



HAL
open science

Etude du rôle des facteurs d'épissage à domaines UHM dans la régulation de l'épissage alternatif

Manel Tari

► To cite this version:

Manel Tari. Etude du rôle des facteurs d'épissage à domaines UHM dans la régulation de l'épissage alternatif. Médecine humaine et pathologie. Université Paris Saclay (COMUE), 2018. Français. NNT : 2018SACLE041 . tel-03051836

HAL Id: tel-03051836

<https://theses.hal.science/tel-03051836>

Submitted on 10 Dec 2020

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

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

Etude du rôle des facteurs d'épissage à domaines UHM dans la régulation de l'épissage alternatif

Thèse de doctorat de l'Université Paris-Saclay
Préparée à l'Université d'Evry Val d'Essonne

École doctorale n°577 Structure et Dynamique des Systèmes
Vivants (SDSV)
Spécialité de doctorat : Sciences de la vie et de la santé

Thèse présentée et soutenue à Evry, le 17 Décembre 2018, par

Manel TARI

Composition du Jury :

Dr. Isabelle Behm-Ansmant Chargée de recherche, Université de Lorraine	Présidente
Dr. Eric Allemand Chargé de recherche, Institut Pasteur	Rapporteur
Dr. Marc-Henri Stern Directeur de recherche, Institut Curie	Rapporteur
Sandrine Baghdoyan Ingénieure de recherche, IStem	Examinatrice
Pr. David PASTRE Professeur, Université d'Evry Val d'Essonne	Directeur de thèse
Dr. Alexandre MAUCUER Chargé de recherche, Inserm	Co-Directeur de thèse

Table des matières

Table des matières	3
Liste des abréviations et acronyme	7
Liste des figures	9
Liste des tableaux	12
I. Etat de l'art	13
Chapitre I : L'épissage des pré-ARN messagers	13
1. <i>La réaction d'épissage</i>	15
2. <i>Les séquences en cis nécessaires à l'identification des introns</i>	16
3. <i>La réaction d'épissage par le spliceosome majeur</i>	17
3.1. Les UsnRNP	17
3.2. Les protéines SR	19
3.3. Les étapes d'assemblage du spliceosome	23
3.4. Complexité de la composition du spliceosome	28
Chapitre II : L'épissage alternatif	31
1. <i>Types d'épissages alternatifs</i>	32
2. <i>Régulation de l'épissage alternatif</i>	34
2.1. Les séquences régulatrices en cis	34
2.2. Les protéines majeures de la régulation de l'épissage	38
2.3. Autres facteurs impliqués dans la régulation de l'épissage	47
Chapitre III : Les facteurs d'épissage à domaines UHM	52
1. <i>Les interactions UHM-ULM</i>	53
1.1. Domaine RRM, domaine UHM, similitudes et différences	53
1.2. Interaction UHM-ULM : découverte et structure canonique	54
1.4. Mise en évidence de protéines intervenant dans des interactions UHM-ULM	56
2. <i>Les facteurs d'épissage à domaines UHM (UHMSF)</i>	62
2.1. La protéine U2AF ⁶⁵	63
2.2. La protéine CAPER α	67
2.3. La protéine PUF60	70
2.4. La protéine SPF45	72
II. Objectifs	77
III. Matériel et Méthodes	81
1. Matériel	81
1.1. <i>Souches et plasmides utilisés</i>	81
a) Souches bactériennes utilisées	81
b) Cellules d'organismes eucaryotes utilisées	81
c) Plasmides utilisés	81

1.2. Oligodésoxyribonucléotides amorces et shRNA utilisés	84
1.3. Anticorps utilisés	87
1.4. Tampons, solutions et milieux de culture utilisés	88
2. Méthodes	89
2.1. Méthodes de biologie moléculaire	89
2.1.1. Préparation de constructions de plasmides par la méthode de RF-cloning	89
2.1.2. Purification de protéines recombinantes	94
2.1.3. Transcription des ARN <i>in vitro</i>	96
2.2. Méthodes de biologie cellulaire	96
2.2.1. Culture de cellules	96
2.2.2. Préparation d'extraits à partir de cellules	97
2.2.3. Préparation de surnageants viraux de shRNA	98
2.2.4. Transduction de cellules par les surnageants viraux de shRNA	99
2.2.5. Transfection de cellules	100
2.2.6. Préparation d'ARN totaux	100
2.3. Etude des complexes formés par les facteurs d'épissage à domaines UHM et avec SF3b155.	101
2.3.1. Etude des associations entre les facteurs d'épissage à domaines UHM et avec SF3b155 dans la cellule par Proximity Ligation Assay	101
2.3.2. Etude des interactions directes entre les facteurs d'épissage à domaines UHM et SF3B155 par GST-Pull Down	104
2.4. Etude du rôle des facteurs d'épissage à domaines UHM dans l'épissage alternatif	106
2.4.1. Etude de l'effet de la réduction du taux des facteurs d'épissage à domaines UHM sur 65 exons cassette par le système BioMark HD	106
2.4.1. Etude de l'effet de la réduction du taux de CAPER α et U2AF ⁶⁵ sur des minigènes rapporteurs d'épissage par qRT-PCR	108
2.5. Etude des assemblages macromoléculaires formés par U2AF ⁶⁵ et CAPER α avec l'ARN et le domaine multi-ULM de SF3b155	109
2.5.1. Etude des assemblages macromoléculaires formés par U2AF ⁶⁵ et CAPER α avec l'ARN et le domaine multi-ULM de SF3b155 par des expériences de sédimentation	109
IV. Résultats	112
1. Etude biochimique et en immunofluoromicroscopie des complexes formés par les facteurs d'épissage à domaines UHM avec SF3b155	112
1.1. Constat initial	112
1.2. Les facteurs d'épissage à domaines UHM sont colocalisés dans les noyaux de cellules de mammifères	113
1.3. Les facteurs d'épissage à domaines UHM interagissent de manière directe avec SF3b155	121
1.4. Coopérativité de CAPER α et U2AF ⁶⁵ pour les interactions avec SF3b155 dépendante des interactions UHM-ULM	124
2. Etude fonctionnelle du rôle des facteurs d'épissage à domaines UHM dans l'épissage alternatif.	133
2.1. Constat initial	133
2.2. Sélection des événements d'épissage alternatif étudiés	134
2.3. Etude de l'influence d'un déficit en facteurs d'épissages à domaines UHM sur l'épissage de 65 exons cassette	137

3. Etude des assemblages macromoléculaires formés par U2AF⁶⁵ et CAPERα autour du domaine multi-ULM de SF3b155	151
3.1. <i>Constat initial</i>	151
3.2. <i>Formation des assemblages macromoléculaires de CAPERα et U2AF⁶⁵ avec le domaine multi-ULM de SF3b155 dépendant de la composition de l'ARN en enchainement de polypyrimidine</i>	152
V. Discussion, perspectives et conclusion	160
1. Interactions des facteurs d'épissage à domaines UHM avec SF3b155	161
2. U2AF⁶⁵ renforce le recrutement de CAPERα lors des interactions avec SF3b155	161
3. Rôle des facteurs d'épissage à domaines UHM dans l'épissage des exons cassette	162
4. Modèle d'assemblage des facteurs d'épissage à domaines UHM autour de SF3b155 lors de la reconnaissance des sites 3' d'épissage	163
5. Les facteurs d'épissage à domaines UHM : enjeux et stratégies thérapeutiques ?	166
VI. Références bibliographiques	169

Liste des abréviations et acronyme

°C : degré Celsius

aa: acides(s) amine(s)

ADN : Acide Désoxyribonucléique

ADNc : ADN complémentaire

Ala : Alanine

Arg : Arginine

Phe : Phénylalanine

ARN : Acide Ribonucléique

ARNm : ARN messenger

ARNr : ARN ribosomique

ARNt : ARN de transfert

ATP : Adénosine-Triphosphate

ATX : Ataxine

BSA : Bovine Serum Albumin (Albumine de sérum bovin)

CAPER α : Cancer Associated Protein ER α

CLIP-seq: Cross-Linking ImmunoPrecipitation-sequencing

CTD: C-terminal Domain

C-ter : C-terminal(e)

Cys : Cystéine

Da : Dalton

DAPI :4',6-Diamidine-2'-phenylindole dihydrochloride

DCIS : Ductal Carcinoma *In Situ*

DMEM : Dulbecco's Modified Eagle Medium

dNTP : désoxy Nucléotide Triphosphate

EDTA : Éthylènediaminetétraacétique

EMSA : Electrophoretic Mobility Shift Assay

ER+ : Estrogen Receptor +

ESE : Exonic Splicing Enhancer

ESS : Exonic Splicing Silencer

FIR : FBP-interacting repressor

FLIM : Fluorescence-lifetime Imaging Microscopy

FRET : Fluorescence Resonance Energy Transfer

Glu : Glutamine

Gly : Glycine

GST : Glutathion S Transférase

HDAC : Histone Désacétylase

HEK : Human Embryonic Kidney

HeLa : Henrietta Lacks

hGH : human Growth Hormone

hnRNP : heteronuclear RiboNucleic Particle

Introns : Intragenic Region

IPTG : Isopropyl β -D-1-thiogalactopyranoside

ISE : Intronic Splicing Enhancer

ISS : Intronic Splicing Silencer

ITC : Isothermal Calorimetric Titration

Kb : kilobase

kDa : kilo Dalton

KH : hnRNP K-Homology

KIS : Kinase-Interacting Stathmin
Lys : Lysine
MCF-7 : Michigan Cancer Foundation-7
MCS : Multi Cloning Site
miR, miRNA : microRNA
NLS : Nuclear Localisation Signal
NOVA : Neuro-oncological Ventral Antigen
Nt : nucléotide(s)
N-ter : N-terminal(e)
PBS : Phosphate Buffered Saline
PCNA : Proliferating Cell Nuclear antigen
PCR : Polymerase Chain Reaction
PLA : Proximity Ligation Assay
PMSF : Phenylmethane Sulfonyl fluoride
Pol : polymérase
pré-ARNm : Acide Ribo Nucléique pré-messager
SR : Serine Arginine rich domain
PTB : Pyrimidine Tract Binding protein
PUF60 : Poly(U)-binding-splicing Factor 60
RBM17 : RNA-binding motif protein 17
RBM39 : RNA-binding motif protein 39
RBP : RNA-binding protein
RF-cloning : Restriction Free cloning
RISC : RNA Induced Silencing Complex
RMN : Résonance Magnétique Nucléaire
RNP : Ribo Nucleic Particle
RRM : RNA Recognition Motif
SDS-PAGE : SDS-Polyacrylamide Gel Electrophoresis
shRNA : small hairpin RNA
SMA : Spinal Muscular Atrophy
snRNA : small nuclear RNA
snRNP : small nuclear RNP
SPF45 : Splicing factor 45
SRP : Signal Recognition Particule
SV40 : Simian Virus 40
U2AF : U2 Auxillary Factor
UHM : U2AF Homology Motif
UHMSF : U2AF Homology Motif Splicing Factor
ULM : U2AF Ligand motif
UV : Ultra-Violet
Val : Valine
V-EGF : Vascular Epidermal Growth Factor
Zn : Zinc

Liste des figures

I. Etat de l'art

Figure 1. Représentation schématique des deux étapes de transestérification de la réaction d'épissage	155
Figure 2. Séquences nécessaires à l'identification des introns.....	166
Figure 3. Représentation schématique de l'organisation de la snRNP U1 (haut) et interaction du snRNA U1 avec le pré-ARNm (bas).....	188
Figure 4. Représentation schématique de l'organisation de la snRNP U2 (haut) et interaction du snRNA U2 avec le pré-ARNm (bas).....	199
Figure 5. Représentation schématique de la structure d'un domaine RRM canonique	211
Figure 6. Exemple de protéines SR-Like	233
Figure 7. Les étapes d'assemblage du spliceosome	244
Figure 8. Représentation schématique du complexe E.....	255
Figure 9. Représentation schématique du complexe A	277
Figure 10. Schéma illustrant la complexité et la dynamique de la composition des complexes A, B et C du spliceosome chez l'Homme	3030
Figure 11. Modes d'épissage alternatif des pré-ARNm.....	322
Figure 12. Séquences cis impliquées dans la régulation de l'épissage Alternatif.....	355
Figure 13. Méthode SELEX d'identification des séquences activatrices de l'épissage de Schaal & Maniatis, 1999.....	366
Figure 14. Dualité fonctionnelle des protéines SR.....	411
Figure 15. Régulation de l'épissage alternatif par la machinerie transcriptionnelle et les modifications épigénétiques	499
Figure 16. Structure tridimensionnelle d'une interaction UHM-ULM	499
Figure 17. Complexe ternaire d'interactions UHM-ULM composé des interactions entre SF1-U2AF ⁶⁵ -U2AF ³⁵ permettant la reconnaissance des sites d'épissage en 3'	53
Figure 18. Représentation schématique d'une interaction UHM-ULM.....	54
Figure 19. Représentation schématique des structures des protéines à domaines UHM.....	57
Figure 20. Représentation schématique des structures des protéines à domaines ULM identifiées.....	59
Figure 21. Le facteur d'épissage à multi-ULM SF3b155 durant les étapes de reconnaissance des sites 3' d'épissage.....	60
Figure 22. Modèle de recrutement de la snRNP U2 au pré-ARNm.....	61
Figure 23. Modèle de régulation de l'épissage alternatif par le facteur U2AF proposé par Shao et al, 2014.....	64

III. Matériel et Méthodes

Figure 24. Carte du plasmide pGEX-6P-1 utilisé pour réaliser des constructions de U2AF ⁶⁵ , CAPER α , PUF60 et SPF45	80
Figure 25. Carte du plasmide pGL4.16 utilisé pour la préparation des constructions TJP1, DN2, TJP1iDN2, DN2iTJP1, TJP1_SXL et TJP1_9U.....	81
Figure 26. Schéma représentant l'insertion d'un fragment d'ADN par RF-cloning.....	89
Figure 27. Mécanisme d'action du shRNA.....	97

Figure 28. Principe de la méthode de Proximity Ligation Assay.....	101
Figure 29. Représentation schématique de la méthode de GST-Pull Down.....	102
Figure 30. Exemple d'une plaque microfluidique.....	104

IV. Résultats

Figure 31. Contrôle de l'expression basale des protéines par immunofluorescence.....	113
Figure 32. Contrôle positif de l'expérience de Proximity Ligation Assay.....	114
Figure 33. Contrôles négatifs de l'expérience de PLA.....	115
Figure 34. Résultats obtenus lors de l'étude de la proximité de U2AF ⁶⁵ avec CAPER α , PUF60 et SPF45.....	117
Figure 35. Résultats obtenus lors de l'étude de la proximité de CAPER α avec PUF60 et SPF45 et de PUF60 avec SPF45.....	118
Figure 36. Résultats obtenus lors de l'étude de la proximité de SF3b155 avec CAPER α , PUF60 et SPF45.....	119
Figure 37. Interaction des protéines UHM avec SF3b155.....	121
Figure 38. Interaction de SF3b155 avec les protéines UHM.....	122
Figure 39. U2AF ⁶⁵ renforce le recrutement de CAPER α lors de l'interaction avec SF3b155.....	123
Figure 40. Le renforcement du recrutement de CAPER α par U2AF ⁶⁵ est dépendant de SF3b155.....	125
Figure 41. L'augmentation du recrutement de CAPER α sur SF3b155 est dépendant d'un domaine en N-terminal de U2AF ⁶⁵	127
Figure 42. Représentation schématique des différents mutants de SF3b155 utilisés au cours de l'étude.....	128
Figure 43. Interactions de CAPER α et U2AF ⁶⁵ avec SF3b155 dépendante du nombre de tryptophanes dans les motifs ULM.....	129
Figure 44. Coopérativité de CAPER α et U2AF ⁶⁵ pour les interactions avec SF3b155 dépendante du nombre de tryptophanes dans les motifs ULM.....	131
Figure 45. Méthode de sélection des événements d'épissage alternatif analysés au cours de l'étude.....	134
Figure 46. Sélection de 65 exons cassette sur la base de leur score Δ SPY.....	136
Figure 47. Niveaux d'expression des protéines U2AF ⁶⁵ , CAPER α , PUF60 et SPF45 par transduction virale de shRNA.....	137
Figure 48. Niveaux d'expression des ARNm CAPER α , U2AF ⁶⁵ , PUF60 et SPF45 par qPCR.....	138
Figure 49. Représentation graphique de trois exemples d'index d'épissage et de changements d'épissage induits par la diminution du taux de U2AF ⁶⁵ , CAPER α , PUF60 ou SPF45.....	140
Figure 50. Représentation graphique des changements d'épissage induits par l'inhibition de CAPER α , U2AF ⁶⁵ , PUF60 et SPF45.....	141
Figure 51. Diagramme de Venne représentant les changements d'épissage diminués (gauche) et augmentés (droite) communs aux quatre facteurs d'épissage à domaines UHM.....	142
Figure 52. Tableau représentant les changements d'épissage communs à deux facteurs d'épissage à domaines UHM.....	143
Figure 53. Représentation graphique en nuage de points des valeurs de changements d'épissage observées entre deux facteurs d'épissage à domaines UHM.....	143
Figure 54. Représentation graphique en nuage de points des valeurs de changement d'épissage lors de l'inhibition de CAPER α , U2AF ⁶⁵ , PUF60 et SPF45 par rapport au score Δ SPY.....	144

Figure 55. Séquences des introns des ARNm DNM2 et TJP1 et constructions des minigènes rapporteurs des exons cassette analysés.....	146
Figure 56. Représentation graphique des index d'épissage et des changements d'épissage des minigènes rapporteurs TJP1 ET DNM2 induits par l'inhibition de U2AF ⁶⁵ et CAPER α	148
Figure 57. Représentation graphique des index d'épissage et des changements d'épissage des minigènes rapporteurs TJP1iDNM2 et DNM2iTJP1 induits par l'inhibition de U2AF ⁶⁵ et CAPER α	149
Figure 58. Représentation graphique des index d'épissage des constructions de minigènes rapporteurs TJP1_9U et TJP1_sxl induits par l'inhibition de U2AF ⁶⁵ et CAPER α	150
Figure 59. Sédimentation <i>in vitro</i> de CAPER α et U2AF ⁶⁵	151
Figure 60. Sédimentation différentielle de U2AF ⁶⁵ avec les ARN TJP1 et DNM2 en présence de SF3b155.....	153
Figure 61. Sédimentation différentielle des ARN TJP1 et DNM2 avec U2AF ⁶⁵ en présence de SF3b155.....	154
Figure 62. Formation d'assemblages macromoléculaires par U2AF ⁶⁵ et CAPER α en présence de SF3b155 et d'ARN riche en enchainements de pyrimidines (1).....	156
Figure 63. Formation d'assemblages macromoléculaires par U2AF ⁶⁵ et CAPER α en présence de SF3b155 et d'ARN riche en enchainements de pyrimidines (2).....	157

V. Discussion, perspectives et conclusion

Figure 64. Modèle de formation d'assemblages macromoléculaires de CAPER α et U2AF ⁶⁵ pour la reconnaissance des séquences introniques riches en enchainements de pyrimidines et le recrutement de la snRNP U2.....	166
---	-----

Liste des tableaux

I. Etat de l'art

Tableau 1. Nouvelle nomenclature des protéines SR, anciens noms correspondants et domaines protéiques.....	20
Tableau 2 : Séquences consensus activatrices de l'épissage découvertes par la méthode SELEX, d'après Schaal & Maniatis 1999.....	37
Tableau 3. Séquences reconnues par les protéines SR.....	39
Tableau 4. Classification des protéines hnRNP : nom et représentation schématique des structures de chaque famille d'isoformes.....	43
Tableau 5. Séquences de fixation des hnRNP sur l'ARN.....	45
Tableau 6. Famille des protéines à domaines UHM: nomenclature, position et taille du domaine UHM, partenaires ULM connus, séquences RNP1/RNP2 et motif RXF.....	58
Tableau 7. Protéines à domaines ULM: nom, position du domaine ULM, partenaires UHM connus et séquences ULM.....	60

III. Matériel et Méthodes

Tableau 8 : Les différents plasmides recombinants utilisés au cours de l'étude.....	84
Tableau 9 : Séquences des shRNA utilisés au cours de l'étude.....	85
Tableau 10: Liste des oligoséoxyribonucléotides amorces utilisés pour les expériences de clonage et de transcription <i>in vitro</i>	85
Tableau 11 : Liste des oligoséoxyribonucléotides amorces utilisés pour les expériences de qPCR.....	86
Tableau 12 : Liste des Anticorps utilisés en western blot et immunofluorescence.....	87
Tableau 13 : Tampons et solutions utilisés.....	88
Tableau 14 : Milieux de culture utilisés.....	89
Tableau 15 : Cycles d'amplification des fragments d'ADN par PCR.....	92
Tableau 16 : Cycles d'amplification du plasmide et insertion du megaprimer par PCR.....	93
Tableau 17 : Cycles d'amplification d'ADN provenant de colonies bactériennes par PCR.....	93

IV. Résultats

Tableau 18 : Colocalisations protéiques analysées au cours de l'étude.....	114
---	-----

I. Etat de l'art

Chapitre I : L'épissage des pré-ARN messagers

« Tous les Hommes désirent naturellement savoir ». Cette phrase, qui débute La Métaphysique d'Aristote, affirme ce désir de posséder la science qui guide chaque Homme dans sa quête de la compréhension, de la connaissance et du savoir. Cette phrase, écrite il y a plus de 2500 ans, qui a été interprétée, empruntée, traduite et transformée a retenu mon attention, il y a quelques semaines, lorsque j'ai eu entre les mains un roman de Nina Bouraoui qui avait pour titre cette même citation. Dans sa quête du savoir, l'auteure entreprend une rétrospection croisée de deux étapes clés de sa vie. Dans ce contexte, la notion de désir de savoir prend une toute autre dimension : pour savoir il faut se rappeler, s'enquérir des théories, des notions, des recherches et des méthodes antérieures. Pour savoir, pour comprendre, il faut analyser, décomposer puis reconstituer l'image finale. La quête de connaissance n'est plus une avancée solitaire. Elle s'appuie sur les découvertes passées.

Pour chaque recherche il y a une équipe de chercheurs, un modèle de recherche et une découverte clé. C'est dans le génome de l'Adénovirus que l'équipe de Phillip Allen Sharp et Richard Roberts ont identifié, pour la première fois en 1977, des gènes morcelés [Berget et al., 1977; Chow et al., 1977]. Les ARN messagers n'étaient plus des copies conformes de l'ADN. Une nouvelle organisation venait d'émerger. Entre les séquences d'ARN codant pour des protéines se trouvaient des séquences non codantes. Peu de temps après, en 1978, l'équipe de Walter Gilbert a défini cet arrangement particulier des ARN messagers [Gilbert, 1978]. Dans un ordre précis, les transcrits étaient composés de séquences « informationnelles » entrecoupées de séquences « silencieuses ». A partir de cette organisation, Walter Gilbert a nommé les séquences exprimées « Exons », contraction de EXpressed regiON et les séquences silencieuses « Introns », contraction de INTragenic RegiON. Afin d'obtenir un ARN messenger mature qui serait traduit en protéines, le transcrit devait alors subir un processus où les introns étaient éliminés et les exons joints les uns aux autres. Ce processus fut appelé épissage.

