que celui en 3' sert comme bloqueur pour empêcher des ligations répétées d'adaptateur *en tandem*. Voici les concentrations finales des réactifs de la réaction enzymatique de ligation: 3 µM d'adaptateur 1, 10% DMSO, tampon 1X (50 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT, pH 7,5 à 25°C), 1 mM ATP, 7,5 unités de T4 ARN ligase 1 (fournisseur: *New England BioLabs Inc.*, #cat: M0204S), 20% PEG8000, compléter à 15 µL avec l'ADN avec *ribotail*. Voici les conditions de réaction enzymatique de ligation: une fois les ingrédients ensemble, il est très important de bien homogénéiser le mélange. Le PEG8000 rend le milieu très visqueux, et il est recommandé de l'agiter en tournant l'embout d'une pipette, ensuite par de lentes aspirations. La présence de bulles n'est pas grave, tant que le mélange est bien mélangé. Une fois bien homogène, le mélange a été incubé à 16°C pendant 16 heures (pas d'inactivation nécessaire, cela pourrait endommager la jonction d'ARN). À noter qu'une fragile jonction d'ARN est présente entre la région aléatoire et la région conservée en 3', et ainsi, il ne faut pas trop attendre pour la prochaine étape afin de limiter la dégradation.

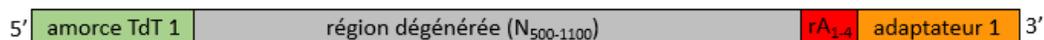


Figure 12: Schématisation de l'anatomie complète de la librairie TdT avant son amplification.

#### 2.1.3.4 Amplification de la librairie par RT-PCR

Les molécules d'ADN possèdent maintenant une région conservée aux extrémités 5' et 3'. Elles sont ainsi amplifiables par PCR. Dû à la petite section d'ARN entre la région aléatoire et la région conservée 3', une RT-PCR a été faite. Un protocole classique de RT a été utilisé pour synthétiser le brin complémentaire d'ADN (trousse: SuperScript™ IV Reverse Transcriptase, Fournisseur: ThermoFisher, #cat: 18090010). L'amorce antisens 1, complémentaire à l'adaptateur 1, a été utilisée, et non pas l'hexamère (N<sub>6</sub>) fourni par le fabricant. L'ADNc a ensuite été amplifié par PCR classique. Voici les concentrations finales des réactifs de la réaction de la PCR: 1 µM d'amorce TdT 1, 1 µM d'amorce antisens 1, tampon 1X (20 mM Tris-HCl pH 8,8, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 10 mM KCl, 20 mM MgSO<sub>4</sub>, 1% Triton X-100), 0,2 mM dNTP, 1,25 unité de Taq ADN polymérase (Fournisseur: Université de Sherbrooke, Service de Purification de Protéines [SPP]), 2 µL du mélange RT, volume ajusté à 50 µL avec de l'eau Milli-Q, mélange répété dans huit tubes différents. Il est à noter que l'amorce antisens possède un phosphate en 5', car cette modification est requise pour pouvoir digérer le brin antisens lors du retour à l'ADN simple-brin au cours des cycles SELEX. Voici les informations sur les cycles de la PCR: dénaturation initiale à 95°C pour 3 minutes. 35 cycles de: 1) dénaturation à 95°C pour 30 secondes, 2) appariement des amorces à 60°C pendant 30 secondes, 3) élongation à 72°C pendant 90 secondes. Et enfin, une élongation finale à 72°C pendant 5 minutes. Les huit tubes PCR ont ensuite été mis ensemble dans un seul tube eppendorf de 1.5 mL.

#### 2.1.3.5 Cycle final

Il est commun d'employer 25 à 35 cycles lors d'une PCR classique. Par contre, tous ces cycles ne mènent pas nécessairement à une amplification d'ADN (Nehdi *et al.*, 2020). Au cours de la PCR, la quantité d'ADN plafonne éventuellement aux alentours du 20<sup>ème</sup> cycle dû à un manque d'amorces (Nehdi *et al.*, 2020). Les cycles restants sont donc simplement des répétitions de dénaturation et de réappariement d'ADN (Nehdi *et al.*, 2020). Dans le contexte d'une PCR amplifiant par exemple un gène spécifique, ceci ne mène pas à des problèmes, car les seules populations d'ADN présentes sont les brins matrices et leur brin complémentaire; retrouver leur copie complémentaire est une simple tâche. Par contre, dans le cas d'une PCR amplifiant une librairie, les molécules initiales d'ADN possèdent chacune leur propre séquence unique. Ceci mène à des appariements imparfaits lors des cycles finaux, puisqu'il n'est pas si simple de retrouver sa séquence complémentaire parfaite parmi plusieurs milliards de différentes séquences environnantes (Nehdi *et al.*, 2020). Les assises d'amorce sont identiques d'une molécule à une autre, mais pas la séquence dégénérée du milieu (Nehdi *et al.*, 2020). Ceci mène à des ADN partiellement double-brin et majoritairement simple-brin, causant des différents motifs de migration (Nehdi *et al.*, 2020). Pour

contourner cette difficulté, il est possible de faire une PCR à 35 cycles normalement, puis prendre une portion de ce mélange PCR, le transférer dans un nouveau mélange PCR (avec de nouvelles amorces), puis effectuer un seul cycle PCR. Voici les concentrations finales des réactifs de la PCR finale: 1  $\mu\text{M}$  d'amorce TdT 1, 1  $\mu\text{M}$  d'amorce antisens 1, tampon 1X (20 mM Tris-HCl pH 8,8, 10 mM  $(\text{NH}_4)_2\text{SO}_4$ , 10 mM KCl, 20 mM  $\text{MgSO}_4$ , 1% Triton X-100), 0.2 mM dNTP, 2,5 unités de TAQ ADN polymérase (Fournisseur: Université de Sherbrooke, Service de Purification de Protéines [SPP]), 10  $\mu\text{L}$  du mélange PCR précédent, volume ajusté à 100  $\mu\text{L}$  avec de l'eau Milli-Q, mélange répété dans 3 tubes différents. Voici les informations sur le cycle unique de la PCR finale: dénaturation à 95°C pour 5 minutes, appariement des amorces à 60°C pendant 5 minutes, élongation à 72°C pendant 5 minutes. Il est crucial de ne pas réchauffer à 95°C après la synthèse du brin complémentaire à 72°C, car l'ADN est maintenant parfaitement double-brin, et chauffer retourne au problème initial. Les trois tubes des PCR finales ont ensuite été mélangés ensemble.

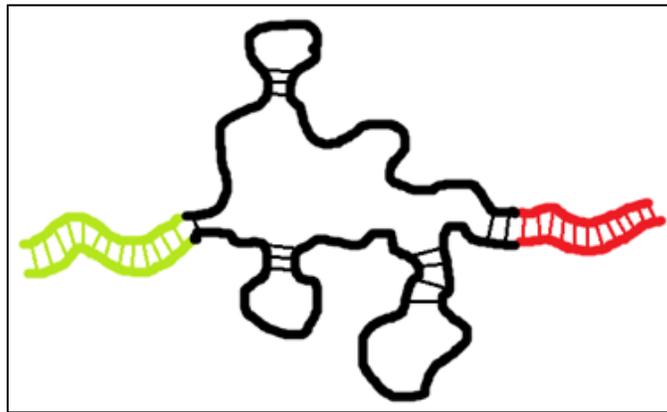


Figure 13: Schématisation du problème d'amplification de bibliothèques par PCR.

Couleur verte: région de l'amorce sens (ADN parfaitement double-brin). Couleur rouge: région de l'amorce antisens (ADN parfaitement double-brin). Couleur noire: région dégénérée non-complémentaire d'ADN non-apparié (ADN surtout simple-brin).

### 2.1.3.6 Gel agarose et quantification

Un échantillon de la PCR finale et de la PCR initiale (sans cycle final) ont été migrés côte à côte dans un gel agarose 2% TAE (120 V, 45 minutes). Ceci permet une comparaison entre de l'ADN simple-brin et double-brin. Le gel a ensuite été visualisé par *Gel Doc* (modèle: Gel Doc XR+ System, compagnie: Bio-Rad, logiciel: Image Lab™, version: 5.2.1). Un signal d'ADN diffus est attendu pour la PCR initiale (*smear* non-défini, tout au long du gel), alors qu'une section plus foncée bien délimitée (de ~500 à ~1500 pb) ressort pour le puits du cycle final. Cette section représente de l'ADN double-brin. Cette vérification

permet de confirmer que l'amplicon de la PCR est bel et bien de la librairie, et non pas des artefacts de mésappariements d'ADN.

Pour avoir une idée de la quantité d'ADN présents dans la librairie, la région foncée bien délimitée du gel a été dosée puis comparée à une des bandes du poids moléculaire (chaque bande a un nombre de ng lui étant propre, disponible sur le site du fournisseur). Le logiciel employé pour le dosage de la photo de gel était ImageJ (version 1.53c), développé par le NIH.

#### **2.1.3.7 Retour à ADN simple-brin (digestion)**

Les amplicons de PCR sont double-brin, mais une librairie simple-brin est requise pour effectuer des cycles SELEX. Le mélange PCR finale a donc été digéré par l'enzyme exonucléase lambda (fournisseur: *New England BioLabs Inc.*, #cat: M0262S). Cette enzyme est compatible avec le tampon de la PCR, donc simplement 40 unités de l'enzyme ont été ajoutées à 202,5 µL du mélange PCR final. Le tube a ensuite été mélangé par aspiration de pipette, puis incubé à 37°C pendant 1 heure, suivi d'une inactivation à 75°C pendant 10 minutes. Le contenu de ce tube est maintenant considéré comme de la librairie d'ADN simple-brin, prête d'utilisation. Seules ~1 pmol de ce tube ont été utilisées pour chaque SELEX (trois différents SELEX en total, un pour chaque cible), le reste a été conservé à -20°C.

#### **2.1.4 Librairie enzymatique - T4 ARN ligase 1 (N<sub>200</sub>)**

##### **2.1.4.1 Synthèse de la région aléatoire**

La deuxième méthode explorée pour la production de librairie de longue taille était la ligation enchaînée de plusieurs fragments d'ADN dégénéré, un à la suite de l'autre. Dans un contexte de ligation classique, ces types de répétitions successives se dénomme « concatémère ». Un concatémère est défini comme étant un fragment d'ADN contenant plusieurs copies d'une même séquence, et ce répétée en série (exemple: ATGATGATGATGATG...). En biologie moléculaire, la formation de concatémères *en tandem* lors des ligations est considérée comme nuisible, et diverses stratégies sont employées afin de limiter le plus possible leur apparition. Dans l'ouvrage présent, ce phénomène a été poussé afin de vérifier s'il était possible de lier plusieurs fragments dégénérés ensemble afin d'obtenir une longue région aléatoire.

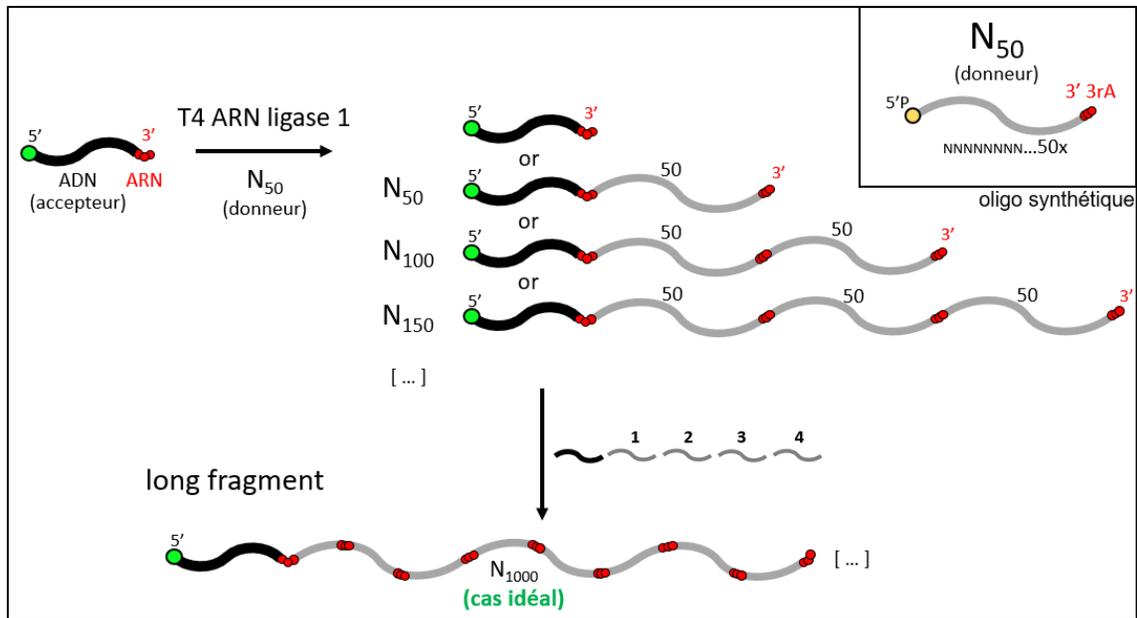


Figure 14: Schématisation de la création d'une longue région dégénérée par ligation.

Les boules vertes représentent une modification fluorescéine, les boules jaunes des phosphates (P), et les boules rouges des riboadénines (rA). Les fragments noirs sont des ADN à séquence connue (accepteur), alors que les fragments gris sont à séquence inconnue (donneur dégénéré).

Rappelons que la T4 ARN ligase 1 requiert quelques ARN terminaux en 3' de l'ADN accepteur, et un phosphate en 5' de l'ADN donneur (Miura *et al.*, 2019). Ainsi, pour que cet assemblage se fasse dans le bon ordre, l'amorce fluorescéine initiale (dénommée « amorce ligation 2 ») ne doit que pouvoir recevoir des fragments d'ADN en son 3', alors que les donneurs  $N_{50}$  doivent pouvoir s'accrocher à d'autres molécules par leur 5' et en 3'. Par conséquent, l'amorce ligation 2 possède une modification fluorescéine en 5' (sert de marquage et de blocage), et trois riboadénine en 3' pour permettre la ligation simple-brin. Les donneurs  $N_{50}$  possèdent un phosphate en 5' et trois riboadénine en 3'. Les conditions de ligation sont identiques à celles de la section 2.1.3.3, à l'exception de ces ajouts ou changements: 0,25  $\mu\text{M}$  ADN accepteur (amorce ligation 2), 2,5  $\mu\text{M}$  ADN donneur ( $N_{50}rArAr$ ), 0,4 mM ATP, 96 unités de T4 ARN ligase 1, volume final ajusté à 240  $\mu\text{L}$  avec de l'eau Milli-Q, et une incubation à 37°C pendant 1 heure sur un bloc chauffant. Le mélange de ligation a été mélangé avec 1 volume de tampon de chargement (15 mM EDTA dans du formamide 100%, sans colorant ajouté, car la couleur interfère avec la fluorescence), puis a été chargé dans plusieurs puits d'un gel dénaturant urée-PAGE 6%. Afin d'obtenir une meilleure résolution de bandes et ne pas surcharger les puits de fluorescence, il est recommandé de ne pas dépasser un volume de 20  $\mu\text{L}$  dans les puits (10  $\mu\text{L}$  échantillon avec 10  $\mu\text{L}$  de tampon de chargement). Dimension des vitres du gel PAGE: 24,5 cm x 32,5 cm (horizontal x vertical). Dimension des puits du gel PAGE: 0,4 cm x 2,0 cm

(largeur x profondeur). La migration a été faite à 500 V (32 mA et 16 W) pendant 60 minutes à température pièce. Une des plaques de verre a ensuite été retirée sans glisser des bulles sous le gel, puis le gel a été scanné au Typhoon™ FLA 9500 (version du logiciel: 1.0). Voici les paramètres utilisés du Typhoon: force du laser à 1000 V, résolution de 200 µm, et filtres 473LPR (473 nm) et BPB1 (520-540 nm). Il est attendu de voir un profil de migration sous forme d'échelle. Une bande initiale de ~20 nucléotides est visible au bas du gel, puis des bandes s'ajoutant à chaque bond de 50 nucléotides (20, 70, 120, 170...), tout en diminuant en intensité plus l'on s'approche du haut du gel. Les bandes majeures de 220 nucléotides (quatre N<sub>50</sub> ou N<sub>200</sub>) des puits ont été extraites du gel avec une lame de rasoir (dimension d'environ 1 cm x 0,5 cm). L'ADN a ensuite été élué du gel avec une solution d'eau Milli-Q contenant 0,3 M de NaCl (500 µL avec deux bouts de gel dans un tube eppendorf 1.5 mL) sur un rotor pendant 16 heures à 4°C. L'éluat a ensuite été précipité par éthanol (**Annexe III**).

#### **2.1.4.2 Ligation d'adaptateur en 3'**

L'ADN possède déjà une courte queue d'ARN (les N<sub>50</sub> synthétiques ont trois rA en 3') et est prêt à recevoir une séquence connue d'ADN simple-brin (aussi appelé adaptateur ou *linker*) en 3' par ligation. Il est recommandé d'utiliser un adaptateur possédant une modification phosphate en 5' et en 3'. Le phosphate en 5' est nécessaire à la réaction enzymatique de ligation, alors que celui en 3' sert comme bloqueur pour empêcher des ligations répétées d'adaptateur *en tandem*. Les conditions de ligation sont identiques à celles de la section 2.1.3.3, à l'exception de ces ajouts ou changements: 0,5 µM d'adaptateur 2, 0,4 mM ATP, 10 unités de T4 ARN ligase 1, compléter à 25 µL avec l'ADN purifié de la section 2.1.4.1 (la plus grande quantité d'ADN possible), mélange incubé à 37°C pendant 1 heure.

#### **2.1.4.3 Amplification de la librairie par RT-PCR**

Les molécules d'ADN possèdent maintenant une région conservée aux extrémités 5' et 3'. Elles sont ainsi amplifiables par PCR. Dû aux multiples jonctions d'ARN entre chacun des N<sub>50</sub>, une RT-PCR a été fait. La réaction de RT a été fait dans les mêmes conditions que la section 2.1.3.4, à l'exception de ces changements ou ajouts: l'amorce antisens 2 (complémentaire à l'adaptateur 2) a été utilisée, ajout de 20 unités d'inhibiteur d'ARNase *RiboLock* (fournisseur: Thermo Scientific™, #cat: EO0381), et le volume a été complété avec 9,95 µL du mélange de ligation N<sub>50</sub> de la section 2.1.4.2. L'ADNc a ensuite été amplifié par PCR avec les mêmes conditions que la section 2.1.3.4 à l'exception des ajouts ou changements suivants : 1 µM d'amorce ligation 2, 1 µM d'amorce antisens 2, 2,5 unités de Taq ADN polymérase, volume ajusté à 100 µL, et élongation à 72°C pendant 30 secondes. Un total de trois tubes PCR a été fait, tous ont

les mêmes concentrations et conditions. Puisque cette librairie ( $N_{200}$ ) n'a pas la même taille que la librairie TdT ( $N_{500-1100}$ ), il n'est pas nécessaire de faire un cycle PCR final pour la visualiser sur gel. Un échantillon de la PCR a été migré dans un gel agarose et puis dosé, dans les mêmes conditions décrites à la section 2.1.3.6.

#### **2.1.4.4 Retour à ADN simple-brin (digestion)**

Les amplicons de PCR sont double-brin, mais une librairie simple-brin est requise pour effectuer des cycles SELEX. Le mélange de PCR finale a donc été digéré par l'enzyme Exonucléase lambda de la même façon décrite qu'à la section 2.1.3.7 à l'exception que l'enzyme a été ajouté à 135  $\mu$ L du mélange PCR. Seules  $\sim 60$  pmol d'ADN de ce tube a été utilisé pour chaque SELEX (3 différents SELEX en total), le reste a été conservé à  $-20^{\circ}\text{C}$ .

#### **2.1.5 Librairies chimiques ( $N_{20}$ et $N_{50}$ )**

Puisque le but des travaux présents est de comparer des aptamères de courtes et longues tailles, des cycles SELEX ont été effectués en parallèle sur plusieurs librairies variant en longueur. En plus des librairies enzymatiques de longue taille présentées ci-haut, deux librairies synthétisées chimiquement ont été commandées du fournisseur: *Integrated DNA Technologies, Inc* (IDT). Une librairie chimique possède une région dégénérée de 20 nucléotides ( $N_{20}$ ) et l'autre de 50 ( $N_{50}$ ). Afin de s'assurer de l'intégrité des librairies chimiques avant le début des SELEX, une vérification PCR a été effectuée sur les librairies  $N_{20}$  et  $N_{50}$ . La même méthodologie que la section 2.1.3.4 a été utilisée à l'exception des changements ou ajouts suivants: pas de RT nécessaire, 1  $\mu$ M d'amorce sens 3 ou 4 (3:  $N_{20}$ , 4:  $N_{50}$ ), 1  $\mu$ M d'amorce antisens 3 ou 4, 0,375 unité de TAQ ADN polymérase, 100 pM de librairie chimique  $N_{20}$  ou  $N_{50}$  comme matrice (paire d'amorce 3 et 4 respectivement), volume ajusté à 15  $\mu$ L avec de l'eau Milli-Q, et élongation à  $72^{\circ}\text{C}$  pendant 30 secondes. Les deux échantillons ont ensuite été migrés sur gel agarose (même méthode que la section 2.1.3.6) simplement pour vérifier si les amplicons avaient la longueur attendue ( $\sim 60$  pb pour la librairie  $N_{20}$  et  $\sim 90$  pb pour la librairie  $N_{50}$ ). À noter que ces échantillons PCR n'ont pas été utilisés comme librairie de départ pour les SELEX, cette étape n'était qu'à titre vérificatif.

## **2.2 Ligands**

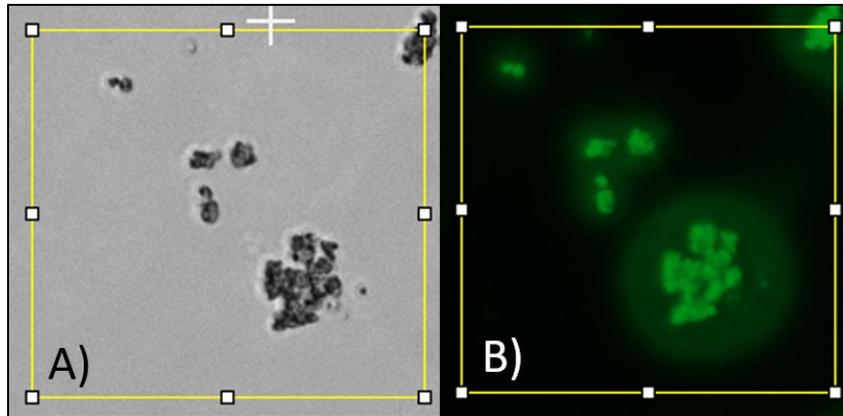
### **2.2.1 Choix des cibles**

L'objectif des travaux présents était de démontrer un avantage à utiliser des librairies de longue taille. Ceci peut être fait en utilisant n'importe quel ligand. Par contre, certains ligands sont plus faciles à

manipuler et à se procurer en grande quantité. Les protéines sont idéales dans le cadre de ce type de travaux, car elles sont moléculairement de grande taille comparativement à un ion, par exemple. Ainsi, la distinction entre un complexe aptamère-ligand et un aptamère seul sera plus facile à discerner. En soi, le choix du ligand n'est pas réellement important, mais nous avons la chance de disposer de plusieurs collaborations avec d'autres laboratoires travaillant sur plusieurs protéines. Trois protéines ont été choisies comme cibles d'intérêts: la galectine-7 (fournisseur: laboratoire de Yves St-Pierre, INRS), les protéiques NheA et HblB (fournisseur: laboratoire de Nadja Jeßberger, Ludwig-Maximilians), et la nouvelle protéine HP0953 (fournisseur: laboratoire de Charles Calmettes, INRS). La galectine-7 est une protéine impliquée dans les cancers. Sa surexpression est souvent associée à la progression et prolifération tumorale, et cause aussi l'apoptose de cellules T (Pham *et al.*, 2021). Les protéines NheA et HblB sont des sous-unités de toxines issues de la bactérie *Bacillus cereus* (Gdoura-Ben Amor *et al.*, 2019). Ces toxines sont souvent associées à des intoxications alimentaires provenant de viandes, de légumes, de riz et de pâtes (Gdoura-Ben Amor *et al.*, 2019). La protéine HP0953 a nouvellement été isolée de la bactérie *Helicobacter pylori*, un pathogène colonisant la muqueuse gastrique (Wen *et al.*, 2003). La protéine semble sécrétée hors de la cellule lors d'un stress acide à des fins de défense contre l'hôte (Wen *et al.*, 2003).

### **2.2.2 Fonctionnalisation de billes magnétiques**

Les ligands choisis, il était maintenant question de trouver une façon permettant de séparer les séquences des bibliothèques capables d'interactions, et celles n'interagissant pas avec la cible. Des billes magnétiques avec des groupements NHS (*N*-hydroxysuccinimide) en leur surface ont été employées pour ceci. Lorsqu'un groupement NHS vient en contact avec un groupement amine ( $\text{NH}_2$ ), une réaction chimique se produit et forme une liaison covalente avec l'amine (Klykov et Weller, 2015) tout en relâchant le NHS dans le surnageant. La grande majorité des protéines contiennent des acides aminés possédant des groupements amines, ainsi ce type de billes sont souvent employées pour la fonctionnalisation de protéines en leur surface (Klykov et Weller, 2015). Ci-dessous se trouvent les étapes du processus de la fonctionnalisation des ligands sur les billes magnétiques. Par contre, avant de procéder à la fonctionnalisation avec les protéines cibles, les conditions de la réaction ont été vérifiées avec un peptide fluorescent.



**Figure 15: Fonctionnalisation des billes magnétiques avec un peptide fluorescent.**

Des billes magnétiques NHS (200 nm de diamètre) ont été fonctionnalisées avec un peptide marqué à la fluorescéine (provenance: laboratoire de Nicolas Doucet, INRS). A) photo sans réflecteur, B) photo avec réflecteur à intensité de 10%. La dimension du carré jaune est d'environ 25 µm x 25 µm. L'excitation du fluorophore a été faite à 495 nm et l'émission à 519 nm. La photo a été prise avec un microscope optique inversé (modèle: Axio Observer.Z1 / 7, objectif utilisé: alpha Plan-Apochromat 100x/1.46 Oil DIC (UV) M27, logiciel: ZEN version 2,6).

