

Université de Sherbrooke

**Étude de protéines liant les G-quadruplexes situés en 5'UTR de gènes impliqués dans
la maladie de Parkinson**

Par
Marc-Antoine Turcotte
Programmes de Biochimie

Mémoire présenté à la Faculté de médecine et des sciences de la santé
en vue de l'obtention du grade de maître ès sciences (M. Sc.)
en Biochimie

Sherbrooke, Québec, Canada
Août 2020

Membres du jury d'évaluation
Jean-Pierre Perreault, Département de biochimie et génomique fonctionnelle
Éric Massé, Département de biochimie et génomique fonctionnelle
François-Michel Boisvert, Département d'immunologie et de biologie cellulaire

Dédicace à Gilles, Julie et Vincent

Pas de pression, pas de diamants
Thomas Carlyle

RÉSUMÉ

Étude de protéines liant les G-quadruplexes situés en 5'UTR de gènes impliqués dans la maladie de Parkinson

Par

Marc-Antoine Turcotte
Programme de Biochimie

Mémoire présenté à la Faculté de médecine et des sciences de la santé en vue de l'obtention du diplôme de maître ès sciences (M.Sc.) en Biochimie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec, Canada, J1H 5N4

Au sein du système nerveux, les mécanismes de régulation post-transcriptionnelle contribuent grandement aux différents phénotypes cellulaires. En effet, de faibles dérégulations peuvent mener à des maladies. Les G-quadruplexes (G4) d'ARN émergent comme étant de nouveaux acteurs dans la régulation post-transcriptionnelle. Les G4 sont des structures secondaires très stables retrouvées dans l'ADN et l'ARN. Ils sont formés par l'empilement d'un minimum de deux tétrades de guanines et stabilisés par des cations monovalents, habituellement le potassium. Au niveau de l'ARN, ils sont impliqués dans différents mécanismes comme l'épissage alternatif, la polyadénylation alternative, l'activation et l'inhibition de la traduction et la maturation des microARN. Selon plusieurs, les G4 sont considérés comme étant des événements plutôt que des entités. Des éléments *trans*, soit des protéines, sont alors en mesure d'assurer leur repliement et dépliement.

Une analyse bio-informatique a confirmé un enrichissement de G4 d'ARN dans les transcrits du système nerveux. Cette analyse a aussi permis d'identifier la maladie de Parkinson comme l'une des maladies ayant la plus grande proportion de potentiel G4 (pG4). En effet, sur 16 gènes reliés à la maladie, 15 d'entre eux contiennent au moins un pG4. Ensuite, différentes techniques biochimiques comme des essais de fluorescence, le dichroïsme circulaire et la cartographie *in-line* ont été utilisées pour confirmer la formation de certains G4 situés dans des régions 5' non traduites (5'UTR). Ces techniques ont permis d'identifier 4 nouveaux G4 d'ARN n'ayant jamais été caractérisés auparavant. Par la suite, des essais luciférase ont conclu que deux de ces G4, chez les ARNm de VPS35 et PRKN, ont un effet répresseur sur la traduction. Les protéines pouvant lier ces deux G4 ont ensuite été analysées par spectrométrie de masse. La protéine Guanine Nucleotide-Binding Protein-Like 1 (GNL1) a alors été identifiée comme étant une protéine pouvant lier les deux G4. Finalement, l'interaction entre GNL1 et le G4 de VPS35 et PRKN a été confirmée par des essais de retardement sur gel.

En bref, cette étude a permis d'identifier une nouvelle protéine, GNL1, pouvant lier deux nouveaux G-quadruplexes se situant dans le 5'UTR de gènes dérégulés dans la maladie de Parkinson. Une régulation de l'interaction entre GNL1 et les G4 pourrait être une option thérapeutique pour contrer les dérégulations présentes dans la maladie de Parkinson.

Mots clés : G-quadruplexe, ARN, Parkinson, Protéine liant l'ARN

SUMMARY

Study of the proteins able to bind G-quadruplexes situated in 5'UTR of genes implicated in Parkinson's disease

By
Marc-Antoine Turcotte
Biochemistry Program

Thesis presented at the Faculty of medicine and health sciences for the obtention of Master degree diploma maitre ès sciences (M.Sc.) in Biochemistry, Faculty of medicine and health sciences, Université de Sherbrooke, Sherbrooke, Québec, Canada, J1H 5N4

Within the nervous system, the post-transcriptional regulatory mechanisms contribute to the different cellular phenotypes. Indeed, weak deregulations can lead to diseases. RNA G-quadruplexes (G4) are emerging as new players in post-transcriptional regulation. G4s are very stable secondary structures found in DNA and RNA. They are formed by a minimum of two tetrads of guanines and stabilized by monovalent cations usually potassium. At the RNA level, they are involved in different mechanisms, such as alternative polyadenylation, alternative splicing, activation and inhibition of translation, and maturation of microRNAs. According to many, G4s are seen as events rather than entities. *Trans* elements, like proteins, are able to ensure their folding and unfolding.

A bioinformatic analysis confirmed that there is indeed an enrichment of RNA G4s in the transcripts of the nervous system. This analysis also identified Parkinson's disease as one of the diseases with the highest proportion of potential G4 (pG4). Out of 16 disease-related genes, 15 of them contain a least one pG4. Then, different biochemical techniques such as fluorescence assay, circular dichroism, and *in-line* probing were used to confirm the formation of G4s located in the different 5' untranslated regions (5'UTR). These techniques identified 4 new RNA G4s that have never been characterized before. Subsequently, luciferase assays concluded that two of these G4s, in the mRNA of VPS35 and PRKN, have a repressive effect on translation. The proteins capable of binding these two G4s were then identified by mass spectrometry. The protein Guanine Nucleotide-Binding Protein-Like 1 (GNL1) was identified as being a protein capable of binding the two G4s. Finally, the interaction between GNL1 and the G4 from VPS35 and PRKN was confirmed by electrophoretic mobility shift assays.

In brief, this study identified a new protein, GNL1, which can bind two new G-quadruplexes located in deregulated genes associated with Parkinson's disease. The regulation of the interaction between GNL1 and the two G4s could be a therapeutic option to counter the deregulations present in Parkinson's disease.

Keywords: G-quadruplex, ARN, Parkinson, RNA-binding protein

TABLE DES MATIERES

1	Résumé	iv
2	Summary	v
3	Table des matières	vi
4	Liste des figures	viii
5	Liste des tableaux	x
6	Liste des abréviations	xi
1	Introduction	1
1	Les acides nucléiques, interactions canoniques et non canoniques	1
2	Structures G-quadruplexes	4
2.1	G-quadruplexes d'ADN	6
2.2	G-quadruplexes d'ARN	6
2.2.1	ARN messenger	7
2.2.1.1	Région 5' non traduite	7
2.2.1.2	Cadre de lecture ouvert	8
2.2.1.3	Région 3' non traduite	8
2.2.2	ARN non codants	9
3	Protéines liant des G-quadruplexes	10
3.1	Hélicases	11
3.2	Protéines chaperons	11
4	Méthodes de détection et de caractérisation des G-quadruplexes	14
4.1	Bio-informatique	14
4.2	Molécules fluorescentes	15
4.3	Anticorps spécifiques	16
4.4	Cartographie	16
4.4.1	<i>In-line</i>	17
4.4.2	Diméthylsulfate	17
4.4.3	L'empreinte T1	17
4.4.4	Transcription inverse	18
4.5	Dichroïsme circulaire	18
4.6	Résonance magnétique nucléaire	19
4.7	Cristallographie de rayons X	20
4.8	Spectrométrie Raman	21

4.9	Essais luciférase	21
5	Maladies neurodégénératives	22
5.1	ARN et G-quadruplexes associés aux maladies neurodégénératives	22
5.2	Maladie de Parkinson	24
5.2.1	Symptômes	24
5.2.2	Causes moléculaires	25
5.2.2.1	Dysfonction du protéasome	26
5.2.2.2	Dysfonction de l'autophagie	26
5.2.2.3	Dysfonction du repliement des protéines	27
5.2.2.4	Dysfonction mitochondriale	27
5.2.2.5	Dysfonction dans la signalisation de l'insuline	28
6	Hypothèse/problématique	30
2	Résultats	31
	Article « Guanine Nucleotide-Binding Protein-Like 1 (GNL1) binds RNA G-quadruplex structures in genes associated with Parkinson's disease »	31
3	Discussion	96
	Analyse de l'enrichissement des G-quadruplexes dans les maladies du système nerveux	96
	Analyse de la formation des G-quadruplexes <i>in vitro</i>	98
	Analyse de l'impact de formation des G-quadruplexes <i>in cellulo</i>	104
	Analyse des partenaires protéiques de PRKN et de VPS35	107
	Analyse de l'interaction entre GNL1 et le G4 de VPS35 et PRKN	112
	Méthodes alternatives	114
	Perspectives au projet	115
4	Conclusion	116
5	Liste des références	117
6	Annexe 1	133
	The Small Nuclear Ribonucleoprotein Polypeptide A (SNRPA) binds to the G-quadruplex of the BAG-1 5'UTR	134

LISTE DES FIGURES

Figure 1	: Structure chimique et liaison hydrogène des différents acides nucléiques	p. 2
Figure 2	: Liaisons non canoniques des bases azotées	p. 3
Figure 3	: Structure d'un G-quadruplexe	p. 4
Figure 4	: Différents types de conformations d'un G-quadruplexe	p. 5
Figure 5	: Fonctions cellulaires des G-quadruplexes	p. 9
Figure 6	: Dysfonctions souvent impliquées dans la pathologie de Parkinson	p. 24
Figure 7	: Enrichment of pG4 in genes associated with the nervous system	p. 73
Figure 8	: Characterization of RNA G-Quadruplexes found in genes deregulated in PD by fluorescence assay	p. 74
Figure 9	: Characterization of the RNA G-Quadruplexes found in genes that are deregulated in PD by <i>in-line</i> probing	p. 75
Figure 10	: <i>In cellulo</i> characterization of RNA G-Quadruplexes in SH-SY5Y cells	p. 76
Figure 11	: RNA pull-down of proteins by the PRKN and VPS35 G-quadruplexes	p. 77
Figure 12	: Electrophoretic mobility shift assays of the GNL1 protein	p. 78
Figure S1	: <i>In-line</i> probing of the five positive pG4	p. 86-87
Figure S2	: Modified version of PsiCHECK-2 used with either the 5'UTR WT or the G/A mutant G4s of VPS35, LRRK2, SNCA and PRKN	p. 88
Figure S3	: Luciferase assays in HEK293T cells	p. 89
Figure S4	: Sequences of the wild type and the G/A mutant G4 constructs of pMA-RQ-VPS35/PRKN-4xS1m	p. 90
Figure S5	: NMM fluorescence assays of the chimeric RNAs that included the 4xS1m aptamer	p. 91
Figure S6	: <i>In-line</i> probing of the chimeric RNAs containing the 4xS1m aptamer	p. 92

Figure S7	: Purification profile of the fused GST-GNL1 protein	p. 93
Figure 13	: Proposition de G-quadruplexes observés par cartographie <i>in-line</i>	p. 98
Figure 14	: Dichroïsme circulaire des structures G-quadruplexes de VPS35 et PRKN	p. 99
Figure 15	: Dénaturation thermique des structures G-quadruplexes de VPS35 et PRKN	p. 100
Figure 16	: Essais d'arrêt de transcription inverse	p. 102
Figure 17	: Comparaison entre l'analyse protéique SAINT stricte et moins stricte	p. 109

LISTE DES TABLEAUX

Tableau 1	: Protéines pouvant lier les G-quadruplexes	p. 12
Tableau 2	: 5'UTR pG4 located in genes related to Parkinson's disease and manual analysis of G4 structures	p. 72
Tableau S1	: Oligonucleotides used in this study	p. 80
Tableau S2	: Pattern of the mutations used for the conception of the G/A mutant oligonucleotides	p. 84
Tableau 3	: Températures de dénaturation déterminées par dénaturation thermique	p. 101

LISTE DES ABRÉVIATIONS

2'OH	2'-Hydroxyle
A	Adénine
AD	<i>Alzheimer Disease</i> , Maladie d'Alzheimer
ADN	Acide désoxyribonucléique
ALS	<i>Amyotrophic lateral sclerosis</i> , Sclérose latérale amyotrophique
ARN	Acide ribonucléique
ARNm	ARN messenger
ARNt	ARN de transfert
ATP	Adénosine Triphosphate
C	Cytosine
CDS	<i>Coding sequence</i> , Séquence codante
DMS	Diméthylsulfate
DHX36	<i>DEAH box protein 36</i> , Protéine 36 possédant une boîte DEAH
eIF4e	<i>Eukaryotic translation initiation factor 4E</i> , Facteur d'initiation de la traduction eucaryote 4E
eIF4G1	<i>Eukaryotic translation initiation factor 4 Gamma 1</i> , Facteur d'initiation de la traduction eucaryote 4 Gamma 1
FBXO7	<i>F-Box Protein 7</i> , Protéine F-Box 7
FMRP	<i>Fragile X Mental Retardation Protein</i> , Protéine responsable de la retardation mentale dans le X fragile
FTD	<i>Frontotemporal degeneration</i> , Dégénération fronto-temporal
G	Guanine
G4	G-quadruplexe
G/A mutant	G-quadruplexe muté
GIGYF2	<i>GRB10 Interacting GYF Protein 2</i> , Protéine GYF interagissant avec GRB10
HEK	<i>Human embryonic kidney cells</i> , Cellules rénales embryonnaires humaines
IGF-1	<i>Insuline growth factor 1</i> , Facteur de croissance de l'insuline 1
IRES	<i>Internal ribosome entry site</i> , Site d'entrée interne du ribosome
K	Potassium
KCl	Chlorure de potassium
KEGG	<i>Kyoto Encyclopedia of Genes and Genome</i> , Encyclopédie des gènes et génomes de Kyoto
Li	Lithium
LiCl	Chlorure de lithium
LRRK2	<i>Leucine-rich repeat kinase 2</i> , Kinase 2 riche en répétition leucine
lncARN	<i>Long non coding ARN</i> , Long ARN non codant
miARN	micro ARN
MPTP	1-méthyl-4-phenyl-1,2,3,6-tetrahydropyridine
MS	<i>Mass spectrometry</i> , Spectrométrie de masse
N	Nucléotides : adénine, cytosine, thymine, uracile ou guanine
ND	<i>Neurodegenerative disease</i> , Maladie neurodégénérative
NMM	N-méthyle mésoporphyrine IX
ORF	<i>Open reading frame</i> , Cadre de lecture ouvert

PARK7/DJ-1	<i>Parkinsonism Associated Deglycase</i> , Déglycase associée au Parkinsonisme
PINK1	<i>PTEN-induced kinase 1</i> , Kinase induite par PTEN 1
pG4	<i>Potential G-quadruplex</i> , G-quadruplexe potentiel
PD	<i>Parkinson Disease</i> , Maladie du Parkinson
Pre-miARN	Pré-micro ARN
PINK1	<i>PTEN-induced kinase 1</i> , Protéine kinase induite par PTEN
PRKN	<i>Parkin RBR E3 Ubiquitin Protein Ligase</i> , Protéine ubiquitine ligase Parkin RBR E3
R	Purine (Adénine ou Guanine)
RBP	<i>RNA-Binding Protein</i> , Protéines liants l'ARN
RT	<i>Reverse transcription</i> , Transcription inverse
RMN	Résonance magnétique nucléaire
RNase T1	Ribonucléase T1
ROS	<i>Reactive oxygen species</i> , Espèces réactives à l'oxygène
S1m	<i>Aptamer S1 modified</i> , Aptamère S1 modifié
SNCA	α -synucléine
SN	<i>Substantia Nigra</i>
T	Thymine
Tm	<i>Melting temperature</i> , Température de dénaturation
TSS	<i>Transcription start site</i> , Site d'initiation de la transcription
U	Uracile
UTR	<i>Untranslated region</i> , Région non traduite
VPS35	Vacuolar protein sorting-associated protein 35, Protéine 35 associée aux protéines vacuolaires sortantes
WT	<i>Wild-type</i> , Type sauvage
Y	Pyrimidine (Cytosine, Thymine ou Uracile)

INTRODUCTION

Malgré les avancées médicales et scientifiques, certaines maladies sont toujours incurables. Certaines d'entre elles, comme les maladies neurodégénératives, peuvent rapidement changer la vie d'une personne. On peut penser à la maladie de Parkinson ou à la maladie d'Alzheimer qui sont grandement présentes dans notre société. Le développement de nouvelles thérapies cellulaires dans le but de soigner celles-ci est donc nécessaire. La complexité de ces maladies est directement associée aux nombreux mécanismes de régulation présents dans les neurones. Dans cette optique, la compréhension des mécanismes de régulation neuronale au niveau fondamental est un facteur clé à un potentiel traitement.

1 Les acides nucléiques, interactions canoniques et non canoniques

Chez tous les organismes vivants, le matériel génétique est contenu dans leur code génétique, soit la séquence nucléotidique formant leur génome. Les 4 principaux nucléotides retrouvés dans l'ADN (acide désoxyribonucléique) sont l'adénine (A), la cytosine (C), la guanine (G) et la thymine (T). Au niveau de l'ARN (acide ribonucléique), l'uracile (U) remplace la thymine (Figure 1A-D). Ces acides nucléiques peuvent s'apparier entre eux grâce à des liaisons hydrogène pour former ce qu'on appelle des paires de bases. La liaison de type Watson-Crick est la liaison la plus observée. Elle est formée par l'interaction entre une purine (R) et une pyrimidine (Y). Plus précisément, il s'agit de l'interaction par deux liaisons hydrogène entre l'adénine et la thymine (ou l'uracile) et par trois liaisons hydrogène entre la guanine et la cytosine. Les 2 liaisons hydrogène entre A et T/U sont formées par la liaison entre l'hydrogène de l'azote 6 du A et l'oxygène 4 du T/U et entre l'azote 1 du A et l'hydrogène de l'azote 3 du T/U (Figure 1E). Les 3 liaisons hydrogène entre la guanine et la cytidine sont quant à elles formées par la liaison entre l'oxygène 6 du G et l'hydrogène de l'amine 4 du C, entre l'hydrogène de l'azote 1 du G et l'azote 3 du C et entre l'hydrogène de l'amine 2 du G et de l'oxygène 2 du C (Figure 1E). Les 3 liaisons hydrogène formant la paire de bases entre la guanine et la cytosine assurent une liaison énergétiquement plus stable et demandent plus d'énergie pour induire une dénaturation que pour l'adénine et de ses deux partenaires. L'ARN et l'ADN peuvent donc se retrouver sous forme simple brin ou double

brin. Par ailleurs, ces brins peuvent former ce qu'on appelle des structures secondaires. Celles-ci sont formées par l'appariement de plus d'un brin ADN ou d'ARN. L'exemple classique est la double hélice d'ADN. Elle est considérée comme étant structure assez stable. Il existe cependant d'autres types de structures secondaires, comme les tiges-boucles, les pseudonœuds (*pseudoknots*) et plusieurs autres.

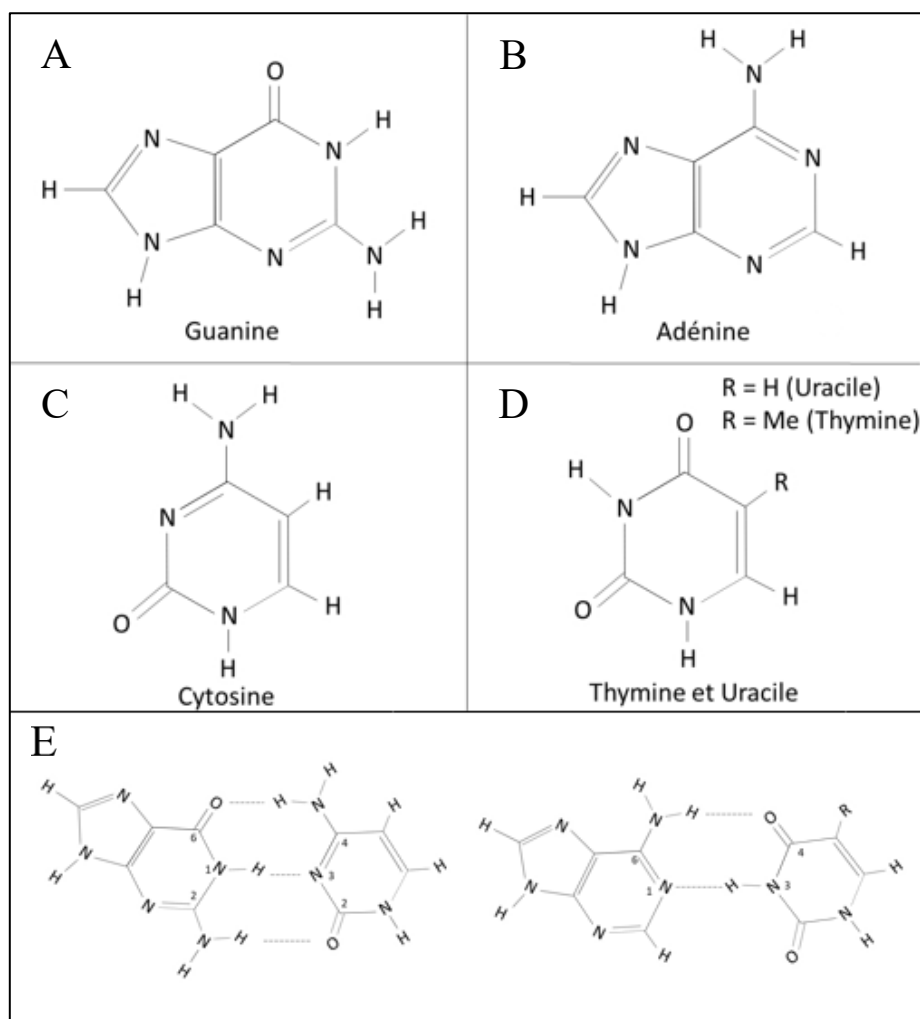


Figure 1 : Structure et liaisons hydrogène des différents acides nucléiques.

A) Guanine, **B)** Adénine, **C)** Cytosine, **D)** Thymine/Uracile, et **E)** Paires de bases Watson-Crick entre la guanine et la cytosine (gauche) et entre l'adénine et la thymine ou l'uracile (droite). Les chiffres représentent la position des atomes participants aux ponts hydrogènes.

Il existe aussi différentes structures secondaires au niveau de l'ADN et de l'ARN qui impliquent d'autres types de liaison. On appelle ces liaisons « non canonique ». L'une d'entre elles est la liaison Hoogsteen (Ghosal et Muniyappa, 2006). Dans cette liaison, l'azote en position N7 et groupement oxygène en position C6 de la purine agissent comme un accepteur de pont hydrogène. Cela a pour impact de former une liaison avec les atomes N3 et l'azote en C4 de la pyrimidine (Figure 2). Bien que toutes les possibilités de liaisons Hoogsteen ne soient pas aussi stables que celles de type Watson-Crick, certaines d'entre elles sont en mesure de se produire et de former des structures qui ont un effet régulateur sur la réplication, la transcription et la traduction en cellule (Rangadurai et al., 2018).

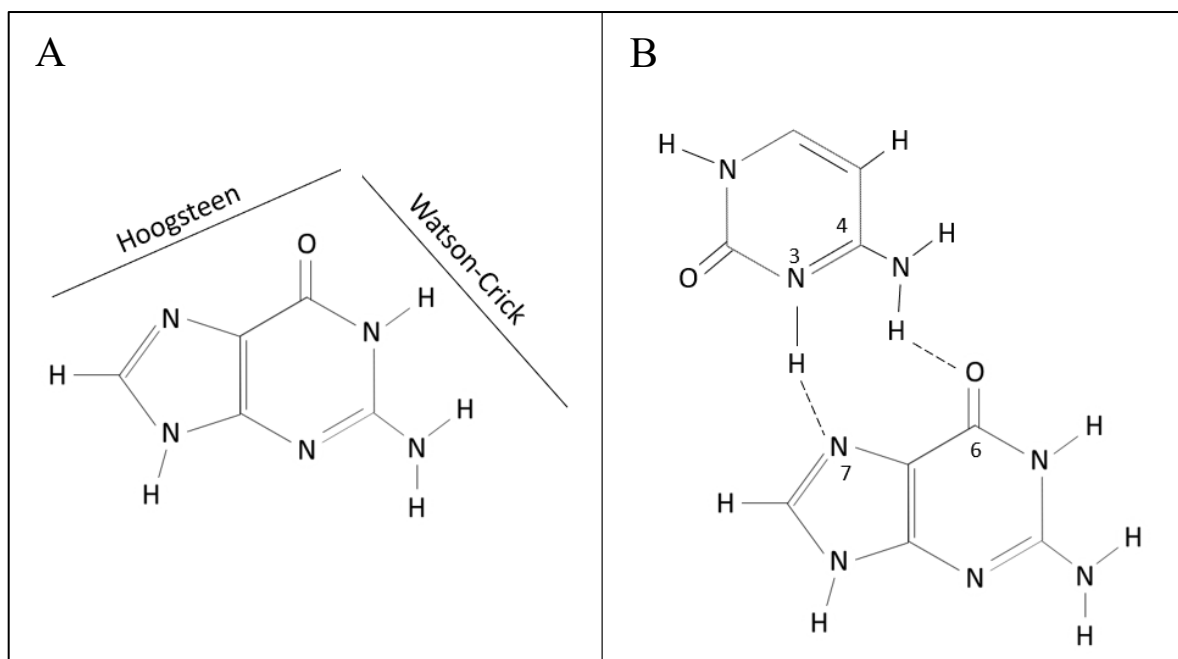


Figure 2 : Liaisons non canoniques des bases azotées

A) Différentes faces d'une base azotée pouvant participer à des liaisons hydrogène, et **B)** liaisons non canoniques Hoogsteen. Les chiffres représentent la position des atomes participant aux ponts hydrogènes.

2 Structures G-quadruplexes

Quatre guanines peuvent s'apparier entre elles grâce à quatre liens Hoogsteen pour former ce qu'on appelle un quartet (Figure 3A, B). Ces quartets peuvent ensuite s'empiler les uns par-dessus les autres pour former ce qu'on appelle un G-quadruplexe (G4). Originellement, on considérait qu'un G4 pouvait se former à partir de l'empilement de 3 quartets de guanine (Figure 3C) (Harkness et Mittermaier, 2017). Ce phénomène peut seulement s'effectuer en présence de certains ions spécifiques de charges et de grosseurs bien précises. Au niveau cellulaire, la formation du G-quadruplexe est hautement stabilisée par le potassium (K) tandis que d'autres ions, comme le lithium (Li), contribue à la déstabilisation. Il existe cependant des ions intermédiaires, comme le sodium, qui sont en mesure de venir stabiliser la structure à un plus faible niveau (Hardin et al., 1992). La fonction primaire du cation monovalent est de stabiliser la haute électronégativité présente au centre du quartet engendrée par la proximité des 4 oxygènes qui amènent une forte densité électronique (Figure 3A) (Chen et Yang, 2012). Les G-quadruplexes de type canonique répondent à la formule générale $G_x-N_{1-7}-G_x-N_{1-7}-G_x$ (où x est supérieur ou égal à 3, G correspond à une guanine et N à l'un des 4 nucléotides) (Wong et al., 2010). L'empilement des quartets va aussi former ce qu'on appelle des boucles (démonstrées par les N dans la formule). Elles sont constituées des nucléotides ne participant pas à l'interaction Hoogsteen, mais se retrouvant entre chacune des séries de guanines (Cevic et Plavec, 2005). C'est donc un total de trois boucles qui seront présentes dans chacun des G4 intramoléculaires, et ce peu importe la quantité de tétrades participant au G4. Celles-ci auront un rôle à jouer dans la stabilité du G4 et dans sa détection par plusieurs facteurs *trans*. Par ailleurs, l'orientation de l'assemblage du G-quadruplexe aura aussi un impact sur sa fonction, sa reconnaissance et sa stabilité. Selon le type d'acide nucléique dans lequel la structure se retrouve (ADN ou ARN), les G4 peuvent se retrouver sous une forme parallèle, antiparallèle de type chaise, antiparallèle de type panier ou bien hybride (Figure 4) (Burge et al., 2006).

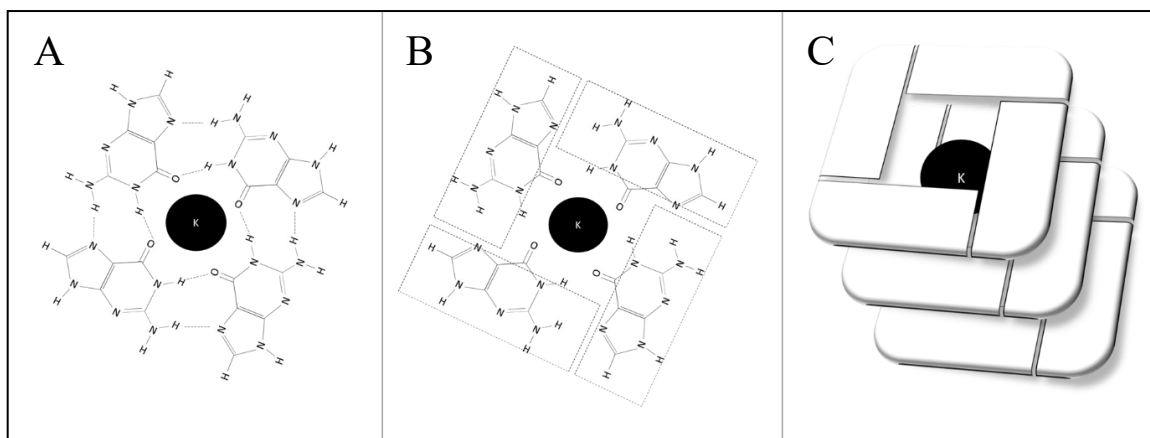


Figure 3: Structure d'un G-quadruplexe

A) Tétrade de guanines stabilisée par un ion potassium (K), et **B)-C)** Structure imagée de l'empilement de trois tétrades de guanines.

Des études récentes ont montré qu'il existait beaucoup de G-quadruplexes ne répondant pas à la formule canonique décrite ci-haut. On dit alors que les G4 sont de type non canonique. Certains peuvent avoir certaines boucles plus longues, des nucléotides autres que des guanines participant aux liens Hoogsteen, des espacements dans une série de guanine (GGNG) ou posséder seulement deux quartets (Bolduc et al., 2016; Meier et al., 2018; Sengar et al., 2019). À ce jour, ceux-ci ont beaucoup plus été ignorés et négligés lors des recherches. Il est donc important de tenir compte de ces G-quadruplexes de type non canonique puisqu'ils pourraient représenter une grande proportion des G4 en cellule. Par ailleurs, comme ils sont non canoniques et présentent plus de diversités, il est possible de penser que ceux-ci pourraient être plus facilement ciblés par des molécules chimiques, des anticorps ou des oligonucléotides antisens.

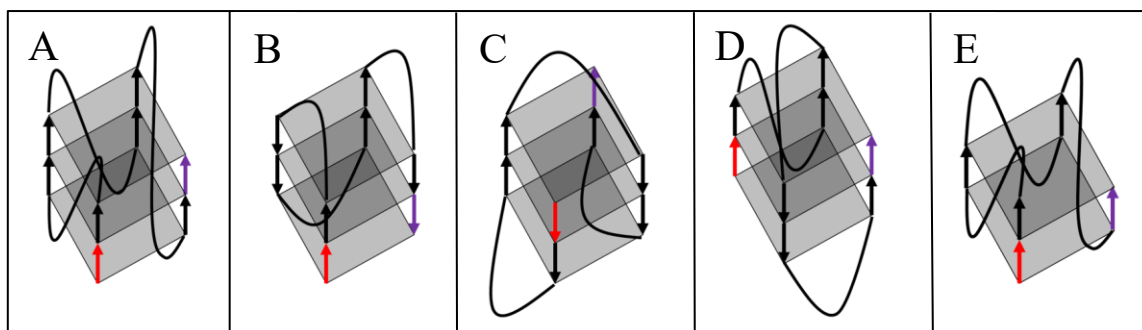


Figure 4 : Différents types de conformation d'un G-quadruplexe

A) Parallèle, **B)** Antiparallèle de type chaise, **C)** Antiparallèle de type panier, **D)** Hybride, et **E)** 2 quartets. Le trait rouge est associé au début du G4, tandis que le trait mauve est associé à la fin du G4.

2.1 *G-quadruplexes d'ADN*

Le G-quadruplexe d'ADN est le type de G4 ayant le plus été étudié au fil des années. L'intérêt envers celui-ci résulte directement du fait qu'il est présent au niveau des télomères de l'ADN (Blackburn, 2000). La topologie des G4 présents dans les télomères est bien unique puisqu'il y a présence de G4 intramoléculaires de différents types de conformations (Dai et al., 2007; Lim et al., 2009). La formation de ces G4 a été associée à une inhibition de l'activité télomérase, ce qui en fait une cible thérapeutique intéressante dans les cancers (Zahler et al., 1991). Par ailleurs, plusieurs G4 d'ADN ont été retrouvés dans des séquences promotrices et ont un rôle de régulateur sur la transcription des gènes d'intérêts. Parmi ceux-ci, on retrouve beaucoup de gènes dérégulés dans le cancer comme c-MYC, VEGF, K-RAS (Cogoi et al., 2008; Siddiqui-Jain et al., 2002; D. Sun et al., 2005). Il a d'ailleurs été montré qu'il existait un enrichissement en séquences pouvant potentiellement former un G4 (pG4) dans les régions promotrices des gènes humains, près de sites de départ de la transcription (*transcription start site*, TSS) (Huppert et Balasubramanian, 2007).

2.2 *G-quadruplexes d'ARN*

Les G-quadruplexes d'ARN ont beaucoup de similarités structurales avec les G4 d'ADN. Cependant, il a été démontré que les G4 d'ARN ont une plus grande stabilité thermodynamique (Joachimi et al., 2009). Cette stabilité accrue serait premièrement due à un

meilleur empilement des tétrades. En effet, une hydratation réduite du G4 est observée dans le cas d'un G4 d'ARN (Olsen et Marky, 2009). Plus l'hydratation d'un G4 est élevée, plus il tend à adopter une forme non repliée. Cette diminution d'hydratation augmente donc sa stabilité globale en le rendant plus compact. Le deuxième facteur influençant de manière positive la stabilité du G4 d'ARN est la présence d'un réseau supplémentaire de ponts hydrogène reliés aux groupes 2'-hydroxyle (2'OH) du ribose (Pagano et al., 2008). En effet, il est possible d'observer un plus long temps de résidence des ponts hydrogène entre les 2'OH et l'accepteur du pont hydrogène situé sur le résidu suivant (les liaisons se produisent pendant un plus grand délai de temps). Une autre différence peut aussi être observée au niveau conformationnel du G4 d'ARN. Contrairement à celui d'ADN, le G4 d'ARN est seulement en mesure d'adopter une conformation parallèle. Ce phénomène est principalement dû au fait qu'il est très coûteux énergétiquement pour un G4 d'ARN avoir des résidus en position *syn* et que ceux-ci sont nécessaire pour avoir des orientations antiparallèles (Fay et al., 2017a). Au point de vue fonctionnel, les G-quadruplexes d'ARN ont des fonctions qui dépendent de leurs emplacements dans les transcrits et du type d'ARN dans lesquels ils se retrouvent.

2.2.1 ARN messenger

Les G-quadruplexes retrouvés dans les ARN messenger (ARNm) sont ceux ayant été le plus étudiés au cours des années. La fonction exercée par les G4 au niveau post-transcriptionnel d'ARNm est dépendante de leurs positions dans le transcrit d'ARN (S. Rouleau et al., 2017). En effet, un même G4 situé dans différentes positions pourra avoir diverses fonctions. Cependant, ces fonctions pourront être récurrente à plusieurs positions d'un même transcrit (Figure 5C).

2.2.1.1 Région 5' non traduite

Les G-quadruplexes situés dans la région 5' non traduite (*5' untranslated region*, 5'UTR) sont ceux étant les plus documentés dans la littérature. Ceux-ci sont majoritairement reconnus comme ayant un effet répressif sur l'initiation coiffe-dépendante de la traduction (Beaudoin et Perreault, 2010; Jodoin et al., 2014, 2019). Deux mécanismes ont été proposés pour expliquer cet effet. Le premier mécanisme implique la haute stabilité du G4. Celle-ci amènerait un arrêt lors de la progression de la sous-unité du ribosome dans le 5'UTR (Shahid

et al., 2010). Le second mécanisme implique un effet négatif du G4 sur la reconnaissance de la structure coiffée de l'ARNm par le facteur d'initiation de la traduction eucaryote 4E (*Eukaryotic translation initiation factor 4E*, eIF4e) (Kumari et al., 2007). Cependant, des exceptions existent. En effet, certains G4 situés en 5'UTR ont été identifiés comme ayant un rôle positif sur la traduction des ARNm. Ceux-ci sont cependant beaucoup moins communs. Dans ce cas-ci, le mécanisme pourrait être dû à des sites d'entrée interne du ribosome (*Internal ribosome entry site*, IRES) qui sont favorisés lors de la formation du G4, ou bien à d'autres mécanismes qui demeurent encore à être identifiés (Agarwala et al., 2013; Bonnal et al., 2003; Morris, Negishi, et al., 2010).

2.2.1.2 Cadre de lecture ouvert

Les G-quadruplexes situés dans le cadre de lecture ouverte (*Open reading frame*, ORF) sont très peu nombreux et surtout reconnus comme étant des inhibiteurs stériques à la progression du ribosome (Endoh et al., 2013a). Par ailleurs, certaines études ont montré que la présence de G4 dans l'ORF est essentielle au repliement et à la protéolyse de certaines protéines, et ce, dû à un ralentissement ou un arrêt partiel du ribosome lors de l'élongation (Endoh et al., 2013b). De plus, les G4 situés dans l'ORF sont aussi en mesure d'induire des changements de cadre de lecture (Yu et al., 2014). Ce changement de cadre (+1 ou -1) résulte souvent en un ARNm plus court en raison de codons de terminaison prématurés. Ce phénomène étant plus rare, il est tout de même observé dans certains ARNm (Endoh et Sugimoto, 2013).

2.2.1.3 Région 3' non traduite

Les fonctions des G-quadruplexes situés dans la région 3' non traduite (*3' untranslated region*, 3'UTR) ont été beaucoup moins étudiées et sont donc moins comprises que celles associées aux autres régions. Ces G4 peuvent avoir une fonction répressive ou activatrice sur la traduction de l'ARNm, et ce, sans avoir d'impact au niveau de la stabilité de l'ARN (Bolduc et al., 2016). Ces changements de régulation au niveau traductionnel seraient associés à la liaison de certaines protéines. L'inhibition stérique ou séquence spécifique de la liaison de certains micro ARN (miARN) pourrait aussi être impliquée. Une autre fonction spécifique aux G4 retrouvés dans le 3'UTR a aussi été identifiée. Les G4 d'ARN en 3'UTR sont en mesure de réguler la formation de site de polyadénylation alternatif (Beaudoin et

Perreault, 2013). Les G-quadruplexes ont alors un impact sur la variation de longueur du 3'UTR. La liaison des miARN peut aussi être affectée par cette diminution ou augmentation de la longueur de l'UTR.

2.2.2 ARN non codants

Récemment, plusieurs études ont vu le jour associant une fonction à des G-quadruplexes présents dans des ARN non codants. Au niveau des longs ARN non codants (*Long non coding RNA*, lncARN), les G4 qui y sont présents sont surtout associés à une séquestration de protéines, ce qui a un impact sur la régulation endogène de ces protéines. En effet, un exemple est le lncARN contenant des séquences formant des G-quadruplexes (*G-quadruplex-forming sequence containing lncRNA*, GSEC) qui a été identifié comme étant en mesure de séquestrer la protéine 36 possédant une boîte DEAH (*DEAH box protein 36*, DHX36) et de causer une inhibition de son activité hélicase sur les autres ARN dans la cellule (Matsumura et al., 2017). Un second rôle a aussi été identifié. Les G-quadruplexes pourraient avoir un impact sur l'interaction antisens lncARN-ARNm en venant affecter les sites de complémentarité entre les deux ARN (Gong et Maquat, 2011). Cependant, il existe possiblement bien d'autres rôles associés aux G4 dans les lncARN puisque peu de recherches ont été réalisées dans ce domaine.

Des G-quadruplexes ont aussi été retrouvés au sein de pre-micro ARN (pre-miARN) (Mirihana Arachchilage et al., 2015; Pandey et al., 2015). La formation de ces G4 a un effet direct sur la production totale de miARN. En effet, les G4 formés sont en mesure d'empêcher la coupure du pre-miARN par la protéine Dicer. Cela entraîne donc une diminution du miARN mature dans la cellule (S. Rouleau et al., 2017). Il a aussi été montré que la formation d'un G4 dans un miARN mature peut empêcher sa liaison à son ARN cible (Tan et al., 2016).

Pour finir, certains G4 ont été retrouvés au sein d'ARN de transfert (ARNt). Ceux-ci ne sont pas aussi présents que ceux retrouvés dans les autres ARN non codants, mais ont quand même une fonction au niveau de la régulation post-transcriptionnelle. Plus précisément, certains G4 ont été trouvés dans des ARNt clivés par l'angiogénine (Ivanov et al., 2014). Les G4 situés dans la partie 5' de l'ARNt sont alors requis pour venir inhiber la traduction de certains ARNm.

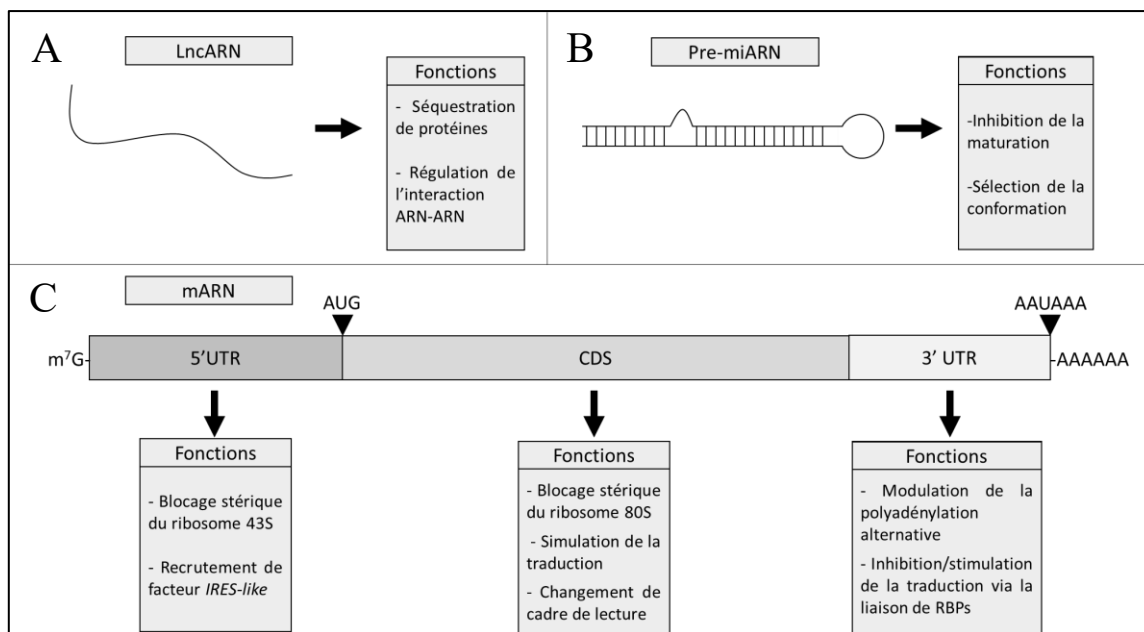


Figure 5 : Fonctions cellulaires des G-quadruplexes

A) Implications dans les longs ARN non codants, **B)** Implications dans la maturation des miARN, et **C)** Implications dans les ARNm.

3 Protéines liant des G-quadruplexes

Un aspect important à comprendre lors de l'étude des G-quadruplexes est que ceux-ci sont constamment en équilibre dans la cellule. Que ce soit pour les G4 d'ADN ou d'ARN, il existe une dynamique où les structures des acides nucléiques alternent entre la forme G4 et linéaire. Certains pensent qu'ils seraient globalement non formés dans la cellule (Guo et Bartel, 2016). Un mécanisme de régulation doit donc être présent pour coordonner la formation et le dépliement de ceux-ci et assurer une bonne régulation post-transcriptionnelle (RPT). Récemment, la communauté scientifique a étudié différents facteurs *trans* pouvant venir réguler la formation des G-quadruplexes (dans l'ADN et dans l'ARN). Ils ont découvert qu'une certaine quantité de protéines était en mesure d'interagir avec les G4 pour les déformer (les hélicases) et les stabiliser (Brázda et al., 2014). Bien qu'il existe des exceptions, ces protéines possèdent, la plupart du temps, des domaines de liaison à l'ARN RGG (tri-RGG, di-RGG, RGG) ou RG (tri-RG, di-RG, RG) dans leurs séquences peptidiques (R représentant ici une arginine et G une glycine) (Thandapani et al., 2013). La compréhension

globale des mécanismes de liaison entre les protéines et les G4 pourrait nous en apprendre plus sur les processus impliqués dans la RPT et nous donner des outils pour la réguler dans une optique thérapeutique. Cependant, aucune protéine ayant une spécificité unique pour un seul G4 n'a été trouvée à ce jour. En revanche, certaines protéines ont été identifiées avec une spécificité pour certaines conformations de G4 (Benhalevy et al., 2017).

3.1 Hélicases

Les hélicases sont des protéines qui ont la capacité de déplier des structures secondaires d'ADN ou d'ARN. La plupart d'entre elles utilisent l'énergie contenue dans l'adénosine triphosphate (ATP) pour effectuer leur fonction (Patel et Donmez, 2006). Il existe plusieurs types d'hélicases, certaines ont comme substrat l'ADN, d'autre l'ARN et certaines peuvent même avoir une activité sur les deux. Elles sont séparées en six superfamilles (SF 1 à 6) et chacune de ces familles possède des substrats et des orientations (5' vers 3' ou 3' vers 5') bien définis (Mendoza et al., 2016). Les hélicases sont le type de protéines ayant le plus été identifié comme étant en mesure de lier les G-quadruplexes (voir Tableau 1). La plus étudiée parmi celles-ci est la protéine DHX36/RAUH. Elle est une hélicase appartenant à la famille des DEAH-box. DHX36 assure la régulation de beaucoup de gènes et d'ARN dans la cellule (Sauer et al., 2019). La popularité de celle-ci est surtout associée à sa grande affinité pour les G-quadruplexes (d'ARN ou d'ARN) qui se situe dans le picomolaire (Giri et al., 2011). C'est aussi l'une des rares protéines ayant été découvertes dont le mécanisme de liaison et de résolution du G4 est bien connu. En effet, pour assurer sa liaison aux G4, il a été démontré qu'il est essentiel que le G4 ait une extrémité simple brin 3' d'ADN ou d'ARN (Chen et al., 2018). Par ailleurs, il a été montré que la présence de ligand à G4 peut interférer avec l'activité de l'hélicase en augmentant la stabilité du G4 à un niveau trop élevé pour la capacité de l'hélicase ou en camouflant le site de liaison. C'est donc un phénomène qui pourrait être utilisé de manière thérapeutique.

3.2 Protéines chaperons

Les protéines pouvant avoir une fonction de chaperon sur la formation du G4 en cellule ont été beaucoup moins étudiées que les hélicases. Seules certaines ont été découvertes à ce jour (voir Tableau 1). La plupart de celles-ci ont de hautes constantes d'affinité pour les G4. Cependant, comme elles n'ont aucune fonction hélicase, elles contribuent à la formation du

G4 en créant un complexe avec celui-ci. Elles augmentent par le fait même leurs stabilités. La protéine la plus connue pour stabiliser les G4 est la protéine responsable de la retardation mentale dans le X fragile (*Fragile X Mental Retardation Protein*, FMRP). Celle-ci a été montrée comme étant capable de lier des G4 avec son domaine RGG et de les stabiliser (Agarwala et al., 2015). Dans plusieurs exemples, FMRP augmente l'effet répressif des G-quadruplexes situés en 5'UTR en les stabilisant, ce qui empêche la progression de la sous-unité ribosomale (Schaeffer et al., 2001). FMRP est d'ailleurs en mesure de délocaliser dans les neurites certains ARNm via leur G4 (Melko et Bardoni, 2010). Cependant, aucune protéine n'a été rapportée pour lier et stabiliser uniquement les G4 d'ARN.

Tableau 1 : Protéines pouvant lier les G-quadruplexes

Protéines rapportées pour lier les G4 d'ARN			
Nom	Espèce	Fonction	Références
AFF	Humain	Épissage	(Melko et al., 2011)
Aven	Humain	Traduction	(Thandapani et al., 2015)
DDX21	Humain	Hélicase	(McRae et al., 2017a)
DDX3X	Humain	Hélicase	(Herdy et al., 2018b)
DHX36	Humain	Hélicase	(Ariyo et al. 2015)
eIF4A	Humain	Hélicase	(Wolfe et al., 2014)
FMRP	Humain	Stabilisation	(Vasilyev et al., 2015)
Gis2	Cryptococcus neoformans	Hélicase	(Leipheimer et al., 2018)
GRSFI	Humain	Dégradosome mitochondrial	(Pietras et al. 2018)
LANA	KSHV	Stabilisation	(Dabral et al., 2019)
hnRNPF	Humain	Épissage	(Huang et al., 2017)
Lin28	Humain	Hélicase	(O'Day et al., 2015)
Mov10L1/Mov10	Humain	Hélicase	(Zhang, Yu, et al., 2019)
NCp7	HIV	Hélicase	(Butovskaya et al., 2019)
PRC2	Humain	-	(Wang et al., 2019)
YB-1	Humain	Stabilisation	(Ivanov et al., 2014)
Znf9	Cryptococcus neoformans	Hélicase	(Leipheimer et al., 2018)

Protéines rapportées pour lier les G4 d'ADN et ARN			
Nom	Espèce	Fonction	Références
CNBP	Humain	Hélicase	(Armas et al., 2008)
Dbp2	Levure	Hélicase (ARN)	(Gao et al., 2019)
DDX5	Humain	Hélicase	(Wu, Xing, et al., 2019)
Ded1	Levure	Hélicase (ARN)	(Gao et al. 2019)
DHX9	Humain	Hélicase	(Chakraborty et Grosse 2011)
hnRNPA1	Humain	Stabilisation	(Krüger et al., 2010)
HP1	Humain	Stabilisation G4 parallèle	(Benhalevy et al., 2017)
Mss116	Levure	Hélicase (ARN)	(Gao et al., 2019)
Nucleolin	Humain	Stabilisation	(Lago et al., 2017)
TLS/FUS	Humain	Échafaud	(Yagi et al., 2018)
TRF2	Humain	Échafaud	(Biffi et al., 2012)

Protéines rapportées pour lier les G4 d'ADN			
Nom	Espèce	Fonction	Références
BLM	Humain	Hélicase	(Sun et al., 1999)
DNA2	Souris	Hélicase	(Lin et al., 2013)
FANCI	Humain	Hélicase	(Wu et al., 2008)
HMGB1	Humain	Stabilisation	(Amato et al., 2018)
hnRNPA2	HIV	Hélicase	(Scalabrin et al., 2017)
Pif1	Levure	Hélicase	(Sanders, 2010)
RecQ	Bactérie	Hélicase	(Wu et Maizels, 2001)
Rif1	Levure	Association multiple de G4 (multimer)	(Masai et al., 2019)
Sgs1	Levure	Hélicase	(Sun et al., 1999)
Slx9	Levure	Reconnaissance et protection G4 stables	(Götz et al. 2019)

4 Méthodes de détection et de caractérisation des G-quadruplexes

Beaucoup de techniques *in vitro*, *in cellulo* et *in silico* ont été développées pour étudier les G-quadruplexes. Chacune de ces techniques possède des avantages et des inconvénients. Lors de l'étude des G-quadruplexes, il faut donc combiner plusieurs d'entre elles pour confirmer la présence d'un G4 dans une séquence donnée.