Comme dans tout domaine de recherche, des études supplémentaires viennent s'ajouter, compléter ou parfois contredire les premières découvertes. Contrairement à ce que pensaient Phillip Allen Sharp, Richard Roberts, Walter Gilbert et leurs équipes respectives, les introns n'étaient pas spécifiques des ARN codants pour des protéines. Ces séquences silencieuses font également partie des précurseurs d'ARN non codants tel que les ARN de transfert et les ARN

ribosomiques. Par ailleurs, cette organisation de séquences ne se limite pas exclusivement aux précurseurs d'ARN nucléaires, puisque des introns sont également présents dans les ARN d'organites cellulaires comme les mitochondries ou les chloroplastes.

Si tous les précurseurs d'ARN matures présentent une organisation similaire, tous les introns ne sont pas éliminés de la même manière. Certains introns sont dit « auto-catalytiques » c'est-à-dire que l'élimination de ces séquences peut s'effectuer en l'absence de protéines et par une action catalytique inhérente au pré-ARN. Dans la famille des introns auto-catalytiques, on distingue trois groupes différents, selon la réaction chimique générée lors de l'élimination de l'intron.

Les introns de groupe I sont présents dans les précurseurs d'ARN ribosomique, les précurseurs d'ARN de transfert et les pré-ARN messagers des protozoaires, des mitochondries de fungus, des chloroplastes d'algues, des bactériophages et des eubactéries. L'épissage de ces introns est réalisé par deux étapes de transestérification initiées par une guanosine ou une guanosine triphosphate externe à la séquence d'ARN [Haugen et al., 2005].

Les introns de groupe II sont quant à eux présents dans les précurseurs d'ARN des chloroplastes et mitochondries de cellules eucaryotes [Lambowitz and Zimmerly, 2011]. L'épissage des introns de groupe II est également réalisé par deux réactions de transestérification successives. Cependant, contrairement aux introns de groupe I, cette réaction est initiée par une adénosine présente dans la séquence de l'intron.

Enfin, les introns de groupe III, ont été identifiés dans les pré-ARNm de chloroplastes de protistes [Montandon and Stutz, 1984]. Le mécanisme d'excision de ces introns est similaire à celui du groupe II. Cependant, ils se distinguent des introns de groupe II par la taille de leurs séquences, considérablement plus courtes et qui présentent une abondance de nucléotides A et U.

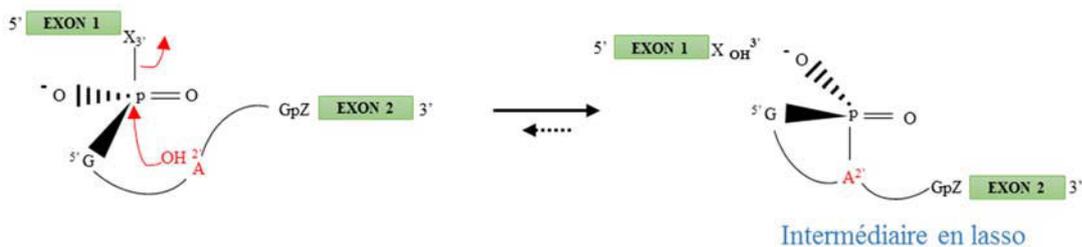
Pour certains précurseurs d'ARN, le mécanisme d'excision des introns est réalisé par un complexe ribonucléoprotéique appelé RNase P. C'est notamment le cas des introns d'ARN de transfert, de la sous-unité 5S de l'ARN ribosomique et de la SRP (particule de reconnaissance de signal). La RNase P possède une activité endonucléase. Elle a la capacité de cliver les molécules d'ARN qui présentent des structures en tige-boucle. Ce clivage est réalisé à la jonction de l'ARN simple brin et l'extrémité 5' de l'ARN double brin [Altman, 2011].

L'élimination des introns des pré-ARN messagers implique comme pour les introns de groupe II, deux réactions de transestérification successives. Cependant, ces ARN semblent avoir perdu leurs propriétés auto-catalytiques au cours de l'évolution et requièrent un nombre important de protéines et de particules ribonucléoprotéiques pour éliminer correctement ces séquences. En effet, cinq particules ribonucléoprotéiques appelées UsnRNP (Uridine-rich small nuclear RiboNucleoProteins) sont responsables de la catalyse de la réaction d'épissage, et forment le spliceosome. La réaction d'épissage d'une grande majorité des pré-ARNm est prise en charge par le spliceosome dit « majeur ». Celui-ci est composé des cinq UsnRNP U1, U2, U4-U6 et U5. L'épissage de moins de un pour cent des pré-ARN messagers humains est réalisé par un autre type de spliceosome dit « mineur », constitué des UsnRNP U11, U12, U4atac, U5 et U6atac [Patel and Steitz, 2003]. Durant l'épissage, les UsnRNP interagissent avec le pré-ARNm et définissent les frontières entre les séquences des exons et des introns.

1. La réaction d'épissage

L'épissage des pré-ARN messagers nucléaires se décompose en deux réactions de transfert de groupements phosphodiester (transestérifications) (Figure 1). Tout d'abord, l'hydroxyle (OH) d'un résidu Adénosine, appelé point de branchement, situé dans la séquence de l'intron, engage

Etape 1: Attaque nucléophile en 5' et formation d'un intermédiaire en lasso



Etape 2: Attaque nucléophile en 3' et ligation des deux exons

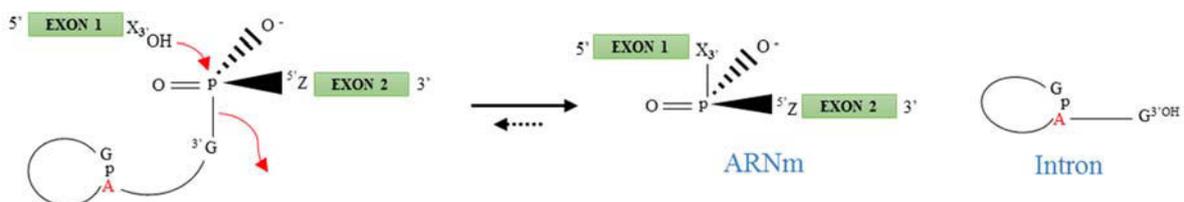


Figure 1. Représentation schématique des deux étapes de transestérification de la réaction d'épissage. Dans l'étape 1, l'attaque du 2' OH de l'Adénosine du point de branchement sur le phosphate du résidu à l'extrémité 5' de l'intron provoque la formation d'un intermédiaire en lasso. Dans l'étape 2, l'attaque du 3' OH de l'exon 1 sur le phosphate à l'extrémité 3' de l'intron permet la libération de l'intron et de l'ARNm épissé. (D'après Moore et Sharp, 1993)

une attaque nucléophile vers un phosphate (P) à l'extrémité 5' de l'intron à la jonction de l'exon 1. Cette réaction génère un exon 1 avec une extrémité 3'OH libre et une structure intermédiaire contenant l'exon 2 toujours lié à l'intron qui est sous forme de lasso.

La deuxième étape de transestérification consiste en l'attaque nucléophile par l'extrémité 3'OH de l'exon 1 du phosphate en 3' de l'intron à la jonction de l'exon 2. L'intron est ainsi libéré sous forme de lasso et les deux exons sont unis [Moore and Sharp, 1993].

2. Les séquences en *cis* nécessaires à l'identification des introns

Les premières études cherchant à définir les séquences consensus des jonctions entre les exons et les introns n'ont été que peu concluantes étant donné la très faible homologie de ces séquences. Malgré cela, deux paires de nucléotides conservés dans toutes les séquences ont été identifiées. Celles-ci consistaient en un dinucléotide GU situé à l'extrémité 5' de l'intron, et un dinucléotide AG à son extrémité 3' [Breathnach et al., 1978]. Quelques années plus tard, Shapiro et Senapathy [Shapiro and Senapathy, 1987] [Senapathy et al., 1990] ont réussi à identifier des séquences consensus en réalisant des analyses systématiques de centaines de séquences de gènes eucaryotes codant pour des protéines. Les alignements de ces séquences ont permis de définir les jonctions intron/exon de manière plus stricte (Figure 2). Le site d'épissage en 5' de l'intron est alors défini par une séquence consensus AG\GURAGU (R = purine, \ = jonction exon-intron). Le site 3' d'épissage est quant à lui subdivisé en 3 séquences distinctes : une séquence consensus YAG\G (Y= pyrimidine) située à la jonction intron/exon, précédée d'une séquence composée d'un enchainement de pyrimidines et un point de

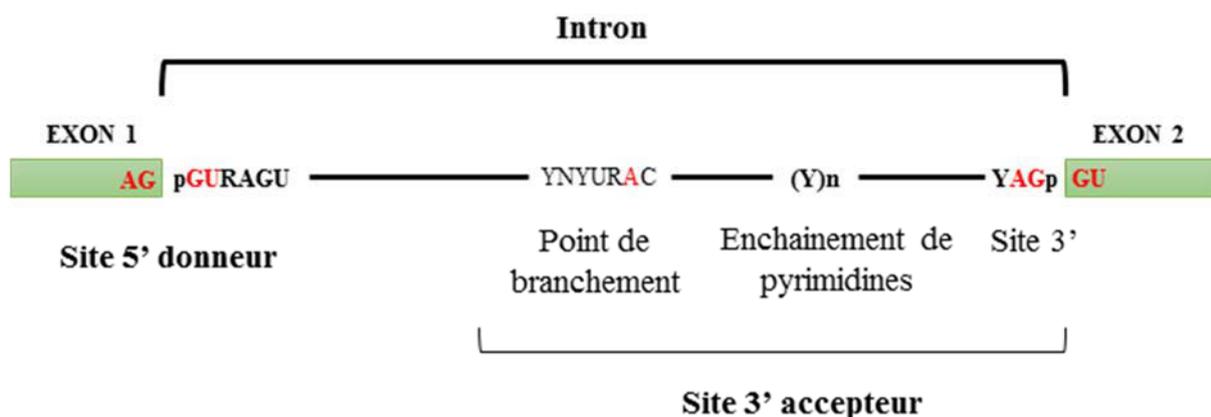


Figure 2. Séquences nécessaires à l'identification des introns. Représentation schématique de l'organisation d'un intron situé entre deux exons. Les séquences conservées jouant un rôle important dans la mise en place de la machinerie d'épissage sont indiquées Y : pyrimidine ; R : purine. (Pour revue, Cartegni et al, 2002)

branchement constitué d'une adénosine dans une séquence consensus YNYURAC (R = purine, Y = pyrimidine et N = A, C, G ou U) entre 10 et 40 nucléotides en amont de la jonction intron/exon.

3. La réaction d'épissage par le spliceosome majeur

Le spliceosome majeur s'assemble sur le pré-ARNm et modifie sa structure et sa composition en protéines au cours de la réaction d'épissage. Les cinq particules snRNP U1, U2, U4, U5 et U6 qui composent le complexe spliceosomal, ainsi qu'un grand nombre de facteurs d'épissage reconnaissent les séquences consensus présentes dans les introns et engagent des interactions ARN-ARN ou Protéine-ARN afin d'accomplir les deux réactions de transestérification.

3.1. Les UsnRNP

Chaque UsnRNP qui compose le spliceosome est constituée d'un ARN spécifique appelé UsnRNA et de plusieurs protéines, certaines communes et d'autres spécifiques. Les séquences des UsnRNA sont fortement conservées chez les eucaryotes [Branlant et al., 1980]. Pendant leurs processus de maturation, ils subissent des modifications post-transcriptionnelles tel que des 2'-O-méthylations et des pseudo-uridylation [Massenet and Branlant, 1999] nécessaires à leurs fonctions dans l'épissage et à leur assemblage en UsnRNP. Sept protéines, appelées protéines Sm, sont communes à quatre des cinq UsnRNP. Elles sont retrouvées dans la composition des UsnRNP U1, U2, U4 et U5, dont le processus d'assemblage s'effectue dans le cytoplasme. Les sept protéines Sm s'organisent en anneau heptamérique autour du snRNA et se lient à des séquences spécifiques [Will et al., 2001]. La snRNP U6, dont l'assemblage s'effectue dans le noyau, est composée de sept protéines appelée protéines LSm qui s'organisent de la même manière que les protéines Sm autour du UsnRNA en se liant à des séquences spécifiques [He and Parker, 2000].

En plus de l'anneau heptamérique de protéines Sm, la snRNP U1 est composée des trois protéines U1-A, U1-70K et U1-C. La structure et l'organisation de la particule est représentée schématiquement dans la Figure 3. Les protéines U1-A, U1-70K et les protéines Sm sont liées à l'ARN de manière directe, tandis que la protéine U1-C est liée à la particule par l'intermédiaire d'interactions protéine-protéine. La séquence en 5' du snRNA U1 est très conservée et intervient dans la liaison avec la jonction intron/exon au niveau du site 5' d'épissage. La protéine U1-70K renforce cette liaison par l'intermédiaire d'interactions protéine-protéine [Will and Lührmann, 2011].

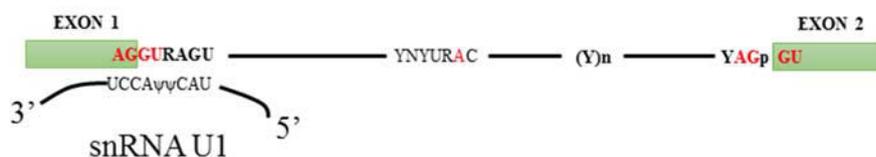
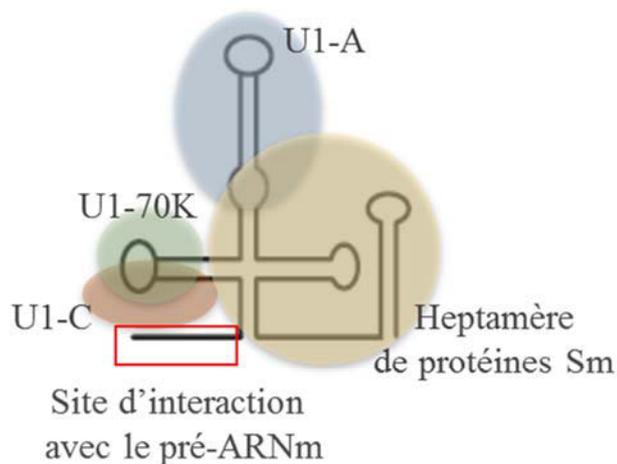


Figure 3. Représentation schématique de l'organisation de la snRNP U1 (haut) et interaction du snRNA U1 avec le pré-ARNm (bas). La snRNP U1 est composée d'un anneau de protéines Sm et des protéines U1-A, U1-70K et U1-C. L'extrémité 5' du snRNA U1 intervient dans la liaison avec l'extrémité 5' de l'intron du pré-ARNm (pour revue, Will and Luhrmann, 2011)

La snRNP U2 dont la composition est schématisée dans la Figure 4 est quant à elle composée, en plus des protéines Sm, des protéines U2A', U2B'' et du complexe SF3. Ce complexe multi-protéique est composé de deux sous-unités SF3a et SF3b qui à leur tour sont composées de plusieurs protéines. La sous unité SF3a contient les protéines SF3a 61, 62 et 114 tandis que la protéine SF3b est composée des sous-unités SF3b49, SF3b130, SF3b145 et SF3b155 et des protéines p10, p14a et p14b.

Le snRNA U2 se lie à la séquence du point de branchement par appariements de bases. Cette interaction est stabilisée par les protéines SF3a et SF3b. la protéine p14 vient ensuite interagir directement avec le point de branchement [Wahl et al., 2009] [Will and Lührmann, 2011].

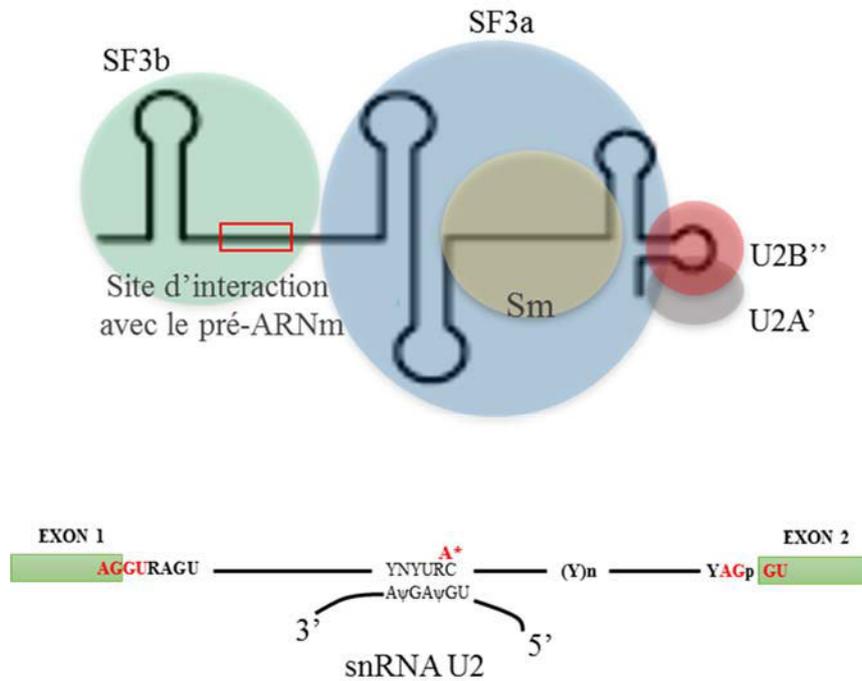


Figure 4. Représentation schématique de l'organisation de la snRNP U2 (haut) et interaction du snRNA U2 avec le pré-ARNm (bas). La snRNP U2 est composée d'un anneau de protéines Sm et des protéines U2A', U2B'' et des complexes SF3a et SF3b. Le snRNA U2 interagit avec la séquence du point de branchement de l'intron du pré-ARNm (pour revue, Wahl et al., 2009).

Les deux snRNP U4 et U6 s'associent dans le noyau dès la fin de leur maturation pour former la di-snRNP U4/U6. La snRNP U5 vient ensuite s'associer pour former la tri-particule U4/U6-U5. La tri-snRNP U4/U6-U5 est composée d'un nombre important de protéines. Elle subit différents réarrangements structuraux au cours de l'épissage, induisant des changements dans sa composition protéique.

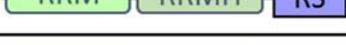
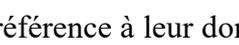
Une fois assemblée, la tri-snRNP rejoint le spliceosome, ce qui conduit à des réarrangements structuraux durant lesquels la snRNP U6 s'associe au site 5' d'épissage et interagit simultanément avec la snRNP U2. La snRNP U5 quant à elle, interagit avec les deux exons de part et d'autre de l'intron, induisant un rapprochement spatial des deux sites d'épissage 5' et 3' [Wahl et al., 2009; Will and Lührmann, 2011].

3.2. Les protéines SR

Les protéines SR composent une famille de protéines qui possèdent des similitudes structurales et qui interviennent à différents moments de l'épissage. Chez les eucaryotes, ces protéines possèdent des séquences très conservées entre les espèces. Chez l'homme, douze protéines SR ont été identifiées. Elles ont récemment été classifiées et renommées afin d'homogénéiser la

nomenclature [Manley and Krainer, 2010]. Le tableau 1 regroupe la nouvelle nomenclature des protéines SR, leurs anciens noms respectifs et leurs domaines protéiques.

Tableau 1. Nouvelle nomenclature des protéines SR, anciens noms correspondants et domaines protéiques. RS : domaine riche en arginines et en sérines ; RRM : domaine d'interaction avec l'ARN ; RRMH : RRM Homologue ; Zn : motif doigt à zinc.[Manley and Krainer, 2010]

Nouvelle nomenclature	Noms précédents	Domaines protéiques
SRSF1	ASF/SF2SF2 ; ASF ; SRp30a	
SRSF2	SC35 ; PR264 ; SRp30b	
SRSF3	SRp20	
SRSF4	SRp75	
SRSF5	SRp40 ; HRS	
SRSF6	Rp55 ; B52	
SRSF7	9G8	
SRSF8	SRp46	
SRSF9	SRp30c	
SRSF10	SRp38 ; TASR1 Srrp40	
SRSF11	SRp54 ; p54	
SRSF12	Srrp35	

Ces protéines sont nommées ainsi en référence à leur domaine d'interaction avec les protéines ou domaine RS. Dans toutes les protéines SR, le domaine RS est situé à l'extrémité C-terminale. Ce domaine se compose d'un nombre variable de dipeptides Arginine-Serine et permet à ces protéines d'établir des interactions avec les domaines RS d'autres facteurs protéiques, par l'intermédiaire d'interactions de type ionique, impliquant des arginines chargées positivement et des sérines phosphorylées [Zahler et al., 1992].

La phosphorylation du domaine RS est indispensable à l'assemblage du spliceosome [Cao et al., 1997; Fetzer et al., 1997]. Ce domaine est soumis à de nombreux événements de

phosphorylation qui permettent de réguler l'activité des protéines SR et leur localisation cellulaire [Sanford and Bruzik, 2001]. Plusieurs types de protéines kinases agissant sur les protéines SR ont été identifiées, notamment la famille des SRPK (SRPK1 et SRPK2) pour *SR protein kinase*, la famille Clk/Sty, Akt, et l'ADN topoisomérase I [Colwill et al., 1996; Patel et al., 2005; Prasad et al., 1999; Rossi et al., 1996]. Les protéines SR sont également soumises à l'activité de phosphatases spécifiques PP1A, PP2A et PP2C [Shi et al., 2006].

Dans la séquence des protéines SR, on retrouve également en N-terminal un ou deux domaines d'interaction avec l'ARN, appelés domaines RRM pour *RNA Recognition Motif*. En fonction du nombre de domaines RRM qu'elles contiennent, les protéines SR sont classées en deux groupes distincts. Le premier groupe est composé des protéines comprenant deux domaines RRM, à savoir SFRS1 (ASF/SF2), SFRS4 (SRp75), SFRS5 (SRp40), SFRS6 (SRp55) et SFRS9 (SRp30c). Le deuxième groupe est composé des protéines qui ne contiennent qu'un seul domaine RRM et rassemble les protéines SFRS2 (SC35), SFRS3 (SRp20), SFRS2B (SRp46), SFRS7 (9G8), SFRS13B (Srrp35), SFRS5 (Srrp40) et SFRS11 (SRp54). Les deux domaines RRM1 et RRM2 des protéines SR sont constitués principalement de deux hélices α et d'un feuillet β composé de 4 brins β antiparallèles avec une topologie $\beta\alpha\beta\beta\alpha\beta$ formant un cœur hydrophobe. Le feuillet β joue un rôle de plateforme de fixation à l'ARN [Zahler et al., 1992]. Les deux brins β centraux contiennent deux séquences RNP-1 (séquence consensus : RGFVGFVxF) et RNP-2 (séquence consensus : LFVGNL) [Bandziulis et al., 1989] qui interviennent dans l'établissement de contacts directs avec l'ARN (Figure 5). En fonction du

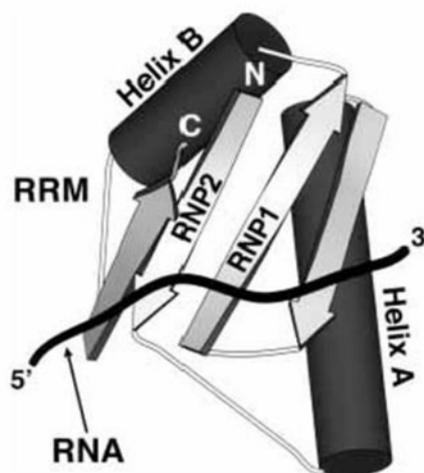


Figure 5. Représentation schématique de la structure d'un domaine RRM canonique. Un domaine RRM canonique est composé de deux hélices α (A et B) contre quatre feuillets β antiparallèles (flèches) avec une topologie $\beta\alpha\beta\beta\alpha\beta$. Les motifs RNP 1 et 2 forment les deux feuillets β centraux : RNP1 étant β 3 et RNP2 étant β 1 (pour revue, Kielkopf et al. 2004)

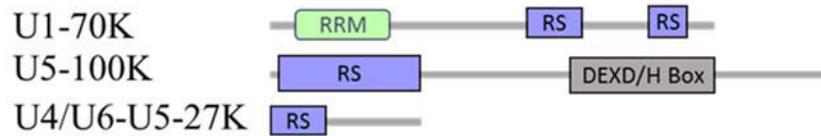
type de protéines SR, d'autres éléments de séquences situés dans des boucles flexibles reliant les brins du feuillet β peuvent avoir un rôle très important dans la spécificité de la reconnaissance de l'ARN.