Information sur les billes magnétiques NHS: concentration: 3% m/v (30 mg/mL) dans du DMSO, fournisseur: Galenvs, diamètre: ~200 nm. Le tube mère a bien été brassé, puis un volume de 83 µL a été transféré dans un tube eppendorf de 1.5 mL. Les billes ont été lavées deux fois avec 1000 µL de tampon de couplage afin d'éliminer le DMSO (tampon de couplage: 0,5 M de NaHCO<sub>3</sub>, 0,5 M de NaCl, pH ajusté à 8,3 avec du HCl et KOH, dans de l'eau Milli-Q). Un lavage consiste à ajouter la solution aux billes, vortexer le tube, placer le tube sur un support magnétique (les billes se collent à la paroi), puis enlever le surnageant par pipette. Il est bien important de vortexer les billes après chaque nouvel ajout de solution afin de bien répartir les billes dans le tube (elles ont tendance à sédimenter). Trois ligands ont été utilisés pour trois aliquots de billes séparées (1 mg/mL de galectine-7, 0,01 mg/mL de NheA et HblB chaque, et 1 mg/mL de HP0953). Une fois le tampon de couplage enlevé, les billes ont été incubées avec 1000 µL de ligand pendant 2 heures sur un rotor à 21°C. Les tubes ont été brièvement vortexés à chaque 15 minutes lors de l'incubation. Le surnageant a ensuite été retiré, puis les billes ont été lavées à deux reprises avec 1000 µL de tampon de couplage. Les billes ont ensuite été incubées pendant 1 heure à 21°C avec 1000 µL de tampon de blocage (tampon de blocage: 0,1 M Tris-base, pH 8.0 ajusté avec HCl/NaOH, dans de l'eau Milli-Q). Le Tris du tampon de blocage contient des groupements NH<sub>2</sub>, qui saturent les sites de liaison des billes non-couvertes par les protéines (il est très important de ne pas utiliser un tampon ayant des NH<sub>2</sub> avant l'incubation avec les ligands, comme le tampon dans lequel baignent les protéines par exemple). Le tampon de blocage a ensuite été retiré, puis les billes ont été lavées avec 1000 µL de tampon de lavage (tampon de lavage: 0,1 M Tris-base, 1 M NaCl, pH 8.0 ajusté avec HCl/NaOH, dans de l'eau Milli-Q). Les

billes ont ensuite été lavées avec 1000  $\mu\text{L}$  de tampon acide (tampon acide: 0,1 M d'acétate de sodium, 1 M de NaCl, pH 4.0 ajusté avec de l'acide acétique glacial, dans de l'eau Milli-Q). La fonctionnalisation maintenant complétée, les billes ont été resuspendues dans 1000  $\mu\text{L}$  de tampon de SELEX (tampon de SELEX: 20 mM Tris-base, 100 mM NaCl, 5 mM KCl, 2 mM  $\text{CaCl}_2$ , 1 mM  $\text{MgCl}_2$ , pH 7,4 ajusté avec du HCl/NaOH, dans de l'eau Milli-Q). Il faut noter que le tampon de SELEX est simplement le tampon utilisé lors des futurs cycles SELEX, il est entièrement à la discrétion de l'utilisateur. Le tube contient maintenant 2,5 mg/mL de billes magnétiques fonctionnalisées avec les cibles protéiques. En plus des trois tubes contenant chacune des cibles, un quatrième tube a été préparé en parallèle en n'utilisant aucune protéine. Ce dernier contient simplement des billes bloquées par le tampon de blocage (à base de  $\text{NH}_2$ ) et elles ont été utilisées lors des cycles négatifs pour se débarrasser des séquences se liant directement aux billes, plutôt qu'aux cibles sur les billes.

## 2.3 SELEX

### 2.3.1 Étapes des cycles

Les cycles SELEX sont en fait des incubations successives cycliques en changeant quelques paramètres à chaque cycle. Un cycle SELEX comprend plusieurs étapes répétées pendant 15 cycles, chacun des cycles amenant une modification de paramètre afin d'augmenter la stringence de la sélection. La première étape consiste à incuber les billes magnétiques fonctionnalisées avec une librairie. Ces incubations ont été faites dans un tube eppendorf de 1,5 mL à température pièce (21°C) sur rotor (voir le **Tableau 2** pour le temps d'incubation). Le tube contenait les billes fonctionnalisées (préalablement brassé avant le pipetage), la librairie, les protéines négatives (à partir du cycle 5), le tout dans un volume final ajusté à 500  $\mu\text{L}$  avec du tampon SELEX (tampon de SELEX: 20 mM Tris-base, 100 mM NaCl, 5 mM KCl, 2 mM  $\text{CaCl}_2$ , 1 mM  $\text{MgCl}_2$ , pH 7,4 ajusté avec du HCl/NaOH, dans de l'eau Milli-Q). Suite à l'incubation, le tube a été placé sur un support magnétique (les billes se collent contre la paroi) afin de permettre d'éliminer le surnageant par pipette. Le surnageant contenant les séquences d'ADN non-collés aux billes a ensuite été archivé à 4°C. Les billes « sèches » ont ensuite été lavées afin de libérer les séquences faiblement accrochées. Un lavage consiste à ajouter 500  $\mu\text{L}$  de tampon SELEX au tube, vortexer le tube pendant 15 secondes, incuber le tube sur rotor à 21°C (voir **Tableau 2** pour les temps d'incubation), replacer le tube sur le support magnétique, jeter (archiver) le surnageant, puis répéter le lavage au besoin (le nombre de lavages à chaque cycle est mentionné dans le **Tableau 2**). Un volume de 15  $\mu\text{L}$  d'eau Milli-Q est ajouté aux billes « sèches », suivi d'un brassage par vortex puis d'une brève centrifugation par *minifuge*. Cette étape

permet de récupérer toutes les billes, puis de les placer en suspension dans un petit volume de liquide dans le fond du tube. Le tube est ensuite chauffé à 95°C sur bloc chauffant pendant 10 minutes afin d'éluer les ADN collés sur les billes. Suite au chauffage, le tube est rapidement sorti du bloc chauffant (la température dans le tube baisse rapidement), puis centrifugé pour 1 seconde (afin d'envoyer les gouttes des parois vers le fond) avant d'être immédiatement placé sur le support magnétique (les billes devraient sortir du liquide dans le fond). Ce liquide (~15 µL) contient les ADN qui seront amplifiés par PCR. Ce liquide est séparé en deux: 5 µL ont été archivés et 10 µL sont utilisés comme matrice PCR. Les mêmes conditions PCR de la section 2.1.3.4 ont été utilisées à l'exception des ajouts ou changements suivants: 1 µM d'amorce sens, 1 µM d'amorce antisens, 2,5 unités de Taq ADN polymérase, 10 µL d'ADN élué des billes comme matrice, volume ajusté à 100 µL avec de l'eau Milli-Q, et élongation à 72°C pendant 30 secondes (90 secondes pour la librairie N<sub>800</sub>). Les paires d'amorces sens et antisens 1, 2, 3, et 4 de l'**Annexe I** ont été utilisées pour amplifier les librairies N<sub>800</sub>, N<sub>200</sub>, N<sub>20</sub> et N<sub>50</sub> respectivement. Les amorces sens possèdent toutes une modification 5' fluorescéine (permet un dosage futur), et les amorces antisens un 5' phosphate (permet une digestion pour retour à de l'ADN simple-brin). Un échantillon du mélange PCR a ensuite été migré sur gel agarose à titre vérificatif. Le cycle SELEX est maintenant complété, mais un retour à de l'ADN simple-brin est requis afin de débiter le prochain cycle SELEX. Pour ce faire, 50 µL du mélange PCR a été incubé avec 10 unités d'enzyme lambda exonucléase à 37°C pendant une heure suivie d'une inactivation thermique à 95°C pendant 15 minutes sur thermocycleur (détails à la section 2.1.3.7). Le 52 µL contenu dans le tube est maintenant considéré comme une librairie enrichie simple-brin prête à être utilisée pour le prochain cycle SELEX.

Effectuer des PCR successives au cours des cycles SELEX amène souvent à la formation d'artefacts, tels que des bandes non-attendues ou diffuses. Une extraction de gel des bandes attendues, suivie d'une élution d'ADN, puis une précipitation par éthanol permet de limiter l'impact de ce phénomène. L'ADN a ensuite été resuspendu dans 45 µL d'eau Milli-Q et 5 µL de tampon PCR 1X (ceci est équivalent à 50 µL de mélange PCR). Ces étapes ont été faites sur les cycles 5, 7, 9, 10, 11, 12, 13, 14, 15.

### **2.3.1 Cycles négatifs et protéines négatives**

Les séquences se liant à d'autres entités que les ligands ont été éliminées par des sélections négatives. Les cycles négatifs contre des billes bloquées (sans ligands) ont été effectués suite au cycle 4, 8, 11 et 13. Suite à ces cycles, les librairies ont été amplifiées par PCR suivie d'une digestion exonucléase. Le 52 µL de librairie simple-brin résultant a ensuite été incubé avec 20 µL de billes vides et 428 µL de tampon SELEX sur rotor à 21°C. Les temps d'incubation de chaque cycle négatif sont disponibles au **Tableau 2**. Suite

à l'incubation, le surnageant (comprenant l'ADN ne s'étant pas collé aux billes vides) a directement été transféré dans un nouveau tube eppendorf afin d'effectuer le cycle SELEX suivant.

Malgré les cycles négatifs, sélectionner un aptamère envers une cible ne garantit pas une spécificité à cette cible; il est possible que cet aptamère interagisse avec une autre protéine. Une façon de limiter ceci est d'employer des ligands négatifs. Similairement aux cycles négatifs, la fraction contenant les ADN se liant aux protéines négatives sera éliminée. Lors des cycles normaux de SELEX, le surnageant est constamment éliminé, seules les billes sont conservées. Ainsi, des protéines négatives présentes dans le surnageant permettent une élimination systémique des séquences capables d'interactions avec les ligands négatifs sans incubation additionnelle. Le but des travaux présents étant plus axé sur le type de librairie que sur les ligands, le choix des ligands négatifs n'était pas d'une grande importance. Trois protéines ont été employées: la protéine BSA (*bovine serum albumin*) est couramment utilisée en biologie moléculaire et est peu dispendieuse. Le lysozyme (protéine de blanc d'œuf) est aussi très peu couteux. La galectine-1 (morphologiquement similaire à la galectine-7) a été utilisée comme protéine négative pour permettre une discrimination à l'intérieur du groupe des galectines. Pour le ligand gal-7, la gal-1 et le BSA ont été utilisés comme protéines négatives. Le BSA et le lysozyme ont été employés pour les ligands toxines et HP0953. Les protéines négatives ont été introduites à partir du cinquième cycle inclusivement, toujours à une concentration de 100 nM. Puisque le surnageant est la fraction récupérée lors des cycles négatifs, aucune protéine négative n'est employée lors des cycles négatifs.

### 2.3.2 Tableau des cycles SELEX

Tableau 1: Paramètres des étapes de SELEX à chaque cycle.

cycle	quantité de billes (µL)	concentration de BSA (nM)	concentration de gal1 ou lysozyme (nM)	temps d'incubation (mins)	nombre de lavages	temps de lavage (mins)	commentaires
C1	200	0	0	120	1	30	-
C2	20	0	0	120	1	30	-
C3	20	0	0	120	1	30	-
C4	20	0	0	60	1	30	-
<b>C4,5</b>	<b>20</b>	<b>0</b>	<b>0</b>	<b>30</b>	<b>0</b>	<b>-</b>	<b>billes bloquées</b>
C5	20	100	100	60	1	30	extraction
C6	20	100	100	60	2	5→45	extraction
C7	10	100	100	45	2	5→45	extraction
C8	10	100	100	45	2	5→45	extraction
<b>C8,5</b>	<b>20</b>	<b>0</b>	<b>0</b>	<b>60</b>	<b>0</b>	<b>-</b>	<b>billes bloquées</b>
C9	10	100	100	45	2	5→45	extraction
C10	10	100	100	30	2	5→45	extraction
C11	10	100	100	30	3	5→5→60	extraction
<b>C11,5</b>	<b>20</b>	<b>0</b>	<b>0</b>	<b>90</b>	<b>0</b>	<b>-</b>	<b>billes bloquées</b>
C12	5	100	100	30	3	5→5→60	extraction
C13	5	100	100	15	3	5→5→60	extraction
<b>C13,5</b>	<b>20</b>	<b>0</b>	<b>0</b>	<b>120</b>	<b>0</b>	<b>-</b>	<b>billes bloquées</b>
C14	5	100	100	10	3	5→5→60	extraction
C15	5	100	100	5	3	5→5→60	extraction

Le volume final contenu dans chaque tube était de 500 µL (ajusté avec du tampon SELEX). Les quatre cycles en rouge (4,5, 8,5, 11,5 et 13,5) représentent des cycles négatifs effectués avec des billes bloquées sans ligands. Une extraction des bandes, suivi d'une élution puis d'une précipitation à l'éthanol a été effectué aux cycles 5, 7, puis 9 à 15 inclusivement. Les protéines BSA, galectine-1, et lysozyme étaient les protéines négatives employées. La gal-1 était pour le ligand gal-7, alors que le lysozyme a été utilisé pour les deux autres cibles (la BSA a été utilisée pour les trois bibliothèques). Un temps de lavage indiquant 5→5→45 signifie deux lavages de cinq minutes suivies par un troisième lavage de 45 minutes.

### 2.3.3 Dosage des fractions

Il est possible de suivre le niveau d'enrichissement des bibliothèques à travers les cycles en faisant des dosages de fractions. Le surnageant de l'incubation, des lavages, et une partie de l'éluat ont été conservés à tous les cycles. Ces fractions contiennent de l'ADN fluorescent et peuvent être dosées afin de suivre dans le temps des changements de ratio d'ADN. Il est attendu d'avoir de plus en plus d'ADN dans l'éluat et moins d'ADN dans les surnageants d'incubation et de lavages plus l'on avance dans les cycles.

Plus la valeur augmente, plus l'enrichissement est important. Pour obtenir des valeurs numériques à partir des surnageants, un dosage de la fluorescence des fractions a été effectué. Un volume de 2 µL de chacune des fractions a été placé sur un verre hydrophobe (verre traité par liquide hydrofuge, Marque: Rain-X, #cat: 36231) préalablement délimité avec un crayon permanent. Il est important de travailler rapidement pour limiter l'évaporation des gouttes. Le verre a ensuite été scanné au Typhoon avec les paramètres suivants: force du laser à 500 V, résolution de 200 µm, et filtres 473LPR (473 nm) et BPB1 (520-

540 nm). Il est important de ne pas utiliser un voltage trop élevé, car une tache complètement saturée en fluorescence (point noir) ne mènera pas à un dosage significatif. Il faut plutôt opter pour des gouttelettes variant en teintes de gris.

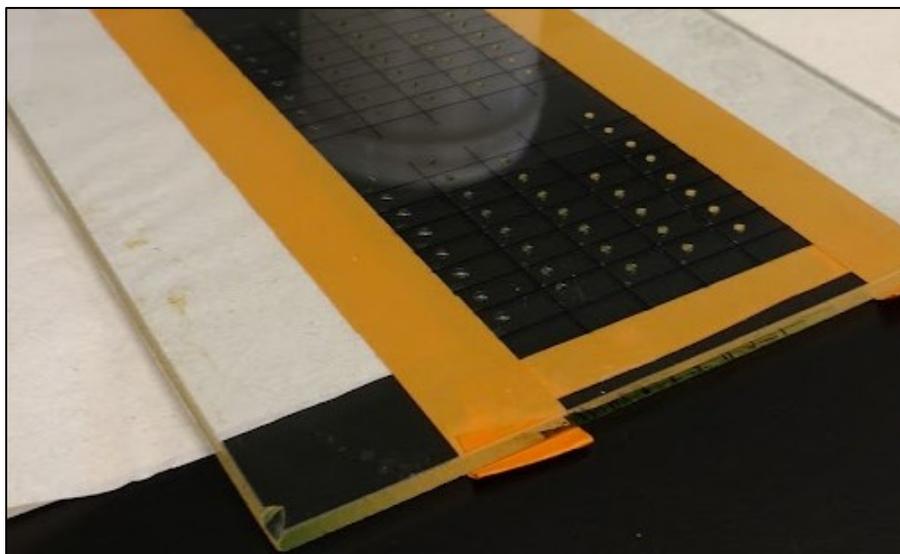


Figure 16: Photo de la plaque de verre contenant les gouttes des fractions d'un cycle SELEX.

Le fichier .tif a ensuite été ouvert avec le logiciel ImageJ. Chacune des gouttelettes a été dosée puis soustraite à la valeur du bruit de fond du verre. Les valeurs numériques résultantes (surnageant, lavages, éluat) se situent entre 0 (blanc absolu) et 65535 (noir absolu), et les valeurs entre 1 et 65534 correspondent à différentes intensités de gris. Il est important que la valeur de toutes les régions de la photo du gel n'arrive pas à ces deux extrêmes, car ceci indiquerait une saturation. Les valeurs numériques ont ensuite été placées dans la formule ci-dessous.

$$\text{Pourcentage de liaison} = \frac{\text{Éluat}}{(\text{Éluat} + \text{Surnageant} + \text{Lavages})} \times 100\%$$

#### 2.3.4 Préparation pour le séquençage

Afin de déterminer les séquences présentes dans les bibliothèques enrichies du cycle SELEX final, les bibliothèques ont été séquencées à haut débit. Les bibliothèques N<sub>20</sub> et N<sub>50</sub> ont été envoyées à la plateforme de séquençage CERMO-FC à l'Université du Québec à Montréal (UQAM). Leur séquençage de type Miseq (trousse V3, 600 cycles) n'offre malheureusement pas la possibilité de séquencer de longs fragments, tel que pour les deux bibliothèques enzymatiques. Les bibliothèques N<sub>200</sub> et N<sub>800</sub> ont ainsi été séquencées avec la

technologie Nanopore, capable de séquencer des fragments d'ADN de plusieurs kb. Le laboratoire de Frédéric Veyrier (INRS) a effectué le séquençage avec le modèle MinION Mk1C en suivant le protocole de la compagnie *Nanopore Technologies*. Concernant les librairies N<sub>20</sub> et N<sub>50</sub>, l'indexation des librairies avec les code-barres (« *barcoding* » en anglais) a été effectué avant l'envoi des échantillons à CERMO-FC. Les amorces indexées utilisées sont présentées dans le **Tableau 3** ci-dessous (les séquences sont à l'**Annexe I**). Une fois indexés, les amplicons ont été purifiés par colonne à l'aide de la trousse de la compagnie *Geneaid* (#cat: DFH300) avant d'être envoyés pour séquençage.

**Tableau 2: Paires d'amorces employées lors de l'indexation des courtes librairies N<sub>20</sub> et N<sub>50</sub>.**

ligands	librairies	amorce sens (index i5)	amorce antisens (index i7)
galectine-7	N <sub>20</sub>	S518	N703
	N <sub>50</sub> (1 pmol)	S518	N702
	N <sub>50</sub> (10 nmol)	S518	N701
NheA et HblB	N <sub>20</sub>	S518	N707
	N <sub>50</sub> (1 pmol)	S518	N706
	N <sub>50</sub> (10 nmol)	S518	N705
HP0953	N <sub>20</sub>	S518	N714
	N <sub>50</sub> (1 pmol)	S518	N712
	N <sub>50</sub> (10 nmol)	S518	N711

### 2.3.5 Analyse des séquences

Les courtes librairies séquencées par Miseq (N<sub>20</sub> et N<sub>50</sub>) ont été analysées à l'aide du logiciel AptaSUITE v0.0.1 (Hoinka *et al.*, 2018). Les informations recueillies étaient les séquences, le nombre de lectures associées à chaque séquence, le nombre de lectures total et le pourcentage des séquences à lecture unique. Pour chacune des librairies, les trois séquences (à longueur attendue) ayant le plus de lectures ont été choisies comme des candidats potentiels d'aptamères.

Concernant les longues librairies séquencées par Nanopore (N<sub>200</sub> et N<sub>800</sub>), un contrôle qualité a été effectué pour vérifier la longueur de chaque lecture. Les séquences ayant une longueur de nucléotide sous le minimum attendu ont été éliminées du fichier de sortie (240 nucléotides pour N<sub>200</sub> et 500 pour N<sub>800</sub>). Les séquences restantes ont ensuite été comparées avec celles de leurs amorces respectives. Ceci a été fait à l'aide de la fonction « rechercher et remplacer » du logiciel Word. Trois tranches de huit nucléotides de chaque amorce (début, milieu, fin de l'amorce) ont été recherchées à travers le texte, puis surlignées afin de faire ressortir les séquences correspondantes aux amorces. Les lectures correspondant entièrement à des amorces ont été éliminées du fichier. Des régions correspondantes à des amorces ne

sont pas souhaitables dans le contexte du séquençage d'une librairie. En les visualisant sur le logiciel Word côte à côte, un seul candidat a été choisi pour chacune des librairies N<sub>200</sub> et N<sub>800</sub>. Cette sélection a été faite en prenant la séquence ayant la plus petite proportion de séquence correspondant à ses amorces.

## 2.4 Tests de liaison

Les candidats obtenus après l'analyse ont été testés *in vitro* afin de vérifier s'ils sont réellement capables de se lier à leur cible respective. Quatre différents tests de liaisons ont été effectués: le retard sur gel, la microscopie inverse, l'essai à la goutte, et la thermophorèse à micro-échelle.

### 2.4.1 Retard sur gel

Le principe d'un retard sur gel (*gel shift assay* en anglais) est d'observer une différence de migration sur gel entre un oligonucléotide seul et un oligonucléotide attaché à une cible. Ceci a été fait en incubant les candidats (ADN fluorescent) et leur cible dans du tampon de SELEX pendant 15 minutes à température pièce, puis en les migrant dans un gel natif (PAGE TBE 6%, migration à 150 V pendant 6 heures à 4°C). Seuls trois différents ratios ADN:protéine ont été testés: 1:0, 1:5 et 1:100. L'ADN était à une concentration fixe de 1 µM. Le gel a ensuite été migré puis scanné au Typhoon (1 pmol d'ADN a été migré dans chaque puits). Il est ensuite possible de refaire la migration utilisant des concentrations de protéine moins élevées et plus homogènes pour les candidats obtenant une migration différente afin d'observer un meilleur dégradé, mais aucun retard sur gel reproductible n'a été observé.

### 2.4.2 Microscope inversé

Un microscope à statif inversé a été employé afin d'observer une différence de fluorescence (Cy5) émise par les candidats (ADN) recouvrant des billes magnétiques NHS (200 nm de diamètre) fonctionnalisées par les protéines cibles. Un même candidat a été incubé avec des billes bloquées (sans protéine), des billes contenant une protéine négative (la BSA) et des billes contenant la protéine cible (Gal7 ou HP0953). Les résultats attendus étaient d'observer et mesurer une fluorescence plus intense sur les billes cibles comparativement aux billes négatives. Les mélanges suivants ont été faits: 2 µL de billes, 1 µL d'ADN à ~1 µM (fluorescence Cy5), 7 µL de tampon SELEX. Les mélanges ont ensuite été incubés pendant 30 minutes sur rotor à température pièce (21°C). Un volume de 500 µL de tampon SELEX a ensuite été ajouté au mélange, puis vortexé pendant 30 secondes. Les tubes ont été placés sur un support magnétique afin d'éliminer le surnageant. Un volume de 5 µL de tampon de SELEX a été ajouté afin de resuspendre les billes, puis le tube a été vortexé brièvement. Après avoir secoué le tube vers le bas une fois (pour amener

le liquide vers le fond du tube), 1  $\mu\text{L}$  du fond du tube (contenant les billes et un reste de tampon SELEX) a été placé entre lame et lamelle pour observation au microscope. Pour observer une plus grande concentration de billes, le tube a été centrifugé par minifuge pendant 5 secondes avant de prélever le 1  $\mu\text{L}$ . Information sur le microscope: modèle: Axio Observer.Z1 / 7, filtre utilisé: Cy5, objectif utilisé: alpha Plan-Apochromat 100x/1.46 Oil DIC (UV) M27, intensité du laser: 10%, logiciel: ZEN version 2,6). Les mêmes paramètres de contraste ont été utilisés pour chaque type de bille pour un même candidat.

### 2.4.3 Essai à la goutte

L'essai à la goutte (*spotting test* en anglais) est très similaire à la section « 2.4.3 Dosage des fractions ». Des gouttelettes contenant de l'ADN fluorescent ont été déposées sur une plaque de verre, puis ont été scannées au Typhoon. L'idée de base était de mettre une quantité initiale connue d'ADN fluorescent dans un tube, ajouter ensuite des billes magnétiques contenant les protéines cibles en leur surface, incuber l'ADN et les billes, enlever les billes, puis de remesurer l'ADN restant dans le tube. Le rôle des billes est de pêcher l'ADN capable d'interactions. Ainsi, une diminution de fluorescence (ou déplétion) indique une liaison entre le candidat et sa cible. Le surnageant est plus facile à mesurer que les billes elles-mêmes, car il est difficile de prélever une quantité égale de billes à travers différents pipetages. Les billes ont tendance à sédimenter très rapidement, alors un grand écart de signal serait visible d'un échantillon à l'autre. Dans un volume de 5  $\mu\text{L}$ , 0,5  $\mu\text{L}$  d'ADN fluorescent (Cy5 en 5') à 100 nM (10 nM final) a été incubé avec 4,5  $\mu\text{L}$  de billes fonctionnalisées (concentration plus haut). Quatre différentes fonctionnalisations de billes ont été testées pour chaque candidat: pas de bille, billes bloquées, billes avec BSA (protéine négative), et des billes avec la protéine cible (Gal7 ou HP0953). Il était attendu d'observer une déplétion de fluorescence plus importante avec les protéines cibles comparativement aux trois autres. Après 15 minutes d'incubation à température pièce, les tubes ont été placés sur un support magnétique afin de séparer les billes de la partie liquide, puis 2  $\mu\text{L}$  de la partie liquide ont été placés sur une plaque de verre quadrillée. La plaque a été scannée au Typhoon (force du laser à 500 V, résolution de 100  $\mu\text{m}$ , excitation à 635 nm et filtre d'émission LPR >665 nm). Les gouttes ont ensuite été dosées par ImageJ afin de déterminer le pourcentage de liaison.