4.1 Bio-informatique

Les G-quadruplexes peuvent être prédits selon leur séquence pour former ce qu'on appelle des G4 potentiels (pG4). Plusieurs algorithmes informatiques ont été développés afin de maximiser la sensibilité et la spécificité de détection, c'est-à-dire de maximiser la probabilité de pouvoir trouver un véritable pG4 parmi des séquences aléatoires.

D'une part, les premiers algorithmes se basaient sur une séquence de référence, majoritairement $G_x-N_{1-7}-G_x-N_{1-7}-G_x-N_{1-7}-G_x$ (où $x \geq 3$). Un exemple de ceux-ci est QCRS mapper (Kikin et al., 2006). Assez rapidement, la communauté impliquée dans l'étude des G4 est venue à la conclusion que plusieurs G4 ne répondaient pas à ces critères (G4 non canoniques). Les algorithmes ont alors évolué en une seconde génération donnant une valeur à chacun des nucléotides présents dans la séquence. C'est le cas du *cGcC score*, un outil informatique donnant un score en fonction du ratio entre le nombre de guanines consécutives sur le nombre de cytosines consécutives (Beaudoin et al., 2014). Un second exemple est G4 Hunter, un outil bio-informatique similaire au *cGcC score*, qui donne des valeurs positives à chacune des guanines consécutives et des valeurs négatives à chacune des cytosines consécutives (Bedrat et al., 2016).

Bien que ces outils soient grandement utilisés, l'intelligence artificielle a permis la création d'un nouvel algorithme plus performant, G4NN, basé sur l'entraînement d'un réseau de neurones face à des séquences positives et négatives (Garant et al., 2017). L'algorithme a été entraîné avec des séquences ayant été confirmées en laboratoire comme étant en mesure de former un G4 et d'autres séquences ne pouvant pas adopter cette structure secondaire. Plutôt que de se baser sur chacun des nucléotides, la séquence est transformée en plusieurs vecteurs étant rattachés à des trinucleotides. L'utilisation de l'apprentissage profond a un grand

avantage par rapport aux autres techniques puisque la valeur attribuée à chaque nucléotide/trinucléotide n'est pas biaisée par le pointage donné par l'expérimentateur. Cependant, un biais peut être amené lors de la sélection des séquences d'entraînement. Récemment, G4 RNA screener a été créé dans le but de réunir tous ces prédicteurs (cGcC, G4NN et G4 Hunter) et d'en tirer les éléments positifs et négatifs propres à chacun des algorithmes (Garant et al., 2018).

4.2 Molécules fluorescentes

Les molécules fluorescentes pouvant lier les G4 sont grandement utilisées pour confirmer la formation de pG4 (Umar et al., 2019). Ces molécules sont en mesure d'émettre à une longueur d'onde bien précise lorsqu'elles sont excitées à une longueur d'onde donnée. Pour ce faire, il y a utilisation d'un spectrofluoromètre qui envoie une onde à une longueur d'excitation sur l'échantillon et un détecteur à 90° qui capte à la longueur d'onde d'émission. Habituellement, les molécules utilisées pour la détection d'un G4 ont des constantes d'affinités assez élevées avec le G4 (Vummidi et al., 2013). En plus d'être utilisées pour la confirmation de formation d'un G4, ces molécules fluorescentes ont aussi été utilisées pour la visualisation des G4 en cellule (Lu et al., 2016).

Il existe différents types de molécules fluorescentes utilisées pour la détection de G4. Elles sont séparées en différentes classes selon leur squelette moléculaire. Un exemple de classes de ligands est celle des porphyrines. Elles sont des molécules existant dans la nature et utilisées par les organismes comme cofacteurs (Yingfu Li et al., 1996). Celles-ci interagissent avec les G4 en venant se superposer sur les quartets et amènent, par le fait même, une stabilisation de la structure (Vummidi et al., 2013). Un exemple de cette classe est la molécule N-méthyle mésoporphyrine IX (NMM) qui est utilisée pour l'identification de G4 d'ADN et d'ARN (Nicoludis et al., 2012a; Tang & Shafer, 2006). Le NMM est sensible et sélectif pour les G4 et permet l'étude en temps réel de ceux-ci. Cependant, il ne permet pas la discrimination visuelle entre plusieurs types de G4 (Umar et al., 2019). Ce composé a une excitation qui se produit à 393 nm et une émission à 610 nm lorsqu'il est lié à un G4 (Gabelica et al., 2013). Une seconde classe de ligand est celle des benzothiozoles. Celles-ci sont des composés organiques cationiques créés en synthèse organique ayant plusieurs fonctions. Un exemple est la Thioflavin T (ThT), un ligand ayant un signal sans ligand très faible, ce qui

diminue le bruit de fond total. Il émet en revanche beaucoup plus de faux positifs et de faux négatifs que le NMM (Tong et al., 2013). Le ThT a une excitation qui se produit à 425 nm et a une émission à 490 nm (Gabelica et al., 2013). Ce ligand est reconnu comme étant en mesure de lier les G4 d'ADN et d'ARN sur la partie supérieure du quartet (Xu et al., 2016). Beaucoup d'autres classes de ligands fluorescents sont aussi disponibles commercialement, mais beaucoup moins utilisées. Un point important à garder en tête est que les ligands fluorescents développés pour lier spécifiquement les G4 sont souvent basés sur des G4 de type canonique. Leur efficacité envers les structures non canoniques n'est donc pas connue. De plus, aucune molécule fluorescente spécifique à un seul G4 n'a été développée à ce jour.

4.3 *Anticorps spécifiques*

Un anticorps (1H6) a récemment été développé contre les G-quadruplexes d'ADN en utilisant une librairie de motifs intramoléculaires (Biffi et al., 2013). Cet anticorps a aussi été utilisé pour la détection de G4 d'ARN (Biffi et al., 2014). L'utilisation de cet anticorps *in cellulo* a montré que des structures G4 sont présentes au niveau du noyau et du cytoplasme de cellules humaines. Cela amène une preuve importante de la formation des G4 dans le génome et le transcriptome. Cependant, il faut mentionner que la présence d'une liaison anticorps-G4 pourrait faire en sorte qu'un G4 instable soit en mesure de se former grâce à la liaison de l'anticorps (Fay et al., 2017a). Par ailleurs, certains groupes de recherche s'interrogent sur la sélectivité de cet anticorps, en particulier pour la reconnaissance de G4 non canonique. Par exemple, la présence d'ADN riche en thymine entraînerait un phénomène de la réactivité croisée de 1H6 (Kazemier et al., 2017).

4.4 *Cartographie*

La cartographie de séquence est un type de technique ayant pour but de visualiser les résidus d'une séquence d'ARN et d'identifier certaines caractéristiques propres à la structure adoptée. Il est souvent possible d'observer des hydrolyses ou des modifications chimiques uniques à la structure de l'ARN. Généralement, l'analyse est réalisée par électrophorèse sur gel d'acrylamide à des résolutions allant jusqu'au nucléotide. Les bandes peuvent alors être visualisées et quantifiées par densitométrie.

4.4.1 *In-line*

La technique de cartographie *in-line* est basée sur l'hydrolyse spontanée des liens d'une séquence d'ARN. En effet, en présence de magnésium, il peut se produire une transestérification entre le 2'OH et le phosphore adjacent qui vient rompre le lien entre deux nucléotides (Nahvi et Green, 2013). Dans un G4, les guanines faisant partie des quartets forment une structure assez rigide. Elles sont alors beaucoup moins accessibles et subissent moins de clivage. Cependant, d'autres nucléotides comme ceux présents dans les boucles sont davantage flexibles et plus susceptibles à l'hydrolyse. Les G4 sont les seules structures se formant en présence de potassium. Il est alors possible de s'attendre à un enrichissement de clivage au niveau des boucles en présence de potassium, comparativement à en présence de lithium. Un ratio arbitraire de 2 fois est associé à un enrichissement considéré comme étant significatif (Beaudoin et al., 2013). Le principal avantage de cette technique est l'utilisation d'ARN marqué de façon radioactive au phosphore gamma qui permet de travailler à faible concentration. Cela mime bien les phénomènes se produisant *in cellulo* et réduit la probabilité d'observer des G4 intermoléculaires.

4.4.2 *Diméthylsulfate*

Le marquage de l'ARN avec le diméthylsulfate (DMS) est aussi une technique grandement utilisée dans le domaine des G4. L'incubation de cette molécule avec l'ARN aura pour effet de méthyler les guanines en position N7. Comme dans un G4 les résidus guanines participant à la tétrade effectuent un pont hydrogène par la face Hoogsteen, ceux-ci ne seront pas accessibles pour la méthylation. Par la suite, une transcription inverse (*reverse transcription*, RT) est effectuée sur l'ARN. Dû à la modification chimique, la polymérase subit un arrêt aux sites ayant été modifiés par le DMS. Ces sites d'arrêts apparaissent alors avec une densitométrie plus élevée sur gel. Tout comme dans le cas du *in-line*, la réaction est effectuée en deux conditions : l'une favorisant la formation du G4 et l'autre la défavorisant. Les guanines effectuant des liens Hoogsteen participants aux tétrades peuvent alors être identifiées avec cette technique (Sun et Hurley, 2010).

4.4.3 *L'empreinte T1*

L'utilisation de certaines enzymes, comme la ribonucléase T1 (RNase T1) est aussi utilisée pour déterminer les guanines impliquées dans le G4. Celle-ci possède la capacité d'effectuer

un clivage en 3' de résidus guanines sous forme simple brin (Eun, 1996). Les structures secondaires, comme les G-quadruplexes, confèrent alors une protection aux guanines participant aux différents quartets face au clivage de l'enzyme. Il est ensuite possible de réaliser l'expérience dans plusieurs conditions et de visualiser quelles guanines subissent une réduction de clivage lorsqu'elles sont incubées en présence de potassium (Agarwala et al., 2017).

4.4.4 *Transcription inverse*

L'arrêt de la polymérase est une technique basée sur la stabilité du G4 vis-à-vis le passage de la polymérase. En effet, dans cette technique, une extension nucléotidique (G4 d'ADN) ou une transcription inverse (G4 d'ARN) est effectuée sur de l'ADN ou l'ARN marqué au phosphore radioactif. La présence d'un G-quadruplexe aura pour effet d'arrêter la polymérase (*Reverse stalling assay*, RTS) (Kwok et Balasubramanian, 2015). Plus le G4 aura une grande stabilité, plus cet effet sera élevé. Il est donc possible d'effectuer une RTS en présence de différents sels, molécules organiques ou bien avec des séquences mutantes pour identifier les sites d'arrêt de la polymérase (Zhang, Yang, et al., 2019).

4.5 *Dichroïsme circulaire*

Le dichroïsme circulaire est une technique utilisée en biophysique pour la détermination de structure secondaire des macromolécules (Ranjbar et Gill, 2009). Cette méthode est basée sur l'absorption de la lumière polarisée par les molécules biologiques. L'appareil émet une première lumière polarisée avec une direction horaire et une seconde avec une direction antihoraire. Lorsque la molécule présente une différence d'absorption pour l'un des deux types de direction, on dit qu'elle possède des propriétés de dichroïsme circulaire (Bishop et Chaires, 2002). Cette différence, aussi appelée ellipticité molaire, est calculée à une multitude de longueurs d'onde. Tout changement conformationnel dans une molécule peut alors être observé avec cette technique. Différentes conditions peuvent alors être utilisées pour voir comment la structure secondaire se comporte. Il est alors possible de faire la différence entre un G4 qui est formé et un qui ne l'est pas. Par exemple, un G4 parallèle d'ARN donne un spectre avec une ellipticité molaire possédant un maximum à environ 264 nm et un minimum à environ 240 nm (Del Villar-Guerra et al., 2018).

Une autre façon d'utiliser le dichroïsme circulaire est d'observer à une longueur d'onde bien précise, soit 264 nm pour les G4 d'ARN, et d'effectuer une dénaturation thermique du G4. Le G-quadruplexe va alors passer d'une forme repliée à dépliée, ce qui va entraîner un changement dans l'ellipticité molaire à cette longueur d'onde. Avec cette méthode, il est possible de déterminer la stabilité d'un G4 et d'en extraire plusieurs paramètres physico-chimiques, dont la température de dénaturation (*Melting temperature*, T_m) (Hao et al., 2019).

Le grand désavantage de ces deux techniques est que les acides nucléiques en solution sont à des concentrations très élevées (dans l'ordre du micro à millimolaire). À ces concentrations, il est possible qu'il y ait formation de G-quadruplexes intermoléculaires. Cela ne récapitule donc pas le scénario se déroulant dans la cellule. Un autre point est que cette technique est surtout utilisée avec des séquences contenant seulement le G4. Des séquences plus longues diminuent grandement le signal associé au G4 et peuvent même, dans certains cas, le masquer totalement.

4.6 Résonance magnétique nucléaire

La résonance magnétique nucléaire (RMN) a plusieurs utilités dans le domaine des G-quadruplexes. Les principales sont la détermination à haute résolution de structure en solution et l'étude des liens chimiques (Adrian et al., 2012). Une des caractéristiques importantes des G-quadruplexes au niveau structural est la présence de protons imines impliqués dans les liaisons Hoogsteen. Dans un spectre RMN proton, ce type de liaison est responsable d'un décalage chimique dans les 10-12 ppm tandis que ceux impliqués dans les liaisons de type Watson-Crick ont plutôt un décalage chimique allant dans les 12-15 ppm (Fay et al., 2017a). Un spectre RMN comportant 12 pics distincts dans la région 10-12 ppm est caractéristique d'un G-quadruplexe à trois tétrades (où 12 guanines sont en interaction dans le corps du G4). Il existe cependant des exceptions, comme dans le cas d'un mélange conformationnel (souvent le cas dans les G4 d'ADN), où plus de 12 pics distincts associés aux protons imine peuvent être observés (Adrian et al., 2012).

La RMN peut aussi être utilisée pour élucider les propriétés structurales de certaines macromolécules. Cela requiert alors de très grandes concentrations en acides nucléiques.

L'analyse de la structure par l'effet *Overhauser* de spectrométrie nucléaire (*Nuclear Overhauser effect spectroscopy*, NOESY) est souvent utilisée pour détecter la distance entre des paires de protons (Mathad et Yang, 2011). Il est alors possible de voir les protons d'une tétrade qui se retrouvent relativement près de ceux d'une autre tétrade. La conformation du G4, soit la position des guanines et l'orientation de leur lien glycosidique (*syn* ou *anti*), peut alors être assignée de façon très précise.

Pour finir, il est aussi possible de résoudre la structure du G-quadruplexe à une très haute résolution. Cela est généralement réalisé à l'aide des spectres 3D protons et carbones. Pour ce faire, il y a premièrement génération de plusieurs structure basées sur les spectres RMN déterminés. Les structures sont par la suite affinées à l'aide de contraintes de dynamique moléculaire. Pour finir, les structures acceptées sont choisies selon leur énergie totale, le nombre de violations, la convergence et la superposition des atomes de deux structures (Adrian et al., 2012). Il est habituellement beaucoup plus facile d'obtenir la structure des tétrades du G4 tandis que la flexibilité présente dans les boucles amène parfois certains problèmes de résolution structurale.

4.7 Cristallographie de rayons X

La cristallographie de rayon X est une méthode utilisée pour la résolution structurale de macromolécules comme les protéines et les acides nucléiques, et ce à un niveau atomique. Le principe général repose sur l'émission de plusieurs rayons X sur le cristal du composé d'intérêt. Celui-ci émet ensuite un patron de diffraction qui est unique à sa structure. Un modèle 3D peut ensuite être réalisé à l'aide du patron (Smyth et Martin, 2000). Ce modèle peut posséder une très grande résolution allant jusqu'aux Angströms (Clark et al., 2012). La difficulté de cette technique se trouve au niveau de la formation du cristal. En effet, lors de l'étude de structure G-quadruplexes, il y a souvent présence de certaines contraintes comme la pureté chimique, l'homogénéité, la topologie, le choix de solvant limité et l'imposition d'une compaction des molécules d'ARN/ADN (Parkinson et Collie, 2019). Plusieurs structures de G-quadruplexes d'ARN et d'ADN ont cependant été résolues grâce à cette technique (pour des exemples consulter : Collie et al., 2010; Parkinson et al., 2002). Par ailleurs, des structures de G4 étant en interaction avec certaines protéines ont tout récemment vu le jour et permettent de mieux comprendre l'interaction entre les deux macromolécules

(Chen et al., 2018). L'avantage de cette technique est qu'il est plus facile d'insérer un ligand dans le cristal (comparativement à la RMN) et de déterminer les interactions chimiques qu'il effectue avec le G-quadruplexe (Schüttelkopf et van Aalten, 2004). Sa haute précision est grandement utilisée pour le développement de potentiels ligands (Micco et al., 2013).

4.8 *Spectrométrie Raman*

La spectrométrie Raman est une autre technique utilisée pour l'étude de la conformation des G-quadruplexes. Celle-ci est plus spécifiquement basée sur l'information des vibrations provenant des molécules. La méthode repose sur la différence de fréquence entre la lumière émise par la source et celle émise après avoir passé par une substance (Yang Li et al., 2018). L'analyse de ce décalage de fréquence permet de trouver certaines propriétés propres au milieu étudié. Chacune des molécules étudiées aura des spectres Raman uniques. Des études ont permis la caractérisation des G-quadruplexes parallèles et antiparallèles (Miura et al., 1995). Les pics associés à chacune des différences vibrationnelles des différentes conformations ont été résolus et confirmés par plusieurs groupes de recherche (Friedman et Terentis, 2016). L'avantage de cette technique, comparativement à d'autres comme la RMN, est qu'elle requiert relativement moins d'échantillons et elle peut se permettre des échantillons moins purs (Yang Li et al., 2018).

4.9 *Essais luciférase*

L'essai luciférase est une technique en cellule utilisée pour comprendre l'effet de la formation d'un G-quadruplexe sur la traduction de son ARNm. La luciférase est un type d'enzyme étant en mesure de produire de la bioluminescence à partir de certaines molécules chimiques (Lee et al., 2019). Deux luciférases sont particulièrement utilisées en biologie moléculaire. La première est la *Renilla* luciférase qui est en mesure de convertir la Coelenterazine en Coelenteramide, et ce en émettant de la lumière (Lorenz et al., 1991). La seconde est la *Firefly* luciférase qui est en mesure de convertir la D-Luciférine en Oxyluciférine (de Wet et al., 1985). Les séquences de ces deux enzymes sont clonées dans un seul plasmide et exprimées en cellules eucaryotes. Il est ensuite possible d'ajouter des séquences en amont ou en aval de l'une des deux luciférases, comme celles provenant des 5'UTR ou bien de 3'UTR et d'étudier l'effet de ces derniers. Dans le cas des G-quadruplexes, les séquences ajoutées contiennent habituellement le G4 ou le G4 muté (G/A mutant). Il est alors possible d'évaluer

l'impact de la présence du G4 sur la quantité totale de protéines luciférase en solution. Le niveau des différentes protéines peut ensuite être mesuré de façon séquentielle avec leur substrat spécifique. La deuxième luciférase quant à elle agit en tant que contrôle (Huang et al., 2011; Lee et al., 2008). En effet, comme elle n'est pas sous contrôle de la séquence insérée, il est possible de se servir de son niveau pour corriger l'efficacité de transfection, la mort cellulaire et la quantité de cellules totale. Cela se fait normalement en effectuant un ratio entre la luciférase étant sous influence d'un insert et la luciférase n'étant pas sous aucune influence.

5 Maladies neurodégénératives

5.1 *ARN et G-quadruplexes associés aux maladies neurodégénératives*

De nos jours, les maladies neurodégénératives (*Neurodegenerative diseases*, ND) prennent beaucoup d'importance dans notre société. Cette augmentation de la prévalence est principalement due au vieillissement de la population. Les ND sont caractérisées par la mort progressive et irréversible d'une population de neurones ou de cellules gliales (Dugger et Dickson, 2017). Les plus connues sont l'Alzheimer (*Alzheimer disease*, AD), la maladie de Parkinson (*Parkinson disease*, PD), la dégénération fronto-temporal (*Frontotemporal degeneration*, FTD) et la sclérose latérale amyotrophique (*Amyotrophic lateral sclerosis*, ALS). Récemment, beaucoup d'études ont tenté de comprendre les mécanismes moléculaires reliés au phénomène de la neurodégénérescence. Certains se sont concentrés sur le rôle de l'ARN au sein de ces maladies (E. Y. Liu et al., 2017). Selon eux, l'implication de l'ARN au sein des ND peut se diviser en cinq grands groupes. Le premier est la perte d'interaction entre certaines protéines pouvant lier l'ARN (*RNA-binding proteins*, RBP) et l'ARN. Un des exemples est la perte de la protéine 43 liant le domaine de l'ARN TAR (*TAR DNA-binding protein 43*, TDP-43) dans le FTD et l'ALS qui a un impact direct sur l'épissage alternatif de certains ARN (Buratti et al., 2001; Wu, Cheng, et al., 2019). Le second groupe est la répétition de microsatellites. Dans ce cas-ci, la formation de foci d'ARN peut souvent être observée dans la cellule. Un exemple est la répétition de l'hexanucléotide GGGGCC se retrouvant dans la région C9ORF72 du chromosome 9 qui est responsable de l'agrégation de

l'ARN au noyau (Wang et al., 2019). Le troisième est la dérégulation de certains microARN. Il y a alors un impact direct sur les différents niveaux d'ARN et la stabilité de ceux-ci en cellule. Dans l'ALS et la FTD, la mauvaise localisation de TDP-43 entraîne une dérégulation de la maturation des miARN. Cela est principalement dû au fait que TDP-43 interagit avec les pri-miARN et la protéine Drosha. La biogenèse des miARN est en effet facilitée par la présence de celui-ci (Kawahara et Mieda-Sato, 2012). Le quatrième groupe est une présence de mutations dans certaines protéines responsables de la maturation des ARN de transfert. L'effet d'une diminution d'activité ou d'une inactivation de ces protéines peut se faire sentir niveau traductionnel. C'est le cas de certains individus dans la maladie de Charcot-Marie-Tooth où des mutations ont été retrouvées dans des aminoacyles ARNt synthétases (Antonellis et al., 2003). Le cinquième est la présence de mutation dans les lncARN. Dans l'AD, le lncARN BACE1-AS est surexprimé en cellule et cause l'augmentation de stabilité de la Beta-sécrétase 1 (BACE1). Ceci qui provoque une plus grande traduction de l'ARNm. BACE1 est responsable de la formation de plaques amyloïdes, il est alors possible d'observer une augmentation des plaques dans la cellule (Faghihi et al., 2008).

L'implication de certains G-quadruplexes dans les maladies neurodégénératives a récemment commencé à être étudiée (Simone et al., 2015). Par exemple, il a été montré que l'expansion de l'hexanucléotide dans la région C9orf72 forme une population mixte de G4 parallèles et antiparallèles dans l'ADN (Šket et al., 2015). Il a aussi été montré que les G4 sont en mesure de se former au niveau de l'ARN de C9orf72. Ceux-ci pourraient même être liés par certaines RBP (Conlon et al., 2016; Mori et al., 2013; Wang et al., 2019). Par ailleurs, l'interaction protéines : G4 est impliquée dans certaines maladies neuronales. FMRP, l'une des protéines les plus connues pour lier les G4 d'ARN est impliquée dans certaines maladies neuronales comme le syndrome du X-Fragile (Melko et Bardoni, 2010). Dans le neurone, cette protéine est en effet responsable du transport axonal des ARNm et participe aussi à la répression traductionnelle, et ce par la liaison des G4 de certains ARNm (Darnell et al., 2011; Subramanian et al., 2011). Il est possible de trouver des G-quadruplexes présents dans tous les cinq groupes d'ARN précédemment mentionnés.

5.2 *Maladie de Parkinson*

5.2.1 *Symptômes*

La maladie de Parkinson est l'une des maladies où plusieurs efforts sont concentrés pour comprendre les mécanismes reliés à sa pathologie. Cette maladie touche une grande partie de la population mondiale. En Amérique du Nord et en Europe, la maladie de Parkinson a une incidence de 9-19/100 000 habitants par année, une prévalence de 30-190/100 000 habitants, une moyenne d'âge au moment du diagnostic de 61 ans et cette maladie touche en moyenne les patients pendant une durée de 13 ans (Ellison et al., 2013). Bien que certaines formes de la maladie soient héréditaires, la plupart des cas sont sporadiques. Ils sont donc associés à de nouvelles mutations dans le génome du patient (Selvaraj et Piramanayagam, 2019).

Il y a principalement trois symptômes associés à la PD. Ceux-ci sont une bradykinésie, c'est-à-dire une lenteur inégale au niveau des mouvements, des tremblements au repos et une rigidité musculaire. L'ensemble de ces symptômes doit être présent pour qu'un patient soit diagnostiqué comme étant atteint de la PD (Reich et Savitt, 2019). Il existe plusieurs types de traitements pour la PD, mais ceux-ci ciblent seulement les symptômes de la maladie. Ils amènent un ralentissement de l'apparition du processus neurodégénératif, mais sont incapables de renverser le processus. Certains traitements augmentent la concentration de dopamine intracérébrale tandis que d'autres viennent moduler la stimulation des récepteurs dopaminergique (Kalia et Lang, 2015). La maladie de Parkinson est donc une maladie incurable.

Au niveau pathologique, il est possible d'observer ce qu'on appelle un corps de Lewy situé dans le corps ou dans les neurites des cellules dopaminergiques du *Substantia Nigra* (SN). Ce corps de Lewy est caractérisé par une agrégation d'une forme anormalement repliée d'une protéine appelée α -synucléine (aussi présent dans la maladie de Lewy). Le SN est la région du cerveau responsable de la modulation des mouvements et de la fonction de récompense (Chinta et Andersen, 2005; Lammel et al., 2014). Il est donc facilement possible de faire le lien entre la zone affectée et les symptômes associés. Par ailleurs, une perte de pigmentation peut être observée au niveau du *Substantia Nigra*. Cela est principalement causé par la

dégénérescence des neurones dopaminergiques. La pigmentation est initialement associée à l'accumulation de neuromélanine (Zecca et al., 2001).

5.2.2 Causes moléculaires

Il a été montré que dans 20 à 30% des cas de Parkinson, la maladie était reliée à des causes génétiques. Celles-ci sont soit dominantes ou récessives (Ellison et al., 2013). Les mécanismes moléculaires reliés au Parkinson peuvent être divisés en plusieurs types de dysfonction en fonction de quelles organelles ou voie de signalisation ils touchent. Ces types de dysfonction sont ceux du protéasome, de la mitochondrie, de l'autophagie, du repliement de protéines et de la voie de signalisation de l'insuline (Figure 6). Ultimement, toutes ces mutations vont entraîner l'augmentation du stress oxydatif dans la cellule et la mort irréversible des neurones dopaminergiques. Ce stress oxydatif cause une augmentation des radicaux libres dans la cellule et entraîne une perte de l'homéostasie de la phosphorylation oxydative dans la membrane de la mitochondrie (Dias et al., 2013).

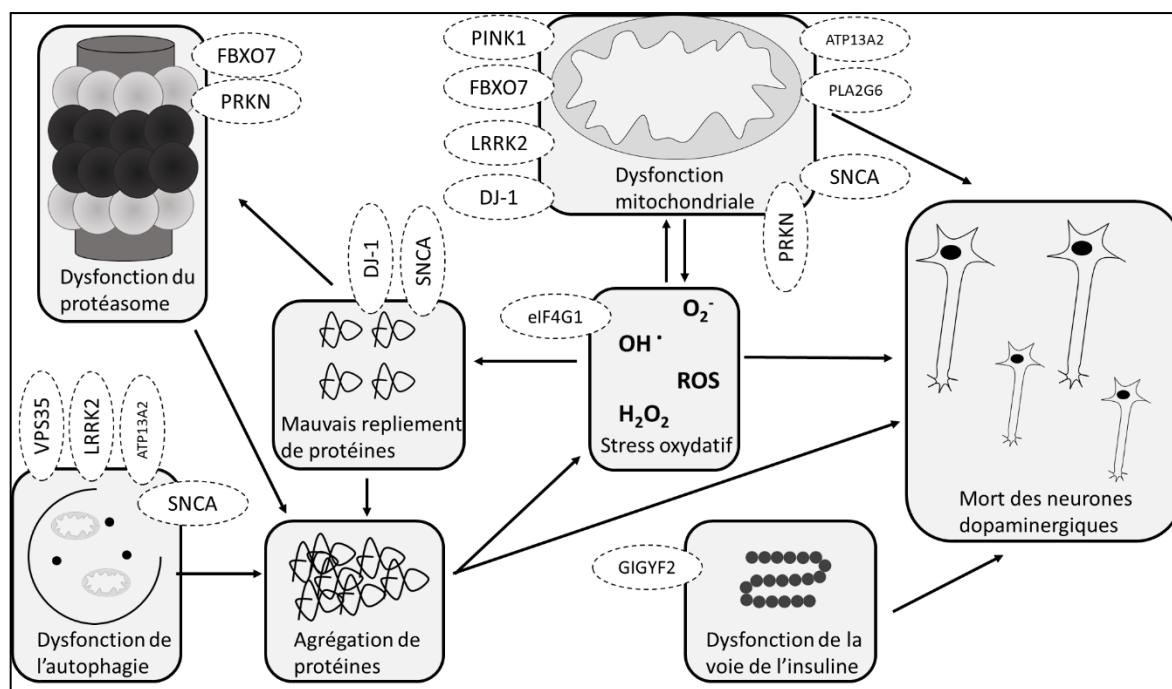


Figure 6 : Dysfonctions souvent impliquées dans la pathologie de Parkinson.

Les flèches représentent les relations entre différentes dysfonctions. Le nom de gènes importants pour lesquels des mutations ont souvent été rapportées dans la maladie de Parkinson est encadré et relié à leur type de dysfonction.

5.2.2.1 Dysfonction du protéasome

Une anomalie dans le fonctionnement du protéasome est souvent responsable de la mort cellulaire chez les patients atteints du Parkinson. Le protéasome est un complexe protéique responsable de la dégradation des protéines. La plupart des protéines sont reconnues par celui-ci grâce à un signal d'ubiquitination (Tanaka, 2009). Lorsque certaines mutations touchent les mécanismes d'ubiquitinations ou bien l'efficacité du protéasome en lui-même, une accumulation de protéines se produit dans la cellule. Cette accumulation crée un phénomène de toxicité et entraîne la mort des neurones dopaminergiques (Kalia et al., 2013). Un exemple de protéine grandement mutée dans la PD est la protéine ubiquitine ligase Parkin RBR E3 (*Parkin RBR E3 Ubiquitin Protein Ligase*, PRKN). Celle-ci est normalement responsable de l'ubiquitinylation de certaines protéines en cellule. Son inactivation ou une diminution de son activité amène une diminution de la quantité totale de protéines ubiquitinylées, l'accumulation de celles-ci et, par le fait même, l'agrégation de certaines protéines en cellule, comme l' α -synucléine (Snyder et al., 2003). Un autre exemple est la protéine F-Box 7 (*F-Box Protein 7*, FBXO7) qui est une protéine dont les mutations ont un impact direct sur les fonctions du protéasome. En effet, celle-ci est une sous-unité du complexe « cullin-1-based SCF ». Ce complexe participe avec d'autres protéines à la formation d'une E3 ubiquitine ligase. Sa fonction pathologique dans la PD est encore une fois basée sur une diminution des protéines étant acheminées au protéasome dû à une réduction de l'ubiquitinylation (Kirk et al., 2008; Vingill et al., 2016).

5.2.2.2 Dysfonction de l'autophagie

L'accumulation de vacuoles autophagiques est observée dans le cerveau de patients atteints de la maladie Parkinson (Anglade et al., 1997). L'accumulation de celles-ci est une caractéristique morphologique importante de cellules en apoptoses (González-Polo et al., 2005). Par ailleurs, l' α -synucléine (SNCA) a été montrée comme étant dégradé aussi par autophagie (Webb et al., 2003). La mutation de SNCA et son agrégation aurait un impact direct sur la macroautophagie dans la cellule en augmentant la quantité de structure vésiculaire autophagique (Stefanis et al., 2001). Une seconde protéine souvent mutée dans la PD est la déglycase associée au Parkinsonisme (*Parkinsonism Associated Deglycase*, DJ-1). Cette protéine est impliquée dans la polarisation mitochondriale, la fragmentation

mitochondriale et l'accumulation de marqueurs autophagiques dans les neurones dopaminergiques (Thomas et al. 2011). Une dernière protéine importante étant souvent mutée dans les dysfonctions de l'autophagie dans la maladie de Parkinson est la protéine 35 associée aux protéines vacuolaires sortantes (*Vacuolar protein sorting-associated protein 35*, VPS35). Celle-ci fait partie du complexe *WASH* (un groupe de protéines faisant partie du rétromère) et est responsable du transport entre les endosomes vers la membrane plasmique et le trans-golgi (Williams, Chen, et Moore 2017). Dans les cas du PD où VPS35 est mutée, l'autophagie serait négativement affectée dans la cellule au niveau de la formation des autophagosomes (Zavodszky et al. 2014).

5.2.2.3 Dysfonction du repliement des protéines

Une protéine importante impliquée dans l'agrégation de protéines est l' α -synucléine. Il s'agit d'une protéine normalement responsable dans les vésicules synaptiques (Al-Wandi et al., 2010). Plusieurs types de modifications dans cette protéine peuvent être négatives pour la cellule. Par exemple, une oxydation plus prononcée de l' α -synucléine est observée dans la PD. Cette oxydation augmente directement la probabilité d'agrégation de la protéine (Giasson et al., 2000). La phosphorylation de l' α -synucléine augmente aussi ses chances de s'agréger en cellule (Koukouraki et Doxakis, 2016). Ces agrégats de protéines sont par la suite ciblés par le protéasome. Cependant, comme mentionné plutôt, le protéasome est souvent touché dans la maladie de Parkinson. Il se produit alors une accumulation de la protéine en cellule et création d'un corps de Lewis (Dauer et Przedborski, 2003). Au niveau fonctionnel, ces oligomères ont été montrés comme ayant un impact sur la dégradation protéique globale de la cellule et sur la fonction de plusieurs organelles comme le réticulum endoplasmique et la mitochondrie (Ingelsson, 2016).

5.2.2.4 Dysfonction mitochondriale

Des anomalies dans les complexes de la chaîne respiratoire mitochondriale ont été détectées chez les patients atteints de la PD (Greenamyre et al., 2001). Ceux-ci ont pour impact d'augmenter le stress oxydatif dans la cellule et amènent ainsi une perte d'énergie produite par les mitochondries. L'inhibition du complexe est donc responsable d'une augmentation des radicaux super oxyde en cellule. Ces molécules peuvent ensuite entraîner des dommages

sur les protéines, les lipides, l'ADN et même sur la chaîne respiratoire en elle-même (Cohen, 2000). Les dommages sur la chaîne respiratoire entraîneraient davantage de mitochondries endommagées et une augmentation globale des espèces réactives à l'oxygène (*Reactive oxygen species*, ROS) en cellule. Un cercle vicieux pourrait alors être observé en cellule. Comme mentionné plutôt, ces oxydations contribuent directement à l'augmentation de l'agrégation de l' α -synucléine. Des mutations dans la protéine kinase induite par PTEN (*PTEN-induced kinase 1*, PINK1) sont souvent impliquées dans les mécanismes reliés au Parkinson. En effet, au niveau de la mitochondrie, celle-ci est responsable de l'activation de PRKN par phosphorylation lors d'une dépolarisation de la membrane de la mitochondrie. PRKN peut à son tour ubiquitinyler la mitochondrie pour entraîner sa mitophagie (Gladkova et al., 2018; Sliter et al., 2018; Wauer et al., 2015). Une mutation dans un de ces deux gènes diminue donc la mitophagie dans la cellule et contribue à l'accumulation de mitochondries endommagées. Encore une fois, FBXO7 exerce un rôle sur l'ubiquitination de la mitochondrie (Zhou et al., 2018). Une autre protéine dérégulée dans le PD est la kinase 2 riche en répétition leucine (*Leucin-rich repeat kinase 2*, LRRK2). Cette protéine est localisée à la mitochondrie et affecte sélectivement le complexe 4 de la chaîne respiratoire. L'intégrité de la mitochondrie et des problèmes dans la chaîne respiratoire peuvent alors se faire ressentir. (Aufschnaiter et al., 2018).

5.2.2.5 Dysfonction dans la signalisation de l'insuline

Pour finir, il est possible d'observer une diminution de l'activité des voies de signalisation reliées à l'insuline chez les patients atteints de la PD. En effet, une diminution des récepteurs de l'insuline dans le *Substantia Nigra* est observée chez les patients atteints comparativement à des individus de même âge sains. Cette diminution de récepteur est directement associée à une augmentation de la résistance à l'insuline (Morris et al., 2014). Ce phénomène serait relié à une augmentation de la phosphorylation de certaines protéines pro-insuline, ce qui aurait pour impact de désactiver la signalisation de l'insuline (Moroo et al., 1994). Des études ont par ailleurs montré que des souris nourries à un régime riche en gras ont une dysfonction dans la signalisation de l'insuline et sont victimes d'une augmentation de la vulnérabilité vis-à-vis le 1-méthyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), un agent chimique pouvant récapituler la PD dans les neurones dopaminergiques. Une augmentation du phénomène de

neurodégénération dans les neurones dopaminergique est alors observée au niveau du *Substantia Nigra* (Choi et al., 2005; Morris, Bomhoff, et al., 2010). Une seconde évidence de cette dérégulation est l'implication d'une mutation Arg610Gly dans la protéine GYF interagissant avec GRB10 (*GRB10 Interacting GYF Protein 2*, GIGYF2) qui est responsable du recrutement du facteur de croissance de l'insuline 1 (*Insuline growth factor 1*, IGF-1) et du récepteur à l'insuline. Celle-ci a été identifiée comme un variant génétique dans la PD et suggère aussi une dérégulation dans la voie de l'insuline (Giovannone et al., 2003; Ruiz-Martinez et al., 2015).

6 Hypothèse/problématique

Les mécanismes de régulation post-transcriptionnelle sont des thématiques de recherche émergentes en biologie moléculaire. La compréhension de ces régulations par les G-quadruplexes d'ARN de gènes dérégulés dans les maladies neurodégénératives pourrait donner des indices sur les mécanismes affectés lors de certaines maladies comme la maladie de Parkinson. Par ailleurs, les différents traitements ciblant certains G-quadruplexes sont à ce jour non spécifiques. Notre hypothèse est qu'il serait possible de moduler le niveau protéique de gènes impliqués dans la maladie de Parkinson, et ce par la modulation de la liaison entre des protéines et des G-quadruplexes situés dans des régions 5'UTR. L'objectif générale de ce projet est l'étude de protéines liant les G-quadruplexes situés dans les régions 5'UTR de gènes impliqués dans la maladie de Parkinson. Pour accomplir ce travail, nous proposons comme objectifs spécifiques d' :

- Identifier et étudier des G-quadruplexes d'ARN impliqués dans des gènes associés à la maladie de Parkinson;
- Identifier des protéines étant en mesure de lier ces G-quadruplexes; et,
- Caractériser l'interaction entre les protéines et les G-quadruplexes d'ARN trouvés; et de,
- Moduler la formation du complexe protéine-G4.

Ces objectifs nous permettront de mieux comprendre l'importance des G-quadruplexes au sein de la régulation post-transcriptionnelle.

RÉSULTATS

Article « Guanine Nucleotide-Binding Protein-Like 1 (GNL1) binds RNA G-quadruplex structures in genes associated with Parkinson's disease »

Auteurs de l'article: Marc-Antoine Turcotte, Jean-Michel Garant, Hélène Cossette-Roberge et Jean-Pierre Perreault

Statut de l'article: Soumis à *Journal of RNA biology* le 20 Août 2020

Avant-propos: Marc-Antoine Turcotte a écrit la totalité de l'article et celui-ci a été révisé et corrigé par le directeur de recherche, Jean-Pierre Perreault. Marc-Antoine Turcotte a réalisé la presque totalité des expériences, mis à part les analyses sur G4 RNA *Screeners* (Jean-Michel Garant) et les expériences de fluorescence au NMM sur les 11 pG4 (Hélène Cossette-Roberge). Le projet de recherche s'est échelonné sur les deux ans de maîtrise.

Résumé :

Au niveau post-transcriptionnel, les ARN sont hautement régulés dans les maladies neurodégénératives et la présence de quelques mutations peut affecter le sort des cellules neuronales. À ce jour, l'impact de la régulation des G-quadruplexes (G4) dans certaines maladies neurodégénératives, comme la maladie de Parkinson (PD), n'a jamais été analysé. Dans cette étude, nous avons initialement identifié *in silico* des G4 potentiels (pG4) présents dans les gènes dérégulés liés au système nerveux et montré qu'il y existait un enrichissement significatif en pG4. Ensuite, il a été démontré par des tests biochimiques que plusieurs des séquences G4 récupérées dans la région 5' non traduite (5'UTR) d'ARNm associés à la maladie de Parkinson se repliaient en G4 *in vitro*. Le clonage de la longueur totale du 5'UTR de ces candidats en amont d'un système rapporteur de luciférase a permis de montrer que le G4 de la protéine ubiquitine ligase Parkin RBR E3 (PRKN) et de la protéine 35 associée aux protéines vacuolaires sortantes (VPS35) réprimait de manière significative la traduction des deux gènes dans les cellules SH-SY5Y. Par la suite, une stratégie de purification par affinité en utilisant l'une ou l'autre de ces deux séquences G4 comme appâts a isolé la protéine Guanine Nucleotide-Binding Protein-Like 1 (GNL1). Cette dernière s'est avérée avoir une affinité plus élevée pour les séquences sauvages G4 que mutantes. Cette étude met en lumière un nouveau mécanisme moléculaire important pour la maladie de Parkinson et propose de nouvelles avenues pour moduler la traduction des maladies.

Guanine Nucleotide-Binding Protein-Like 1 (GNL1) binds RNA G-Quadruplex structures in genes associated with Parkinson's disease

Marc-Antoine Turcotte, Jean-Michel Garant, H el ene Cossette-Roberge, and Jean-Pierre Perreault*

Department of Biochemistry, Pavillon de recherche appliqu ee sur le cancer,
Universit e de Sherbrooke, Sherbrooke, Qu ebec, Canada, J1E 4K8

*To whom correspondence should be addressed:

Email: jean-pierre.perreault@usherbrooke.ca

Telephone Number: 819-821-8283

Fax Number: 819-564-5340

ORCID ID: 0000-0002-4559-5541

Running Title: RNA G4s as translational regulators

Key words: G-quadruplex, translation, RNA binding proteins, transcriptome, neurosciences

List of abbreviations: CDS, Coding sequence; EMSA, electrophoretic mobility shift assay; G4, G-Quadruplex; GNL1, Guanine Nucleotide-Binding Protein-Like 1; HGNC, HUGO Gene Nomenclature Committee; K_d , dissociation constant; KEGG, Kyoto Encyclopedia of Genes and Genomes; ND, Neurodegenerative diseases; NMM, N-Methyl Mesoporphrine; PD, Parkinson's disease; pG4, Potential G4; PRKN, Parkin RBR E3 Ubiquitin Protein Ligase; UTR, Untranslated region; VPS35, Vacuolar Protein Sorting-Associated Protein 35

ABSTRACT

RNAs are highly regulated at the post-transcriptional level in neurodegenerative diseases and just a few mutations can significantly affect the fate of neuronal cells. To date, the impact of G-quadruplex (G4) regulation in neurodegenerative diseases like Parkinson's disease (PD) has not been analyzed. In this study, *in silico* potential G4s located in deregulated genes related to the nervous system were initially identified and were found to be significantly enriched. Several G4 sequences found in the 5' untranslated regions (5'UTR) of mRNAs associated with Parkinson's disease were demonstrated to in fact fold *in vitro* by biochemical assays. Subcloning of the full-length 5'UTRs of these candidates upstream of a luciferase reporter system led to the demonstration that the G4s of both Parkin RBR E3 Ubiquitin Protein Ligase (PRKN) and Vacuolar Protein Sorting-Associated Protein 35 (VPS35) significantly repressed the translation of both genes in SH-SY5Y cells. Subsequently, a strategy of using label-free RNA affinity purification assays with either of these two G4 sequences as bait isolated the Guanine Nucleotide-Binding Protein-Like 1 (GNL1). The latter was shown to have a higher affinity for the G4 sequences than for their mutated version. This study sheds light on a new molecular mechanism important in Parkinson's disease, and proposes new avenues for modulating disease.

INTRODUCTION

G-quadruplexes (G4) are non-canonical secondary structures found in both DNA and RNA molecules. In cells, as well as *in vitro*, four guanine residues can be linked together by Hoogsteen bonds in order to form a structure called a G-quartet that can be stabilized by the presence of a monovalent cation, usually potassium (Rouleau *et al.* 2017b). Canonical G4 structures respect the motif G₃-N₁₋₇-G₃-N₁₋₇-G₃-N₁₋₇-G₃ (where N indicates A, U, C or G). However, some studies have shown that non-canonical G4s may have longer loops, only two quartets, or the presence of bulges in G-runs (Bolduc *et al.* 2016; Jodoin *et al.* 2014; Xiao *et al.* 2018). While DNA G4s can be present in many conformations, RNA G4s, which are more stable than their DNA counterparts, are restricted to folding into parallel conformations (Małgowska 2016).

Studies on G4s have demonstrated that they exhibit many regulatory functions in cells. For example, DNA G4s have been demonstrated to be implicated in DNA replication, transcription and in telomere elongation (Kim, 2019; Lerner et Sale, 2019; Neidle et Parkinson, 2003). In mRNA, G4s have been located in the 5' untranslated region (UTR), the coding sequence (CDS) and the 3'UTR (Rouleau *et al.* 2017b). G4s located in the 5'UTR have been shown to impair ribosome scanning, leading to repression of the cap-dependent translation (Beaudoin et Perreault, 2010; Bugaut et Balasubramanian, 2012). However, some G4s have been also demonstrated to enhance cap-independent translation by the stabilization of internal ribosome entry sites (Jodoin *et al.* 2019). On the other hand, CDS G4s have been shown to impair translational elongation and to induce frameshifting (Endoh *et al.* 2013; Yu *et al.* 2014). Finally, G4s located in the 3'UTR have been shown to be associated with both translational repression and the regulation of polyadenylation (Arora et Suess, 2011; Beaudoin et Perreault, 2013). Furthermore, G4s were found to have an impact on other types

of RNAs, like pre-miRNA, where they have been shown to regulate RNA processing (Rouleau *et al.* 2017a). Lastly, in long non-coding RNA, they often sequester G4 RNA binding proteins (Matsumura *et al.* 2017).

It has been proposed that RNA G4s are globally unfolded in eukaryotes (Guo *et al.* 2016). Since they are known to be relatively stable structures, intracellular mechanisms must regulate their formation. Yet only a few proteins have been shown to be able to fold, unfold or stabilize G4s (Mendoza *et al.* 2016). Examples of these RNA binding proteins are the G-Rich RNA Sequence Binding Factor 1 (GRSF1), a protein known to unfold RNA G4s and to facilitate degradosome-mediated decay, and nucleolin, a protein known to stabilize RNA G4s (Masuzawa *et al.* 2020; Pietras *et al.*, 2018). Since each G4 function is thought to be regulated by specific binding proteins, targeting them appears to be an attractive avenue for controlling the associated post-translational events (Sun *et al.* 2019).

Over the years, several G4 motifs have been associated with important mechanisms in neurological diseases (Ishiguro *et al.*, 2016; Koukouraki *et al.* 2016). For example, the RNA foci C9orf72, which forms multiple repeats of G4s, has been linked to both frontotemporal dementia and amyotrophic lateral sclerosis (Schludi *et al.* 2018). A mutated miRNA associated with Alzheimer's disease has been demonstrated to fold into a G4 (Imperatore *et al.* 2020). In fragile X syndrome, fragile X mental retardation protein, a protein known to both bind to mRNAs containing a G4 structure and to regulate their transport, was shown to be less abundant. This reduction also directly affects the translation of several other mRNAs (Melko *et al.* 2010). In the present study, the original identification of the Guanine Nucleotide-Binding protein-like 1 (GLN1) as an RNA binding protein (RBP) that binds to the G4 structures located in the 5'UTRs of both VPS35 and PRKN, two genes that are deregulated in Parkinson's disease, is reported.

MATERIALS AND METHODS

Bioinformatic analyses

In order to retrieve genes associated with nervous system diseases, BRITE hierarchy files (br08402) were downloaded from the disease data-oriented entry points of the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.* 2015). The gene symbols from each line associated with the nervous system were conserved and then transformed into a list file. Next, the symbols were corrected using the Multi-symbol checker of the HUGO Gene Nomenclature Committee (HGNC) (Braschi *et al.* 2019). Then, the RNA.fa.gez files were downloaded from Refseq (NCBI, March 2018), and each gene from the HGNC list was associated with a Fasta transcript sequence. The Fasta files were then submitted to the G4 RNA Screener software using the default parameters, and all windows with a G4NN score >0.5 were considered as being positive (Garant *et al.* 2017). Finally, pG4 windows were annotated by their transcript position (5'UTR, CDS and 3'UTR) using the RefSeq CDS annotation "ncbiRefSeqCds.txt.gz "(May 2018).