Le domaine RRM2 présente des séquences RNP-1 et 2 dégénérées par rapport à celles des domaines RRM1 mais contient tous les acides aminés conservés qui forment le cœur hydrophobe du domaine RRM1 [Zhou and Fu, 2013]. Ce domaine RRM2, aussi appelé RRMH (pour RRM Homologue), participe également à la spécificité de reconnaissance des séquences régulatrices sur le pré-ARNm [Cáceres et al., 1997; Chandler et al., 1997]. La protéine SRSF7 (9G8) possède un motif en doigt de zinc, en plus de son unique domaine RRM. Ce motif contribue lui aussi à la spécificité de fixation à l'ARN.

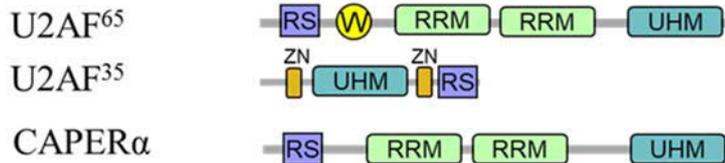
En plus de leurs similarités structurales, les protéines SR présentent des caractéristiques fonctionnelles communes. En effet, toutes les protéines SR ont la capacité d'induire la réaction d'épissage des pré-ARNm *in vitro* dans la fraction cytoplasmique S100, qui présente une déficience pour cette fonction [Zahler et al., 1992]. Cette caractéristique fonctionnelle est un critère de sélection et d'identification des protéines appartenant à la famille des protéines SR.

Les protéines SR établissent également des interactions Protéine-Protéine avec des protéines possédant des domaines RS mais qui ne font pas partie de la famille des protéines SR. Ces protéines sont appelées SR-like ou SRrp pour *SR related proteins* (Figure 6) [Graveley, 2000]. Les protéines SR-like interviennent directement dans l'épissage. Ces protéines peuvent être des facteurs d'épissage, comme c'est le cas de U2AF³⁵, U2AF⁶⁵ et CAPER α qui interviennent dans la reconnaissance des sites 3' d'épissage, ou des protéines composantes des UsnRNP, comme U1-70K, U5-100K et [U4/U6-U5]-27K. Elles peuvent être également des protéines régulatrices comme hTra2 α et hTra2 β ou co-activatrices de l'épissage comme SRm160 et SRm300. A la différence des protéines SR, les protéines SRm160 et SRm300 ne possèdent pas de domaine RRM. Elles s'associent en un complexe SRm160/SRm300 et favorisent l'épissage par l'établissement d'interactions entre les protéines SR et les constituants du spliceosome comme la snRNP U1 [Blencowe et al., 2000; Eldridge et al., 1999]. Les protéines kinases Clk1, 2, 3, capables de phosphoryler le domaine RS des protéines SR, sont aussi des protéines SR-like.

Protéines des UsnRNP



Facteurs d'épissage



Protéines régulatrices



Protéines co-activatrices



Protéines kinases



Figure 6. Exemple de protéines SR-Like ; RS : domaine riche en arginines et en serine pour les interactions protéine-protéine ; RRM : "RNA-Recognition Motif", domaine d'interaction avec l'ARN ; Zn : motif doigt à Zinc ; DEXD/H : caractéristique des hélicases ; UHM : U2AF-Homology Motif pour les interactions protéine-protéine ; W : Tryptophane.

3.3. Les étapes d'assemblage du spliceosome

Avant l'assemblage du spliceosome, le pré-ARNm est associé à des protéines qui se lient à l'ARN. Cette liaison particulière des protéines avec le pré-ARNm forme un complexe, appelé complexe H. Dans ce complexe, on retrouve à l'extrémité 5' du pré-ARNm la protéine CBP80, grande sous-unité du complexe nucléaire de liaison à la coiffe, et des protéines SR qui interviennent dans la reconnaissance des jonctions intron-exon.

Des études *in vitro* de l'assemblage du spliceosome sur le pré-ARNm ont mis en évidence cinq complexes qui correspondent à des étapes successives de l'assemblage du spliceosome. Ces complexes présentent des propriétés différentes de migration sur gel natif témoignant de leur composition variable en protéines. Durant ces étapes, des combinaisons d'UsnRNP et des

facteurs d'épissage s'associent au pré-ARNm. [Wahl et al., 2009]. Il s'agit des complexes E, A, B, B* et C (Figure 7).

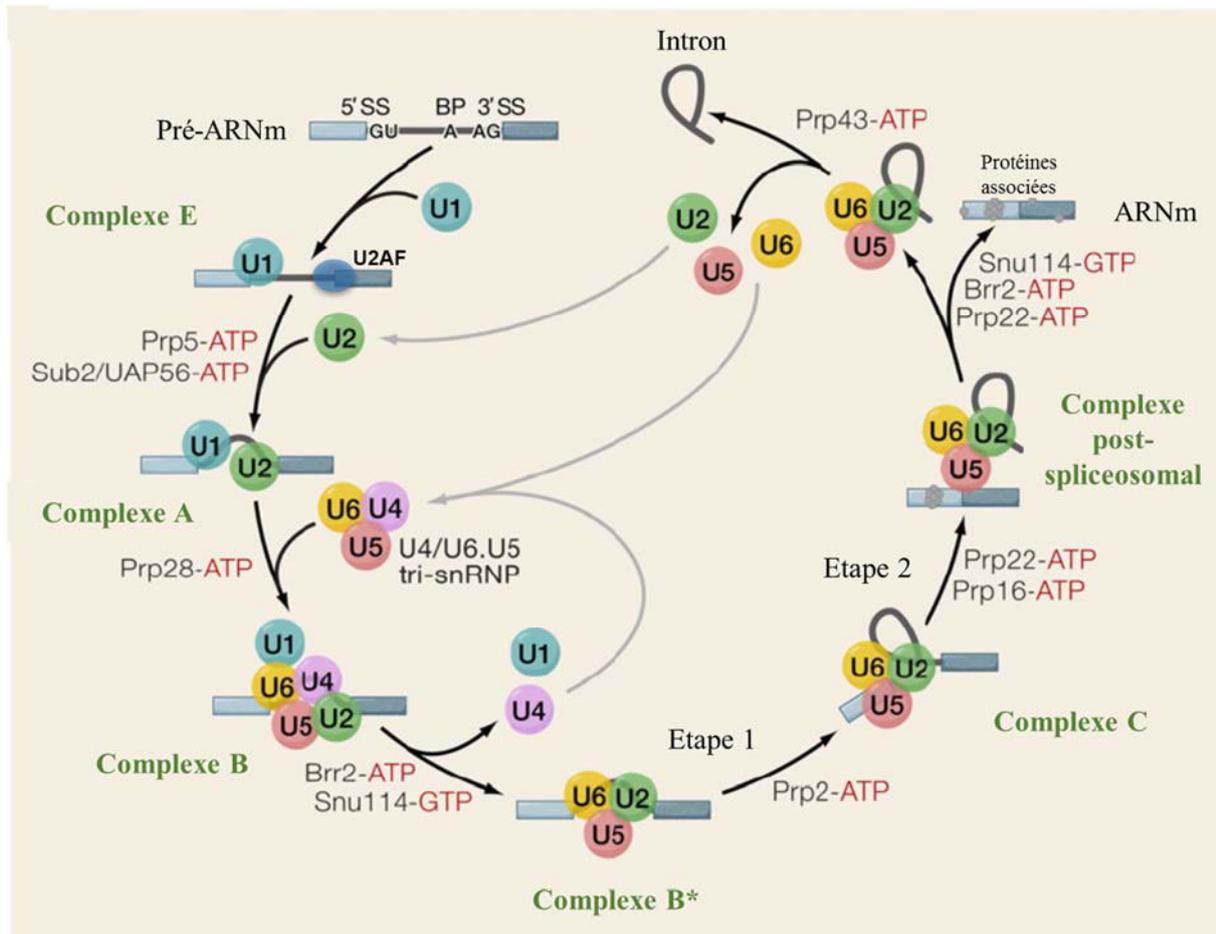


Figure 7. Les étapes d'assemblage du spliceosome. *Complexe E*: Association de la snRNP U1 avec le site 5' d'épissage et du facteur U2AF avec la séquence de répétitions de polypyrimidines et le dinucléotide AG situé au site 3' ; *Complexe A*: Fixation de la snRNP U2 ; *Complexe B*: Association de la triparticule U4/U6•U5 et changements conformationnels du pré-ARNm ; *Complexe B**: Dissociation des snRNP U1 et U4 et interaction des snRNP U2 et U6 entre elles et avec l'intron (au niveau de la boîte de branchement et du site 5' d'épissage) ; *Complexe C*: Le snRNA U5 aligne les extrémités des deux exons pour faciliter leur ligation au cours de la seconde étape de la réaction d'épissage. L'ARNm et l'intron sont libérés sous forme de lasso. Les ATPases et Hélicases nécessaires au réarrangements de l'ARN et au remodelage des RNP à chaque étape sont indiquées. (D'après Wahl et al, 2009)

3.3.1. Reconnaissance des sites d'épissage : formation du complexe E

L'initiation de l'épissage, le recrutement des UsnRNP et leur assemblage séquentiel autour de l'intron repose sur la reconnaissance des sites d'épissage.

Cette reconnaissance des sites d'épissage est initiée en 5' par la fixation de la snRNP U1 et en 3' par l'intermédiaire du facteur d'épissage U2AF composé de ses deux sous-unités U2AF⁶⁵ et U2AF³⁵ qui se lient respectivement à la séquence Polypyrimidine et à la jonction intron-exon. Enfin, la protéine SF1 reconnaît et se lie à la séquence de branchement. L'assemblage de ces différentes protéines sur le pré-ARNm conduit à la formation du complexe d'engagement, ou complexe E (Figure 8).

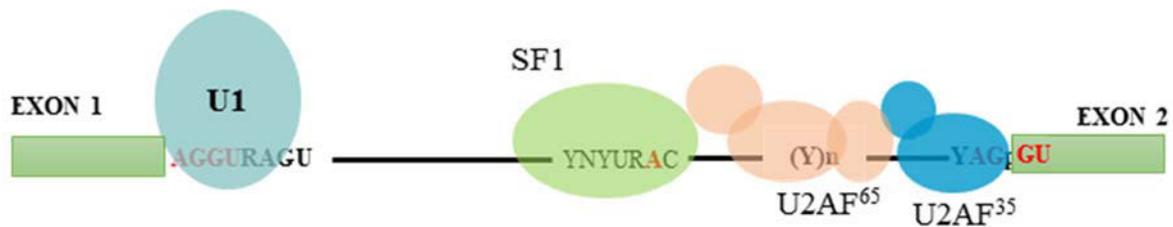


Figure 8. Représentation schématique du complexe E. La première étape de l'épissage est initiée par l'interaction de la snRNP U1 au site d'épissage en 5'. Les deux sous-unités du facteur d'épissage U2AF, U2AF⁶⁵ et U2AF³⁵ interagissent respectivement avec la séquence de répétitions en polypyrimidines et le site 3' d'épissage. Simultanément, le facteur d'épissage SF1 se lie à la séquence du point de branchement.

3.3.1.1. Reconnaissance du site 5' d'épissage par la snRNP U1

L'assemblage du spliceosome débute par l'assemblage de la snRNP U1 à la séquence en 5' de l'intron. Le snRNA U1 possède une séquence complémentaire à celle du site d'épissage en 5'. De plus, la protéine U1-C stabilise cette interaction en interagissant à la fois avec le pré-ARNm et les autres protéines constituantes de la snRNP U1 [Muto et al., 2004].

Il se trouve parfois que la séquence de la jonction exon-intron en 5' soit dégénérée par rapport à la séquence consensus, compromettant ainsi la force de l'interaction de la séquence avec la snRNP U1. Dans ces conditions, des protéines SR, dont SC35 et ASF/SF2 vont alors interagir avec la protéine U1-70K de la snRNP U1 par l'intermédiaire de leurs domaines RS et ainsi, stabiliser la fixation de la particule sur le pré-ARNm [Guirou and O'Reilly, 2015].

3.3.1.2. Reconnaissance du site 3' d'épissage par les facteurs d'épissage U2AF et SF1

La séquence du point de branchement étant particulièrement peu conservée chez les eucaryotes supérieurs, la fixation de la snRNP U2 sur les séquences 3' nécessite plusieurs facteurs d'épissage auxiliaires. Les facteurs d'épissages les plus importants pour la reconnaissance du site 3' d'épissage sont les deux sous-unités de la protéine U2AF (*U2 Auxillary Factor*), U2AF⁶⁵ et U2AF³⁵. La protéine U2AF⁶⁵ se lie à la séquence Polypyrimidine par l'intermédiaire de ses deux domaines RRM, tandis que la sous-unité U2AF³⁵ se fixe au dinucléotide AG au niveau de

la jonction intron/exon [pour revue, Wu and Fu, 2015]. U2AF⁶⁵ possède également un domaine RS, même si celui-ci ne semble pas être impliqué dans les interactions protéine-protéine, il permettrait de stabiliser la snRNP U2 lors de son recrutement au point de branchement [Valcárcel et al., 1996].

La protéine SF1 est le premier facteur d'épissage à reconnaître le point de branchement [Berglund et al., 1997]. La reconnaissance du point de branchement est réalisée par l'intermédiaire d'interactions ARN-protéines entre le domaine de liaison à l'ARN de SF1 et la séquence du pré-ARNm. Ce domaine appelé KH-QUA2 interagit préférentiellement avec la séquence du point de branchement *in vitro* [Berglund et al., 1998b; Peled-Zehavi et al., 2001].

L'affinité de liaison de SF1 avec l'ARN au niveau du point de branchement est augmentée par l'interaction non spécifique de la protéine avec l'ARN par l'intermédiaire de son domaine à doigt de Zinc (Zn) [Berglund et al., 1998b].

L'interaction de SF1 avec le point de branchement est donc dépendante de la conservation de la séquence consensus dans laquelle se trouve le point de branchement. En effet, plus la séquence est conservée, plus l'affinité de l'interaction est forte [Peled-Zehavi et al., 2001].

Enfin, SF1 interagit également avec U2AF⁶⁵ grâce à une structure d'interactions dite UHM-ULM (U2AF-Homology-Motif-U2AF-Ligand Motif). Je détaillerai cette structure d'interactions dans le Chapitre III. L'interaction de SF1 avec U2AF⁶⁵ permet de renforcer la liaison des protéines avec la séquence de l'intron et se fait de manière coopérative avec le pré-ARNm [Berglund et al., 1998a; Guth and Valcárcel, 2000].

L'ensemble de ces étapes, incluant l'assemblage de la snRNP U1 sur la jonction exon-intron en 5', la liaison de U2AF³⁵ au dinucléotide AG et de U2AF⁶⁵ à la séquence polypyrimidine et enfin le recrutement de SF1 au niveau du point de branchement conduisent à la formation du complexe d'engagement, ou complexe E.

3.3.2. Recrutement de la snRNP U2 : Formation du complexe A

La formation du complexe A, appelé pré-spliceosome, est caractérisée par la fixation de la snRNP U2 au point de branchement [Das et al., 1999]. Cette fixation est ATP-dépendante et conduit au déplacement du facteur SF1 [Berglund et al., 1998a]. Elle fait intervenir des hélicases, tel que la protéine UAP56, induisant des changements de conformation du pré-ARNm. A ce stade de l'assemblage du spliceosome, un réseau d'interactions s'établit alors entre

les protéines liées au pré-ARNm et les UsnRNP [Will and Lührmann, 2011], ce qui pourrait permettre le rapprochement des exons de part et d'autre de l'intron.

L'interaction de la snRNP U2 avec le pré-ARNm est réalisée par la liaison du snRNA U2 à la séquence du point de branchement (Figure 4 et Figure 9). Cette interaction ARN-ARN est stabilisée par les complexes SF3a et SF3b [Gozani et al., 1996]. En effet, les protéines des deux sous-unités (à l'exception de SAP130) interagissent de manière non spécifique avec l'ARN, en amont du point de branchement [Will and Lührmann, 2011].

La sous-unité SF3b155 de la protéine SF3b se lie à l'ARN de part et d'autre du point de branchement et interagit avec U2AF⁶⁵ [Champion-Arnaud and Reed, 1994]. La protéine p14/SF3b 14a se fixe sur le point de branchement et favorise l'interaction du snRNA U2 avec le pré-ARNm [Will et al., 2001].

La fixation de la snRNP U2 et l'appariement du snRNA U2 à la séquence du point de branchement, permet une modification de la structure de l'ARN et un dégagement de l'adénosine du point de branchement hors de la double hélice d'ARN, ce qui permettra le déclenchement de la première attaque nucléophile du site d'épissage en 5' par le groupement hydroxyle 2'OH de l'adénosine du point de branchement.

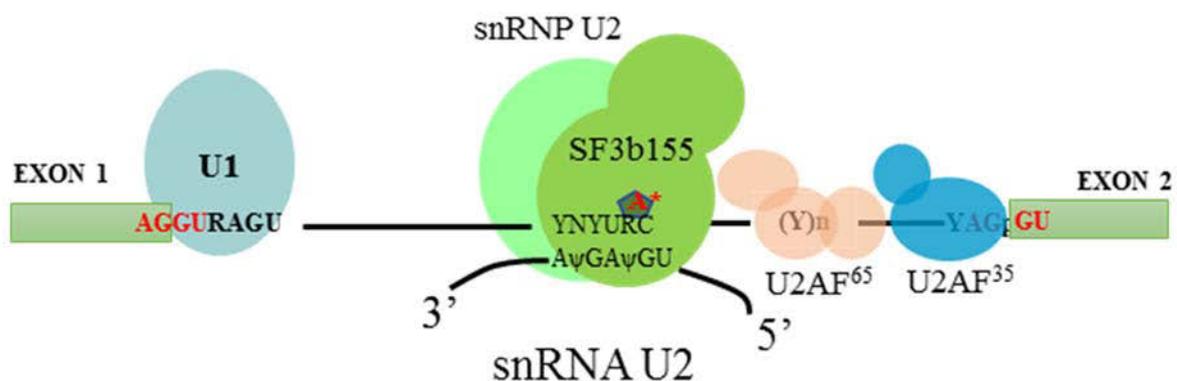


Figure 9. Représentation schématique du complexe A. Le snRNA U2 se lie à la séquence du point de branchement. Cette liaison est stabilisée par les complexes SF3a et SF3b. La sous-unité SF3b155 se fixe de part et d'autre du point de branchement, tandis que p14 se fixe sur le point de branchement. L'ensemble de ces interactions induit une restructuration du pré-ARNm et le dégagement de l'adénosine du point de branchement.

3.3.4. Recrutement de la tri-snRNP U4/U6-U5 : Formation des complexes B et B*

La tri-snRNP U4/U6-U5 est préassemblée et rejoint le spliceosome en formation. Cette tri-snRNP interagit avec le pré-ARNm pour former le complexe B (Figure 7). Durant ces

interactions, la snRNP U5 se lie aux séquences du pré-ARNm simultanément en 5' et en 3'. La snRNP U1 se détache du site 5' et laisse place à la snRNP U6. Des réarrangements structuraux ATP-dépendants conduisent à la dissociation des snRNP U4 et U6 en faveur de l'interaction entre les snRNP U6 et U2. L'interaction entre les snRNA U4 et U6 inhibe le site actif du snRNA U6 qui se lie au site 5' d'épissage au cours de la première étape de transestérification. L'inactivation initiale de la snRNP U6 et sa libération par des changements conformationnels assurent que l'intron du pré-ARNm n'est pas clivé prématurément. Au cours de ces réarrangements conformationnels, le snRNA U2 reste associé au point de branchement. Lorsque le duplexe de snRNA U6-U2 est formé, le site actif du snRNA U6 est activé [Dvinge, 2018].

La liaison de la snRNP U6 au site 5' et son interaction avec la snRNP U2 liée au point de branchement, induisent la formation d'une structure en lasso de l'intron entre le site 5' et le point de branchement. [Wahl et al., 2009; Will and Lührmann, 2011]

La snRNP U5 favorise le rapprochement entre les deux exons. Cet assemblage conduit à la formation du complexe actif, ou complexe B*. C'est à cette étape qu'a lieu la première réaction de transestérification.

3.3.5. Changements conformationnels et formation du complexe catalytique C et du complexe post-spliceosomal

Des réarrangements structuraux conduisent à la formation du complexe C, où a lieu la deuxième réaction de transestérification. A cette étape, l'intron est libéré sous forme de lasso et les interactions entre UsnRNA sont rompues. Les exons sont rapprochés à l'aide de la snRNP U5 et les deux fragments sont joints par la deuxième réaction de transestérification [Will and Lührmann, 2011]

3.4. Complexité de la composition du spliceosome

La dynamique d'association et de dissociation du spliceosome dont la composition en protéines est sans cesse modifiée au cours de la maturation du pré-ARNm, témoigne d'une grande diversité d'interactions protéines-protéines et ARN-protéines. Les protéines impliquées dans la réaction d'épissage sont libérées et déstabilisées à chaque étape du processus (Figure 10). Par exemple, durant la transition du complexe A au complexe B, plus de 25 protéines composantes de la tri-snRNP U4/U6-U5 et plus de 35 protéines non-snRNP sont recrutées simultanément. Cette dynamique d'association-dissociation de complexes protéiques a tout d'abord été étudiée

sur des complexes purifiés à partir d'extraits nucléaires de cellules HeLa, par des techniques de chromatographie d'affinité et de filtration sur gel [Bennett et al., 1992; Reed et al., 1988]. Au total, la réaction d'épissage fait intervenir entre 150 et 300 protéines et complexes ribonucléoprotéiques [Jurica and Moore, 2003; Rappsilber et al., 2002]. Cette dynamique est indispensable à la reconnaissance des sites d'épissage, et à la réaction de transestérification. Elle contribue à pallier à la grande diversité des séquences définissant les sites d'épissage. Les protéines cœur essentielles à la réaction d'épissage sont ainsi supplémentées par des protéines régulatrices, principalement du fait de la dégénérescence des séquences qui définissent les introns [Fabrizio et al., 2009].