#### 2.4.4 Thermophorèse à micro-échelle

La thermophorèse à micro-échelle (*microscale thermophoresis*, ou MST) est une technologie permettant de mesurer l'interaction entre des molécules (Jerabek-Willemsen *et al.*, 2014). Il est possible d'observer une différence d'intensité de fluorescence, souvent une diminution due à un effet de *quenching* lorsque l'ADN s'attache à son ligand (Jerabek-Willemsen *et al.*, 2014). De plus, un complexe ADN:protéine est plus encombrant que de l'ADN ou une protéine libre (Jerabek-Willemsen *et al.*, 2014). Ainsi, lorsque soumis à une augmentation de température, les complexes migreront vers les endroits plus frais, mais en se déplaçant plus lentement que les ADN et protéines libres (Jerabek-Willemsen *et al.*, 2014). Cette différence de temps de migration est mesurée par l'appareil en détectant l'intensité de fluorescence à l'endroit chauffé (Jerabek-Willemsen *et al.*, 2014). Une concentration fixe des candidats d'ADN à 2,5 nM (possédant tous un fluorophore Cy5 en 5') a été incubée avec une concentration de protéine cible variant de 1,5 nM à 50 µM dans 8 microtubes différents. L'incubation a été faite dans du tampon SELEX pendant 15 minutes à température pièce (21°C). Du Tween-20 (concentration finale à 0.1%) a ensuite été ajouté aux tubes avant de charger les échantillons à l'intérieur de capillaires (compagnie: Nanotemper, capillaires pour Monolith™ NT.115, #cat: MO-K022) fournis par le fabricant de la machine. Les capillaires ont été lus par l'appareil Monolith NT.115Pico en utilisant le logiciel MO.Control v1.6.1. La température « medium » a été choisie, avec l'intensité du laser variant entre 5-50%, afin d'obtenir un compte de fluorescence d'au moins 3000 pour chaque capillaire de chaque candidat.

## 3 RÉSULTATS

---

### 3.1 Synthèse des librairies

Afin de comparer la différence entre de longues et courtes librairies, il était nécessaire en premier lieu de se les procurer. Les courtes librairies peuvent être commandées, mais les longues doivent être synthétisées sur place. Ci-dessous sont présentés quelques gels démontrant des preuves de concept ou de l'optimisation de conditions menant aux librairies enzymatiques  $N_{200}$  et  $N_{800}$ .

#### 3.1.1 Librairie $N_{800}$ par terminale transférase

Puisque la synthèse des librairies produira des ADN simple-brin, il était nécessaire d'avoir recours à un gel dénaturant afin d'empêcher la formation de structures. Les poids moléculaires conventionnels sont conçus pour des gels natifs, il était donc nécessaire de produire nous-mêmes nos propres marqueurs de poids moléculaires marqués par fluorescence (le même marquage que les librairies). Le gel ci-bas (**Figure 17**) présente des longueurs précises d'amplicons fluorescents pouvant être utilisés comme marqueurs de taille lorsque migré à côté d'échantillons à longueur inconnue. Les amplicons obtenus ont une taille théorique de 528, 815, 1110, et 1502 pb. Ces ADN peuvent maintenant être utilisés comme marqueur de poids moléculaire dans des gels dénaturants PAGE puisque l'un des deux brins est marqué à la fluorescéine. Un colorant intercalant d'ADN n'est donc plus nécessaire. Lors de leur utilisation, les marqueurs ont été chargés directement dans les gels PAGE après avoir été chauffés à 95°C dans du formamide (aucune digestion enzymatique nécessaire).

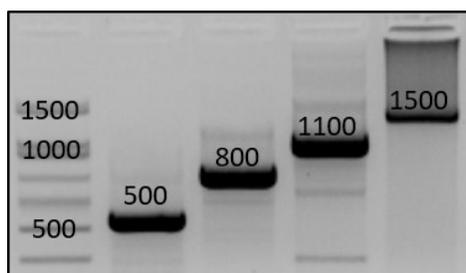


Figure 17: Migration agarose d'amplicons fluorescents indexés issus du gène ribosomal 16S bactérien.

Une PCR a été faite sur le gène 16S de *E. coli* DH5 $\alpha$ . L'amorce sens 8F avec les amorces antisens 518R, 805R, 1092R, ou 1492 ont été utilisées (quatre amplicons de longueurs variables). Une deuxième PCR a ensuite été effectuée sur ces amplicons en utilisant des amorces complémentaires aux amorces 16S, mais avec une séquence additionnelle protubérante aux extrémités 5'. Une troisième PCR a ensuite été faite en utilisant des amorces fluorescentes correspondant à ces protubérances. Les résultats présentés ci-haut correspondent à cette troisième PCR migré dans un gel TAE agarose 2% à 120 V pendant 45 minutes.

### 3.1.1.1 Synthèse de la région aléatoire

La prochaine étape était de tester l'enzyme terminale transférase afin de voir sa portée de synthèse en termes de longueur (nombre de nucléotides ajoutés). Le temps d'incubation, le ratio ADN:enzyme, et la quantité de dNTP ont été testés. Dans la **Figure 18** se retrouve plusieurs profils de migration typique résultant d'une incubation avec l'enzyme terminale transférase. Un signal est observable partout dans le puits (du haut du puits jusqu'à beaucoup plus bas dans le gel), bien que l'intensité varie. Une augmentation de la quantité d'ADN pour une même quantité d'enzyme mène à un allongement moins important des oligonucléotides (« *smear* » ou tache plus étalée). Une synthèse plus importante est aussi observable lorsque l'on augmente le temps d'incubation. La quantité de dNTP aide l'allongement dans les temps d'incubation plus bas, mais aucune différence ne semble ressortir pour les temps d'incubation plus élevés. Par contre, une augmentation du temps d'incubation semble contrer ces deux phénomènes puisque le signal est plus compact et plus élevé.

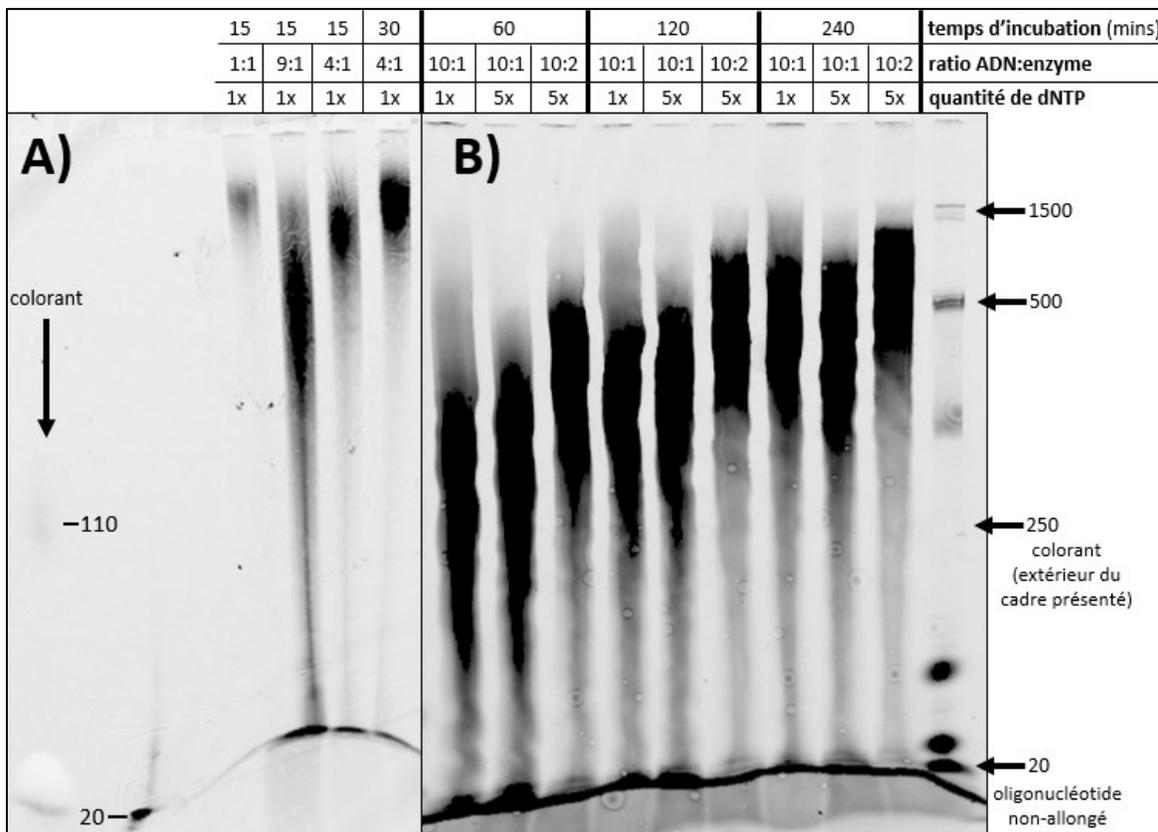


Figure 18: Longueur des ADN en fonction du temps d'incubation, du ratio ADN:enzyme, et de la quantité de dNTP.

Un oligonucléotide d'ADN (20 nt) marqué en 5' par la fluorescéine a été incubé avec l'enzyme TdT. De différentes quantités d'enzyme, temps d'incubation et quantité de dNTP ont été testés. Les mélanges ont ensuite été migrés dans un gel dénaturant urée-PAGE 6% (A) ou 3% (B) (500 V pendant 1 heure) avant d'être visualisés au Typhoon (force du laser à 500 V (A) ou 1000 V (B), résolution de 200 µm, excitation à 473 nm et filtre d'émission BPB1 à 520-540 nm).

Puisque le contexte de synthèse vise à créer une librairie, une grande quantité d'ADN est requise. Il faut donc envisager de mesurer la quantité maximale d'ADN que ce genre de réaction peut tolérer sans sacrifier la longueur de la région aléatoire. Une forte intensité de diffusion a été observée dans la **Figure 19A** lorsqu'il y a un surchargement d'ADN. Une augmentation de la quantité d'ADN pour une même quantité d'enzyme diminue la longueur moyenne des fragments (**Figure 19A**). Il semble aussi y avoir une augmentation de longueur en augmentant le temps d'incubation, mais celle-ci plafonne rendue à 4 heures (**Figure 19B**). Le signal semble plus se concentrer lorsque l'on passe de 30 minutes à 60 minutes d'incubation pour un même ratio moléculaire ADN:enzyme. Suivant ces résultats, un ratio ADN:enzyme de 1:1 avec un temps d'incubation de deux heures a été employé pour synthétiser la librairie N<sub>800</sub> de départ.

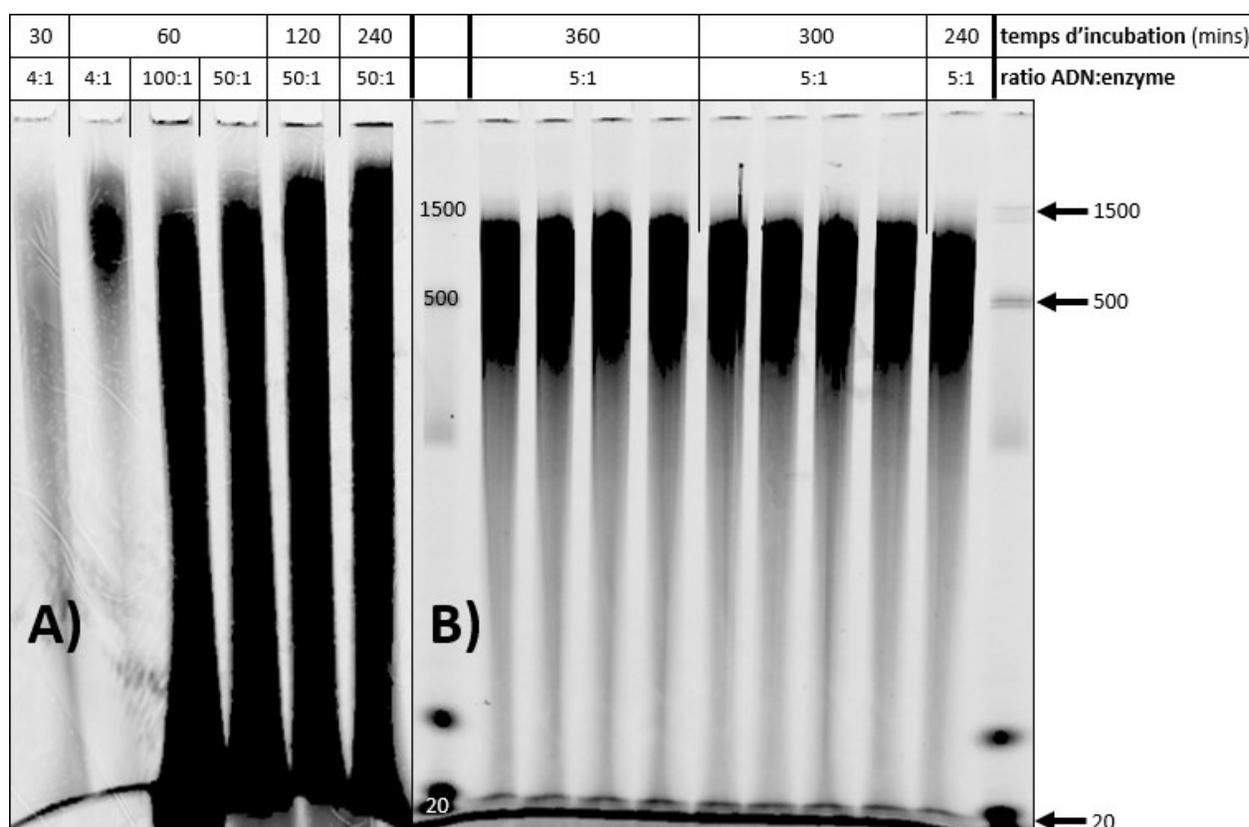
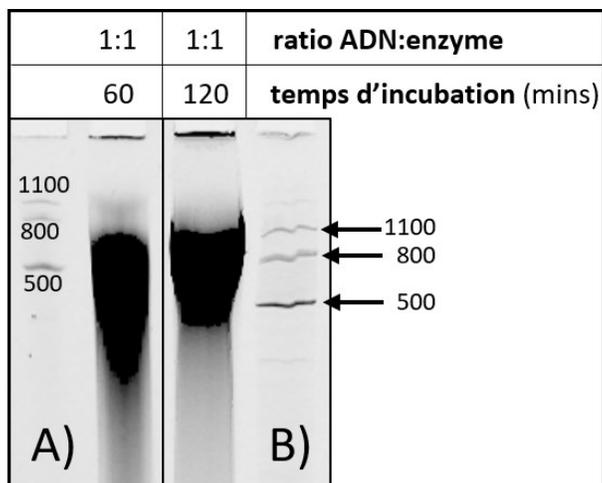


Figure 19: Longueur des ADN en fonction du temps d'incubation et du ratio ADN:enzyme.

Un oligonucléotide d'ADN (20 nt) marqué en 5' par la fluorescéine a été incubé avec 2 pmol d'enzyme TdT. Différentes quantités d'ADN initial allant de 8 pmol à 200 pmol et temps d'incubation ont été testés. Les mélanges ont ensuite été migrés dans des gels dénaturants urée-PAGE 6% (A) ou 3,5% (B) (500 V pendant 1 heure) avant d'être visualisés au Typhoon (force du laser à 1000 V, résolution de 200  $\mu$ m, excitation à 473 nm et filtre d'émission BPB1 à 520-540 nm).

La quantité d'ADN de la **Figure 20** a été limitée afin de ne pas causer une surcharge dans les puits et afin de rendre le signal le moins étalé possible. En effet, un signal plus compact (moins diffus) a été obtenu. Comparativement à l'incubation d'une heure, une incubation de deux heures permet d'obtenir une forme plus compacte et légèrement plus élevée. La longueur finale varie de 500 à 1100 nucléotides approximativement. Ces paramètres ont été utilisés afin de synthétiser la région aléatoire de la librairie N<sub>800</sub>.



**Figure 20:** Profil de migration de la librairie optimisée finale TdT (région aléatoire, sans adaptateur).

Un oligonucléotide d'ADN (20 nt) marqué en 5' par la fluorescéine a été incubé avec l'enzyme TdT. A) 60 minutes d'incubation, B) 120 minutes d'incubation (condition choisie pour la librairie). Les mélanges ont ensuite été migrés dans un gel dénaturant urée-PAGE 6% (500 V pendant 80 minutes) avant d'être visualisés au Typhoon (force du laser à 1000 V, résolution de 200 µm, excitation à 473 nm et filtre d'émission BPB1 à 520-540 nm). Six puits ont été chargés afin d'augmenter la quantité d'ADN récupérée, mais seulement un est présenté.

### 3.1.1.2 Ajout de l'adaptateur et vérification

Ayant des fragments d'ADN simple-brin avec une première région connue (oligonucléotide de 20 nt) suivie de leur longue région aléatoire (N<sub>500</sub> à N<sub>1100</sub>), il était maintenant question de rendre ces fragments amplifiables par PCR. Ceci a été fait en ajoutant une deuxième région connue à l'extrémité 3'. La première étape était de tester les prérequis de l'enzyme T4 ARN ligase 1 et de faire la preuve de concept. Une ligation entre deux oligonucléotides d'ADN ne semble pas fonctionner en absence de courtes séquences d'ARN en 3' de l'accepteur (hauteur identique à l'accepteur seul, **Figure 21**). Par contre, la présence d'ARN en 3' permet une bonne ligation, avec la majorité des oligonucléotides de départ se faisant lier aux donneurs. La présence d'une courte séquence d'ARN en 3' est donc nécessaire afin de lier des fragments ensemble. L'ajout d'une riboqueue à des fins de ligation était inspiré d'un article ayant déjà publié cette optimisation de ligation (Miura *et al.*, 2019).

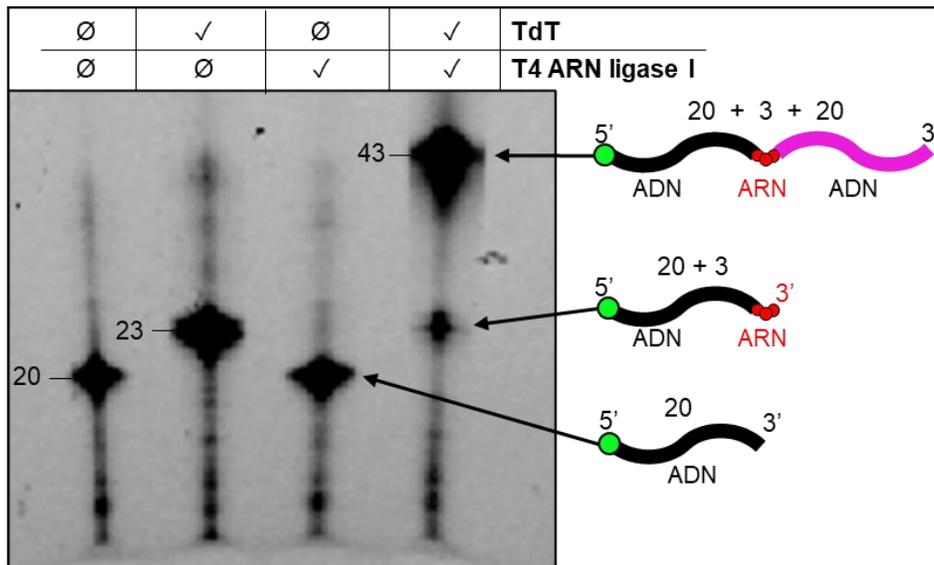


Figure 21: Preuve de concept démontrant qu'une courte queue d'ARN en 3' de l'ADN accepteur est nécessaire pour lier des ADN simple-brin.

Un oligonucléotide d'ADN (20 nt) marqué en 5' par la fluorescéine a été incubé avec l'enzyme TdT. Deux types d'oligonucléotides ont été testés: avec ou sans *ribotail*. La queue d'ARN (*ribotail*) a été ajoutée par incubation avec la TdT et du rATP. Les oligonucléotides de départ ont ensuite été incubés avec un adaptateur d'ADN (phosphate en 5') et la T4 ARN ligase 1 (ratio accepteur:donneur de 1:10). Les mélanges ont ensuite été migrés dans un gel dénaturant urée-PAGE 10% (500 V pendant 30 minutes) avant d'être visualisés au Typhoon (force du laser à 1000 V, résolution de 200 µm, excitation à 473 nm et filtre d'émission BPB1 à 520-540 nm).

Cette vérification maintenant faite, les librairies ont reçu une courte queue d'ARN en 3' à l'aide de la terminale transférase et de rATP. Ces ADN ont ensuite été ligés avec un adaptateur (ADN à séquence connue). Les librairies résultantes ont finalement été amplifiées par PCR à titre de vérification et aussi pour obtenir une quantité d'ADN suffisante pour débiter les SELEX (**Figure 22**). Un important artefact de PCR est visible à environ 40 pb, et celui-ci a persisté dans le temps à travers toutes les expériences subséquentes. La PCR 1 a obtenu un signal diffus à travers le puits qui semble graduellement pâlir jusqu'à ~1000 pb. Ceci est probablement un assemblage chaotique d'ADN simple-brin ou des artefacts de PCR. La PCR 2 a ce même profil de migration, mais il est plus pâle et contient aussi une région plus foncée aux alentours de ~500-1100 pb. Cette région correspond exactement à la longueur attendue de la librairie TdT de la **Figure 20B**. De plus, la **Figure 22B** montre qu'une digestion exonucléase retourne à une migration identique à celle de la PCR 1, indiquant que la région plus foncée correspond fort probablement à de l'ADN parfaitement double-brin. La librairie est donc amplifiable par PCR et sera utilisée pour les étapes de SELEX par la suite (section 3.2). En faisant une règle de trois, en éliminant le signal de bruit de fond, puis en multipliant la valeur par le volume restant du mélange PCR final non-migré, la quantité utilisable d'ADN de la librairie TdT produite était de ~5,1 pmol (dans 270  $\mu$ L au total). Une autre observation est qu'une matrice sans adaptateur 3' mène à un signal beaucoup plus intense que celui de la PCR 1 avec adaptateur. Par contre, lorsqu'un cycle final est fait en absence d'adaptateur 3', le signal reste identique, indiquant qu'il est question d'amplification non-spécifique lorsqu'il y a absence d'une des deux assises.

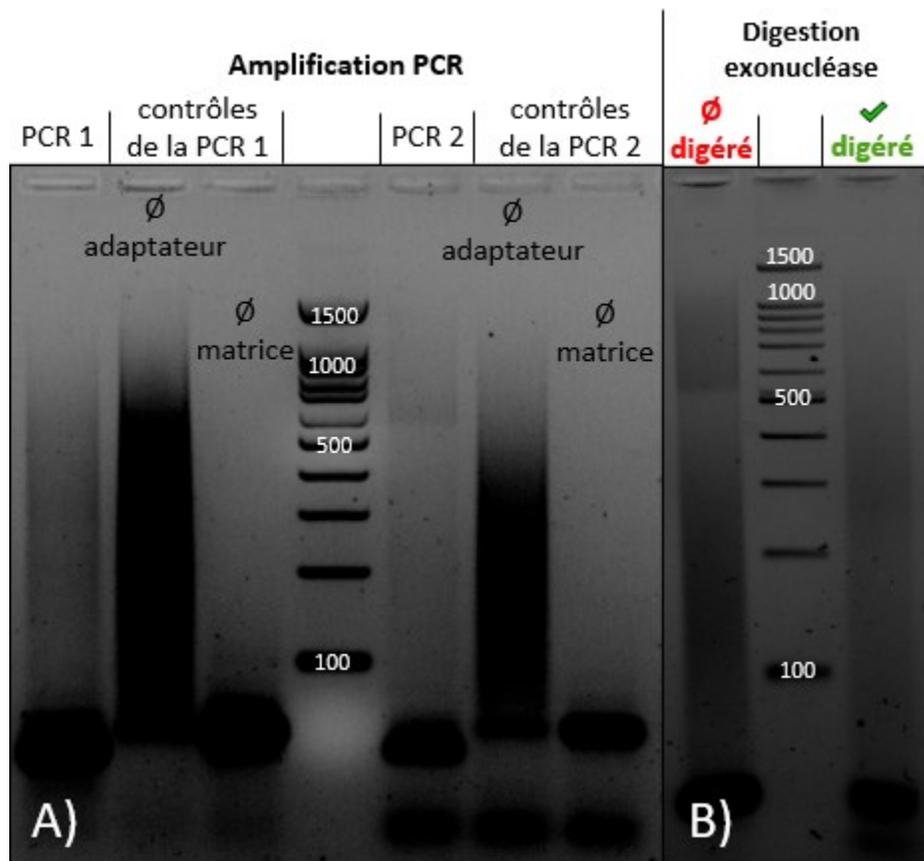


Figure 22: PCR finale pour le retour à de l'ADN double-brin et digestion avec une exonucléase.

La librairie de la Figure 20B a été extraite du gel puis purifiée. L'ADN a ensuite reçu une courte queue d'ARN (par TdT et rATP). Ce mélange a été amplifié par PCR (PCR 1), suivie d'un cycle final (PCR 2). La PCR 2 est issue d'une dilution de la PCR 1 dans un nouveau mélange suivi d'un seul cycle PCR (avec nouveaux réactifs) (la section 2.1.3.5 fournit plus de détails). En B) est présentée la digestion du brin complémentaire par une exonucléase à même le mélange PCR. Les échantillons ont été migrés dans un gel TAE agarose 2% à 120 V pendant 45 minutes. L'expérience entière a été faite deux fois (n=2, mêmes résultats).

### 3.1.2 Librairie N<sub>200</sub> par ligation

#### 3.1.2.1 Ligation des oligonucléotides dégénérés

Une deuxième approche de synthèse de région aléatoire a été explorée: par ligation de régions dégénérées N<sub>50</sub> à l'aide de la T4 ARN ligase 1. Similairement à la TdT, il était question d'allonger un oligonucléotide de départ de façon enzymatique, mais cette fois-ci en ajoutant de façon successive des fragments simple-brin N<sub>50</sub> en 3'. Les premières expériences tentaient de tester différentes quantités de N<sub>50</sub> pour une même quantité d'oligonucléotides de départ. La preuve de concept a été réussie, il est possible d'ajouter des fragments simple-brin *en tandem* (**Figure 23**). La ligation en chaîne de N<sub>50</sub> est démontrée par la présence de plusieurs bandes montantes à travers un même puits par bonds de 50 nt. Le ratio accepteur:donneur de 1:10 détient le meilleur rendement de ligation.