For the purpose of comparing the KEGG results to other databases, tab-separated values files were retrieved from the Human Protein Atlas for the enriched neuronal genes. They were then associated with a Fasta transcript file and submitted to the G4 RNA screener. Genes with a G4NN score >0.5 were considered as being positive.

Design of the oligonucleotides

Oligonucleotides corresponding to the potential G4 sequences were designed to include both upstream and downstream wild type flanking sequences of up to 20 to 25 nucleotides (nts) in length. G/A mutants (several Gs substituted by As) were also synthesized for each WT G4 sequence (see in Tables S1 and S2 for sequence details). When required for *in vitro*

transcription, a T7 promoter sequence (5'-TAATACGACTCACTATAG_{G₁₋₃}-3', where G₁₋₃ indicates a stretch of 1 to 3 consecutive guanines) was added to the sense strand. All oligonucleotides were purchased from Biobasic.

***In vitro* transcription**

In order to reconstitute the full-length WT G4 and G/A mutant sequences (which are 70-90 nts in size), DNA templates were synthesized via a PCR filling strategy using both forward and reverse primers at a concentration of 2 mM. Then, purified Pfu DNA polymerase (2 μ L) was added to a final volume of 100 μ L containing 2 mM MgSO₄, 0,2 mM dNTPs, 10 mM (NH₄)₂SO₄, 10 mM KCl, 20 mM Tris-HCl (pH 8.8), 0,1 % Triton-X-100 and 5 % DMSO. The PCR program used was 12 cycles of 95°C (1 min), 54°C (1 min) and 72°C (1 min), and was terminated with a final elongation of 5 min at 72°C. The DNA was then ethanol precipitated and dissolved in 10 μ L of ultrapure water. DNA integrity and size were then verified by 2% agarose gel electrophoresis.

In order to synthesize RNA *in vitro*, 100 μ L reactions containing 5 μ L of the PCR filling DNA reactions, 2 μ L of purified T7 RNA polymerase, 0.01 U of pyrophosphatase (Roche Diagnostics) and final concentrations of 80 mM HEPES-KOH (pH 7.5), 24 mM MgCl₂, 5 mM NTPs, 2 mM spermidine and 40 mM DTT were prepared and then were incubated at 37°C for 3 h. After this incubation, a DNase treatment (5U, Promega) was performed for 20 min at 37°C. The proteins were then removed by phenol/chloroform extraction, and the RNA ethanol precipitated. The RNA samples were dissolved in 30 μ L ultrapure water, mixed with 60 μ L of loading buffer (94 % formamide, 0.1 % EDTA and 0.025% w/v each of bromophenol blue and xylene cyanol) and were then purified using denaturing (8 M urea) 5% polyacrylamide gels (19:1). The RNA bands of the appropriate

sizes were cutted and nucleic acids eluted overnight in a solution that contained 1 mM EDTA, 0.1 % SDS and 0.5 M ammonium acetate, ethanol precipitated, dried and dissolved in 100 μ L of ultrapure water. For the transcriptions of both pMA-RQ VPS35-4xS1m and pMA-RQ PRKN-4xS1m, the plasmids were digested with *EcoR1* and *HindIII* prior to the transcription reaction.

N-Methyl Mesoporphyrin (NMM) fluorescent assays

With the aim of probing G4 formation, fluorescence assays using NMM were performed as described previously (Jodoin et Perreault, 2018). Briefly, either the WT or a G/A mutant G4 RNA (200 pmol) was added to folding buffer (20 mM Li-cacodylate (pH 7.5) and 100 mM of either LiCl or KCl) in a final volume of 50 μ L. The reactions were then heated to 70°C and allowed to slowly cool for 1 h at room temperature. Next, 50 μ L of 2X *in-line* probing buffer (40 mM Li-cacodylate (pH 8.5), 40 mM MgCl₂ and 200 mM of either LiCl or KCl) were added, yielding a final volume of 100 μ L. The NMM ligand (1 μ L at a concentration of 0.5 mM; N-Methyl-Mesoporphyrin IX, NMM580, Frontier Scientific Inc., Logan, Utah) was then added and the reaction incubated for 30 min at room temperature while being protected from light. The fluorescence intensity was monitored using an Hitachi F-2500 fluorescence spectrophotometer with an excitation wavelength of 399 nm and the emission wavelength being measured between 500 nm and 650 nm in a 10 mm quartz cuvette. The fluorescence at 605 nm was used for quantification. All NMM assays were performed at least in duplicate.

Dephosphorylation and radioactive labeling

The Antarctic Phosphatase kit (New England Biolabs) was used to prepare radioactive transcripts by incubating the RNA samples (50 pmol) at 37°C for 45 min. Upon completion

of the reactions, the phosphatase was inactivated at 70°C for 5 min and the reaction was then immediately chilled on ice. A fraction of the dephosphorylated RNA (10 pmol) was then incubated at 37°C for 1 h with the T4 Polynucleotide Kinase (PNK) kit (Promega) in the presence of [$\gamma^{32}\text{P}$]ATP (2 μL ; 6 000 Ci (222 TBq)/mmol in 50 mM Tricine (pH 7.6), PerkinElmer). Two volumes of loading buffer (94 % formamide, 0.1 % EDTA and 0.025% (w/v) each of bromophenol blue and xylene cyanol) were then added and the resulting samples were then electrophoresed through denaturing (8 M urea) 5% polyacrylamide gels (19:1). The bands of appropriated sizes were cut out of the gels, and the transcripts were then eluted overnight in a solution containing 1 mM EDTA, 0.1 % SDS and 0.5 M ammonium acetate. The eluted transcripts were then ethanol precipitated and the resulting pellets dissolved in 20 μL of ultrapure water.

In-line probing

In order to confirm G4 folding, *in-line* probing was performed as described previously (Beaudoin et al., 2013; Jodoin et Perreault, 2018). Briefly, either WT or G/A mutant G4 RNA labeled with ^{32}P (50 000 counts per minute (CPM) per reaction) was incubated in folding buffer (20 mM Li-Cacodylate (pH 7.5) and 100 mM of either LiCl or KCl) in a final volume of 10 μL for 5 min at 70°C before being slow cooled for 1 h at room temperature. Subsequently, 50 μL of 2X *in-line* buffer (40 mM Li-Cacodylate (pH 8.5), 40 mM MgCl_2 and 200 mM of either LiCl or KCl) and 40 μL of ultrapure water were added to the mixtures before incubating them for 40 h at room temperature. The RNAs were then ethanol precipitated in the presence of glycogen, ethanol washed, dried and dissolved in 20 μL of loading buffer (94 % formamide, 0.1 % EDTA and 0.025% (w/v) of xylene cyanol). The nucleotide ladder was prepared by incubating 25 000 CPM of ^{32}P -labeled RNA for 1 min at

room temperature with 1 μ L of 2 N NaOH. The reactions were then neutralized by adding 3 μ L of 1 M Tris-HCl (pH 7.5), and the RNA ethanol precipitated in the presence of glycogen before being dissolved in 20 μ L of loading buffer (94 % formamide, 0.1 % EDTA and 0.025% (w/v) of xylene cyanol). For the RNase T1 ladder, both WT and G/A mutant G4 RNAs were incubated for 2 min at 37 °C with 0.6 U of T1 RNase (Roche Diagnostic) in 20 mM Tris-HCl (pH 7.5), 10 mM MgCl₂ and 100 mM LiCl buffer. The reactions were then stopped by the addition of 20 μ L of formamide dye loading buffer which contained only xylene cyanol as colourant. Equal amounts of CPMs for each condition and ladder were used, with the exception of the T1 ladder for which three times less CPM were used. The reactions were then adjusted to 10 μ L by addition of formamide loading dye buffer and then electrophoresed through denaturing (8 M urea) 10% polyacrylamide gels. The resulting gels were dried for 50 min at 60°C and were then exposed to a phosphoscreen for 16 h. The results were visualized using a Typhoon Trio imaging system (GE Healthcare), and the band quantification was performed using the SAFA semi-automated software (Laederach *et al.* 2008). The hydrolysis ratio for each nucleotide position is represented by the band intensity in the presence of KCl divided by that in the presence of LiCl. All experiments were performed at least in duplicate.

Cell cultures

SH-SY5Y cells (ATCC, CRL-2266) were cultivated in Dulbecco's modified Eagle medium/F12 with Glutamax (Gibco) supplemented with 10% fetal bovine serum (FBS; Wisent). Human Kidney Embryo 293T (HEK293T) cells (ATCC, CRL-3216) were cultivated in DMEM (Multicell) supplemented with 10 % FBS. When the cells reached a

confluence of 80-90 % they were subcultured at dilutions of 1:2 and 1:20, respectively. All cultures were incubated at 37°C with a 5 % CO₂ atmosphere.

Molecular cloning

The PsiCHECK-2 plasmids used in the luciferase assays were prepared as follows. Puc57 plasmids containing either the full-length WT or G/A mutant 5'UTR G4 sequences of VPS35, LRRK2, SNCA and PRKN were purchased from Biobasics (see Table S1 for the full sequences). The 5'UTR sequences were codigested with *SpeI* and *SalI*, and were then ligated upstream of the Renilla luciferase in a modified version of the PsiCHECK-2 plasmid (Figure S2) that had previously been digested with the same restriction enzymes. The insertion was then confirmed by PCR, and the plasmids were sequenced at the Plateforme de Séquençage of Université Laval.

The PsiCHECK-2 plasmids prepared previously were used as templates in PCR reactions in order to generate G4 regions containing *EcoRI* and *BamHI* restriction sites located in 5' and 3', respectively. The resulting DNA templates were codigested with *EcoRI* and *BamHI* and ligated into pMA-RQ plasmid containing the 4xS1m aptamer (Leppek et Stoecklin, 2014). The insertion was confirmed by PCR, and all constructs were sequenced at the Plateforme de Séquençage of Université Laval.

Luciferase assay

SH-SY5Y and HEK293 cells were subcultured at densities of 80 000 and 22 500 cells per cm² 24 h, respectively, before transfection. They were then transfected in triplicate with Lipofectamine 2000 as described by the manufacturer (Thermo Fisher Scientific) using 500 ng of pUC19 and 50 ng of PsiCHECK-2 in each well of 24 well plates. A day later, the

medium was removed and the cells were washed once with PBS. The cells were then dislodged and suspended by scraping in the presence of 500 μ L of PBS. Four hundred microliters of the resulting suspension were then used to extract proteins, while the remaining 100 μ L were used to extract total RNA. Both fractions were centrifuged at 4°C for 10 min at 2400 x g and the supernatants removed. The cell pellets for RNA isolation were frozen at -80°C until treated, while 100 μ L of protein lysis buffer (Dual-luciferase reporter assay system, Promega) was added to the cell pellets destined for protein extraction. In order to lyse the cells, the solutions were then incubated for 15 min at room temperature. Only a fraction (5 μ L) of the resulting solution was then used for the luciferase assays as suggested by the manufacturer's protocol. Luminescence was then monitored on a Glomax 20/20 luminometer with an integration time of 5 sec. The ratios of the *Renilla/Firefly* luciferases were then calculated and presented as the means and standard deviation of 3 experiments.

RNAs were extracted with Trizol according to the manufacturer's protocol (Thermo Fisher Scientific). When required, the RNAs were quantified by digital RT-PCR at the Laboratoire de Génomique Fonctionnel de Université de Sherbrooke.

G4 ligands

SH-SY5Y cells at a confluence of 70% were incubated for 24 h with either 10 μ M of BRACO19 or 5 μ M of Phen-DC3. Ultrapure water was used as negative control for BRACO19 and DMSO for Phen-DC3. After the 24 h were complete, the cells were washed once with 1X PBS and were then harvested by centrifugation for 5 min at 600 x g. The resulting supernatants were discarded and the cells were incubated with 20 μ L of KCl-lysis buffer (50 mM Tris-HCl (pH 7.5), 100 mM KCl, 50 mM NaCl, 1 mM DTT, 0,1 % Triton X-100, 10 % glycerol and 1 mM MgCl₂) for 30 min at 4°C. The cellular lysates were then

centrifuged at 17 000 x g for 20 min at 4°C, and the resulting supernatants were kept. The protein concentrations were determined by the method of Bradford (BIO-RAD). Western blots were then performed as described previously (Jodoin *et al.* 2019) in order to monitor the expression of VPS35 and PRKN. Anti-PRKN (ab77924; Abcam), anti-VPS35 (ab226180; Abcam) and anti- β -Actin (A5441; Sigma-Aldrich) were used as primary antibodies, and both anti-Mouse (A-21057; Thermo Fisher Scientific) and anti-Rabbit (926-32213; Licor) were used as secondary antibodies. The densities of the resulting bands were determined using the Image Studio Lite software (Licor). Both the PRKN and VPS35 protein levels were adjusted with the β -Actin protein levels.

Pull-down assays and mass spectrometry

Proteins able to bind PRKN and VPS35 G4 were isolated using a previously described strategy (Leppek and Stoecklin 2014; Bolduc *et al.* 2020). Briefly, cellular protein extracts were prepared from 4 confluent 15 cm dishes of untransfected SH-SY5Y cells. The cell culture dishes were placed on ice and washed once with ice-cold PBS (10 mL). Ice-cold KCl-lysis buffer (4 mL; 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 50 mM NaCl, 1 mM DTT, 0,1 % Triton X-100, 10 % glycerol, 2 mM MgCl₂, Ribonucleoside Vanadyl complex RNase inhibitors (New England Biolabs) and Mini Complete Protease inhibitors (Roche)) was then added. The cells in the first dish were scrapped off using a cold cell scraper and then were transferred into the second dish. The steps were repeated with the three other dishes until the cells from all 4 dishes were combined. One final wash was performed with 1 mL of ice-cold KCl-lysis buffer, after which the cells were transferred to a pre-cooled 15 mL tube and incubated for 30 min at 4°C. The lysate was then clarified by centrifugation at 17 000 x g for 20 min at 4°C. The supernatant was then recovered, divided into 2 fractions and concentrated

using Amicon Ultra-4 MWCO 3 000 centrifuge tubes (EMD Millipore) at 1 600 x g until the final volumes reached 0.5-1 mL. The protein concentrations of the cellular extracts were then determined using the method of Bradford (BIO-RAD). The extracts were then pre-cleared by the addition of 50 μ L of a 50 % slurry of Streptavidin Sepharose High Performance beads (GE Healthcare) that had previously been washed 3 times with 1X KCl-lysis buffer and incubating for 16 h at 4°C. At this point, the beads were discarded. Next, either the *in vitro* transcribed 4xS1m WT or 4xS1m G/A mutant G4 RNA (30 μ g) was dissolved in 140 μ L of KCl-lysis buffer without MgCl₂ or RNase inhibitors and then heated to 70°C prior to cooling over a period of 1 h until it reached room temperature. Subsequently, the solution was supplemented to a final concentrations of 1 mM MgCl₂ and 120 U of RNase inhibitors were added (Protein purification platform from Université de Sherbrooke). The RNA solutions were then mixed with 100 μ L of a pre-washed (with KCl-lysis buffer) Streptavidin Sepharose bead slurry and incubated for 2 h at 4°C. After this incubation, the beads were centrifuged at 20 x g for 1 min and the supernatant was discarded. The beads were then washed once with ice-cold KCl-lysis buffer. Next, the pre-cleared lysate (e.g. the one incubated 16 h at 4°C) was supplemented with 1.5 μ L of RNase inhibitor (Protein purification platform from Université de Sherbrooke), added to the prepared RNA-coupled Streptavidin Sepharose beads and incubated for 3 h at 4°C. Lastly, the beads were washed five times with 1 mL of KCl-wash buffer (50 mM Tris-HCl (pH 7.5), 100 mM KCl, 200 mM NaCl, 1 mM DTT, 0,1 % Triton X-100, 10 % glycerol, 5 mM MgCl₂, Ribonucleoside Vanadyl complex RNase inhibitors (NEB) and Mini Complete Protease inhibitors (Roche)).

In order to prepare the proteins for mass spectrometry, those linked to RNA-beads were washed 5 times with 500 μ L of 20 mM ammonium bicarbonate. The samples were

gently shaken and centrifuged at 20 x g for 1 min between each wash. DTT (10 mM) dissolved in 20 mM ammonium bicarbonate was then added to the beads (2 volumes, 100 μ L) and the samples were incubated at 60°C for 30 min with constant agitation. After cooling, iodoacetamide (15 mM) dissolved in 20 mM ammonium bicarbonate was then added (2 volumes, 100 μ L) and a further incubation for 60 min in the dark (or protected from light) with constant agitation at room temperature was performed. The concentration of DTT was then increased to 15 mM, in order to quench the iodoacetamide, and the reaction incubated for a further 10 min. Trypsin (1 μ g) was then added and the reaction was incubated for 16 h at 37°C with constant agitation. The trypsin digestion was stopped by acidifying to a final concentration of 1 % formic acid and the beads were centrifuged at 20 x g for 1 min. The supernatant was then transferred to a new tube, the beads were washed with 200 μ L of a 60 % acetonitrile-0.1 % formic acid solution and the beads were centrifuged at 20 x g for 1 min. The second supernatant was pulled into the first one and dried in a speed vacuum for 20 h with high heating. The samples were dissolved in 30 μ L of 0.1 % Trifluoroacetic acid (TFA), desalted using Ziptips (Millipore Sigma), dried using a speed vacuum and finally were dissolved in 30 μ L of 1% formic acid solution.

Approximately 2 μ g of previously prepared peptides were injected into an Acclaim PepMap100 C18 column (0,3 mm id \times 5 mm; Dionex Corporation) at a flow rate of 4 μ L/min in order to identify any potential G4 protein partners. The peptides were then eluted into a PepMap C18 nanocolumn (75 μ m \times 50 cm; Dionex Corporation) coupled to an Orbitrap QExactive mass spectrometer (Thermo Fisher Scientific). The HPLC and MS were performed as described previously (Telekawa *et al.* 2018). The Xcalibur software was used to acquire the data (Thermo Fisher Scientific). Quantification was performed using the

MaxQuant version 1.5.2.8 software, and the Uniprot human database (10/04/2018, 74,811 entries) as described previously (Boisvert *et al.* 2010; Tyanova *et al.* 2016). Proteins with a coverage >5 % and a score of >15 were conserved for analysis. PRKN was analyzed with the triplicate pull-downs, while VPS35 was analyzed with 2 of the triplicate pull-downs because of the lower amount of protein present. The MS/MS counts were then analyzed using SAINT (a probabilistic scoring of affinity purification) in order to compare the WT G4 and G/A mutant G4 conditions (Choi *et al.* 2011). Proteins with a score higher than 0.9 and with a 10-fold or more enrichment were kept.

In order to confirm the proteins enrichment of GNL1 by western blot, 5% of the protein input was removed from the cellular lysate, dosed with DC reagent (BIO-RAD) and mixed with Laemmli buffer in order to obtain a final concentration of 50 mM Tris-HCl (pH 6.8), 1 % SDS, 10 % glycerol, 0.002 % bromophenol blue and 2.5 % β -mercaptoethanol. After the fives washes of the proteins bound to the RNA-bead complexes, 50 μ l of Laemmli buffer (50 mM Tris-HCl (pH 6.8), 1 % SDS, 10 % glycerol, 0.002 % bromophenol blue and 2.5 % β -mercaptoethanol) were added to the beads to dissolve the proteins. An SDS-PAGE was then performed on 30 μ g of the protein input and 50% of each pull-down. Western blots were performed using anti-GNL1 as primary antibody (ab245634; Abcam) and anti-rabbit as secondary antibody (926-32213; Licor).

Protein purification

In order to produce recombinant protein, the GNL1 sequence was fused to an N-terminal glutathione-S-transferase (GST) gene in the plasmid pcDNA3.1 (-) and the resulting plasmid was then transfected into ten 15-cm dishes of 60% confluent SH-SY5Y cells. Two days

post transfection, the cells were washed once with PBS and then were scraped in PBS. The cells were then harvested by centrifugation at 1 600 x g for 10 min, and the resulting supernatant discarded. The cell pellet was resuspended in 20 mL of lysis buffer (1X PBS, 1mM EDTA, 250 mM NaCl, 10% glycerol, 1 mM DTT and 1 tablet of Mini Complete Protease inhibitors (Roche)). Triton X-100 (210 μ L) was then added and the solution was sonicated at 20% power for 20 cycles of 5 sec ON – 5 sec OFF using a Branson Digital Sonifier 450. The resulting solution was then incubated for 30 min at 4°C prior to being centrifuged at 17 000 x g for 10 min at 4°C. The supernatant was then carefully transferred into a new tube. In parallel, of glutathione Sepharose 4B beads (250 μ L, GE healthcare) were washed three times with lysis buffer. The beads were then added to the cellular lysate and the sample incubated for 3 h at 4°C. The sample was then centrifuged at 225 x g for 2 min at 4°C and the supernatant discarded. The beads were then washed twice with lysis buffer containing 1% Triton X-100), twice with cold PBS and once with wash buffer (50 mM Tris-HCl pH 8.0, 250 mM of NaCl and 0.1% Triton X-100). The GST-GNL1 was then eluted by incubating for 15 min with 250 μ L of elution buffer (10 mM reduced glutathione, 50 mM Tris-HCl pH 8.0, 250 mM NaCl, 10% Glycerol and 0.1% Triton X-100). The elution step was repeated four times. The different fractions were then combined and dialyzed for 16 h using a 3 mL Slide-A-Lyzer Dialysis Cassette (10K MWCO; Thermo Fisher) with dialyze buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 2 mM DTT and 10% glycerol). Finally, the protein concentration was estimated using the DC protein assay (BIO-RAD), and the purity was evaluated by SDS-PAGE.

Electrophoretic mobility shift assay (EMSA)

In order to confirm the binding of GST-GNL1 to the G4 structures of both the PRKN and the VPS35 5'UTRs, EMSA were performed with the same 5'-labeled transcripts as were used in the *in-line* probing experiments. For each assay, 500 CPM of labeled RNA were mixed with 1 nM of unlabeled transcript in folding buffer (20 mM Tris-HCL pH 8.0 and 100 mM KCl), and the samples were then incubated for 5 min at 70°C prior to slow cooling at room temperature over a period of 1 h. Next, for every 7,5 µL of RNA solution of, 11.25 µL of binding buffer (200 mM Tris-HCl pH 8.0, 900 mM KCl, 10 mM MgCl₂ and 50% glycerol), 0.94 µL of RNase inhibitor (40 U/µL), 7.5 µL of BSA (3 mg/ul), 7.5 µL of yeast total RNA (2 ug/µL) and 6,5 µL of water were added. The recombinant protein was then added at concentrations ranging from 0 to 1816 nM. The samples were then incubated for 20 min at room temperature and then electrophoresed through a 5% native polyacrylamide gel in TBE buffer at 20 V/cm for 1 h at 4°C. The shift percentages were evaluated using the formula «Shift/(Shift + No shift RNA)». The dissociation constants (K_d) were calculated with the GraphPad software (Prism) using the "specific binding with hill slope nonlinear regression function".

Statistical analysis

Statistical analyses were performed using the GraphPad software (Prism). Depending on the experiments, either the One-sample Kolmogorov-Smirnov test or the Two-way ANOVA test was used (see the figure legends for a description). A p-value of less than 0.05 was considered statically significant. Proteins statistically enriched in the pull-down experiments were identified with SAINT (Choi *et al.* 2011). * represented a p-value < 0.05, ** a p-value <0.01, *** a P-value <0.001 and **** a p-value < 0.0001.

RESULTS

G-quadruplexes are globally enriched in diseases associated with the nervous system

Initially, potential G4s (pG4) related to nervous system diseases were investigated using the disease database in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.* 2015). In this database, each gene annotated to be deregulated for a specific disease is presented. The Fasta files of the mature transcripts associated with each variant of the nervous system genes were retrieved, and the sequences were scanned using the G4Screener software, a predictor of potential G4 that uses three scores (Garant *et al.* 2018). Two of these scores (specifically the G4H and cG/cC scores) estimate the G4 propensity, including measuring the competing nucleotide context, while the third score is a result of a neuronal network training using experimentally proved G4 sequences (G4NN) (Beaudoin *et al.* 2014; Bedrat *et al.* 2016; Garant *et al.* 2017). The use of a threshold of 0.5 or higher for the G4NN score (the default threshold proposed for the G4Screener) showed that 85 % of the 758 genes analyzed included at least one pG4 sequence. However, the different isoforms resulting from alternative splicing were not taken into account with the KEGG database. In order to remove this limitation, a new search for all isoforms of these genes was performed in the Refseq database using the accession numbers associated with each gene from the disease database. According to this search, 1 985 out of the 2 460 mRNA isoforms include at least one pG4 sequence (i.e. 81% of all of the isoforms include at least one pG4). The level of mRNAs that include at least one pG4 appears to be significantly higher than expected. In order to verify whether or not this was a true enrichment of pG4 in the mRNAs of nervous system diseases, the mRNA isoforms were compared to random sets of both mRNA and all other types of RNA (Figure 7). Alternative splicing isoforms from genes deregulated in nervous system diseases that include at least one pG4 are statistically enriched when compared to both

random sequence RNA sets. The use of a second database, specifically the Human Protein Atlas, also led to the conclusion that there is an higher proportion of pG4 within the enriched alternatively spliced isoform variants that are related to the nervous system (Figure 7).

Since pG4 are abundant in the genes associated with nervous system diseases, it was decided to restrict the analysis to the neurodegenerative diseases (ND), a sub-group in the KEGG disease database. Out of the 380 genes in the ND sub-group, 255 of them included at least one pG4 sequence (67%). In terms of the alternative splicing isoforms from the RefSeq database, 846 variants out of the 1038 in the ND subgroup included at least one pG4 sequence (82%).

The location of a G4 in an mRNA is a predictor of its potential contribution in terms of post-transcriptional regulation. In order to investigate the location of the pG4 within the mRNA, their positions were established for each of the isoform variants associated with the neurodegenerative diseases: 21% were located in the 5'UTR, 46% in the CDS region and 33% in the 3'UTR. At first view, the abundance of pG4 located within the CDS region seems surprising, because G4s are globally known to be enriched in both the 5' and 3'UTRs (Vannutelli *et al.* 2020). However, it is important to keep in mind that we are looking at the quantity, and not the density, of G4. As CDS are significantly longer than 5'UTRs, this may explain the observed bias.

Interestingly when one looks more closely at each mRNA associated with neurodegenerative diseases, an important enrichment of pG4 was observed for those associated with the Parkinson disease (PD) sub-group. Fifteen out of the sixteen associated genes (94%) included at least one pG4 sequence (46 out of the 49 isoform variants). More specifically, 29% of the G4s were located in the 5'UTRs, 51% in the coding regions and 20% in the 3'UTRs.

A G4Screener search using the default windows of 60 nucleotides (nts) yielded an output that consisted of a score indicating the probability of a pG4 sequence being located within the scanned region. Consequently, when moving from one position to another the output looks like a belt curve for which the maximal value should be indicative of the sequences most likely to fold into a G4 structure. The most probable sequence of the pG4 located in the 5'UTR (i.e. that located at the top of the belt shape), were selected for further characterization (see Table 2 for the detailed nucleotide sequences). Surprisingly, visual analysis of these eleven sequences revealed that almost all sequences did not fulfill the definition of a canonical pG4 ($G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$). Instead, only the pG4 sequence of the PLA2G6 mRNA appeared to satisfy the criteria of three G-quartets and three loops of 7 nucleotides or less (see Table 2). Most of the other sequences included pG4 sequences with only two G-quartets (eIF4G1 and SNCA). Several include central loops longer than 7 nucleotides (VPS35 and GIGYF2), and one was proposed to be formed by three G-quartets that included a bulge (PRKN). All of these features are proposed to be unfavorable to the stability of a G4 structure. Moreover, it is not clear which runs of guanosines would be forming the G4 structures. The demonstration of the folding of these pG4s into stable G4 structures was therefore mandatory before any further work could be undertaken.

***In vitro* validation of the pG4**

The binding of fluorescence ligands *in vitro* is a technique that is frequently used for the validation of G4 structures (Umar *et al.* 2019). Initially, the binding of the fluorescent G4-specific ligand N-Methyl Mesoporphrine (NMM) was tested in order to verify whether or not the eleven pG4 sequence candidates from the 5'UTRs of the genes related to PD can fold into a G4 structure. Briefly, in addition to the candidate sequences that can potentially form

pG4s, the sequences that are located up to 25 nucleotides both upstream and downstream of the PG4 were also considered in the *in vitro* transcription synthesis (See Tables S1 and S2 for the detailed sequences). The goal of this was to conserve the natural context of the tested pG4s as the neighboring nucleotides are known to influence the folding of a G4 (Beaudoin *et al.* 2014). For example, it has been shown that a surrounding sequence rich in cytosine residues may form CG Watson-Crick base pairs that prevent the formation of the G4 structures. Alternatively, a surrounding sequence including a run of guanosines may also contribute to the formation of an unexpected G4 structure. For each pG4, a G/A mutant version was also constructed. The mutants consist in the substitution of several guanosine residues by adenosines in order to prevent the formation of the G4 motifs. After gel purification, the RNA species were dissolved in a buffer that contained 100 mM of either LiCl or KCl, that is to say conditions that do and do not support G4 folding, respectively. The RNA was then heat denatured and slowly renatured before the addition of the NMM. Once bound to a G4 structure, the emission signal of NMM was detected with a peak at 605 nm (Nicoludis *et al.* 2012). G4s were considered as being positive when the signal was higher than 25 fluorescence units, which is an arbitrarily identified value for the wildtype (WT) sequence in the presence of K^+ . The results for each of the pG4 candidates are presented in the histogram in Figure 8. Six out of the eleven WT sequences exhibited a signal over 25 fluorescence units (VPS35, LRRK2, PAPK7, SNCA, PRKN, and eiF4G1A). In all cases, the WT sequences in the presence of Li^+ had a signal lower than 10 units, with the exception of the SNCA candidate that exhibited an increase of the fluorescent signal of around 15 units. The fluorescent signals for the G/A mutants were at background levels no matter what the salt conditions were. The only exception was the SNCA candidate where a signal of 22 units was detected in the presence of K^+ , while it remained at the background level in the presence

of Li^+ . Most likely this is the result of complex mixture of the many distinct structures formed by this RNA species. In brief, NMM binding suggested that more than half of the pG4 sequences do indeed fold into G4 structures.

In order to increase the confidence of the pG4 folding into a G4 structure, as well as to identify which guanosine residues are involved in the G-quartets, *in-line* probing analyses were performed for each of the six retained candidates. This technique, which is often used in the characterization of RNA G4 structures, is based on the spontaneous hydrolysis of flexible RNA nucleotides (e.g. those located in single-stranded positions) in the presence of Mg^{2+} (Beaudoin *et al.* 2013). Nucleotides involved in interactions (usually those located in double-stranded positions) are less susceptible to hydrolysis. In the case of a G4 structure, the guanosines involved in the G-quartets are usually not flexible, while the nucleotides that form the loops are and are therefore highly susceptible to hydrolysis. As compared to classical biophysical approaches like the circular dichroism spectrum, this approach requires only trace amounts of RNA, a fact which is considerably more biologically relevant. RNAs (5'-radiolabeled) were incubated at room temperature for 40 h in a buffer containing 20 mM Mg^{2+} and 100 mM of either K^+ or Li^+ . The resulting mixtures were subsequently analyzed at a single nucleotide resolution by denaturing polyacrylamide gel electrophoresis. A typical gel for the VPS35 candidate is shown in Figure 9A. The hydrolysis of the nucleotides involved in the loops, specifically C_{13} of loop 1, C_{18} , U_{19} and U_{20} of loop 2 (the central loop) and A_{23} of loop 3, was observed for the WT sequence in the presence of K^+ , but not in the presence of Li^+ , nor for the G/A mutant. The cleavage intensities under the different conditions were evaluated by densitometry, and the K^+/Li^+ ratios for each nucleotide are reported in the associated histograms (Figure 9B and S1). A nucleotide was considered to be significantly accessible when the ratio K^+/Li^+ was greater than 2. The K^+/Li^+ ratio was also

compared to that of the G/A mutant sequence in order to show that the increased accessibility was due to the presence of the G4 structure. For VPS35, the G4 involves the nucleotides located between positions 10 to 27 that are located near the 5' end of the synthetic transcript. In the case of LRRK2, the G4 involves the residues G₂₉ to G₄₂ and most likely forms a structure that contains only two quartets (Figure 9C). The pG4 of PARK7 was not folded according to the *in-line* probing data (Figure 9D), while the band profile of eIF4GA1 was so complex that it could not be analyzed and most likely results from the presence of a mixture of different structures in solution (data not shown). In the latter cases, it is important to mention that these two were the candidates with the lowest fluorescent NMM signals (see Figure 8). For SNCA, a mixture of two quartets can be observed from positions G₁₃ to G₃₉ (Figure 9E) resulting from the presence of many consecutive guanines. Finally, for PRKN, a 3 quartet G4 was formed by the residues G₁₀ to G₃₀ (Figure 9F). Thus, the *in-line* probing results confirmed the folding of four out of the six candidates tested: VPS35, LRRK2, SNCA and PRKN.

***In cellulo* evaluation of the G4 folding**

The four G4 validated by the biochemical approaches were further investigated *in cellulo*. Briefly, the full-length 5'UTR, as well as the G/A mutant versions, of the four genes were cloned upstream of the *Renilla* luciferase (RLuc) open reading frame (ORF) in the PsiCHECK-2 plasmid, which also includes a *Firefly* luciferase (FLuc) ORF for normalization purposes (see Table S1 and Figure S2 for the detailed sequences). The resulting plasmids were transfected into SH-SY5Y cells, a well-documented model cell line for Parkinson's disease studies. This cell line originates from humans, is catecholaminergic and is easy to work with (Xicoy *et al.* 2017). Twenty-four hours post-transfection, the RLuc and FLuc

luciferase levels were evaluated by measuring their respective activities in the different cellular lysates. Interestingly, a significant decrease (~1.5 fold) in the activity of the Rluc was observed for the WT G4s of both the PRKN and the VPS35 5'UTRs as compared to those of their respective G/A mutant versions. RT-ddPCR monitoring showed that the RNA quantities remained unchanged, confirming that the observed differences occurred at the translational and not the transcriptional level. No significant impact was observed for the G4 of the 5'UTR of SNCA, and a slight decrease was observed for the LRKK2 5'UTR G4.

The same experiment was repeated using HEK293T cells, which are often used for neuronal cell studies because of their similarity to neuronal cells in terms of development (Stephanenko et Dmitrenko, 2015). Similar results were obtained: the presence of the WT 5'UTR G4s of both PRKN and VPS35 significantly decreased the Rluc level (2.5-fold). Unlike the experiments performed in the SH-SY5Y cells, those using the HEK293T cells revealed a slight decrease in luciferase expression in the presence of the SNCA 5'UTR G4, while no effect was observed in the case of the LRKK2 5'UTR G4 (Figure S3).

Since the G4s of the 5'UTRs of both PRKN and VPS35 were the only ones which successfully passed both the *in vitro* and the *in vivo* tests, all subsequent experiments were limited to these two candidates. At this point, it became important to confirm that the folding of the G4 in the 5'UTR has a similar effect on the endogenous PRKN and VPS35. In order to do so, the expression levels of both proteins were evaluated in both the presence and the absence of two distinct G4 ligands, BRACO19 and Phen-DC3. Both of these ligands are known to stabilize G4s structures in cells (Burger *et al.* 2005; Chung *et al.* 2014; Yang *et al.* 2018; Zyner *et al.* 2019). Briefly, SH-SY5Y cells were incubated for 24 h in the presence of either BRACO19 or Phen-DC3, and the cells were then recovered and lysed. The proteins

were then extracted and analyzed by western blot using VPS35 and PRKN specific antibodies. The presence of Phen-DC3 significantly reduced the expression levels of both proteins, while BRACO19 only had an effect for PRKN (Figure 10B). These results confirmed not only that the presence of the G4 can repress the translation of both the endogenously expressed VPS35 and PRKN, but also the equivalent results obtained in the from the luciferase assays.

Isolation of proteins binding the G4 of either VPS35 or PRKN

It has been proposed that G4 motifs are only transiently folded in eucaryotes (Yang *et al.* 2018), highlighting the fact that RNA binding proteins (RBP) are important regulators in both their folding and their unfolding. Consequently, it is of interest to identify proteins able to bind the G4s of the VPS35 and/or PRKN 5'UTRs. Label-free RNA affinity purification assays were performed in order to identify the repertoire of potential binding partners (see Figure 11A). Initially, both the WT and G/A mutant G4 sequences of both VPS35 and PRKN were fused to four repeats of the S1m aptamer (4xS1m; see Figure S4). This RNA aptamer has a high affinity for streptavidin (Leppek et Stoecklin, 2014) and its presence does not impair the G4 folding as the NMM fluorescence is not affected (Figure S5). Moreover, *in-line* probing experiments also confirmed the proper folding of both G4s when they are fused to the S1m aptamer (Figure S6). In order to identify any potential protein partners, total proteins from SH-SY5Y cells were extracted and then pulled down using either the chimeric WT G4 or the G/A mutant RNA bound to streptavidin-coated sepharose beads. After the removal of any non-specific interactions, the proteins bound to the RNAs were recovered, trypsin digested and analyzed by MS/MS. These label-free RNA affinity purification assays were performed in triplicate, and were analyzed using the MaxQuant software (Cox et Mann,

2008), followed by the probabilistic scoring of affinity purification software using the SAINT software (Choi *et al.* 2011) by comparing proteins observed in the WT G4 fraction to those found in the G/A mutant fraction. The SAINT output consists of a protein fold enrichment and a calculated enrichment score from 0 to 1 where the most significant proteins are at 1. The proteins with a SAINT score higher than 0.9 and with a fold enrichment of at least 10-fold were selected for both PRKN and VPS35 (Choi *et al.* 2011). The SAINT data identified 3 proteins that were able to bind VPS35: specifically GNL1, a GTPase; and, two splicing factors, PRPF19 and YBX1. The later RBP is a protein already known to bind G4 (Lyons *et al.* 2016), which increased the confidence in the results presented here (Fig. 5B). In the case of PRKN, 13 proteins were identified: three GTP binding proteins (GNL1, LSG1 and GTPGF4); two splicing factors (SF3B2 and SF3B3); three mitochondrial ribosomal proteins (MRPL13, MRPL19 and MRPL45; all of which are highly unlikely and most likely simply result from the fact that cellular extracts without prior compartment isolation were used); three other RBP (FAM120, RALY and GRSF1; the later has been previously reported to bind a G4 RNA structure (Pietras *et al.*, 2018)); one transcription factor (ZNF622); and, one DNA binding protein (NKR1) (Figure 11B). Interestingly, GNL1 was the only protein common to both datasets of G4 binding proteins. This protein also has one of the highest enrichments and SAINT scores in both cases. To confirm that GNL1 was indeed enriched, western blots were performed after the 4xS1m protein pull-down. As expected, a clear enrichment of GNL1 for VPS35 and PRKN was observed when the WT G4 was compared to the G/A mutants (i.e. ~ 4.5 folds; Figure 11C). It was therefore decided to further investigate the interaction of GNL1 with both VPS35 and PRKN.

Characterization of the interaction between GNL1 and either the VPS35 or PRKN proteins

In order to confirm that GNL1 indeed binds to both the VPS35 and the PRKN G4s, electrophoretic mobility shift assays (EMSA) were performed. Initially, recombinant GST-GNL1 proteins were produced in HEK293T cells and were purified on glutathione-Sepharose-Beads (Figure S7). The production of recombinant proteins in human cells increases the probability that GNL1 has the post-translational modifications found in SH-SY5Y cells. The EMSA assays were performed using the same RNA molecules as were used for both the NMM assays and the *in-line* probings (i.e. both the WT and G/A mutant versions). The transcripts were initially heated and then slowly cooled to room temperature in order to ensure proper folding of the G4 structures. Subsequently, the RNA molecules were incubated with different concentrations of GST-GNL1 protein prior to performing the EMSA. As expected, GST-GNL1 had a significantly higher affinity for the WT G4s of both the VPS35 and PRKN 5'-UTRs as compared to those for their respective G/A mutant versions (Figure 12). The dissociation constant of the WT VPS35 was 2.5-fold greater than that for the G/A mutant (152 nM and 360 nM, respectively), while that for the WT PRKN was 35 fold smaller than that of its G/A mutant (194 nM versus 7339 nM, respectively). The results were also confirmed using a commercial GST-GNL1 obtained from Abnova (data not shown). In addition, it was confirmed that neither the GST protein alone, nor the impurities retrieved with the GST beads during the purification process, induced any shift with the various RNA constructs (data not shown). In brief, these data confirm that GNL1 binds directly both the VPS35 and PRKN G4s.

DISCUSSION

The neuronal network is a complex system with many different levels of post-transcriptional and post-translational mechanisms of regulation. Several post-transcriptional or post-translational deregulations at either the RNA or the protein levels have been shown to result in neurodegenerative diseases (Liu *et al.* 2017). Hence, the observation of an enrichment of pG4 sequences in mRNA isoforms associated with neuronal diseases, as compared to what is seen in both a random RNA and mRNA samples, was not surprising (Figure 7). Indeed G4 motifs have so far been implicated in a wide variety of post-transcriptional regulation types, including splicing, pre-miRNA processing and both cap-dependant and -independent translations, to name only a few examples (Rouleau *et al.* 2017b; Varshney *et al.* 2020). In fact, the general belief of the research community is that G4 structures are intimately associated with fine-tuned mechanisms of both transcriptional and post-transcriptional regulation. Therefore, it was expected that many pG4 would be found in the present study since the nervous system is the most differentiated system in the human body. More specifically in PD, pG4 were detected in fifteen genes, highlighting the potential importance of gene regulation through the presence of RNA G4 structures in PD. Prior to this bioinformatic search, only a single G4 motif located in the SNCA mRNA had been reported to be associated with PD (Koukouraki and Doxakis 2016). This candidate was also retrieved in the present study. The RNA G4 subgroup associated with PD includes pG4s located in both the 3'UTR and in the coding region, but the majority were located in the 5'UTR of the mRNA (11 out of 16 genes). It was therefore decided to focus on these 11 candidates as they provided the advantage that the mechanism is relatively easy to study using a luciferase reporter assay. However, it is important to keep in mind the possibility that other candidates may contribute to the regulation of other important post-transcriptional mechanisms.

The predicted G4 sequences located in the 5'UTRs of PD associated genes appeared to be relatively unstable. Only the pG4 from the PLA2G6 candidate fulfilled the criteria for a canonical G4 motif, all others did not (Table 2). Specifically, all of the others harboured several variations, variations such as the presence of a 2-quartet instead of the classical 3-quartet core, the presence of a mismatch between the G-quartets, or the presence of a larger than 7-nts loop, to name but a few examples. This level of non-canonical pG4 is quite surprising, and may be considered as being a potential characteristic of the transcriptome of nervous cells. Further biochemical, as well as cellular and molecular biology studies, have the potential of providing physical support for this hypothesis. That said, considering that G4s have been suggested to be “events” and not be “things”, (i.e. they are transiently folded in human cells), the abundance of non-canonical RNA G4s suggests that RBP may play a significant role in either the folding or the unfolding of these structures as most of the time non-canonical RNA G4s are less stable than canonical ones. Such RBP proteins may manage the RNA G4s by being a dynamic switch for post-transcriptional regulation.

Because of the above, the finding of a large proportion of non-canonical G4 structures makes mandatory the identification of the pG4 candidates that actually fold into a G4 structure. NMM fluorescent binding revealed that six of the candidates did indeed fold into G4 structures, although they did so in different proportions (according to the fluorescent signals). The *in-line* probing experiments confirmed that four of these six did indeed fold into G4 structures (Figures 8 and 9). The *in vitro* validation permitted the screening of a large number of potential candidates in order to further focus the efforts on the more time consuming *in cellulo* experiments. Unfortunately, only two out the four candidates were in fact confirmed to have a significant impact *in cellulo*. This could be an indication that RBP in cells may, in fact, have an impact on the post-transcriptional regulation. While the *in vitro*

assays were performed using pG4 sequences surrounded by up to 25 nts, in order to not be restricted to the positions that may specifically contribute to the G4 structures, the *in cellulo* luciferase assays were performed using the full-length 5'UTR sequences in order to be as relevant as possible to the natural genomic context. The luciferase assays revealed that the G4 motifs of both the VPS35 and PRKN 5'UTRs significantly repressed the translation (Figure 10A). In this study, the pG4 of the SNCA 5'UTR that had previously been reported to show some limited repression of translation (Koukouraki and Doxakis 2016), exhibited only a light repression that cannot be deemed as being significant. This discrepancy may simply depend on the construction used, the experimental conditions, or, alternatively, on the fact that only a small proportion of the pG4 is actually folded (maybe because it is a motif based solely on 2 G-quartets; Table 2). More importantly, the folding of the G4 from both the endogenous VPS35 and PRKN mRNAs received physical support from ligand binding experiments in cells (Figure 10B). The ligands may not have reacted similarly to both rG4 because their binding to non-canonical RNA G4 structures has not yet been well studied. All of these experiments provide much information on the stability of the G4s related to PD. The vast majority of them seem to show that these G4 are globally unstable. Most are formed by non-canonical G4s, thus reducing the stability of the G4 structures. Also, some structures seem to be formed by an equilibrium of a mixture of different G4s (e.g. SNCA and eIF4G1).

PRKN and VPS35 were the two most promising candidates identified by the screening methodology used here. These proteins exhibit distinct functions and are differently deregulated in PD (Nguyen *et al.* 2019). On one hand, Parkin RBR E3 Ubiquitin Protein Ligase (PRKN) is an E3 ubiquitin ligase that is activated by PINK1 to promote the ubiquitination of the outer mitochondrial membrane proteins, leading to mitophagy. In PD, there is an inactivation and/or reduction of the ubiquitination activity of PRKN that alters the

mitochondria recycling pathways (Sliter *et al.* 2018; Gatica *et al.* 2018). On the other hand, Vacuolar Protein Sorting-Associated Protein 35 (VPS35) is a component of the the WASP and SCAR homologue (WASH) retromer that is implicated in endosomal trafficking. In PD, VPS35 deficiency or mutation impairs both mitochondrial fusion in dopaminergic neurons and endosome transportation (Tang *et al.* 2015). In brief, both are important proteins in PD and controlling their expression may provide a new therapeutic avenue, or at least minimally provide a functional genomic approach for learning more about their contribution to the disease state. To do so, targeting G4 structures may *a priori* be an interesting approach (Sun *et al.* 2019). However, one of the important hurdles of such an approach is that G4s are structurally similar and, consequently, targeting sequence specifically a G4 motif will likely be difficult to achieve using small molecules. Moreover, the presence of non-canonical RNA G4 does not simplify the situation.

Clearly, the results obtained both *in vitro* and *in cellulo*, as well as the fact that it appears to be complicated to specifically target a G4 structure in the nervous system, campaign for the identification of the RBPs that interact with both candidates. In order to do so, a pull-down approach independently using the RNA G4s of PRKN and VPS35 was performed (Figure 11). Even though crude cellular extracts were used (i.e. there was no cellular compartment pre-isolation), most of the identified proteins were RBP, and only a few DNA binding proteins were retrieved (e.g. NKRF; Figure 11B) even though RNA and DNA G4 structures are globally similar (Asamitsu *et al.* 2019). The use of a crude cellular extract is also most likely the reason that the mitochondrial proteins MRPL13, MRPL19 and MRPL45 were retrieved (Figure 11B). Clearly, this is a limit with this approach. Importantly, the finding that YBX1 bound to the G4 of VPS35 and that GRSF1 bound to the G4 of PRKN, two RBP previously reported to interact with other G4 structures (Lyons *et al.* 2016; Pietras

et al. 2018), provided confidence in the results presented here. There was also another interesting protein, FAM120A, which had previously been identified as a protein binding to G-Rich regions located in 3'UTRs (Kelly *et al.* 2019). In the latter report, the authors showed that mutating some of the guanine residues of the G-Rich regions reduced the binding of FAM120A, a characteristic of the G4 binding proteins. The only protein that was retrieved which bound with both the VPS35 and PRKN G4 structures was GNL1, a protein known to be a GTPase (Boddapati *et al.* 2012). To our knowledge GNL1 has never been reported in the literature to interact with a G4 structure. The sequence of GNL1 includes a Di-RG domain, a domain type previously described as being a G4 RNA binding domain (Thandapani *et al.* 2013; Ozdilek *et al.* 2017; Huang *et al.* 2018). The direct binding of GNL1 to both the VPS35 and PRKN RNA G4s was confirmed by EMSA. A higher affinity of GNL1 was observed for the WT G4 versions as compared to the respective G/A mutant ones. Interestingly, in both cases, the K_d for the WT sequence was in the nanomolar range, supporting a strong affinity of GNL1 towards the RNAs. Another interesting fact was that the G/A mutation seems to affect the binding of GNL1 to PRKN more (35-fold versus 2.5-fold for VPS35). PRKN harbors a canonical G4, while VPS35 has a non-canonical one. Therefore, the difference between the WT and the G/A mutant versions of PRKN is more significant in terms of stability, and, most likely, a higher fraction of the G4 is truly folded. Once again this hypothesis pointed out the importance of further investigating the consequences associated with the canonical and non-canonical G4 motifs that appear to be a particular feature of the transcriptome of the nervous cells.

This study reinforces the possible importance of G4 structures as potential *cis*-acting features in the post-transcriptional regulation of the nervous system, more specifically that in neurological diseases. A new RBP has also been identified, GNL1, that is able to bind to the

5'UTR RNA G4s of two genes associated with Parkinson's disease (VPS35 and PRKN). GNL1 is a *trans*-acting RBP which may constitute a suitable target for controlling the translation of both genes. In brief, this study sheds light on a new molecular mechanism important for Parkinson's disease, and proposes new avenues for modulating disease translation.

ACKNOWLEDGEMENTS

We thank D. Lévesque for his support with the MS instrument and for the label-free analysis. J.P.P. holds the Research Chair of the Université de Sherbrooke in RNA Structure and Genomics and is a member of the Centre de Recherche du CHUS.

FUNDINGS

This project was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC; **155219-17**) to JPP. MAT received student fellowships from the Canadian Institutes of Health Research (CIHR), the Fonds de Recherche Québec Santé (FRQS) and from the Faculty of medicine and health sciences of Université de Sherbrooke. J.M.G received a student fellowship from NSERC. H.C.R received a student fellowship from Fonds de Recherche Québec Nature et Technologie (FRQNT). The funders had no role in study design, data collection and analysis, decision to publish, nor in the preparation of the manuscript.

CONTRIBUTIONS TO THE RESEARCH

M.A.T and J.P.P designed all of the experiments. MAT, J.M.G and H.R.C performed the experiments. MAT, J.M.G, H.R.C and J.P.P analysed data. M.A.T and J.P.P wrote the manuscript.