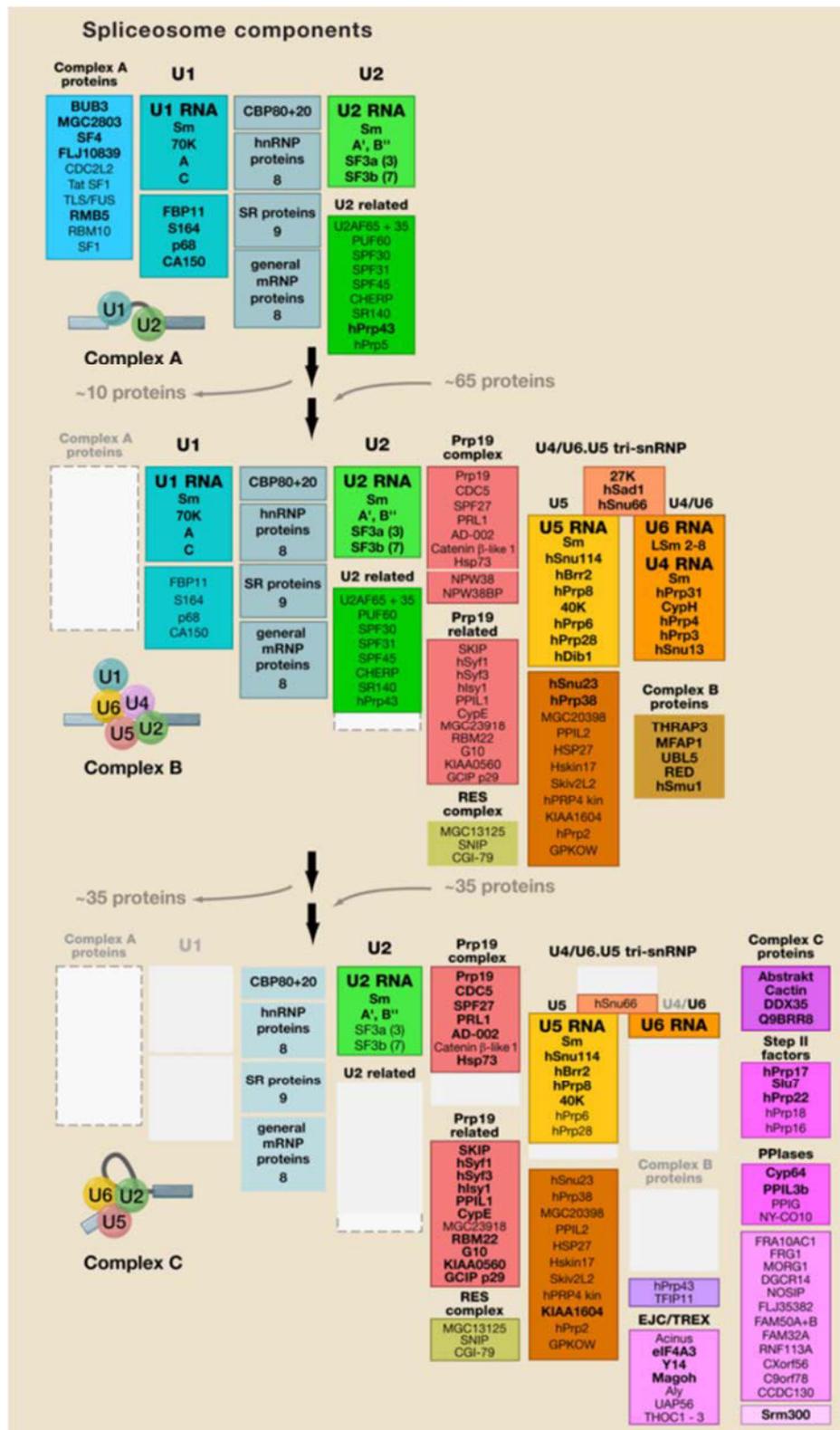


Figure 10. Schéma illustrant la complexité et la dynamique de la composition des complexes A, B et C du spliceosome chez l'Homme. Les protéines sont classées selon leur fonction, leur association stable avec les snRNP, ou leur présence dans un autre complexe spliceosomal. La composition en protéines des complexes a été déterminée par spectrométrie de masse. Les protéines présentes en quantités substoechiométriques, conformément au nombre de peptides détectés par spectrométrie de masse, sont grisés. (D'après Wahl et al, 2009)

Chapitre II : L'épissage alternatif

Les années 1970 ont certainement bouleversé la recherche, les connaissances et l'approche que l'on avait de l'ARN. Après la découverte des introns et de la réaction d'épissage en 1977 par Phillip Allen Sharp et Richard Roberts, l'équipe de Philip Allen Sharp expose quelques mois plus tard un autre mécanisme essentiel faisant intervenir des réarrangements des séquences de l'ARN : L'épissage alternatif [Berk and Sharp, 1978].

C'est en analysant les séquences de deux ARN viraux correspondants au petit et grand antigène T, produits lors de la phase précoce de l'infection par le virus SV40, que les chercheurs ont constaté que ces deux transcrits étaient originaires de la même séquence génomique. C'est à partir de ces observations que l'équipe de Phillip Allen Sharp décrit un mécanisme d'expression de gènes contrôlé en partie par l'épissage.

Quelques années plus tard, en 1982, l'équipe de RM. Evans [Amara et al., 1982] propose un modèle où la maturation des ARNm participe à l'augmentation de la diversité de l'expression des gènes chez les mammifères. Ce modèle est basé sur leur observation de la maturation du pré-ARNm issu du gène de la calcitonine. A partir de ce transcrit primaire, étaient produits deux ARNm matures. L'un codait pour l'hormone calcitonine présente dans la thyroïde et l'autre pour une protéine nommée CGRP pour *calcitonine gene-related peptide* qui était exprimée dans l'hypothalamus.

Dans la littérature, une distinction est toujours faite entre un épissage dit « constitutif » et un épissage dit « alternatif ». Pendant de longues années, la communauté scientifique a considéré l'épissage alternatif comme un mécanisme rare, survenant dans des cas particuliers. En 1994, Phillip Allen Sharp déclare dans son exposé pour la réception du prix Nobel de Physiologie à l'institut Karolinska à Stockholm, qu'un vingtième des transcrits humains étaient sujet à un épissage alternatif. En prenant pour exemple le gène de la fibronectine dont l'épissage alternatif permet de produire vingt protéines différentes, Phillip Allen Sharp illustre l'importance de l'épissage alternatif dans l'augmentation de la capacité codante du génome sans modifier la taille de celui-ci, dans la diversité protéique et dans la spécificité tissulaire.

A l'heure actuelle, les études de séquençage à haut débit des transcriptomes de tissus divers suggèreraient que les transcrits issus de l'épissage alternatif seraient largement plus répandus que ce qui était précédemment considéré. Chez l'homme, entre 92 et 94% des gènes multi-

exons sont sujets à un épissage alternatif dans un ou plusieurs types cellulaires [Wang et al., 2008].

Dans quelques années, la barrière entre les épissages constitutifs et alternatifs sera sans doute remise en question en faveur d'un mécanisme inclusif de maturation des ARN faisant intervenir des éléments modulateurs en fonction du contexte cellulaire.

1. Types d'épissages alternatifs

Les types d'épissages alternatifs sont classés en fonctions des caractéristiques de ou des exons qui sont sujets à épissage et par conséquent de la composition de l'ARNm mature résultant de la réaction d'épissage. La Figure 11 est une représentation schématique des cinq types d'épissages alternatifs dépendants de la nature de l'exon.

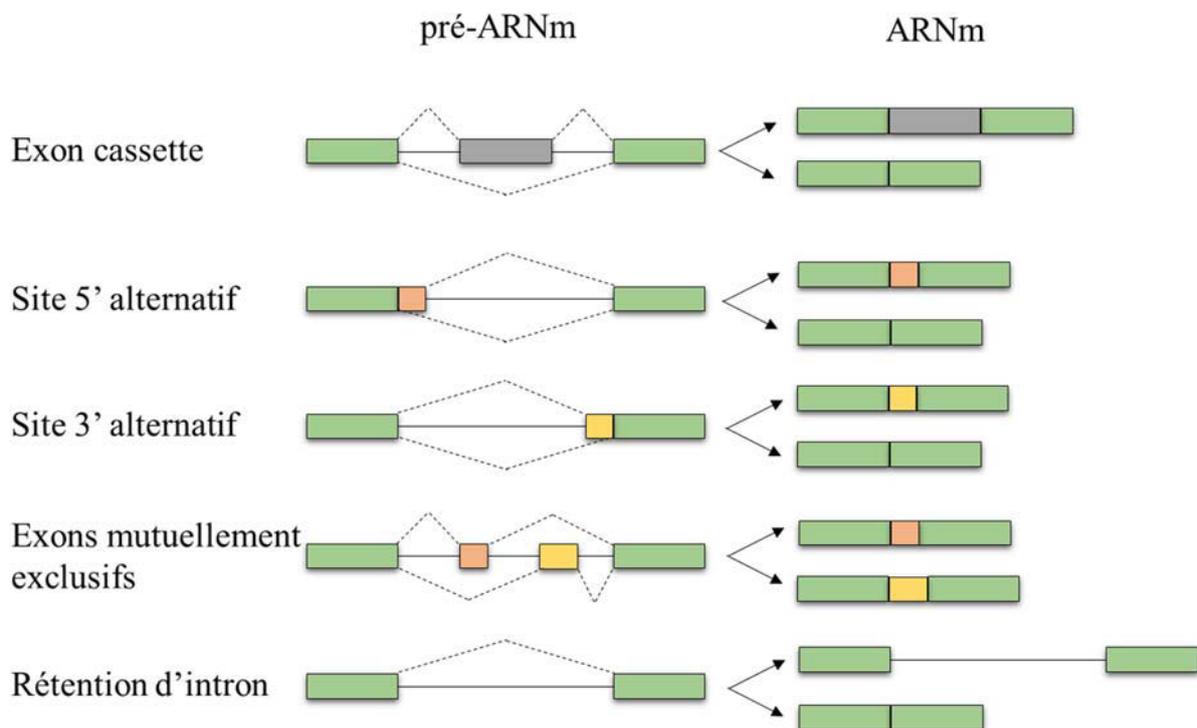


Figure 11. Modes d'épissage alternatif des pré-ARNm. Les exons constitutifs sont représentés en vert et les exons épissés alternativement sont représentés en gris, orange et jaune.

Dans le cas d'un exon cassette, un exon alternatif est soit intégré, soit éliminé lors de la réaction d'épissage. Ces types d'exons sont souvent tissu-spécifiques, car ce mode d'épissage peut conférer une propriété supplémentaire à la protéine en fonction du phénomène physiologique spécifique à la cellule qui compose le tissu, ou alors servir de régulateur dans des types cellulaires où la fonction de la protéine n'est pas nécessaire. C'est le cas par exemple de l'exon

N1 du gène de la tyrosine kinase SRC qui est inclus dans les cellules neuronales et exclu dans les cellules non-neuronales. Cet exon de 18 nucléotides situé entre les exons 3 et 4, forme le transcrit n-src dans les neurones, tandis que son excision forme le transcrit c-src dans les autres types cellulaires. L'inclusion de l'exon N1 dans les neurones conduit à l'insertion de 6 acides aminés dans le domaine SH3 de la protéine kinase et altère sa capacité d'interaction avec d'autres protéines [Brown and Cooper, 1996].

Les exons définis par des sites 5' ou 3' alternatifs possèdent dans le même intron plus d'une copie des séquences consensus qui définissent les sites d'épissage. Ces sites entrent en compétition lors de la définition des exons et la reconnaissance des sites d'épissage par le spliceosome. L'épissage de ces pré-ARNm conduit à la production d'ARNm matures de tailles différentes pouvant induire des décalages de cadre de lecture et donc, produire des protéines aux fonctions distinctes. Ce type d'épissage alternatif est souvent observé chez les virus [Ropers et al., 2004]. Chez l'Homme, il est présent dans certains ARNm, comme par exemple le pré-ARNm du gène Bcl-x1 où sont présents deux sites d'épissage en 5' [Boise et al., 1993] ou le pré-ARNm du gène PTB dont l'exon 9 contient deux sites d'épissage en 3' [Wollerton et al., 2001].

Les exons mutuellement exclusifs sont des exons qui ne sont jamais simultanément présents dans l'ARNm mature et dont les séquences se succèdent dans le pré-ARNm. Dans ce cas-là, les sites 5' et 3' d'épissage d'un exon sont en compétition avec les sites 5' et 3' d'un autre exon. Ces exons codent en général pour des domaines alternatifs de la protéine et sont exprimés de manière tissu-spécifique. C'est le cas notamment du pré-ARNm de la β -tropomyosine où les deux exons 6A et 6B sont exprimés différemment selon le type cellulaire. En effet, l'exon 6A est exprimé dans les cellules non-musculaires, tandis que l'exon 6B est exprimé dans les cellules musculaires [Libri et al., 1989]. C'est aussi le cas de l'exon 2 de l' α -tropomyosine qui est exprimé spécifiquement dans les cellules qui composent le muscle lisse, alors que l'exon 3 du gène est exprimé dans les autres tissus.

Les exons résultants de la rétention d'intron sont des introns non épissés car non reconnus par la machinerie d'épissage. Ces introns sont alors inclus dans la séquence de l'ARNm mature. C'est le cas par exemple de l'intron 3e de l'élément P chez la drosophile, dont l'excision produit une transposase dans les cellules germinales, tandis que sa rétention dans les cellules somatiques conduit à la production d'une protéine qui réprime la transposition [Laski et al., 1986].

2. Régulation de l'épissage alternatif

Chez les vertébrés, les séquences qui définissent les sites d'épissage sont souvent dégénérées et s'éloignent des séquences consensus [Wahl et al., 2009]. Ainsi, les séquences en 5', en 3' et autour du point de branchement ne sont pas suffisantes pour définir les introns et les exons des pré-ARNm. Pour faciliter la reconnaissance des sites d'épissage, des séquences ciblées par des protéines régulatrices sont présentes dans les exons et les introns.

L'épissage alternatif survient lorsque le spliceosome doit faire un choix entre plusieurs sites d'épissage en compétition. Le choix du spliceosome sera alors dépendant du contexte et des caractéristiques des séquences régulatrices qui favoriseront ou inhiberont la reconnaissance des sites d'épissage et l'assemblage de ce spliceosome. Ce choix peut également se produire de manière indirecte, faisant intervenir la machinerie transcriptionnelle ou des éléments épigénétiques.

Un autre aspect de régulation de l'épissage est apparu ces dernières années, impliquant les composants constitutifs du spliceosome qui étaient considérés précédemment comme des éléments communs à tous les épissages et qui par conséquent n'avaient pas ou peu de rôle dans la régulation de ces événements.

2.1. Les séquences régulatrices en *cis*

En fonction des caractéristiques des séquences régulatrices en *cis* du pré-ARNm, celles-ci sont classées en plusieurs catégories : les sites d'épissages consensus nécessaires à l'assemblage du spliceosome et à la réaction catalytique de l'épissage, les séquences reconnues par les facteurs d'épissage activateurs ou répresseurs et enfin les séquences qui conduisent à la formation de structures secondaires.

Le contexte dans lequel les séquences d'épissage consensus sont incluses est également important pour la réaction d'épissage. Les sites d'épissage sont souvent désignés comme « forts » ou « faibles » et coïncident fréquemment avec respectivement les mécanismes d'épissages constitutifs et alternatifs [Sugnet et al., 2004]. Les séquences entourant les sites faibles divergent de celles que l'on retrouve autour des sites forts, tout en gardant une forte conservation entre les espèces, ce qui suggère un rôle de ces séquences dans la régulation de l'épissage.

Les sites d'épissage faibles sont particulièrement présents dans les cas de sites 5' ou 3' alternatifs. Dans ce cas-là, une des deux jonctions intron/exon est plus difficilement reconnue par la machinerie d'épissage, induisant une forte dominance d'un site au détriment d'un autre. Deux sites d'épissage peuvent également être de force équivalente. Il en résulte alors un équilibre dans le choix d'un des deux sites d'épissage.

Les pré-ARNm présentant des sites d'épissage faibles contiennent également des éléments de séquences, qui agissent en *cis* pour stimuler ou réprimer l'épissage. Ces séquences sont des sites d'interaction avec des facteurs protéiques et agissent en facilitant ou en empêchant l'assemblage du spliceosome, ou encore en modifiant l'accessibilité de séquences adjacentes dans le pré-ARNm.

Ces séquences sont classées selon leur localisation dans l'ARN et leurs caractéristiques activatrices ou inhibitrices (Figure 12). Les séquences interagissant avec les protéines activatrices sont appelées ISE, pour "*Intronic Splicing Enhancer*" ou ESE, pour "*Exonic Splicing Enhancer*" tandis que les séquences interagissant avec des protéines inhibitrices sont appelées ISS, pour "*Intronic Splicing Silencer*" et ESS, pour "*Exonic Splicing Silencer*".

Elles sont reconnues par des protéines régulatrices de l'épissage, parmi lesquelles les protéines SR dont les fonctions sont principalement activatrices et les hnRNPs qui sont inhibitrices de l'épissage [Chen and Manley, 2009].

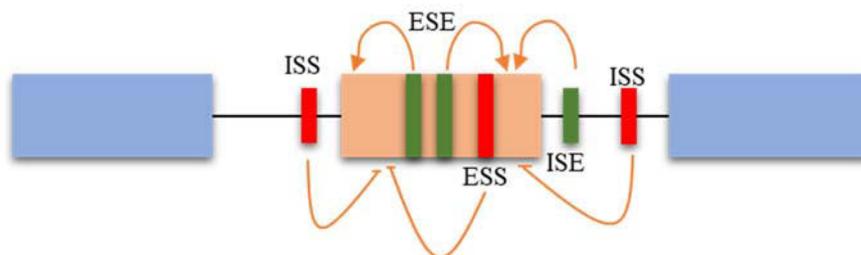


Figure 12. Séquences *cis* impliquées dans la régulation de l'épissage Alternatif. En plus des séquences consensus des sites d'épissage, des séquences régulatrices sont présentes dans les introns et les exons. Elles sont activatrices (ESE et ISE) ou inhibitrices (ESS et ISS). ESE: *Exonic Splicing Enhancer*; ISE: *Intronic Splicing Enhancer*; ESS: *Exonic Splicing Silencer*; ISS: *Intronic Splicing Silencer*. (D'après Chen et Manley, 2009)

Cependant, il est difficile de caractériser précisément ces séquences régulatrices, car il n'y a pas de séquences consensus proprement dites. Plusieurs approches ont été utilisées afin d'identifier les séquences en *cis*, notamment l'approche de Schaal & Maniatis, [Schaal and Maniatis, 1999] appelée méthode SELEX, qui a permis d'identifier les séquences activatrices de l'épissage (Figure 13). Cette méthode repose sur l'identification *in vitro* de séquences d'ARN activatrices de l'épissage par une technique de sélection exponentielle. Pour cela, des

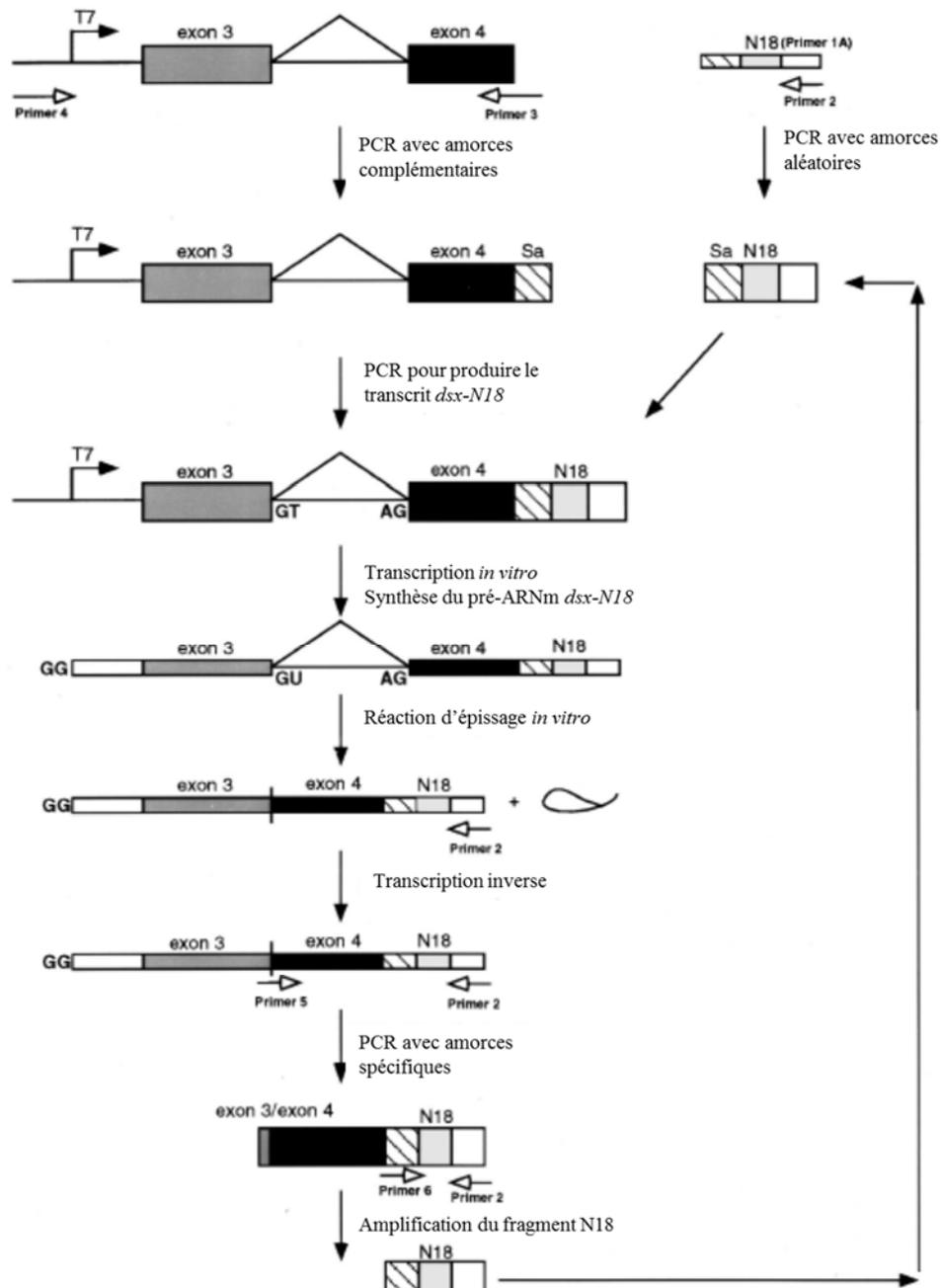


Figure 13. Méthode SELEX d'identification des séquences activatrices de l'épissage de Schaal & Maniatis, 1999.

séquences aléatoires de 18 nucléotides (N18) ont été insérées 90 nucléotides en aval du site d'épissage en 3' du pré-ARNm *dsx* de la drosophile, à la position d'un site ESE originellement présent dans la séquence du gène. Les $6,9 \times 10^6$ séquences obtenues ont ensuite été transcrites *in vitro* afin de générer une population de transcrits primaires *dsx-N18*. Ces transcrits primaires sont ensuite incubés dans des extraits nucléaires afin de réaliser l'épissage *in vitro*. Les ARNm produits grâce à la présence de séquences activatrices sont ensuite isolés et amplifiés par RT-PCR en utilisant des amorces spécifiques des jonctions d'épissage. La séquence activatrice est ensuite isolée à son tour par PCR puis clonée dans le pré-ARNm *dsx*, afin de débiter un nouveau cycle de sélection. Après six cycles de sélection, les séquences consensus activatrices de l'épissage ont été mises en évidence et classées en 7 groupes (Tableau 2). Ces séquences activatrices se sont avérées être aussi bien riches en purines qu'en pyrimidines.