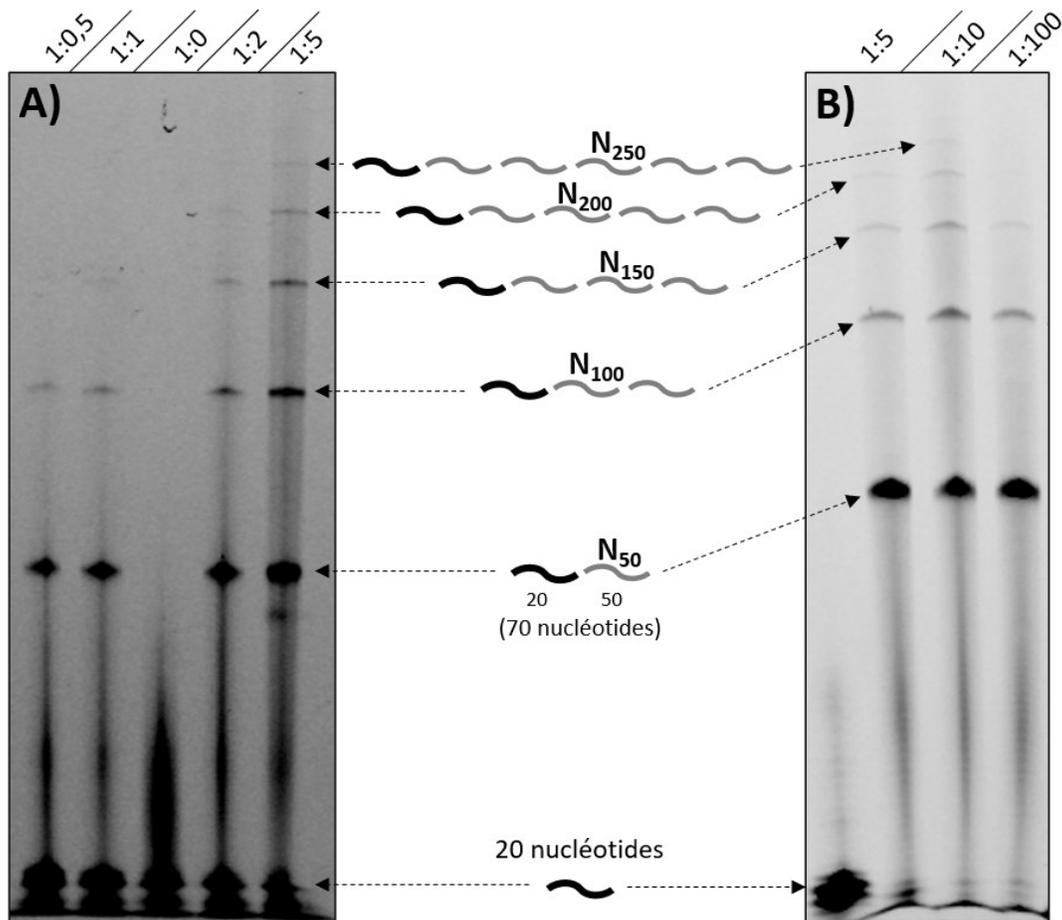
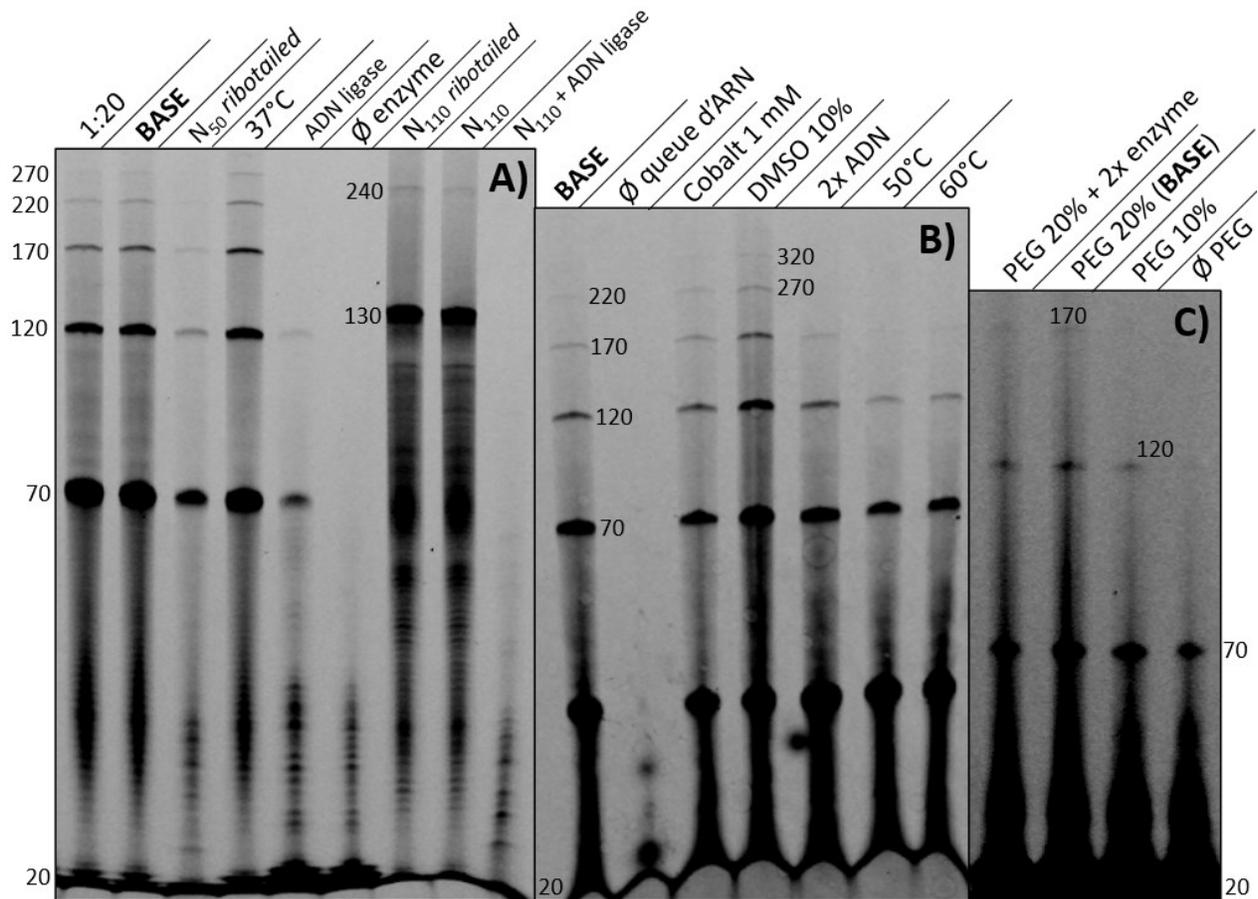


Figure 23: Nombre de fragments d'ADN simple-brins ligués en fonction du ratio d'ADN accepteur:donneur.

Des N<sub>50</sub> (phosphate en 5' et 1rA en 3') ont été ligués (par T4 ARN ligase 1) à des oligonucléotides d'ADN (20 nt) marqués en 5' par la fluorescéine. Le mélange a été migré dans un gel dénaturant urée-PAGE 6% (A) ou 10% (B) à 500 V pendant 1 heure (A) ou 2 heures (B). Le gel a été scanné au Typhoon (force du laser à 1000 V, résolution de 200  $\mu$ m, excitation à 473 nm et filtre d'émission BPB1 à 520-540 nm). Une quantité de 0,8 (A) ou de 1,6 (B) pmol d'ADN fluorescent est présente dans chaque puits.

Il était maintenant question de tester plusieurs conditions de réactions afin de rendre le procédé plus efficace. Ceci permettrait d'atteindre plus d'ajouts successifs de  $N_{50}$  en tandem, ce qui résulterait en des régions aléatoires plus longues et diversifiées. Une incubation à 37°C semble idéale comparativement à 21 (puits « BASE »), 50 et 60°C (**Figure 24**). La présence du cobalt et de DMSO semble aussi légèrement augmenter le rendement de la ligation (**Figure 24B**). La présence de 20% PEG (**Figure 24C**) et d'une queue d'ARN semble aussi primordiale à la réaction. Une plus grande quantité d'ADN dans le mélange ne semble pas aider. Tout ceci étant dit, il ne semble pas possible d'aller au-delà de 300 nt aléatoires, et le signal diminue fortement à chaque bond successif de  $N_{50}$ .



**Figure 24:** Diverses conditions testées ayant pour but d'améliorer le rendement de ligation d'ADN simple-brin.

Des  $N_{50}$  (phosphate en 5' et un rA en 3') ont été ligués (par T4 ARN ligase 1) à un oligonucléotide d'ADN (20 nt) marqués en 5' par la fluorescéine. Plusieurs différentes conditions ont été testées par rapport au mélange de base (désigné « BASE »). Les mélanges ont été migrés dans des gels dénaturants urée-PAGE 10% (A), 6% (B), ou 3,5% (C) à 500 V pendant 1 heure (B et C) ou 2,5 heures (A). Le gel a été scanné au Typhoon (force du laser à 1000 V, résolution de 200  $\mu\text{m}$ , excitation à 473 nm et filtre d'émission BPB1 à 520-540 nm). La désignation « *ribotailed* » signifie un ajout d'une queue d'ARN de façon enzymatique plutôt que chimiquement par un fournisseur d'oligonucléotides. Une quantité de 1 pmol d'ADN fluorescent est présente dans chaque puits.

Des essais additionnels d'optimisation de conditions ont été effectués. Cette fois-ci, divers agents dénaturants et plus de ratios accepteur:donneur ont été testés (**Figure 25**). Dans leur concentration testée, la bétaine et l'urée semblent nuire à la réaction de ligation. La présence de DMSO de la **Figure 25B** obtient des bandes  $N_{50}$  et  $N_{100}$  plus intenses que la condition sans DMSO. Une quantité d'ADN (même ratio accepteur:donneur) 2,5 fois plus élevée obtient un rendement plus efficace.

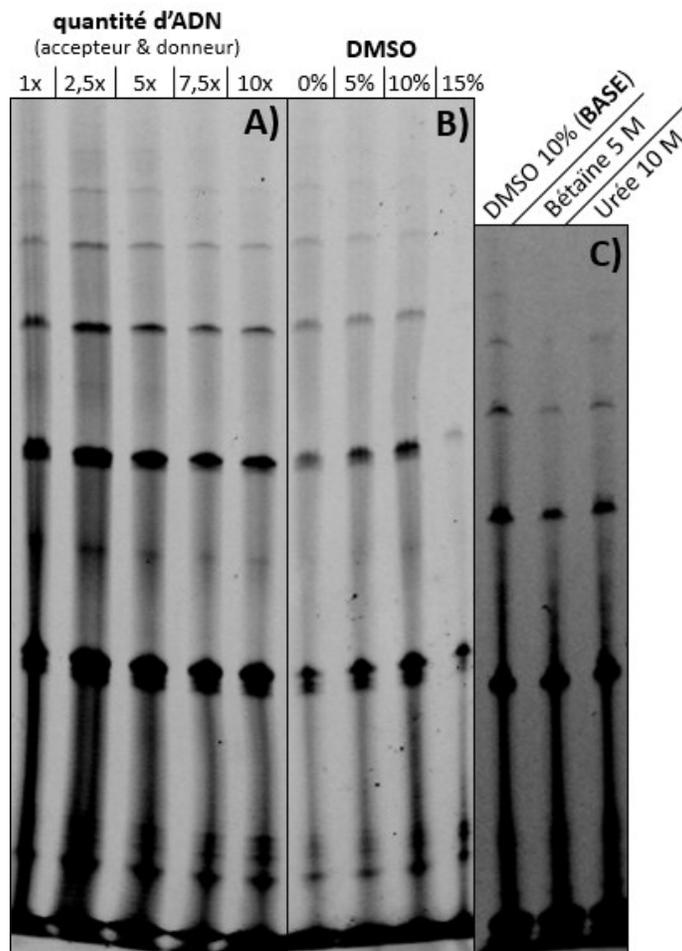


Figure 25: Effet de la quantité de  $N_{50}$  et d'agents dénaturants sur le nombre de fragments d'ADN simple-brin ligés.

Des  $N_{50}$  (phosphate en 5' et un rA en 3') ont été ligués (par T4 ARN ligase 1) à des oligonucléotides d'ADN (20 nt) marqués en 5' par la fluorescéine. Plusieurs différentes conditions ont été testées par rapport au mélange de base (désigné « BASE »). Les échantillons en A) ont été dilués de façon pondérée juste avant le chargement dans le gel afin d'avoir une quantité d'ADN identique dans chacun des cinq puits (permet une bonne comparaison). Les mélanges ont été migrés dans des gels dénaturants urée-PAGE 6% à 500 V pendant 75 minutes. Le gel a été scanné au Typhoon (force du laser à 1000 V, résolution de 200  $\mu$ m, excitation à 473 nm et filtre d'émission BPB1 à 520-540 nm). Une quantité de 2,5 pmol d'ADN fluorescent est présente dans chaque puits.

Une autre condition pertinente à tester était le nombre de ribonucléotides en 3' des fragments d'ADN. Trois riboadénines semblent la meilleure quantité à utiliser, à la fois dans le cas de l'accepteur et du donneur (**Figure 26**). Ces oligonucléotides (amorce initiale et N<sub>50</sub>) ont donc été choisis pour les expériences ultérieures, cette fois-ci comportant trois rA en 3' plutôt qu'un seul.

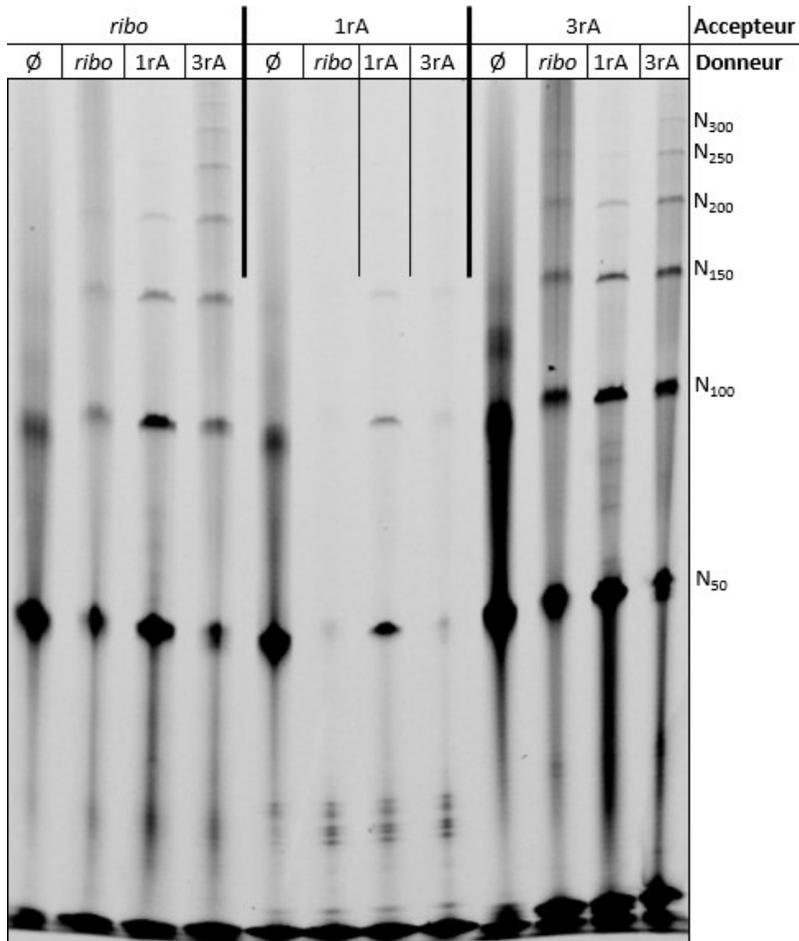


Figure 26: Effet du nombre de riboadénine (rA) en 3' du donneur et de l'accepteur sur le rendement de ligation d'ADN simple-brin.

Des N<sub>50</sub> (phosphate en 5' et un ou trois rA en 3', tel qu'écrit au-dessus du gel) ont été ligués (par T4 ARN ligase 1) à des oligonucléotides d'ADN (20 nt) marqués en 5' par la fluorescéine. Un nombre variable de riboadénine en 3' a été testé à la fois pour l'accepteur et le donneur. Les mélanges ont été migrés dans des gels dénaturants urée-PAGE 6% à 500 V pendant 75 minutes. Le gel a été scanné au Typhoon (force du laser à 1000 V, résolution de 200 µm, excitation à 473 nm et filtre d'émission BPB1 à 520-540 nm). La désignation « *ribo* » signifie un ajout d'une queue d'ARN de façon enzymatique (par la TdT) plutôt que chimiquement par un fournisseur d'oligonucléotides. Une quantité de 2,5 pmol d'ADN fluorescent est présente dans chaque puits.

L'étape suivant l'ajout de la région aléatoire serait de rendre le fragment amplifiable en ajoutant une deuxième assise d'amorce à séquence connue (adaptateur). Dans le but de diminuer les pertes d'ADN, il a été tenté de combiner l'ajout des fragments  $N_{50}$  simultanément à celui de l'adaptateur en une seule réaction (**Figure 27**). Malheureusement, la combinaison des deux réactions ne semble pas efficace. La présence d'adaptateur diminue fortement le nombre de  $N_{50}$  entre les deux assises. Un important étalement des bandes semble aussi apparaître plus l'on augmente la quantité d'adaptateurs 3'.

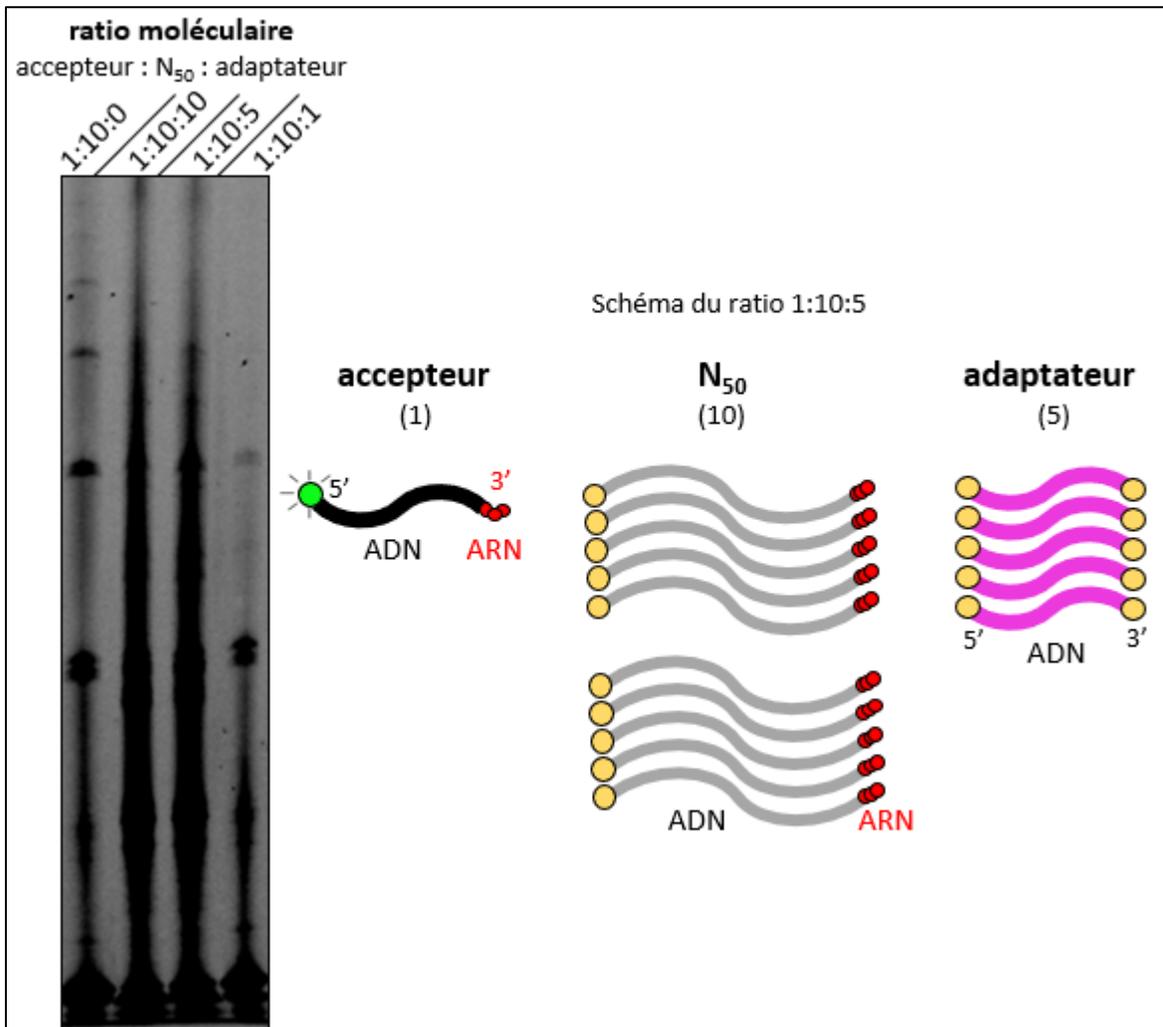


Figure 27: Essai de ligation d'ADN simple-brin de région aléatoire et d'adaptateur en une seule réaction.

Des  $N_{50}$  (phosphate en 5' et trois rA en 3') et des adaptateurs (20 nt, phosphate en 5' et 3') ont été ligués (par T4 ARN ligase 1) à des oligonucléotides d'ADN (20 nt) marqués en 5' par la fluorescéine. Des quantités variables d'adaptateurs ont été testées. Les mélanges ont été migrés dans des gels dénaturants urée-PAGE 6% à 500 V pendant 75 minutes. Le gel a été scanné au Typhoon (force du laser à 1000 V, résolution de 200  $\mu$ m, excitation à 473 nm et filtre d'émission BPB1 à 520-540 nm). Une quantité de 2,5 pmol d'ADN fluorescent est présente dans chaque puits.

### 3.1.2.2 Ajout de l'adaptateur et vérification

Comme l'ajout de l'adaptateur (assise à séquence connue en 3') dans la réaction de multimérisation de la **Figure 27** était inefficace, la préparation de la librairie s'est faite en deux étapes. Un grand volume de réaction de ligation a été fait dans les conditions optimisées, puis migré sur gel (non présenté). La bande correspondant à quatre ajouts  $N_{50}$  ( $N_{200}$ ) a été extraite du gel, puis purifiée par précipitation à l'éthanol avant l'ajout de l'adaptateur par une deuxième réaction de ligation. Le mélange de réaction a ensuite été amplifié par RT-PCR. Une seule bande de 240 pb était attendue ( $20+200+20$ ), mais un profil d'escalier est visible (**Figure 28**). Chaque bond de l'escalier correspond à une augmentation de 50 pb, ce qui semble indiquer qu'il s'agit d'un représentant de chaque  $N_{50}$  ajouté ( $N_0$ ,  $N_{50}$ ,  $N_{100}$ ,  $N_{150}$  et  $N_{200}$ ) flanqué par deux assises d'amorces de 20 nt. L'hypothèse émise est qu'il y aurait de la dégradation à chaque jonction d'ARN entre les  $N_{50}$  avant la ligation avec l'adaptateur. Ainsi, lors de cette deuxième ligation, un adaptateur serait ajouté à plusieurs types de molécule plutôt qu'exclusivement la molécule  $N_{200}$ . L'expérience a été répétée en travaillant proprement et à froid afin de limiter la dégradation d'ARN, mais le même résultat est survenu. Il est aussi important de mentionner qu'un important dimère de  $\sim 40$  pb survient en absence de matrice.

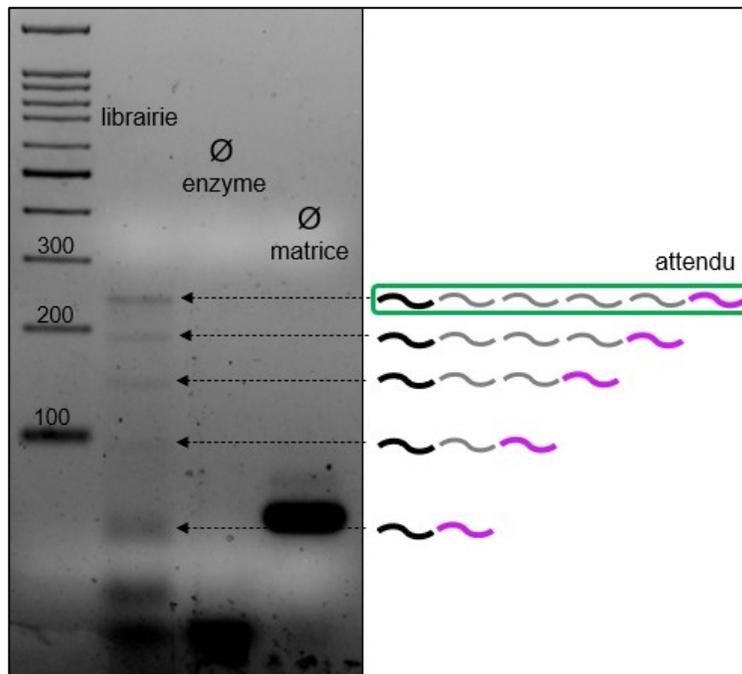


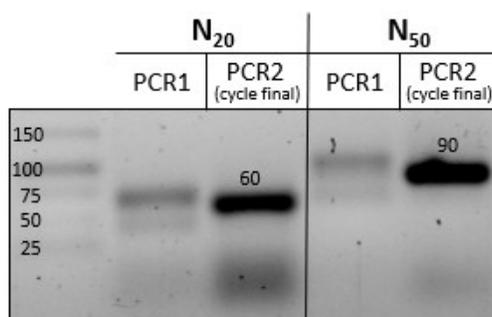
Figure 28: Amplification par RT-PCR de la librairie enzymatique  $N_{200}$ .

Une ligation de  $N_{50}$  à un oligonucléotide a été effectuée dans des conditions optimisées, puis migrée sur gel PAGE pour purifier la bande  $N_{200}$ . L'ADN a ensuite reçu un adaptateur en 3' par une deuxième ligation. Subséquemment, le mélange a été amplifié par RT-PCR avant d'être finalement migré dans un gel TAE agarose 2% à 120 V pendant 45 minutes. L'expérience entière a été faite deux fois ( $n=2$ , mêmes résultats).

Ce profil de migration en escalier n'est pas idéal dans le cadre d'un SELEX, mais cette librairie a tout de même été utilisée lors des SELEX (voir section 3.2). Une analyse plus poussée sera abordée lors de la discussion. La quantité utilisable d'ADN dans la librairie de ligation produite était de  $\sim 360$  pmol dans 270  $\mu\text{L}$  au total (méthodologie identique à la section 3.1.1.2, **Figure 22**).