CONFLICT OF INTEREST

The authors declare no conflicts of interest for this work.

REFERENCES

- Arora A., Suess B. (2011) An RNA G-quadruplex in the 3' UTR of the proto-oncogene PIM1 represses translation. *RNA Biol.* **8**, 802–805.
- Asamitsu S., Takeuchi M., Ikenoshita S., Imai Y., Kashiwagi H., Shioda N. (2019) Perspectives for Applying G-Quadruplex Structures in Neurobiology and Neuropharmacology. *Int. J. Mol. Sci.* **20**.
- Beaudoin J.-D., Jodoin R., Perreault J.-P. (2013) In-line probing of RNA G-quadruplexes. *Methods* **64**, 79–87.
- Beaudoin J.-D., Jodoin R., Perreault J.-P. (2014) New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Res.* **42**, 1209–1223.
- Beaudoin J.-D., Perreault J.-P. (2010) 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* **38**, 7022–7036.
- Beaudoin J.-D., Perreault J.-P. (2013) Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.* **41**, 5898–5911.
- Bedrat A., Lacroix L., Mergny J.-L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759.
- Boddapati N., Anbarasu K., Suryaraja R., Tendulkar A. V., Mahalingam S. (2012) Subcellular Distribution of the Human Putative Nucleolar GTPase GNL1 is Regulated by a Novel Arginine/Lysine-Rich Domain and a GTP Binding Domain in a Cell Cycle-Dependent Manner. *J. Mol. Biol.* **416**, 346–366.
- Boisvert F.-M., Lam Y. W., Lamont D., Lamond A. I. (2010) A Quantitative Proteomics Analysis of Subcellular Proteome Localization and Changes Induced by DNA Damage. *Mol. Cell. Proteomics MCP* **9**, 457–470.
- Bolduc F., Garant J.-M., Allard F., Perreault J.-P. (2016) Irregular G-quadruplexes Found in the Untranslated Regions of Human mRNAs Influence Translation. *J. Biol. Chem.* **291**, 21751–21760.
- Bolduc F., Turcotte M.-A., Perreault J.-P. (2020) The Small Nuclear Ribonucleoprotein Polypeptide A (SNRPA) binds to the G-quadruplex of the BAG-1 5'UTR. *Biochimie.*
- Braschi B., Denny P., Gray K., Jones T., Seal R., Tweedie S., Yates B., Bruford E. (2019) Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res.* **47**, D786–D792.
- Bugaut A., Balasubramanian S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **40**, 4727–4741.
- Burger A. M., Dai F., Schultes C. M., Reszka A. P., Moore M. J., Double J. A., Neidle S. (2005) The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Res.* **65**, 1489–1496.
- Choi, Larsen B., Lin Z.-Y., Breitkreutz A., Mellacheruvu D., Fermin D., Qin Z. S., Tyers M., Gingras A.-C., Nesvizhskii A. I. (2011) SAINT: Probabilistic Scoring of Affinity Purification - Mass Spectrometry Data. *Nat. Methods* **8**, 70–73.
- Chung W. J., Heddi B., Hamon F., Teulade-Fichou M.-P., Phan A. T. (2014) Solution Structure of a G-quadruplex Bound to the Bisquinolinium Compound Phen-DC3. *Angew. Chem. Int. Ed.* **53**, 999–1002.

- Cox J., Mann M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372.
- Endoh T., Kawasaki Y., Sugimoto N. (2013) Suppression of Gene Expression by G-Quadruplexes in Open Reading Frames Depends on G-Quadruplex Stability. *Angew. Chem. Int. Ed.* **52**, 5522–5526.
- Garant J.-M., Perreault J.-P., Scott M. S. (2017) Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. *Bioinformatics* **33**, 3532–3537.
- Garant J.-M., Perreault J.-P., Scott M. S. (2018) G4RNA screener web server: User focused interface for RNA G-quadruplex prediction. *Biochimie* **151**, 115–118.
- Gatica D., Lahiri V., Klionsky D. J. (2018) Cargo recognition and degradation by selective autophagy. *Nat. Cell Biol.* **20**, 233.
- Guo J. U., Bartel D. P. (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**.
- Huang Z.-L., Dai J., Luo W.-H., Wang X.-G., Tan J.-H., Chen S.-B., Huang Z.-S. (2018) Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions. *J. Am. Chem. Soc.*
- Imperatore J. A., Then M. L., McDougal K. B., Mihailescu M. R. (2020) Characterization of a G-Quadruplex Structure in Pre-miRNA-1229 and in Its Alzheimer's Disease-Associated Variant rs2291418: Implications for miRNA-1229 Maturation. *Int. J. Mol. Sci.* **21**.
- Ishiguro A., Kimura N., Watanabe Y., Watanabe S., Ishihama A. (2016) TDP-43 binds and transports G-quadruplex-containing mRNAs into neurites for local translation. *Genes Cells* **21**, 466–481.
- Jodoin R., Bauer L., Garant J.-M., Mahdi Laaref A., Phaneuf F., Perreault J.-P. (2014) The folding of 5'-UTR human G-quadruplexes possessing a long central loop. *RNA N. Y. N* **20**, 1129–1141.
- Jodoin R., Carrier J. C., Rivard N., Bisailon M., Perreault J.-P. (2019) G-quadruplex located in the 5'UTR of the BAG-1 mRNA affects both its cap-dependent and cap-independent translation through global secondary structure maintenance. *Nucleic Acids Res.* **47**, 10247–10266.
- Jodoin R., Perreault J.-P. (2018) G-quadruplexes formation in the 5'UTRs of mRNAs associated with colorectal cancer pathways. *PLoS One* **13**, e0208363–e0208363.
- Kanehisa M., Sato Y., Kawashima M., Furumichi M., Tanabe M. (2015) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462.
- Kelly T. J., Suzuki H. I., Zamudio J. R., Suzuki M., Sharp P. A. (2019) Sequestration of microRNA-mediated target repression by the Ago2-associated RNA-binding protein FAM120A. *RNA* **25**, 1291–1297.
- Kim N. (2019) The Interplay between G-quadruplex and Transcription. *Curr. Med. Chem.* **26**, 2898–2917.
- Koukouraki P., Doxakis E. (2016) Constitutive translation of human α -synuclein is mediated by the 5'-untranslated region. *Open Biol.* **6**, 160022.
- Laederach A., Das R., Vicens Q., Pearlman S. M., Brenowitz M., Herschlag D., Altman R. B. (2008) Semi-automated and rapid quantification of nucleic acid footprinting and structure mapping experiments. *Nat. Protoc.* **3**, 1395–1401.

- Leppek K., Stoecklin G. (2014) An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res.* **42**, e13.
- Lerner L. K., Sale J. E. (2019) Replication of G Quadruplex DNA. *Genes* **10**.
- Liu E. Y., Cali C. P., Lee E. B. (2017) RNA metabolism in neurodegenerative disease. *Dis. Model. Mech.* **10**, 509–518.
- Lyons S. M., Achorn C., Kedersha N. L., Anderson P. J., Ivanov P. (2016) YB-1 regulates tiRNA-induced Stress Granule formation but not translational repression. *Nucleic Acids Res.* **44**, 6949–6960.
- Małgowska M. (2016) Overview of RNA G-quadruplex structures. *Acta Biochim. Pol.* **63**, 609–621.
- Masuzawa T., Oyoshi T. (2020) Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization. *ACS Omega* **5**, 5202–5208.
- Matsumura K., Kawasaki Y., Miyamoto M., Kamoshida Y., Nakamura J., Negishi L., Suda S., Akiyama T. (2017) The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene* **36**, 1191–1199.
- Melko M., Bardoni B. (2010) The role of G-quadruplex in RNA metabolism: Involvement of FMRP and FMR2P. *Biochimie* **92**, 919–926.
- Mendoza O., Bourdoncle A., Boulé J.-B., Brosh R. M., Mergny J.-L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.* **44**, 1989–2006.
- Neidle S., Parkinson G. N. (2003) The structure of telomeric DNA. *Curr. Opin. Struct. Biol.* **13**, 275–283.
- Nguyen M., Wong Y. C., Ysselstein D., Severino A., Krainc D. (2019) Synaptic, Mitochondrial, and Lysosomal Dysfunction in Parkinson's Disease. *Trends Neurosci.* **42**, 140–149.
- Nicoludis J. M., Barrett S. P., Mergny J.-L., Yatsunyk L. A. (2012) Interaction of human telomeric DNA with N- methyl mesoporphyrin IX. *Nucleic Acids Res.* **40**, 5432–5447.
- Ozdilek B. A., Thompson V. F., Ahmed N. S., White C. I., Batey R. T., Schwartz J. C. (2017) Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res.* **45**, 7984–7996.
- Pietras Z., Wojcik M. A., Borowski L. S., Szewczyk M., Kulinski T. M., Cysewski D., Stepień P. P., Dziembowski A., Szczesny R. J. (2018) Dedicated surveillance mechanism controls G-quadruplex forming non-coding RNAs in human mitochondria. *Nat. Commun.* **9**.
- Rouleau S. G., Garant J.-M., Bolduc F., Bisailon M., Perreault J.-P. (2017a) G-Quadruplexes influence pri-microRNA processing. *RNA Biol.* **15**, 198–206.
- Rouleau S., Jodoin R., Garant J.-M., Perreault J.-P. (2017b) RNA G-Quadruplexes as Key Motifs of the Transcriptome. **170**, 1–20.
- Schludi M. H., Edbauer D. (2018) Targeting RNA G-quadruplexes as new treatment strategy for C9orf72 ALS/FTD. *EMBO Mol. Med.* **10**, 4–6.
- Sliter D. A., Martinez J., Hao L., Chen X., Sun N., Fischer T. D., Burman J. L., et al. (2018) Parkin and PINK1 mitigate STING-induced inflammation. *Nature* **561**, 258.
- Stephanenko A. A., Dmitrenko V. V. (2015) HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene* **569**, 182–190.

- Sun Z.-Y., Wang X.-N., Cheng S.-Q., Su X.-X., Ou T.-M. (2019) Developing Novel G-Quadruplex Ligands: From Interaction with Nucleic Acids to Interfering with Nucleic Acid-Protein Interaction. *Molecules* **24**.
- Tang F.-L., Liu W., Hu J.-X., Erion J. R., Ye J., Mei L., Xiong W.-C. (2015) VPS35 deficiency or mutation causes dopaminergic neuronal loss by impairing mitochondrial fusion and function. *Cell Rep.* **12**, 1631–1643.
- Telekawa C., Boisvert F.-M., Bachand F. (2018) Proteomic profiling and functional characterization of post-translational modifications of the fission yeast RNA exosome. *Nucleic Acids Res.* **46**, 11169–11183.
- Thandapani P., O'Connor T. R., Bailey T. L., Richard S. (2013) Defining the RGG/RG Motif. *Mol. Cell* **50**, 613–623.
- Tyanova S., Temu T., Cox J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319.
- Umar M. I., Ji D., Chan C.-Y., Kwok C. K. (2019) G-Quadruplex-Based Fluorescent Turn-On Ligands and Aptamers: From Development to Applications. *Molecules* **24**.
- Vannutelli A., Belhamiti S., Garant J.-M., Ouangraoua A., Perreault J.-P. (2020) Where are G-quadruplexes located in the human transcriptome? *NAR Genomics Bioinforma.* **2**.
- Varshney D., Spiegel J., Zyner K., Tannahill D., Balasubramanian S. (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.*, 1–16.
- Xiao C.-D., Shibata T., Yamamoto Y., Xu Y. (2018) An intramolecular antiparallel G-quadruplex formed by human telomere RNA. *Chem Commun* **54**, 3944–3946.
- Xicoy H., Wieringa B., Martens G. J. M. (2017) The SH-SY5Y cell line in Parkinson's disease research: a systematic review. *Mol. Neurodegener.* **12**, 10.
- Yang S. Y., Lejault P., Chevrier S., Boidot R., Robertson A. G., Wong J. M. Y., Monchaud D. (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.* **9**, 1–11.
- Yu C.-H., Teulade-Fichou M.-P., Olsthoorn R. C. L. (2014) Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res.* **42**, 1887–1892.
- Zyner K. G., Mulhearn D. S., Adhikari S., Martínez Cuesta S., Di Antonio M., Erard N., Hannon G. J., Tannahill D., Balasubramanian S. (2019) Genetic interactions of G-quadruplexes in humans. *eLife* **8**, e46793.

FIGURES

Figure 7: Enrichment of pG4 in genes associated with the nervous system.

pG4s with a G4NN threshold of at least 0.5 were considered as being positive pG4s. pG4s from both the nervous system section in the KEGG disease (KEGG) and the proteins enriched in the nervous system from the Human Proteins Atlas were compared with a set of random mRNAs and with random RNA. All distributions of random passed the "One-sample Kolmogorov-Smirnov test" for normal distribution with a p-value below 0.01. Within these samples, a value must have a z-score above 2.262 in order to be considered as not being part of the distribution with a confidence level of 95 % ($\alpha=0.05$), or above 3.250 in order to be considered as not being part of the distribution with a confidence level of 99 % ($\alpha=0.01$).

Figure 8: Characterization of RNA G-Quadruplexes found in genes deregulated in PD by fluorescence assay

The fluorescent assays were performed with either WT or G/A mutant G4 RNAs (300 pmol) of the 11 5'UTR pG4 candidates. The RNAs were incubated for 30 min with 0.5 mM NMM and 100 mM of either KCl or LiCl at room temperature. The samples were excited at 399 nm and the emission was detected 605 nm. The error bars represent the SEM. All experiments were performed in duplicate.

Figure 9: Characterization of the RNA G-Quadruplexes found in genes that are deregulated in PD by *in-line* probing

The *in-line* probing experiments were analyzed by both polyacrylamide gel electrophoresis and densitometry. A) Denaturing polyacrylamide gel of the *in-line* probing reactions of VPS35. Lanes 1 to 4 are for the WT, while lanes 5 to 8 are for the G/A mutant. Lanes 1 and

5: NaOH Ladder; Lanes 2 and 6: RNase T1 Ladder; Lanes 3 and 7: the Li⁺ condition; and, Lanes 4 and 8: the K⁺ condition. Histograms for the B) VPS35; C) LRRK2; D) PARK7; E) SNCA; and, F) PRKN RNA G4s were traced from the *in-line* probing data. The results were derived from the densitometric analysis of each band using the SAFA software (Laederach *et al.* 2008). The sequence is indicated on the y-axis. The X-axis represents the K⁺/Li⁺ ratios of each band's intensity for both the WT and the G/A mutant G4s. The K⁺/Li⁺ ratios are shown in black for the WT G4s and in grey for the G/A mutant G4s. The nucleotides are identified by their position with respect to the 5' end of the synthetic transcript. The nucleotides marked with arrowheads are those substituted for adenine in the G/A mutant. The square guanines are those which could be expected to form a G4. " II " indicates the minimal limit of gel quantification. The dotted line represents the twofold threshold that denotes a significant gain in flexibility. Each bar represents the average of two independent experiments, and the error bars represent the standard deviations.

Figure 10: *In cellulo* characterization of RNA G-Quadruplexes in SH-SY5Y cells

The impact of the G4s on translation was analyzed using both luciferase assays and specific G4 ligands. A) The relative RLuc/FLuc ratios for the luciferase activities of the 5'UTRs of VPS35, LRRK2, SNCA and PRKN are shown on the left y-axis. The relative RNA levels for both the WT G4 and G/A mutant, as compared to the RNA level of the control FLuc, are shown on the right y-axis. The results shown are the means of at least three independent experiments, and the error bars represent the SEM. The P-values were calculated by two-way ANOVA. B) Endogenous protein expression levels of VPS35 and PRKN as quantified by western blot after a 1 day incubation with 10 μ M BRACO19 and 35 μ M PHEN-DC. The relative expression levels of the VPS35 and PRKN proteins were corrected to that of β -actin.

The results shown are the means of two independent experiments, and the error bars represent the SEM. The P-values were calculated by two-way ANOVA. (*) P < 0.05, (**) P < 0.01, (***) P < 0.001, (****) P < 0.0001.

Figure 11: RNA pull-down of proteins by the PRKN and VPS35 G-quadruplexes

Proteins able to bind to the VPS35 and PRKN RNA G4s were identified by a pull-down experiment that used the 4xS1m aptamer and the total proteins extract from SH-SY5Y cells. A) Illustration of the pull-down experiments. B) The proteins that were statistically enriched by the VPS35 and PRKN RNA transcripts. Only the genes with a SAINT score >0.9 and a ≥ 10 -fold enrichment were kept. The G/A mutants were considered as being the control group. The fold enrichment was determined by the SAINT software. The colors facilitate the visual association of the different functions: GTP binding (Blue), Splicing (Orange), Mitochondrial ribosomal protein (Black), RNA binding protein (Red), DNA binding protein (Green) and Transcription factor (White). C) A western blot using anti-GNL1 antibody was performed on 30 μ g of the protein input (Lane 1) and 50% of the proteins eluted from the beads (Lanes 2 and 3). The folds were calculated using Image Studio Lite (Licor).

Figure 12: Electrophoretic mobility shift assays of the GNL1 protein

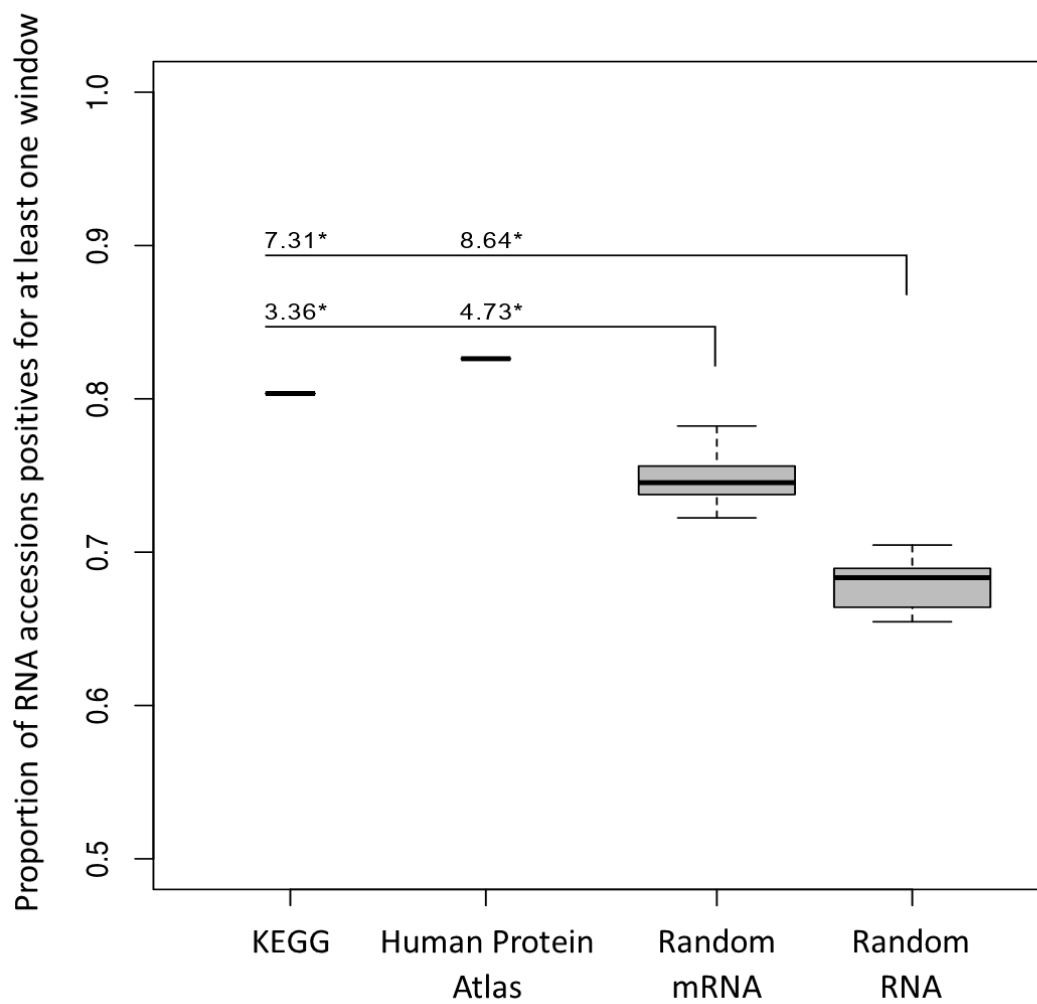
Electrophoretic mobility shift assays of the A) VPS35 and B) PRKN RNA G4s. The EMSA were performed using 0, 50, 100, 250, 500, 750, 1000 and 1816 nM of protein. The bands were analyzed by densitometry (on the top), and the data plotted graphically (on the bottom). The GraphPad software (Prism) "specific binding with hill slope nonlinear regression" function was used to determine the K_d values. Each assay was performed in duplicate.

TABLES

Genes	Nucleotide sequence 5' to 3'	G4NN Scores
VPS35		0,843
PINK1		0,821
FBXO7		0,817
LRRK2		0,663
ATP13A2		0,624
GIGYF2		0,618
PLA2G6		0,599
SNCA		0,558
PRKN		0,549
eIF4G1		0,542

Table 2: 5'UTR pG4s located in genes related to Parkinson's disease and manual analysis of the G4 structures

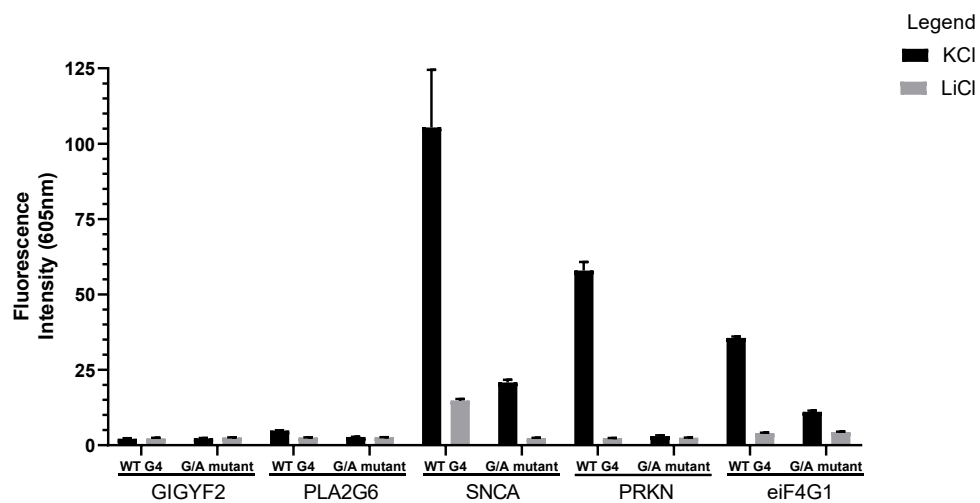
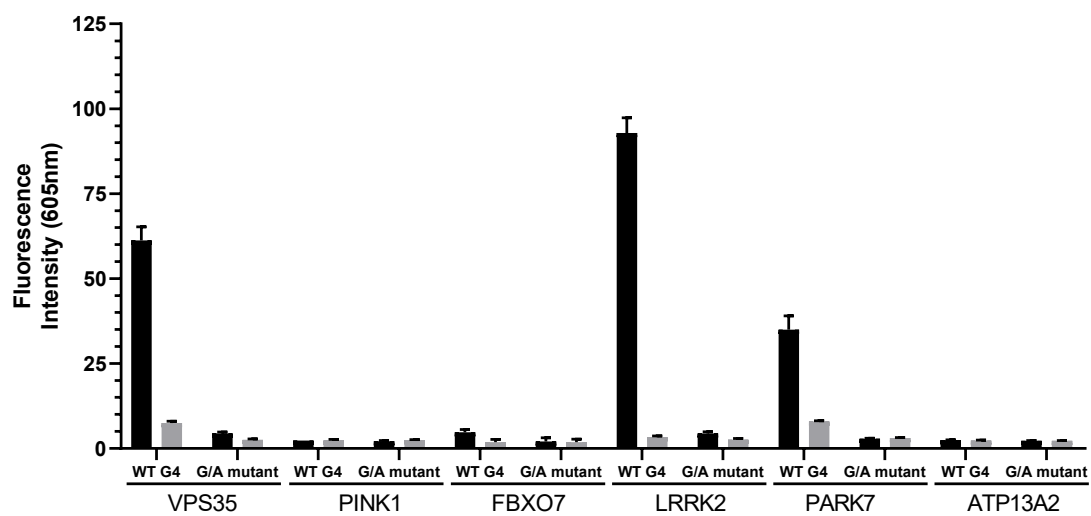
The 5'UTR pG4s associated with Parkinson's disease are listed in the first column. In the second column, the boxes indicate the G4 possessing the highest score, and the most abundant alternative mRNAs are shown. A minimum of one G4 structure had to have been manually found in each sequence. In order to improve the clarity, a maximum of 2 potential structures are shown per sequence. These structures are not exhaustive, many others could occur. The positions of the nucleotides are shown respective to their location in the 5'UTR. The G4NN scores are shown in the third column.



M-A Turcotte *et al.*

G4 RNA as translation regulator

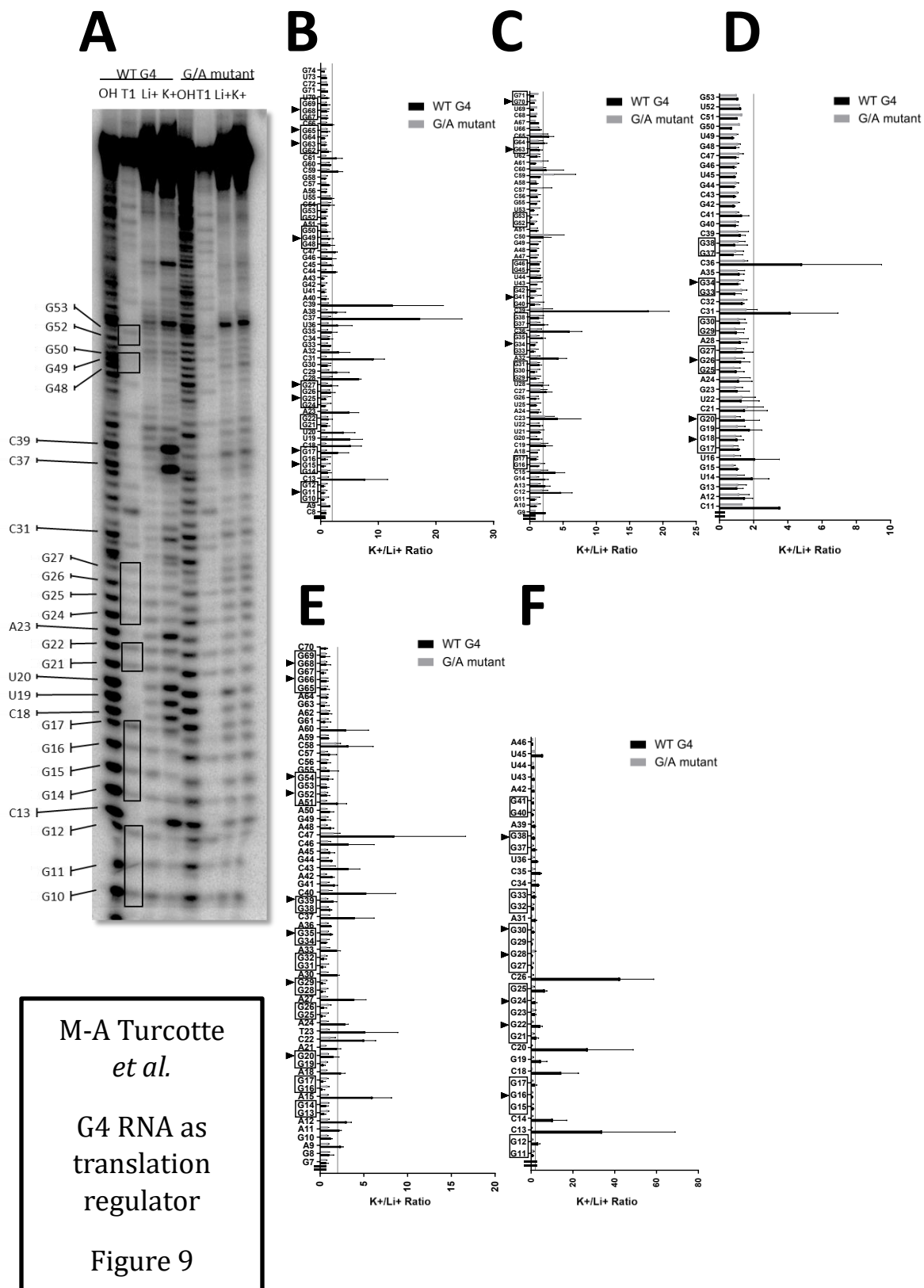
Figure 7

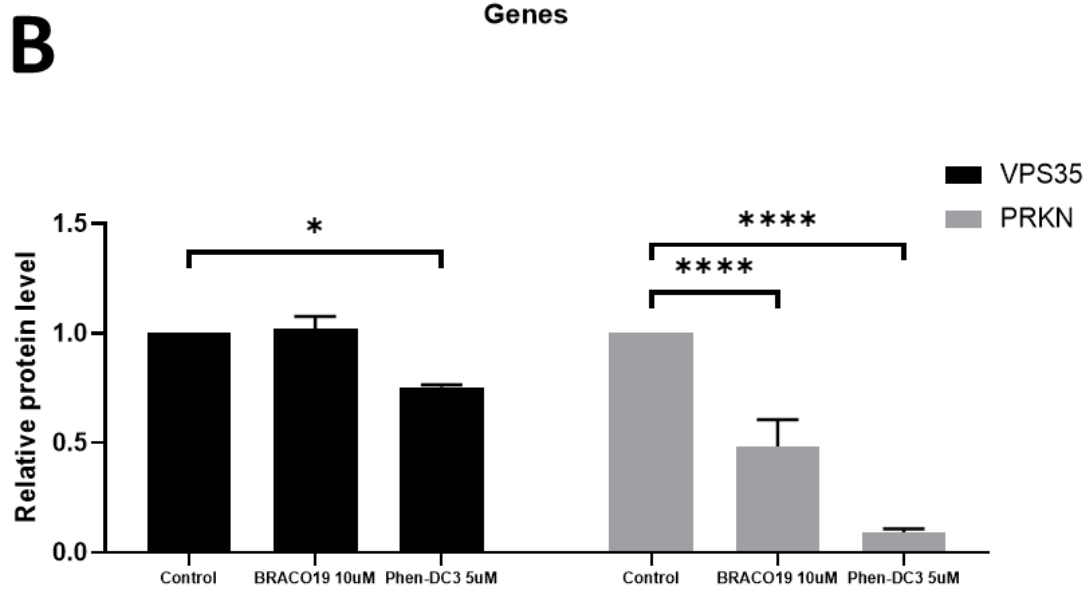
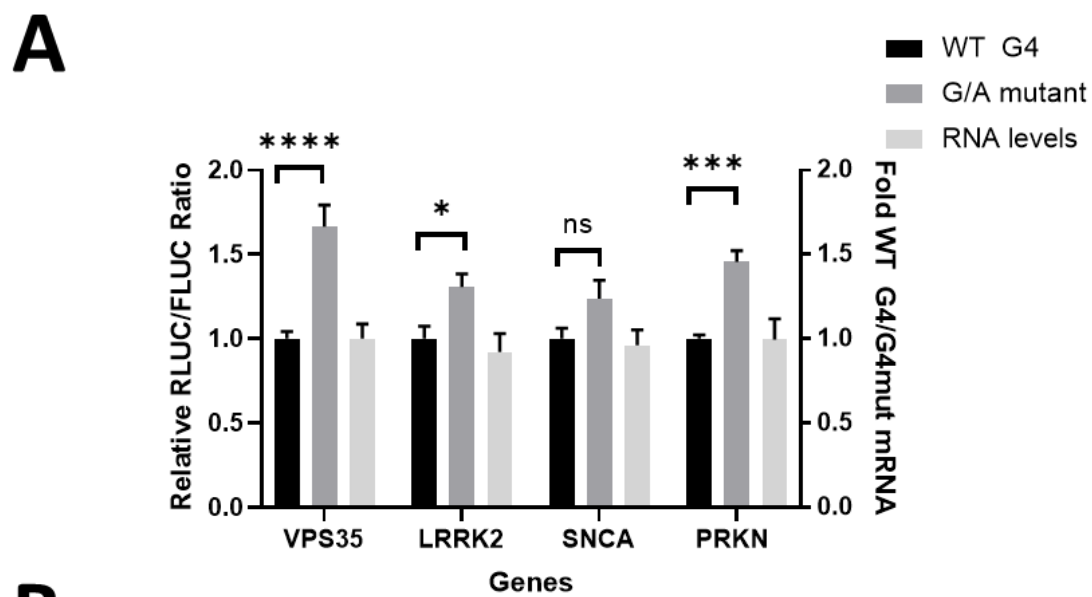


M-A Turcotte *et al.*

G4 RNA as translation regulator

Figure 8

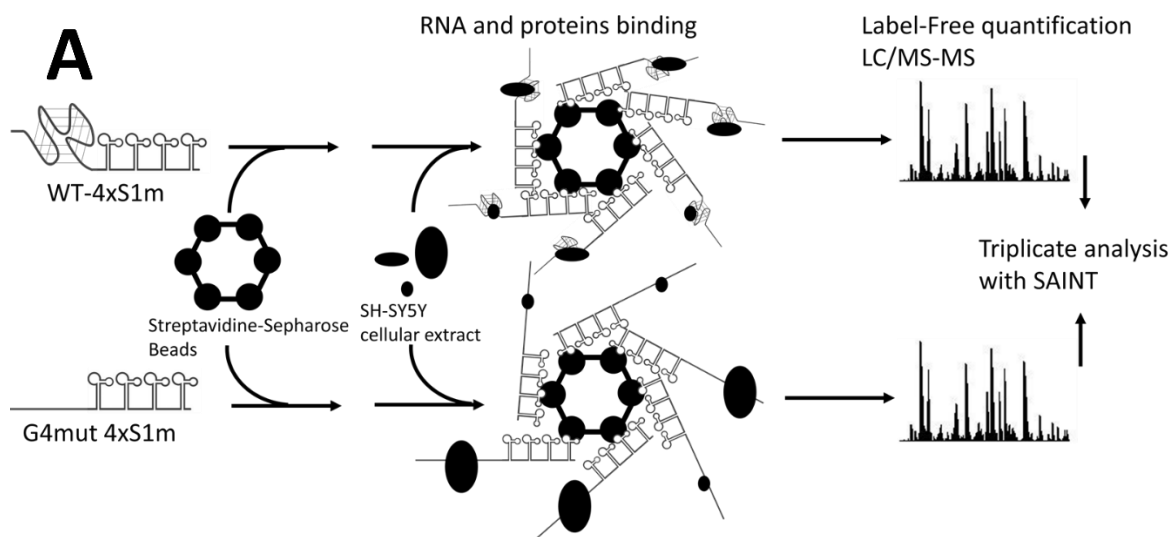




M-A Turcotte *et al.*

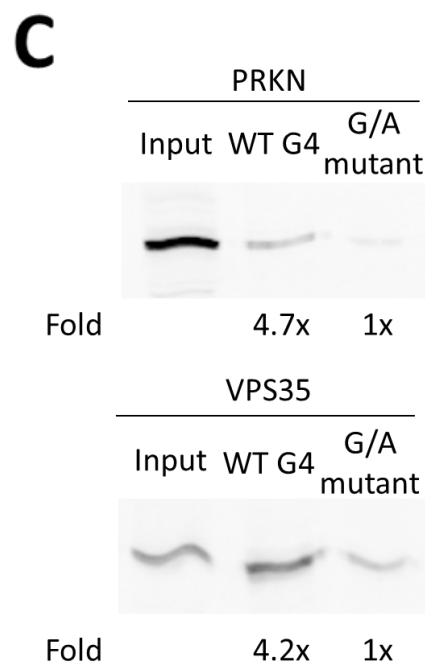
G4 RNA as translation regulator

Figure 10



B

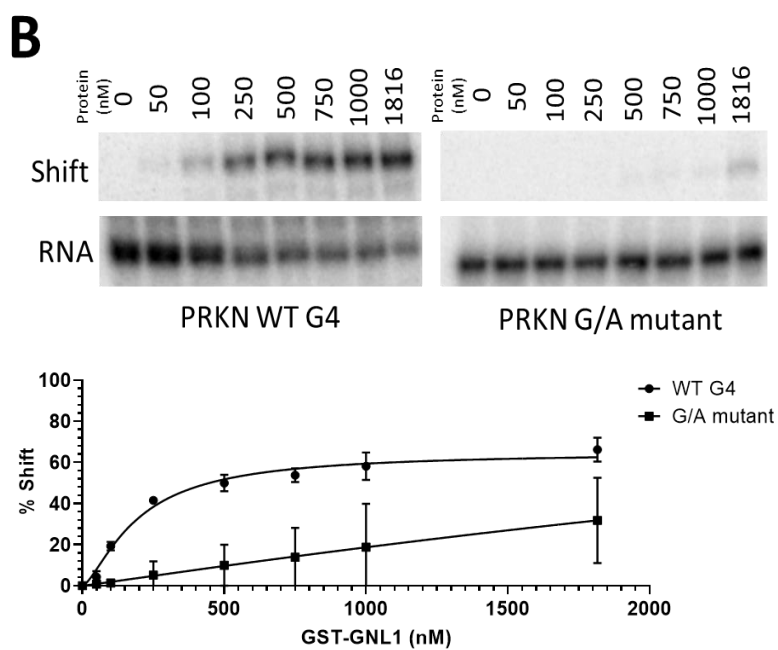
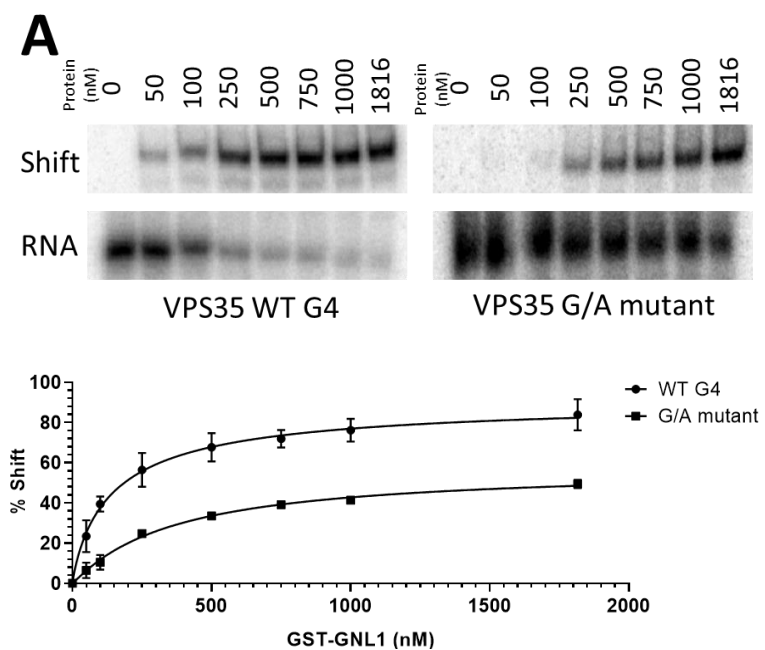
VPS35			
Genes	Saint score	Fold	Function
GNL1	1	12	GTP binding
PRPF19	0,98	40	Splicing
YBX1	0,92	11	Splicing
PRKN			
Genes	Saint score	Fold	Function
GNL1	1,00	143	GTP binding
SF3B2	1,00	63	Splicing
MRPL13	1,00	60	Mitochondrial ribosomal protein
FAM120A	1,00	50	RNA binding protein
ZNF622	1,00	23	Transcription factor
LSG1	1,00	15	GTP Binding
GRSF1	0,99	33	RNA binding protein
NKRF	0,97	27	DNA binding protein
MRPL19	0,96	10	Mitochondrial ribosomal protein
RALY	0,95	23	RNA binding protein
MRPL45	0,95	23	Mitochondrial ribosomal protein
SF3B3	0,93	12	Splicing
GTPBP4	0,91	10	GTP binding



M-A Turcotte *et al.*

G4 RNA as translation regulator

Figure 11



M-A Turcotte *et al.*

G4 RNA as translation regulator

Figure 12

Supplementary data section

Guanine Nucleotide-Binding Protein-Like 1 (GNL1) binds RNA G-Quadruplex structures in genes associated with Parkinson's disease

Marc-Antoine Turcotte, Jean-Michel Garant, H el ene Cossette-Roberge, and Jean-Pierre Perreault*

Supplementary Tables

Supplementary Table S1: Oligonucleotides used in this study

Gene	Type	Sequence	Nucleotides
ATP13A2 WT G4	Forward	5'-TAATACGACTCACTATAGGG-3'	20
	Reverse	5'-CGGGGCACCTGGGCCACCAGGCTCGCCCTCCGTGCGCGCCGCAAGCC CCGCCCCGGCCCTATAGTGAGTCGTATTA-3'	75
ATP13A2 G/A mutant	Forward	5'-TAATACGACTCACTATAGG-3'	20
	Reverse	5'-CGGGGCACCTGGGCCACCAGGCTCGCTCTCCGTGCGCGCCGCAAGTC TCGCTCGGTCCTATAGTGAGTCGTATTA-3'	75
eIF4G1 WT G4	Forward	5'-TAATACGACTCACTATAGGGCCAGGCCCGAACCCGGTGTCCCCGGGTG GGGGGTGGGGACGCCACGGCCGAAGCAGCTAGCTC-3'	83
	Reverse	5'-TTCCAGGTCTCGCTGGCACCAGGCTCCCGATCACGAACGGAGCTAGC TGCTTCGGCCGTGGCGTCCCCACCCCCACCCGGG-3'	83
eIF4G1 G/A mutant	Forward	5'-TAATACGACTCACTATAGGGCCAGGCCCGAACCCGGTGTCCCCGAGTG AGAGATGAGAACGCCACGGCCGAAGCAGCTAGCTC-3'	83
	Reverse	5'-TTCCAGGTCTCGCTGGCACCAGGCTCTCGGATCACGAACGGAGCTAGC TGCTTCGGCCGTGGCGTTCTCATCTCTCACTCGGG-3'	83
FBXO7 WT G4	Forward	5'-TAATACGACTCACTATAGGGCCTTCACGGGAGGCCCGGGCTCTTCCGG GCGTCGCGGAGCCGGAGGGTGCAGGCGACGGGAAG-3'	83
	Reverse	5'-CGGGCCATGTTCTCCAGGAGCCGGACCCAGCCGACCACCCGCGCTTC CCGTCGCTGCACCCTCCGGCTCCGCGACGCCCGG-3'	83
FBXO7 G/A mutant	Forward	5'-TAATACGACTCACTATAGGGCCTTCACGGGAGGCCCGGGCTCTTCCGA GCGTCGCGGAGCCGGAGAGTGCAGGCGACGAGAAG-3'	83
	Reverse	5'-CGGGCCATGTTCTCCAGGAGCCGGATCTCAGCCGACCACTCGCGCTTC TCGTCGCTGCACTCTCCGGCTCCGCGACGCTCGG-3'	83
GIGYF2 WT G4	Forward	5'-TAATACGACTCACTATAGG-3'	19
	Reverse	5'-CGGTCCGGGACAGGGAGTCTCAGAACCCAGTCAGTCCTCAGCCTCAA CAACACAAGATGGCCGACACGTCGCCTATAGTGAGTCGTATTA-3'	83
GIGYF2	Forward	5'-TAATACGACTCACTATAGG-3'	19

G/A mutant	Reverse	5'-CGGTCCGGGACAGGGAGTCTCAGAATCTCAGTCAGTTCTCAGCCTCAA CAACACAAGATGGTCGACACGTCGCCTATAGTGAGTCGTATTA-3'	83
LRRK2 WT G4	Forward	5'-TAATACGACTCACTATAGGTTCCCTGAGCAGCGGACGTTTCATGCTGGG AGGGCGGGCGGTTGGAAGCAGGTGCCACCATGG-3'	81
	Reverse	5'-TCAGAGTTTCTCGTCTCTTCGCACCCCTGACAGCTGCCACTAGCCAT GGTGGCACCTGCTTCCAACCCGCCGCCCTCCC-3'	81
LRRK2 G/A mutant	Forward	5'-TAATACGACTCACTATAGGTTCCCTGAGCAGCGGACGTTTCATGCTGGG AGAGCGGCGAGTTGGAAGCAGGTGCCACCATAG-3'	81
	Reverse	5'-TCAGAGTTTCTCGTCTCTTCGCATCTCTGACAGCTGCTACTAGCTATG GTGGCACCTGCTTCCAACCTCGCCGCTCTCCC-3'	81
PARK7 WT G4	Forward	5'-TAATACGACTCACTATAGG-3'	19
	Reverse	5'-ACACCAGGCTGAAAATGAACGCACGCCAGCACGCACGCGCGCCGTCC GGCCTCCCTCAGCCCCACACTGCGCAGACTCCTATAGTGAGTCGTATTA-3'	96
PARK7 G/A mutant	Forward	5'-TAATACGACTCACTATAGG-3'	19
	Reverse	5'-ACACCAGGCTGAAAATGAACGCACGCCAGCACGCACGCGCGCCGTTC GGCCTCTCTCAGTCTCACACTGCGCAGACTCCTATAGTGAGTCGTATTA-3'	96
PINK1 WT G4	Forward	5'-TAATACGACTCACTATAGGCACCGCCCCAAGTTTGTGACCGGCGG GGGACGCCGGTGGTGGCGGCAGCGGCGGCTGCGGG-3'	83
	Reverse	5'-TCGCACCGCCATGGTGGCGCCGCGGCCCGGTGCCCCGCAGCCGCCGC TGCCGCCACCACCGGCGTCCCCCGCCGGTCACAAC-3'	83
PINK1 G/A mutant	Forward	5'-TAATACGACTCACTATAGGCACCGCCCCAAGTTTGTGACCGGCGAG AGACGCCGGTGATGGCGACAGCGGCGACTGCGAG-3'	83
	Reverse	5'-TCGCACCGCCATGGTGGCGCCGCGGCTCGGTGCTCTCGCAGTCGCCGC TGTCGCCATCACCAGGCGTCTCTCGCCGGTCACAAC-3'	83
PLA2G6 WT G4	Forward	5'-TAATACGACTCACTATAGG-3'	19
	Reverse	5'-GCGCCGAGGAAGTTGGGAACGGACCCCCAGGCCCGCCACCCGCG AGGTCACTCACCTCGGCTTACCTATAGTGAGTCGTATTA-3'	86
PLA2G6	Forward	5'-TAATACGACTCACTATAGG-3'	19

G/A mutant	Reverse	5'-GCGCCGAGGAAGTTGGGGAACGGACTCTCAGGTCTCGCTCACTCGCG AGGTCACCTCGGCTTACCTATAGTGAGTCGTATTA-3'	86
PRKN WT G4	Forward	5'-TAATACGACTCACTATAGGATTCTAGGGCCGGGCGCGGGGGCGGGG AGGCCTGGAGGATTTAACCCAGGAGAGCCGCTGG-3'	81
	Reverse	5'-GCGCGCAGCGGCCAGCCGCGCCTCCCACCAGCGGCTCTCCTGGGTT AAATCCTCAGGCCTCCCCGCCCCGCGCCCGG-3'	81
PRKN G/A mutant	Forward	5'-TAATACGACTCACTATAGGATTCTAGGGCCGAGCGGAGAGCGAGAA GGCCTGAAGGATTTAACCCAGAAGAGCCGCTGG-3'	81
	Reverse	5'-GCGCGCAGCGGCCAGCCGCGCCTCTCACCAGCGGCTCTTCTGGGTT AAATCCTCAGGCCTTCTCGCTCTCGCGCTCGG-3'	81
SNCA WT G4	Forward	5'-TAATACGACTCACTATAGGAGAAGGAGAAGGAGGAGGACTAGGAGGA GGAGGACGGCGACGACCAGAAGGGGCCAAGAG-3'	80
	Reverse	5'-CGCGTCGCGGCGCTCGGTCGCTCGCCCCCTCTCTGGGCCCTTCTGGT CGTCGCCGTCTCCTCCTCCTAGTCCTCCTC-3'	80
SNCA G/A mutant	Forward	5'-TAATACGACTCACTATAGGAGAAGGAGAAGGAGAAGGACTAGAAGGA GAAGGACGACGACGACCAGAAGAGACCCAAGAG-3'	80
	Reverse	5'-CGCGTCGCGGCGCTCGGTCGCTCGCTCTCTCTTGGGTCTCTTCTGGT CGTCGTCGCTTCTCCTTCTAGTCCTTCTC-3'	80
VPS35 WT G4	Forward	5'-TAATACGACTCACTATAGGCTAGAGAGGGCGGGGCTTGGAGGGGCCG CAGCGTCACATGACCGCGGGAGGCTACGCGGGGGC-3'	83
	Reverse	5'-TCCCCAGAGCCTGCAGCAAGCAGCACCCGCCCCGCGCGTAGCCTCCCG CGGTCATGTGACGCTGCGGCCCTCCAAGCCCCGC-3'	83
VPS35 G/A mutant	Forward	5'-TAATACGACTCACTATAGGCTAGAGAGGGCGAGACTTGGAGAGACCG CAGCGTCACATGACCGCGAGAGGCTACGCGGAGAC-3'	83
	Reverse	5'-TCCCCAGAGCCTGCAGCAAGCAGCACTCGTCTCGCGGTAGCCTCTCG CGGTCATGTGACGCTGCGGTCTCTCCAAGTCTCGC-3'	83
5'UTR LRRK2	WT G4	5'-ACTAGTGCGCTGGCTGCGGGCGGTGAGCTGAGCTCGCCCCGGGGAG CTGTGGCCGGCGCCCCTGCCGTTCCCTGAGCAGCGGACGTTTCATGCTGGG AGGGCGGCGGGTTGGAAGCAGGTGCCACCGTCGAC-3'	133