Tableau 2 : Séquences consensus activatrices de l'épissage découvertes par la méthode SELEX, d'après Schaal & Maniatis 1999.

Type d'activateur	Séquence consensus
Classe I Séquences riches en purines	AGGAG
	GGGGA
	GGAGGA
	GGAGA
	AGAGG
Classe II Séquences riches en pyrimidines	UCUCC
	UCUUC
	UCCUC
Classe III	(U)GGACCNG
Classe IV	(C)CACC(C)
Classe V	YGCCGCC
Classe VI	RGAACYU
Classe VII Activateurs forts sans consensus	CCGACGCCAUGGACGACGAG
	GGCUGCCAGACGGAAUUGG
	CCGUGACAGCAUCGGCCGG
	CGUCGGCAGGUGGUCCCG
	UCUGGAUCCUGCGGAUGG

Une étude similaire a été réalisée afin d'identifier les séquences inhibitrices et leurs partenaires protéiques [Fairbrother and Chasin, 2000]. Des séquences de 100 pb ont été insérées dans l'exon cassette d'un minigène à trois exons et le profil d'épissage du minigène a ensuite été analysé dans les cellules. Cette étude a permis d'identifier plusieurs types de séquences inhibitrices : parmi elles, des « pseudo-sites d'épissage » correspondant à des séquences ayant de fortes similitudes avec les sites d'épissage canoniques et qui fixent des éléments du spliceosome afin

de le rendre inactif, des séquences inhibitrices riches en pyrimidines et enfin des motifs YAG et GGG qui correspondent à des sites de fixation des hnRNP.

D'autres séquences inhibitrices ont ensuite été identifiées. C'est le cas des séquences riches en répétition du dinucléotide CA qui sont spécifiquement des sites de liaisons pour la protéine hnRNP L [Hui et al., 2005].

Les motifs de régulation de l'épissage peuvent être de petite taille et présenter des séquences dégénérées, comme c'est le cas du motif YCAY reconnu par les protéines régulatrices NOVA dont le mécanisme peut être activateur ou inhibiteur de l'épissage [Ule et al., 2005].

La formation de structures secondaires par le pré-ARNm peut également jouer un rôle dans la modulation de l'épissage en jouant sur l'accessibilité du spliceosome aux sites d'épissage ou des protéines liant l'ARN aux séquences régulatrices et ainsi agir comme élément inhibiteur ou activateur de l'épissage [Estes et al., 1992]. Ceci est notamment le cas dans la régulation de l'épissage de hGH, l'hormone de croissance humaine qui est régulée par la formation d'une structure en tige-boucle sur l'ARN, contrôlant l'utilisation de deux sites accepteurs de l'épissage en 5' correspondant à deux isoformes distincts hGH-N et hGH-V de la même protéine [Estes et al., 1992].

2.2. Les protéines majeures de la régulation de l'épissage

Deux principales familles de protéines sont impliquées dans la régulation de l'épissage alternatif : les protéines SR [Zhou and Fu, 2013] et les hnRNP [Han et al., 2010]. Les constituants de ces deux familles de protéines se lient à des séquences spécifiques présentes dans le pré-ARNm.

2.2.1. Les protéines SR

Comme décrit précédemment dans le chapitre I, les protéines SR jouent un rôle essentiel à la fois dans la définition des exons et dans l'assemblage de la machinerie d'épissage, mais également dans la régulation de cet assemblage. Les protéines SR sont principalement activatrices de l'épissage des pré-ARNm et se fixent à des séquences spécifiques dans l'exon.

2.2.1.2. Sites de fixation des protéines SR

Les études permettant la mise en évidence de séquences spécifiques de fixation des protéines SR se basent sur la technique SELEX. Cette approche consiste à sélectionner les sites de liaison de haute affinité de la protéine SR étudiée parmi un ensemble de combinaisons de séquences

d'ARN. Les séquences sélectionnées sont ensuite introduites dans un minigène afin d'identifier leur rôle dans la régulation de l'épissage en présence de la protéine SR étudiée. C'est ainsi que les séquences spécifiques de chaque protéine SR ont été identifiées (Tableau 3) [Schaal and Maniatis, 1999].

Cependant, malgré les études approfondies sur ces modèles de minigènes afin d'identifier les séquences cibles des protéines SR, des obstacles mécanistiques restent encore à surmonter. Même s'il est à présent possible de prédire les sites de liaisons de la plupart des protéines SR en fonction des motifs consensus identifiés *in vitro* [Cartegni et al., 2003] ainsi que par des analyses *in vivo* de CLIP-seq [Ankō and Neugebauer, 2012; Pandit et al., 2013a; Sanford et al., 2009], l'interaction des protéines SR avec l'ARN dépend de plusieurs facteurs tels que la structure secondaire de l'ARN et une compétition avec d'autres protéines régulatrices liant l'ARN [Long and Caceres, 2009].

Des études récentes tentant d'adresser le problème de spécificité de liaison des protéines SR, indiquent qu'il serait possible d'identifier ces interactions par des analyses structurales [Daubner et al., 2013] couplées à une cartographie des interactions ARN-protéine à l'échelle génomique [Pandit et al., 2013b].

Tableau 3. Séquences reconnues par les protéines SR (pour revue, Long and Caceres, 2009 SELEX de liaison (SL) ou SELEX fonctionnel (SF) en extrait nucléaire (EN) ou cytoplasmique (S100) par CLIP (C), EMSA (E), test d'épissage (TE) ou pontage UV (pUV). R: A/G; Y: C/U; S: C/G; K: G/U; W: A/U; M: A/C; D: A/G/U; N: nucléotide quelconque.

Nouvelle nomenclature	Sites de fixation identifiés
SRSF1	RGAAGAAC (SL) AGGACRRAGC (SL) SRSASGA (S100) (SF) UGRWG (C)
SRSF2	AGSAGAGUA (SL) GUUCGAGUA (SL) GRYY C/A G/C YR (S100) (SF) UGUUCSAGWU (SL) GWUWCCUGCUA (SL) GGGUAUGCUG (SL) GAGCAGUAGKS (SL) AGGAGAU (SL) UGCNGYY (EN) (SF)
SRSF3	GGUCCUCUUC (E) WCWWC (TE)

	CUCKUCY (EN) (SF)
SRSF4	GAAGGA (pUV)
SRSF5	GAGCAGUCGGCUC (SL) ACDGS (SI00) (SF)
SRSF6	USCGKM (SI00) (SF) UCAACCAGGCGAC (SL)
SRSF7	UCAACA (pUV) ACGAGAGAY (SL) GGACGACGAG (EN) (SF)
SRSF8	GACGAC (SF) AAAGAGCUCGG (SF) CUGGAUU (E)
SRSF9	GACGAC (SF) AAAGAGCUCGG (SF) CUGGAUU (E)
SRSF10	ACAAAGACAAA (SL)
SRSF11	Riche en C (pUV)
SRSF12	ND

2.2.1.3. Rôle des protéines SR dans la régulation de l'épissage

Le rôle et le mécanisme d'action des protéines SR dans la régulation de l'épissage est l'un des plus étudiés. De nombreuses études biochimiques ont mis en évidence la capacité des protéines SR à promouvoir la sélection des sites d'épissage par leur interaction avec des sites ESE [Cavaloc et al., 1999; Liu et al., 2000; Schaal and Maniatis, 1999].

Cependant, des études récentes ont révélé que le rôle des protéines SR ne se limitait pas seulement à l'activation de l'inclusion de l'exon, mais qu'elles pouvaient également dans certains cas induire des sauts d'exon et ce en interagissant avec le pré-ARNm au niveau des sites ISS [Dembowski et al., 2012; Erkelenz et al., 2013]. Cette dualité fonctionnelle des protéines SR peut s'expliquer en appréhendant le mécanisme d'épissage alternatif comme avant tout un choix entre deux sites d'épissage en compétition.

Ceci est notamment illustrée dans les travaux de l'équipe de Y.Feng [Zhou et al., 2014] qui a analysé les événements d'épissage régulés par la protéine SRSF10. Par des analyses du transcriptome par RNAseq des cellules où le taux de SRSF10 a été préalablement diminué, suivi de la recherche des séquences de liaison de la protéine par des analyses bio-informatiques

à partir des séquences consensus identifiées par SELEX , l'équipe a observé dans certains évènements d'épissage que la liaison de la protéine à l'exon cassette induisait une inclusion de cet exon. Cependant, dans d'autres évènements d'épissage, la même protéine se liait à l'exon constitutif en aval de l'exon cassette et induisait son exclusion (Figure 14).

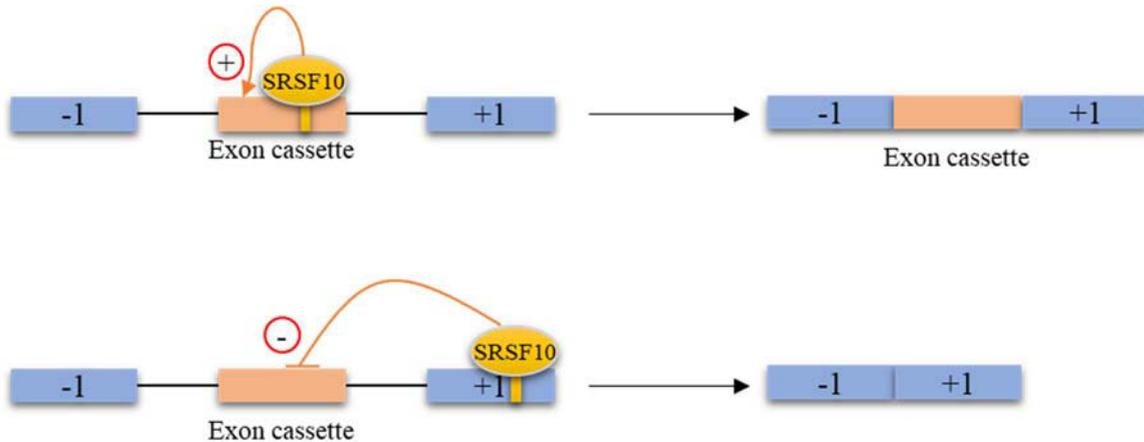


Figure 16. Dualité fonctionnelle des protéines SR. Lorsque la protéine SRSF10 se lie à une séquence spécifique dans l'exon alternatif, elle induit l'inclusion de l'exon. Cependant, lorsqu'elle se lie à une séquence spécifique dans l'exon constitutif en aval de l'exon alternatif, elle inhibe l'inclusion de cet exon. (D'après Zhou et Al, 2014)

Le phénomène complexe de régulation de l'épissage alternatif par les protéines SR pourrait alors dépendre de plusieurs paramètres combinant plusieurs mécanismes de contrôle de l'épissage. Ces protéines SR interagiraient à la fois entre elles et avec l'ARN au niveau des sites de régulation spécifiques et agiraient ainsi en synergique ou entreraient en compétition.

2.2.2. Les hnRNP

Les hnRNP pour *heterogeneous nuclear RiboNucleoProteins* sont des complexes composés de protéines assemblées autour du pré-ARNm. Ces complexes sont formés immédiatement après la transcription. Les protéines qui constituent ces complexes sont également appelées hnRNP [Görlach et al., 1993]. Contrairement aux protéines SR, les hnRNP sont classiquement considérées comme répresseurs de l'épissage et sont fréquemment liées aux sites ISS.

2.2.2.1. Définition et structure des hnRNP

La famille des hnRNP compte plus de 25 protéines hnRNP différentes. Elles présentent des structures très conservées et des poids moléculaires variables entre 34 et 120 kDa (Tableau 4). Elles sont classées par ordre alphabétique de A à U. Ces protéines ont été identifiées grâce à des expériences d'immuno-sélection, de pull down et de cross-link par UV [van Eekelen et al.,

1981; Görlach et al., 1993]. Certaines des protéines hnRNP tel que les hnRNP A1 et C1 font partie des protéines les plus abondantes dans le noyau. Elles sont majoritairement retrouvées dans le nucléoplasme [Piñol-Roma and Dreyfuss, 1993], mais certaines d'entre elles (les hnRNP A/B, D, E, J, et K) transitent entre le noyau et le cytoplasme. Dans le noyau, elles s'associent aux pré-ARNm et dans le cytoplasme aux ARNm matures.

Les hnRNP interviennent principalement dans la structuration, la maturation et l'épissage alternatif des ARNm. Elles jouent également un rôle dans l'activation de la transcription et le remodelage de la chromatine. Dans le cytoplasme, elles agissent sur la stabilité, le transport et la traduction des ARN [Han et al., 2010].

Elles possèdent dans leur séquence en N-terminal un signal d'import nucléaire M9. Ce domaine interagit avec la transportine, un récepteur d'import nucléaire. Pour certaines hnRNP, comme la hnRNP A1, le domaine M9 peut également agir comme un signal d'export nucléaire. La hnRNP A1 peut alors sortir du noyau en étant associée à l'ARN.

Quatre domaines de fixation à l'ARN ont été identifiés dans les séquences des hnRNP. Ces domaines sont de types RRM, quasi-RRM, RGG (Arginine-Glycine-Glycine) ou KH. Le domaine le plus fréquent est le domaine RRM, composé de quatre feuillets β et deux hélices α avec deux séquences conservées RNP1 et RNP2. Une boucle variable connectant les feuillets β est responsable de la spécificité d'interaction à l'ARN [Dvinge, 2018].

Le domaine RGG est caractérisé par des clusters adjacents contenant des résidus aromatiques compris entre deux tripeptides Arg-Gly-Gly (RGG) [Kiledjian and Dreyfuss, 1992]. Ce domaine est souvent considéré comme auxiliaire et interviendrait dans les interactions homologues ou hétérologues entre hnRNPs [Cartegni et al., 1996].

Le domaine KH (*pour hnRNP K-Homology*) est caractérisé par 3 brins β antiparallèles faisant face à trois hélices α . Une boucle flexible qui lie deux hélices α est responsable de l'interaction spécifique avec l'ARN [Lewis et al., 2000].

Tableau 4. Classification des protéines hnRNP : nom et représentation schématique des structures de chaque famille d'isoformes, RRM : RNA Recognition Motif ; RRM H : RRM Homologue ; Gly : riches en répétitions Glycine ; KH : K Homology. En plus des domaines communs, les protéines hnRNP présentent d'autres domaines spécifiques (en jaune) (d'après Han et al., 2010).

Nom	Domaines protéiques
A/B	RRM RRM GLY
C1/C2	RRM [jaune]
D1 à D4	RRM RRM GLY
E1, E2	KH KH KH
F	RRMH RRMH RRMH
G	RRM GLY
H, H'	RRMH RRMH GLY RRMH GLY
H3 (2H9)	RRMH RRMH
HAP (SAF-B)	RRM [jaune] [jaune]
I (PTB)	RRMH RRMH RRMH RRMH
K1, K2	KH KH GLY KH
L	GLY RRMH RRMH RRMH RRMH
M1 à M4	RRM RRM [jaune] RRM
P	GLY RRM GLY
Q1 à Q5	RRM RRM RRM GLY
R	[jaune] RRM RRM RRM GLY [jaune]
U (SAF-A)	[jaune] [jaune] GLY GLY

Les hnRNP possèdent également des domaines auxiliaires tels que les domaines riches en glycine qui confèrent une flexibilité à la protéine. La spécificité de liaison des hnRNP à l'ARN est principalement générée par la structure 3D de la protéine alors que les régions flanquant les domaines de liaison à l'ARN renforcent l'interaction [Dreyfuss et al., 2002].

2.2.2.2. Sites de fixation des hnRNP

Les hnRNP présentent une grande spécificité de reconnaissance de l'ARN, ce qui leur permet de s'associer à tous les pré-ARNm (Tableau 5). Certaines reconnaissent préférentiellement des

séquences riches en cytosines (hnRNP E, 1 et K) et d'autres des séquences riches en guanines (hnRNP NB, E, H, H', 2H9 et M). Par ailleurs, certaines hnRNP comme la hnRNP B, reconnaissent des séquences spécifiques avec une très forte affinité [Dreyfuss et al., 2002].

Tableau 5. Séquences de fixation des hnRNP sur l'ARN : établies par SELEX (d'après Han et al., 2010).

Nom	Sites de fixation
A/B	G4(AU3)4A (A0) UAGGG, (Py)UAG (A1, A1B) GCCAAGGAGCC (A2, B1) GCCAAGGAGCC(A3)
C1/C2	Riche en U, AGUA(U)SGUGGA
D1 à D4	Riche en A/U, UUAG/G
E1, E2	Poly(C), Poly(G)
F	Poly(G)
G	AAGU
H, H'	GGGA/C, Poly(G)
H3 (2H9)	GGGA
HAP (SAF-B)	ND
I (PTB)	Riche en C/U
K1, K2	Poly(C), AU(C)3/4(U/A)2
L	Répétitions C/A
M1 à M4	GAGGAGG
P	ND
Q1 à Q5	ND
R	ND
U (SAF-A)	Poly(G), riche en A/U

2.2.2.3. Rôle des hnRNP dans la régulation de l'épissage

La formation de complexes ribonucléoprotéiques entre les hnRNP et le pré-ARNm influence la fixation des facteurs d'épissage et l'assemblage du spliceosome. Les hnRNP ont ainsi principalement un rôle inhibiteur dans l'épissage car elles empêchent la fixation des protéines constituantes du spliceosome. Au niveau de l'intron, elles interagissent avec les séquences ISS et dans l'exon avec des séquences ESS. Elles agissent également sur les sites définissant les sites d'épissage en 5' et en 3'.

La hnRNP A1 par exemple, agit sur l'épissage en oligomérisant sur des séquences ISS contenant des tri-nucléotides UAG. Cette oligomérisation engendre des encombrements stériques sur le pré-ARNm, masque les sites de liaison des UsnRNP et des facteurs d'épissage et empêche ainsi l'assemblage du spliceosome sur les sites d'épissage ou au niveau du point de branchement [Hallay et al., 2006].

Les hnRNP peuvent également réguler l'épissage en formant des boucles d'ARN et ainsi séquestrer les sites d'épissage. C'est le cas des hnRNP H, PTB (hnRNP I) ou de la hnRNP A1 qui régule l'épissage de son propre ARNm. L'exon alternatif 7b de la hnRNP A1 est flanqué par deux introns possédant des sites de haute affinité pour la protéine. La hnRNP A1 se fixe sur ces sites de haute affinité. La dimérisation des deux protéines entraîne ensuite la formation d'une boucle engendrant une séquestration spatiale de l'exon alternatif, ce qui induit le rapprochement des sites 5' et 3' d'épissage des exons 7 et 8 et l'excision de l'exon 7b [Martinez-Contreras et al., 2006].

Certaines hnRNP sont impliquées à la fois dans l'activation et l'inhibition de l'épissage. C'est le cas des protéines hnRNP H/F qui présentent une dualité fonctionnelle dépendante de la cible d'épissage. En effet, lors de l'épissage de l'exon N1 du pré-ARNm *c-src*, les hnRNP H/F interagissent avec la protéine KSRP et se lient à une séquence intronique riche en guanines. Ce complexe active alors l'inclusion de l'exon NI alternatif [Markovtsov et al., 2000]. L'effet inverse est observé dans le cas de la régulation de l'épissage alternatif de l'exon 7 du pré-ARNm de la tropomyosine de rat. En effet, en se fixant à un site ESS, la hnRNP H empêche l'inclusion de l'exon alternatif [Chen et al., 1999].

La hnRNP C est, quant à elle, un facteur d'activation général de l'épissage. Les 2 isoformes de la protéine, hnRNP C1 et C2, s'associent pour former un tétramère qui se fixe de manière coopérative aux ARN nouvellement synthétisés, indépendamment de leur séquence [Görlach et al., 1994]. Les interactions avec l'ARN se produisent par l'intermédiaire d'un domaine RRM

et d'un domaine leucine zipper sur des séquences du pré-ARNm riches en pyrimidines [Görlach et al., 1994]

2.2.3. Dualité d'interactions entre protéines SR et hnRNP

Comme nous venons de le voir, les protéines SR étaient au départ considérées comme des éléments activateurs de l'épissage tandis que les hnRNP étaient plus souvent identifiées comme des répresseurs.

Cependant, cette séparation entre régulateurs positifs d'un côté et régulateur négatif de l'autre s'est avérée peu représentative de la réalité. La fonction des deux familles de protéines varie considérablement. L'un des facteurs dépendants de cette variation est l'effet spécifique de chaque protéine en fonction du site de liaison. C'est le cas notamment de la protéine SRSF10 qui augmente l'inclusion d'un exon lorsqu'elle est liée à celui-ci, mais a un effet opposé lorsqu'elle se lie à un autre exon en aval [Zhou et al., 2014] C'est le cas également de la hnRNP L qui se lie à la fois à une séquence ESS et diminue l'inclusion de l'exon [Rothrock et al., 2005], mais peut aussi interagir avec des séquences ISS et ISE et ainsi induire ou réprimer l'inclusion de l'exon en aval [Hui et al., 2003, 2005].

Par ailleurs, des effets coopératifs ou compétitifs entre les membres des protéines SR et hnRNP entrent en jeu dans la régulation de l'épissage. Par exemple, la hnRNP A1 peut contrecarrer l'effet d'une protéine SR en limitant sa fixation à l'ARN. De même, l'oligomérisation de la hnRNP A1 le long de l'ARNm empêche la snRNP U1 d'accéder au site d'épissage [Eperon et al., 2000]. C'est le cas également lorsqu'il y a juxtaposition entre une séquence ESE recrutant une protéine SR, et une séquence ESS recrutant une protéine hnRNP [Smith and Valcárcel, 2000]. Les deux protéines entrent alors en compétition. C'est alors le ratio de ces protéines dans la cellule qui détermine quel effet sera dominant sur l'épissage. Ainsi la régulation de l'épissage alternatif dépend du ratio SR / hnRNP présent dans les cellules. Ce ratio peut alors varier en fonction du tissu, ou du processus physiologique.

2.3. Autres facteurs impliqués dans la régulation de l'épissage

2.3.1. Régulation de l'épissage par des facteurs tissu-spécifiques

La nature des isoformes d'ARN retrouvés dans la cellule est extrêmement variable en fonction du tissu. Il est donc évident que l'on retrouve des facteurs régulateurs de l'épissage qui agissent spécifiquement dans des tissus donnés, induisant une régulation tissu-spécifique de leurs pré-ARNm cibles. C'est le cas des familles de protéines NOVA, RBFOX et CELF.