### 3.1.3 Librairie chimique $N_{20}$ et $N_{50}$

Ayant en main deux librairies enzymatiques à tester, il était maintenant question de les tester en parallèle à des librairies classiques de courtes tailles synthétisées chimiquement par un fournisseur d'ADN (IDT). Une PCR de vérification a été effectuée afin de confirmer que ces courtes librairies étaient amplifiables (**Figure 29**). Les bandes obtenues ont la longueur attendue: 60 pb pour la librairie  $N_{20}$  ( $20+20+20$ ), et 90 pb pour la librairie  $N_{50}$  ( $20+50+20$ ). Un cycle final de PCR a été effectué afin d'obtenir de l'ADN parfaitement double-brin, mais contrairement à la librairie  $N_{800}$ , ce cycle ne semble pas nécessaire pour observer des amplicons sur gel. Les bandes de la PCR1 ne sont pas autant compactes que la PCR2 (cycle final), sans doute dû à la présence d'ADN partiellement simple-brin modifiant le profil de migration (voir section 2.1.2.5 pour plus de détails). Par contre, ceci n'empêchera pas de faire de futures extractions de gel puisque le signal est bien visible, même s'il est légèrement dissipé.



**Figure 29: Amplification PCR des librairies chimiques  $N_{20}$  et  $N_{50}$  avec cycle final.**

Deux courtes librairies ( $N_{20}$  et  $N_{50}$ ) synthétisées chimiquement ont été commandées (fournisseur: IDT). À titre vérificatif, ces oligonucléotides ont été amplifiés par PCR (PCR1), suivie d'un cycle final (PCR2). La PCR2 est issue d'une dilution de la PCR1 dans un nouveau mélange suivi d'un seul cycle PCR (avec nouveaux réactifs). Les échantillons ont été migrés sur gel TAE agarose 2% à 120 V pendant 45 minutes.

### 3.1.4 Récapitulatif des librairies et des ligands

Ayant les longues librairies enzymatiques et les courtes librairies chimiques en main, il était question de pondérer les quantités initiales de molécules; il serait injuste de comparer 1 pmol de librairie à 10000 pmol. Puisque seulement 1 pmol d'ADN de la librairie TdT ( $N_{800}$ ) peut être employée, il a fallu utiliser cette même quantité pour les deux librairies de plus courte taille ( $N_{20}$  et  $N_{50}$ ). Pour la librairie  $N_{20}$ , 1 pmol correspond environ à une molécule pour chaque séquence possible. Concernant la librairie de ligation  $N_{200}$ , puisque le profil de migration semblait très hétérogène et que le résultat des cycles SELEX était incertain, le 60 pmol entier a été utilisé afin d'augmenter les chances d'en retirer un aptamère. Dans un SELEX classique, il est typique d'utiliser 10 nmol de librairie initiale (Kohlberger et Gadermaier, 2021), et ainsi, 10 nmol de la librairie chimique  $N_{50}$  a été utilisée pour comparer à ces conditions plus « standards ». Concernant les billes magnétiques, trois types de ligands ont été employés: la galectine-7, les protéines NheA et HblB, de même que la protéine HP0953. Des billes bloquées ont aussi été produites, mais elles ne sont pas considérées comme des ligands; elles ne serviront que lors des cycles négatifs. Ci-dessous se trouve une compilation visuelle des ligands et des librairies (**Tableau 3**).

Tableau 3: Récapitulatif des trois ligands et cinq librairies employées dans les cycles SELEX.

ligands	librairies	quantité (pmoles)	type de synthèse
galectine-7	$N_{50}$	10000	chimique
	$N_{50}$	1	chimique
	$N_{20}$	1	chimique
	$N_{200}$	60	enzymatique (ligation)
	$N_{800}$	1	enzymatique (TdT)
NheA et HblB	$N_{50}$	10000	chimique
	$N_{50}$	1	chimique
	$N_{20}$	1	chimique
	$N_{200}$	60	enzymatique (ligation)
	$N_{800}$	1	enzymatique (TdT)
HP0953	$N_{50}$	10000	chimique
	$N_{50}$	1	chimique
	$N_{20}$	1	chimique
	$N_{200}$	60	enzymatique (ligation)
	$N_{800}$	1	enzymatique (TdT)

Les librairies  $N_{800}$  font en fait 500 à 1100 pb en longueur, mais le terme «  $N_{800}$  » (longueur moyenne) est employé pour sauver de l'espace et garder le tout simple. Il est important de se rappeler que les librairies enzymatiques  $N_{200}$  et  $N_{800}$  sont issues d'une PCR, donc leur nombre de pmol ne représente pas des molécules uniques, mais plutôt des copies d'un nombre plus petit de molécules uniques.

### 3.2 Cycles SELEX

Avant le départ des cycles SELEX, un test a été réalisé pour vérifier si l'ADN en surface des billes pouvait être récupéré à des fins d'amplification PCR. Le gel ci-dessous (**Figure 30**) souligne l'importance de chauffer les billes magnétiques afin de récupérer l'ADN collé en surface. Lorsqu'une PCR utilise les billes directement (sans chauffage) comme matrice plutôt que l'éluat, il ne semble pas y avoir d'amplification de la matrice attendue. À la lumière de ce résultat, un chauffage a été fait systématiquement sur toutes les billes juste avant la PCR.

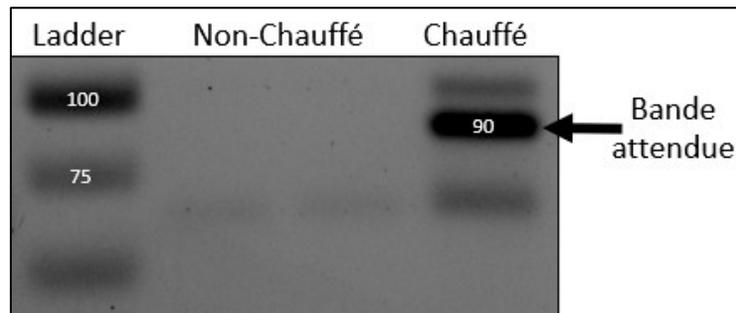


Figure 30: Importance du chauffage des billes avant la PCR.

Un cycle SELEX test a été effectué sur la librairie N50. Suite à l'incubation de la librairie sur la bille magnétique, deux PCR ont été effectués: la première utilisait les billes non-chauffées comme matrice, alors que la deuxième utilisait l'éluat de billes chauffées à 95°C pendant 5 minutes. Les amplicons PCR ont ensuite été migrés sur gel agarose 2%. Les deux puits « Non-Chauffé » sont le même échantillon.

Ayant les bibliothèques amplifiables en main, les cycles SELEX ont été entamés. Le gel ci-dessous présente un résultat typique d'amplicons PCR des différentes bibliothèques migrés dans un gel agarose. Dans le cas d'une PCR typique, des bandes de 90 pb ( $N_{50}$ ), 60 pb ( $N_{20}$ ), escalier de 40 à 240 pb ( $N_{200}$ ) et le signal diffus de la  $N_{800}$  ont été obtenus. Ce type de gel est attendu à travers les cycles SELEX. Par contre, des problèmes de PCR (présence de grandes régions diffuses) ont commencé à survenir dès le cycle 6 (**Figure 31**). Pour limiter le plus possible la présence de ces artefacts, des extractions de gel ont été faites au cycle 6 et à partir du cycle 9 jusqu'au cycle 15 final. L'ADN contenu dans ces bouts de gel a ensuite été élué puis digéré par l'exonucléase lambda afin de poursuivre les SELEX.

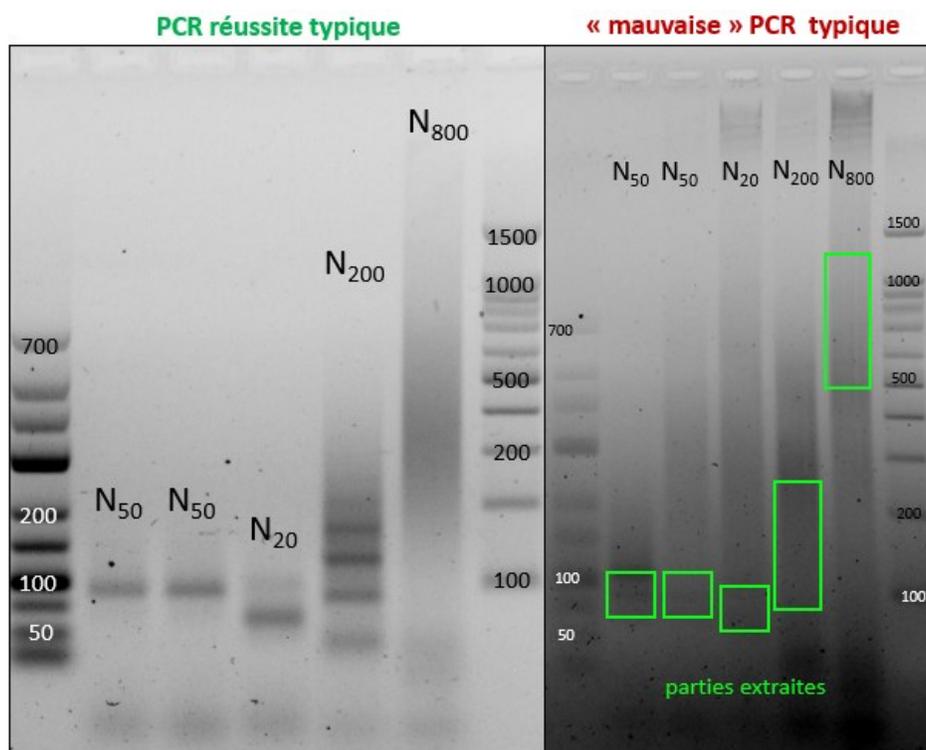


Figure 31: PCR typique (réussie ou avec artefacts) lors des cycles SELEX.

Un total de 15 cycles SELEX ont été effectués en utilisant en parallèle cinq bibliothèques ( $N_{50}$  à 10 nmol,  $N_{50}$  à 1 pmol,  $N_{20}$ ,  $N_{200}$  et  $N_{800}$ ) sur trois cibles (Gal7, toxines NheA et HblB, et HP0953). Seul le gel d'un SELEX pour une protéine a été présenté. Les bibliothèques ont été incubées avec des billes fonctionnalisées, puis lavées. Les billes ont ensuite été chauffées afin de relâcher et d'amplifier par PCR l'ADN retenu. Les échantillons PCR ont été migrés sur gel TAE agarose 2% à 120 V pendant 45 minutes. Les rectangles verts représentent les zones extraites du gel lorsque des purifications étaient nécessaires.

### 3.3 Dosages des fractions

Il est commun de vérifier l'enrichissement des bibliothèques dans un contexte de SELEX. Il est attendu d'obtenir de meilleurs rendements de liaison envers les cibles plus l'on avance dans les cycles. L'intensité de la goutte semble très intense au cycle 1 dans toutes les fractions, sauf pour les lavages (**Figure 32**). La fraction de surnageant et du lavage 1 semble diminuer en force au fil des cycles. Par contre, des variations sont observables pour les éluats et sont causées par la différence des paramètres des cycles SELEX (exemple: moins de billes magnétiques). Les lavages 2 et 3 ne semblent pas comporter une grande différence d'ADN entre eux à travers des cycles.

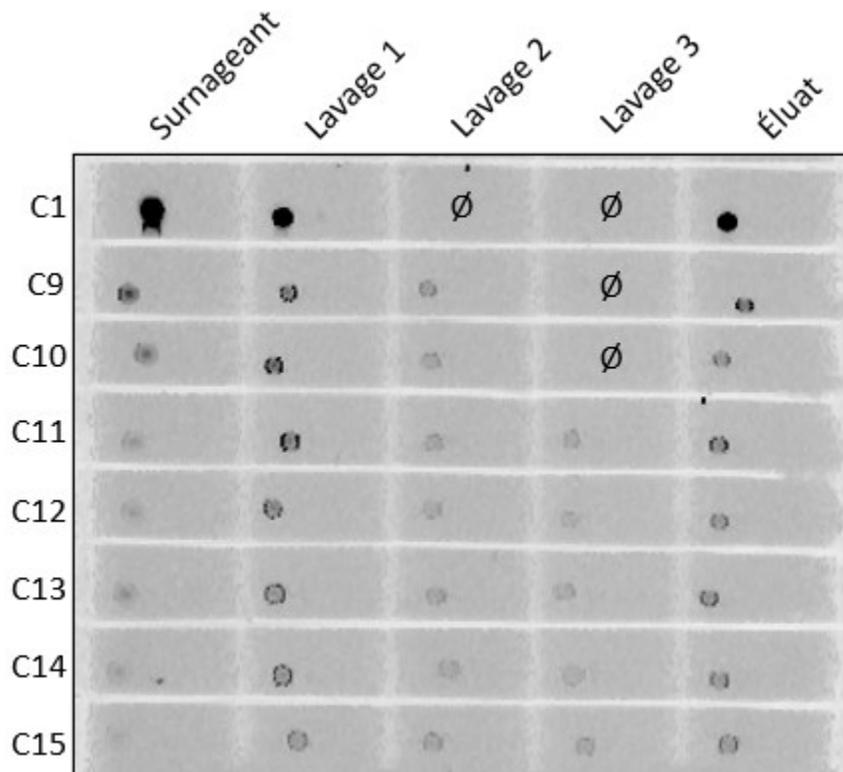
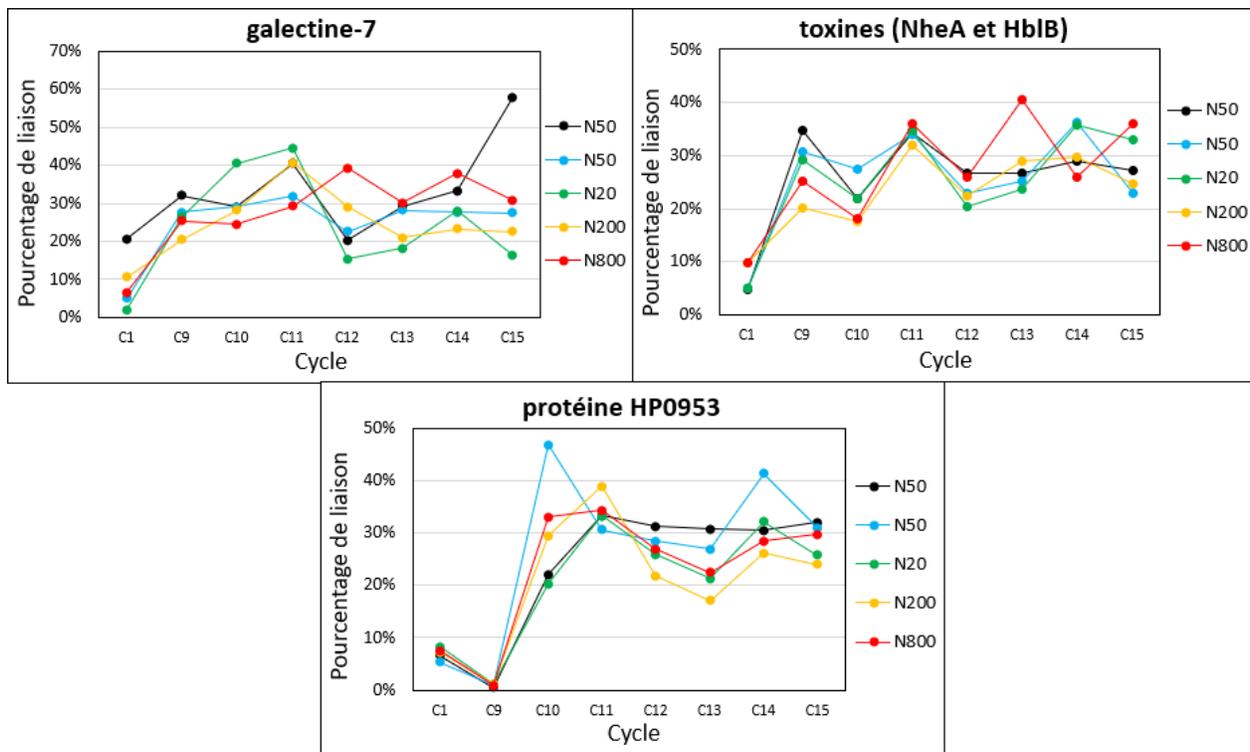


Figure 32: Dosage sur verre de gouttelettes contenant de l'ADN fluorescent issues des fractions des cycles SELEX.

Les fractions des cycles SELEX (surnageant de l'incubation, surnageant des lavages 1 à 3, et l'éluat de bille) ont été conservés puis dosés afin de mesurer un changement de ratio fluorescence (ADN avec fluorescéine en 5') à travers les cycles. Des gouttes de 2  $\mu$ L de chaque fraction ont été placées sur une plaque de verre hydrophobe quadrillée. La vitre a été scannée au Typhoon (force du laser à 500 V, résolution de 200  $\mu$ m, excitation à 473 nm et filtre d'émission BPB1 520-540 nm). Seule une bibliothèque d'un ligand est présentée, mais l'expérience a été répétée pour les cinq bibliothèques et les trois ligands. Les codes C1 à C15 en marge de l'image correspondent au numéro du cycle SELEX (C9 = cycle 9 du SELEX).

Une fois tous les dosages de fractions effectués, les intensités de chaque goutte (valeurs numériques) ont été entrées dans la formule présentée au chapitre « Matériel et méthode », à la section « Dosage des fractions ». Elle permet de déterminer le pourcentage des molécules collées aux cibles comparativement aux molécules totales. Les courbes d'enrichissements (**Figure 33**) ont toutes une allure de dents de scie. Ce phénomène peut être expliqué par les changements de paramètres à travers les cycles (exemples: lavage plus intense ou diminution de la quantité de ligands, décrits dans le **Tableau 2**). Par contre, si l'on observe la tendance générale des courbes, elles commencent toutes à un pourcentage de liaison bas au cycle 1 et finissent à un pourcentage plus élevé au cycle 15, ce qui pointe vers un enrichissement des séquences envers leur cible.



**Figure 33: Enrichissement des cinq bibliothèques au cours du SELEX.**

Les pourcentages de liaisons issus des gouttelettes de la Figure 32 ont été tracés sur un graphique. Ces graphiques mettent en relation la valeur des pourcentages de liaison en fonction du cycle SELEX. Chaque graphique représente un ligand (Gal7, toxines NheA et HbIB, et HP0953) ainsi que ses cinq bibliothèques attirées. La bibliothèque N<sub>50</sub> en noire représente la bibliothèque à 10 nmol, alors que la bibliothèque N<sub>50</sub> bleu pâle est celle à 1 pmol.

### 3.4 Séquençage et analyse

Les bibliothèques finales ayant passé à travers 15 cycles SELEX, il était maintenant question de les envoyer à séquencer afin d'identifier les séquences candidates potentielles d'aptamère. Les courtes bibliothèques N<sub>20</sub> et N<sub>50</sub> ont reçu un séquençage de type Miseq, alors que le type Nanopore a été employé pour les longues bibliothèques N<sub>200</sub> et N<sub>800</sub>.

#### 3.4.1 Préparation des bibliothèques N<sub>20</sub> et N<sub>50</sub>

Le séquençage à haut-débit Illumina requiert des séquences code-barres indexées de part et d'autre des échantillons à traiter. L'ajout des index contenant les code-barres Illumina à l'amplicon de base est nettement visible pour les deux PCR. L'ajout des séquences protubérantes en 5' à chaque brin semble donc bien réussi (**Figure 34**). Il y a aussi une absence de bruits de fond dans les puits sans matrice. Les échantillons ont ensuite été envoyés à la plateforme CERMO-FC de l'UQAM afin de recevoir un séquençage Miseq.

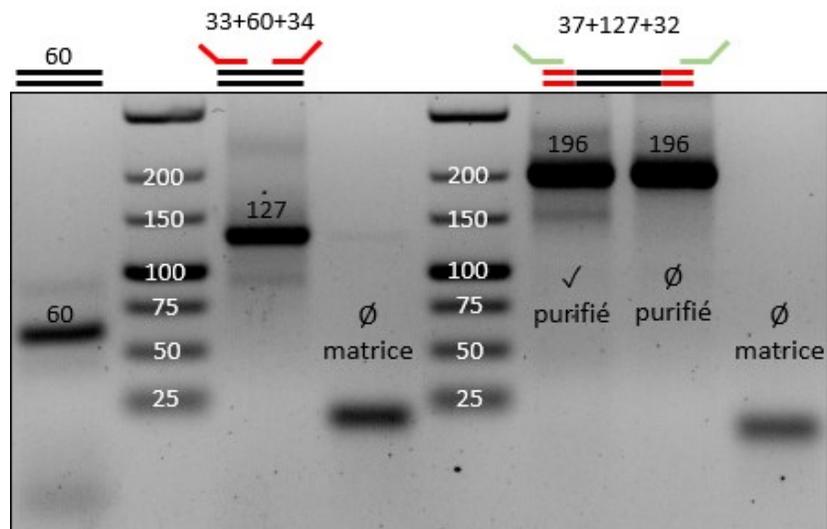
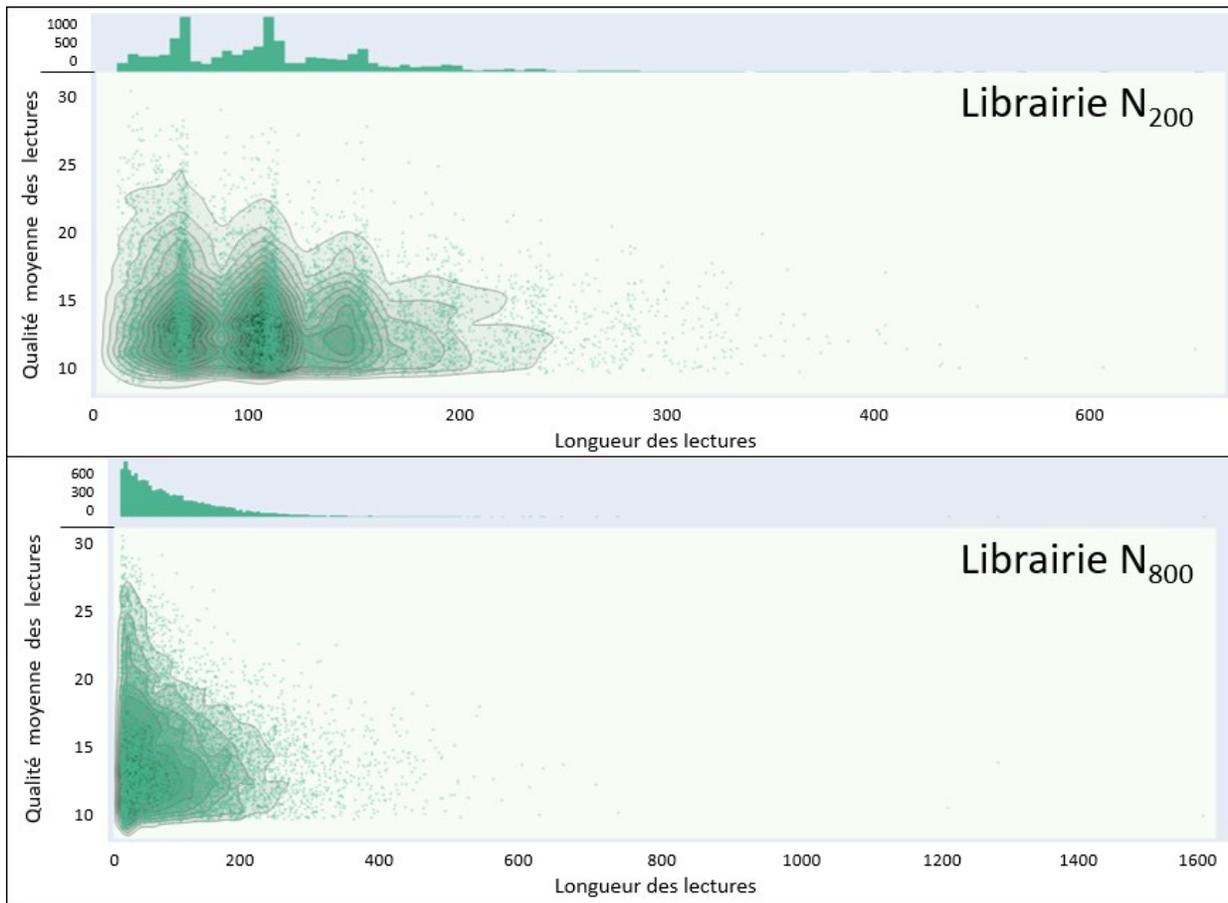


Figure 34: Indexation par PCR des bibliothèques en guise de préparation au séquençage à haut débit.

La bande de 60 pb représente une bibliothèque N<sub>20</sub> typique du cycle 15 du SELEX (sans code-barres). Deux PCR successives de 25 cycles ont été effectuées. La première utilisait des amorces contenant une séquence complémentaire à la fois à la bibliothèque et à l'amorce contenant l'index. La deuxième PCR utilisait strictement les amorces complètes d'indexation (i5-S518 et i7-N701 à N715). Les échantillons PCR ont été migrés sur gel TAE agarose 2% à 120 V pendant 45 minutes. Une fois vérifiés, les échantillons ont été envoyés pour être séquencés.

### 3.4.2 Contrôle qualité des librairies N<sub>200</sub> et N<sub>800</sub> après séquençage Nanopore

Les librairies N<sub>200</sub> et N<sub>800</sub> sont beaucoup plus longues, et ainsi ne peuvent pas être séquençées par Miseq. Un séquençage de type Nanopore a donc été fait, et la **Figure 35** ci-dessous représente le contrôle qualité de l'échantillon d'un seul ligand (deux librairies). Les librairies N<sub>200</sub> ont obtenu beaucoup de séquences regroupées en foyers se situant à 70, 110, et 150 pb. Ceci ressemble grossièrement au profil de migration en forme d'escalier des gels PCR N<sub>200</sub>. Les librairies N<sub>800</sub> possèdent toutes les trois un énorme foyer de lectures aux alentours de 40-50 pb. Similairement aux librairies N<sub>200</sub>, plus l'on monte dans la longueur des séquences, plus le nombre de lectures diminue avec quelques séquences uniques entre 500 et 1600 pb (longueurs attendues). Après analyse des fichiers .fasta N<sub>200</sub> et N<sub>800</sub>, plus de 99% des lectures correspondaient à des répétitions de bouts d'amorce. Ayant éliminé celles-ci, quelques séquences ressemblant à des séquences aléatoires ont été conservées (potentiels candidats d'aptamère).