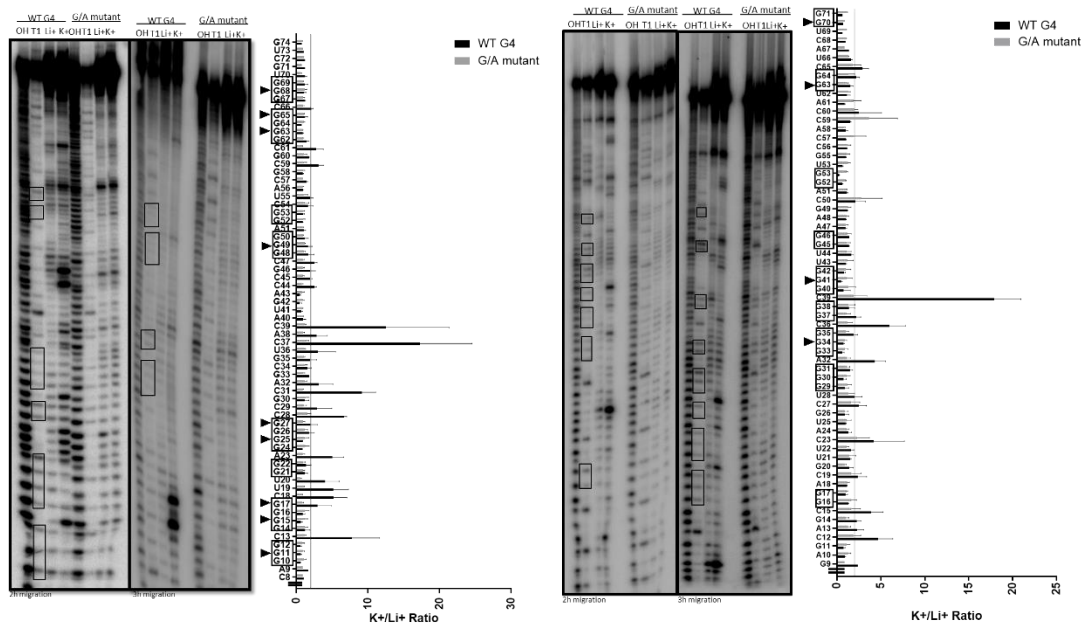
	G/A mutant	5'-ACTAGTGCCTGGCTGCGGGCGGTGAGCTGAGCTCGCCCCGGGGAG CTGTGGCCGCGCCCCCTGCCGTTCCCTGAGCAGCGGACGTTTCATGCTGGG AGAGCGGCGAGTTGGAAGCAGGTGCCACCGTCGAC-3'	133
5'UTR PRKN	WT G4	5'-ACTAGTATTCTAGGGCCGGGCGCGGGGCGGGGAGGCCTGGAGGAT TTAACCCAGGAGAGCCGCTGGTGGGAGGCGCGGCTGGCGCCGCTGCGCGC ATGGGCCTGTTCTGGCCCGCAGCCGCCACCTACCCAGTGACCGTCGAC-3'	146
	G/A mutant	5'-ACTAGTATTCTAGGGCCGAGCGGAGAGCGAGAAGGCCTGAAGGAT TTAACCCAGAAGAGCCGCTGGTGGAGAGGCGCGGCTGGCGCCGCTGCGCGCA TGGGCCTGTTCTGGCCCGCAGCCGCCACCTACCCAGTGACCGTCGAC-3'	146
5'UTR SNCA	WT G4	5'-ACTAGTAGGAGAAGGAGAAGGAGGAGGACTAGGAGGAGGAGGACGG CGACGACCAGAAGGGGCCCAAGAGAGGGGGCGAGCGACCGAGCGCCGCG ACGCGGAAGTGAGGTGCGTGCGGGCTGCAGCGCAGACCCCGGCCGGCCCC CTCCGAGAGCGTCCTGGGCGCTCCCTCACGCCTTGCCTTCAAGCCTTCTGCCT TTCCACCCTCGTGAGCGGAGAACTGGGAGTGGCCATTCGACGACAGTGTGG TGTAAGGAATTCATTAGCCGTCGAC-3'	275
	G/A mutant	5'-ACTAGTAGGAGAAGGAGAAGGAGAAGGACTAGAAGGAGAAGGACGA CGACGACCAGAAGAGACCCAAGAGAGAGAGCGAGCGACCGAGCGCCGCGA CGCGGAAGTGAGGTGCGTGCGGGCTGCAGCGCAGACCCCGGCCGGCCCCCT CCGAGAGCGTCCTGGGCGCTCCCTCACGCCTTGCCTTCAAGCCTTCTGCCTTT CCACCCTCGTGAGCGGAGAACTGGGAGTGGCCATTCGACGACAGTGTGGTG TAAAGGAATTCATTAGCCGTCGAC-3'	275
5'UTR VPS35	WT G4	5'-ACTAGTGCTAGAGAGGGCGGGGCTTGGAGGGGCCGAGCGTCACATG ACCGCGGAGGCTACGCGGGGGCGGGTGTGCTTGTGCTGACAGGCTCTGGG GAGTCGCCGTCGAC-3'	111
	G/A mutant	5'-ACTAGTGCTAGAGAGGGCGGAGACTTGGAGAGACCGCAGCGTCACATG ACCGCGAGAGGCTACGCGGAGACGAGTGTGCTTGTGCTGACAGGCTCTGGGG AGTCGCCGTCGAC-3'	111

Supplementary Table S2: Pattern of the mutations used for the conception of the G/A mutant oligonucleotides

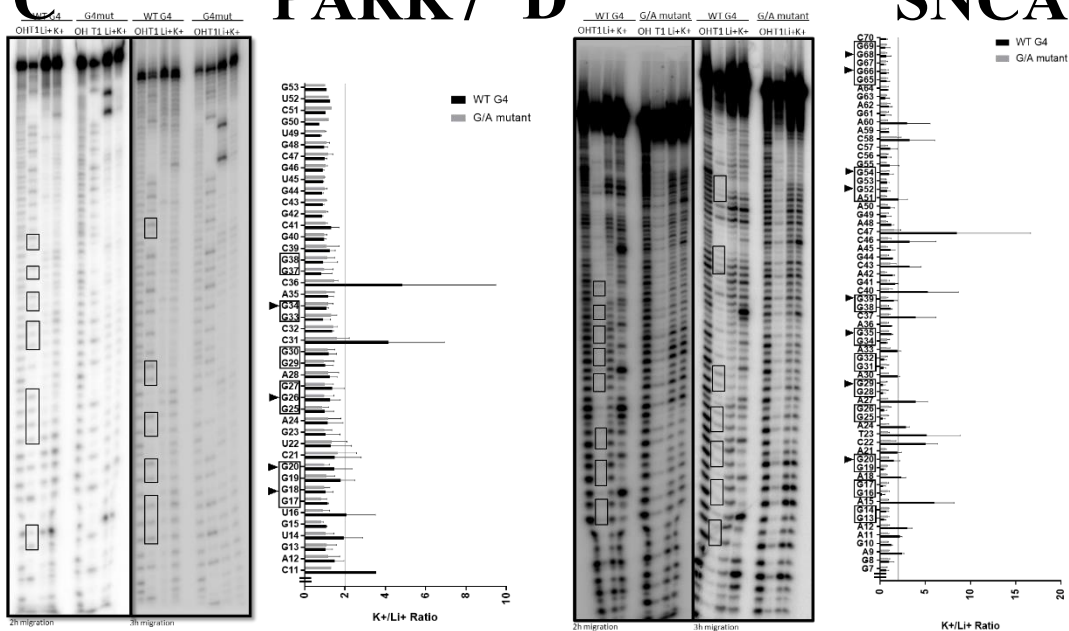
<i>GAG</i>
<i>GAGA</i>
<i>GAGAG</i>
<i>GAGAGA</i>
<i>GAG GG GAG</i>
<i>GAG GG GG GAG</i>
<i>GAG GG GA GG GAG</i>
<i>GAG GG GA GG GA GAG</i>

Supplementary Figures

A VPS35 B LRRK2



C PARK7 D SNCA



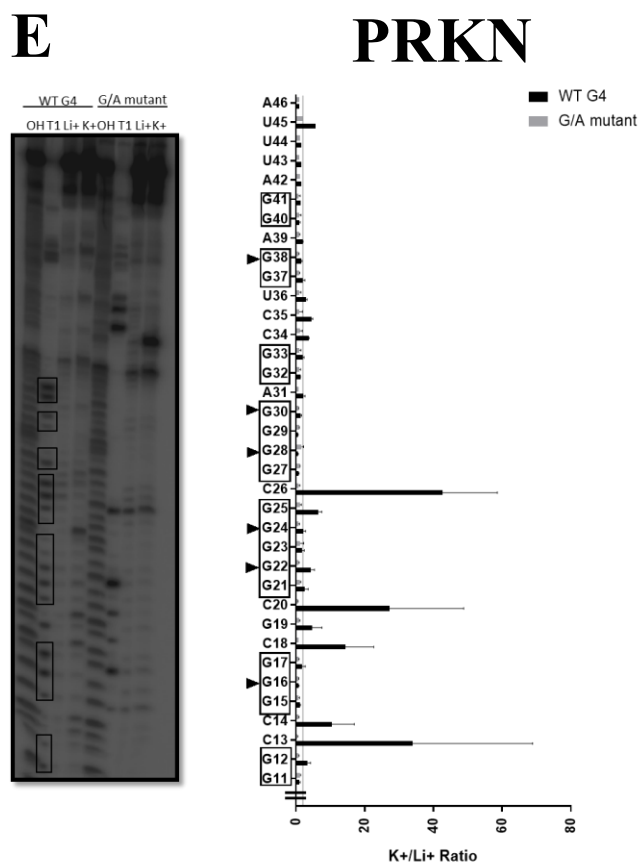


Figure S1: *In-line* probing of the five positive pG4s

The RNA G4 foldings of: A) VPS35; B) LRRK2; C) PARK7; D) SNCA; and, E) PRKN were evaluated by *in-line* probing. Left: Denaturing polyacrylamide gels of the *in-line* probing reactions. Lanes 1 to 4 are the WT G4, and lanes 5 to 8 are the G/A mutant version. Lanes 1 and 5 are the NaOH Ladder; lanes 2 and 6 are the T1 Ladder; lanes 3 and 7 were obtained under the Li⁺ condition; and, lanes 4 and 8 were obtained under the K⁺ condition. The conversion of the radioactive gel into data was made using the SAFA software and is presented on the right. The sequence is indicated on the y-axis. The x-axis represents the K⁺/Li⁺ ratios of the relative band intensities of the WT G4 and the G/A mutant G4 for each nucleotide. The K⁺/Li⁺ ratios are shown black for the WT G4 and in grey for the G/A mutant G4. The nucleotides are identified with their *in vitro* transcript position. The nucleotides denoted with an arrowhead are those substituted for adenine in the G/A mutant versions. "||"

indicates the minimal limit of quantification. The boxed guanines are those that could be expected to form a G4. The dotted line represents the twofold threshold that denotes a significant gain in flexibility. Each bar represents the average of two independent experiments, and the error bars represent the standard deviations.

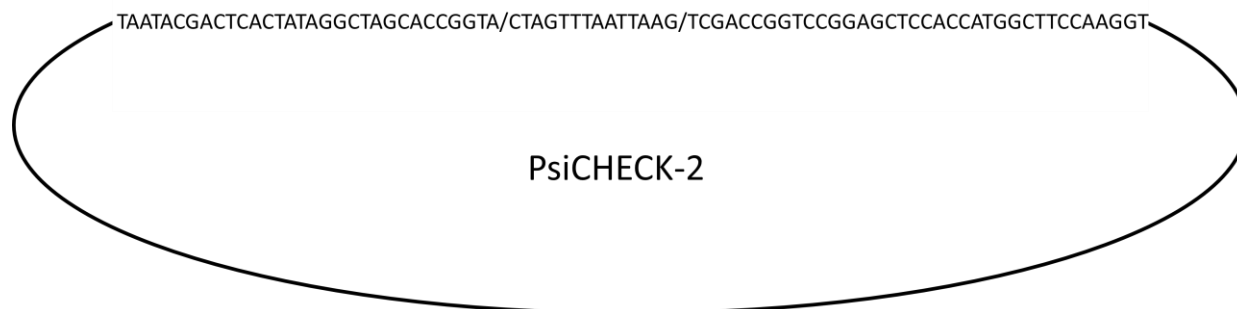


Figure S2: Modified version of PsiCHECK-2 used with either the 5'UTR WT or the G/A mutant G4s of VPS35, LRRK2, SNCA and PRKN

Schematic of the modified version of PsiCHECK-2. The slash symbols (/) indicate the two cleavage sites where the 5'UTR were inserted.

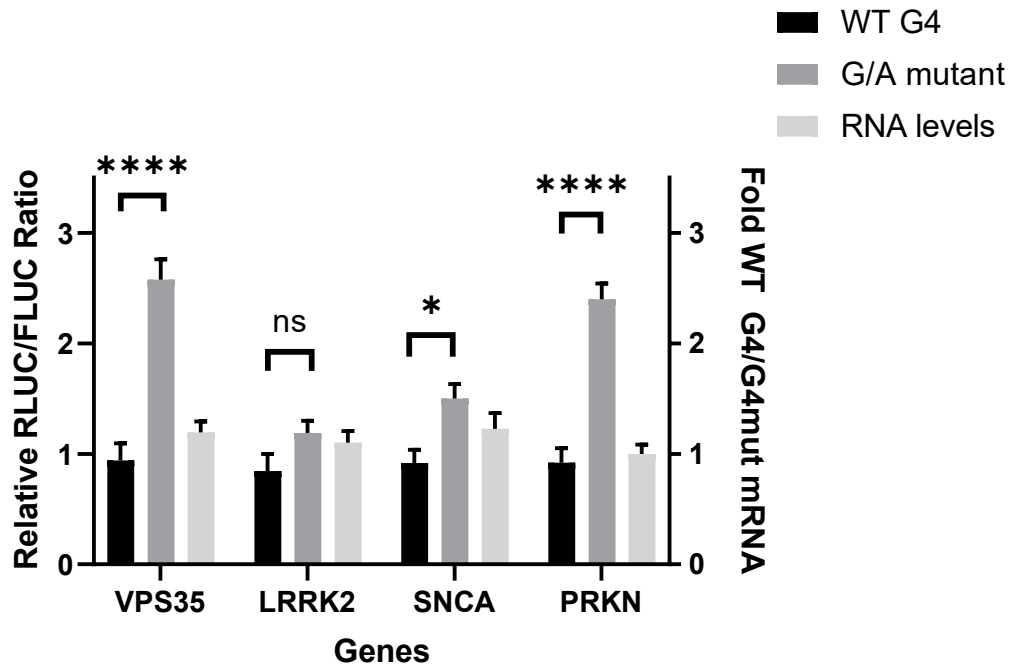
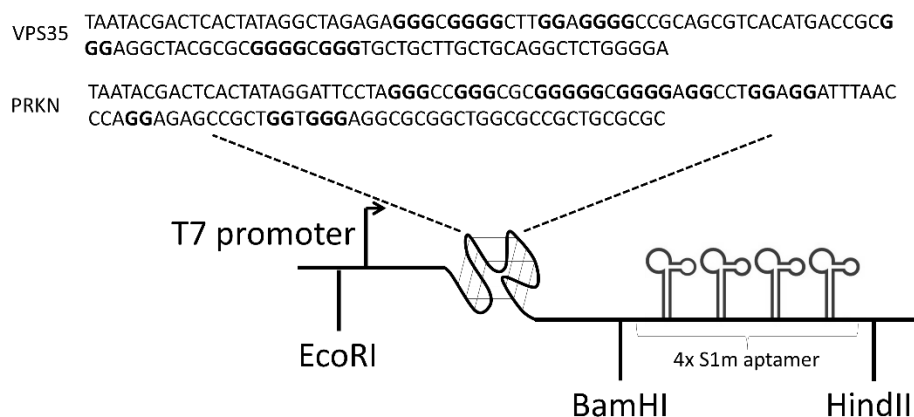


Figure S3: Luciferase assays in HEK293T cells

The relative RLUC/FLUC ratios for the luciferase activities of the 5'UTRs of VPS35, LRRK2, SNCA and PRKN in the HEK293T cell line are shown in the left y-axis. The 5'UTRs were cloned in the Psicheck 2 plasmid. The RNA levels for the « WT / G/A mutant » for RLUC are shown on the left y-axis. The results are the means of at least two independent experiments, and the error bars represent the SEM. P-values were calculated by two way ANOVA. (*) $P < 0.05$, (**) $P < 0.01$, (***) $P < 0.001$, (****) $P < 0.0001$.

pMA-RQ VPS35-4xS1m and pMA-RQ PRKN-4xS1m



pMA-RQ VPS35 G/A mutant-4xS1m and pMA-RQ PRKN G/A mutant-4xS1m

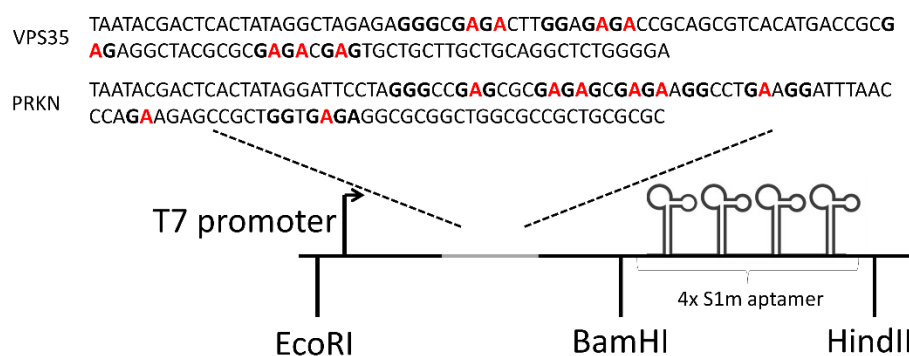


Figure S4: Sequences of the wild type and the G/A mutant G4 constructs of pMA-RQ-VPS35/PRKN-4xS1m

The DNA sequences used for both the NMM and the *in-line* probing experiments contained 94 and 93 nts in the cases of VPS35 and PRKN, respectively, that were fused to the pMA-RQ 4xS1m plasmid. Top: pMA-RQ VPS35-4xS1m and pMA-RQ PRKN-4xS1m. Bottom: pMA-RQ VPS35 G/A mutant-4xS1m and pMA-RQ PRKN G/A mutant-4xS1m.

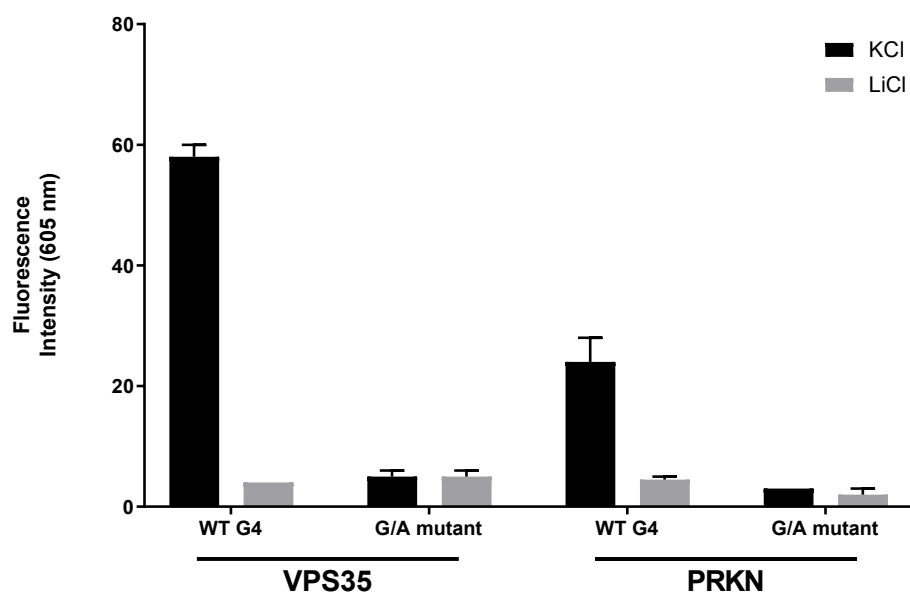


Figure S5: NMM fluorescence assays of the chimeric RNAs that included the 4xS1m aptamer

Fluorescence assays of either the WT or the G/A mutant G4s. The RNAs (300 pmol) were mixed with 0.5 mM NMM and 100 mM of either KCl or LiCl and were then incubated for 30 min at room temperature. The samples were excited at 399 nm and the emission spectra were monitored at 605 nm. All experiments were performed in duplicate. The error bars represent the SEM.

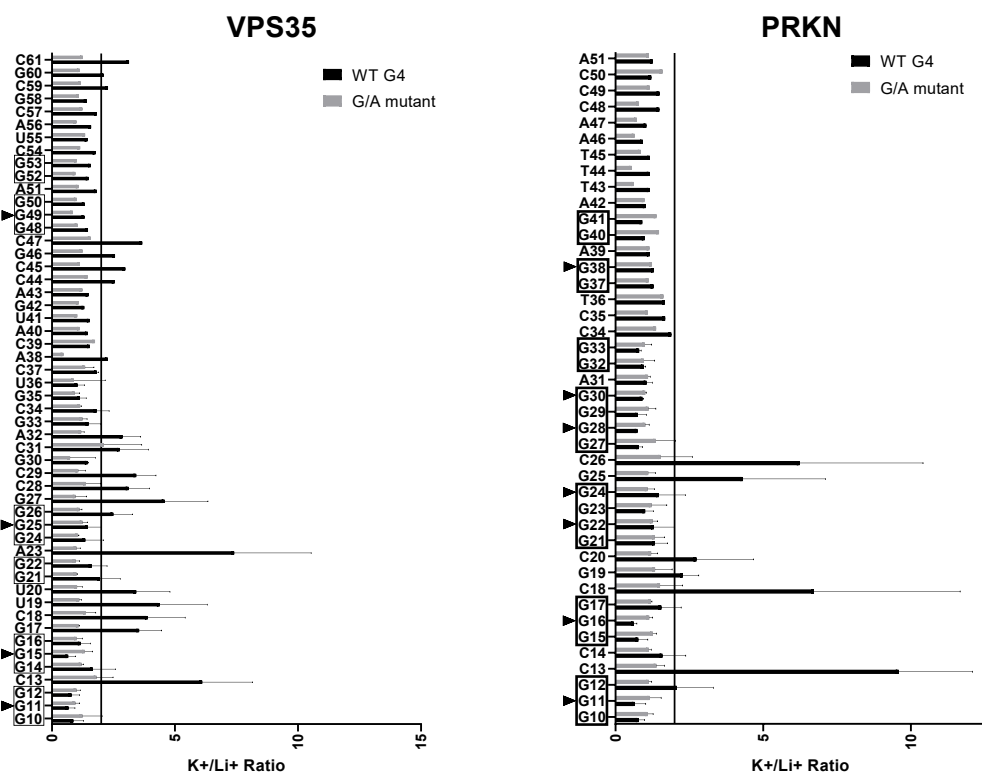


Figure S6: *In-line* probing of the chimeric RNAs containing the 4xS1m aptamer

In-line probing results for the VPS35 (left) and PRKN (right) RNA G4s. The results are presented after having converted the radioactive gel results into data using the SAFA software. The sequences are indicated on the y-axes. The x-axes represent the K^+/Li^+ ratios of each band's intensity for the WT and the G/A mutant G4s. The K^+/Li^+ ratios are shown in black for WT and in gray for the G/A mutant. The nucleotides are identified with their *in vitro* transcript position. The nucleotides denoted with an arrowhead are those substituted for adenine in the G/A mutant version. The boxed guanines are those that could be expected to form a G4. The dotted line represents the twofold threshold that denotes a significant gain in flexibility. Each bar represents the average of two independent experiments, and the error bars represent the standard deviation.

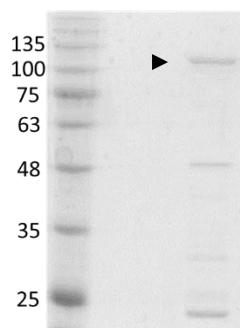


Figure S7: Purification profile of the fused GST-GNL1 protein

The recombinant protein was electrophoresed through a 12% SDS-PAGE gel which was then colored with Coomassie Blue. Lane 1: Labeled Blue Elf Pre-Stained Protein Ladder (KDa, FroggaBio); Lane 2: 2 μ g of GST-GNL1 purification. The arrowhead indicates the GST-GNL1 protein (111 kDa).

DISCUSSION

Analyse de l'enrichissement des G-quadruplexes dans les maladies du système nerveux

La détection de G-quadruplexes est un domaine qui prend beaucoup d'importance dans différents types de maladies. Il est en effet possible d'associer plusieurs d'entre eux à des dérégulations pathologiques. Par ailleurs, les neurones font parties des types de cellule possédant beaucoup de mécanismes de régulation post-transcriptionnels. En effet, cette complexité est directement due au fait que les neurones possèdent différentes fonctions très spécialisées. De plus, la différenciation des neurones est un phénomène faisant appel à une multitude de mécanismes transcriptionnels et post-transcriptionnels (Hamby et al., 2008). C'est par ailleurs le cerveau qui présente les plus grands niveaux d'épissage alternatif et d'édition (Kiebler et al., 2013). De simples problèmes dans la régulation peuvent donc rapidement amener la mort de ceux-ci. C'est ce qui a porté notre groupe à vouloir étudier s'il existait un enrichissement de pG4 dans les maladies associées au système nerveux. Pour ce faire, il y a utilisation d'une base de données comprenant les différentes protéines étant régulées dans les maladies du système nerveux. Plus précisément, les données requises ont été trouvées dans le *Kyoto Encyclopedia of Genes and Genomes* (KEGG). Le *KEGG Disease* a été utilisé pour extraire chacune des protéines associées aux différentes maladies neuronales. Cette base de données a été choisie puisqu'elle comporte la plupart des protéines étant modifiées dans chacune des maladies. Par ailleurs, il est possible d'y retrouver certaines sous-catégories, comme les maladies neurodégénératives. La séquence nucléotidique FASTA reliée aux protéines dérégulées dans le système nerveux pour chacun des variants (ARNm alternatifs) a alors pu être associée grâce aux données annotées de transcrit mature du NCBI.

Ensuite, le prédicteur G4NN a été déterminé comme étant l'outil idéal pour la détection de pG4. En effet, celui-ci possède beaucoup moins de biais que les autres outils puisqu'il n'est pas basé sur un motif, mais bien sur l'apprentissage profond avec différentes séquences positives et négatives ayant été testées en laboratoire. Pour identifier un pG4 positif, un score égal ou supérieur à 0.5 a été utilisé pour maximiser la sensibilité et la spécificité de la détection (Garant et al., 2017). Avec l'utilisation de ce paramètre, 85% de tous les gènes

présents dans la base de données du système nerveux ont été identifiés comme pouvant former un pG4. Lorsque les ARNm alternatifs de toutes les protéines dérégulées (Refseq Accession Numbers) étaient évalués, c'était plutôt 80% des variants qui pouvaient former un pG4. À première vue, ce résultat semblait être relativement élevé. De ce fait, une analyse comparant la quantité de pG4 associée aux variants à des ARN et des ARNm sélectionnés de façon aléatoire a permis de conclure qu'il existe en effet un enrichissement de pG4 (Figure 7). Par ailleurs, les transcrits des protéines décrites par une seconde base de données, le *human protein atlas*, comme étant enrichis dans le système nerveux ont eux aussi un enrichissement au niveau de la quantité de pG4. Il est alors possible de penser que l'enrichissement observé dans les gènes dérégulés du système nerveux soit dû à un enrichissement global.

Par la suite, un sous-groupe a été sélectionné puisqu'une trop grande quantité de pG4 était présente dans la base de données pour être analysés manuellement *in vitro*. C'est pour cette raison que les gènes reliés aux maladies neurodégénératives, un sous-groupe présent dans la base de données du KEGG, ont été isolés. Il est possible de voir que dans ce sous-groupe 67% des gènes peuvent former un pG4. Étrangement, l'enrichissement de pG4 au niveau des gènes possédant un G4 semble être perdu. Cependant, la position des pG4 dans ces gènes semble être assez similaire dans les deux groupes. Il est en effet possible d'observer un enrichissement des G4 dans le CDS, suivi du 3'UTR et du 5'UTR. Ces résultats peuvent apparaître à priori comme étranges, puisque les G4 sont bien connus pour être enrichis dans le 5'UTR et le 3'UTR. Une hypothèse est que la quantité de G4 est ici évaluée plutôt que la densité en G4. En effet, une analyse de la taille moyenne des différentes régions (CDS 1663 nts, 3'UTR 1470 nts et 5'UTR 259 nts) permet d'observer une grandeur inférieure pour les 5'UTR et le 3'UTR à celle du CDS (Piovesan et al., 2016). Cela peut donc être responsable de l'enrichissement observé dans le CDS.

Ensuite, puisque le groupe étudié avait toujours une grande taille, les maladies présentant dans la catégorie des ND ont manuellement été étudiées pour en sélectionner seulement une possédant une grande proportion en G4. C'est pour cette raison que la maladie de Parkinson a été sélectionnée. En effet, 94% des gènes dérégulés dans cette maladie pouvaient former

au moins un pG4. De plus, environ 80% des gènes (soit 11 gènes) qui possédaient des pG4 en avaient au moins un dans leur 5'UTR. Cette fois-ci, la plupart des pG4 étaient situés dans le CDS et le 5'UTR. La quantité de gènes à étudier était alors raisonnable pour passer à des analyses *in vitro*.

Les 11 séquences pouvant former un pG4 dans leur 5'UTR ont été retenues pour des analyses plus approfondies (Tableau 2). Une analyse manuelle des séquences a permis de conclure qu'étonnamment, seule une d'entre elles (PLA2G6) avait les nucléotides nécessaires pour former un G4 canonique à trois tétrades et ayant des boucles allant de 1 à 7 nucléotides (Tableau 2). Cependant, il était tout de même possible de retrouver des G4 à deux quartets dans toutes les autres séquences. Certaines d'entre elles avaient même la possibilité de former des G4 possédant des longues boucles et d'autres possédant des quartets avec un espacement (*bulge*). Une hypothèse pour expliquer la présence de ces G-quadruplexes de type non canonique pourrait être que ceux-ci sont associés à différents mécanismes de régulation post-transcriptionnelle. Une caractéristique de ces structures est que leur stabilité est habituellement plus faible, ce qui fait d'eux des G4 étant plus facilement régulés par des facteurs en *trans*. Par ailleurs, il pourrait exister des protéines spécifiques étant en mesure de lier ce type de structure. Comme il est très difficile de cibler les G4 canoniques, l'obtention de structures non canoniques est très intéressante puisque cela laisse envisager une régulation spécifique par le développement de molécules organiques ou d'oligonucléotides. Cela pourrait donc être une option pour contrebalancer la dérégulation de certaines protéines dans la maladie de Parkinson.

Analyse de la formation des G-quadruplexes *in vitro*

La formation des 11 pG4 en 5'UTR a ensuite été étudiée *in vitro*. Les données de sortie du logiciel G4 *Screeener* donnent des fenêtres nucléotidiques de 60 nucléotides et se déplacent de 10 nucléotides pour chacune des nouvelles fenêtres. Comme plusieurs fenêtres étaient présentes pour un même pG4, celles possédant le plus grand score pour chacun des pG4 et étant situées dans l'ARNm alternatif le plus abondant ont été utilisées pour la recherche (Tableau 2). Des nucléotides en amont et en aval de la séquence (jusqu'à 25 nucléotides) ont par ailleurs été ajoutés pour mieux mimer l'environnement de la cellule. En effet, il a été

montré que l'ajout d'un contexte nucléotidique peut avoir un impact sur la formation d'un G4 (Beaudoin et al., 2014).

Les pG4 ont tout d'abord été étudiés par essais de fluorescence avec le N-méthyle mésoporphyrine IX (NMM). Ce ligand a été sélectionné dû à sa sensibilité et spécificité élevée envers les G4 d'ARN. Une intensité de fluorescence arbitraire minimale de 25 unités de fluorescence a été choisie. En effet, une large quantité de G4 *in vitro* a été étudié au laboratoire au cours des dernières années et cette valeur est habituellement en mesure de discriminer les positifs des négatifs. L'utilisation du ligand NMM a permis la détection de six G4 (Figure 8). Parmi ceux-ci, quatre possédaient un signal relativement élevé d'au moins 50 unités de fluorescence à 605 nm, ce qui est souvent associé à de vrai G4. Les deux autres possédaient une intensité intermédiaire entre 25 et 40 unités de fluorescence, ce qui demande habituellement une confirmation avec une seconde technique. La fluorescence a donc permis d'identifier que 55% des potentiels pG4 sélectionnés seraient en mesure de former des G4. À première vue, ce résultat semble être assez faible puisque le score G4NN est reconnu pour bien performer pour la détection de pG4. Cependant, il est important de mentionner qu'un G4 qui ne se forme pas en conditions *in vitro* pourrait quand même se former dans des conditions *in cellulo*, que ce soit grâce à des conditions différentes de tampon ou bien grâce à certaines protéines spécifiques. Par ailleurs, des structures non canoniques de G4 pourraient être en mesure de donner des faux négatifs en empêchant la liaison du ligand. Dans le cas de l'étude, cela ne cause pas de problème puisque la fluorescence au NMM est utilisée en phase exploratoire pour trouver des candidats potentiels. Il n'est donc pas surprenant de ne pas trouver tous les G4 présents dans le 5'UTR des gènes.

L'utilisation d'au moins deux techniques pour la confirmation de la formation d'un G4 est assez courante. Tous les G-quadruplexes détectés en fluorescence ont alors été confirmés avec une seconde technique, la cartographie *in-line*. Avec cette technique, quatre G4 ayant une intensité élevée en fluorescence ont été montrés comme étant en mesure de former un G4 en cartographie *in-line* (Figure 9 et S1). La présence de G4 dans les ARNm de eiF4G1 et PARK7 n'a pas pu être démontrée avec cette technique. En effet, les résultats de eiF4G1 ont donné un dédoublement dans les bandes, ce qui peut être caractéristique d'une population

mixte d'ARN présents en solution (gel non montré). Pour ce qui est de PARK7, il était possible de n'observer aucun enrichissement significatif dans le clivage des résidus en conditions de potassium. Encore une fois, rien n'exclut que ces deux derniers aient la capacité de se former *in cellulo*. Les résultats de cartographie nous ont donc motivés à abandonner ces deux candidats pour la suite des expériences.

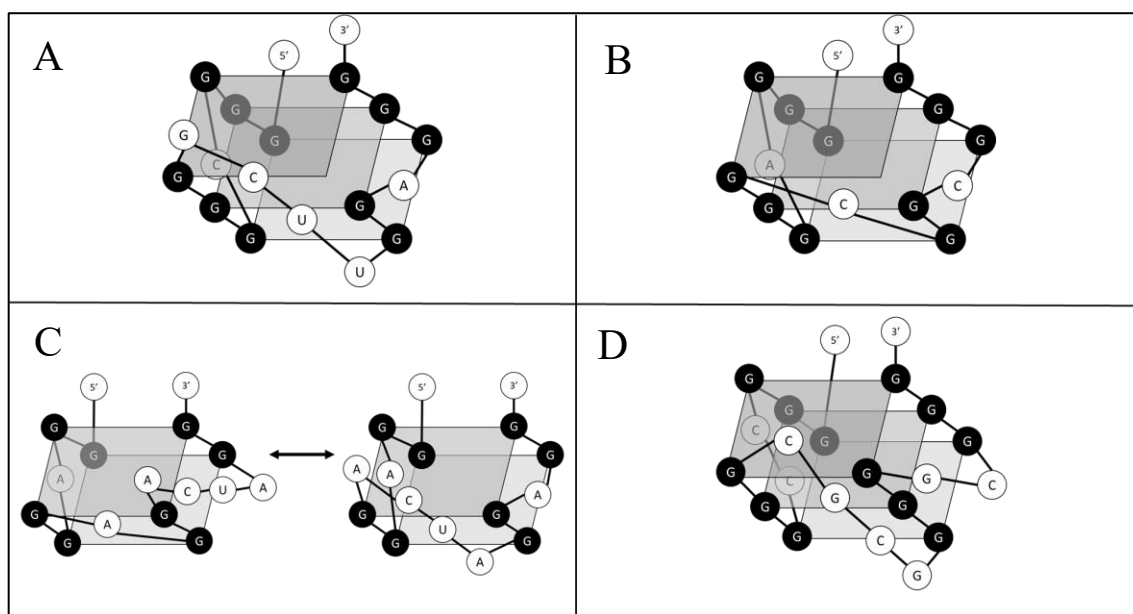


Figure 13 : Propositions de G-quadruplexes observées par cartographie *in-line*

Représentation imagée des différents candidats: **A) VPS35 B) LRRK2 C) SNCA D) PRKN.**

La double flèche représente un état d'équilibre.

La cartographie *in-line* a aussi permis d'obtenir une idée globale du type de G-quadruplexe présent dans chacun des ARN (Figure 13). Les résultats n'étant cependant pas absolus, d'autres options auraient pu être représentées. Dans l'ARN de PRKN, un G4 canonique à 3 tétrades semble être présent. Pour ce qui est de celui présent dans SNCA, il semble y avoir un équilibre observé entre plusieurs G4 à deux tétrades. Pour finir, l'ARN LRRK2 et VPS35, il semble y avoir des G-quadruplexes non canoniques, où il y aurait un 3 tétrades, mais l'un des brins serait seulement formé par un doublet de guanines. Un point intéressant à noter est que les G4 formé par PRKN et VPS35 semblent se former avec des guanines présentes dans le contexte nucléotidique ajouté en amont de la séquence. Cela est une bonne preuve que

l'ajout de ce contexte a un impact sur la formation des G4. Cette analyse donne une vue des nucléotides accessibles présents dans les G4.

Les techniques *in vitro* ont donc permis l'identification de quatre nouveaux G4 d'ARN situés dans les 5'UTR, soit PRKN, VPS35, LRRK2 et SNCA. Au niveau fonctionnel, PRKN code pour une protéine, qui à la suite de son activation par PINK1, est responsable de l'ubiquitination des protéines se retrouvant à la membrane mitochondriale pour entraîner la mitophagie (Gladkova et al., 2018). Quant à elle, VPS35 code pour une des sous-unités du complexe WASH responsable du trafic endosomal (Williams et al., 2017). De son côté, LRRK2 code pour une composante importante de la chaîne respiratoire mitochondriale (Aufschnaiter et al., 2018). Finalement, SNCA code normalement pour une protéine étant impliquée dans les vésicules synaptiques (Al-Wandi et al., 2010).

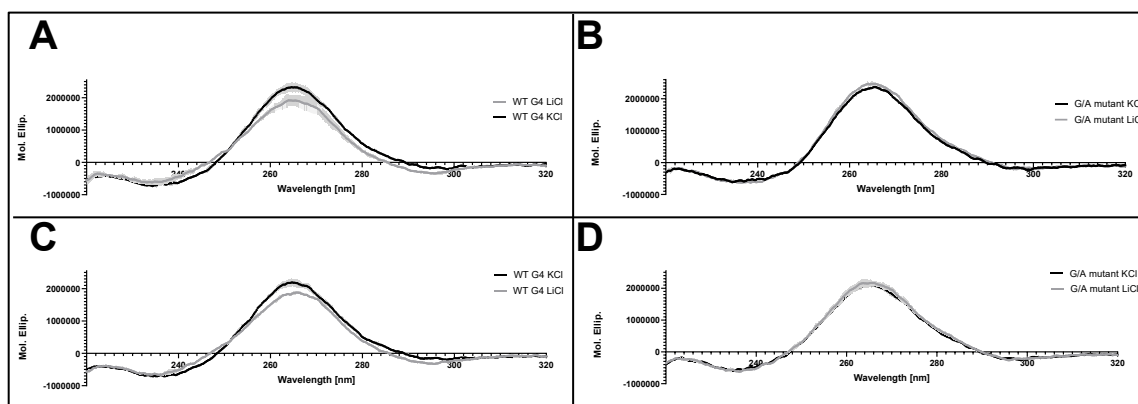


Figure 14: Dichroïsme circulaire des structures G-quadruplexes de VPS35 et PRKN Spectres de dichroïsme circulaire allant de 220 à 320 nm réalisés avec 4 μ M d'ARN de VPS35 (A, B) ou PRKN (C, D) réalisé avec la séquence de type sauvage (WT) (A, C) ou le G/A mutant (B, D) en présence de 100 mM LiCl ou KCl à 25°C. Les régions ombragées représente l'erreur sur chacune des valeurs.

Une analyse plus avancée de la formation du G4 de PRKN et de VPS35 (dû aux résultats *in cellulo*) a été réalisée à l'aide de trois techniques supplémentaires. Pour commencer, les deux structures secondaires du G4 ont été évaluées en dichroïsme circulaire (Figure 14). Le tout a été réalisé avec la même séquence que celle utilisée pour les essais de fluorescence et la cartographie *in-line*. Lorsque les spectres des deux gènes WT sont analysés, il y a présence d'un pic qui est plus négatif à 240 nm et un plus positif à 264 nm en présence de potassium

qu'en présence de lithium. Cela est caractéristique de la présence d'un G4. Par ailleurs, il est possible de voir que cet effet n'est pas observé avec la séquence mutante.

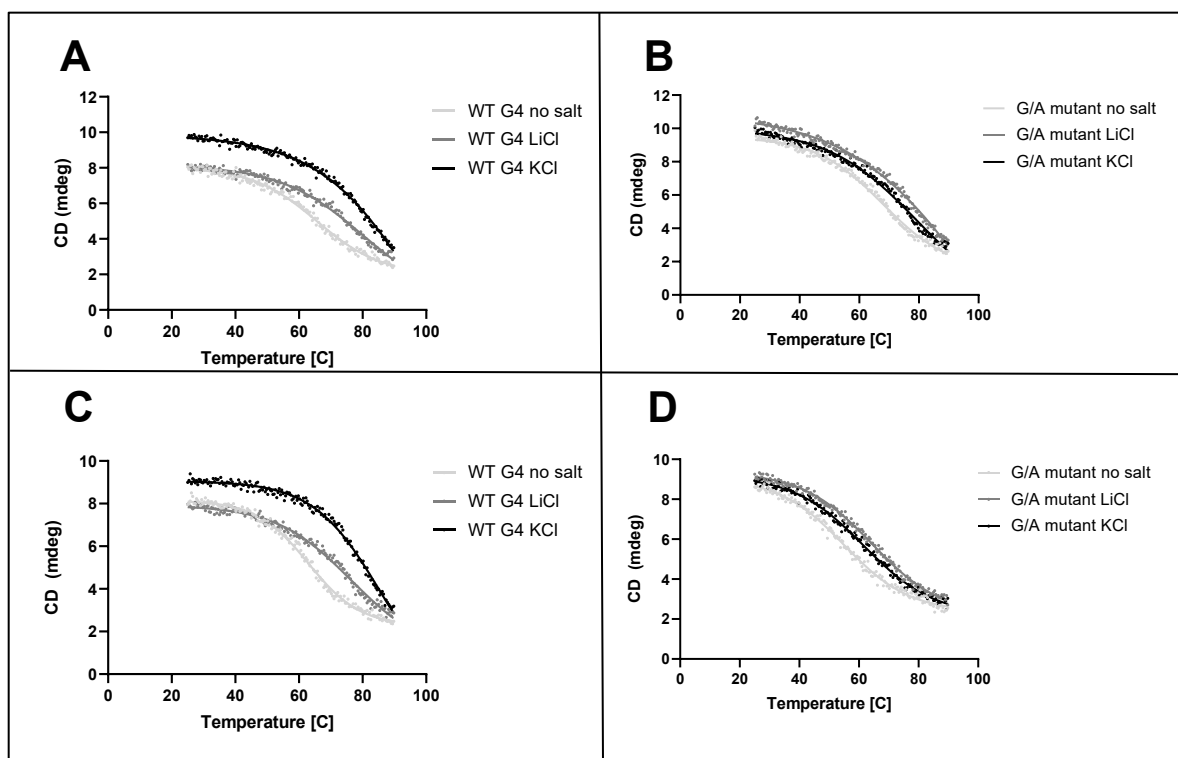


Figure 15: Dénaturation thermique des structures G-quadruplexes de VPS35 et PRKN
Dénaturation thermique monitorée à 264 nm de 4 μ M d'ARN pour le G4 de VPS35 (A, B) et de PRKN (C, D) fait avec le WT G4 (A, C) ou le G/A mutant (B, D) en présence de 100 mM LiCl ou KCl ou en absence de sel de 25 °C à 90 °C.

Ensuite, la stabilité thermique de ces séquences a été évaluée à 264 nm sur une variation de température allant de 25°C à 90°C (Figure 15). Il est possible de voir que les séquences d'ARN de type sauvage (WT) dénaturées en présence de potassium ont une plus grande résistance envers la dénaturation thermique que les séquences mutantes. Pour finir, il ne semble y avoir qu'un seul état d'équilibre dans chacune des structures observées, soit un équilibre entre le G4 formé et la séquence linéaire. Des régressions non linéaires (*Hill equation*) ont ensuite été effectuées pour déterminer précisément les températures de dénaturation. Cependant, ces résultats sont à prendre avec une certaine réserve puisque dans plusieurs cas, aucune température ne pouvait être déterminées avec certitude due à une dénaturation incomplète à 90°C (Tableau 3). Globalement, les données montrent une

augmentation de la température de dénaturation dans l'échantillon de type sauvage qui n'est pas observé en dans le G/A mutant. Un exemple de donnée incorrecte est la courbe de dénaturation de VPS35 G/A mutant en lithium où il semble y avoir une stabilisation. L'analyse des paramètres montre que cette valeur est due à un manque de points dans le plateau inférieur. Il est impossible avec cette technique d'augmenter la température sans changer les conditions (due à l'évaporation du solvant). Par ailleurs, un changement de solvant ne serait pas représentatif de la solubilisation des G4 *in cellulo*. Bien que la valeur soit incorrecte, les graphiques montrent bien le phénomène de stabilisation accru en potassium seulement pour l'ARN de type sauvage.

Tableau 1 : Températures de dénaturation déterminées par dénaturation thermique

Gènes		Sans sel (°C)	LiCl (°C)	KCl (°C)
PRKN	WT G4	62,9	76,3	85,2
	G/A mutant	55,0	64,3	61,6
VPS35	WT G4	65,6	80,1	99,7
	G/A mutant	67,9	99,5	79,0

La capacité des deux G4 à entrainer un blocage a ensuite été évaluée par des essais d'arrêt de la transcriptase inverse (RTS) (Figure 16). Premièrement, pour PRKN, un site d'arrêt est observé seulement en potassium dans la condition WT, ce qui montre bien que le G4 est assez stable pour induire un arrêt. Deuxièmement, pour VPS35, il est possible cette fois-ci de voir un site commun d'arrêt aux deux conditions lorsque le WT G4 est comparé avec le G/A mutant. Cela peut signifier que le G4 possède une certaine capacité à se former en présence de lithium. Cependant, seule la condition WT en potassium possède un deuxième site d'arrêt. Ce phénomène est probablement dû à l'augmentation de la stabilité du G4 par le potassium. Un point intéressant à noter est que la présence des G4 ne permet pas d'arrêter totalement la progression de la transcriptase. Cela laisse donc penser que *in cellulo*, le ribosome ne doit pas être totalement arrêté lors de sa progression sur l'ARNm. En bref, la réalisation de cinq techniques différentes *in vitro* a permis de montrer dans tous les cas la formation du G-quadruplexe d'ARN de PRKN et de VPS35.

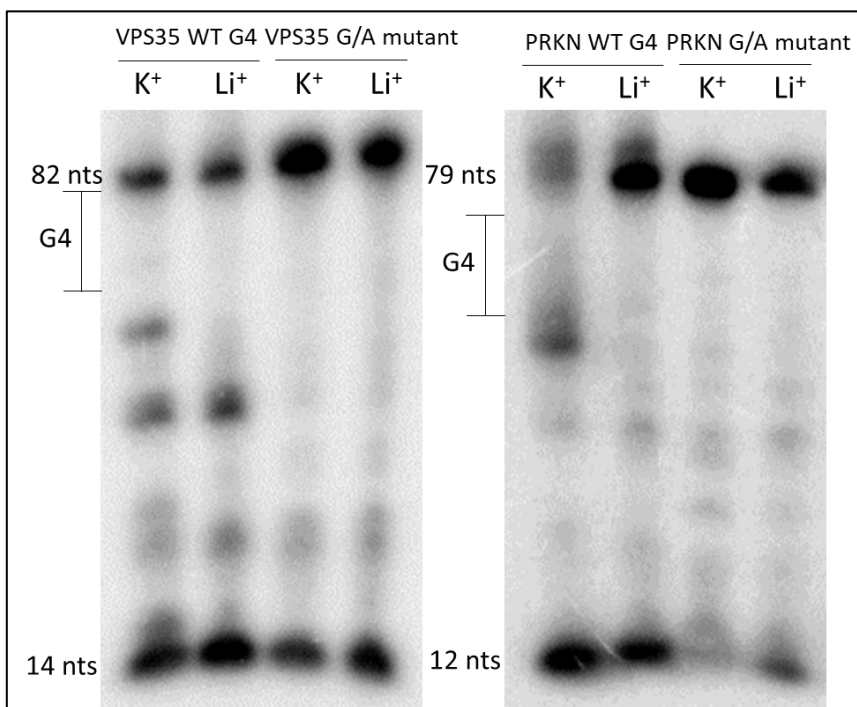


Figure 16: Essais d'arrêt de transcription inverse

Les ARN WT or G/A mutant de VPS35 (gauche) et PRKN (droite) ont été transcrit à l'inverse avec l'enzyme SuperScript III (Introvigen™) en présence de KCl ou LiCl pendant 15 minutes. La position théorique du G4 est annotée G4. La longueur des nucléotides pour la forme courte (l'amorce seule) et dans la forme longue (transcription inverse complète) est identifiée.

Analyse de l'impact de formation des G-quadruplexes *in cellulo*

La formation des G-quadruplexes est un phénomène complexe devant être confirmé à plusieurs niveaux. À la suite de la confirmation des G4 d'ARN *in vitro*, ceux-ci doivent être testés en cellule pour évaluer quel est l'impact sur les mécanismes d'intérêt. Dans le cas présent, l'impact de ces G4 sur la traduction a été évalué. Cela est important puisque ce ne sont pas tous les G4 situés dans le 5'UTR qui vont avoir un impact sur la traduction.

L'analyse *in cellulo* de l'impact des G4 a premièrement été réalisée dans des cellules SH-SY5Y. Ce type de cellule est une lignée dérivée des SK-N-SH qui a été établie à partir d'une biopsie de la moelle osseuse contenant des métastases de neuroblastome d'un enfant de 4 ans (Biedler et al., 1978). Les cellules SH-SY5Y sont souvent utilisées lors des études portant

sur la maladie de Parkinson, principalement due à leurs propriétés neuronales et dopaminergiques (Xicoy et al., 2017). Cependant, au cours de cette étude, le choix des cellules a souvent été remis en question dû à leur difficulté de mise en culture, leur faible résistance aux stress, leur lente vitesse de croissance, leur difficulté de transfection et l'incapacité d'effectuer un *knock-down* avec des siARN. Une seconde lignée, une lignée embryonnaire de rein humain (HEK293T), a été utilisée à titre exploratoire pour évaluer l'effet des G4. En effet, celles-ci ont été utilisées puisqu'elles comportent toutes les caractéristiques pour être facilement adaptées à différents protocoles. Par ailleurs, les HEK293T sont étroitement liés au niveau développemental avec les cellules neuronales, ce qui fait d'eux d'excellents candidats pour de la sorte (Stephanenko & Dmitrenko, 2015).