Les protéines neuronales NOVA-1 et NOVA-2 font partie des régulateurs de l'épissage tissu-spécifique les plus étudiés. Ces deux protéines qui possèdent des domaines de liaison à l'ARN de type KH, sont presque exclusivement exprimées dans les neurones du système nerveux central et régulent l'épissage alternatif de cibles spécifiques [Ule et al., 2005].

La protéine PTBP2 pour *pyrimidine tract binding protein 2* est un homologue de la protéine PTB. Cette protéine est spécifiquement présente dans les neurones et intervient dans la régulation de l'épissage alternatif de cibles neuronales [Wollerton et al., 2001].

Les facteurs d'épissage tissu-spécifique peuvent être à la fois activateurs et répresseurs, selon la nature du pré-ARNm et de la présence ou non d'autres facteurs de régulation de l'épissage. Certains facteurs tissu-spécifique, comme c'est le cas des protéines de la famille NOVA, se lient aux séquences introniques ISE et induisent ainsi l'assemblage du spliceosome.

Un autre exemple est le cas de la protéine MBNL-1, pour *muscleblind-like protein-1* qui active l'inclusion de l'exon en se liant au site 3' d'épissage, mais qui a un effet inverse lorsqu'elle interagit avec la séquence 5' en aval de l'exon alternatif [Konieczny et al., 2014]. MBNL-1 peut également interagir avec les structures secondaires de l'ARN, notamment les structures introniques en tige-boucle et ainsi bloquer l'accès du facteur d'épissage U2AF⁶⁵ au site de Polypyrimidine lors des étapes précoces de l'assemblage du spliceosome [Warf and Berglund, 2010].

La plupart des facteurs d'épissage tissu-spécifique étudiés étaient jusque-là impliqués dans le développement des tissus musculaires ou des neurones. Ceci est principalement dû à leur identification par l'intermédiaire de leur implication dans des maladies telles que les dystrophies musculaires ou les troubles neurodégénératifs. Cependant, l'émergence ces dernières années d'un criblage systématique de protéines liant l'ARN a permis d'identifier un grand nombre de régulateurs de l'épissage autres que neuronaux [Grosso et al., 2008]. Il reste

encore cependant à identifier les mécanismes d'action de ces protéines sur le pré-ARNm et leurs rôles dans la différenciation et la spécificité du tissu.

2.3.2. Influence d'éléments extérieurs

Dans un environnement cellulaire, l'épissage est sujet à de nombreuses pressions résultant des interactions croisées entre plusieurs processus physiologiques qui se déroulent de manière simultanée. La régulation de l'épissage alternatif des pré-ARNm peut alors être influencée par la transcription, des facteurs épigénétiques tels que le positionnement du nucléosome, la modification post traductionnelle des histones ou la méthylation de l'ADN.

2.3.2.1. Influence de la transcription

Le mécanisme d'épissage est étroitement lié à la transcription, la majorité des épissages chez l'homme se déroulant de manière co-transcriptionnelle. Le recrutement de facteurs régulateurs de l'épissage et l'assemblage du spliceosome sur le pré-ARNm sont alors initiés simultanément à la transcription de l'ARN naissant par le complexe de l'ARN polymérase II. Deux modèles ont été proposés, expliquant l'effet de la transcription sur la régulation de l'épissage (Figure 15).

Le modèle « cinétique » [de la Mata et al., 2003] propose que des modulations de la vitesse d'élongation de l'ARN polymérase II pourrait induire ou empêcher la reconnaissance de sites d'épissages compétiteurs et des séquences régulatrices. Par conséquent, plus l'élongation par l'ARN polymérase II est rapide, plus l'exon aura des chances d'être exclu. Par ailleurs, une transcription lente pourrait promouvoir le recrutement du spliceosome à des sites d'épissage faibles et ainsi induire l'inclusion d'un exon.

La vitesse de synthèse du pré-ARNm peut aussi influencer la formation de structures secondaires, et par conséquent moduler l'accessibilité à des sites de régulation, induisant l'activation ou la répression de l'épissage [Eperon et al., 1988].

Le modèle « de recrutement » quant à lui, propose une régulation de l'épissage alternatif par des recrutements directs ou indirects de facteurs d'épissage agissant en *trans* sur l'ARNm naissant. C'est le cas par exemple des protéines SR, qui sont recrutées par l'ARN polymérase II de manière co-transcriptionnelle sur l'ARNm naissant [Das et al., 2007]. Ce recrutement s'effectue par l'intermédiaire du domaine CTD de la sous-unité catalytique de l'ARN polymérase II. Par ailleurs, le recrutement de certaines protéines composantes des UsnRNP U1

et U2, ainsi que le facteur d'épissage U2AF, s'effectue également par l'intermédiaire du domaine CTD de l'ARN polymérase II [Hsin and Manley, 2012].

2.3.2.2. Influence des facteurs épigénétiques

L'architecture de la chromatine, les modifications post-traductionnelles des histones et de l'ADN peuvent avoir un impact sur la régulation de l'épissage alternatif (Figure 15). En effet, Un grand nombre de facteurs impliqués dans le remodelage de la chromatine sont associés aux complexes spliceosomaux et peuvent impacter l'épissage des pré-ARNm [Allemand et al., 2016].

Par ailleurs, des études ont montré que les exons étaient plus enrichis en nucléosomes, par comparaison aux introns [Dvinge, 2018]. Ce positionnement particulier des nucléosomes pourrait alors jouer un rôle dans la définition des exons et pourrait intervenir dans la modulation de la vitesse d'élongation en forçant le ralentissement de l'ARN polymérase II et de ce fait, influencer le choix des sites d'épissage. Le complexe de remodelage du nucléosome SW1/SNF

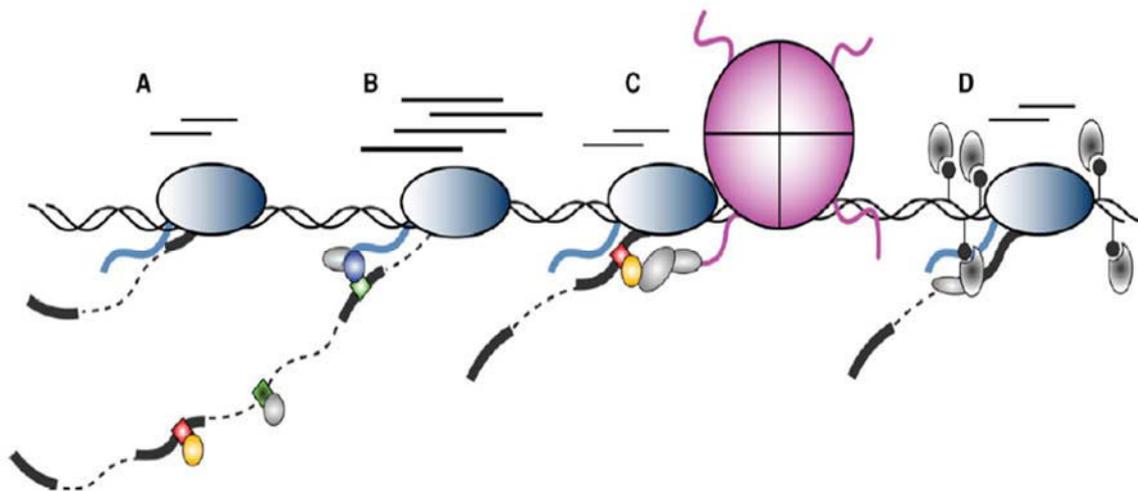


Figure 17. Régulation de l'épissage alternatif par la machinerie transcriptionnelle et les modifications épigénétiques. A : Une élongation lente de l'ARN polymérase II (bleu) favorise la reconnaissance des sites d'épissage faibles. Les lignes au-dessus de l'ARN polymérase II représentent la vitesse d'élongation relative. Sur l'ARN naissant, les traits gros représentent les exons et les pointillés les introns. B : Une ARN polymérase II plus rapide conduit à plus de sites d'épissages et au recrutement de plus de facteurs d'épissage sur le préARNm induisant la reconnaissance préférentielle de sites d'épissage forts. Le domaine CTD de l'ARN polymérase II (ligne bleue) peut également recruter des facteurs d'épissage sur le préARNm, indépendamment de la vitesse d'élongation. C : la présence d'un nucléosome (violet) ralentit la vitesse de transcription de l'ARN polymérase II et agit directement et indirectement dans le recrutement de facteurs d'épissages. D : Des méthylations de l'ADN (cercles noirs et trait fins) influencent la vitesse d'élongation et de recrutement de facteurs d'épissage via des protéines qui se lient aux nucléotides méthylés. (Pour revue, d'après Heidi Dvinge, 2018)

peut également agir comme régulateur de l'épissage en réduisant la vitesse d'élongation de l'ARN polymérase II et ainsi faciliter l'inclusion des exons avec des sites d'épissages faibles [Batsché et al., 2006].

Tout comme les nucléosomes, des histones présentant des modifications post-traductionnelles distinctes sont retrouvés spécifiquement dans les exons. Ces modifications épigénétiques peuvent influencer la reconnaissance des sites d'épissage en recrutant des facteurs d'épissage de manière directe ou indirecte.

C'est notamment le cas de la triméthylation de l'histone H3K4 qui favorise l'épissage en recrutant la snRNP U2 sur le pré-ARNm via des interactions avec l'ATPase CHD1 [Sims et al., 2007]. Les histones peuvent également intervenir dans le recrutement de protéines régulatrices, tel que la SRSF1 qui interagit avec la lysine triméthylée de H3K4. La méthylation des histones peut également impacter l'épissage via d'autres mécanismes, comme c'est le cas de la triméthylation de l'histone H3 (H3K9, H3K27 et H3K36) qui favorise la rétention d'introns en empêchant la fixation de la snRNP U5 et par conséquent la réaction d'épissage.

L'acétylation des histones peut également influencer l'épissage. En effet, l'activité des histones désacétylases HDAC peut impacter l'épissage alternatif en modulant la vitesse de transcription de l'ARN polymérase II mais également en recrutant des facteurs de régulation tels que SRSF5 sur leurs sites de fixation sur le pré-ARNm [Hnilicová et al., 2011].

Enfin, la méthylation de l'ADN peut elle aussi influencer la régulation de l'épissage alternatif, par exemple en modulant la transcription ou encore le recrutement des facteurs d'épissage. L'épissage alternatif peut de ce fait être influencé par l'état de méthylation d'un gène et ce d'une part, par la modulation de l'élongation durant la transcription, et d'autre part par la régulation du recrutement des facteurs d'épissage.

Par ailleurs, l'état de méthylation d'un gène est étroitement lié à d'autres facteurs épigénétiques. En effet, la méthylation de l'ADN influe sur le positionnement du nucléosome [Chodavarapu et al., 2010] et sur les modifications post-traductionnelles des histones et ainsi indirectement réguler la vitesse de l'élongation durant la transcription.

L'épissage alternatif peut également être régulé par des protéines qui reconnaissent les nucléotides méthylés, tel que par exemple la protéine MeCP2 qui interagit avec des nucléotides méthylés dans les exons et favorise l'inclusion des exons, par l'intermédiaire d'un recrutement de HDAC, réduisant la vitesse d'élongation de l'ARN polymérase II [Maunakea et al., 2013].

2.3.3. Influence du spliceosome

L'épissage alternatif peut être régulé par la machinerie de l'épissage constitutif, incluant directement les éléments du spliceosome, l'assemblage des snRNP ou la modulation de l'assemblage des complexes sur le pré-ARNm.

L'efficacité de la formation de complexes spliceosomaux sur le pré-ARNm peut de ce fait impacter de manière directe la sélection des sites d'épissage. La présence de séquences régulatrices inhibitrices telles que les ISS ou ESS peut empêcher l'interaction de la snRNP U1 avec le pré-ARNm. En effet, la perturbation de cette interaction, favorisera la reconnaissance d'un site d'épissage adjacent et ainsi le recrutement de la snRNP U2.

La régulation de l'épissage alternatif peut également être influencée par le niveau d'expression des UsnRNP. Une abondance d'UsnRNP peut influencer l'épissage de manière tissu-spécifique. Ceci a été notamment démontré en étudiant les mécanismes impliqués dans l'apparition d'une dystrophie musculaire appelée amyotrophie spinale, ou SMA pour *Spinal Muscular Atrophie* [Pellizzoni et al., 1998]. La SMA est caractérisée par un défaut d'épissage de la protéine SMN, responsable de la biogénèse des UsnRNP, particulièrement l'assemblage des protéines cœurs Sm sur l'UsnRNA. Cependant, même si l'expression de la protéine SMN est ubiquitaire, une déficience du taux de cette protéine affecte spécifiquement les motoneurons. Une des explications possibles serait que les pré-ARNm entreraient en compétition face à une déficience d'UsnRNP ce qui pourrait ainsi favoriser l'épissage d'exons sans spécificité tissulaire avec des sites d'épissage « forts » au détriment d'exons présentant une spécificité neuronale et possédant des sites d'épissage « faibles ».

Une étude systématique du rôle de 200 protéines dans l'épissage de 36 cibles différentes [Papasaïkas et al., 2015] a révélé un nombre important de facteurs d'épissage potentiels. La sélection des sites d'épissage était non seulement influencée par les protéines impliquées dans les premières étapes de reconnaissance des sites d'épissage, mais également par des protéines impliquées à des étapes plus tardives de la formation du spliceosome. Cependant le mécanisme impliquant les composants du spliceosome dans la régulation de l'épissage reste à élucider.

Chapitre III : Les facteurs d'épissage à domaines UHM

Chez les eucaryotes supérieurs, un grand nombre de protéines interviennent dans la régulation de l'expression des gènes. Comme nous l'avons vu dans les Chapitres I et II, l'épissage est un processus hautement complexe qui dépend de centaines de facteurs impliqués dans la reconnaissance et le choix des sites d'épissage et dans l'activité catalytique conduisant à l'excision des introns. Ces protéines interviennent dans la définition des sites d'épissage par l'intermédiaire d'interactions spécifiques avec l'ARN. Elles modulent l'épissage en se fixant sur des séquences régulatrices exoniques ou introniques (ESS, ESE, ISS, ISE). Ces interactions spécifiques des protéines avec l'ARN s'effectuent par l'intermédiaire de leurs domaines de liaison à l'ARN. Le domaine de liaison à l'ARN le plus répandu et le plus caractérisé est le domaine RRM pour *RNA Recognition Motif* [Burd and Dreyfuss, 1994].

Les interactions protéine-protéine sont également importantes au bon déroulement de l'épissage. Ce type d'interactions intervient notamment lors du recrutement des UsnRNP mais aussi dans les mécanismes de régulation de l'épissage comme nous l'avons vu dans le chapitre II pour la famille des protéines SR qui engagent des interactions protéine-protéine par l'intermédiaire de leur domaine RS et interagissent simultanément avec des séquences régulatrices sur le pré-ARNm.

En 1993, l'équipe d'Adrian Krainer [Birney et al., 1993] entreprend une analyse systématique des domaines RRM des protéines impliquées dans l'épissage. Par des alignements de séquences, l'équipe identifie un domaine RRM atypique présent dans les séquences des sous-unités U2AF⁶⁵ et U2AF³⁵ du facteur U2AF et propose que la structure de ce domaine aurait évolué afin d'accomplir un rôle différent de la liaison avec l'ARN.

Après avoir étudié les structures de ces domaines RRM non canoniques par RMN, les équipes de M.Green [Kielkopf et al., 2001] et M.Sattler [Selenko et al., 2003] ont caractérisé un nouveau type d'interactions intervenant lors des étapes précoces de l'épissage des pré-ARNm. Cette interaction s'effectue entre les protéines U2AF³⁵ et U2AF⁶⁵ d'une part et U2AF⁶⁵ et SF1 d'autre part. Le domaine RRM non canonique présent dans les séquences d'U2AF³⁵ et U2AF⁶⁵ interagit avec un ligand présent respectivement dans les séquences d'U2AF⁶⁵ et SF1.

Le terme UHM pour *U2AF-Homology-Motif* a alors été proposé pour distinguer ce domaine RRM non canonique du domaine RRM classique [Kielkopf et al., 2004] et le nom ULM pour

UHM-Ligand-Motif a été donné au ligand impliqué dans une interaction avec un domaine UHM [Corsini et al., 2007].

Des études ultérieures ont ensuite identifié une famille de protéines à domaines UHM. Parmi ces protéines, PUF60 [Page-McCaw et al., 1999], CAPER α [Dowhan et al., 2005] et SPF45 [Corsini et al., 2007] qui possèdent toutes des domaines UHM similaires à ceux de U2AF⁶⁵ et U2AF³⁵.

En parallèle, sept domaines ULM ont été identifiés dans la séquence de la protéine SF3b155, composante de la snRNP U2. Parmi ces sept ULM, cinq avaient la capacité d'interagir avec le domaine UHM de U2AF⁶⁵ *in vitro* [Thickman et al., 2006]

De nombreuses études visant à rechercher des mutations affectant les facteurs d'épissage dans les cancers ont permis d'identifier des mutations dans le facteur d'épissage à domaine UHM U2AF³⁵ mais également la protéine à domaines ULM SF3b155. Ces mutations sont présentes dans des tumeurs de syndromes myélodysplasiques et dans certaines leucémies myéloïdes aiguës. [Yoshida et al., 2011]. Ces découvertes mettant en évidence une corrélation entre les mutations des facteurs d'épissage à domaines UHM et ULM et de nombreux cancers renforce l'importance de caractériser ces interactions protéines-protéines, d'identifier de nouveaux partenaires et de comprendre leur rôle dans la régulation de l'épissage.

1. Les interactions UHM-ULM

1.1. Domaine RRM, domaine UHM, similitudes et différences

La structure tridimensionnelle d'un domaine RRM canonique est composée de deux hélices α apposées à un feuillet β composé de quatre brins β antiparallèles avec une topologie $\beta\alpha\beta\alpha\beta$ (Figure 5). Les motifs RNP1 et RNP2 forment les deux brins β centraux, RNP1 étant β_3 et RNP2 étant β_1 . Ces motifs RNP1 et RNP2 maintiennent le repliement correct de la protéine. Par ailleurs, les chaînes latérales aromatiques en seconde position de RNP2 et en troisième/cinquième position de RNP1, sont exposées à la surface et sont responsables de la reconnaissance de l'ARN [Bauer et al., 2012; McLaughlin et al., 2011; Sickmier et al., 2006]

Dans la plupart des cas, plusieurs domaines RRM sont observés dans la séquence d'une protéine, avec une moyenne de deux domaines RRM par protéine [Kielkopf et al., 2004] comme nous l'avons vu dans le chapitre I pour certaines protéines SR.

Le domaine UHM présente la même topologie structurale $\beta\alpha\beta\beta\alpha\beta$ que le domaine RRM. C'est au niveau du site de reconnaissance de l'ARN que l'on observe une divergence. En effet, le premier résidu du motif RNP1 et le second résidu du motif RNP2 sont exposés à la surface du feuillet β alors qu'ils sont typiquement impliqués dans la reconnaissance de l'ARN dans un domaine RRM canonique. Ces résidus sont des acides aminés aliphatiques (Ala 47, et Val 110 pour U2AF³⁵ ou Cys 379, Cys 429 pour U2AF⁶⁵), alors que ce sont des résidus basiques et aromatiques dans les domaines RRM typiques. [Kielkopf et al., 2001; Selenko et al., 2003].

1.2. Interaction UHM-ULM : découverte et structure canonique

La première interaction de Type UHM-ULM a été identifiée par les équipes de P. Legrain et A.Krämer [Rain et al., 1998]. Par des techniques de GST- pull down en utilisant différents mutants de U2AF⁶⁵ et SF1, l'équipe met en évidence une interaction entre le domaine UHM de U2AF⁶⁵ et une région de 137 acides aminées en C-terminal de SF1. L'équipe de M.Sattler s'est ensuite intéressée à la structure tridimensionnelle de cette interaction [Selenko et al., 2003]. Des études par RMN ont permis de caractériser un type d'interaction dans laquelle un tryptophane de SF1 présent dans un domaine riche en acides aminés chargés positivement, interagissait avec le domaine UHM de U2AF⁶⁵ en s'insérant dans une poche hydrophobe. En parallèle, l'équipe de S. Burley [Kielkopf et al., 2001] caractérise le même type d'interaction entre le domaine UHM de U2AF³⁵ et un tryptophane de U2AF⁶⁵ (Figure 16).

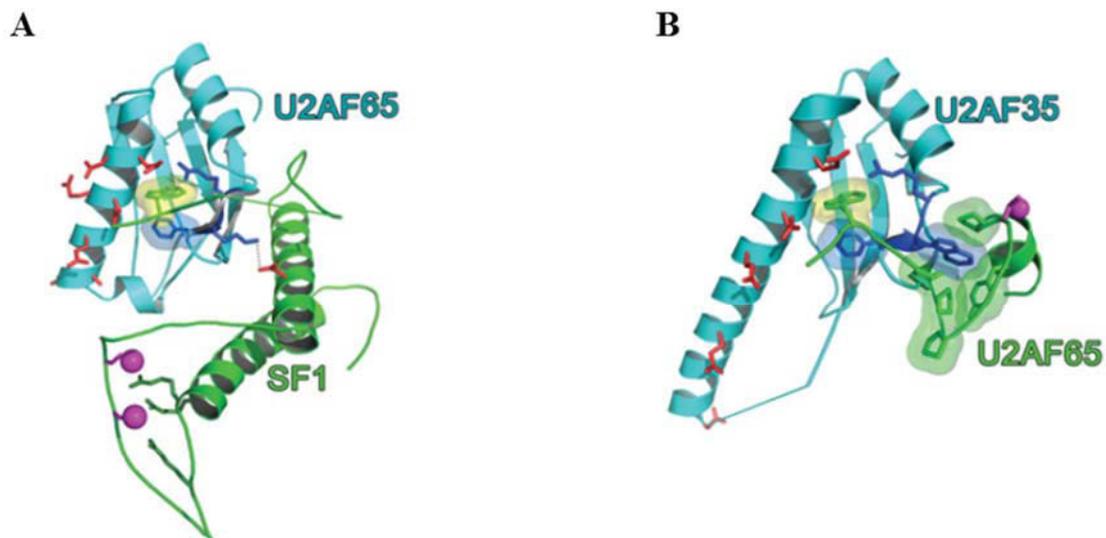


Figure 16. Structure tridimensionnelle d'une interaction UHM-ULM. A: Interaction entre le domaine UHM de U2AF⁶⁵ et un tryptophane de SF1 ; B : Interaction entre le domaine UHM de U2AF³⁵ et un tryptophane de U2AF⁶⁵ (D'après Loerch and Kielkopf, 2016).

C'est ainsi qu'un complexe ternaire SF1-U2AF⁶⁵-U2AF³⁵ a été identifié (Figure 17). Comme nous l'avons vu dans le chapitre I, ce complexe protéique intervient dans les étapes précoces de l'épissage, notamment lors de l'étape de formation du complexe E, lors de la reconnaissance des sites d'épissage en 3'. Ce complexe s'assemble au niveau du site d'épissage en 3' par l'intermédiaire d'interactions spécifiques avec la séquence du pré-ARNm. L'interaction de type UHM-ULM entre SF1-U2AF⁶⁵-U2AF³⁵ vient ensuite renforcer la liaison avec l'ARN.