**Figure 35: Exemple typique de contrôle qualité du séquençage Nanopore (NanoPlot Report) pour les librairies enzymatiques N<sub>200</sub> et N<sub>800</sub>.**

Chaque point vert correspond à une lecture. En axe des X se retrouve la longueur des lectures alors que l'axe des Y présente la qualité de la lecture. Ce contrôle qualité a été généré pour les librairies N<sub>200</sub> et N<sub>800</sub> des trois ligands (un seul est présenté). Une qualité moyenne de lectures élevée indique un bon résultat.

### 3.4.3 Candidats d'aptamères potentiels obtenus

Une fois les librairies séquencées, les séquences ne comportant pratiquement que des régions d'amorces ont été éliminées. Quelques séquences ont été retenues afin d'être testées en laboratoire en tant que candidat d'aptamère. Le **Tableau 4** ci-bas présente tous les candidats retenus. Ils représentent de potentiels candidats d'aptamère à tester en laboratoire. Certains candidats N<sub>20</sub> et N<sub>50</sub> n'avaient pas une longueur de région aléatoire attendue, mais ont tout de même été retenus dû à leur nombre élevé de lectures. Les candidats N<sub>200</sub> et N<sub>800</sub> possédaient tous certaines sections de leur séquence correspondant à des amorces, mais la majorité de la séquence entière semblait aléatoire (avec certains biais d'homopolymères ou de séquences répétées de dinucléotides). Les séquences entières sont disponibles aux **Annexes IV, V et VI** pour la Gal7, les toxines, et HP0953 respectivement.

**Tableau 4: Candidats d'aptamères potentiels obtenus suite à l'analyse des fichiers de sortie de séquençage Miseq et Nanopore pour toutes les librairies et cibles.**

librairie	galectine-7			toxines (NheA et HblB)			protéine HP0953		
	longueur	lectures	% lecture unique (total)	longueur	lectures	% lecture unique (total)	longueur	lectures	% lecture unique (total)
N <sub>50</sub> (10 nmol)	24	176	<b>91%</b> (11201)	26	18	<b>76%</b> (6371)	50	57	<b>73%</b> (6474)
	50	47		28	12		16	12	
	50	20		17	6		10	11	
	24	16		50	4		50	9	
	50	12		50	4		50	7	
	-	-		50	3		-	-	
N <sub>50</sub> (1 pmol)	27	107	<b>84%</b> (8955)	-	-	<b>52%</b> (9050)	49	68	<b>90%</b> (7308)
	50	83		-	-		50	25	
	24	69		-	-		50	17	
	32	65		-	-		49	14	
	50	41		-	-		50	13	
	50	16		-	-		-	-	
N <sub>20</sub>	20	32	<b>86%</b> (11417)	20	52	<b>92%</b> (12372)	20	51	<b>86%</b> (11368)
	20	20		20	22		30	45	
	20	19		20	20		20	28	
	-	-		-	-		20	16	
N <sub>200</sub>	273	1	-	359	1	-	118	1	-
N <sub>800</sub>	569	1	-	580	1	-	595	1	-

Les librairies N<sub>20</sub> et N<sub>50</sub> ont été séquencées par Miseq et analysées par le logiciel AptaSUITE. Les séquences ayant le plus de lectures et ne correspondant pas à des biais de PCR ont été retenues comme candidats. Les librairies N<sub>200</sub> et N<sub>800</sub> ont été séquencées par Nanopore. Les lectures ayant une trop importante couverture correspondant à des d'amorces répétées ont été éliminées. Seule la séquence ayant le moins gros pourcentage d'identité d'amorce a été retenue (une pour chaque librairie pour chaque ligand).

### 3.4.4 Contenu nucléotidique de la librairie initiale N<sub>800</sub>

La librairie initiale N<sub>800</sub> (avant les SELEX) a également été séquencée en parallèle aux librairies finales (après les SELEX) afin de déterminer leur contenu nucléotidique. Rappelons que cette librairie est issue d'une réaction d'élongation par la terminale transférase. Ceci communiquerait une information utile sur les possibles biais enzymatiques lors de la synthèse des régions aléatoires. En effet, il est notable que chaque type de nucléotide demeure présent dans la librairie N<sub>800</sub> initiale (trois séquences complètes sont disponibles à l'**Annexe VII**), ce qui n'est pas le cas des librairies finales N<sub>800</sub>. La préférence pour la librairie N<sub>800</sub> (région aléatoire synthétisée par la terminale transférase) se porte sur les purines (**Tableau 5**), ayant une prévalence égale à la fois pour les adénines (A) et les guanines (G). Le test t de Student révèle que les nucléotides A et G (purines) sont significativement plus abondants que les T et C (pyrimidines).

Tableau 5: Proportion nucléotidique (A, T, G, ou C) de la librairie enzymatique N<sub>800</sub> initiale avant les cycles SELEX.

séquence	A	T	G	C
1	34%	16%	33%	17%
2	37%	14%	34%	15%
3	32%	14%	36%	18%
4	34%	13%	35%	18%
5	38%	14%	33%	15%
6	38%	13%	35%	14%
7	34%	14%	35%	17%
8	37%	13%	34%	16%
9	36%	13%	34%	17%
10	37%	16%	31%	16%
11	33%	14%	36%	17%
12	32%	14%	37%	17%
13	34%	16%	34%	16%
14	35%	17%	32%	16%
15	30%	12%	40%	18%
16	32%	13%	38%	17%
17	33%	13%	36%	18%
18	34%	14%	36%	16%
19	38%	13%	33%	16%
20	34%	12%	36%	18%
<b>moyenne</b>	<b>35%</b>	<b>14%</b>	<b>35%</b>	<b>17%</b>

Vingt séquences du fichier de sortie de séquençage N<sub>800</sub> ont été prises afin de calculer leur pourcentage nucléotidique (A, T, G, et C). Ces séquences ressemblaient le plus possible à ce qui était recherché, c'est-à-dire les fragments ayant une région aléatoire flanquée par une région d'amorce avec le moins de biais d'amorce possible (c'est-à-dire avec peu ou pas de séquences ayant plus de sept bases consécutives similaires aux amorces). Les nucléotides entre les deux amorces ont ensuite été comptés sur un site web ([www.biologicscorp.com](http://www.biologicscorp.com), outils: *GC content calculator*). Plusieurs tests statistiques ont été effectués sur ces données: test de Shapiro-Wilk pour vérifier la normalité des données (valeur-p=6.019<sup>-9</sup>), test d'analyse de variances (ANOVA) pour vérifier l'égalité des variances (valeur-p=<0.00001), et finalement un test t de Student (valeur-p=<0.0001) afin de vérifier s'il y a une différence significative entre les différents nucléotides. Trois séquences complètes de ces 20 lectures sont présentées à l'**Annexe VII**.

Ci-dessous se retrouve une séquence de la région aléatoire produite par la TdT (deux autres sont à l'Annexe VII).

```
AGGCACCGGCGCAGGAGCGAAAACTCGGAAGGCAACGGAGAGACACGTAGAAGGAATGGGGAGGAAAGTGAAGTTATG
CCCCGGGAGGGATAGCTGGCACGGCCTGGCGGGGATGGGAAGGGGCATGCACAAGATAAACTGGAAAGGAGCTGGGAAAG
AACGGAGGTGAGATGCTGAAGGCATTACAAGGAAAGCTCGGGGATTGGCGCCGGTGGATAAGGGTTAGTGGAGACAAGGC
CGGGAATGAGATAGTGAATGAGAGGTAGGTGTCCACGAAAAGGAAGGGGAGACAGATTGCAAAAAAGGGAATATTGATG
TGTCACAGCAGAAAACGTATAAACGGCAGAAGAGCACGTCCAGGGGCGAGAAGGGGAACTCAAAGTTGCAGTATGGGACC
GGAGAGGAATATAGTGAGCGCCGACGCCAGAGCGTCCGCCAAAGGCACCATTGAGCCAGAGAAAGTATCAAACACTACTGG
ACGGCGAGCCGTCGGAAAGAGCGCAATGTGCGAGCGCAGATGGTAACGTAGCACAAGAAGTCAGAAAGACC GCGTGTGAT
CAATTAAGCTATTGGCGAGGGCGTTGTACTACACGCGCAGAGTGGGCTAAACAGGGACAGGGATTGAGCGATCGAGCGAC
TG TAGAATCAGGACTCGACGTGCAGGCTTAACACGAGCAACAACAAGTATGGTGGGTAAAGACTAGGGCGGTAAGGGCG
CGCGGACGATAAAGCCGGTGAAGGCGCGCCGCGCAAAAGTGCAA
```

Malgré la prévalence plus élevée des adénines et guanines, le contenu nucléotidique semble bien réparti. La bonne diversité nucléotidique de librairie N<sub>800</sub> initiale (avant le SELEX) n'est pas présente chez les librairies N<sub>800</sub> finales (après le SELEX). Ceci confirme que la terminale transférase peut produire des régions dégénérées *de novo*, ce qui répond au premier objectif. Lorsque l'on compare la séquence ci-haut à celle des N<sub>800</sub> des **Annexe IV, V, et VI**, de nombreuses répétitions mononucléotidiques ou dinucléotidiques deviennent évidentes. Cette comparaison suggère que l'apparition de ce genre de biais survient au cours des PCR successives des SELEX, et non pas lors de la synthèse de la librairie.

### 3.5 Tests de liaison

Il était maintenant question de vérifier si les candidats choisis sont réellement capables d'interaction avec leur cible. Quatre différents tests ont été effectués afin de vérifier ceci. Ces tests permettraient de qualifier les candidats comme étant de réels aptamères.

#### 3.5.1 Retard sur gel (*gel shift assay*)

Le retard sur gel est un test d'électrophorèse sur gel natif permettant de comparer deux différentes formes de molécules: seule ou liée avec une cible. Le profil de migration d'un ADN seul sera différent de celui du même ADN collé à sa cible. Les candidats ont donc été migrés en présence de différentes concentrations de protéine (**Figure 36**). Aucun retard ne semble observable en présence des cibles respectives des candidats comparativement au candidat seul. Plusieurs autres gels (non présentés) ont été effectués afin d'optimiser les conditions, et aucun retard n'a été repéré.

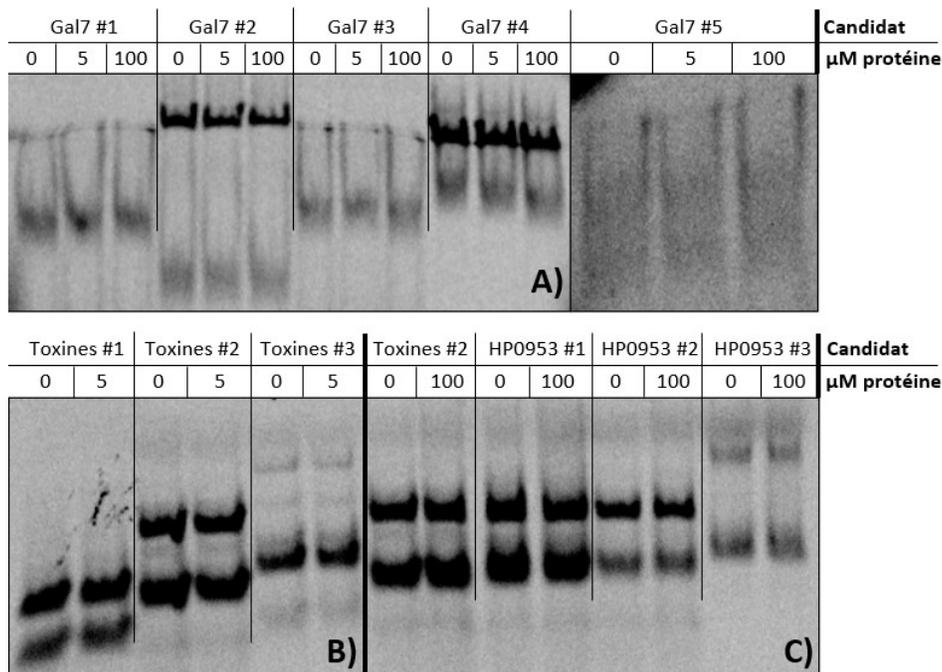
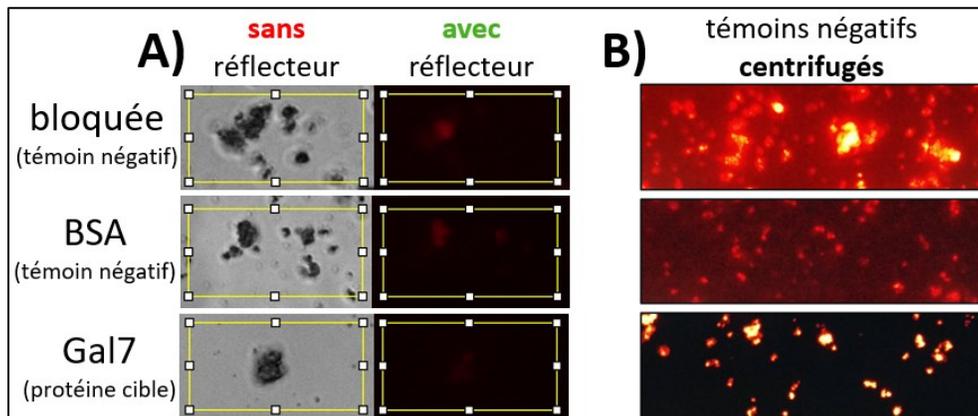


Figure 36: Retard sur gel natif PAGE comparant 12 candidats d'ADN (Cy5) seuls ou en présence de leur cible (protéine Gal7, toxines, et HP0953).

Les 12 candidats (ADN fluorescents Cy5) ont été incubés avec leur ligand respectif pendant 30 minutes à 21°C. Les mélanges ont ensuite été migrés dans des gels TBE 6% PAGE natif à 150 V pendant 6 heures à 4°C. Les gels ont été scannés au Typhoon. Chaque puits contient ~1 pmol d'ADN. Ligands testés: A) Galectine-7 (deux gels distincts), B) toxines NheA et HblB, C) HP0953 (candidat de toxines présent comme contrôle).

### 3.5.2 Microscope inversé

Le deuxième test de liaison est par observation microscopique. Le principe est d'incuber les candidats d'ADN fluorescents avec des billes magnétiques contenant en leur surface diverses protéines, laver les billes, puis les observer au microscope à statif inversé (détecte la fluorescence des ADN). Une fluorescence plus intense indiquerait une meilleure interaction. De la fluorescence est visible sur tous les types de billes (**Figure 37**). Malheureusement, il est difficile de discerner l'intensité de fluorescence des billes d'un échantillon à l'autre. Ceci pourrait être dû au fait que les billes sont trop petites (diamètre: 200 nm) pour le microscope. Or, il n'est pas possible de savoir si une forte coloration coïncide à plus d'ADN couvrant des billes, ou simplement à un amas de billes plus important. La comparaison entre des billes bloquées, des billes avec BSA et les billes positives n'est donc pas possible dans les conditions actuelles. De plus petits amas de billes ont été observés (non présenté), mais obtiennent un signal très faible, presque indétectable. Lorsque centrifugées avant l'observation, toutes les billes brillent fortement en fluorescence.

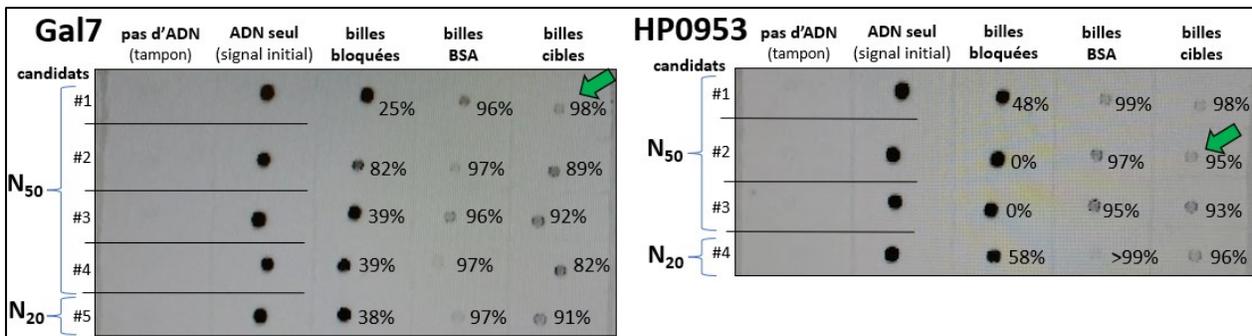


**Figure 37:** Billes magnétiques (fonctionnalisées avec des protéines) incubées avec de l'ADN fluorescent (candidats). Images prises par un microscope à statif inversé.

Des billes magnétiques NHS (200 nm de diamètre) ont été fonctionnalisées de trois différentes façons: aucune protéine (billes bloquées), avec la BSA (protéine négative), ou avec la Gal7 (protéine cible). Ces trois types de billes ont été incubés avec un aptamère candidat de la Gal7 sur rotor pendant 30 minutes à température pièce. Les billes ont ensuite été lavées, puis placées entre lame et lamelle avant d'être visualisées par un microscope inversé. A) sans centrifugation B) avec centrifugation avant le prélèvement des billes à placer sur la lame. Intensité du laser à 10%. La dimension du rectangle jaune est d'environ 20  $\mu\text{m}$  x 10  $\mu\text{m}$ . L'excitation du fluorophore des ADN (Cy5) a été faite à 673 nm et l'émission à 650 nm. La photo a été prise avec un microscope optique inversé (modèle: Axio Observer.Z1 / 7, objectif utilisé: alpha Plan-Apochromat 100x/1.46 Oil DIC (UV) M27, logiciel: ZEN version 2,6).

### 3.5.3 Essai à la goutte (*spotting test*)

Ce troisième test de liaison dose la quantité d'ADN présent dans un milieu avant et après une incubation avec des billes magnétiques fonctionnalisées (bloquées, BSA ou cible). Ainsi, une diminution d'intensité indique que de l'ADN a été pêché par les billes, pointant vers une interaction ADN-cible. La présence de bille bloquée (sans protéine en leur surface) ne semble pas diminuer la quantité d'ADN (**Figure 38**), ce qui pointe vers une absence d'interaction ADN-bille vide. Les billes fonctionnalisées avec la protéine cible des candidats semblent causer une déplétion du milieu. Cependant, ceci semble aussi le cas pour les billes BSA (protéine négative). Or, il semble y avoir un manque de sélectivité protéique pour la plupart des candidats. Par contre, certains candidats semblent préférer leur cible à la BSA (candidat Gal7 #1 et HP0953 #2), ce qui pointe potentiellement vers des aptamères plus sélectifs.



**Figure 38:** Essai à la goutte mesurant l'interaction entre les candidats (ADN fluorescent) et des billes magnétiques fonctionnalisées. Mesure prise par le niveau de déplétion d'ADN du milieu lorsqu'il est pêché par les billes.

Les candidats (ADN fluorescent Cy5) ont été dosés avant et après une incubation avec des billes magnétiques fonctionnalisées de différentes façons (bloquée, BSA, ou cible Gal7 / HP0953). L'incubation a durée 15 minutes à température pièce, puis les billes ont été éliminées du surnageant en plaçant les tubes sur un support magnétique. Des gouttes de 2 µL du surnageant ont été placées sur une plaque de verre hydrophobe, puis scannées au Typhoon (force du laser à 500 V, résolution de 100 µm, excitation à 635 nm et filtre d'émission LPR >665 nm). La cible toxine est absente, car seule une très petite quantité de ce ligand était disponible. Les pourcentages indiquent la proportion d'ADN s'étant collée aux billes comparativement à l'ADN seul. Les flèches vertes indiquent les meilleurs candidats (BSA plus foncé comparativement à la cible).

### 3.5.4 Thermophorèse à micro-échelle

Le quatrième test de liaison est axé sur la thermophorèse. Le principe repose sur une différence de migration entre des candidats (ADN fluorescents) seuls et des candidats sous forme de complexe avec leur cible. Les complexes migrent plus lentement, donc il est possible de distinguer la fluorescence locale entre des échantillons exposés à différentes concentrations de cibles. Malheureusement, il ne semble pas y avoir d'association entre la concentration du ligand et le signal émis (**Figure 39**). Une disposition descendante (ou ascendante) était attendue, mais les points semblent répartis de façon aléatoire, indiquant qu'il n'y a pas d'interaction entre les candidats et leur cible. Certains échantillons avaient un semblant de tendance, mais de différents signaux ont été obtenus lors d'une deuxième mesure (non reproductible). Puisque la BSA semblait indiquer une interaction avec les candidats, elle a aussi été testée comme cible avec cette approche, mais a aussi obtenu des résultats négatifs (bruit de fond).

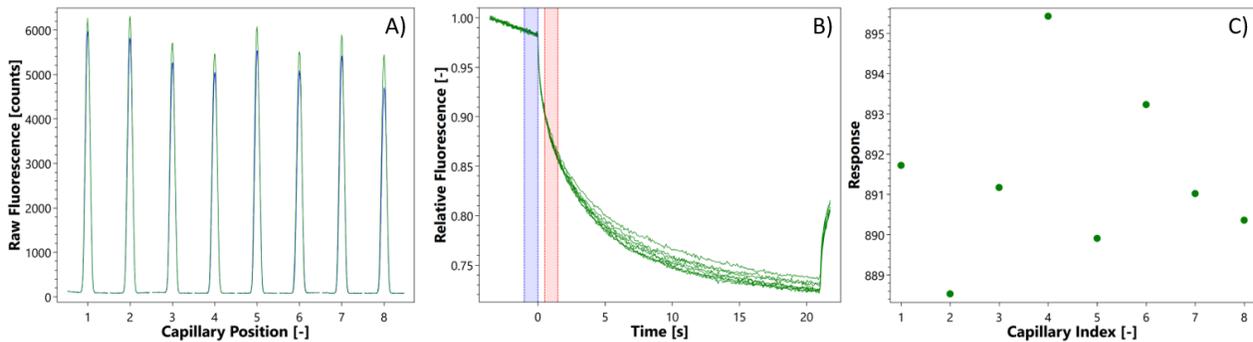


Figure 39: Fichier de sortie typique de thermophorèse à micro-échelle présentant une courbe de liaison entre une concentration fixe de candidat (ADN fluorescent) et une concentration grandissante de sa cible respective.

Les candidats (ADN fluorescents Cy5) ont été incubés avec une quantité croissante de protéine (15 minutes dans du tampon SELEX). Respectivement aux capillaires 1 à 8, les concentrations suivantes de protéine ont été employées: 0, 0,032, 0,16, 0,8, 4, 20, 100 et 500 nM (l'ADN restait à 2,5 nM). Du Tween-20 à 0,5% a été ajouté aux mélanges avant le chargement et la lecture des capillaires. A) fluorescence initiale de chaque capillaire avant chauffage, B) fluorescence relative avant chauffage (région bleue) et après chauffage (région rouge) en fonction du temps, C) courbe de liaison mettant en relation la fluorescence de l'avant et l'après chauffage du graphique en B. Un seul candidat est présenté, mais cette expérience a été faite pour chacun des 12 candidats, incluant ceux de la toxine (NheA et HblB). Les expériences ont aussi été répétées en employant la BSA plutôt que les cibles, tous obtenant des résultats comparables, c'est-à-dire sans interaction apparente.

## 4 DISCUSSION

---

### 4.1 Synthèse des bibliothèques enzymatiques

#### 4.1.1 Terminale transférase

Un aspect flagrant des résultats est la petite quantité initiale des bibliothèques enzymatiques  $N_{800}$  à  $\sim 1$  pmol. Encore pire, cette quantité est issue d'une réaction PCR. Ceci signifie que dès le départ, les molécules ne sont pas uniques, mais plutôt que chaque séquence est présente en plusieurs copies. Cette PCR était nécessaire, car seule une très faible quantité d'ADN était récupérable après la réaction enzymatique TdT. En effet, une grande difficulté amenée par les fragments d'ADN de longue taille est un faible rendement d'élution une fois migré dans un gel. Dans le cas présent, le gel était jugé nécessaire, car il fallait éliminer les courts fragments et seulement récolter les longs fragments. Par contre, d'autres techniques pourraient accomplir ceci sans avoir recours à des éluions. Par exemple, une séparation de taille d'ADN par chromatographie par filtration de gel de type sephadex (Ó'Fágáin *et al.*, 2011). Ce type de purification permet de séparer les grandes entités des plus petites (Ó'Fágáin *et al.*, 2011). Les colonnes sont jugées comme étant chimiquement « inerte », donc seule la séparation physique des molécules en fonction de leur grosseur est prise en compte (Ó'Fágáin *et al.*, 2011). Par conséquent, il serait possible de récupérer plusieurs fractions ayant passé dans la colonne, chacune comportant une longueur décroissante d'ADN (vérifiable sur gel par la suite). Pour ces colonnes, seule la pression du piston d'une seringue est requise pour la séparation des différentes tailles. Il est ainsi possible de faire un mélange à grand volume (dans l'ordre des mL), charger le tout dans une seringue de 10 mL, puis simplement passer l'échantillon dans la résine afin de fractionner les longueurs des ADN allongés. Possiblement que la diversité et quantité d'ADN de ce genre de méthode ne demanderait pas de PCR afin d'arriver à une quantité acceptable d'ADN dans la bibliothèque initiale. Ce pourrait être une avenue à explorer puisque l'exactitude de la longueur n'est pas aussi importante que de simplement séparer les longs des petits brins.

Plusieurs gels ont été présentés dans les résultats démontrant que la terminale transférase est très sensible à divers changements de conditions (**Figures 18, 19, et 20**). Il serait sans doute possible de faire plus d'essais additionnels, surtout axés sur les ions. Un exemple concret est que l'activité de l'enzyme semble diminuer de façon importante lorsque le fragment d'ADN allongé forme des structures supérieures importantes, comme la forme d'épingle à cheveux (*hairpins* en anglais) (Barthel *et al.*, 2020). Ceci peut diminuer la longueur possible des régions aléatoires. Par contre, ce phénomène peut être atténué en utilisant certains cofacteurs additionnels (Barthel *et al.*, 2020).