L'impact des G4 sur l'expression protéique a été étudié avec des essais luciférase. Plus précisément, elles ont été réalisées à l'aide d'un plasmide (PsiCheck 2.0) codant pour deux différentes luciférase, soit la *Renila* et la *Firefly* (voir section « Essais luciférase » dans l'introduction). Ensuite, il a été décidé d'utiliser le 5'UTR complet de chacun des ARN pour bien récapituler les phénomènes se produisant normalement dans l'ARN des gènes d'intérêts. Cependant, il aurait tout de même été intéressant d'aussi cloner les séquences plus courtes utilisées pour les essais de fluorescence et de cartographie *in-line* pour évaluer si des résultats semblables auraient pu être observés. Les quatre ARN d'intérêt possédaient chacun des 5'UTR considérés comme étant assez courts, ce qui a facilité les clonages de ceux-ci.

Les essais luciférase ont bien fonctionné dans les deux lignées cellulaires. En effet, après 24 heures de transfection, il était déjà possible d'observer une variation significative dans l'expression de la luciférase dans les cellules SH-SY5Y et HEK293T (Figure 10 et S2). Des temps de culture plus élevés n'ont alors pas été tentés. Comme présenté dans les résultats, le patron d'expression de la luciférase obtenu dans les deux lignées de culture est assez similaire. Il est en effet possible d'observer un effet inhibiteur assez élevé exercé par le G-quadruplexe de PRKN et de VPS35. Cet effet peut être déduit du fait que lorsque le G-quadruplexe est muté, il y a une augmentation de l'activité totale de la *Renila* luciférase, ce qui se traduit par une augmentation de la quantité de luciférase sous contrôle de l'insert dans le plasmide. Les résultats pour SNCA et LRRK2 montrent quant à eux aucun ou peu d'effet

dans les deux types cellulaires. Par ailleurs, l'effet observé dans les HEK293T pour SNCA est assez similaire à ce qui a été observé auparavant chez un autre groupe de recherche (Koukouraki & Doxakis, 2016). Il est tout de même possible de comparer les résultats obtenus entre les deux types de cellules. L'effet observé dans les HEK293T pour PRKN et VPS35 est en moyenne 1,55 plus fort que celui observé dans les SH-SY5Y. Une des hypothèses est que ce phénomène pourrait être dû à une différence dans le protéome des deux lignées cellulaires, ce qui produit une variation dans les niveaux des protéines étant en mesure de lier les G4.

Par la suite, il a été montré que l'effet observé en luciférase n'était pas dû à un effet transcriptionnel (Figures 10 et S2). En effet, lorsque les différents niveaux d'ARN entre les WT G4 et les G/A mutants sont évalués, il est possible de voir qu'ils sont tous égaux. Cela veut donc dire que le niveau d'ARN ne change pas entre les différentes conditions. Cependant, il est encore trop tôt, avec les expériences effectuées, pour conclure que le phénomène observé est dû seulement à un effet traductionnel. D'autres expériences comme un gradient de polysomes pourraient être effectuées pour contribuer à répondre à cette question.

La fonction d'inhibition du G4 de PRKN et de VPS35 a ensuite été testée avec une seconde technique. Cette fois-ci, des ligands spécifiques en mesure de lier les G4 ont été utilisés dans les SH-SY5Y (Figures 10 et S2). Le premier est une acridine trisubstituée, BRACO19, reconnue comme étant en mesure de lier et de stabiliser les G-quadruplexes d'ADN et d'ARN (Burger et al., 2005; S. Y. Yang et al., 2018). Le second est un bisquinolinium, Phen-DC3, pouvant lui aussi stabiliser les G-quadruplexes d'ADN et d'ARN (Chung et al., 2014). L'utilisation de BRACO19 à 10 μM a montré une diminution de 50 % du niveau endogène de la protéine PRKN tandis qu'aucun effet ne pouvait être observé sur le niveau endogène de VPS35. Cela pourrait être expliqué par le fait que le ligand BRACO19 a beaucoup plus d'affinité pour les G-quadruplexes canoniques, comme celui de PRKN. Ensuite, l'utilisation du Phen-DC3 à 5 μM a diminué le niveau endogène de la protéine PRKN de 90% et de VPS35 de 25% (Figure 10B). Cela montre bien que ces ligands à G4 ont une spécificité envers certains types de G4. Cette expérience confirme bien l'effet observé dans les essais

luciférase, soit que les deux G4 ont un effet inhibiteur sur la traduction. Cependant, il est important de mentionner que l'utilisation de ligands à G4 peut donner des effets qui sont indirects. Il est donc important de toujours confirmer avec d'autres méthodes, comme avec l'essai luciférase.

Analyse des partenaires protéiques de PRKN et de VPS35

Par la suite, les protéines étant en mesure de lier ces deux G4 ont été identifiées par purification par affinité sans marquage (*Label-free affinity purification*; Figure 11A). Notre hypothèse était que certaines protéines étant dérégulées dans la PD sont sous régulation de G-quadruplexe d'ARN dont la formation et la déformation est régulée par la liaison de certaines protéines. La régulation entre la protéine et le G4 pourrait permettre de contrebalancer le déséquilibre de certaines voies de signalisation observées dans la maladie de Parkinson. Pour ce faire, quatre copies d'un aptamère d'ARN étant en mesure de lier la streptavidine ont été ajoutées en aval de chacune des séquences d'ARN (les mêmes que celles utilisées pour les essais *in vitro*) (Leppek & Stoecklin, 2014a). D'ailleurs, ma participation au développement de cette technique dans l'identification des protéines pouvant lier le G4 de BAG-1 m'a permis d'obtenir un article second auteur (Annexe 1). Il a ensuite été important de confirmer que l'ajout du 4xS1m n'avait aucune influence sur la formation des deux G-quadruplexes (Figures S4 et S5). Les ARN ont par la suite été liés sur les billes de streptavidine-Sépharose. Ceux-ci ont ensuite été incubés avec des extraits totaux protéiques de SH-SY5Y. Un des points importants à mentionner est que l'utilisation d'un extrait cellulaire total amène une rupture des compartiments cellulaires. Cela fait donc en sorte qu'il est possible d'identifier une liaison de l'ARN avec certaines protéines qui, en conditions cellulaires, n'auraient pas été en mesure d'interagir avec le G4 puisqu'elles ne se retrouvent pas au même endroit dans la cellule. Les protéines ont ensuite été envoyées en spectrométrie de masse et l'enrichissement de protéines pouvant lier la séquence sauvage a été identifié avec le programme *Significance Analysis of INteractome* (SAINT) (Choi et al., 2011). L'utilisation du programme SAINT permet la comparaison de données de spectrométrie de masse entre deux conditions ayant été réalisées dans des conditions dites *Label-Free*. En effet, celui-ci compare les spectres MS/MS d'un échantillon à un contrôle (le contrôle étant G/A mutant dans ce cas-ci) et donne un score indiquant la probabilité qu'il y ait un

enrichissement significatif (allant de 0 à 1). Pour notre étude, des paramètres stricts ont été utilisés pour augmenter l'identification de vrais partenaires protéiques. Pour ce faire, un score minimal de 0.9 et un ratio d'enrichissement d'au moins 10 fois entre le WT G4 et le G/A mutant ont été utilisés. Ce score est celui recommandé par SAINT et l'enrichissement a été décidé de façon arbitraire. Cette analyse a permis l'identification de 13 protéines étant enrichies pour l'ARN WT de PRKN et 3 protéines pour celui de VPS35 (Figure 10). Selon nous, cette différence au niveau de la quantité de protéines repêchées pour les deux G4 peut s'expliquer par le fait que le G4 de PRKN est canonique contrairement à celui de VPS35. En effet, en ayant un G4 plus canonique, une plus grande quantité de protéines serait en mesure de lier sa structure. Par ailleurs, il est important de mentionner que la méthode utilisée ne permet pas seulement d'identifier les interacteurs directs. En effet, certaines interactions indirectes peuvent se retrouver dans les protéines identifiées.

Ensuite, l'analyse des différentes fonctions reliées à chacune de ces protéines a permis d'identifier 4 groupes de protéines. Tout d'abord, il est possible de voir que 3 protéines sont des protéines pouvant lier l'ARN. Comme l'appât de notre technique est de l'ARN, c'est majoritairement des protéines appartenant à ce groupe qui étaient attendues. Cependant, les protéines appartenant aux RBP ne représentent que 19% des protéines ayant été identifiées. Ce très bas pourcentage pourrait s'expliquer par le fait que les propriétés RBP de certaines protéines n'aient pas encore été étudiées. Un second groupe relié à la liaison du GTP contenant 3 différentes protéines est aussi présent. Une hypothèse est que l'enrichissement en guanine dans l'ARN WT comparativement au G/A mutant pourrait causer cet enrichissement. Cependant, comme les guanines sont ici polymérisées en ARN, les résultats restent assez surprenants puisque l'environnement moléculaire est assez différent du GTP seul (un polymère d'acides nucléiques vs un monomère). Ensuite, 4 protéines reliées à l'épissage ont été identifiées. Ce résultat n'est pas étonnant, bien qu'aucun intron ne se retrouve en périphérie du G4. En effet, certaines d'entre elles possèdent certainement des propriétés RBP pour pouvoir exercer leur fonction sur l'ARN (Van Nostrand et al., 2020). Comme mentionné dans l'introduction, il est déjà connu que plusieurs protéines sont en mesure de lier les G4 pour réguler l'épissage. Dans cette expérience, une lyse cellulaire totale est réalisée. Des protéines participant à l'épissage (principalement situés dans le noyau) ont beaucoup plus de chance d'interagir avec l'ARN pour agir en tant que RBP. Pour finir, 4

protéines ayant une fonction mitochondriale ribosomale ont été identifiées. Une hypothèse est que la rupture des membranes lors de la lyse cellulaire a fait interagir ces protéines avec l'ARN. En effet, ces protéines en condition cellulaire ne devraient pas être en contact avec l'ARN de PRKN.

Une analyse attentive des protéines liées à VPS35 a premièrement permis de détecter YBX1, une protéine ayant été identifiée comme ayant la capacité de lier certains types de G4. En effet, celle-ci lie les G4 se retrouvant dans des tRNA clivés par l'angiogenine pour inhiber la traduction et induire des granules de stress (Ivanov et al., 2014). L'identification de cette protéine montre bien que des RBP à G4 sont en mesure d'être identifiées avec cette technique pour VPS35. La seconde est PRPF19, une protéine étant impliquée dans le complexe d'épissage et ayant des propriétés d'action sur l'épissage (Makarova et al., 2004). Pour finir, il est possible de retrouver une GTPase, Guanine Nucleotide-Binding Protein-like 1 (GNL1), la seule protéine étant commun aux deux ensembles de données. Au niveau fonctionnel, cette protéine favorise la transition G2/M dans le cycle cellulaire grâce à son interaction avec RPS20 (Krishnan et al., 2018). Au niveau structural, aucun domaine de liaison à l'ARN n'est prédit. Selon les prédictions, cette protéine possède seulement un domaine de liaison au GTP (Hunter et al., 2009). Cependant, lorsque les acides aminés présents dans la séquence sont analysés, il est possible de retrouver la présence d'un Di-RG (Thandapani et al., 2013). Ce Di-RG pourrait être responsable de la liaison avec les différents G4. Au niveau pathologique, bien qu'aucun polymorphisme dans cette protéine n'a été associé aux maladies neurodégénératives, certains ont été associés à la maladie de Grave (Y.-H. Liu et al., 2014). Celle-ci est une hyperthyroïdie auto-immune. Ensuite, une analyse plus complexe des résultats permet de voir que cette protéine à un score SAINT parfait de 1. Cela supporte bien que les probabilités qu'il y ait vraiment une interaction entre la protéine et les ARN sont assez élevées. Par ailleurs, celle-ci se trouve à être la protéine la plus enrichie avec le G4 de PRKN (143 fois) et dans les 3 protéines les plus enrichies pour VPS35 (12 fois). Cependant, encore une fois, il est important de garder en tête que bien que l'enrichissement soit très élevé, cela pourrait être dû à des interactions secondaires. GNL1 pourrait donc être une protéine qui est en mesure de lier les G-quadruplexes de VPS35 et de PRKN pour réguler la traduction dans les neurones.

Pour ce que qui est des protéines liées à PRKN, il est encore une fois possible d'en identifier une connue pour lier les G4 d'ARN. En effet, GRSF1 a été montré comme étant en mesure de déplier les structures G4 d'ARN pour entraîner la dégradation de l'ARN par le dégradosome (Pietras et al., 2018). Cette protéine confirme que des RBP à G4 peuvent être identifiées avec cette technique pour PRKN. Ensuite, une autre protéine intéressante ayant été identifiée est FAM120A. Il a déjà été montré que celle-ci est en mesure de lier des régions riches en guanines situées dans le 3'UTR d'ARNm (Kelly et al., 2019). Une diminution de son affinité envers ces séquences serait aussi observée lorsque certaines guanines sont mutées pour des adénines. Il est donc possible de penser que des G4 peuvent être formés dans ces régions et que la protéine FAM120A ait la capacité de les lier. Une troisième protéine pouvant lier l'ARN, RALY, a aussi été identifiée (Tenzer et al., 2013). Cependant, aucune information ne laisse entendre que cette protéine serait en mesure de lier les G4. Une seconde catégorie de protéine reliée cette fois-ci à l'épissage alternatif a été identifiée. SF3B2 et SF3B3, deux protéines impliquées dans le complexe SF3b responsable dans la liaison en liant le 5' et le 3' du site d'épissage, ont été identifiées (Yokoi et al., 2011).

Une analyse moins stricte a aussi été réalisée pour vérifier comment se comportaient les deux ensembles de données (Figure 17). Cette fois-ci, seul un score SAINT de 0.8 a été utilisé. Cela a permis d'identifier 3 nouvelles protéines pour VPS35 et 12 nouvelles pour PRKN (portant le total à 6 et 25). Parmi celles-ci, il est possible d'identifier de nouvelles protéines pouvant lier l'ARN. La première, SSB, est une RBP possédant un domaine de liaison à l'ARN (RRM) reconnaissant la portion 5' des ARN. Par ailleurs, il a été montré que cette protéine chaperon a des propriétés hélicases pouvant défaire des structures tige-boucle (Sommer et al., 2020). La seconde est UPF1, ayant des fonctions hélicases menant à la dégradation de l'ARN (Singh et al., 2019). Il est donc possible que ces deux protéines soient aussi en mesure de déplier des structures G4. Deux autres RBP, HNPNPC et TROVE2, ont par ailleurs été identifiées (Fischl et al., 2019; Quinones-Valdez et al., 2019). Celles-ci ne semblent cependant pas avoir de lien avec les structures secondaires. Étrangement, la technique a aussi permis l'identification de plusieurs protéines ribosomales (RPL 5, 17, 19, 29, 36, 36A, 37A et 38) en plus des protéines ribosomales associées à la mitochondrie précédemment identifiées (MRPL 13, 19 et 45). Une hypothèse est que des ribosomes se retrouvant dans l'extrait cellulaire aient été en mesure de reconnaître l'ARN et de commencer

à faire une numérisation de l'ARN. Comme le G4 de PRKN se retrouve au début du 5'UTR, celui-ci aurait été victime d'un arrêt structural par le G4. Pour finir, une troisième protéine, SF3B1, appartenant au complexe Sf3b a aussi été identifiée (Yokoi et al., 2011). Cela laisse croire que ce complexe d'épissage pourrait avoir un rôle à jouer avec l'interaction du G4 (directe ou indirecte).

<p style="text-align: center;">Protéines liant l'ARN</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Stricte</th> <th style="text-align: left;">Score 0.8</th> </tr> </thead> <tbody> <tr> <td>FAM120A(1,00; 50)</td> <td>UPF1 (0,99; 7)</td> </tr> <tr> <td>GRSF1 (0,99; 33)</td> <td>HNRNPC (0,98; 7)</td> </tr> <tr> <td>RALY (0,95; 23)</td> <td>TROVE2 (0,86; 13)</td> </tr> </tbody> </table>		Stricte	Score 0.8	FAM120A(1,00; 50)	UPF1 (0,99; 7)	GRSF1 (0,99; 33)	HNRNPC (0,98; 7)	RALY (0,95; 23)	TROVE2 (0,86; 13)	<p style="text-align: center;">Protéines liant le GTP</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Stricte</th> <th style="text-align: left;">Score 0.8</th> </tr> </thead> <tbody> <tr> <td>GNL1 (1,00; 143)</td> <td>Aucune</td> </tr> <tr> <td>LSG1 (1,00; 15)</td> <td></td> </tr> <tr> <td>GNL1 (1,00; 12)</td> <td></td> </tr> <tr> <td>GTPBP4 (0,91; 10)</td> <td></td> </tr> </tbody> </table>		Stricte	Score 0.8	GNL1 (1,00; 143)	Aucune	LSG1 (1,00; 15)		GNL1 (1,00; 12)		GTPBP4 (0,91; 10)					
Stricte	Score 0.8																								
FAM120A(1,00; 50)	UPF1 (0,99; 7)																								
GRSF1 (0,99; 33)	HNRNPC (0,98; 7)																								
RALY (0,95; 23)	TROVE2 (0,86; 13)																								
Stricte	Score 0.8																								
GNL1 (1,00; 143)	Aucune																								
LSG1 (1,00; 15)																									
GNL1 (1,00; 12)																									
GTPBP4 (0,91; 10)																									
<p style="text-align: center;">Protéines liant l'ADN</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Stricte</th> <th style="text-align: left;">Score 0.8</th> </tr> </thead> <tbody> <tr> <td>NKRF (0,97; 27)</td> <td>SSB (0,84; 12)</td> </tr> </tbody> </table>		Stricte	Score 0.8	NKRF (0,97; 27)	SSB (0,84; 12)	<p style="text-align: center;">Protéines ribosomales</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Stricte</th> <th style="text-align: left;">Score 0.8</th> </tr> </thead> <tbody> <tr> <td>Aucune</td> <td>RPL5 (1,00; 5)</td> </tr> <tr> <td></td> <td>RPL19 (1,00; 5)</td> </tr> <tr> <td></td> <td>RPL38 (0,99; 5)</td> </tr> <tr> <td></td> <td>RPL17 (0,90; 4)</td> </tr> <tr> <td></td> <td>RPL36 (0,90; 4)</td> </tr> <tr> <td></td> <td>RPL37A (0,87; 7)</td> </tr> <tr> <td></td> <td>RPL29 (0,82; 4)</td> </tr> <tr> <td></td> <td>RPL36A (0,81; 8)</td> </tr> </tbody> </table>		Stricte	Score 0.8	Aucune	RPL5 (1,00; 5)		RPL19 (1,00; 5)		RPL38 (0,99; 5)		RPL17 (0,90; 4)		RPL36 (0,90; 4)		RPL37A (0,87; 7)		RPL29 (0,82; 4)		RPL36A (0,81; 8)
Stricte	Score 0.8																								
NKRF (0,97; 27)	SSB (0,84; 12)																								
Stricte	Score 0.8																								
Aucune	RPL5 (1,00; 5)																								
	RPL19 (1,00; 5)																								
	RPL38 (0,99; 5)																								
	RPL17 (0,90; 4)																								
	RPL36 (0,90; 4)																								
	RPL37A (0,87; 7)																								
	RPL29 (0,82; 4)																								
	RPL36A (0,81; 8)																								
<p style="text-align: center;">Protéines participant à l'épissage</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Stricte</th> <th style="text-align: left;">Score 0.8</th> </tr> </thead> <tbody> <tr> <td>SF3B2 (1,00; 63)</td> <td>SF3B1 (0,86; 17)</td> </tr> <tr> <td>PRPF19 (0,98; 40)</td> <td></td> </tr> <tr> <td>SF3B3 (0,93; 12)</td> <td></td> </tr> <tr> <td>YBX1 (0,92; 11)</td> <td></td> </tr> </tbody> </table>		Stricte	Score 0.8	SF3B2 (1,00; 63)	SF3B1 (0,86; 17)	PRPF19 (0,98; 40)		SF3B3 (0,93; 12)		YBX1 (0,92; 11)															
Stricte	Score 0.8																								
SF3B2 (1,00; 63)	SF3B1 (0,86; 17)																								
PRPF19 (0,98; 40)																									
SF3B3 (0,93; 12)																									
YBX1 (0,92; 11)																									
<p style="text-align: center;">Autres protéines</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Stricte</th> <th style="text-align: left;">Score 0.8</th> </tr> </thead> <tbody> <tr> <td>ZNF622 (1,00; 23)</td> <td>TRIM28 (0,86; 7)</td> </tr> <tr> <td></td> <td>CARM1 (0,84; 6)</td> </tr> </tbody> </table>		Stricte	Score 0.8	ZNF622 (1,00; 23)	TRIM28 (0,86; 7)		CARM1 (0,84; 6)	<p style="text-align: center;">Protéines ribosomales mitochondriales</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Stricte</th> <th style="text-align: left;">Score 0.8</th> </tr> </thead> <tbody> <tr> <td>MRPL13 (1,00; 60)</td> <td>Aucune</td> </tr> <tr> <td>MRPL10 (0,96; 10)</td> <td></td> </tr> <tr> <td>MRPL45 (0,95; 23)</td> <td></td> </tr> </tbody> </table>		Stricte	Score 0.8	MRPL13 (1,00; 60)	Aucune	MRPL10 (0,96; 10)		MRPL45 (0,95; 23)									
Stricte	Score 0.8																								
ZNF622 (1,00; 23)	TRIM28 (0,86; 7)																								
	CARM1 (0,84; 6)																								
Stricte	Score 0.8																								
MRPL13 (1,00; 60)	Aucune																								
MRPL10 (0,96; 10)																									
MRPL45 (0,95; 23)																									

Figure 17 : Comparaison entre l'analyse protéique SAINT stricte et moins stricte
 Comparaison entre l'analyse stricte (score SAINT de 0,90 et plus; score d'enrichissement de 10 et plus) et l'analyse moins stricte (score SAINT de 0,80 et plus seulement). Le premier chiffre entre les parenthèses identifie le score SAINT tandis que le deuxième indique l'enrichissement (WT / G/A mutant). Les protéines identifiées en vert sont celles ayant été retrouvées dans la purification par affinité de VPS35, tandis que ceux de PRKN sont en noir. Catégories des autres protéines : ZNF622 (facteur de transcription), TRIM28 (corepresseur nucléaire) et CARM1 (arginine méthyle transférase).

Étonnamment, aucune protéine n'ayant été annotée comme étant une hélicase n'a été identifiée dans le pull-down avec un score SAINT de 0.9, et ce pour les deux ensembles de données. Une hypothèse pour expliquer cela est que la cinétique associée à la liaison des

hélicases est trop rapide pour être détectée par ce type de purification par affinité, ce qui se traduit par un score SAINT plus bas. Pour répondre à cette hypothèse, le score SAINT a été abaissé à 0.8. Il est alors possible d'identifier deux nouvelles hélicases, ce qui laisse croire que notre hypothèse est bonne.

Analyse de l'interaction entre GNL1 et le G4 de VPS35 et PRKN

L'analyse de l'interaction entre GNL1 et le G4 de VPS35 et de PRKN a permis de montrer que cette protéine possède une plus grande affinité pour les séquences contenant des G4 que pour leur séquence mutée (Figure 12). Étonnamment, GNL1 est en mesure de lier les deux G-quadruplexes à de basses constantes d'affinités assez semblables (152 nM vs 198 nM pour VPS35 et PRKN, respectivement). Il aurait été attendu que la différence entre les deux valeurs soit plus élevée puisque deux types de G4 (canonique et non canonique) sont ici comparés. Il est donc possible de penser que GNL1 est en mesure de lier des G4 avec des conformations différentes. Cela pourrait facilement être confirmé en effectuant des essais de retardement sur gel avec d'autres candidats. Comme mentionné plus tôt, le domaine probable de liaison entre GNL1 et les G4 est la présence d'un RG dans sa séquence. Il serait alors intéressant de produire un mutant non fonctionnel pour ensuite vérifier les différences que cela causerait en essais de retardement ou encore en cellule.

Un autre point important est la différence d'affinité de GNL1 pour les deux séquences WT et G/A mutants (152 nM vs 360 nM, pour PRKN et 194 nM versus 7339 nM pour VPS35). Clairement, les mutations induites dans PRKN ont un impact largement supérieur sur l'interaction. Une hypothèse est que la mutation d'une structure canonique en structure linéaire amène beaucoup plus de différence au niveau conformationnel qu'une mutation d'une structure non canonique (souvent moins stable) en structure linéaire. Ce raisonnement amène encore plus d'importance au fait de bien distinguer entre les G-quadruplexes de type canonique et non canonique dans des systèmes complexes comme dans les neurones.

La technique utilisée dans cette étude pour évaluer l'affinité de GNL1 envers les G4 est relative plutôt qu'absolue. En effet, bien que nous ayons été en mesure de montrer que la GST et les différentes impuretés ne contribuent pas au retardement sur gel, ceux-ci ont quand même un impact au niveau de la concentration obtenue après la purification (données non

montrées). Il est donc possible de s'attendre à des constantes d'affinité qui sont légèrement plus basse que ce qui a été obtenu dans cette étude. Cependant, dans le cadre de l'étude, l'élément important était de noter une différence entre la forme WT et G/A mutant et non d'obtenir une valeur absolue pour comparer avec d'autres protéines ou avec d'autres études.

Finalement, l'interaction entre le G4 et l'ARN ne semble pas engendrer la formation de duplex ou d'autre structure, comme nous avons déjà observé au laboratoire avec SNRPA, une protéine ayant une affinité pour les G4 de BAG-1 (Annexe 1). Un duplex en essais de retardement sur gel est normalement expliqué par une dimérisation des protéines, mais pourrait aussi être dû au fait que la protéine se lie de façon non spécifique sur une séquence de l'ARN, ce qui diminuerait sa spécificité envers les G4. Un point à noter est que GNL1 n'a pas une constante d'affinité aussi élevée que SNRPA (K_d en nM vs μ M). Des duplex pourraient peut-être survenir à de plus haute concentration en protéine. Malheureusement, notre quantité totale de protéines obtenue après purification ne nous permettait pas d'effectuer des essais à ces concentrations.

Méthodes alternatives

Une alternative à la technique d'identification des protéines pouvant lier les G4 d'ARN utilisée dans cette étude aurait été la réalisation de la purification par affinité *in cellulo*. En effet, il aurait été possible de fabriquer un plasmide codant pour notre ARN d'intérêt suivi d'une séquence étant reconnu par la protéine MS2 (Yoon & Gorospe, 2016). Il aurait alors été possible d'effectuer une fixation entre les protéines et l'ARN *in cellulo*. Cela aurait permis de seulement identifier les protéines pouvant interagir avec le G-quadruplexe en condition cellulaire. Cependant, il aurait probablement aussi fallu utiliser une grande quantité de cellules puisque le nombre de protéines repêchées avec cette technique aurait sans doute été moindre. Par ailleurs, le fait que l'ARN soit surexprimé pourrait aussi avoir un impact sur les protéines étant en mesure de lier les G4 puisque ceux-ci seraient présentes en grande quantité dans la cellule. Le phénomène d'identification de protéines avec interaction indirecte serait cependant toujours présent.

Une seconde alternative aurait été d'effectuer une comparaison des protéines pouvant lier la séquence sauvage avec une autre séquence similaire où les guanines aurait été changées pour des 7-déaza-guanine. Les déaza-guanine sont des guanines possédant un carbone en position 7, plutôt qu'un azote. Cela permet d'abolir la formation du G4 sans venir changer la séquence nucléotidique. En effet, il serait alors possible d'être certain que les protéines n'aient pas été recrutées en raison d'une différence au niveau de la séquence. Ce type de purification par affinité a déjà été rapporté pour l'étude des G4 (Herviou et al., 2020a). Cependant, le point négatif de cette technique est que la perte du groupement azote pourrait aussi avoir un impact sur la liaison de certaines protéines dû à la perte d'une liaison hydrogène. Il serait donc intéressant de comparer les protéines retrouvées dans la purification par affinité avec celles présentes dans la purification G/A mutant réalisée dans ce mémoire.

Perspectives au projet

Une première avenue au projet serait d'effectuer un *knock-down* ou *knock-out* de la protéine GNL1 en utilisant le système CRISPR-Cas9, des siARN ou bien des shARN. Cela permettrait d'observer si une diminution du niveau de la protéine a un impact sur les niveaux protéiques de VPS35 et de PRKN. Des tentatives de knock-down avec des siARN et des shARN ont été effectuées lors de ma maîtrise, mais nous n'avons pas été en mesure d'obtenir une diminution du niveau de GNL1 dans les SH-SY5Y. De l'optimisation est donc nécessaire. Par exemple, il serait possible d'utiliser un nouveau système d'infection pour produire des shARN ou bien d'acheter de nouveaux siARN ayant été confirmés comme fonctionnant dans les cellules SH-SY5Y.

Deuxièmement, il serait intéressant d'effectuer des mutations dans les boucles de chacun des G4 pour évaluer l'impact sur la liaison protéique. En s'assurant que ces mutations n'abolissent pas la formation du G4, cela pourrait permettre d'identifier les résidus essentiels à l'interaction protéine-G4. L'impact pourrait alors facilement être évalué en effectuant des gels de retardement et en déterminant les nouvelles constantes d'affinité. Cela aiderait à mieux comprendre l'interaction entre les deux macromolécules.

Troisièmement, l'affinité de GNL1 pourrait être déterminée sur d'autre type de structure G4. Cela permettrait de déterminer si certains types de G4 sont plus propices à être régulés par cette protéine. Par exemple, la liaison de GNL1 au G4 de LRRK2 et de SNCA pourrait pointer vers une régulation fonctionnelle des ARN de gènes étant dérégulés dans la maladie de Parkinson.

Finalement, s'il est montré que GNL1 a un impact sur la traduction, il serait possible d'utiliser l'interaction protéine-G4 pour tenter de trouver des molécules organiques, des anticorps ou des ARN antisens pour induire un changement dans l'expression protéique de VPS35 et de PRKN. En effet, il est connu qu'il existe un grand manque de spécificité dans le ciblage des G4 par des molécules organiques dû à leurs structures semblables. En utilisant des molécules pouvant venir affecter la liaison de la protéine au G4, cela permettrait une approche beaucoup plus spécifique.

CONCLUSION

Cette étude constitue la première analyse de G-quadruplexes se retrouvant dans les ARNm des gènes étant dérégulés dans la maladie de Parkinson. Tout d'abord, cette étude a démontré qu'il existait un enrichissement en G-quadruplexes potentiels dans les différents transcrits étant dérégulés dans les maladies neurodégénératives. Ensuite, quatre nouveaux G-quadruplexes se retrouvant dans le 5'UTR de VPS35, LRRK2, SNCA et PRKN ont été découverts par des essais de fluorescence et cartographie *in-line*. La formation de deux d'entre eux (VPS35 et PRKN) ont montré, par essais luciférases et par des ligands spécifiques, une répression de la traduction de leurs protéines. Finalement, la protéine GNL1 a été identifiée, par des essais de précipitation par affinité et retardement sur gel, comme étant une nouvelle protéine pouvant lier ces G-quadruplexes.

Les résultats obtenus dans ce mémoire amènent une grande quantité de résultats préalables pour de nombreux projets. D'un côté fondamental, la caractérisation de l'interaction entre GNL1 et les G-quadruplexes pourrait amener la découverte de nouveaux mécanismes de liaisons et de régulation conformationnelle. D'un côté pathologique, il serait intéressant de moduler les niveaux de GNL1 dans des modèles simulant la maladie de Parkinson pour comprendre si cette protéine pourrait être une cible lors des dérégulations de VPS35 et de PRKN.

LISTE DES RÉFÉRENCES

- Adrian, M., Heddi, B., & Phan, A. T. (2012). NMR spectroscopy of G-quadruplexes. *Quadruplex DNA Methods*, 57(1), 11-24.
- Agarwala, P., Pal, G., Pandey, S., & Maiti, S. (2017). Mutagenesis Reveals an Unusual Combination of Guanines in RNA G-Quadruplex Formation. *ACS Omega*, 2(8), 4790-4799.
- Agarwala, P., Pandey, S., & Maiti, S. (2015). The tale of RNA G-quadruplex. *Organic & Biomolecular Chemistry*, 13(20), 5570-5585.
- Agarwala, P., Pandey, S., Mapa, K., & Maiti, S. (2013). The G-Quadruplex Augments Translation in the 5' Untranslated Region of Transforming Growth Factor β 2. *Biochemistry*, 52(9), 1528-1538.
- Al-Wandi, A., Ninkina, N., Millership, S., Williamson, S. J. M., Jones, P. A., & Buchman, V. L. (2010). Absence of α -synuclein affects dopamine metabolism and synaptic markers in the striatum of aging mice. *Neurobiology of Aging*, 31(5-19), 796-804.
- Amato, J., Madanayake, T. W., Iaccarino, N., Novellino, E., Randazzo, A., Hurley, L. H., & Pagano, B. (2018). HMGB1 binds to the KRAS promoter G-quadruplex: A new player in oncogene transcriptional regulation? *Chemical Communications*, 54(68), 9442-9445.
- Anglade, P., Vyas, S., Javoy-Agid, F., Herrero, M. T., Michel, P. P., Marquez, J., Mouatt-Prigent, A., Ruberg, M., Hirsch, E. C., & Agid, Y. (1997). Apoptosis and autophagy in nigral neurons of patients with Parkinson's disease. *Histology and Histopathology*, 12(1), 25-31.
- Antonellis, A., Ellsworth, R. E., Sambuughin, N., Puls, I., Abel, A., Lee-Lin, S.-Q., Jordanova, A., Kremensky, I., Christodoulou, K., Middleton, L. T., Sivakumar, K., Ionasescu, V., Funalot, B., Vance, J. M., Goldfarb, L. G., Fischbeck, K. H., & Green, E. D. (2003). Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V. *American Journal of Human Genetics*, 72(5), 1293-1299.
- Armas, P., Nasif, S., & Calcaterra, N. B. (2008). Cellular nucleic acid binding protein binds G-rich single-stranded nucleic acids and may function as a nucleic acid chaperone. *Journal of Cellular Biochemistry*, 103(3), 1013-1036.
- Aufschnaiter, A., Kohler, V., Walter, C., Tosal-Castano, S., Habernig, L., Wolinski, H., Keller, W., Vögtle, F.-N., & Büttner, S. (2018). The Enzymatic Core of the Parkinson's Disease-Associated Protein LRRK2 Impairs Mitochondrial Biogenesis in Aging Yeast. *Frontiers in Molecular Neuroscience*, 11(205).
- Barnes, J. D., Arhel, N. J., Lee, S. S., Sharp, A., Al-Okail, M., Packham, G., Hague, A., Paraskeva, C., & Williams, A. C. (2005). Nuclear BAG-1 expression inhibits apoptosis in colorectal adenoma-derived epithelial cells. *Apoptosis*, 10(2), 301-311.
- Beaudoin, J.-D., Jodoin, R., & Perreault, J.-P. (2013). In-line probing of RNA G-quadruplexes. *Methods*, 64(1), 79-87.
- Beaudoin, J.-D., Jodoin, R., & Perreault, J.-P. (2014). New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Research*, 42(2), 1209-1223.
- Beaudoin, J.-D., & Perreault, J.-P. (2010). 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Research*, 38(20), 7022-7036.

- Beaudoin, J.-D., & Perreault, J.-P. (2013). Exploring mRNA 3'-UTR G-quadruplexes : Evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Research*, *41*(11), 5898-5911.
- Bedrat, A., Lacroix, L., & Mergny, J.-L. (2016). Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Research*, *44*(4), 1746-1759.
- Benhalevy, D., Gupta, S. K., Danan, C. H., Ghosal, S., Sun, H.-W., Kazemier, H. G., Paeschke, K., Hafner, M., & Juranek, S. A. (2017). The Human CCHC-type Zinc Finger Nucleic Acid-Binding Protein Binds G-Rich Elements in Target mRNA Coding Sequences and Promotes Translation. *Cell Reports*, *18*(12), 2979-2990.
- Biedler, J., Roffler-Tarlov, S., Schachner, M., & Freedman, L. (1978). Multiple Neurotransmitter Synthesis by Human Neuroblastoma Cell Lines and Clones. *Cancer Research*, *38*(11), 3751-3757.
- Biffi, G., Di Antonio, M., Tannahill, D., & Balasubramanian, S. (2014). Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nature Chemistry*, *6*(1), 75-80.
- Biffi, G., Tannahill, D., & Balasubramanian, S. (2012). An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *Journal of the American Chemical Society*, *134*(29), 11974-11976.
- Biffi, G., Tannahill, D., McCafferty, J., & Balasubramanian, S. (2013). Quantitative Visualization of DNA G-quadruplex Structures in Human Cells. *Nature Chemistry*, *5*(3), 182-186.
- Bishop, G. R., & Chaires, J. B. (2002). Characterization of DNA Structures by Circular Dichroism. *Current Protocols in Nucleic Acid Chemistry*, *11*(1), 7.11.1-7.11.8.
- Blackburn, E. H. (2000). Telomere states and cell fates. *Nature*, *408*(6808), 53-56.
- Bolduc, F., Garant, J.-M., Allard, F., & Perreault, J.-P. (2016). Irregular G-quadruplexes Found in the Untranslated Regions of Human mRNAs Influence Translation. *Journal of Biological Chemistry*, *291*(41), 21751-21760.
- Bonnal, S., Schaeffer, C., Créancier, L., Clamens, S., Moine, H., Prats, A.-C., & Vagner, S. (2003). A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *The Journal of Biological Chemistry*, *278*(41), 39330-39336.
- Brázda, V., Hároníková, L., Liao, J. C. C., & Fojta, M. (2014). DNA and RNA Quadruplex-Binding Proteins. *International Journal of Molecular Sciences*, *15*(10), 17493-17517.
- Buratti, E., Dörk, T., Zuccato, E., Pagani, F., Romano, M., & Baralle, F. E. (2001). Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *The EMBO Journal*, *20*(7), 1774-1784.
- Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., & Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research*, *34*(19), 5402-5415.
- Burger, A. M., Dai, F., Schultes, C. M., Reszka, A. P., Moore, M. J., Double, J. A., & Neidle, S. (2005). The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Research*, *65*(4), 1489-1496.
- Butovskaya, E., Soldà, P., Scalabrin, M., Nadai, M., & Richter, S. N. (2019). The HIV-1 nucleocapsid protein unfolds stable RNA G-quadruplexes in the viral genome and is inhibited by G-quadruplex ligands. *ACS Infectious Diseases*, *5*(12), 2127-2135.
- Cammas, A., & Millevoi, S. (2017). RNA G-quadruplexes : Emerging mechanisms in disease. *Nucleic Acids Research*, *45*(4), 1584-1595.

- Cevec, M., & Plavec, J. (2005). Role of Loop Residues and Cations on the Formation and Stability of Dimeric DNA G-Quadruplexes. *Biochemistry*, *44*(46), 15238-15246.
- Chakraborty, P., & Grosse, F. (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair*, *10*(6), 654-665.
- Chen, Tippana, R., Demeshkina, N. A., Murat, P., Balasubramanian, S., Myong, S., & Ferré-D'Amaré, A. R. (2018). Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*, *558*(7710), 465-469.
- Chen, Y., & Yang, D. (2012). Sequence, Stability, Structure of G-Quadruplexes and Their Drug Interactions. *Current protocols in nucleic acid chemistry*, CHAPTER, Unit17.5.
- Chinta, S. J., & Andersen, J. K. (2005). Dopaminergic neurons. *The International Journal of Biochemistry & Cell Biology*, *37*(5), 942-946.
- Choi, J.-Y., Jang, E.-H., Park, C.-S., & Kang, J.-H. (2005). Enhanced susceptibility to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine neurotoxicity in high-fat diet-induced obesity. *Free Radical Biology and Medicine*, *38*(6), 806-816.
- Choi, Larsen, B., Lin, Z.-Y., Breikreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z. S., Tyers, M., Gingras, A.-C., & Nesvizhskii, A. I. (2011). SAINT : Probabilistic Scoring of Affinity Purification - Mass Spectrometry Data. *Nature Methods*, *8*(1), 70-73.
- Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M.-P., & Phan, A. T. (2014). Solution Structure of a G-quadruplex Bound to the Bisquinolinium Compound Phen-DC3. *Angewandte Chemie International Edition*, *53*(4), 999-1002.
- Clark, G. R., Pytel, P. D., & Squire, C. J. (2012). The high-resolution crystal structure of a parallel intermolecular DNA G-4 quadruplex/drug complex employing syn glycosyl linkages. *Nucleic Acids Research*, *40*(12), 5731-5738.
- Cogoi, S., Paramasivam, M., Spolaore, B., & Xodo, L. E. (2008). Structural polymorphism within a regulatory element of the human KRAS promoter : Formation of G4-DNA recognized by nuclear proteins. *Nucleic Acids Research*, *36*(11), 3765-3780.
- Cohen, G. (2000). Oxidative Stress, Mitochondrial Respiration, and Parkinson's Disease. *Annals of the New York Academy of Sciences*, *899*(1), 112-120.
- Coldwell, M. J., deSchoolmeester, M. L., Fraser, G. A., Pickering, B. M., Packham, G., & Willis, A. E. (2001). The p36 isoform of BAG-1 is translated by internal ribosome entry following heat shock. *Oncogene*, *20*(30), 4095-4100.
- Collie, G. W., Haider, S. M., Neidle, S., & Parkinson, G. N. (2010). A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Research*, *38*(16), 5569-5580.
- Conlon, E. G., Lu, L., Sharma, A., Yamazaki, T., Tang, T., Shneider, N. A., & Manley, J. L. (2016). The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *ELife*, *5*, e17820.
- Cox, J., Hein, M. Y., Lubner, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & Cellular Proteomics: MCP*, *13*(9), 2513-2526.
- Dabral, P., Babu, J., Zareie, A., & Verma, S. C. (2019). LANA and hnRNP A1 regulate the translation of LANA mRNA through G-quadruplexes. *Journal of Virology*, *7*(13), 16793-16805.
- Dai, J., Carver, M., Punchedewa, C., Jones, R. A., & Yang, D. (2007). Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K⁺ solution :

- Insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Research*, 35(15), 4927-4940.
- Darnell, J. C., Van Driesche, S. J., Zhang, C., Hung, K. Y. S., Mele, A., Fraser, C. E., Stone, E. F., Chen, C., Fak, J. J., Chi, S. W., Licatalosi, D. D., Richter, J. D., & Darnell, R. B. (2011). FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell*, 146(2), 247-261.
- Dauer, W., & Przedborski, S. (2003). Parkinson's Disease: Mechanisms and Models. *Neuron*, 39(6), 889-909.
- de Wet, J. R., Wood, K. V., Helinski, D. R., & DeLuca, M. (1985). Cloning of firefly luciferase cDNA and the expression of active luciferase in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 82(23), 7870.
- Del Villar-Guerra, R., Trent, J. O., & Chaires, J. B. (2018). G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angewandte Chemie (International Ed. in English)*, 57(24), 7171-7175.
- Dias, V., Junn, E., & Mouradian, M. M. (2013). The role of oxidative stress in Parkinson's disease. *Journal of Parkinson's Disease*, 3(4), 461-491.
- Dou, N., Yang, D., Yu, S., Wu, B., Gao, Y., & Li, Y. (2018). SNRPA enhances tumour cell growth in gastric cancer through modulating NGF expression. *Cell Proliferation*, 51(5), e12484.
- Dugger, B. N., & Dickson, D. W. (2017). Pathology of Neurodegenerative Diseases. *Cold Spring Harbor Perspectives in Biology*, 9(7).
- Ellison, D., Love, S., Chimelli, L., Harding, B. N., Lowe, J. S., Vinters, H. V., Brandner, S., & Yong, W. H. (2013). Parkinsonism and akinetic-rigid disorders. Dans *Neuropathology* (3^e éd., p. 567-585). <https://www.clinicalkey.com/#!/content/book/3-s2.0-B9780723435150000283>
- Endoh, T., Kawasaki, Y., & Sugimoto, N. (2013a). Suppression of Gene Expression by G-Quadruplexes in Open Reading Frames Depends on G-Quadruplex Stability. *Angewandte Chemie International Edition*, 52(21), 5522-5526.
- Endoh, T., Kawasaki, Y., & Sugimoto, N. (2013b). Stability of RNA quadruplex in open reading frame determines proteolysis of human estrogen receptor α . *Nucleic Acids Research*, 41(12), 6222-6231.
- Endoh, T., & Sugimoto, N. (2013). Unusual -1 Ribosomal Frameshift Caused by Stable RNA G-Quadruplex in Open Reading Frame. *Analytical Chemistry*, 85(23), 11435-11439.
- Eun, H.-M. (1996). Nucleases. Dans H.-M. Eun (Éd.), *Enzymology Primer for Recombinant DNA Technology* (p. 145-232).
- Faghihi, M. A., Modarresi, F., Khalil, A. M., Wood, D. E., Sahagan, B. G., Morgan, T. E., Finch, C. E., St Laurent, G., 3rd, Kenny, P. J., & Wahlestedt, C. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nature Medicine*, 14(7), 723-730.
- Fay, M. M., Lyons, S. M., & Ivanov, P. (2017a). RNA G-quadruplexes in biology: Principles and molecular mechanisms. *Journal of molecular biology*, 429(14), 2127-2147.
- Fay, M. M., Lyons, S. M., & Ivanov, P. (2017b). RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *Journal of Molecular Biology*, 429(14), 2127-2147.

- Fischl, H., Neve, J., Wang, Z., Patel, R., Louey, A., Tian, B., & Furger, A. (2019). HnRNPC regulates cancer-specific alternative cleavage and polyadenylation profiles. *Nucleic Acids Research*, *47*(14), 7580-7591.
- Friedman, S. J., & Terentis, A. C. (2016). Analysis of G-quadruplex conformations using Raman and polarized Raman spectroscopy. *Journal of Raman Spectroscopy*, *47*(3), 259-268.
- Gabelica, V., Maeda, R., Fujimoto, T., Yaku, H., Murashima, T., Sugimoto, N., & Miyoshi, D. (2013). Multiple and Cooperative Binding of Fluorescence Light-up Probe Thioflavin T with Human Telomere DNA G-Quadruplex. *Biochemistry*, *52*(33), 5620-5628.
- Gao, J., Byrd, A. K., Zybailov, B. L., Marecki, J. C., Guderyon, M. J., Edwards, A. D., Chib, S., West, K. L., Waldrip, Z. J., Mackintosh, S. G., Gao, Z., Putnam, A. A., Jankowsky, E., & Raney, K. D. (2019). DEAD-box RNA helicases Dbp2, Ded1 and Mss116 bind to G-quadruplex nucleic acids and destabilize G-quadruplex RNA. *Chemical Communications*, *55*, 4467-4470.
- Garant, J.-M., Perreault, J.-P., & Scott, M. S. (2017). Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. *Bioinformatics*, *33*(22), 3532-3537.
- Garant, J.-M., Perreault, J.-P., & Scott, M. S. (2018). G4RNA screener web server : User focused interface for RNA G-quadruplex prediction. *Biochimie*, *151*, 115-118.
- Ghosal, G., & Muniyappa, K. (2006). Hoogsteen base-pairing revisited : Resolving a role in normal biological processes and human diseases. *Biochemical and Biophysical Research Communications*, *343*(1), 1-7.
- Giasson, B. I., Duda, J. E., Murray, I. V. J., Chen, Q., Souza, J. M., Hurtig, H. I., Ischiropoulos, H., Trojanowski, J. Q., & Lee, V. M.-Y. (2000). Oxidative Damage Linked to Neurodegeneration by Selective α -Synuclein Nitration in Synucleinopathy Lesions. *Science*, *290*(5493), 985-989.
- Giovannone, B., Lee, E., Laviola, L., Giorgino, F., Cleveland, K. A., & Smith, R. J. (2003). Two Novel Proteins That Are Linked to Insulin-like Growth Factor (IGF-I) Receptors by the Grb10 Adapter and Modulate IGF-I Signaling. *Journal of Biological Chemistry*, *278*(34), 31564-31573.
- Giri, B., Smaldino, P. J., Thys, R. G., Creacy, S. D., Routh, E. D., Hantgan, R. R., Lattmann, S., Nagamine, Y., Akman, S. A., & Vaughn, J. P. (2011). G4 resolvase 1 tightly binds and unwinds unimolecular G4-DNA. *Nucleic Acids Research*, *39*(16), 7161-7178.
- Gladkova, C., Maslen, S. L., Skehel, J. M., & Komander, D. (2018). Mechanism of parkin activation by PINK1. *Nature*, *559*(7714), 410.
- Gong, C., & Maquat, L. E. (2011). LncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature*, *470*(7333), 284-288.
- González-Polo, R.-A., Boya, P., Pauleau, A.-L., Jalil, A., Larochette, N., Souquère, S., Eskelinen, E.-L., Pierron, G., Saftig, P., & Kroemer, G. (2005). The apoptosis/autophagy paradox : Autophagic vacuolization before apoptotic death. *Journal of Cell Science*, *118*(14), 3091-3102.
- Götz, S., Pandey, S., Bartsch, S., Juranek, S., & Paeschke, K. (2019). A Novel G-Quadruplex Binding Protein in Yeast-Slx9. *Molecules (Basel, Switzerland)*, *24*(9), 1774.
- Greenamyre, J. T., Sherer, T. B., Betarbet, R., & Panov, A. V. (2001). Complex I and Parkinson's Disease. *IUBMB Life*, *52*(3-5), 135-141.