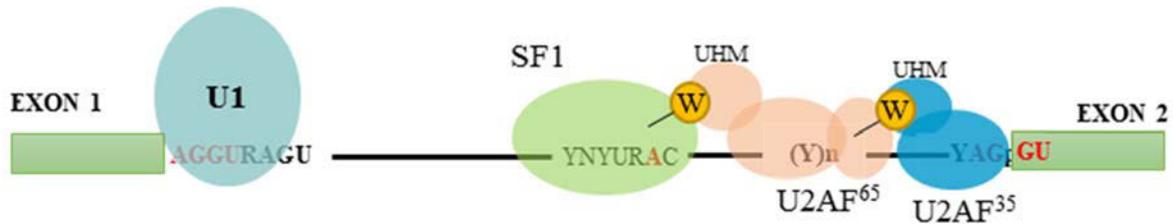


Figure 17. Complexe ternaire d'interactions UHM-ULM composé des interactions entre SF1-U2AF⁶⁵-U2AF³⁵ permettant la reconnaissance des sites d'épissage en 3'. UHM : U2AF-Homology Motif.

La poche hydrophobe du domaine UHM se trouve entre les hélices α et les motifs RNP1 et RNP2. Un motif Arg-X-Phe (X : n'importe quel acide aminé) conservé a été identifié dans la boucle reliant l'hélice α et le brin β des domaines UHM. Ce motif interviendrait dans le renforcement de la liaison du tryptophane du domaine ULM dans la poche hydrophobe du domaine UHM (Figure 18). En effet, le ligand Tryptophane de SF1 s'insère entre les résidus Arginine du domaine UHM de U2AF⁶⁵ (Arg 452), qui forme deux liaisons non covalentes intramoléculaires avec le dernier résidu acide glutamique de la première hélice α (Glu 405), et le résidu Phénylalanine (Phe 454) [Selenko et al., 2003].

Des contacts électrostatiques entre des résidus acides chargés négativement de l'hélice α du domaine UHM de U2AF⁶⁵ et des résidus basiques de SF1 chargés positivement, contribuent à cette interaction. En effet, des interactions entre le résidu Asparagine 401 du domaine UHM de U2AF⁶⁵ et le résidu Arginine 21 du domaine ULM de SF1 ont été observées.

Un motif conservé Arg-X-Phe a également été retrouvé dans la séquence du domaine UHM de U2AF³⁵ (Arg133-X-Phe135). Comme observé pour le domaine UHM de U2AF⁶⁵, le résidu Arginine forme des liaisons non covalentes intramoléculaires avec le dernier résidu Glu 88 de l'hélice α . Des contacts électrostatiques ont également été observés entre le résidu Glu 84 de l'hélice α du domaine UHM de U2AF³⁵ et le résidu Lys 90 du domaine ULM de U2AF⁶⁵ [Kielkopf et al., 2001].

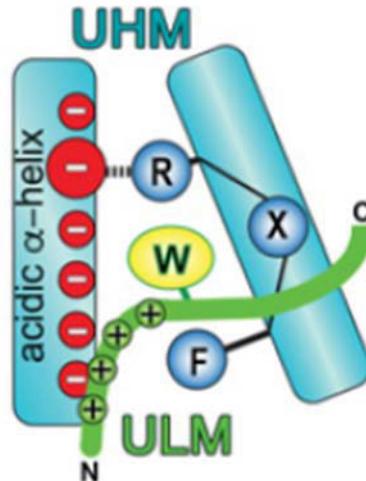


Figure 18. Représentation schématique d'une interaction UHM-ULM. La poche hydrophobe du domaine UHM se trouve entre les hélices α et les motifs RNP1 et RNP2. Un motif R-X-F dans la boucle reliant la première hélice α et le brin β du domaine UHM renforce la liaison du tryptophane du domaine ULM dans la poche hydrophobe du domaine UHM. Le ligand Tryptophane s'insère entre les résidus Arginine (R) du domaine UHM, qui forme deux liaisons non covalentes intramoléculaires avec le dernier résidu acide glutamique de la première hélice α , et le résidu Phénylalanine (F). Des contacts électrostatiques entre des résidus acides chargés négativement de l'hélice α du domaine UHM et des résidus basiques chargés positivement du domaine ULM contribuent à cette interaction (D'après Loerch and Kielkopf, 2016)

1.4. Mise en évidence de protéines intervenant dans des interactions UHM-ULM

En se basant sur le modèle d'interactions U2AF³⁵/U2AF⁶⁵ d'une part et U2AF⁶⁵/SF1 d'autre part, la recherche d'autres protéines possédant des domaines UHM similaires et pouvant engager des interactions du même type a été mise en œuvre. Par ailleurs, il était également important d'identifier des partenaires protéiques possédant des tryptophanes contenus dans des motifs ULM pouvant intervenir dans ces interactions.

1.4.1. Identification de protéines possédant des domaines UHM

La première protéine présentant des homologues de structures et de fonctions avec U2AF⁶⁵ à avoir été identifiée est PUF60. L'équipe de P.A. Sharp [Page-McCaw et al., 1999] découvre ce nouveau facteur d'épissage capable de se lier, comme U2AF⁶⁵ aux séquences polypyrimidine au niveau des sites d'épissage en 3'. De plus, PUF60 en présence de U2AF⁶⁵ était capable de restaurer l'épissage *in vitro*.

Dans la même publication, l'équipe identifie un domaine en région C-terminale de la séquence de PUF60 similaire au domaine UHM de U2AF⁶⁵. Le nom PUMP pour *PUF60-U2AF65 MUD Protein-protein interactions* avait été proposé pour décrire ce domaine.

En s'appuyant sur la structure caractéristique des interactions protéine-protéine des domaines UHM des protéines U2AF⁶⁵ et U2AF³⁵, des équipes de chercheurs ont entrepris la caractérisation d'autres protéines possédant des domaines similaires et pouvant être impliqués dans ce type d'interactions.

Plusieurs approches ont été utilisées afin d'étendre la famille des protéines à domaines UHM. Un motif de séquence caractéristique du domaine UHM a été défini manuellement en se basant sur des résidus conservés qui à la fois maintiennent le repliement caractéristique du domaine et qui sont également capable de médier des interactions protéine-protéine de manière similaire que dans interactions U2AF³⁵-UHM/U2AF⁶⁵-ULM et U2AF⁶⁵-UHM/SF1-ULM [Kielkopf et al., 2001; Selenko et al., 2003].

Trois critères d'identification des domaines UHM ont alors été déterminés en se basant sur la similarité avec le domaine RRM canonique :

- Faible conservation des acides aminés dans les séquences RNP1 et RNP2
- Un motif Arg–X–Phe dans la dernière boucle du repliement RRM-like
- Des résidus acides conservés dans l'hélice α

A partir de ces critères, 19 protéines candidates possédant potentiellement des domaines similaires aux domaines UHM de U2AF⁶⁵ et U2AF³⁵ ont été identifiées. Des comparaisons phylogéniques de ces séquences avec des séquences de domaines RMM canoniques d'autres facteurs d'épissage comme U1A, SXL, PAB, HuD et la Nucléoline ont montré que douze de ces protéines se regroupaient dans une branche distincte, reflétant une nette divergence de la séquence d'un domaine RRM classique [Kielkopf et al., 2004].

Certaines de ces protéines identifiées étaient des homologues présentes chez des espèces différentes. C'est le cas par exemple de MUD2, qui est l'homologue de U2AF⁶⁵ chez la levure *S.cerevisiae*.

Chez l'Homme, sept protéines ont ainsi été identifiées comme faisant partie de la famille des protéines à domaines UHM (Tableau 6 et Figure 19), comme c'est le cas de SPF45 (ou RBM17) [Corsini et al., 2007], PUF60 [Corsini et al., 2009] et CAPER α (ou RBM39) [Loerch et al., 2014]. Ces trois protéines interviennent dans la régulation de l'épissage alternatif [Dowhan et al., 2005; Hastings et al., 2007; Huang et al., 2012; Lallena et al., 2002; Page-McCaw et al., 1999]. D'autres protéines à domaines UHM interviennent de manière directe ou indirecte dans la régulation de l'épissage alternatif, comme c'est le cas de la protéine KIS (ou UHMK1). KIS est une protéine Kinase [Maucuer et al., 1997] qui phosphoryle spécifiquement la protéine SF1 et initie des interactions de types UHM-ULM avec SF1 et SF3b155 [Manceau et al., 2008]. La phosphorylation de SF1 par KIS augmenterait ainsi la force de l'interaction U2AF⁶⁵-UHM/SF1-ULM [Chatrikhi et al., 2016; Manceau et al., 2006; Wang et al., 2013].

Tableau 6. Famille des protéines à domaines UHM: nomenclature, position et taille du domaine UHM, partenaires ULM connus, séquences RNP1/RNP2 et motif RXF (d'après Loerch et Kielkopf, 2016)

Nom	RefSeq	Position	Partenaires ULM	Séquence RNP1/RNP2	Motif RXF
URP	NP_005080	197-303	N.D	LLIKSM/RGNVYVQ	RWY
KIS	NP_787062	316-415	SF1 SF3b155	VLRLLN/RGQVFVEY	RMF
PUF60	NP_510965	460-559	SF3b155	MVLRNM/IVKIFVEF	RWF
SPF45	NP_116294	301-401	SF3B155	VLLRNM/AVRIFLEF	RYF
TAT-SF1	NP_055315	260-353	N.D	VIKKNM/DGVASVSF	RWF
MAN1	NP_055134	775-880	MAN1	LKIRNM/EGCVYVKC	SWF
CAPERα	NP_004893	411-524	SF3B155	FQLSNM/QGNVYVKC	RWF

U2AF ⁶⁵	NP_009210	375-475	SF1, SF3B155, Atx1	LCLMNM/CGKIFVEF	RKF
U2AF ³⁵	NP_006749	43-146	U2AF ⁶⁵	IALLNI/VGNVYVKF	RWF

La protéine MAN1, retrouvée dans la membrane nucléaire interne, contient à la fois un domaine UHM et un motif ULM. Une interaction intramoléculaire entre ces deux motifs intervient dans la voie de signalisation du facteur de croissance TGF β [Kondé et al., 2010].

Les protéines à domaines UHM présentent des arrangements similaires. En effet, le domaine UHM est souvent retrouvé en C-terminal (Figure 19). Cette exposition faciliterait la reconnaissance moléculaire et l'établissement d'interactions UHM-ULM [Kielkopf et al., 2004]. De plus, plusieurs de ces protéines possèdent des motifs similaires à ceux retrouvés dans les facteurs d'épissage, tels que des domaines RRM canoniques, des domaines RS, des motifs à doigts de zinc et des régions riches en glycine.

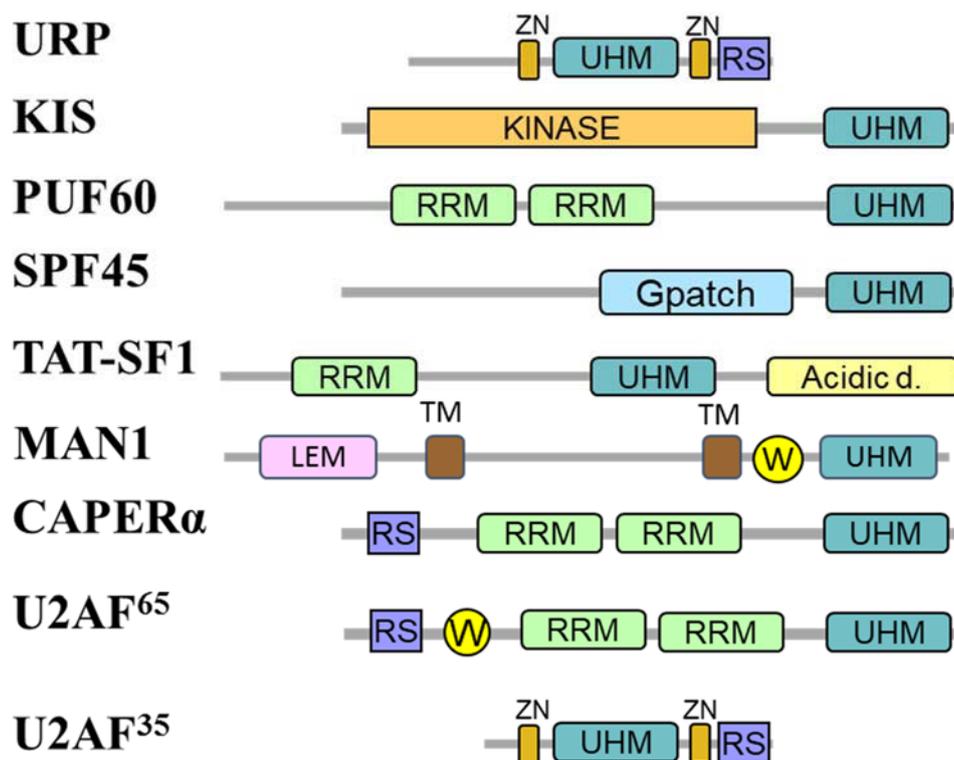


Figure 19. Représentation schématique des structures des protéines à domaines UHM. RRM : RNA Recognition Motif ; UHM : U2AF Homology Motif ; RS : riches en répétitions Serine et Arginine ; ZN : domaine à doigts de zinc ; W : tryptophane ; LEM : LA2P-EMERIN-MANI homology ; TM : domaine transmembranaire ; G-Patch : domaine riche en Glycine ; Acidic d. : domaine riche en Glutamine et Asparagine.

Un grand nombre de ces protéines sont des facteurs d'épissage, comme c'est le cas de SPF45, PUF60 et CAPER α . Certaines, comme PUF60, CAPER α et TAT-SF1 peuvent intervenir également dans la régulation de la transcription.

1.4.2. Identification de protéines possédant des motifs ULM

La séquence des motifs ULM est très peu conservée. Elle se limite à un tryptophane central capable d'interagir dans une poche hydrophobe entre les hélices α et β du domaine UHM précédé en N-terminal d'une série d'acides aminés chargés positivement et suivi en C-terminal d'une courte séquence riche prolines et acides aminés aromatiques. En prenant ces deux critères en compte, une recherche des motifs similaires dans la base de données PROSITE (Swiss-Prot, ScanProsite) [de Castro et al., 2006] donne un résultat de plus de 9500 protéines candidates.

Malgré ce nombre important, seules cinq protéines contenant des motifs ULM ont été identifiées comme interagissant avec des domaines UHM (Tableau 7 et Figure 20). Parmi ces protéines, on retrouve U2AF⁶⁵ et SF1 sur lequel se base le modèle d'interactions UHM-ULM, qui interagissent respectivement avec U2AF³⁵ et U2AF⁶⁵ [Kielkopf et al., 2001; Selenko et al., 2003], MAN1 qui homodimérise [Kondé et al., 2010], et la protéine ATX-1 (Ataxin-1) responsable du développement d'une maladie neurodégénérative appelée ataxie spinocerebellaire de type I [de Chiara et al., 2009]. Le motif ULM de la protéine ATX-1 situé en C-terminal interagit avec le domaine UHM de U2AF⁶⁵ et SPF45 [de Chiara et al., 2009; Lim et al., 2008].

Tableau 7. Protéines à domaines ULM: nom, position du domaine ULM, partenaires UHM connus et séquences ULM (d'après Loerch et Kielkopf, 2016).

Nom	RefSeq	Position	UHM-int	Séquence ULM
MAN1	NP_055134	756 - 771	MAN1	KILVIPSKV W QGQAFH
ATX1	NP_000323	766-781	U2AF ⁶⁵	KPAATR KRR WSAPESR
U2AF ⁶⁵	NP_009210	83-98	U2AF ³⁵	E KKKK V RKY W DV PPPG
SF1	NP_004621	14-28	U2AF ⁶⁵	PS KKR K RSR W NQ DTME

SF3b155	NP_036565	191-206, 209-224, 223-238, 245-260, 284-299, 301-316, 329-344	U2AF ⁶⁵ PUF60 SPF45 CAPER α	QPPSKRKRR W DQTADQ (ULM1) GATPKKLSS W DQAETP (ULM2) TPGHTPSLR W DETPGR (ULM3) PGATPGSKI W DPTPSH ATSSARKNR W DETPKT(ULM4) DTPGHGSG W AETPRT PGASKRKS R W DETPAS(ULM5)
----------------	-----------	---	--	---

Par ailleurs, plusieurs équipes ont remarqué la présence de sept répétitions de motifs ULM dans la région N-terminale de la protéine SF3b155, composante de la snRNP U2 [Cass and Berglund, 2006; Corsini et al., 2007, 2009; Loerch et al., 2014; Thickman et al., 2006].

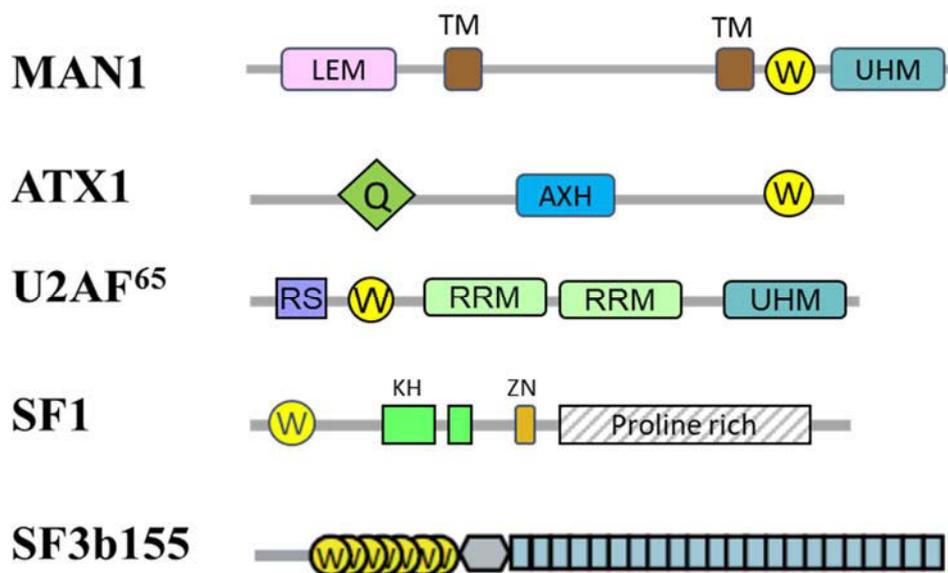


Figure 20. Représentation schématique des structures des protéines à domaines ULM identifiées. RRM : RNA Recognition Motif ; UHM: U2AF Homology Motif ; RS : riches en répétitions Serine et Arginine; ZN: domaine à doigts de zinc; W: tryptophane; LEM: LA2P-EMERIN-MAN1 homology ; TM : domaine transmembranaire ; Q : polyQtract ; AXH : motif AXH

La présence de sept tryptophanes précédés de résidus basiques chargés positivement de manière similaire à l'organisation retrouvée dans le domaine ULM de SF1 a soulevé la question d'une possible interaction de SF3b155 avec plusieurs protéines, par l'intermédiaire d'interactions UHM-ULM (Figure 21). Des expériences de double-hybride chez la levure *S.cerevisiae* ont montré qu'une région de 97 acides aminés de ce domaine (résidus 267–369) était suffisante pour établir une interaction avec U2AF⁶⁵. Cependant, l'interaction était renforcée avec un fragment plus grand de SF3b155 (résidus 1–480). Des expériences *in vitro* ont ensuite confirmé que U2AF⁶⁵ pouvait interagir avec SF3B155 par l'intermédiaire d'interactions UHM-ULM

avec cinq des sept tryptophanes contenus dans des motifs ULM de SF3b155 [Corsini et al., 2007, 2009; Loerch et al., 2014; Thickman et al., 2006].

Ces interactions UHM-ULM entre U2AF⁶⁵ et SF3b155 pourraient intervenir dans la stabilisation de l'association de la snRNPU2 sur l'ARN lors des étapes d'initiation de l'épissage. Cependant, le rôle de cet enchainement de motifs ULM dans la séquence de SF3b155 et leurs partenaires protéiques restent encore à définir.

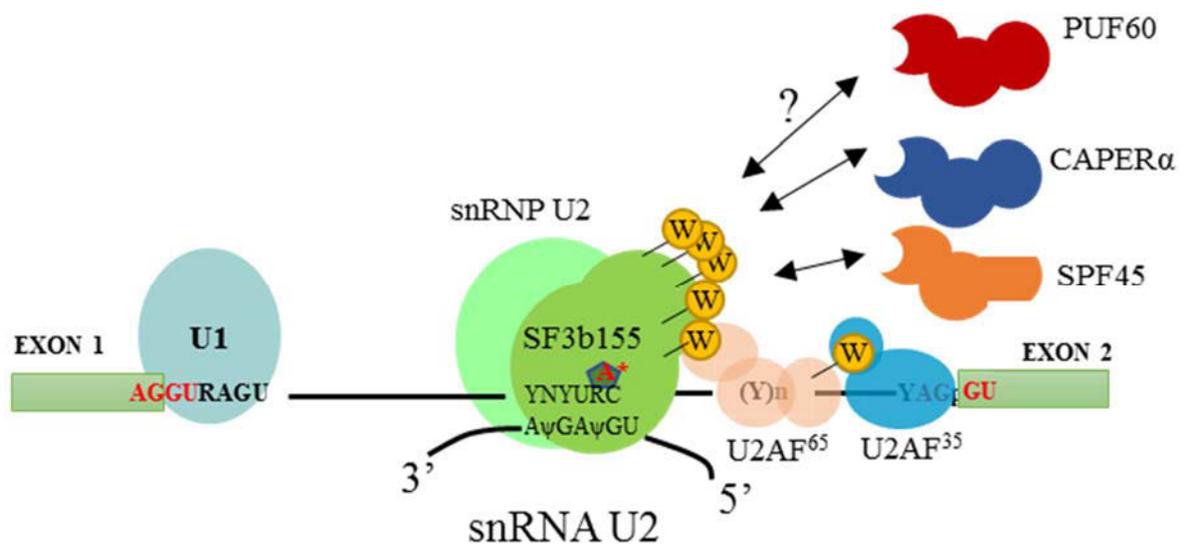


Figure 21. Le facteur d'épissage à multi-ULM SF3b155 durant les étapes de reconnaissance des sites 3' d'épissage. U2AF⁶⁵ interagit avec SF3b155 par l'intermédiaire d'interactions UHM-ULM. Ces interactions UHM-ULM entre U2AF⁶⁵ et SF3b155 pourraient intervenir dans la stabilisation de l'association de la snRNPU2 sur l'ARN lors des étapes d'initiation de l'épissage. L'enchainement de motifs ULM dans la séquence de SF3b155 pourrait indiquer le recrutement d'autres partenaires protéiques à motifs UHM durant cette étape, comme PUF60, CAPER α ou SPF45.

2. Les facteurs d'épissage à domaines UHM (UHMSF)

La recherche par homologie de structures de protéines possédant des domaines UHM a révélé un potentiel rôle fonctionnel commun à ces protéines. En effet, la majorité d'entre elles semblent être impliquées de manière directe ou indirecte dans les mécanismes d'épissage. Durant ma thèse, je me suis intéressée aux interactions biochimiques et au rôle fonctionnel dans la régulation de l'épissage du facteur U2AF⁶⁵ ainsi que de trois autres protéines possédant des domaines UHM, CAPER α , PUF60 et SPF45.