En parcourant la littérature disponible sur cette enzyme, beaucoup d'articles traitent sur sa description et ses comportements en présence de différentes conditions (Ashley *et al.*, 2021; Chua *et al.*, 2020; Sarac et Hollenstein, 2018; Tang *et al.*, 2017), mais aucune n'explore des améliorations mutagéniques afin de rendre l'enzyme plus efficace. C'est en partie l'intérêt de travailler avec cette enzyme; elle est bien décrite, mais très peu utilisée expérimentalement malgré son grand potentiel, indiquant de bonnes opportunités de publications. Par contre, depuis quelques années, plusieurs nouvelles applications biologiques semblent être explorées (Chua *et al.*, 2020; Sarac et Hollenstein, 2018). Une récente publication présente une évolution dirigée thermostable de la TdT, permettant de faire la réaction à 58°C plutôt que 37°C (Chua *et al.*, 2020). Ceci diminue la prévalence des structures secondaires des ADN pouvant interférer avec l'incorporation de nouveaux nucléotides (Chua *et al.*, 2020). Possiblement qu'un futur proche amènera une version plus productive de l'enzyme permettant de synthétiser de plus abondantes librairies. Entre-temps, une nouvelle application de la terminale transférase a été découverte dans l'ouvrage présent, permettant de synthétiser de longs fragments dégénérés amplifiables par PCR. L'article se rapprochant le plus de cette nouveauté est celui de Ashley *et al.* (2021) où les auteurs décrivent un nouvel aptamère issu d'une réaction de terminal transférase suivie d'une seule incubation avec un ligand. Une queue polyA, rendant l'ADN amplifiable, n'a été ajoutée qu'après la sélection des séquences s'appariant à leur ligand (Ashley *et al.*, 2021). Ne pas avoir recourt à des PCR successives diminue grandement la chance d'obtenir des artefacts d'amplification. On pourrait s'imaginer qu'un seul cycle de sélection offrirait des candidats à faible affinité pour leur ligand, mais au contraire, l'article rapporte des  $K_D$  atteignant parfois l'ordre des picomolaires. Cette approche semble bien adaptée pour obtenir des aptamères efficaces rapidement, mais tout comme la grande majorité des articles, ces aptamères demeurent de courte taille (~20-150 nt). Ainsi, nos travaux qui permettent de rendre de longs ADN simple-brin amplifiables offrent donc une nouveauté pouvant potentiellement être exploitée ultérieurement. Quelques essais de plus seraient de mise afin de rendre la longueur des ADN moins hétérogènes, mais cette innovation demeure tout de même utile à avoir accès parmi le répertoire des outils en biologie moléculaire.

#### **4.1.2 T4 ARN ligase 1**

Comparativement à la terminale transférase, les performances de la T4 ARN ligase 1 ne semblent pas autant varier lorsque les conditions de réaction changent. Il est possible de légèrement améliorer le rendement de ligation, mais pas de plusieurs ordres de magnitude. Souvent, il n'est question que de variation de 10 à 50%. Or, il est possible d'augmenter la quantité de la librairie, mais pas d'obtenir de longs

fragments. Effectivement, la région aléatoire la plus longue obtenue était à 300 nucléotides, mais en très faible quantité (à peine détectable). Il est important de rappeler que cette enzyme est couramment utilisée pour circulariser des oligonucléotides simple-brin ayant une longueur minimale de six nucléotides (Petkovic et Müller, 2015). Par conséquent, il est fort possible qu'une grande portion des fragments  $N_{50}$  employés (possédant tous un 5' phosphate et des ARN en 3') se circularisent sur eux-mêmes plutôt que de se coller en chaîne. Il y a donc une réelle limite physique à l'enzyme de liguer successivement plusieurs fragments en tandem de façon répétée. Le processus freine rapidement.

Une autre lacune de cette technique est la présence de dégradation (hypothétique) des jonctions d'ARN entre chaque  $N_{50}$ . Malgré que seule la bande  $N_{200}$  ait été extraite d'un gel, une représentation de toutes les bandes ( $N_0$ ,  $N_{50}$ ,  $N_{100}$ , et  $N_{150}$ ) est visible suite à l'amplification PCR (**Figure 28**). L'hypothèse amenée est qu'après la ligation des  $N_{50}$ , mais avant l'ajout de l'adaptateur (deuxième assise d'amorce), il y aurait eu de la dégradation entre les fragments  $N_{50}$ . Ainsi, l'ajout d'adaptateurs modifierait tous les types de molécules et non pas seulement les  $N_{200}$ . Même lorsque l'expérience a été répétée de façon rapide, avec du matériel sans RNase, et à une température froide, cette dégradation est reproductible. Il est notable que même si la longueur attendue est présente parmi un amalgame d'autres longueurs, les courtes séquences sont très favorisées lors des PCR (Kohlberger et Gadermaier, 2021) ce qui biaiserait les résultats. Cette librairie a tout de même été testée à travers un SELEX malgré cette difficulté connue.

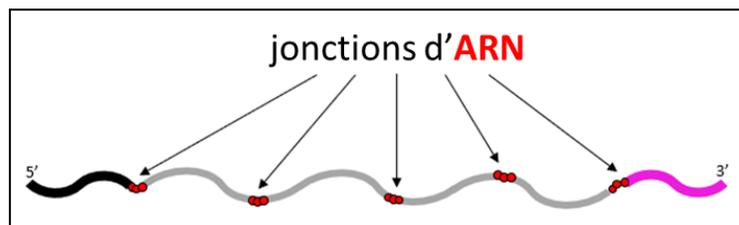


Figure 40: Schéma visuel des jonctions d'ARN entre chaque fragment  $N_{50}$  de la librairie  $N_{200}$ .

Deux articles ont déjà exploré la construction d'énorme concatémère d'ADN double-brin par ligation, atteignant des longueurs de plusieurs milliers de pb (Fujishima *et al.*, 2015; Kuhn et Frank-Kamenetskii, 2005). Cependant, ces ouvrages se basent sur l'assemblage organisé de blocs d'ADN à séquence connue, ce qui ne se qualifie pas comme étant une librairie dégénérée aléatoire. Or, dans un contexte de création d'une longue librairie aléatoire simple-brin, la ligation simple-brin s'avère peu efficace comparativement à simplement commander un tube mère d'oligonucléotides synthétisés chimiquement.

#### **4.1.3 Potentielles méthodes additionnelles**

La terminale transférase et la T4 ARN ligase 1 ont été employés afin d'explorer leur capacité à construire de longues régions aléatoires. D'autres approches seraient aussi envisageables. Par exemple, il serait possible d'utiliser des blocs de nucléotides aléatoires double-brin ayant des protubérances simple-brin (sites de restriction) afin de bâtir de larges fragments, similairement à la technologie présentée par Fujishima *et al.* (2015).

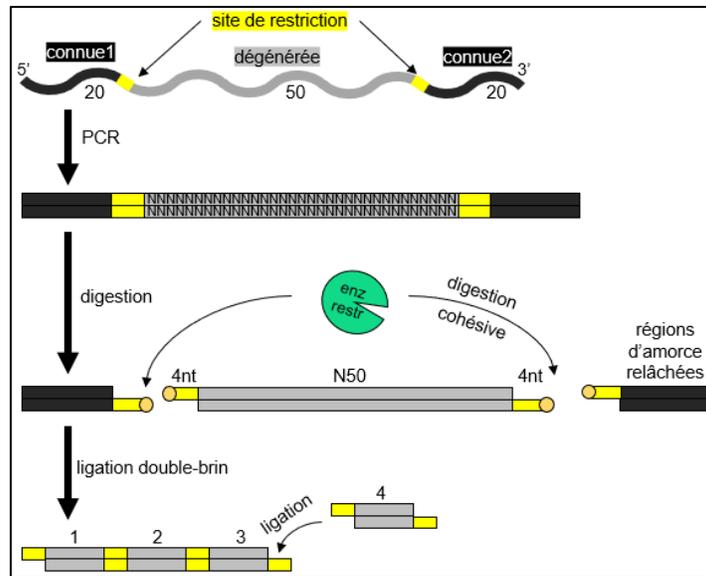


Figure 41: Schématisation d'une méthode alternative de création d'une large région dégénérée par ligation double-brin avec la T4 ADN ligase.

Une deuxième approche de synthèse pourrait se baser sur l'ADN polymérase phi29. Cette polymérase isothermale opère sur une matrice circulaire et est capable de synthétiser successivement des répétitions linéaires simple-brin (Johns *et al.*, 2009). Ainsi, en utilisant une seule amorce, il serait possible d'allonger une matrice aléatoire à une longueur voulue. Il serait aussi envisageable de combiner cette méthode en insérant un site de restriction dans la région d'amorce. Ceci permettrait de partiellement digérer le long brin final (en limitant la quantité d'enzyme) et de contrôler approximativement la longueur désirée. On peut s'imaginer que plusieurs  $N_{200}$  successifs se répétant pourraient mener à un aptamère d'une grande envergure ayant plusieurs points d'attache pour un même ligand.

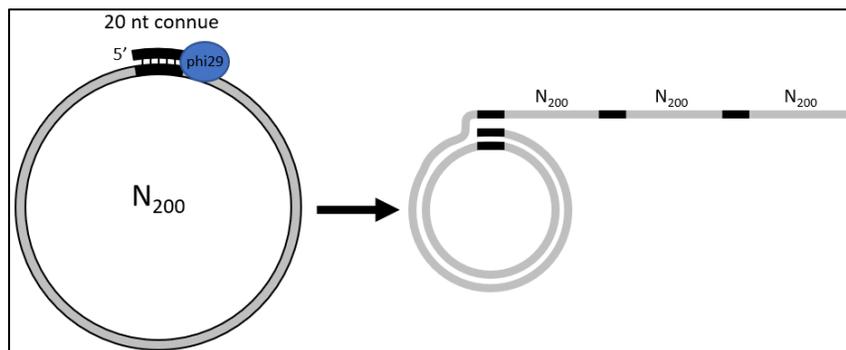


Figure 42: Schématisation d'une méthode alternative de création d'une large région dégénérée par l'ADN polymérase phi29.

## 4.2 Cycle SELEX

### 4.2.1 Billes magnétiques

L'utilisation de billes magnétiques NHS a été simple et efficace, mais certaines difficultés sont survenues. La première est qu'il n'est pas possible de savoir exactement la quantité de protéines fonctionnalisées en surface. Des essais de dosage par le réactif de Bradford et Nanodrop (non présentés) ont été tentés, mais sans résultat positif. Un article mentionne que les groupements NHS relâchés lors de la fonctionnalisation des billes interfèrent avec la mesure d'absorbance à 220-260 nm (Klykov et Weller, 2015). Ce masquage rend difficile la mesure des ADN ou protéines restants dans un surnageant avant et après l'incubation avec les billes. Il est possible de savoir la surface disponible théorique de la bille et le nombre maximal de protéines pouvant s'y coller puisqu'on connaît la taille (kDa) des protéines. Néanmoins, ces estimations sont très approximatives et pourraient facilement varier d'un facteur de 10 à 100. Une solution à cette difficulté est d'employer de plus grosses billes, idéalement ayant un diamètre d'au moins 1  $\mu\text{m}$ . Ceci rendrait les billes facilement visibles au microscope même lorsqu'isolées (une seule plutôt que dans un amas), ce qui permettrait de mesurer et doser la quantité de ligands marqués fonctionnalisés sur la bille.

La deuxième difficulté est que la majorité des billes perdent leur magnétisme lorsqu'on les vortex. En effet, une perte d'approximativement >80% des billes est observable après trois vortex à 30 secondes chaque. Une hypothèse est que ceci est dû au frottement des billes contre la paroi du tube, ce qui les endommage. Ce phénomène n'a été remarqué qu'à la fin des cycles SELEX. Plusieurs brassages par vortex ont été faits lors des lavages des SELEX et au cours de leur fonctionnalisation, et ainsi une importante perte de quantité de ligands a eu lieu au cours des expériences. En rétrospection, il serait raisonnable de s'en tenir qu'à un seul brassage par vortex plutôt que trois, ou de ne faire que des brassages par pipettage ou inversion. Il faut tout de même garder en tête qu'un bon brassage est probablement important, puisqu'il permet de bien déloger les molécules faiblement attachées, ou coincées dans les espaces entre les amas de billes.

### 4.2.2 Séquençage des bibliothèques enzymatiques N<sub>200</sub> et N<sub>800</sub>

Un résultat observable à travers tous les fichiers de séquençages est une forte prévalence de séquences répétées correspondant aux amorces de PCR. Plus de 99% de tous les fragments séquencés à la fin des SELEX correspondaient à ces répétitions. Même une fois éliminé, le <1% des séquences restantes comportait aussi une importante proportion de séquence correspondant à des amorces. De plus, les

séquences restantes ne comportaient qu'une seule lecture, indiquant une absence d'enrichissement mesurable. Une autre importante observation est le biais de séquence des régions entre les régions d'amorce. Une considérable prévalence de répétition de nucléotide est observable, allant souvent dans la centaine (ex: AGAGAG... ou GGGGGG...) (**Annexes IV, V, et VI**). Ces régions mononucléotide ou dinucléotides sont connues comme causant des erreurs de réplication lors d'amplification PCR (Clarke *et al.*, 2001). Il est possible que ces lectures ne soient donc pas des candidats d'aptamères, mais plutôt le résultat de plusieurs PCR erronées ayant accumulés des erreurs de réplication lors des SELEX. Au contraire, ce biais n'est pas présent sur les fichiers de sortie du séquençage des librairies N<sub>200</sub> et N<sub>800</sub> initiales (avant les SELEX). Cette disparité suggère que ce phénomène apparaît lors des PCR successives des SELEX, et non pas à la synthèse initiale des librairies. Malgré cette possibilité, des tests de ces candidats ont tout de même été tentés, mais un autre problème associé à ces répétitions est leur difficulté de synthèse commerciale (Czar *et al.*, 2009; Kosuri et Church, 2014). Rappelons que les longs fragments (>200 nt ou pb) pouvant être commandés sont généralement utilisés à des fins de synthèse protéique issues de gènes (Czar *et al.*, 2009; Kosuri et Church, 2014). La synthèse de ces longs oligonucléotides se fait par PCR par assemblage (*assembly-PCR*) (Czar *et al.*, 2009; Kosuri et Church, 2014). Un arrangement dans un ordre bien précis est requis pour pouvoir former ces longs fragments si l'on veut commander les candidats. Or, si un bloc peut être mis à plusieurs endroits dans un même fragment, cela mènera nécessairement à une construction fautive. Par conséquent, il n'est pas possible de se procurer des candidats d'aptamère consistant en de longues séquences ayant ce type de répétition de dinucléotide par l'entremise d'un fournisseur industriel. Afin de tester les candidats de longues tailles ayant ces répétitions, il a fallu conceptualiser de nouvelles longues amorces afin d'aller chercher à même la librairie finale chaque candidat. Mais malgré plusieurs essais, ces efforts n'ont pas porté fruit. On n'a donc jamais pu tester les candidats d'origine enzymatique de longue taille.

Un fait intéressant pour la librairie N<sub>800</sub> initiale est une forte préférence pour les purines (A et G à 30-40% chaque) par l'enzyme terminale transférase (**Tableau 5**). Ce phénomène est bien décrit dans la littérature dépendamment des cations divalents utilisés (Sarac et Hollenstein, 2018). Plus spécifiquement, le Mg<sup>2+</sup> du tampon amène une préférence vers les purines, alors que le Co<sup>2+</sup> (aussi utilisé) augmente l'incorporation des pyrimidines (Sarac et Hollenstein, 2018). Si l'on se fie aux résultats de séquençage obtenus, ces deux ions combinés dans leur concentration respective favorisent les purines de façon significative (**Tableau 5**). Conséquemment, une option serait de modifier le ratio A:T:G:C du dNTP, ou d'augmenter la quantité de cobalt dans la réaction afin d'atteindre un meilleur équilibre nucléotidique.

#### 4.2.3 Séquençage des librairies chimiques N<sub>20</sub> et N<sub>50</sub>

Contrairement aux librairies enzymatiques, les librairies commerciales N<sub>20</sub> et N<sub>50</sub> ont obtenu plusieurs séquences à lectures multiples (**Annexes IV, V, et VI**). À première vue, la région centrale entre les assises d'amorce semble correspondre à des régions aléatoires, mais plusieurs de ces régions étaient à longueur attendue. Par exemple, une librairie N<sub>50</sub> obtenait parfois une séquence ayant une région aléatoire de 20 nucléotides plutôt que 50. De plus, ces séquences écourtées faisaient souvent partie des séquences ayant le plus de lectures. Malgré cette anomalie, une douzaine de candidats N<sub>20</sub> et N<sub>50</sub> ont été testés lors de tests de liaisons, car ils ont tout de même été sélectionnés lors d'un SELEX. Malheureusement, malgré quatre différents tests de liaisons (**Figures 36, 37, 38 et 39**), aucun de ces candidats ne semble être un aptamère. Ceci souligne l'importance de bien tester les candidats obtenus suite à un séquençage de SELEX. Même l'analyse bioinformatique suggère un bon potentiel pour certaines séquences, il n'y a aucune garantie que ces séquences collent réellement à leur cible; elles peuvent simplement dériver d'un biais de PCR.

#### 4.2.4 Optimisation des PCR

Voyant les résultats de séquençage obtenu après les cycles SELEX, il est évident que les réactions de PCR ont mené à la formation de grands nombres d'artefacts. De nombreuses séquences correspondantes aux amorces se sont retrouvées un peu partout dans les lectures. Ce genre de difficulté est bien connu dans la sphère des SELEX (Kohlberger et Gadermaier, 2021; Sun *et al.*, 2015; Tolle *et al.*, 2014). En effet, les matrices d'oligonucléotides dégénérés sont difficiles à amplifier de façon efficace et ont tendance à former plusieurs produits non-désirables plutôt que des aptamères spécifiques (Sun *et al.*, 2015). En plus de ce phénomène bien connu pour les librairies à longueurs normales (<100 nt), on peut s'imaginer une augmentation des possibles difficultés lorsqu'on a affaire avec une librairie beaucoup plus complexe comme le cas présent.

Plusieurs types d'artefacts peuvent se produire ayant chacun leur cause et complexité (Kohlberger et Gadermaier, 2021), mais les types de migrations obtenues à la **Figure 31** ressemblent beaucoup aux artefacts décrits par Tolle *et al.* (2014). Dans leur article, des sous-produits artificiels apparaissaient dès le 5<sup>ème</sup> cycle SELEX. Ceux-ci avaient tendance à monter en hauteur à travers le gel, similairement à la **Figure 31** dans la section « mauvais gel ». Exactement comme c'était le cas présent, même après des extractions de gels, ces artefacts revenaient sans cesse (Tolle *et al.*, 2014). Fait très important à noter est que ce groupe de recherche ont noté une perte d'enrichissement complète de leur librairie à travers les

cycles SELEX dû à ces artefacts (Tolle *et al.*, 2014), ce qui expliquerait parfaitement la raison pour laquelle les tests de liaison n'ont pas ressorti d'aptamère. L'explication fournie concernant la formation de ces sous-produits est un appariement partiel intermoléculaire des régions aléatoires et des régions d'amorces des librairies (Kohlberger et Gadermaier, 2021; Tolle *et al.*, 2014). Il est possible que ceci soit encore plus problématique pour des librairies ayant une très grande région aléatoire ( $N_{200}$  et  $N_{800}$ ), puisque la région aléatoire représente une proportion beaucoup plus importante comparativement aux régions d'amorces. Ceci causerait un allongement graduel des librairies, comportant chacune plusieurs répétitions de petites sections d'amorces (Tolle *et al.*, 2014). Malheureusement, aucune solution n'existe en ce moment pour complètement contrer ce phénomène. Par contre, plusieurs solutions ont été amenées qui semblent réduire l'apparition de ces répétitions. La plus efficace semble être d'employer plusieurs sites d'amorce et différentes combinaisons d'amorces lors des cycles SELEX (Tolle *et al.*, 2014).

Pour ajouter à cet article, puisque les problèmes semblent survenir à cause d'appariement complémentaire de quelques nucléotides (<10), peut-être serait-il judicieux d'employer des stratégies empêchant le plus possible les courts appariements. Deux articles traitant de l'amplification de séquence riche en GC offrent des pistes intéressantes pour les matrices difficiles à bien amplifier (Green et Sambrook, 2019; Hubé *et al.*, 2005). Le premier article combine trois méthodes afin d'empêcher le plus possible les appariements partiels et les importantes structures secondaires des matrices (Hubé *et al.*, 2005). Ces trois méthodes sont: l'ajout de bêtaïne (agent dénaturant *GC-enhancer*), un programme PCR par *touchdown* (5-10 premiers cycles à appariement à plus haute température), et la *Taq Hot Start* (Hubé *et al.*, 2005). Étonnamment, chacune des trois méthodes individuellement n'ont pas réussi à amplifier des matrices avec de fortes structures secondaires, mais les trois combinées menaient à des bandes bien compactes et denses (Hubé *et al.*, 2005; Kohlberger et Gadermaier, 2021; Shore et Paul, 2010). Le deuxième article suggérait d'employer un programme PCR *slowdown*, qui consiste à faire varier les températures lentement à 2°C/seconde (Green et Sambrook, 2019). En plus de la PCR modifiée, le composé deaza-dGTP (remplaçant le dGTP dans le dNTP) est aussi employé dans la réaction PCR (Green et Sambrook, 2019). Ce dGTP modifié manque un groupement d'azote à la position 7 de l'anneau de la purine (Shore et Paul, 2010). Cette modification aide l'amplification de matrice diluée ou de mauvaise qualité en diminuant les liaisons de type Hoogsteen (mauvais angle de liaison) sans interférer avec les appariements de bases Watson-Crick normaux (Green et Sambrook, 2019; Jung *et al.*, 2002; Shore et Paul, 2010).

De possibles solutions diminuant la formation de sous-produits non-désirés seraient de faire 10 à 20 cycles PCR (Kohlberger et Gadermaier, 2021) plutôt que 35, et potentiellement d'employer un programme PCR à deux étapes (pas d'étape d'appariement, le tout fait à 95°C et 72°C). Une autre

suggestion est de faire des PCR à émulsion (Kohlberger et Gadermaier, 2021). Le principe est de confiner les réactifs de la PCR dans des micro-gouttelettes d'huiles afin de limiter les interactions intermoléculaires entre différents amplicons (Kohlberger et Gadermaier, 2021). Le but de ces suggestions serait de limiter les fortes structures supérieures des librairies complexes et de fortement diminuer les courts appariements de quelques nucléotides (*mispriming*).

Un réflexe face à ces artefacts, très délétères pour l'issue du SELEX, serait de commencer à extraire du gel les bandes à longueurs attendues, mais ceci ne semble pas prévenir la réapparition de ces mêmes artefacts (Tolle *et al.*, 2014). En effet, même en purifiant à chaque cycle, certains groupes ont tout de même observé une perte totale d'enrichissement de librairie lors de SELEX (Kohlberger et Gadermaier, 2021; Tolle *et al.*, 2014). Ceci souligne l'importance de bien optimiser la PCR avant même le départ des cycles SELEX.

L'enzyme piPolB (*primer-independent polymerase B*) récemment décrite par Redrejo-Rodríguez *et al.* (2017) offre aussi un bon potentiel pour les amplifications SELEX. Cette polymérase a une activité de polymérisation complémentaire tout comme les polymérases Taq communes, mais elle ne requiert pas d'amorce comme point de départ (Redrejo-Rodríguez *et al.*, 2017). Elle a été décrite comme étant très processive, avec une activité de déplacement de brin, et capable de synthétiser des fragments ayant plusieurs milliers de pb (Redrejo-Rodríguez *et al.*, 2017). Pouvoir amplifier de l'ADN de façon efficace sans avoir recours à des amorces pourrait présenter de nombreux avantages dans l'optique d'un SELEX, notamment la réduction de mésappariement ou dimères issus d'amorces. Par contre, seuls quelques essais enzymatiques ont été faits avec l'enzyme et jamais dans un contexte de longs fragments (Redrejo-Rodríguez *et al.*, 2017). Il serait intéressant de garder un œil sur les futures publications sur cette enzyme, ou faire quelques tests afin de vérifier sa limite. Un aspect important à tester est de voir si elle a tendance à s'accrocher un peu partout sur la matrice (bande diffuse) ou si elle commence sa polymérisation à partir des extrémités (bande compacte).

#### **4.2.5 Erreur humaine**

Il faut aussi garder en tête qu'une erreur de manipulation peut avoir d'énormes implications dans le SELEX. Dans le cas présent, trois ligands à cinq librairies chaque (15 différents tubes, à 15 cycles SELEX, à 3-5 fractions chaque) révèlent une grande quantité de tubes à gérer. Évidemment, il est possible de prendre des mesures pour limiter au maximum les erreurs d'inattention (exemple: travailler sur différents supports à tube et s'organiser logiquement dans l'espace), mais il est facile de s'imaginer qu'on peut

mettre un échantillon dans un mauvais tube (d'un autre ligand par exemple) et compromettre le SELEX à mi-chemin. Ce genre d'expérience se poursuit sur plusieurs mois et peut seulement être validé à la toute fin. Or, travailler avec une grande quantité de tubes est un risque à prendre en compte; même un excellent technicien de laboratoire commettra occasionnellement des erreurs d'inattention en laboratoire. L'objectif de travailler avec plusieurs ligands et librairies était d'augmenter le plus possible la chance de succès et de permettre des comparaisons de l'efficacité de plusieurs librairies, mais peut-être aurait-il été plus judicieux d'avoir moins d'échantillons. Ceci diminue la chance d'erreurs et la durée des manipulations, ce qui permet conséquemment de faire plus d'expériences et d'optimisations sur ces quelques échantillons, poussant l'analyse au maximum.

### 4.3 Tests de liaison

Il est toujours préférable d'employer plusieurs méthodes lors d'une validation. Dans le cas de tests de liaison, toucher plusieurs techniques empêche de conclure à tort que les candidats sont (ou pas) des aptamères. Dans le cas présent, les quatre techniques employées ne démontrent aucune interaction, ou des interactions non-spécifiques, entre les 12 candidats testés et leurs cibles. S'il n'était question que d'un seul test à résultats négatifs, il serait justifiable d'en douter. Il est difficile de distinguer si un résultat négatif est dû à une méthode non-applicable ou à une réelle absence d'interaction. Cependant, la combinaison des quatre tests pointe vers une absence d'interaction pour les candidats testés, ce qui ne permet pas de les qualifier comme étant des aptamères. En rétrospective, un témoin positif fiable (ayant aussi une cible protéique) aurait été fort utile à travers les tests. Par exemple, employer un aptamère bien connu ayant une protéine peu coûteuse comme cible. Un tel témoin en main permet de rapidement identifier les tests applicables aux candidats à tester.

Tester plus de candidats serait aussi envisageable plutôt que les quelques premiers pour chaque cible. La charge de travail augmenterait sans garanti de succès, mais seul le nombre de lectures a été prise en compte lors de l'analyse bioinformatique. Il faut aussi garder en tête que le SELEX est une méthode qui parfois, pour des raisons entièrement inconnues, ne fonctionne simplement pas (Kohlberger et Gadermaier, 2021; McKeague *et al.*, 2015; Tolle *et al.*, 2014). De base, le principe d'un SELEX est très aléatoire et chaotique, et certaines difficultés ou artefacts peuvent prendre le dessus et complètement noyer la présence des aptamères contenus dans la librairie initiale (McKeague *et al.*, 2015; Tolle *et al.*, 2014). Le facteur hasard joue ainsi un grand rôle, et ceci souligne l'importance de bien optimiser chaque étape du SELEX, avec une attention particulière à la PCR.