- Gueddouda, N. M., Mendoza, O., Gomez, D., Bourdoncle, A., & Mergny, J.-L. (2017). G-quadruplexes unfolding by RHAU helicase. *Biochimica Et Biophysica Acta. General Subjects*, *1861*(5 Pt B), 1382-1388.
- Guo, J. U., & Bartel, D. P. (2016a). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, *353*(6306).
- Guo, J. U., & Bartel, D. P. (2016b). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science (New York, N.Y.)*, *353*(6306).
- Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A. J., Shaw, P. G., Kim, M.-S., Maragakis, N. J., Troncoso, J. C., Pandey, A., Sattler, R., Rothstein, J. D., & Wang, J. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, *507*(7491), 195-200.
- Hamby, M. E., Coskun, V., & Sun, Y. E. (2008). Transcriptional regulation of neuronal differentiation : The epigenetic layer of complexity. *Biochimica et biophysica acta*, *1779*(8), 432-437.
- Hao, F., Ma, Y., & Guan, Y. (2019). Effects of Central Loop Length and Metal Ions on the Thermal Stability of G-Quadruplexes. *Molecules*, *24*(10), 1863.
- Hardin, C. C., Watson, T., Corregan, M., & Bailey, C. (1992). Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). *Biochemistry*, *31*(3), 833-841.
- Harkness, R. W., & Mittermaier, A. K. (2017). G-quadruplex dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1865*(11, Part B), 1544-1554.
- Herdy, B., Mayer, C., Varshney, D., Marsico, G., Murat, P., Taylor, C., D'Santos, C., Tannahill, D., & Balasubramanian, S. (2018a). Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Research*, *46*(21), 11592-11604.
- Herdy, B., Mayer, C., Varshney, D., Marsico, G., Murat, P., Taylor, C., D'Santos, C., Tannahill, D., & Balasubramanian, S. (2018b). Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Research*, *46*(21), 11592-11604.
- Herviou, P., Le Bras, M., Dumas, L., Hieblot, C., Gilhodes, J., Cioci, G., Hugnot, J.-P., Ameadan, A., Guillonneau, F., Dassi, E., Cammas, A., & Millevoi, S. (2020a). HnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma. *Nature Communications*, *11*(1), 2661.
- Herviou, P., Le Bras, M., Dumas, L., Hieblot, C., Gilhodes, J., Cioci, G., Hugnot, J.-P., Ameadan, A., Guillonneau, F., Dassi, E., Cammas, A., & Millevoi, S. (2020b). HnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma. *Nature Communications*, *11*(1), 2661.
- Huang, Vider, J., Serganova, I., & Blasberg, R. G. (2011). ATP-Binding Cassette Transporters Modulate Both Coelenterazine- and D-Luciferin-Based Bioluminescence Imaging. *Molecular Imaging*, *10*(3), 215-226.
- Huang, Zhang, J., Harvey, S. E., Hu, X., & Cheng, C. (2017). RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes & Development*, *31*(22), 2296-2309.
- Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn,

- D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., ... Yeats, C. (2009). InterPro : The integrative protein signature database. *Nucleic Acids Research*, 37(Database issue), D211-D215.
- Huppert, J. L., & Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Research*, 35(2), 406-413.
- Ingelsson, M. (2016). Alpha-Synuclein Oligomers—Neurotoxic Molecules in Parkinson’s Disease and Other Lewy Body Disorders. *Frontiers in Neuroscience*, 10(408).
- Ivanov, P., O’Day, E., Emara, M. M., Wagner, G., Lieberman, J., & Anderson, P. (2014). G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, 111(51), 18201-18206.
- Joachimi, A., Benz, A., & Hartig, J. S. (2009). A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorganic & Medicinal Chemistry*, 17(19), 6811-6815.
- Jodoin, R., Bauer, L., Garant, J.-M., Mahdi Laaref, A., Phaneuf, F., & Perreault, J.-P. (2014). The folding of 5’-UTR human G-quadruplexes possessing a long central loop. *RNA*, 20(7), 1129-1141.
- Jodoin, R., Carrier, J. C., Rivard, N., Bisailon, M., & Perreault, J.-P. (2019a). G-quadruplex located in the 5’UTR of the BAG-1 mRNA affects both its cap-dependent and cap-independent translation through global secondary structure maintenance. *Nucleic Acids Research*, 47(19), 10247–10266.
- Jodoin, R., Carrier, J. C., Rivard, N., Bisailon, M., & Perreault, J.-P. (2019b). G-quadruplex located in the 5’UTR of the BAG-1 mRNA affects both its cap-dependent and cap-independent translation through global secondary structure maintenance. *Nucleic Acids Research*.
- Jodoin, R., & Perreault, J.-P. (2018). G-quadruplexes formation in the 5’UTRs of mRNAs associated with colorectal cancer pathways. *PLOS ONE*, 13(12), e0208363.
- Kalia, L. V., Kalia, S. K., McLean, P. J., Lozano, A. M., & Lang, A. E. (2013). α -Synuclein oligomers and clinical implications for Parkinson disease. *Annals of Neurology*, 73(2), 155-169.
- Kalia, L. V., & Lang, A. E. (2015). Parkinson’s disease. *The Lancet*, 386(9996), 896-912.
- Kambach, C., & Mattaj, I. W. (1992). Intracellular distribution of the U1A protein depends on active transport and nuclear binding to U1 snRNA. *The Journal of Cell Biology*, 118(1), 11-21.
- Kawahara, Y., & Mieda-Sato, A. (2012). TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), 3347-3352.
- Kazemier, H. G., Paeschke, K., & Lansdorp, P. M. (2017). Guanine quadruplex monoclonal antibody 1H6 cross-reacts with restrained thymidine-rich single stranded DNA. *Nucleic Acids Research*, 45(10), 5913-5919.
- Kelly, T. J., Suzuki, H. I., Zamudio, J. R., Suzuki, M., & Sharp, P. A. (2019). Sequestration of microRNA-mediated target repression by the Ago2-associated RNA-binding protein FAM120A. *RNA*, 25(10), 1291-1297.
- Kiebler, M. A., Scheiffele, P., & Ule, J. (2013). What, where, and when : The importance of post-transcriptional regulation in the brain. *Frontiers in Neuroscience*, 7.
- Kikin, O., D’Antonio, L., & Bagga, P. S. (2006). QGRS Mapper : A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Research*, 34(Web Server issue), W676-W682.

- Kirk, R., Laman, H., Knowles, P. P., Murray-Rust, J., Lomonosov, M., Meziane, E. K., & McDonald, N. Q. (2008). Structure of a Conserved Dimerization Domain within the F-box Protein Fbxo7 and the PI31 Proteasome Inhibitor. *Journal of Biological Chemistry*, 283(32), 22325-22335.
- Klein Gunnewiek, J. M., Hussein, R. I., van Aarssen, Y., Palacios, D., de Jong, R., van Venrooij, W. J., & Gunderson, S. I. (2000). Fourteen residues of the U1 snRNP-specific U1A protein are required for homodimerization, cooperative RNA binding, and inhibition of polyadenylation. *Molecular and Cellular Biology*, 20(6), 2209-2217.
- Koukouraki, P., & Doxakis, E. (2016). Constitutive translation of human α -synuclein is mediated by the 5'-untranslated region. *Open Biology*, 6(4), 160022.
- Krishnan, R., Boddapati, N., & Mahalingam, S. (2018). Interplay between human nucleolar GNL1 and RPS20 is critical to modulate cell proliferation. *Scientific Reports*, 8.
- Krüger, A. C., Raarup, M. K., Nielsen, M. M., Kristensen, M., Besenbacher, F., Kjems, J., & Birkedal, V. (2010). Interaction of hnRNP A1 with telomere DNA G-quadruplex structures studied at the single molecule level. *European Biophysics Journal*, 39(9), 1343-1350.
- Kumari, S., Bugaut, A., Huppert, J. L., & Balasubramanian, S. (2007). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nature Chemical biology*, 3(4), 218-221.
- Kwok, C. K., & Balasubramanian, S. (2015). Targeted Detection of G-Quadruplexes in Cellular RNAs. *Angewandte Chemie*, 54(23), 6751-6754.
- Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S., & Balasubramanian, S. (2016). RG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nature Methods*, 13(10), 841-844.
- Lago, S., Tosoni, E., Nadai, M., Palumbo, M., & Richter, S. N. (2017). The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochimica et Biophysica Acta*, 1861(5Part B), 1371-1381.
- Lambert, É., Babeu, J.-P., Simoneau, J., Raisch, J., Lavergne, L., Lévesque, D., Jolibois, É., Avino, M., Scott, M. S., Boudreau, F., & Boisvert, F.-M. (2020). Human Hepatocyte Nuclear Factor 4- α encodes isoforms with distinct transcriptional functions. *Molecular & Cellular Proteomics: MCP*.
- Lammel, S., Lim, B. K., & Malenka, R. C. (2014). Reward and aversion in a heterogeneous midbrain dopamine system. *Neuropharmacology*, 76(Part B), 351-359.
- Lee, Kim, S., Hwang, D. W., Jeong, J. M., Chung, J.-K., Lee, M. C., & Lee, D. S. (2008). Development of a Dual-Luciferase Reporter System for In Vivo Visualization of MicroRNA Biogenesis and Posttranscriptional Regulation. *Journal of Nuclear Medicine*, 49(2), 285-294.
- Lee, Müller, F., & Visser, A. J. W. G. (2019). The Sensitized Bioluminescence Mechanism of Bacterial Luciferase. *Photochemistry and Photobiology*, 95(3), 679-704.
- Leipheimer, J., Bloom, A. L. M., Baumstark, T., & Panepinto, J. C. (2018). CNBP Homologues Gis2 and Znf9 Interact with a Putative G-Quadruplex-Forming 3' Untranslated Region, Altering Polysome Association and Stress Tolerance in *Cryptococcus neoformans*. *MSphere*, 3(4), e00201-18.
- Leppek, K., & Stoecklin, G. (2014a). An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Research*, 42(2), e13.

- Leppek, K., & Stoecklin, G. (2014b). An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Research*, *42*(2), e13.
- Li, Yang, Han, X., Zhou, S., Yan, Y., Xiang, X., Zhao, B., & Guo, X. (2018). Structural Features of DNA G-Quadruplexes Revealed by Surface-Enhanced Raman Spectroscopy. *The Journal of Physical Chemistry Letters*, *9*(12), 3245-3252.
- Li, Yingfu, Geyer, R., & Sen, D. (1996). Recognition of Anionic Porphyrins by DNA Aptamers. *Biochemistry*, *35*(21), 6911-6922.
- Lim, K. W., Amrane, S., Bouaziz, S., Xu, W., Mu, Y., Patel, D. J., Luu, K. N., & Phan, A. T. (2009). Structure of the human telomere in K⁺ solution : A stable basket-type G-quadruplex with only two G-tetrad layers. *Journal of the American Chemical Society*, *131*(12), 4301-4309.
- Lin, W., Sampathi, S., Dai, H., Liu, C., Zhou, M., Hu, J., Huang, Q., Campbell, J., Shin-Ya, K., Zheng, L., Chai, W., & Shen, B. (2013). Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *The EMBO Journal*, *32*(10), 1425-1439.
- Liu, E. Y., Cali, C. P., & Lee, E. B. (2017). RNA metabolism in neurodegenerative disease. *Disease Models & Mechanisms*, *10*(5), 509-518.
- Liu, Y.-H., Chen, Y.-J., Wu, H.-H., Wang, T.-Y., & Tsai, F.-J. (2014). Single Nucleotide Polymorphisms at the PRR3, ABCF1, and GNL1 Genes in the HLA Class I Region Are Associated with Graves' Ophthalmopathy in a Gender-Dependent Manner. *Ophthalmology*, *121*(10), 2033-2039.
- Lorenz, W. W., McCann, R. O., Longiaru, M., & Cormier, M. J. (1991). Isolation and expression of a cDNA encoding Renilla reniformis luciferase. *Proceedings of the National Academy of Sciences*, *88*(10), 4438.
- Lu, Y.-J., Hu, D.-P., Zhang, K., Wong, W.-L., & Chow, C.-F. (2016). New pyridinium-based fluorescent dyes : A comparison of symmetry and side-group effects on G-Quadruplex DNA binding selectivity and application in live cell imaging. *Biosensors and Bioelectronics*, *81*, 373-381.
- Makarova, O. V., Makarov, E. M., Urlaub, H., Will, C. L., Gentzel, M., Wilm, M., & Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *The EMBO Journal*, *23*(12), 2381-2391.
- Malgowska, M., Czajczynska, K., Gudanis, D., Tworak, A., & Gdaniec, Z. (2016). Overview of the RNA G-quadruplex structures. *Acta Biochimica Polonica*, *63*(4), 609-621.
- Masai, H., Fukatsu, R., Kakusho, N., Kanoh, Y., Moriyama, K., Ma, Y., Iida, K., & Nagasawa, K. (2019). Rif1 promotes association of G-quadruplex (G4) by its specific G4 binding and oligomerization activities. *Scientific Reports*, *9*(1), 8618-8618.
- Mathad, R. I., & Yang, D. (2011). G-quadruplex structures and G-quadruplex-interactive compounds. *Methods in Molecular Biology*, *735*, 77-96.
- Matsumura, K., Kawasaki, Y., Miyamoto, M., Kamoshida, Y., Nakamura, J., Negishi, L., Suda, S., & Akiyama, T. (2017). The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene*, *36*(9), 1191-1199.
- McRae, E. K. S., Booy, E. P., Moya-Torres, A., Ezzati, P., Stetefeld, J., & McKenna, S. A. (2017a). Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Research*, *45*(11), 6656-6668.

- McRae, E. K. S., Booy, E. P., Moya-Torres, A., Ezzati, P., Stetefeld, J., & McKenna, S. A. (2017b). Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Research*, *45*(11), 6656-6668.
- McRae, E. K. S., Dupas, S. J., Booy, E. P., Piragasam, R. S., Fahlman, R. P., & McKenna, S. A. (2020). An RNA guanine quadruplex regulated pathway to TRAIL-sensitization by DDX21. *RNA (New York, N.Y.)*, *26*(1), 44-57.
- Meier, M., Moya-Torres, A., Krahn, N. J., McDougall, M. D., Orriss, G. L., McRae, E. K. S., Booy, E. P., McEleney, K., Patel, T. R., McKenna, S. A., & Stetefeld, J. (2018). Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic Acids Research*, *46*(10), 5319-5331.
- Melko, M., & Bardoni, B. (2010). The role of G-quadruplex in RNA metabolism : Involvement of FMRP and FMR2P. *Biochimie*, *92*(8), 919-926.
- Melko, M., Douguet, D., Bensaid, M., Zongaro, S., Verheggen, C., Gecz, J., & Bardoni, B. (2011). Functional characterization of the AFF (AF4/FMR2) family of RNA-binding proteins : Insights into the molecular pathology of FRAXE intellectual disability. *Human Molecular Genetics*, *20*(10), 1873-1885.
- Mendoza, O., Bourdoncle, A., Boulé, J.-B., Brosh, R. M., & Mergny, J.-L. (2016). G-quadruplexes and helicases. *Nucleic Acids Research*, *44*(5), 1989-2006.
- Micco, M., Collie, G. W., Dale, A. G., Ohnmacht, S. A., Pazitna, I., Gunaratnam, M., Reszka, A. P., & Neidle, S. (2013). Structure-Based Design and Evaluation of Naphthalene Diimide G-Quadruplex Ligands As Telomere Targeting Agents in Pancreatic Cancer Cells. *Journal of Medicinal Chemistry*, *56*(7), 2959-2974.
- Millevoi, S., Moine, H., & Vagner, S. (2012). G-quadruplexes in RNA biology. *Wiley Interdisciplinary Reviews. RNA*, *3*(4), 495-507.
- Mirihana Arachchilage, G., Dassanayake, A. C., & Basu, S. (2015). A Potassium Ion-Dependent RNA Structural Switch Regulates Human Pre-miRNA 92b Maturation. *Chemistry & Biology*, *22*(2), 262-272.
- Miura, T., Benevides, J. M., & Thomas, G. J. (1995). A phase diagram for sodium and potassium ion control of polymorphism in telomeric DNA. *Journal of Molecular Biology*, *248*(2), 233-238.
- Mori, K., Lammich, S., Mackenzie, I. R. A., Forné, I., Zilow, S., Kretschmar, H., Edbauer, D., Janssens, J., Kleinberger, G., Cruts, M., Herms, J., Neumann, M., Van Broeckhoven, C., Arzberger, T., & Haass, C. (2013). HnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathologica*, *125*(3), 413-423.
- Moroo, L., Yamada, T., Makino, H., Tooyama, I., McGeer, P. L., McGeer, E. G., & Hirayama, K. (1994). Loss of insulin receptor immunoreactivity from the substantia nigra pars compacta neurons in Parkinson's disease. *Acta Neuropathologica*, *87*, 343-348.
- Morris, Bomhoff, G. L., Stanford, J. A., & Geiger, P. C. (2010). Neurodegeneration in an animal model of Parkinson's disease is exacerbated by a high-fat diet. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *299*(4), R1082-R1090.
- Morris, M. J., Negishi, Y., Pázsint, C., Schonhoft, J. D., & Basu, S. (2010). An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES. *Journal of the American Chemical Society*, *132*(50), 17831-17839.

- Morris, Vidoni, E. D., Perea, R. D., Rada, R., Johnson, D. K., Lyons, K., Pahwa, R., Burns, J. M., & Honea, R. A. (2014). Insulin resistance and gray matter volume in neurodegenerative disease. *Neuroscience*, *270*, 139-147.
- Nahvi, A., & Green, R. (2013). Structural Analysis of RNA Backbone Using In-Line Probing. Dans J. Lorsch (Éd.), *Methods in Enzymology* (Vol. 530, p. 381-397).
- Nicoludis, J. M., Barrett, S. P., Mergny, J.-L., & Yatsunyk, L. A. (2012a). Interaction of human telomeric DNA with N- methyl mesoporphyrin IX. *Nucleic Acids Research*, *40*(12), 5432-5447.
- Nicoludis, J. M., Barrett, S. P., Mergny, J.-L., & Yatsunyk, L. A. (2012b). Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Research*, *40*(12), 5432-5447.
- O'Day, E., Le, M. T. N., Imai, S., Tan, S. M., Kirchner, R., Arthanari, H., Hofmann, O., Wagner, G., & Lieberman, J. (2015). An RNA-binding Protein, Lin28, Recognizes and Remodels G-quartets in the MicroRNAs (miRNAs) and mRNAs It Regulates. *The Journal of Biological Chemistry*, *290*(29), 17909-17922.
- Olsen, C. M., & Marky, L. A. (2009). Energetic and Hydration Contributions of the Removal of Methyl Groups From Thymine to Form Uracil in G-Quadruplexes. *The Journal of Physical Chemistry B*, *113*(1), 9-11.
- Pagano, B., Mattia, C. A., Cavallo, L., Uesugi, S., Giancola, C., & Fraternali, F. (2008). Stability and Cations Coordination of DNA and RNA 14-Mer G-Quadruplexes : A Multiscale Computational Approach. *The Journal of Physical Chemistry B*, *112*(38), 12115-12123.
- Pandey, S., Agarwala, P., Jayaraj, G. G., Gargallo, R., & Maiti, S. (2015). The RNA Stem-Loop to G-Quadruplex Equilibrium Controls Mature MicroRNA Production inside the Cell. *Biochemistry*, *54*(48), 7067-7078.
- Parkinson, G. N., & Collie, G. W. (2019). X-Ray Crystallographic Studies of G-Quadruplex Structures. Dans D. Yang & C. Lin (Éds.), *G-Quadruplex Nucleic Acids : Methods and Protocols* (p. 131-155). Springer New York.
- Parkinson, G. N., Lee, M. P. H., & Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, *417*(6891), 876-880.
- Patel, S. S., & Donmez, I. (2006). Mechanisms of Helicases. *Journal of Biological Chemistry*, *281*(27), 18265-18268.
- Pietras, Z., Wojcik, M. A., Borowski, L. S., Szewczyk, M., Kulinski, T. M., Cysewski, D., Stepień, P. P., Dziembowski, A., & Szczesny, R. J. (2018). Dedicated surveillance mechanism controls G-quadruplex forming non-coding RNAs in human mitochondria. *Nature Communications*, *9*.
- Piovesan, A., Caracausi, M., Antonaros, F., Pelleri, M. C., & Vitale, L. (2016). GeneBase 1.1 : A tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. *Database: The Journal of Biological Databases and Curation*, *2016*.
- Price, S. R., Evans, P. R., & Nagai, K. (1998). Crystal structure of the spliceosomal U2B'-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature*, *394*(6694), 645-650.
- Quinones-Valdez, G., Tran, S. S., Jun, H.-I., Bahn, J. H., Yang, E.-W., Zhan, L., Brümmer, A., Wei, X., Van Nostrand, E. L., Pratt, G. A., Yeo, G. W., Graveley, B. R., & Xiao, X. (2019). Regulation of RNA editing by RNA-binding proteins in human cells. *Communications Biology*, *2*(1), 19.

- Rangadurai, A., Zhou, H., Merriman, D. K., Meiser, N., Liu, B., Shi, H., Szymanski, E. S., & Al-Hashimi, H. M. (2018). Why are Hoogsteen base pairs energetically disfavored in A-RNA compared to B-DNA? *Nucleic Acids Research*, *46*(20), 11099-11114.
- Ranjbar, B., & Gill, P. (2009). Circular Dichroism Techniques: Biomolecular and Nanostructural Analyses- A Review. *Chemical Biology & Drug Design*, *74*(2), 101-120.
- Reich, S. G., & Savitt, J. M. (2019). Parkinson's Disease. *Neurology for the Non-Neurologist*, *103*(2), 337-350.
- Rouleau, S. G., Garant, J.-M., Bolduc, F., Bisailon, M., & Perreault, J.-P. (2018). G-Quadruplexes influence pri-microRNA processing. *RNA Biology*, *15*(2), 198-206.
- Rouleau, S., Jodoin, R., Garant, J.-M., & Perreault, J.-P. (2017). RNA G-Quadruplexes as Key Motifs of the Transcriptome. *Catalytically Active Nucleic Acids*, *170*, 1-20.
- Ruiz-Martinez, J., Krebs, C. E., Makarov, V., Gorostidi, A., Martí-Massó, J. F., & Paísán-Ruiz, C. (2015). GIGYF2 mutation in late-onset Parkinson's disease with cognitive impairment. *Journal of human genetics*, *60*(10), 637-640.
- Sanders, C. M. (2010). Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochemical Journal*, *430*(1), 119-128.
- Sauer, M., Juranek, S. A., Marks, J., De Magis, A., Kazemier, H. G., Hilbig, D., Benhalevy, D., Wang, X., Hafner, M., & Paeschke, K. (2019). DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nature Communications*, *10*(1), 1-15.
- Scalabrin, M., Frasson, I., Ruggiero, E., Perrone, R., Tosoni, E., Lago, S., Tassinari, M., Palù, G., & Richter, S. N. (2017). The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes. *Scientific Reports*, *7*, 45244.
- Schaeffer, C., Bardoni, B., Mandel, J.-L., Ehresmann, B., Ehresmann, C., & Moine, H. (2001). The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *The EMBO Journal*, *20*(17), 4803-4813.
- Schüttelkopf, A. W., & van Aalten, D. M. F. (2004). PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D: Biological Crystallography*, *60*(8), 1355-1363.
- Selvaraj, S., & Piramanayagam, S. (2019). Impact of gene mutation in the development of Parkinson's disease. *Genes & Diseases*, *6*(2), 120-128.
- Sengar, A., Vandana, J. J., Chambers, V. S., Di Antonio, M., Winnerdy, F. R., Balasubramanian, S., & Phan, A. T. (2019). Structure of a (3+1) hybrid G-quadruplex in the PARP1 promoter. *Nucleic Acids Research*, *47*(3), 1564-1572.
- Serikawa, T., Spanos, C., von Hacht, A., Budisa, N., Rappsilber, J., & Kurreck, J. (2018). Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs. *Biochimie*, *144*, 169-184.
- Shahid, R., Bugaut, A., & Balasubramanian, S. (2010). The BCL-2 5' Untranslated Region Contains an RNA G-Quadruplex-Forming Motif That Modulates Protein Expression. *Biochemistry*, *49*(38), 8300-8306.
- Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., & Hurley, L. H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences*, *99*(18), 11593.

- Simone, R., Fratta, P., Neidle, S., Parkinson, G. N., & Isaacs, A. M. (2015). G-quadruplexes : Emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Letters*, *589*(14), 1653-1668.
- Singh, A. K., Choudhury, S. R., De, S., Zhang, J., Kissane, S., Dwivedi, V., Ramanathan, P., Petric, M., Orsini, L., Hebenstreit, D., & Brogna, S. (2019). The RNA helicase UPF1 associates with mRNAs co-transcriptionally and is required for the release of mRNAs from gene loci. *eLife*, *8*.
- Šket, P., Pohleven, J., Kovanda, A., Štalekar, M., Župunski, V., Zalar, M., Plavec, J., & Rogelj, B. (2015). Characterization of DNA G-quadruplex species forming from C9ORF72 G4C2-expanded repeats associated with amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Neurobiology of Aging*, *36*(2), 1091-1096.
- Sliter, D. A., Martinez, J., Hao, L., Chen, X., Sun, N., Fischer, T. D., Burman, J. L., Li, Y., Zhang, Z., Narendra, D. P., Cai, H., Borsche, M., Klein, C., & Youle, R. J. (2018). Parkin and PINK1 mitigate STING-induced inflammation. *Nature*, *561*(7722), 258.
- Smyth, M. S., & Martin, J. H. (2000). X ray crystallography. *Molecular Pathology : MP*, *53*(1), 8-14.
- Snyder, H., Mensah, K., Theisler, C., Lee, J., Matouschek, A., & Wolozin, B. (2003). Aggregated and Monomeric α -Synuclein Bind to the S6' Proteasomal Protein and Inhibit Proteasomal Function. *Journal of Biological Chemistry*, *278*(14), 11753-11759.
- Sommer, G., Sendlmeier, C., & Heise, T. (2020). Salt-Dependent Modulation of the RNA Chaperone Activity of RNA-Binding Protein La. Dans T. Heise (Éd.), *RNA Chaperones : Methods and Protocols* (p. 121-136). Springer US.
- Stefanis, L., Larsen, K. E., Rideout, H. J., Sulzer, D., & Greene, L. A. (2001). Expression of A53T Mutant But Not Wild-Type α -Synuclein in PC12 Cells Induces Alterations of the Ubiquitin-Dependent Degradation System, Loss of Dopamine Release, and Autophagic Cell Death. *The Journal of Neuroscience*, *21*(24), 9549-9560.
- Stephanenko, A. A., & Dmitrenko, V. V. (2015). HEK293 in cell biology and cancer research : Phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene*, *569*(2), 182-190.
- Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J.-L., & Moine, H. (2011). G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO reports*, *12*(7), 697-704.
- Sun, Bennett, R. J., & Maizels, N. (1999). The *Saccharomyces cerevisiae* Sgs1 helicase efficiently unwinds G-G paired DNAs. *Nucleic Acids Research*, *27*(9), 1978-1984.
- Sun, D., Guo, K., Rusche, J. J., & Hurley, L. H. (2005). Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Research*, *33*(18), 6070-6080.
- Sun, & Hurley, L. H. (2010). Biochemical Techniques for the Characterization of G-Quadruplex Structures : EMSA, DMS Footprinting, and DNA Polymerase Stop Assay. *Methods in molecular biology*, *608*, 65-79.
- Sun, N.-F., Meng, Q.-Y., Hu, S.-Y., Tian, A.-L., Wang, R.-H., Liu, Z.-X., & Xu, L. (2011). Correlation between the expression of the BAG-1 gene and clinicopathologic factors in colorectal cancer. *Journal of Cancer Research and Clinical Oncology*, *137*(10), 1419-1424.

- Tan, W., Zhou, J., Gu, J., Xu, M., Xu, X., & Yuan, G. (2016). Probing the G-quadruplex from hsa-miR-3620-5p and inhibition of its interaction with the target sequence. *Talanta*, *154*, 560-566.
- Tanaka, K. (2009). The proteasome : Overview of structure and functions. *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, *85*(1), 12-36.
- Tang, C.-F., & Shafer, R. H. (2006). Engineering the Quadruplex Fold. *Journal of the American Chemical Society*, *128*(17), 5966-5973.
- Tenzer, S., Moro, A., Kuharev, J., Francis, A. C., Vidalino, L., Provenzani, A., & Macchi, P. (2013). Proteome-Wide Characterization of the RNA-Binding Protein RALY-Interactome Using the in Vivo-Biotinylation-Pulldown-Quant (iBioPQ) Approach. *Journal of Proteome Research*, *12*(6), 2869-2884.
- Thandapani, P., O'Connor, T. R., Bailey, T. L., & Richard, S. (2013). Defining the RGG/RG Motif. *Molecular Cell*, *50*(5), 613-623.
- Thandapani, P., Song, J., Gandin, V., Cai, Y., Rouleau, S. G., Garant, J.-M., Boisvert, F.-M., Yu, Z., Perreault, J.-P., Topisirovic, I., & Richard, S. (2015). Aven recognition of RNA G-quadruplexes regulates translation of the mixed lineage leukemia protooncogenes. *ELife*, *4*, e06234.
- Thomas, K. J., McCoy, M. K., Blackinton, J., Beilina, A., van der Brug, M., Sandebring, A., Miller, D., Maric, D., Cedazo-Minguez, A., & Cookson, M. R. (2011). DJ-1 acts in parallel to the PINK1/parkin pathway to control mitochondrial function and autophagy. *Human Molecular Genetics*, *20*(1), 40-50.
- Tong, L., Li, L., Chen, Z., Wang, Q., & Tang, B. (2013). Stable label-free fluorescent sensing of biothiols based on ThT direct inducing conformation-specific G-quadruplex. *Biosensors and Bioelectronics*, *49*, 420-425.
- Umar, M. I., Ji, D., Chan, C.-Y., & Kwok, C. K. (2019). G-Quadruplex-Based Fluorescent Turn-On Ligands and Aptamers : From Development to Applications. *Molecules*,
- Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J.-Y., Cody, N. A. L., Dominguez, D., Olson, S., Sundararaman, B., Zhan, L., Bazile, C., Bouvrette, L. P. B., Bergalet, J., Duff, M. O., Garcia, K. E., Gelboin-Burkhart, C., ... Yeo, G. W. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature*, *583*(7818), 711-719.
- Vasilyev, N., Polonskaia, A., Darnell, J. C., Darnell, R. B., Patel, D. J., & Serganov, A. (2015). Crystal structure reveals specific recognition of a G-quadruplex RNA by a β -turn in the RGG motif of FMRP. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(39), E5391-E5400.
- Vingill, S., Brockelt, D., Lancelin, C., Tatenhorst, L., Dontcheva, G., Preisinger, C., Schwedhelm-Domeyer, N., Joseph, S., Mitkovski, M., Goebbels, S., Nave, K., Schulz, J. B., Marquardt, T., Lingor, P., & Stegmüller, J. (2016). Loss of FBXO7 (PARK15) results in reduced proteasome activity and models a parkinsonism-like phenotype in mice. *The EMBO Journal*, *35*(18), 2008-2025.
- von Hacht, A., Seifert, O., Menger, M., Schütze, T., Arora, A., Konthur, Z., Neubauer, P., Wagner, A., Weise, C., & Kurreck, J. (2014). Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Research*, *42*(10), 6630-6644.
- Vummidi, B. R., Alzeer, J., & Luedtke, N. W. (2013). Fluorescent Probes for G-Quadruplex Structures. *ChemBioChem*, *14*(5), 540-558.

- Wang, X., Goodrich, K. J., Conlon, E. G., Gao, J., Erbse, A. H., Manley, J. L., & Cech, T. R. (2019). C9orf72 and Triplet Repeat Disorder RNAs: G-quadruplex formation, binding to PRC2 and implications for disease mechanisms. *RNA*, *25*, 935-947.
- Wauer, T., Simicek, M., Schubert, A., & Komander, D. (2015). Mechanism of phospho-ubiquitin induced PARKIN activation. *Nature*, *524*(7565), 370-374.
- Webb, J. L., Ravikumar, B., Atkins, J., Skepper, J. N., & Rubinsztein, D. C. (2003). α -Synuclein Is Degraded by Both Autophagy and the Proteasome. *Journal of Biological Chemistry*, *278*(27), 25009-25013.
- Williams, E. T., Chen, X., & Moore, D. J. (2017). VPS35, the Retromer Complex and Parkinson's Disease. *Journal of Parkinson's Disease*, *7*(2), 219-233.
- Wolfe, A. L., Singh, K., Zhong, Y., Drewe, P., Rajasekhar, V. K., Sanghvi, V. R., Mavrikis, K. J., Jiang, M., Roderick, J. E., Van der Meulen, J., Schatz, J. H., Rodrigo, C. M., Zhao, C., Rondou, P., de Stanchina, E., Teruya-Feldstein, J., Kelliher, M. A., Speleman, F., Porco, J. A., Jr, ... Wendel, H.-G. (2014). RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature*, *513*(7516), 65-70.
- Wong, H. M., Stegle, O., Rodgers, S., & Huppert, J. L. (2010). A Toolbox for Predicting G-Quadruplex Formation and Stability. *Journal of Nucleic Acids*, *2010*, 564946.
- Wu, Cheng, W.-C., Chen, C.-Y., Wu, M.-C., Wang, Y.-C., Tseng, Y.-H., Chuang, T.-J., & Shen, C.-K. J. (2019). Transcriptomopathies of pre- and post-symptomatic frontotemporal dementia-like mice with TDP-43 depletion in forebrain neurons. *Acta Neuropathologica Communications*, *7*(50).
- Wu, & Maizels, N. (2001). Substrate-specific inhibition of RecQ helicase. *Nucleic Acids Research*, *29*(8), 1765-1771.
- Wu, Shin-ya, K., & Brosh, R. M., Jr. (2008). FANCI helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Molecular and Cellular Biology*, *28*(12), 4116-4128.
- Wu, Xing, Z., Tran, E. J., & Yang, D. (2019). DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation. *Proceedings of the National Academy of Sciences*, 201909047.
- Xicoy, H., Wieringa, B., & Martens, G. J. M. (2017). The SH-SY5Y cell line in Parkinson's disease research: A systematic review. *Molecular Neurodegeneration*, *12*(1), 10.
- Xu, S., Li, Q., Xiang, J., Yang, Q., Sun, H., Guan, A., Wang, L., Liu, Y., Yu, L., Shi, Y., Chen, H., & Tang, Y. (2016). Thioflavin T as an efficient fluorescence sensor for selective recognition of RNA G-quadruplexes. *Scientific Reports*, *6*, 24793.
- Yagi, R., Miyazaki, T., & Oyoshi, T. (2018). G-quadruplex binding ability of TLS/FUS depends on the β -spiral structure of the RGG domain. *Nucleic Acids Research*, *46*(12), 5894-5901.
- Yang, S. Y., Lejault, P., Chevrier, S., Boidot, R., Robertson, A. G., Wong, J. M. Y., & Monchaud, D. (2018). Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nature Communications*, *9*(1), 1-11.
- Yang, X., Chernenko, G., Hao, Y., Ding, Z., Pater, M. M., Pater, A., & Tang, S. C. (1998). Human BAG-1/RAP46 protein is generated as four isoforms by alternative translation initiation and overexpressed in cancer cells. *Oncogene*, *17*(8), 981-989.
- Yokoi, A., Kotake, Y., Takahashi, K., Kadowaki, T., Matsumoto, Y., Minoshima, Y., Sugi, N. H., Sagane, K., Hamaguchi, M., Iwata, M., & Mizui, Y. (2011). Biological validation that SF3b is a target of the antitumor macrolide pladienolide. *The FEBS Journal*, *278*(24), 4870-4880.

- Yoon, J.-H., & Gorospe, M. (2016). Identification of mRNA-interacting factors by MS2-TRAP (MS2-tagged RNA affinity purification). *Methods in molecular biology (Clifton, N.J.)*, 1421, 15-22.
- Yu, C.-H., Teulade-Fichou, M.-P., & Olsthoorn, R. C. L. (2014). Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Research*, 42(3), 1887-1892.
- Zahler, A. M., Williamson, J. R., Cech, T. R., & Prescott, D. M. (1991). Inhibition of telomerase by G-quartet DMA structures. *Nature*, 350(6320), 718-720.
- Zavodszky, E., Seaman, M. N. J., Moreau, K., Jimenez-Sanchez, M., Breusegem, S. Y., Harbour, M. E., & Rubinsztein, D. C. (2014). Mutation in VPS35 associated with Parkinson's disease impairs WASH complex association and inhibits autophagy. *Nature Communications*, 5, 3828.
- Zecca, L., Tampellini, D., Gerlach, M., Riederer, P., Fariello, R. G., & Sulzer, D. (2001). Substantia nigra neuromelanin: Structure, synthesis, and molecular behaviour. *Molecular Pathology : MP*, 54(6), 414-418.
- Zhang, Yang, M., Duncan, S., Yang, X., Abdelhamid, M. A. S., Huang, L., Zhang, H., Benfey, P. N., Waller, Z. A. E., & Ding, Y. (2019). G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Research*, 47(22), 11746-11754.
- Zhang, Yu, L., Ye, S., Xie, J., Huang, X., Zheng, K., & Sun, B. (2019). MOV10L1 Binds RNA G-Quadruplex in a Structure-Specific Manner and Resolves It More Efficiently Than MOV10. *IScience*, 17, 36-48.
- Zhou, Z. D., Lee, J. C. T., & Tan, E. K. (2018). Pathophysiological mechanisms linking F-box only protein 7 (FBXO7) and Parkinson's disease (PD). *Mutation Research/Reviews in Mutation Research*, 778, 72-78.

ANNEXE 1

Bolduc, F., Turcotte M.A., & Perreault J.P. (2020). The Small Nuclear Ribonucleoprotein Polypeptide A (SNRPA) binds to the G-quadruplex of the BAG-1 5'UTR. *Biochimie*, 176, 122-127.

**The Small Nuclear Ribonucleoprotein Polypeptide A (SNRPA) binds to
the G-quadruplex of the BAG-1 5'UTR**

François Bolduc, Marc-Antoine Turcotte, Jean-Pierre Perreault*

RNA Group/Groupe ARN, Département de biochimie et de génomique fonctionnelle,
Pavillon de Recherche Appliquée au Cancer, Université de Sherbrooke, 3201 Jean-
Mignault, Sherbrooke (Québec) Canada
J1E 4K8

*Corresponding author : Jean-Pierre Perreault

E-mail : jean-pierre.perreault@usherbrooke.ca

Abstract

The protein "BCL-2-associated athanogene-1" (BAG-1), which exists in multiple isoforms, promotes cancer cell survival and is overexpressed in many different cancers. As a result, BAG-1-targeted therapy appears to be a promising strategy with which to treat cancer. It has previously been shown that the 5'UTR of the BAG-1 mRNA contains a guanine rich region that folds into a G-quadruplex structure which can modulate both its cap-dependent and its cap-independent translation. Accumulating data regarding G-quadruplex binding proteins suggest that these proteins can play a central role in gene expression. Consequently, the identification of the proteins that could potentially bind to the G-quadruplex of the BAG-1 mRNA was undertaken. Label-free RNA pulldown assays were performed using protein extracts from colorectal cancer cells and this leads to the detection of RNA G4 binding proteins by LC-MS/MS. The use of G-quadruplex containing RNA, as well as of a mutated version, ensured that the proteins identified were specific for the RNA G-quadruplex structure and not just general RNA binding proteins. Following confirmation of the interaction, the Small Nuclear Ribonucleoprotein Polypeptide A (SNRPA) was shown to bind directly to the BAG-1 mRNA through the G-quadruplex, and knock down experiments in colorectal cancer cells suggested that it can modulate the expression level of BAG-1.

Keywords

G-quadruplex, SNRPA, gene regulation, RNA binding proteins, RNA structure, Cancer

1. Introduction

RNAs contain multiple *cis*- (ex. sequence and structural motifs) and *trans*- (ex. RNA binding proteins; RBPs) acting features that are involved in the regulation of gene expression. One interesting example is the RNA G4 (rG4) structure that involves the stacking of Hoogsteen hydrogen-bonded G-tetrads which, once stabilized by the chelation of monovalent metal ions such as potassium, form a stable four-stranded helical structure (Fay et al., 2017b; Malgowska et al., 2016). G-quadruplexes are highly abundant (Kwok et al., 2016) and are involved in many post-transcriptional regulatory mechanisms including the regulation of translation, alternative splicing, polyadenylation, mRNA localization and microRNA binding and maturation (Millevoi et al., 2012; S. G. Rouleau et al., 2018). The dysfunction of G4s suggests they may be involved in the pathogenesis of diseases such as neurological disorders and cancer (Cammass & Millevoi, 2017).

The protein "BCL-2-associated athanogene-1" (BAG-1) is a multifunctional protein that is overexpressed in many cancers, and whose expression is correlated with both the pathologic grade and distant metastasis. Consequently, this protein is an interesting potential drug target for cancer treatment (N.-F. Sun et al., 2011). The BAG-1 mRNA is alternatively translated into three main isoforms: the long, BAG-1L (50 kDa); the medium, BAG-1M (46 kDa); and, the short, BAG-1S (36 kDa). These isoforms are translated from the same mRNA transcript via two mechanisms. The first mechanism is an alternative translational initiation called "leaky scanning" that occurs at the three alternative in-frame start codons that are present in the 5'UTR of the mRNA (X. Yang et al., 1998). The second is an internal translational initiation mechanism which uses an internal ribosome entry site (IRES) secondary structure that is located upstream of the third start codon (Coldwell et al., 2001). Recently, it was observed that the BAG-1 protein is overexpressed in colorectal cancer (CRC) tissue while the RNA level is decreased, suggesting some sort of a control at the translational level (Jodoin et al., 2019b). In the same report, two novel structural features found in the 5'UTR of the BAG-1 mRNA, an rG4 and a repressive upstream open reading frame (uORF) (Jodoin et al., 2019b; Jodoin & Perreault, 2018), were also described. For the first time it was demonstrated that an rG4 located in the 5'UTR acts at the same time on both cap-dependent and cap-independent translation. In fact, this rG4 acts like most of the 5'UTR rG4, by halting the ribosome, thereby repressing the cap-dependent

translation. However, it also modulates translation by maintaining the global 5'UTR secondary structure, which is essential for the IRES translational mechanism.

RNA G4 are commonly believed “to be events and not things” (Guo & Bartel, 2016b). In other words, rG4s are folded transiently. In recent years, an increasing number of proteins with binding affinities for G4s have been identified (Fay et al., 2017b). In order to decipher and understand the complex expression of BAG-1, the identification and characterization of the G4 RNA binding proteins (G4 RBP) able to bind the G4 located in the 5'UTR of the BAG-1 mRNA was undertaken. One of the interesting candidates that appeared was the Small Nuclear Ribonucleoprotein Polypeptide A (SNRPA).

2. Materials and methods

The details of all protocols are available in the Supplementary File.

2.1 Plasmids

The four repeats of the S1m aptamer (S1mX4) were constructed by gene synthesis (GeneArt, Life Technologies) and were then cloned into the plasmid pMARQ, generating the plasmid pMARQ_S1mX4. The plasmids pMARQ_BAG1G4_S1m4X and pMARQ_BAG1G/Amut_S1m4X were created by PCR-amplifying the first portions of the 5'UTRs of BAG-1 and of a G/A mutated version contained in the psiCHECK-2 plasmid, respectively (Jodoin et al., 2019b). The PCR-fragments were digested with *EcoRI* and *BamHI* and were inserted into the *EcoRI* and *BamHI* restriction sites of the plasmid pMARQ_S1mX4 (Supplementary Fig. S1).

2.2 Label-free RNA affinity purification and LC-MS/MS analysis

RNAs were synthesized by *in vitro* transcription and were gel purified on 5% denaturing polyacrylamide gels as described previously (Jodoin & Perreault, 2018). Cellular extracts were prepared from 4 confluent 15-cm dishes of HCT-116 cells. The cell extracts were mixed with streptavidin-coated sepharose beads pre-bonded with chimeric RNAs, and the bound peptides were analyzed using an LC-MS/MS as illustrated in Figure 1A. The proteins were then identified by MaxQuant analysis (Cox et al., 2014).

2.3 Western blot validation

The interactions of the potential RNA binding proteins (RBP) with the RNA were validated by western blot using commercially available specific antibodies and both the WT and the G/A mutant sequences.

2.4 Gene knock down using siRNAs in HCT-116 cells

The reverse transfection of siRNAs was performed in 6-well plates using Lipofectamine RNAiMAX (Thermo Fisher) according to manufacturer's protocole. After the cells had recovered, total RNA extracts were prepared and the specific mRNAs were analyzed by RT-qPCR at the Laboratory of Functional Genomics of the Université de Sherbrooke (LGFUS). The proteins were monitored by western blotting.

2.5 Recombinant SNRPA purification and electrophoretic mobility shift assays (EMSA)

E. coli Rosetta-2 cells transformed with the plasmid pET42a_SNRPA-cHis (GeneScript) were used for the production of recombinant SNRPA that was subsequently purified using a Ni-NTA column. The purified SNRPA was used in electrophoretic mobility assays that were performed in the presence of either the short or the long 5'-labeled WT or G/A mutant BAG-1 5'UTR versions.

3. Results

3.1 Identification of the proteins that are able to bind to the G-quadruplex of the BAG-1 5'UTR

Label-free RNA affinity purification assays using a chimeric RNA composed of the first 83 nucleotides (nts) of the BAG-1 5'UTR, which contained either the wild type rG4 (WT) or a G/A mutated version fused to four repeats of the S1m aptamer, were performed. The S1m aptamer is a modified streptavidin-binding RNA aptamer (Leppek & Stoecklin, 2014b). *In-line* probing experiments were performed in order to confirm that the rG4 was

correctly folded in the chimeric constructs. The results confirmed the efficient folding of the G4 in the presence of potassium, but not in the presence of lithium. An increased accessibility of the nucleotides located in the loops of the rG4 structure was clearly observed in the presence of potassium, but not lithium ions, a hallmark of rG4 folding. This accessibility was not encountered with the G/A mutant version, regardless the nature of the salt present (see Supplementary Fig. S2). Specifically, total protein extracts from HCT116 CRC cells were incubated with the chimeric RNA that had previously been bound to streptavidin-coated sepharose beads (Fig. 1A). Any non-specific interactions were washed away, and the bound proteins were then subjected to trypsin digestion. The resulting peptides were injected into an LC-MS/MS. The data were analyzed in order to determine which proteins are enriched in the WT samples relative to the G/A mutant samples. The experiment was performed in triplicate, and identified 37 potential candidates that were presented in at least two of the replicates and whose fold enrichment was >2.5 (Fig. 1B and see Supplementary Table S1 for a detailed list). Curiously, half of the proteins on this list are ones related to DNA (i.e. 10 DNA chromatin binding protein, 4 DNA helicases, etc.), and were not considered further since the objective was to find RBP. More interestingly, 3 RNA helicases, as well as 9 proteins associated with splicing and translation, were retrieved. In order to increase the confidence in these findings, western blotting experiments were performed using both the WT and the G/A mutant RNAs and antibodies specific for proteins known to bind RNA and for which antibodies were available from collaborators. Enrichment was observed in the WT sample for 6 proteins out of the 9 candidates tested (Fig. 1C).

3.2 Impact of the knock down of SNRPA on BAG-1 expression in HCT-116 cells

In order to identify candidates that can be associated with the modulation of the translation of BAG-1 mRNA, gene knock down experiments using specific siRNAs were performed. Two siRNAs targeting the expression of each of the 6 proteins observed to be enriched above were reverse transfected into HCT-116 cells. The expression of BAG-1 at the protein level (by western blots) as well as at the RNA level (by RT-qPCR) was monitored at both 48 and 72 h post-transfection. Consistent translational modulation of BAG-1 was observed only for the Small Nuclear Ribonucleoprotein Polypeptide A (SNRPA). Indeed, a knock

down of the amount of SNRPA protein present of close to 90% at 48 h post-transfection, and close to 100% at 72 h, was observed (Fig. 2A). The intensities of the bands corresponding to the BAG-1 proteins were then measured and were discovered to be consistently increased in cells in which the expression of SNRPA was knocked down (Fig. 2B). The highest levels of expression were recorded for BAG-1S. Consistent increases in the expression of the other two isoforms (i.e. M and L) were also observed, although at lower levels. Concomitantly, level of BAG-1 RNA was not increased, which is characteristic of a translational regulation.

3.3 SNRPA can directly bind the G4 of the BAG-1 mRNA

In order to characterize the interaction between the rG4 of the BAG-1 mRNA and the SNRPA protein, the recombinant histidine-tagged SNRPA was purified on a nickel column and used in electrophoretic mobility shift assays (EMSA). The concentration of the 5'-labelled RNA was kept constant (~1 nM), while the protein concentration ranged from 0 to 45 μ M. Two distinct RNA species of different lengths were tested (i.e. a 30 and an 83 nt versions). The folding of the G4 of both versions was confirmed for the WT of these RNAs using N-methyl mesoporphyrin (NMM) fluorescence assays. NMM is known to bind G4 structures with a parallel topology (Nicoludis et al., 2012b). The ligand by itself emits a very low fluorescence, but upon binding to a parallel G4 its fluorescence can be increased from 2- to 10-folds (see Supplementary Fig. S3). Using the first 83 nts of the BAG-1 5'UTR, the SNRPA has a higher affinity for the WT as compared to the G/A mutant version (Fig. 2C). Dissociation constants (K_d) of 2.45 μ M and 8.74 μ M were calculated for the WT and the G/A mutant, respectively. The presence of several bands corresponding to the complex is most likely due to the fact that the SNRPA homodimerized, a characteristic that has been previously reported by others based on EMSA assays (Klein Gunnewiek et al., 2000). When using a shorter version (30 nts, i.e. without the surrounding nucleotides), the same conclusion was reached (K_d of 4.86 μ M for the WT G4 and 20.00 μ M for the G/A mutant sequence; Fig. 2D). The surrounding nucleotides seem to increase the affinity of SNRPA for the RNA by 2-folds. Moreover, only two bands of RNA/SNRPA complexes were observed, suggesting that there is less homodimerization with this smaller RNA species.

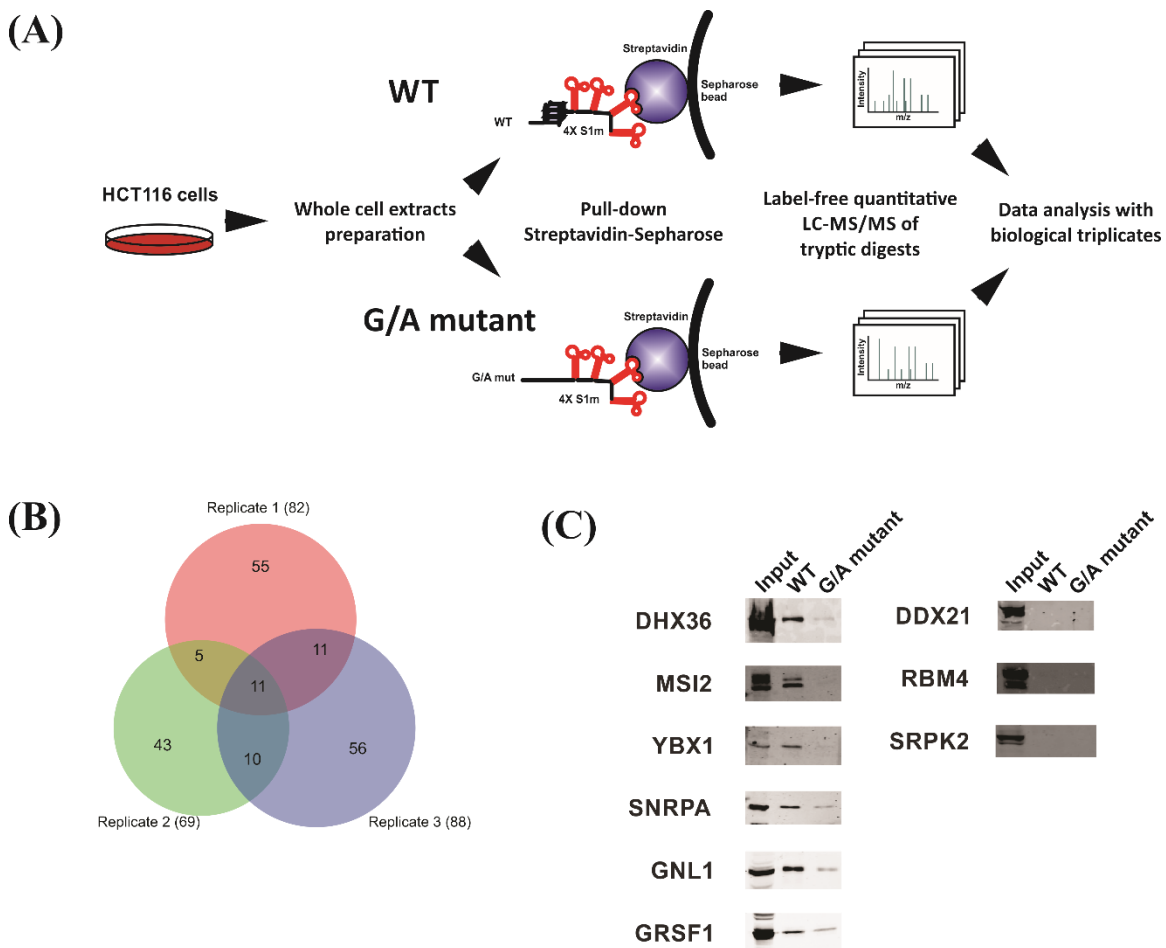


Fig.1. Label-free RNA affinity purification and results from LC-MS/MS. (A) HCT-116 protein extracts were mixed with streptavidin-coated Sepharose beads pre-bonded with chimeric RNAs. Following washes to remove any non-specific interactions, the tryptic digests were eluted and analyzed by LC-MS/MS. (B) Venn diagram generated from three RNA affinity purification assays. The number of proteins that pass the threshold (i.e. >2.5 fold enrichment, >5% coverage and a MaxQuant score >15) is indicated for each replicate. The different colors refer to the different replicates. (C) Western blots experiments used to validate the LC-MS/MS data for some of the RBP candidates.

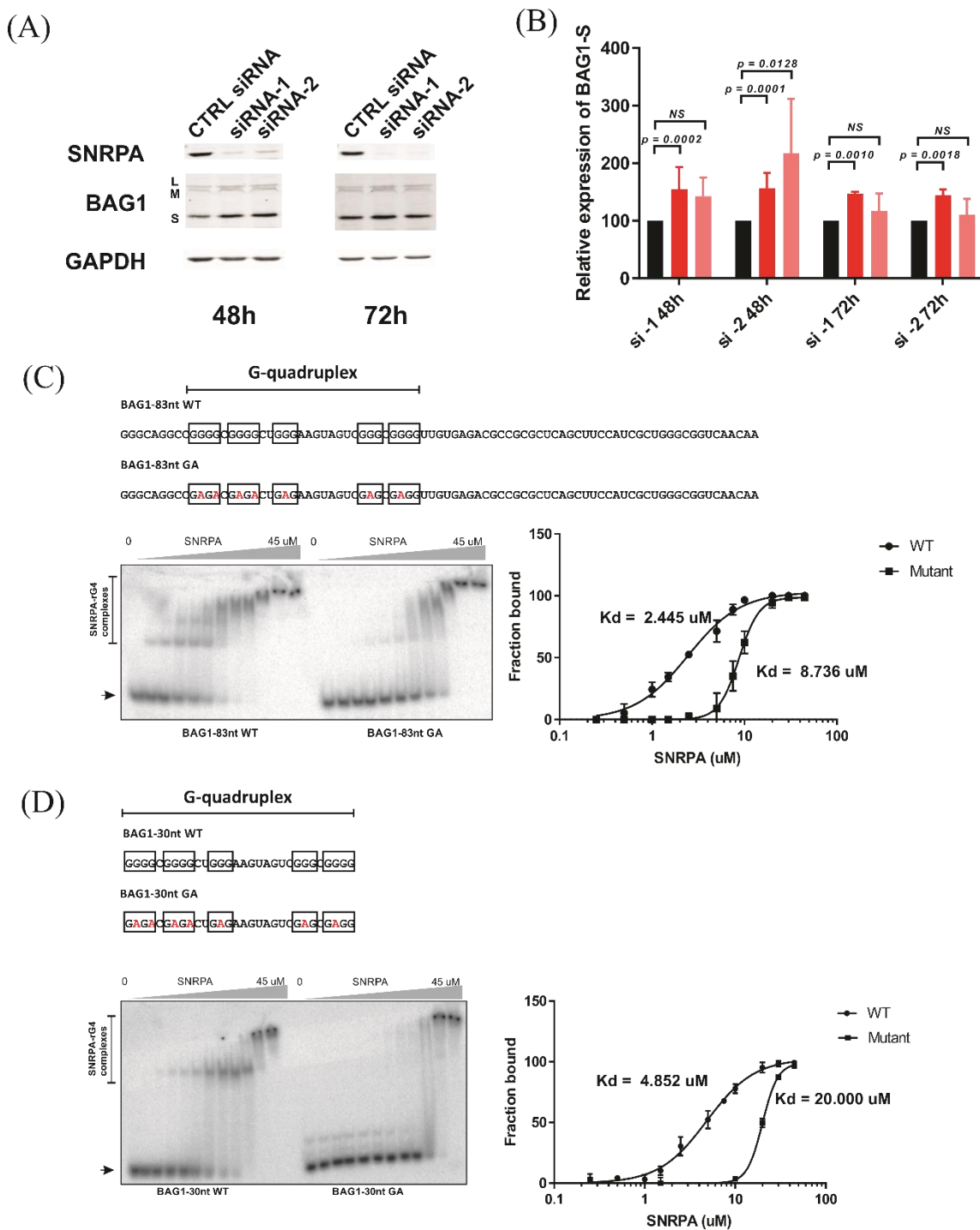


Fig.2. Knock down of SNRPA in HCT-116 cells and EMSA. (A) Western blot using protein extracts from HCT-116 cells knocked down for the expression of SNRPA at both 48 and 72 h post-transfection of siRNAs. Two specific siRNAs (siRNA-1 and siRNA-2) and one control (CTRL) were used. The SNRPA and BAG-1 proteins were detected, as well as GAPDH which was used as a loading control. L, M and S refer to the large, medium

and small BAG-1 isoforms, respectively. (B) Histograms of the relative expression of BAG-1 protein calculated from the western blots (red bars) and the mRNA (pink bars) as determined by RT-qPCR. The Black bars represent the expression levels from cells transfected with the control siRNA. The results presented are the means of three independent experiments, and each time two technical replicates were performed. The error bars indicate the standard deviations. P values were calculated by Dunnett's multiple comparisons test. (C) EMSA using increasing concentrations of purified SNRPA-his protein with a constant concentration of the first 83 nts of either the BAG-1 5'UTR WT or the G/A mutant version (1nM). The arrow indicates the band corresponding to the free RNA, and the different complexes retained on the gel are identified on the left. The mutations of G to A are shown in red. (Right) Hill slope of two independent EMSA assays. For each concentration of SNRPA-his, the disappearance of unbound RNA was measured using the Image Studio Lite software. The error bars represent the standard deviation. (D) Identical to (C) except that the RNA used for EMSA was restricted to the G4 sequence.

4. Discussion

The large scale identification of proteins binding RNA G4s is in its early stages and, despite the fact that some studies have already made it possible to highlight certain proteins of interest, this study reports, for the first time, SNRPA as a potential G4 RBP (Haeusler et al., 2014; Herdy et al., 2018a; Herviou et al., 2020b; McRae et al., 2017b; Serikawa et al., 2018; von Hacht et al., 2014). RNA affinity purification followed by LC-MS/MS analysis revealed the enrichment of 37 potential protein candidates that seem to have an increased affinity for the G4 located in the 5'UTR of the BAG-1 mRNA, as compared to an RNA in which the folding of the G4 has been abolished. Interestingly, although an RNA molecule was used as bait, many chromatin binding proteins ended up being detected. A high representation of members of the SWI/SNI family of proteins that display both helicase and ATPase activities (ex; SMARCC1, SMARCC2, SMARCD2 and SMACA4) was observed. These proteins are thought to regulate the transcription of certain genes by altering the chromatin structure. It is not impossible that these proteins may have dual functions and be able to bind RNA during the transcription of their target genes, which could explain their presence in this experiment. Also, it is important to mention that in total extracts the cellular compartments are broken down during their preparation, a fact which may have favored the detection of so many of these DNA binding proteins. More relevant was the finding of several RBP, including RNA helicases like DHX36 and DDX21 that are

known to unfold RNA G4 (Gueddouda et al., 2017; McRae et al., 2020), which increases the confidence in these candidates and validates the approach. That said, 6 out of the 9 candidates tested by western blot were validated, including DHX36 (Fig. 1C). Among the RBP tested so far, some might be interesting candidates for the regulation of BAG-1 translation.

In order to explain the translational regulation which seems to be observed for BAG-1 in CRC cells, the potential role that these factors could have on this regulation was examined. Using a knock down approach for each candidate, a consistent dysregulation of BAG-1 at the translation level was detected solely for SNRPA. Dysregulation was observed at different levels for other RBPs, but nothing compared to that observed for SNRPA. This protein of 282 amino acids (~34 kD) contains 2 RNA recognition motifs (RRM) and is well-known for its implication in splicing (Price et al., 1998). In fact, SNRPA, through its N-terminal RRM, can bind the loop part of the SL2 sequence AUUGCAC of the U1 RNA (Klein Gunnewiek et al., 2000). This sequence is not part of the BAG-1 5'UTR. Moreover, SNRPA expression has been shown to be higher in tumour tissues than in matched normal gastric mucosa tissues (Dou et al., 2018). When SNRPA was knocked down in HCT-116 cells, an increase in BAG-1 protein expression (1.5-folds) was observed, while the RNA level remained relatively constant. Although mainly known as a splicing factor, SNRPA can shuttle between the nucleus and the cytoplasm (Kambach & Mattaj, 1992). That said, how this splicing factor can modulate the translation of BAG-1 remains an interesting question. Moreover, G4 are known to hinder translation by acting as a roadblock to the scanning of the ribosomes. In other words, they usually decrease translational efficiency. In the present case, the knockdown of SNRPA increases the expression of BAG-1 in HCT-116 cells, suggesting that SNRPA stabilized the G4 and decreased the expression level. As mentioned earlier, helicases like DHX36 and DDX21 are known to resolve G4 structures, and it is not impossible that such a protein can compete for binding the BAG-1 rG4. In this scenario the two would act together to fold/unfold this rG4 in order to modulate BAG-1 expression. Elucidation of this molecular mechanism might open the door to novel therapeutic approaches for CRC since the expression of nuclear BAG-1 is associated with a poorer prognosis and is potentially a useful predictive factor for distant metastasis

(Barnes et al., 2005). Thus, disruption of the nuclear localised BAG-1 would sensitize cells to apoptosis and increase the efficacy of current therapeutic approaches. As shown in this report, the absence of SNRPA increases the expression of the three BAG-1 isoforms, although at different levels. The presence of SNRPA seems to be favorable regarding cancer treatment outcome, by maintaining lower levels of the BAG-1 isoforms.

5. Conclusion

In brief, the present study demonstrated that SNRPA binds directly to the BAG-1 mRNA through the G-quadruplex, and knock down experiments in colorectal cancer cells suggested that it can modulate the expression level of BAG-1.

References

- [1] M. Malgowska, K. Czajczynska, D. Gudanis, A. Tworak, Z. Gdaniec, Overview of the RNA G-quadruplex structures, *Acta Biochim. Pol.* 63 (2016) 609–621. https://doi.org/10.18388/abp.2016_1335.
- [2] M.M. Fay, S.M. Lyons, P. Ivanov, RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms, *J. Mol. Biol.* 429 (2017) 2127–2147. <https://doi.org/10.1016/j.jmb.2017.05.017>.
- [3] C.K. Kwok, G. Marsico, A.B. Sahakyan, V.S. Chambers, S. Balasubramanian, rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome, *Nat. Methods.* 13 (2016) 841–844. <https://doi.org/10.1038/nmeth.3965>.
- [4] S. Millevoi, H. Moine, S. Vagner, G-quadruplexes in RNA biology, *Wiley Interdiscip Rev RNA.* 3 (2012) 495–507. <https://doi.org/10.1002/wrna.1113>.
- [5] S.G. Rouleau, J.-M. Garant, F. Bolduc, M. Bisailon, J.-P. Perreault, G-Quadruplexes influence pri-microRNA processing, *RNA Biol.* 15 (2018) 198–206. <https://doi.org/10.1080/15476286.2017.1405211>.
- [6] A. Cammas, S. Millevoi, RNA G-quadruplexes: emerging mechanisms in disease, *Nucleic Acids Res.* 45 (2017) 1584–1595. <https://doi.org/10.1093/nar/gkw1280>.
- [7] N.-F. Sun, Q.-Y. Meng, S.-Y. Hu, A.-L. Tian, R.-H. Wang, Z.-X. Liu, L. Xu, Correlation between the expression of the BAG-1 gene and clinicopathologic factors in colorectal cancer, *J. Cancer Res. Clin. Oncol.* 137 (2011) 1419–1424. <https://doi.org/10.1007/s00432-011-1011-3>.
- [8] X. Yang, G. Chernenko, Y. Hao, Z. Ding, M.M. Pater, A. Pater, S.C. Tang, Human BAG-1/RAP46 protein is generated as four isoforms by alternative translation initiation and overexpressed in cancer cells, *Oncogene.* 17 (1998) 981–989. <https://doi.org/10.1038/sj.onc.1202032>.
- [9] M.J. Coldwell, M.L. deSchoolmeester, G.A. Fraser, B.M. Pickering, G. Packham, A.E. Willis, The p36 isoform of BAG-1 is translated by internal ribosome entry following heat shock, *Oncogene.* 20 (2001) 4095–4100. <https://doi.org/10.1038/sj.onc.1204547>.
- [10] R. Jodoin, J.C. Carrier, N. Rivard, M. Bisailon, J.-P. Perreault, G-quadruplex located in the 5'UTR of the BAG-1 mRNA affects both its cap-dependent and cap-independent translation through global secondary structure maintenance, *Nucleic Acids Res.* (2019). <https://doi.org/10.1093/nar/gkz777>.
- [11] R. Jodoin, J.-P. Perreault, G-quadruplexes formation in the 5'UTRs of mRNAs associated with colorectal cancer pathways, *PloS One.* 13 (2018) e0208363. <https://doi.org/10.1371/journal.pone.0208363>.
- [12] J.U. Guo, D.P. Bartel, RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria, *Science.* 353 (2016). <https://doi.org/10.1126/science.aaf5371>.
- [13] J. Cox, M.Y. Hein, C.A. Lubner, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, *Mol. Cell Proteomics.* 13 (2014) 2513–2526. <https://doi.org/10.1074/mcp.M113.031591>.

- [14] K. Leppek, G. Stoecklin, An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins, *Nucleic Acids Res.* 42 (2014) e13. <https://doi.org/10.1093/nar/gkt956>.
- [15] J.M. Nicoludis, S.P. Barrett, J.-L. Mergny, L.A. Yatsunyk, Interaction of human telomeric DNA with N-methyl mesoporphyrin IX, *Nucleic Acids Res.* 40 (2012) 5432–5447. <https://doi.org/10.1093/nar/gks152>.
- [16] J.M. Klein Gunnewiek, R.I. Hussein, Y. van Aarssen, D. Palacios, R. de Jong, W.J. van Venrooij, S.I. Gunderson, Fourteen residues of the U1 snRNP-specific U1A protein are required for homodimerization, cooperative RNA binding, and inhibition of polyadenylation, *Mol. Cell. Biol.* 20 (2000) 2209–2217. <https://doi.org/10.1128/mcb.20.6.2209-2217.2000>.
- [17] A.R. Haeusler, C.J. Donnelly, G. Periz, E.A.J. Simko, P.G. Shaw, M.-S. Kim, N.J. Maragakis, J.C. Troncoso, A. Pandey, R. Sattler, J.D. Rothstein, J. Wang, C9orf72 nucleotide repeat structures initiate molecular cascades of disease, *Nature.* 507 (2014) 195–200. <https://doi.org/10.1038/nature13124>.
- [18] T. Serikawa, C. Spanos, A. von Hacht, N. Budisa, J. Rappsilber, J. Kurreck, Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs, *Biochimie.* 144 (2018) 169–184. <https://doi.org/10.1016/j.biochi.2017.11.003>.
- [19] E.K.S. McRae, E.P. Booy, A. Moya-Torres, P. Ezzati, J. Stetefeld, S.A. McKenna, Human DDX21 binds and unwinds RNA guanine quadruplexes, *Nucleic Acids Res.* 45 (2017) 6656–6668. <https://doi.org/10.1093/nar/gkx380>.
- [20] B. Herdy, C. Mayer, D. Varshney, G. Marsico, P. Murat, C. Taylor, C. D'Santos, D. Tannahill, S. Balasubramanian, Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts, *Nucleic Acids Res.* 46 (2018) 11592–11604. <https://doi.org/10.1093/nar/gky861>.
- [21] A. von Hacht, O. Seifert, M. Menger, T. Schütze, A. Arora, Z. Konthur, P. Neubauer, A. Wagner, C. Weise, J. Kurreck, Identification and characterization of RNA guanine-quadruplex binding proteins, *Nucleic Acids Res.* 42 (2014) 6630–6644. <https://doi.org/10.1093/nar/gku290>.
- [22] P. Herviou, M. Le Bras, L. Dumas, C. Hieblot, J. Gilhodes, G. Cioci, J.-P. Hugnot, A. Ameadan, F. Guillonnet, E. Dassi, A. Cammas, S. Millevoi, hnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma, *Nat Commun.* 11 (2020) 2661. <https://doi.org/10.1038/s41467-020-16168-x>.
- [23] E.K.S. McRae, S.J. Dupas, E.P. Booy, R.S. Piragasam, R.P. Fahlman, S.A. McKenna, An RNA guanine quadruplex regulated pathway to TRAIL-sensitization by DDX21, *RNA.* 26 (2020) 44–57. <https://doi.org/10.1261/rna.072199.119>.
- [24] N.M. Gueddouda, O. Mendoza, D. Gomez, A. Bourdoncle, J.-L. Mergny, G-quadruplexes unfolding by RHAU helicase, *Biochim Biophys Acta Gen Subj.* 1861 (2017) 1382–1388. <https://doi.org/10.1016/j.bbagen.2017.01.006>.
- [25] S.R. Price, P.R. Evans, K. Nagai, Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA, *Nature.* 394 (1998) 645–650. <https://doi.org/10.1038/29234>.

- [26] N. Dou, D. Yang, S. Yu, B. Wu, Y. Gao, Y. Li, SNRPA enhances tumour cell growth in gastric cancer through modulating NGF expression, *Cell Prolif.* 51 (2018) e12484. <https://doi.org/10.1111/cpr.12484>.
- [27] C. Kambach, I.W. Mattaj, Intracellular distribution of the U1A protein depends on active transport and nuclear binding to U1 snRNA, *J. Cell Biol.* 118 (1992) 11–21. <https://doi.org/10.1083/jcb.118.1.11>.
- [28] J.D. Barnes, N.J. Arhel, S.S. Lee, A. Sharp, M. Al-Okail, G. Packham, A. Hague, C. Paraskeva, A.C. Williams, Nuclear BAG-1 expression inhibits apoptosis in colorectal adenoma-derived epithelial cells, *Apoptosis.* 10 (2005) 301–311. <https://doi.org/10.1007/s10495-005-0804-8>.

Author's contribution

FB: Designed and performed the experiments, analyzed the data, wrote the manuscript.

MAT: Performed the cloning and participated in the setting up of the RNA affinity purification assays. JPP: Conceived the idea, wrote and edited the manuscript.

Funding

This work was supported by funding from the Natural Sciences and Engineering Research Council of Canada (RGPIN-155219-17) to JPP, who is a member of the Centre de Recherche du CHUS.

Acknowledgements

We thank Katerie Bellemare for technical assistance.

SUPPLEMENTARY INFORMATION**The Small Nuclear Ribonucleoprotein Polypeptide A (SNRPA) binds to the G-quadruplex of the BAG-1 5'UTR**

François Bolduc, Marc-Antoine Turcotte, Jean-Pierre Perreault

SUPPLEMENTARY METHODS

***In-line* probing and N-methyl mesoporphyrin fluorescence assays**

The details of these assays were performed as described previously (Jodoin & Perreault, 2018).

Label-free RNA affinity purification

Two RNAs were synthesized by *in vitro* transcription from the plasmids pMARQ_BAG1G4_S1m4X and pMARQ_BAG1G/Amut_S1m4X: one contained the first 83 nts of the 5'UTR of BAG-1 attached to four copies of the S1m aptamer and the second contained a mutant version where 7 guanosines were mutated to adenosines (see Fig. 2C for specific sequences). The plasmids were digested at both the *EcoRI* site located upstream of the T7 promoter and the *HindIII* site located downstream of the 4 S1m sequence in order to release the T7-tagged DNA templates that were then purified by preparative agarose gel electrophoresis. The DNA was extracted from the gel fragments using the QIAquick Gel Extraction Kit (QIAGEN). The purified DNA templates were then used in run-off *in vitro* transcription reactions using T7 RNA polymerase, and the synthesized RNAs were gel purified on a 5% denaturing polyacrylamide gel as described previously (Jodoin & Perreault, 2018). The RNAs were dissolved in water to a final concentration of 1 µg/µL and were stored at -20°C. For all of the following procedures, low protein binding tubes (Sarstedt) were used in order to reduce the degree of unspecific binding of proteins. Protein extracts from HCT-116 cells were prepared from 4 confluent 15-cm dishes. The cell culture dishes were placed on ice and washed twice with 10 mL of ice-cold PBS. Then, 1 mL of ice-cold 1X KCl-lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM KCl, 50 mM NaCl, 2 mM MgCl₂, 1 mM DTT, 0.1% Triton X-100, 10% glycerol and 2 mM vanadylribonucleoside complex RNase inhibitor (NEB) which contained 1 tablet of the EDTA-free Mini Complete Protease Inhibitors (Roche) per 10 mL of buffer) was added to each culture dish. The adherent cells were scraped off the dish using a cold plastic cell scraper and were gently transferred to a pre-cooled (at 4°C) 15-mL tube. Constant agitation was maintained for 30 min at 4°C followed by centrifugation at 16 000 Xg at 4°C for 20 min. The supernatant was removed and divided into 2 fractions which were then concentrated using two Amicon Ultra-4 (MWCO = 3 000)

centrifuge tubes at 4 000 Xg until final volume was 0.5-1.0 mL in each tube was obtained. The extracts were then combined (an aliquot representing 10% of the total volume was taken at this stage and was labeled "Input") and the protein concentration was measured using the Bradford method. The extract (1.0-2.0 mL) was pre-cleared by the addition of 100 μ L of a 50% slurry of Streptavidin Sepharose High Performance (GE Healthcare) beads that had previously been washed with 1X KCl-lysis buffer (3 x 500 μ L) and incubating for an additional 2–3 h at 4°C. The beads were discarded following centrifugation at 0,02 Xg for 1 min at 4°C. To 30 μ g (30 μ L) of synthesized RNAs (BAG1G4_S1m4X and BAG1G/Amut_S1m4X), 30 μ L of 2X KCL lysis buffer (without magnesium) was added. The samples were incubated at 70°C for 5 min, followed by a slow cool to room temperature so as to permit refolding of the RNA structures (~1h). After slow cooling, 2 μ L of 120 mM MgCl₂ were added for a final concentration of 2 mM. Per sample, 100 μ L a 50% slurry of the streptavidin sepharose beads were washed three times with 1 mL of 1X KCl-lysis buffer. At each step, the beads were gently pelleted at 500 rpm for 1 min at 4°C. The refolded RNA was then added to the beads and the samples were incubated at 4°C for 2–3 h with rotation so as to allow for binding of the RNA. The beads were then sedimented and washed three times with 1 mL of 1X KCL-lysis buffer. The pre-cleared lysates were then added to the freshly prepared RNA-coupled beads and the samples were incubated at 4°C for 3–4 h with rotation. The beads were then washed 6 times for 2-5 min with 1 mL of 1X KCl wash buffer (50 mM Tris-HCl pH 7.5, 100 mM KCl, 200 mM NaCl, 5 mM MgCl₂, 1mM DTT, 0.1% Triton X-100, 10% glycerol and 2 mM vanadylribonucleoside complex RNase inhibitor (NEB) which contained 1 tablet EDTA-free Mini Complete Protease Inhibitors (Roche) per 10 mL of buffer). Lastly, the beads were then prepared for either the LC-MS/MS analysis or the western blotting experiments.

Samples preparation and LC-MS/MS analysis

Sample preparation included the reduction and the alkylation of the proteins followed by trypsin digestion. Digested peptides were then purified on C18 columns. All these steps were performed as described previously (Lambert et al., 2020). The LC-MS/MS analysis was also performed as described (Lambert et al., 2020).

Protein identification by MaxQuant analysis

The raw files from LC-MS/MS were analyzed using the MaxQuant version 1.6.2.2 software and the Uniprot human database (16/07/2013, 88 354 entries). MaxQuant supports the label-free quantification of proteins (Cox et al., 2014). The settings used for the MaxQuant analysis were: i) 2 miscleavages were allowed; ii) enzymes were Trypsin (K/R not before P); and, iii) variable modifications included in the analysis were methionine oxidation and protein N-terminal acetylation. Mass tolerances of 7 ppm for precursor ions and 20 ppm for fragment ions were used. In addition, the following parameters were used: label-free quantification with an LFQ minimum ratio count of 2; identification values “PSM FDR”; “Protein FDR” and “Site decoy fraction” 0.05; and, “Match between runs”. Proteins positive for at least one of the “Reverse” and “Potential contaminant” categories were eliminated. Label-free quantification (LFQ) intensities were used to determine the fold enrichment in WT vs G/A mutant samples for each protein identified. Proteins with a coverage >5 %, a fold enrichment of >2.5 and a MaxQuant score >15 were conserved for subsequent analysis.

Western blots

The interactions of potential RNA binding proteins (RBP) were initially verified by western blot analysis using washed beads that were incubated in 60 μ L of 2X Laemmli loading buffer (125 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 5% β -mercaptoethanol, 0.004% bromophenol blue) in order to elute the bound proteins. The protein concentrations of the input samples were determined by the Bradford method and an equal volume of 2X Laemmli buffer was added. The samples were incubated at 95°C for 3 min before being loaded on a 10% SDS-PAGE. About 30 μ g of protein from the input samples and 10 μ L of the bound proteins were analysed per gel. The proteins were then transferred to nitrocellulose membranes and the membranes were then blocked by incubating in TBST (TBS, 0.1% Tween-20) supplemented with 5% milk for 1 h at room temperature. The primary antibodies (anti-DHX36 (NB200-353, Novus Biologicals), anti-DDX21 (sc-376953, Santa Cruz), anti-MSI2 (NBP2-19443, Novus Biologicals), anti-RBM4 (NBP2-46671, Novus Biologicals), anti-YBX1 (sc-398340, Santa Cruz), anti-SRPK2 (sc-390534, Santa Cruz), anti-SNRPA (sc-

101149, Santa Cruz), anti-GNL1 (sc-514362, Santa Cruz), anti-GRSF1 (ab205531, Abcam), anti-BAG-1 (sc-33704, Santa Cruz) and anti-GAPDH (G9545, Sigma)) were diluted according to the manufacturers' protocols in TBST supplemented with 1% milk. The incubations with the primary antibodies were conducted for 16 h at 4°C, except for anti-GAPDH which was incubated for 1 h at room temperature. After 3 successive washes for 15 min each using TBS-T, the membranes were incubated 1 h at room temperature with the secondary antibodies (IRDye 800CW donkey anti-rabbit IgG (926-32213, Li-Cor) or 690 Alexa Fluor goat anti-mouse IgG (A21057, Thermo Fisher)). The secondary antibodies were diluted 1/10 000 in TBST supplemented with 1% milk. Finally, the membranes were washed 3 x 15 min in TBST and once in PBS before protein detection using an Odyssey infrared imaging system (Li-Cor).

Gene knock down using siRNAs in HCT-116 cells

HCT-116 cells were seeded at a concentration of 200 000 cells per well in 6-well plates. For each well, siRNAs (30 pmoles) and Lipofectamine (7.5 μ L, Lipofectamine RNAi/MAX, Thermo Fisher) were mixed per transfection. Silencer Select siRNAs targeting SNRPA (ID# s13214 and ID# s13214) and Silencer Select Negative Control siRNA #1 (4390843, Life Technologies) were used. The latter is designed with the latest information about miRNA seed regions and sequence alignment algorithms in order to minimize any interactions with mRNA transcripts of the human transcriptome. At both 48 and 72 h post-transfection, the cells were recovered and total lysates were prepared. For each well, the culture media was removed and the cells were washed twice with 1 mL PBS. The cells were then scraped off of the dish using a cold plastic cell scraper in 1mL PBS and were then split into two tubes (roughly 200 μ L were used for RNA detection and 800 μ L for protein detection). The cells were harvested by centrifugation for 5 min at 0,6 Xg. The supernatants were removed and the RNA was then extracted using Qiazol (Qiagen) as described by manufacturer. The RNA was then treated with RNase-free DNase RQ1 (Promega) followed by extraction with phenol-chloroform and ethanol precipitation. The RNA pellets were dissolved in high pure water. The RNA concentration was determined using a NanoDrop Lite apparatus. The RNA was analyzed by RT-qPCR at the Laboratory of Functional Genomics of the Université de Sherbrooke. The pellets for the protein samples were treated with 40 μ L of 1X KCl-lysis

buffer and maintained under constant agitation for 30 min at 4°C followed by centrifugation at 16 000 Xg at 4°C for 20 min. The supernatants were recovered and protein quantification was performed using Bradford method. Finally, an equal volume of 2X Laemmli buffer was added. The proteins were then separated by SDS-PAGE, transferred to nitrocellulose and specific proteins detected by western blotting (see above). The analyses were performed three times on three different days. Each time, duplicate assays were performed and the mean values of these two replicates were used in the analyses. The intensity of each band (i.e. BAG-1 and GAPDH) was measured using the Studio Lite software. The intensities of the BAG-1 bands were normalized against those of the GAPDH bands for loading control purposes. The normalized intensities of BAG-1 bands in the control siRNA assays were then fixed at 100%. Finally, the normalized intensities of the BAG-1 bands in both the siSNRPA-1 and siSNRPA-2 assays were compared to the normalized intensities of the BAG-1 bands detected in the control siRNA assay.

Recombinant SNRPA production and purification

E.coli Rosetta-2 cells transformed with the plasmid pET42a_SNRPA-cHis (GeneScript) were grown to $OD_{600} = 0.5$ in LB medium containing 34 µg/mL of chloramphenicol and 50 µg/mL of kanamycin and were then induced for 3 h at 37°C by the addition of IPTG to a final concentration of 0.4 mM. Protein extracts were prepared in lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM imidazole, 10% glycerol, 0.1% Triton X-100 and 100 µg/mL lysozyme) and were incubated on ice for 30 min. Following sonication on ice (24 cycles of 5 sec burst followed by 5 sec off), the extracts were centrifuged 45 min at 10 000 Xg at 4°C. The supernatants were then loaded on Ni-NTA columns. The bound proteins were washed with lysis buffer, then with wash buffer (50 mM Tris-HCl pH 8.0, 1.5 M NaCl, 10 mM imidazole and 10% glycerol) and finally with lysis buffer that lacked both Triton X-100 and lysozyme. The proteins of interest were eluted with successive passages of elution buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl and 10% glycerol) containing increasing concentrations of imidazole (50, 100, 500 and 1000 mM). The proteins eluted in buffer containing 100 mM imidazole were kept and were dialyzed against storage buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol and 2 mM DTT) for 16 h at 4°C. The purity of the SNRPA-his was evaluated by SDS-PAGE.

Electrophoretic mobility shift assays (EMSA)

Two RNAs of 83 nts in length containing the first nucleotides of the 5'UTR of BAG-1 were *in vitro* transcribed. The PCR templates for the *in vitro* transcriptions were obtained by the PCR filling of 2 complementary oligonucleotides, one consisting of the T7 promoter sequence (5'- TAATACGACTCACTATA-3') and the other containing the sequence of interest (either WT: 5'-TTGTTGACCGCCCAGCGATGGAAGCTGAGCGCGGCGTCTC ACAACCCCGCCCGACTACTTCCCAGCCCCGCCCGGCCTGCCCTATAGTGAGTC GTATTA-3' or G/A mutant:5'TTGTTGACCGCCCAGCGATGGAAGCTGAGCGCGGC GTCTCACAACCTCGCTCGACTACTTCTCAGTCTCGTCTCGGCCTGCCCTATAGTG AGTCGTATTA3'), using purified PFU DNA polymerase (12 cycles of 1 min each at 95°C, 54°C and 72°C, followed by a final elongation at 72°C for 5 min) in buffer containing 0.2 mM dNTPs, 2 mM MgSO₄, 20 mM Tris-HCl pH 8.8, 10 mM KCl, 10 mM (NH₄)₂SO₄, 0.1% Triton-X-100 and 5 % DMSO. The DNA template sizes were verified by agarose gel electrophoresis. The DNA templates were then ethanol precipitated, dried and dissolved in 50 µL H₂O. *In vitro* transcription reactions were carried out as described previously (Jodoin & Perreault, 2018). The 2 short 30 nts long RNAs (containing the G4 sequence of the mutated version without the surrounding nucleotides) were purchased from Integrated DNA Technology. The RNAs were dephosphorylated and 5' labeled as described previously (Jodoin & Perreault, 2018). For each EMSA reaction, the RNA (1 nM corresponding to about 500 counts per minute (CPM)) was first refolded in refolding buffer (20 mM Tris-HCl pH 8.0, 100 mM KCl) by incubating at 70°C for 3 min followed by a slow cool to room temperature over 1 h. Binding buffer (20 mM Tris-HCl pH 8.0, 90 mM KCl, 1 mM MgCl₂, 5% glycerol) supplemented with 0.3U/µL RNase inhibitor, 0.2mg/ml BSA and 0.13 µg/µL yeast total RNA was then added. Finally, recombinant SNRPA at increasing concentrations from 0 to 45 µM was added, all in a final volume of 15 µL. The mixtures were incubated at room temperature for 20 min before being loaded on a 5% acrylamide TBE native gel which was electrophoresed at 20 V/cm. The K_d values were calculated following analysis of the intensities of the free RNA from two independent assays. The disappearance of the free RNA was measured using the Image Studio Lite software. The measure intensities were then analyzed with the GraphPad Prism software using the "specific binding with Hill slope" function.

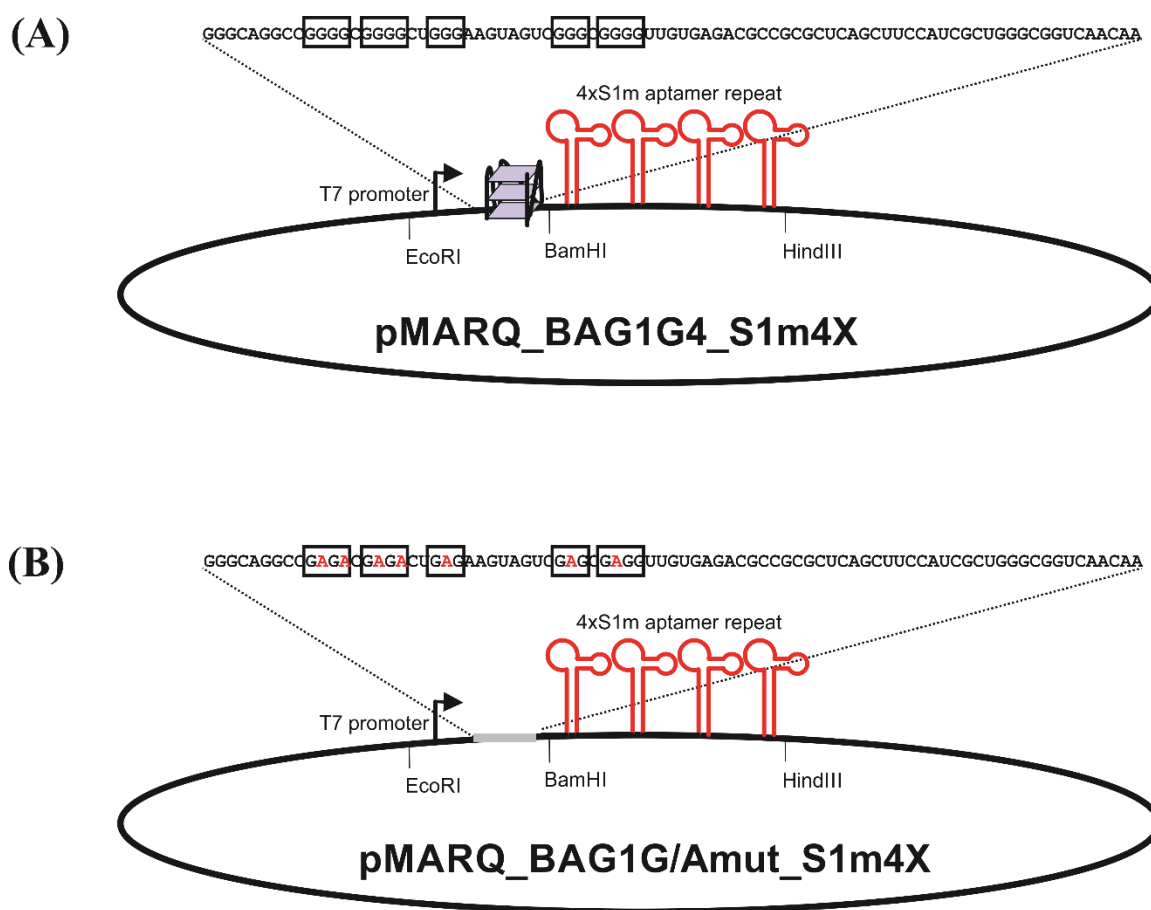
SUPPLEMENTARY TABLE

Supplementary Table S1. Candidate proteins and enrichment for each replicate

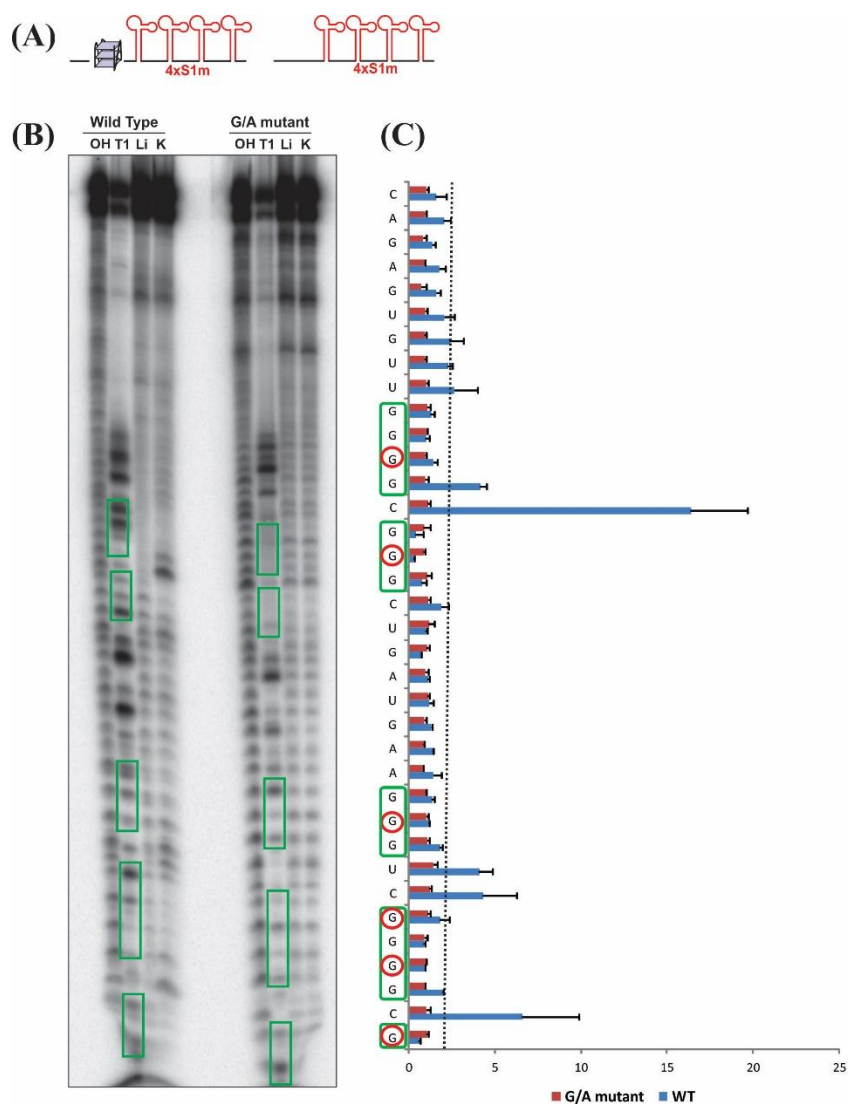
Protein ID	Protein class	Fold enrichment		
		Replicate 1	Replicate 2	Replicate 3
ACTR5	Chromatin binding	0,0	20,0	4,3
GATAD2A	Chromatin binding	20,0	0,0	6,0
GATAD2B	Chromatin binding	0,0	20,0	5,3
MTA1	Chromatin binding	0,0	3,3	14,0
MTA2	Chromatin binding	20,0	20,0	12,0
RBBP4	Chromatin binding	0,0	3,4	4,9
RBBP7	Chromatin binding	3,2	2,6	4,9
SMARCC1	Chromatin binding	3,5	3,6	3,8
SMARCC2	Chromatin binding	20,0	0,0	4,9
SMARCD2	Chromatin binding	20,0	20,0	3,7
GNL1	GTP binding	9,2	5,4	6,4
GTPBP1	GTP binding	4,3	3,0	0,0
GTPBP8	GTP binding	20,0	20,0	0,0
CHD3	Helicase DNA	0,0	20,0	3,0
CHD4	Helicase DNA	3,0	0,0	4,2
INO80	Helicase DNA	0,0	20,0	9,2
SMARCA4	Helicase DNA	20,0	3,2	7,0
DDX21	Helicase RNA	0,0	2,7	3,2
DHX36	Helicase RNA	4,4	2,8	2,8
DHX37	Helicase RNA	4,7	0,0	2,7
HDAC1	Histone deacetylase	0,0	2,7	5,4
TAOK1	Kinase	20,0	4,6	3,9
TAOK2	Kinase	0,0	4,5	11,0
PUS7	Pseudouridine synthase	8,3	2,9	0,0
GRSF1	RNA binding G-Rich	3,2	2,8	0,0
TES	Scaffold	20,0	20,0	0,0
PRPF40B	Splicing	0,0	20,0	3,4
RBM4	Splicing	4,4	0,0	4,4
SNRPA	Splicing	3,2	0,0	3,2

SNRPD1	Splicing	3,5	0,0	2,9
SRPK2	Splicing	20,0	0,0	3,9
YBX1	Splicing	2,5	0,0	5,1
GTF3C5	Transcription factor	0,0	20,0	2,6
PTRF	Transcription factor	2,9	0,0	3,2
YBX3	Translation	3,5	3,5	7,7
MSI2	Translation	7,3	3,2	20,0
RBM4B	Translation	4,4	0,0	4,4

SUPPLEMENTARY FIGURES

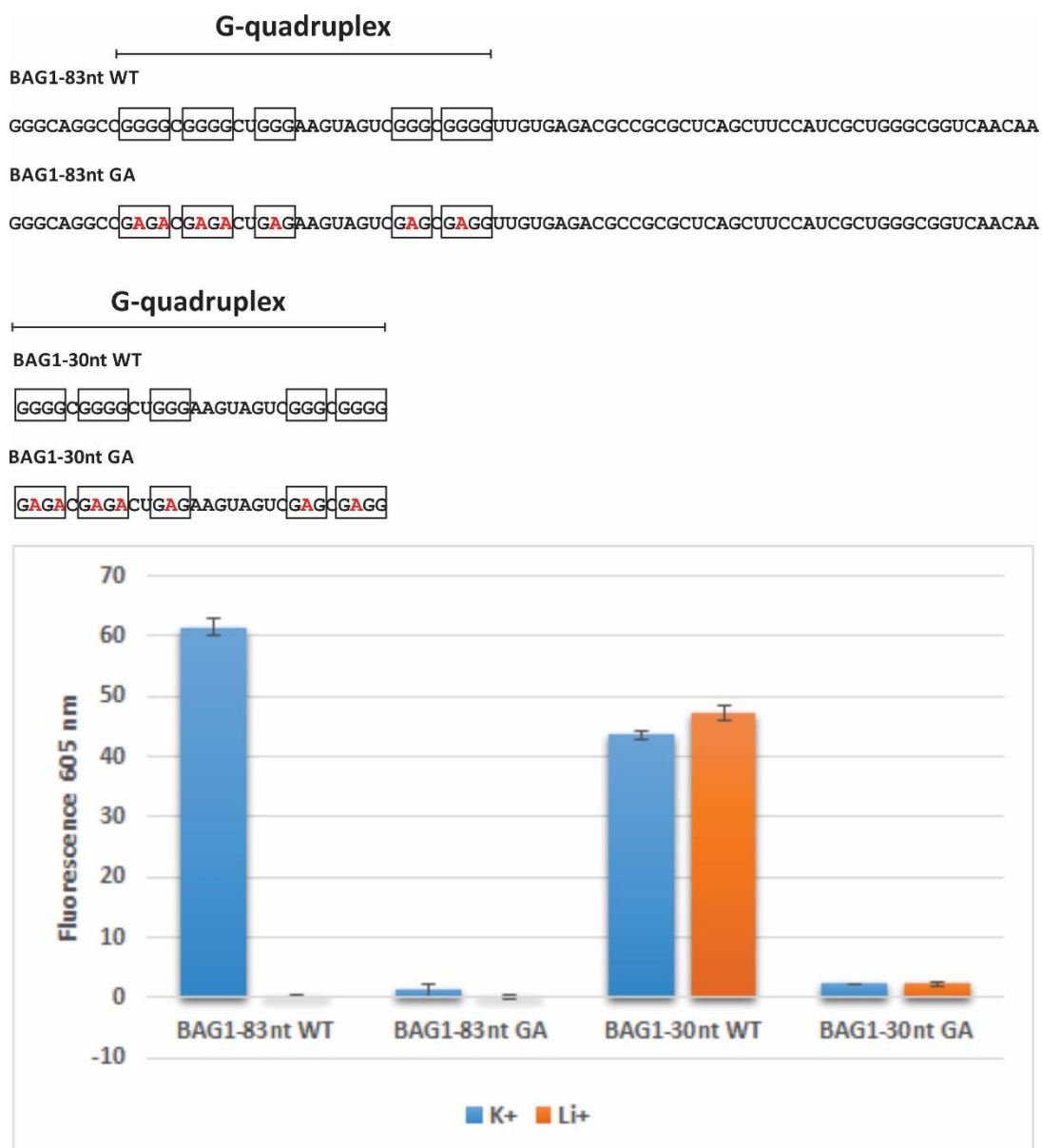


Supplementary Fig. S1. Nucleotide Sequences of the wild type and G/A mutant constructs of BAG-1. Chimeric RNA sequences (WT and G/A mutant) containing the first 83 nts of the BAG-1 5'UTR fused to four repeats of the S1m aptamer (in red). The boxed nucleotides indicate the nucleotides involved in the folding of the G-quadruplex, while the red nucleotides are the G that were mutated to A. (A) pMARQ_BAG1G4_S1m4X, and (B) pMARQ_BAG1G/Amut_S1m4X.



Supplementary Fig. S2. *In-line* probing of the chimeric RNAs used in the RNA affinity purification assays. (A) Chimeric RNA sequences (WT and G/A mutant) containing the first 83 nts of the BAG-1 5'UTR fused to four repeats of the S1m aptamer (in red). (B) Denaturing polyacrylamide gel (10%) of the 5'-radiolabelled chimeric RNA after *in-line* probing. The alkaline hydrolysis (OH) and RNase T1 ladders (T1) indicate the positions of every nucleotide and every guanine, respectively. The *in-line* probing was performed in the presence of either Li⁺ (Li) or K⁺ ions (K). The green boxed nucleotides indicate the guanines involved in the folding of the G-quadruplex. (C) Histograms showing the K⁺/Li⁺ quantification of the *in-line* probing band intensities for each nucleotide for both the WT sequence (blue) and the G/A-mutant sequence (red). Each bar represents the mean of 2 independent experiments, and the error bars represent the standard deviations. The K⁺/Li⁺

cleavage ratio threshold of 2 is indicated by the dotted line. The circled red guanines are those mutated to adenines in the mutant version.



Supplementary Fig. S3. N-methyl mesoporphyrin (NMM) fluorescence assays. The sequences of the 4 *in vitro* transcribed RNAs used in the EMSA are shown at the top of the figure. The bottom part of the figure shows the histograms of the fluorescence emission peaks at 605 nm under the different conditions tested: Orange, Li⁺ ; and, Blue, K⁺ . Each bar represents the mean of 2 independent experiments, and the error bars represent the standard deviations.

References

- [1] R. Jodoin, J.-P. Perreault, G-quadruplexes formation in the 5'UTRs of mRNAs associated with colorectal cancer pathways, *PloS One*. 13 (2018) e0208363. <https://doi.org/10.1371/journal.pone.0208363>.
- [2] É. Lambert, J.-P. Babeu, J. Simoneau, J. Raisch, L. Lavergne, D. Lévesque, É. Jolibois, M. Avino, M.S. Scott, F. Boudreau, F.-M. Boisvert, Human Hepatocyte Nuclear Factor 4- α encodes isoforms with distinct transcriptional functions, *Mol. Cell Proteomics*. (2020). <https://doi.org/10.1074/mcp.RA119.001909>.
- [3] J. Cox, M.Y. Hein, C.A. Lubner, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, *Mol. Cell Proteomics*. 13 (2014) 2513–2526. <https://doi.org/10.1074/mcp.M113.031591>.