#### 4.4 Conclusion

Pour conclure, il n'a malheureusement pas été possible de comparer de longs aptamères à des courts. Il est fortement soupçonné que ceci est dû à une mauvaise optimisation des cycles SELEX, plus particulièrement l'amplification PCR. Il n'est pas rare pour les PCR successives de SELEX de créer certains artefacts menant à une perte totale d'enrichissement de librairie. Plusieurs stratégies sont requises afin de limiter le plus possible ce phénomène. Ceci étant dit, plusieurs intéressantes techniques ont été explorées et ont mené à la création de longues librairies amplifiables par PCR. La ligation de fragment simple-brin ne semble pas très efficace pour ériger une longue molécule, mais la terminale transférase possède un bon potentiel pour la création *de novo* de longs oligonucléotides dégénérés. Cette nouvelle technique ajoute un outil de plus à l'arsenal de la biologie moléculaire. Même si aucun aptamère n'a été découvert, un protocole modifié pourrait mener à des SELEX plus fructueux et ainsi ouvrir l'horizon à plus d'applications.

## 5 BIBLIOGRAPHIE

---

Acquah, C., Agyei, D., Monney, I., Pan, S. and Danquah, M.K., 2018. Aptameric sensing in food safety. *Food Control and Biosecurity*, pp.259-277.

Ashley, J., Schaap-Johansen, A.L., Mohammadniaei, M., Naseri, M., Marcatili, P., Prado, M. and Sun, Y., 2021. Terminal deoxynucleotidyl transferase-mediated formation of protein binding polynucleotides. *Nucleic Acids Research*, 49(2), pp.1065-1074.

Barthel, S., Palluk, S., Hillson, N.J., Keasling, J.D. and Arlow, D.H., 2020. Enhancing terminal deoxynucleotidyl transferase activity on substrates with 3' terminal structures for enzymatic *de novo* DNA synthesis. *Genes*, 11(1), p.102.

Blind, M. and Blank, M., 2015. Aptamer selection technology and recent advances. *Molecular Therapy-Nucleic Acids*, 4, p.e223.

Chen, C., Zhou, S., Cai, Y. and Tang, F., 2017. Nucleic acid aptamer application in diagnosis and therapy of colorectal cancer based on cell-SELEX technology. *NPJ Precision Oncology*, 1(1), pp.1-7.

Christopoulos, T.K. and Diamandis, E.P., 1996. Fluorescence immunoassays. *Immunoassay*, pp.309-335.

Chua, J.P.S., Go, M.K., Osothprarop, T., Mcdonald, S., Karabadzhak, A.G., Yew, W.S., Peisajovich, S. and Nirantar, S., 2020. Evolving a thermostable terminal deoxynucleotidyl transferase. *ACS Synthetic Biology*, 9(7), pp.1725-1735.

Clarke, L.A., Rebelo, C.S., Goncalves, J., Boavida, M.G. and Jordan, P., 2001. PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. *Molecular Pathology*, 54(5), p.351.

Czar, M.J., Anderson, J.C., Bader, J.S. and Peccoud, J., 2009. Gene synthesis demystified. *Trends in Biotechnology*, 27(2), pp.63-72.

Ellington, A.D. and Szostak, J.W., 1990. *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, 346(6287), pp.818-822.

Eun, H.M., 1996. Enzymology primer for recombinant DNA technology. *Elsevier*.

Fowler, J.D. and Suo, Z., 2006. Biochemical, structural, and physiological characterization of terminal deoxynucleotidyl transferase. *Chemical Reviews*, 106(6), pp.2092-2110.

Fujishima, K., Venter, C., Wang, K., Ferreira, R. and Rothschild, L.J., 2015. An overhang-based DNA block shuffling method for creating a customized random library. *Scientific Reports*, 5(1), pp.1-5.

Gdoura-Ben Amor, M., Jan, S., Baron, F., Grosset, N., Culot, A., Gdoura, R., Gautier, M. and Techer, C., 2019. Toxigenic potential and antimicrobial susceptibility of *Bacillus cereus* group bacteria isolated from Tunisian foodstuffs. *BMC Microbiology*, 19(1), pp.1-12.

Gragoudas, E.S., Adamis, A.P., Cunningham Jr, E.T., Feinsod, M. and Guyer, D.R., 2004. Pegaptanib for neovascular age-related macular degeneration. *New England Journal of Medicine*, 351(27), pp.2805-2816.

Green, M.R. and Sambrook, J., 2019. Polymerase chain reaction (PCR) amplification of GC-rich templates. *Cold Spring Harbor Protocols*, 2019(2), pp.pdb-prot095141.

Hall, B., Micheletti, J.M., Satya, P., Ogle, K., Pollard, J. and Ellington, A.D., 2009. Design, synthesis, and amplification of DNA pools for *in vitro* selection. *Current Protocols in Nucleic Acid Chemistry*, 39(1), pp.9-2.

Han, J., Gao, L., Wang, J. and Wang, J., 2020. Application and development of aptamer in cancer: from clinical diagnosis to cancer therapy. *Journal of Cancer*, 11(23), p.6902.

Ho, C.K., Wang, L.K., Lima, C.D. and Shuman, S., 2004. Structure and mechanism of RNA ligase. *Structure*, 12(2), pp.327-339.

Hoinka, J., Backofen, R. and Przytycka, T.M., 2018. AptaSUITE: a full-featured bioinformatics framework for the comprehensive analysis of aptamers from HT-SELEX experiments. *Molecular Therapy-Nucleic Acids*, 11, pp.515-517.

Hube, F., Reverdiau, P., lochmann, S. and Gruel, Y., 2005. Improved PCR method for amplification of GC-rich DNA sequences. *Molecular Biotechnology*, 31(1), pp.81-84.

Hughes, R.A. and Ellington, A.D., 2017. Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harbor Perspectives in Biology*, 9(1), p.a023812.

Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B., 1994. High-resolution molecular discrimination by RNA. *Science*, 263(5152), pp.1425-1429.

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P. and Breitsprecher, D., 2014. Microscale thermophoresis: interaction analysis and beyond. *Journal of Molecular Structure*, 1077, pp.101-113.

Johne, R., Müller, H., Rector, A., Van Ranst, M. and Stevens, H., 2009. Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends in Microbiology*, 17(5), pp.205-211.

Jung, A., Ruckert, S., Frank, P., Brabletz, T. and Kirchner, T., 2002. 7-Deaza-2-deoxyguanosine allows PCR and sequencing reactions from CpG islands. *Molecular Pathology*, 55(1), p.55.

Kim, H.S., Kim, Y.J., Chon, J.W., Kim, D.H., Yim, J.H., Kim, H. and Seo, K.H., 2017. Two-stage label-free aptasensing platform for rapid detection of *Cronobacter sakazakii* in powdered infant formula. *Sensors and Actuators B: Chemical*, 239, pp.94-99.

Klykov, O. and Weller, M.G., 2015. Quantification of N-hydroxysuccinimide and N-hydroxysulfosuccinimide by hydrophilic interaction chromatography (HILIC). *Analytical Methods*, 7(15), pp.6443-6448.

Kohlberger, M. and Gadermaier, G., 2021. SELEX: critical factors and optimization strategies for successful aptamer selection. *Biotechnology and Applied Biochemistry*.

Komarova, N. and Kuznetsov, A., 2019. Inside the black box: what makes SELEX better?. *Molecules*, 24(19), p.3598.

Kosuri, S. and Church, G.M., 2014. Large-scale *de novo* DNA synthesis: technologies and applications. *Nature Methods*, 11(5), pp.499-507.

Kuhn, H. and Frank-Kamenetskii, M.D., 2005. Template-independent ligation of single-stranded DNA by T4 DNA ligase. *The FEBS Journal*, 272(23), pp.5991-6000.

Marimuthu, C., Tang, T.H., Tominaga, J., Tan, S.C. and Gopinath, S.C., 2012. Single-stranded DNA (ssDNA) production in DNA aptamer generation. *Analyst*, 137(6), pp.1307-1315.

McKeague, M., McConnell, E.M., Cruz-Toledo, J., Bernard, E.D., Pach, A., Mastronardi, E., Zhang, X., Beking, M., Francis, T., Giamberardino, A. and Cabecinha, A., 2015. Analysis of *in vitro* aptamer selection parameters. *Journal of Molecular Evolution*, 81(5), pp.150-161.

Michelson, A.M. and Todd, A.R., 1955. Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3': 5'-internucleotidic linkage. *Journal of the Chemical Society (Resumed)*, pp.2632-2638.

Miura, F., Shibata, Y., Miura, M., Sangatsuda, Y., Hisano, O., Araki, H. and Ito, T., 2019. Highly efficient single-stranded DNA ligation technique improves low-input whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Research*, 47(15), pp.e85-e85.

Nehdi, A., Samman, N., Aguilar-Sánchez, V., Farah, A., Yurdusev, E., Boudjelal, M. and Perreault, J., 2020. Novel strategies to optimize the amplification of single-stranded DNA. *Frontiers in Bioengineering and Biotechnology*, p.401.

Ni, S., Zhuo, Z., Pan, Y., Yu, Y., Li, F., Liu, J., Wang, L., Wu, X., Li, D., Wan, Y. and Zhang, L., 2020. Recent progress in aptamer discoveries and modifications for therapeutic applications. *ACS Applied Materials & Interfaces*, 13(8), pp.9500-9519.

Ó'Fágáin, C., Cummins, P.M. and O'Connor, B.F., 2011. Gel-filtration chromatography. *Protein Chromatography*, pp.25-33.

Petkovic, S. and Müller, S., 2015. RNA circularization strategies *in vivo* and *in vitro*. *Nucleic Acids Research*, 43(4), pp.2454-2465.

Pham, N.H., Létourneau, M., Fortier, M., Bégin, G., Al-Abdul-Wahid, M.S., Pucci, F., Folch, B., Rooman, M., Chatenet, D., St-Pierre, Y. and Lagüe, P., 2021. Perturbing dimer interactions and allosteric communication modulates the immunosuppressive activity of human galectin-7. *Journal of Biological Chemistry*, 297(5).

Pollard, T.D., 2010. A guide to simple and informative binding assays. *Molecular Biology of the Cell*, 21(23), pp.4061-4067.

Qi, S., Duan, N., Khan, I.M., Dong, X., Zhang, Y., Wu, S. and Wang, Z., 2022. Strategies to manipulate the performance of aptamers in SELEX, post-SELEX and microenvironment. *Biotechnology Advances*, p.107902.

Redrejo-Rodríguez, M., Ordóñez, C.D., Berjón-Otero, M., Moreno-González, J., Aparicio-Maldonado, C., Forterre, P., Salas, M. and Krupovic, M., 2017. Primer-independent DNA synthesis by a family B DNA polymerase from self-replicating mobile genetic elements. *Cell Reports*, 21(6), pp.1574-1587.

Rittié, L. and Perbal, B., 2008. Enzymes used in molecular biology: a useful guide. *Journal of Cell Communication and Signaling*, 2(1), pp.25-45.

Sabeti, P.C., Unrau, P.J. and Bartel, D.P., 1997. Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chemistry & Biology*, 4(10), pp.767-774.

Sarac, I. and Hollenstein, M., 2019. Terminal deoxynucleotidyl transferase in the synthesis and modification of nucleic acids. *ChemBioChem*, 20(7), pp.860-871.

Shen, L., Bing, T., Zhang, N., Wang, L., Wang, J., Liu, X. and Shangguan, D., 2019. A nucleus-targeting DNA aptamer for dead cell indication. *ACS Sensors*, 4(6), pp.1612-1618.

Shore, S. and Paul, N., 2010. Robust PCR amplification of GC-rich targets with Hot Start 7-deaza-dGTP. *Biotechniques*, 49(5), pp.841-843.

Shuman, S., 2009. DNA ligases: progress and prospects. *Journal of Biological Chemistry*, 284(26), pp.17365-17369.

Silber, R., Malathi, V.G. and Hurwitz, J., 1972. Purification and properties of bacteriophage T4-induced RNA ligase. *Proceedings of the National Academy of Sciences*, 69(10), pp.3009-3013.

Silverman, S.K., 2016. Catalytic DNA: scope, applications, and biochemistry of deoxyribozymes. *Trends in Biochemical Sciences*, 41(7), pp.595-609.

Sriprakash, K.S., Lundh, N., Huh, M.O. and Radding, C.M., 1975. The specificity of lambda exonuclease. Interactions with single-stranded DNA. *Journal of Biological Chemistry*, 250(14), pp.5438-5445.

Stoltenburg, R., Reinemann, C. and Strehlitz, B., 2007. SELEX—a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering*, 24(4), pp.381-403.

Sun, H. and Zu, Y., 2015. A highlight of recent advances in aptamer technology and its application. *Molecules*, 20(7), pp.11959-11980.

Sundaram, P., Kurniawan, H., Byrne, M.E. and Wower, J., 2013. Therapeutic RNA aptamers in clinical trials. *European Journal of Pharmaceutical Sciences*, 48(1-2), pp.259-271.

Svobodová, M., Pinto, A., Nadal, P.O.S.C.K. and O'Sullivan, C.K., 2012. Comparison of different methods for generation of single-stranded DNA for SELEX processes. *Analytical and Bioanalytical Chemistry*, 404(3), pp.835-842.

Tang, L., Navarro Jr, L.A., Chilkoti, A. and Zauscher, S., 2017. High-molecular-weight polynucleotides by transferase-catalyzed living chain-growth polycondensation. *Angewandte Chemie*, 129(24), pp.6882-6886.

Tolle, F., Wilke, J., Wengel, J. and Mayer, G., 2014. By-product formation in repetitive PCR amplification of DNA libraries during SELEX. *PLOS One*, 9(12), p.e114693.

Tuerk, C. and Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249(4968), pp.505-510.

Velez, T.E., Singh, J., Xiao, Y., Allen, E.C., Wong, O.Y., Chandra, M., Kwon, S.C. and Silverman, S.K., 2012. Systematic evaluation of the dependence of deoxyribozyme catalysis on random region length. *ACS Combinatorial Science*, 14(12), pp.680-687.

Wang, L.K., Schwer, B. and Shuman, S., 2006. Structure-guided mutational analysis of T4 RNA ligase 1. *RNA*, 12(12), pp.2126-2134.

Wang, T., Chen, C., Larcher, L.M., Barrero, R.A. and Veedu, R.N., 2019. Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development. *Biotechnology Advances*, 37(1), pp.28-50.

Weerathunge, P., Behera, B.K., Zihara, S., Singh, M., Prasad, S.N., Hashmi, S., Mariathomas, P.R.D., Bansal, V. and Ramanathan, R., 2019. Dynamic interactions between peroxidase-mimic silver nanozymes and chlorpyrifos-specific aptamers enable highly-specific pesticide sensing in river water. *Analytica Chimica Acta*, 1083, pp.157-165.

Wen, Y., Marcus, E.A., Matrubutham, U., Gleeson, M.A., Scott, D.R. and Sachs, G., 2003. Acid-adaptive genes of *Helicobacter pylori*. *Infection and Immunity*, 71(10), pp.5921-5939.

Wu, Y.X. and Kwon, Y.J., 2016. Aptamers: The “evolution” of SELEX. *Methods*, 106, pp.21-28.

Xu, Y., Phillips, J.A., Yan, J., Li, Q., Fan, Z.H. and Tan, W., 2009. Aptamer-based microfluidic device for enrichment, sorting, and detection of multiple cancer cells. *Analytical Chemistry*, 81(17), pp.7436-7442.

Zhang, Y., Lai, B.S. and Juhas, M., 2019. Recent advances in aptamer discovery and applications. *Molecules*, 24(5), p.941.

Zhou, J. and Rossi, J., 2017. Aptamers as targeted therapeutics: current potential and challenges. *Nature Reviews Drug Discovery*, 16(3), pp.181-202.

Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y. and Robb, G.B., 2012. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Research*, 40(7), pp.e54-e54.

Zhuo, Z., Yu, Y., Wang, M., Li, J., Zhang, Z., Liu, J., Wu, X., Lu, A., Zhang, G. and Zhang, B., 2017. Recent advances in SELEX technology and aptamer applications in biomedicine. *International Journal of Molecular Sciences*, 18(10), p.2142.

Zon, G., 2020. Recent advances in aptamer applications for analytical biochemistry. *Analytical Biochemistry*, p.113894.



## 7 ANNEXE II

---

### Protocole de l'électroélution

- Laver la cuve avec de l'eau du robinet.
- Remplir la cuve jusqu'au rebord du plateau central (sans déborder) avec du TBE 0,25X.
- Souffler les bulles dans les tuyaux avec une p1000.
- Migrer l'appareil à 200 V pendant 30 minutes.
- Jeter le tampon, remplir la cuve avec du nouveau tampon TBE 0,25X.
- Souffler les bulles dans les tuyaux avec une p1000.
- Ajouter 200 µL d'ammonium acétate 8 M dans les tuyaux (en forme de « V ») avec une p200.
- Placer les bouts de gel dans les puits en les collant bien au fond (ne doivent pas flotter).
- Migrer l'appareil à 200 V pendant 1 heure.
- Boucher la sortie des tuyaux avec des embouts de pipette p200.
- Enlever le tampon 0,25X à l'aide d'une seringue sans aiguille.
- En utilisant des embouts de chargement effilés, transférer l'ammonium acétate des tuyaux dans un tube eppendorf (volume de ~300 µL contenant l'ADN élué).
- Enlever les embouts p200 des tuyaux de la cuve, et rincer la cuve avec de l'eau du robinet.
- Procéder à une précipitation d'éthanol sans ajouter de l'ammonium acétate (ajouté l'éthanol 100% et précipité directement).

Information sur la cuve: compagnie: IBI (*international biotechnologies, inc.*), modèle: *unidirectional electroelutor (analytical)*, #catalog: 46000

## 8 ANNEXE III

---

### Protocole de la précipitation à l'éthanol

- Extraire la bande du gel et placer dans un tube eppendorf.
- Ajouter 400 µL de tampon d'éluion (0.3 M NaCl dans eau Milli-Q).
- Incuber 1 heure (ou *overnight*) sur un rotor à 21°C (ou 4°C pour ARN).
- Enlever le morceau de gel du tube.
- Ajouter 1/10 volume (40 µL) d'ammonium acétate 3 M.
- Ajouter 2,5 volumes (1000 µL) éthanol 100%, vortexer le tube.
- Incuber à -80°C pendant 1 heure, ou à -20°C pendant 2 heures.
- Centrifuger à force maximale (15000 rpm) pendant 30 minutes.
- Retirer le surnageant (en versant ou avec une pipette).
- Ajouter 500 µL d'éthanol 70%, brasser par aspiration pour déloger le culot (parfois invisible).
- Centrifuger à force maximale (15000 rpm) pendant 10 minutes à 4°C (si possible).
- Retirer le surnageant (en versant ou avec une pipette).
- Sécher le culot à l'air libre afin d'enlever toute trace d'éthanol (placer le tube à l'envers).
- Resuspendre le culot dans un volume désirer d'eau Milli-Q (généralement 10-30 µL).







## 12 ANNEXE VII

Tableau 10: Séquence complète de trois lectures du séquençage Nanopore de la librairie initiale N<sub>800</sub>.

séquence	
1	<p>TATACATGGAGTTTCAGTTTACACTGCCAGGTGTGCTGCTGATGAACCTCTGCAGCACCTACACATTACAAGGACCACCGCCAGGCACCGGCGCAGGAGCGAAAACCTCGAAAGGCAACGGAGAGAGACACGTAGAAGGAATGGGGAGGAAAGTGAAGTTATGCCCGGGAGGGGATAGCTGGCACGGCCTGGCGGGGATGGAAAGGGGCATGCACAAGATAAATGGAAAGGAGCTGGGAAAGAACGGAGGTGAGATGCTGAAGGCATTACAAGGAAAGTCGGGGATTGGCGCGGTGGATAAAGGGTTAGTGGAGACAAGGCCGGGAATGAGATAGTGAAATGAGAGGTAGGTGTCCACGAAAAGGAAAGGGGAGACAGATTGCAAAAAAGGGAATATTGATGTGTACAGCAGAACTGATAAAACGGCAGAAGAGCACGTCCAGGGCCGAGAAGGGGAACTCAAAGTTGCAGTATGGGACCGGAGAGGAATATAGTGA</p> <p>GCGCCGACGCCAGAGCGTGCAGCCAAAGGCACCATTGAGCCAGAGAAAGTATCAAACACTGGACGGCAGCCGTCGAAAGAGCGCAATGTGCGAGCGCAGATGGTAACTGACACAAGAAGTCAAGAACCCGCGTGTGATCAATTAAGCTATTGGCGAGGGCGTGTACTACACGCGCAGAGTGGGTAAACAGGGACAGGGATTAGCGATCGAGCGACTGTAGAATCAGGACTCGACGTGCAGGCTTAAACACGAGCAACAACAAGTATGGTGGGTAAGGACTAGGGCGGTAAAGCGCGCGGACGATAAAGCCGGTGGAAAGGCGCCGCGCAAAAGTGCAAACGATGGATGTGCTGATTGGGTGCGGAGGTTCACTGGGTATGCGTTAACCTTAGCAATACGTAG</p>
2	<p>GCACCTTCGTTCAATTACATGGTACTTCTGGAAGTTAACTTACTACCTCGAACCTCATAACTACATTACAAGGACCACCGCCGGATGGCAAGCTGGAGAA</p> <p>TACCAGGCAGGGAGGATAGCATGGAAAGGGCAAAACAGTACGACAGACGCCCTGAATGAGAAGGTAAAGCAGTGAAGTGGGGGAGAGAAAAAGAAAAGATATCCCGGAAACGGAAGTGACAGCTAGCTGCAGCGTGGCCAGAGATAGCCGCAAAACAGTGTGCGAAGGGGGTATCCAATCAACTGGAAAGAACAGCGGAAGTCCGAGTAAGTAAGTCTGAGAGAGAATAGTACAAGAAGAAAGCCGCAACACAGGGGAAAGAAATTAAGAGGAGGACGTTTCTGGCAACCGGAAACGTTAGGGAAACTTAGAATGAAGAGGAGGAAATGCACCCATGAGAATAAGAGGCTCGGGAGTCGATAGAGCGGGAAACGAGTGGGAAGAAAGATAAAAAAGAGAGCAAGGGGCGGGTAGGAAGAGTCAAGGAAAGCCGGCAGGGAGAGGAAACAGAAAGCGCGATGCGCGTCGCGGGTAGCTGCAGTAAGGCATAGATGCATTCAGATGAGAGAGCTGGAAGTGCAGGCGACAGAATGGCCAAAAAGGGCTCAGAAAAGTGGACTCGGCATAGGTGAAGAAATGAATAGGAGTTGTGAGCACCGCATACGGGACGGACATGCCCGCGCGCCGGGTAGCGGCGGGCAGGTTGGGGAAGACCTTGATACCGCCGCGAGCCGCGCGCGATTGAAATATAACGTAGCCCTTCCGATCAACCTCCGAGCGGCAAGACGAGAAGAGCAGAATGTGATTGCTCGAACTGCTCCCGGTAACGTTTACCCGAAAT</p> <p>TGACATGGCCGTCAACATGGGAAATATAATCGTATAGAAGCAAGTATAGAGGGCAGGTTATTGTCGATGTTATCGTAATTAATCAGGAAAAAGAGATTGAGTTGAATAACGGTAGGGATTAATAACAAAAAAGAACCGGATAACGAATGCACCCAAACGATGGATGTGCTGATTACAGGTGCTGCGAGGGTTCGCTGGGTATTAAGTTAACCTTAGCAATACGTAAC</p>
3	<p>GTATGCTTCGTTTCAGTTAGCACCTACTTACTACCCAGTGAGGGCGCCTCCTCGCAGCACCTACATTACAAGGACCACCGCCCAATATCAAAGATTAACATAGAAGAACAAGAGGTCAGACAGGGACTGGAAATGGGCCAGAGAAGTAAAAAGACACAAGCACTGGGAAAGAACACGGAAAGCCATTGGAAGAGGCAGCGAGGAGCGAGGATGAAAGCGCCCGGAAATAAGCGACGACCGCACCCGAGCTAAAATCCTAGTAAGCGACGAAATGTAGCAAGCAGCGGAGAGAGGGGAAAGGAAGGAGAAGGAAAGGAAAAAGCAATGTGGGAGAAGATGGAGAGTAGGGAAGTAGATCCCGAGGGAGGCCCCAAAGAGCGCGCGGCCGGGGAGAGA</p> <p>AAGAAGCAGTCCAAGCCAGGAGAAGAGCCGGCCGATAAGACGGGCGACGTCCTGACTGTGGATCGGCTGGAAGGGGAGAGATGCACGGGGGCGCCGGGGAGATGACCCGACGTTGACAGATTAGAAATGAAAGGATACCGGCCCGCGGGTGGTAAGGGAGAAGTTTATCGGGAGAAGCGACGTTAGATCACGTGGAGGA</p> <p>AATCTTGTGGGAACGAGGGAATACCCAGGCATTGAAAGAAAACGGATGTGCAAGCAAGAGGATGGTCAAGGCAACGAGAGTGGAAACGACGACGCGAGCCGATAGCAGCGTGCATTTGGAAACAAAAAAGAATCCCTGATGCGTGGATTCGGATGAGAAATGGGACGACCTGCGTATGCGAGGTTAAAGGGAGCCGAGTTAAAGCATGCCGAGAGAGGGAACTTATATCTAAGAGAAAGTAGAGAGGGGAAAGGCTCAGTAACGCAGAGAAGGCCACCCAGACGTTCCGCTAGACTACAGCCCCGCCCCGAGAAATGACCCGCGTGAACGTTGGGAAATAGGGAAACGGATGGATGGATAAAGAGGCACCTGCACGAAGGAGAGAAGAGGGCGCAAACGATGGATGTGCTGATTACGAGGTGCTGCGAGGGGAGGTTCACTGGAGTACAGTAAGTTAACCTTAGCAATACGTAACCT</p>

Ci-haut se retrouve la séquence de trois lectures des 20 présentés au Tableau 5. Ces trois lectures sont issues du fichier de sortie Nanopore de la librairie initiale N<sub>800</sub> avant le SELEX. Les régions rouges correspondent à l'amorce sens, roses à l'amorce antisens, et grises aux adaptateurs Nanopore lors du séquençage.