2.1. La protéine U2AF⁶⁵

2.1.1. Rôles de U2AF⁶⁵ dans l'épissage constitutif

Comme je l'ai décrit dans le Chapitre I, la fixation de la snRNP U2 au site 3' d'épissage dépend de plusieurs facteurs d'épissage. L'un des facteurs les plus importants est la protéine U2AF composée de ses deux sous-unités U2AF⁶⁵ et U2AF³⁵, qui se lient respectivement à la séquence Polypyrimidine et au dinucléotide AG sur la jonction intron/exon et définissent ainsi les sites 3' d'épissage.

Cet hétérodimère a été découvert pour la première fois par l'équipe de M.Green [Ruskin et al., 1988]. C'est en analysant l'assemblage *in vitro* de la snRNP U2 sur le pré-ARNm que l'équipe découvre que contrairement à la snRNP U1, la snRNP U2 était incapable de se lier seule au point de branchement et que cet assemblage nécessitait une autre protéine. A l'aide de techniques de complémentation et d'électrophorèse sur gel dénaturant, ils identifient une

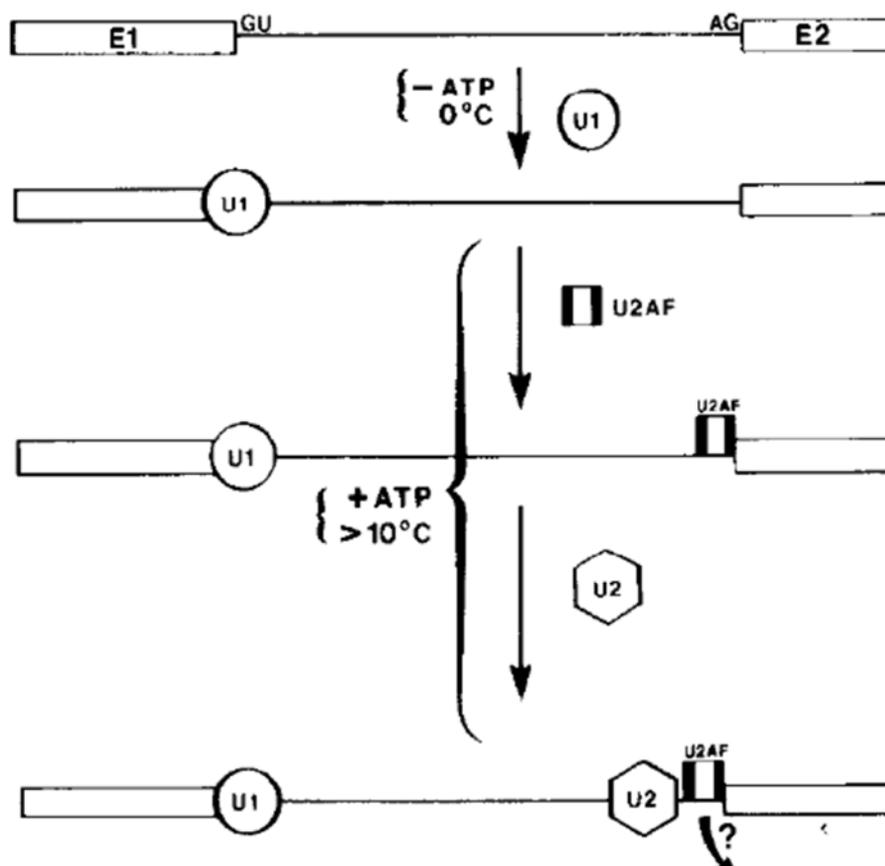


Figure 22. Modèle de recrutement de la snRNP U2 au pré-ARNm (Ruskin et al, 1988). L'équipe propose un modèle où le recrutement de la snRNP U2 au point de branchement nécessite la liaison de la snRNP U1 au site 5' et du facteur U2AF au site 3' d'épissage.

protéine auxiliaire nécessaire à la fixation de la snRNP U2 qu'ils nomment U2AF (U2 auxiliary factor). L'équipe propose alors un modèle d'assemblage du spliceosome (Figure 22), où les premières étapes de reconnaissance des sites d'épissage étaient assurées par la liaison de snRNP U1 au site 5' d'épissage et du facteur U2AF au site 3'. La liaison de U2AF au site 3' permettait ensuite le recrutement de la snRNP U2 au point de branchement.

En purifiant le facteur U2AF, P D. Zamore and M. Green découvrent que la protéine est composée de deux sous-unités de 65 et 35 KD qu'ils nomment respectivement U2AF⁶⁵ et U2AF³⁵. Par des expériences de crosslink aux UV, les chercheurs identifient les séquences de liaison à l'ARN du facteur U2AF et découvrent que U2AF⁶⁵ se lie à la séquence polypyrimidine [Zamore and Green, 1989]. Par des études biochimiques de déplétion et de complémentation du facteur U2AF dans des extraits nucléaires, P D Zamore and M. Green démontrent que U2AF est nécessaire à l'assemblage du spliceosome [Zamore and Green, 1991]. Le site de fixation de U2AF³⁵ au dinucléotide AG du site 3' d'épissage a été identifié quelques années plus tard par l'équipe [Wu et al., 1999]

Cependant, si la présence de U2AF est nécessaire à l'assemblage du spliceosome en général, est-elle indispensable à l'assemblage de tous les spliceosomes ? Par ailleurs, si U2AF est capable de se fixer aux sites d'épissage en 3', cela implique-t-il qu'elle est liée à tous les sites d'épissage en 3' et cette interaction reflète-elle un rôle fonctionnel systématique de la protéine dans l'épissage ?

Chez la levure et dans les cellules de mammifères, on trouve des introns épissés par le spliceosome majeur et qui présentent des séquences polypyrimidine atypiques. Cependant, l'excision de ces introns était toujours dépendante du facteur U2AF [Sridharan and Singh, 2007].

Une récente analyse des interactions d'U2AF avec l'ARN à l'échelle du génome a révélé qu'U2AF était lié à plus de 88% des sites d'épissage en 3' du génome humain [Shao et al., 2014]. Ces observations pourraient être dues aux limites de sensibilités de la technique, mais pourraient également indiquer que, dans certains cas, l'épissage des introns s'effectuerait sans interaction directe avec U2AF⁶⁵ avec l'ARN. D'autres facteurs pourraient alors intervenir pour assurer la fonction de U2AF⁶⁵.

Si l'on sait que le facteur U2AF⁶⁵ n'intervient pas dans l'épissage des pré-ARNm pris en charge par le spliceosome mineur et dépendants de la snRNP U12 [Turunen et al., 2013], l'analyse des

mécanismes impliqués dans l'épissage de ces types d'introns pourrait apporter plus d'information quant au mode de définition des sites d'épissage indépendant de U2AF⁶⁵.

Par ailleurs, des études récentes ont montré que la séquence pauvre en Polypyrimidine de l'intron 5 de CD44 contenait un site de liaison CAUC spécifique de la protéine YB-1 qui interviendrait dans le recrutement de U2AF [Wei et al., 2012]. Ces données démontrent un rôle fonctionnel de U2AF dans l'épissage de cet intron indépendamment de l'interaction de la protéine avec le pré-ARNm.

2.1.2. Rôles de U2AF⁶⁵ dans l'épissage alternatif

Le rôle de U2AF⁶⁵ dans l'épissage alternatif a tout d'abord été identifié par l'équipe de M.Green [Valcárcel et al., 1993] dans un modèle *in vitro* de régulation de l'épissage alternatif d'un gène impliqué dans la différenciation sexuelle somatique de la drosophile. La protéine SXL se lie spécifiquement à la séquence polypyrimidine en 3' du premier intron du pré-ARNm du gène transformer (ou tra). La fixation de SXL bloque ainsi la liaison du facteur U2AF au même site. U2AF se lie alors à des sites de plus faible affinité et induit ainsi l'expression du gène de détermination du sexe chez la femelle.

S'en sont suivi plusieurs autres travaux portant sur l'implication de U2AF dans la régulation de l'épissage alternatif, notamment dans l'épissage de l'exon alternatif N1 de *c-src* [Sharma et al., 2005] régulé par la protéine PTB (polypyrimidine tract binding protein). Cette dernière empêche la fixation de U2AF⁶⁵ au site d'épissage de l'exon +1 en aval, favorisant l'inclusion de l'exon N1. La régulation de l'épissage alternatif de l'exon 6B du pré-ARNm de la β -tropomyosine [Saulière et al., 2006] se fait également par des mécanismes similaires.

Comme c'est le cas dans ces exemples avec PTB, plusieurs études ont montré une compétition d'autres protéines liant l'ARN avec le facteur U2AF, tel que hnRNPC [Zarnack et al., 2013], YB-1 [Wei et al., 2012], TIA1 et TIAR [Wang et al., 2010] dans la régulation de l'épissage.

Par ailleurs, de récentes analyses bio-informatiques suggèrent une implication plus active de U2AF dans la régulation de l'épissage alternatif. Plusieurs sites de liaisons potentiels ont été identifiés dans des exons. En utilisant des constructions de minigènes rapporteurs, il a été démontré que la liaison de U2AF à ces séquences exoniques induisait l'exclusion d'exons alternatifs [Lim et al., 2011].

Les analyses génomiques par CLIP-seq de l'interaction de U2AF⁶⁵ avec le pré-ARNm suivies d'expériences de minigènes rapporteurs ont révélé un effet « polaire » de U2AF⁶⁵ dans la régulation de l'épissage, dépendant du site de fixation de la protéine (Figure 23). En effet, les chercheurs proposent que la liaison de U2AF à une séquence de l'intron en amont de l'exon alternatif induit l'inhibition de l'action du facteur U2AF lié au site d'épissage en 3' et par conséquent l'exclusion de l'exon. Cependant, lorsque U2AF se lie à une séquence de l'intron en aval de l'exon alternatif, l'action de U2AF liée au site d'épissage en 3' n'est pas inhibée et l'exon alternatif est ainsi inclus [Shao et al., 2014].

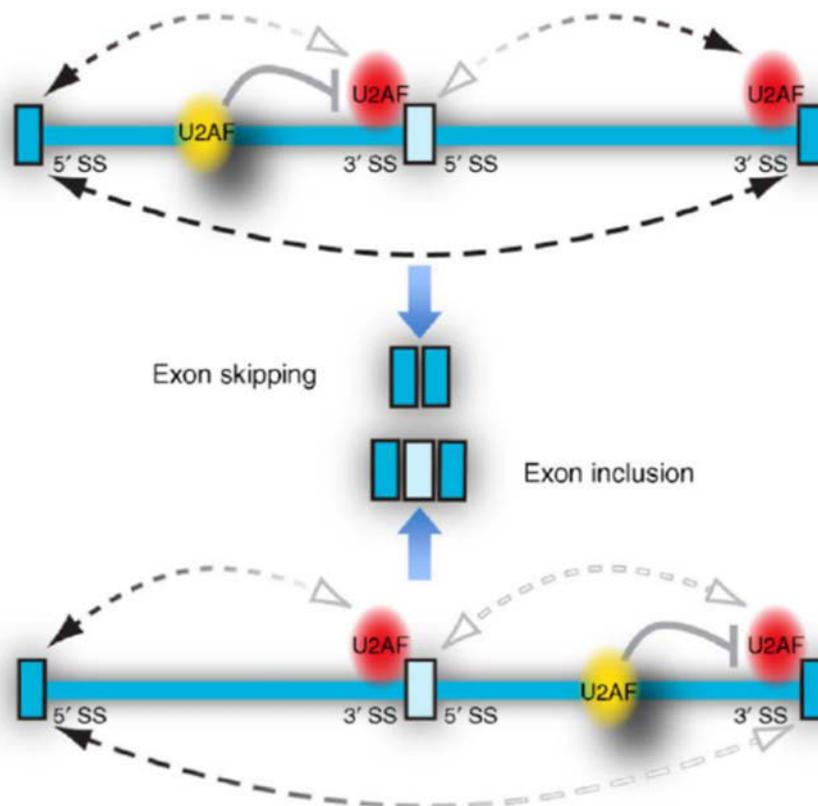


Figure 23. Modèle de régulation de l'épissage alternatif par le facteur U2AF proposé par Fu et collaborateurs (Shao et al, 2014).

2.1.3. Maladies liées à des défauts de U2AF

Des mutations dans les séquences de U2AF⁶⁵ et U2AF³⁵ ont été identifiées et reportées comme associées à des maladies du sang, plus particulièrement à des syndromes myélodysplasiques [Cazzola et al., 2013]. Cependant le mécanisme par lequel ces mutations peuvent induire des maladies reste encore méconnu. La première hypothèse serait que la maladie serait causée par un défaut d'épissage lié à la perte de fonction de U2AF. De récentes études ont d'ailleurs montré que des mutations de U2AF³⁵ engendraient des différences d'épissage alternatif de certains

gènes. Plus particulièrement, ces gènes étaient impliqués dans la prolifération cellulaire [Shao et al., 2014]. Même si des études ont montré que des défauts de U2AF⁶⁵ engendraient des changements dans l'épissage alternatif de certains exons [Cho et al., 2015], aucun lien avec une maladie particulière n'a été observé.

2.2. La protéine CAPER α

CAPER α , aussi connue sous le nom de HCC1 ou RBM39 pour *RNA Binding protein 39*, est la protéine à domaines UHM la plus récemment caractérisée [Loerch et al., 2014]. Ses domaines protéiques sont organisés de manière similaire à ceux d'U2AF⁶⁵. CAPER α a la particularité de posséder une dualité de rôles à la fois dans l'activation de la transcription et dans l'épissage.

2.2.1. Rôles de CAPER α dans la transcription

CAPER α a originellement été découverte pour son rôle dans la régulation de la transcription. Elle a tout d'abord été identifiée comme interagissant spécifiquement avec l'intégrateur transcriptionnel ASC-2 [Jung et al., 2002]. Par la suite un rôle de CAPER α dans la régulation des facteurs de transcription AP-1, c-Jun et des récepteurs à l'œstrogène ER α et ER β a également été mis en évidence [Dowhan et al., 2005]. Par ailleurs, la fonction transactivatrice de CAPER α était dépendante de la liaison des récepteurs à l'œstrogène avec l'estradiol.

2.2.2. Rôles de CAPER α dans l'épissage

CAPER α a été retrouvée associée aux complexes spliceosomaux purifiés [Rappsilber et al., 2002] ce qui pourrait indiquer un potentiel rôle de la protéine dans l'épissage. De plus, des interactions entre CAPER α et la protéine SR, SRP53 ont été identifiées [Cazalla et al., 2005].

Le rôle potentiel de CAPER α dans l'épissage a tout d'abord été analysé dans la publication de Dowhan et al, [Dowhan et al., 2005]. Après avoir découvert que CAPER α jouait un rôle de co-activateur transcriptionnel des récepteurs aux hormones stéroïdes ER α et ER β , l'équipe a recherché si la protéine était également impliquée dans la maturation des pré-ARNm dont la transcription est dépendante de ces récepteurs. En analysant l'épissage alternatif du minigène rapporteur dérivé du gène CT/CGRP suite à l'inhibition de l'expression de CAPER α , l'équipe découvre un mécanisme d'épissage alternatif dépendant de la protéine. De plus, en diminuant le taux d'expression de CAPER α dans des cellules T47D, un mécanisme de régulation de

l'épissage alternatif de VEGF pour *vascular endothelial growth factor* dépendant de la protéine a été identifié.

La similitude d'organisation de CAPER α et U2AF⁶⁵, incluant un domaine RS en N-terminal, deux domaines RRM et un domaine UHM en C-terminal suggère également une implication de cette protéine dans la régulation de l'épissage. En analysant les interactions entre les protéines possédant des domaines RS et les composants impliqués dans la définition des sites d'épissage par une technique de transfert d'énergie entre molécules fluorescentes (FRET) et d'imagerie en temps de vie de fluorescence (FLIM), une interaction entre CAPER α et les deux sous-unités de U2AF a été détectée [Ellis et al., 2008]. De plus, des analyses globales de l'interactome de U2AF⁶⁵ ont montré une interaction de la protéine avec CAPER α [Prigge et al., 2009].

Par ailleurs, des expériences de Gst-pull down en utilisant la protéine GST-SF3b155 produite *in vitro*, incubée avec un extrait cellulaire, ont montré une spécificité d'interaction de CAPER α avec SF3b155 par l'intermédiaire d'interactions UHM-ULM [Loerch et al., 2014].

Enfin, des études d'immunofluorescence ont mis en évidence une colocalisation de CAPER α avec SC35 et les UsnRNA dans les speckles [Anderson et al., 2016], des compartiments nucléaires riches en facteurs d'épissage.

Le rôle de CAPER α dans l'épissage alternatif de VEGF a été plus particulièrement investigué [Huang et al., 2012]. Une diminution de l'expression de CAPER α résultait d'un changement dans le ratio d'expression des isoformes 165/189 de VEGF et une augmentation de l'expression de la forme plus angiogénique VEGF165. CAPER α agirait donc comme suppresseur de l'angiogenèse tumorale. En favorisant la forme moins angiogénique VEGF189.

Des analyses génomiques par CLIP-seq des interactions de CAPER α avec l'ARN ont révélé que la protéine se liait spécifiquement aux sites d'épissage en 3' et en 5' [Mai et al., 2016]. Des analyses d'expression différentielles du transcriptome par séquençage à haut débit (RNAseq) dans des cellules MCF-7 où le taux de CAPER α a été diminué, ont révélé que CAPER α était impliquée dans la régulation de l'épissage de plusieurs gènes aux fonctions variables, et des centaines d'évènements d'épissages alternatifs régulés par la protéine ont été identifiés. La majorité des évènements affectés étaient des exons cassettes et étaient impliqués dans l'endocytose, la réponse aux dommages liés à l'ADN, ou la régulation du cycle cellulaire. De manière intéressante, approximativement 20% des évènements d'épissage alternatif régulés par

CAPER α étaient régulés de manière similaire par U2AF⁶⁵. Ceci suggèrerait une dualité d'action de ces deux facteurs d'épissage à domaine UHM dans la régulation d'évènements d'épissage particuliers [Mai et al., 2016].

De récentes analyses du transcriptome à haut débit par RNAseq dans des cellules HEK où le taux de CAPER α a été diminué par ARN interférence ont montré que 198 exons étaient régulés par CAPER α [Královicová et al., 2018]. Une représentation équivalente d'inclusion et d'exclusion d'exons a été observée parmi ces évènements d'épissage indiquant un rôle plus complexe de CAPER α dans la régulation de l'épissage.

2.2.3. Rôles de CAPER α dans le développement tumoral

Chez l'Homme, il existe trois isoformes de CAPER α et au moins deux de ces isoformes sont caractérisés comme des antigènes associés à des tumeurs induisant des réponses immunitaires humorales chez des patients atteints de cancer du poumon et de carcinome hépatocellulaire [Chai et al., 2014].

La fonction de CAPER α a été liée à plusieurs cancers et progressions tumorales [Sillars-Hardebol et al., 2012]. Par exemple, CAPER α est surexprimée dans les cancers du poumon à petites cellules, les cancers du sein ainsi que dans les carcinomes et les adénomes colorectaux.

Des études tentant d'étudier le rôle de CAPER α dans le développement tumoral ont montré que lorsque l'expression de la protéine était inhibée dans les lymphocytes de patientes atteintes de cancers du sein ER+, cela induisait la suppression de l'activité oncogénique des protéines NF- κ B et v-Rel [Dutta et al., 2008] et de ce fait la prolifération des cellules cancéreuses.

Des études ont également montré que l'expression de CAPER α était presque indétectable dans les cellules mammaires normales tandis qu'elle était augmentée dans les carcinomes *in situ* ductaux (DCIS). La diminution de l'expression de CAPER α dans les cellules cancéreuses MCF-7 a eu pour effet la réduction considérable de la prolifération cellulaire *in vitro*. De plus, lorsque ces cellules tumorales étaient injectées dans des souris, celles-ci développaient des tumeurs de taille plus petite en comparaison aux cellules MCF-7 dans lesquelles CAPER α était exprimée. Par ailleurs, ces cellules présentaient des taux réduits d'expression de protéines régulatrices du cycle cellulaire PCNA, MCM7 et Cycline D1, suggérant que CAPER α favorise la prolifération cellulaire [Mercier et al., 2014].

Cependant, contrairement à l'effet suppresseur de tumeurs de la sous-expression de CAPER α dans les cancers illustrés précédemment, dans les cellules du sarcome d'Ewing, la diminution du taux d'expression de CAPER α induit l'expression d'isoformes plus angiogéniques de VEGF induisant une augmentation de la vascularité tumorale *in vivo* [Huang et al., 2012].

2.3. La protéine PUF60

2.3.1. Rôles de PUF60 dans la régulation de la transcription

La protéine PUF60 pour *Poly-U-binding factor 60 kDa*, aussi connue sous le nom de FIR pour *FBP-interacting repressor*, agit dans certains cas comme activateur de la transcription. En effet, des études ont montré que PUF60 contrôlait l'expression du gène *c-myc* Humain. Une séquence située en aval du site de transcription du gène *myc* et appelée FUSE pour *Far UpStream Element* est nécessaire à l'expression correcte de ce gène. La protéine FBP pour *FUSE-binding protein* reconnaît spécifiquement le site FUSE et stimule l'expression de *c-myc*. Par l'intermédiaire d'interaction protéine/protéine et Protéine/ADN, PUF60 se lie à la fois à la séquence FUSE sur le gène *c-myc* et au facteur FBP et inhibe ainsi la transcription [Liu et al., 2000].

2.3.2. Rôles de PUF60 dans l'épissage

PUF60 est la première protéine homologue à U2AF⁶⁵ à avoir été identifiée. Elle possède deux domaines RRM centraux et un domaine UHM en C-terminal. On note cependant une absence de domaines RS et de motifs ULM dans la séquence de la protéine, contrairement à U2AF⁶⁵.

PUF60 avait à l'origine été identifiée comme un nouveau facteur d'épissage pouvant se lier à la séquence de polypyrimidine au site d'épissage en 3' par l'intermédiaire d'interactions ARN/protéine avec son domaine RRM. PUF60 formait un dimère avec la protéine SRp54 et interagissait avec U2AF⁶⁵ au niveau du site 3' d'épissage. Le domaine d'interaction de PUF60 avec U2AF⁶⁵ avait été identifié et nommé domaine PUMP, pour *PUF60, U2AF⁶⁵, Mud2p protein-protein interaction*. Il avait été caractérisé comme un domaine ressemblant au domaine RRM, mais avec des différences lui conférant la capacité d'effectuer des interactions protéine-protéine [Page-McCaw et al., 1999].

Des techniques de purification du spliceosome couplées à des analyses par spectrométrie de masse ont permis d'identifier un grand nombre de nouveaux facteurs d'épissage, dont PUF60 [Jurica and Moore, 2003].

L'équipe d'A. Krainer a ensuite mis en évidence le rôle fonctionnel de PUF60 dans l'épissage [Hastings et al., 2007] en utilisant comme modèle d'épissage *in vitro* un dérivé du pré-ARNm de la β -globine. L'épissage correct du pré-ARNm de la β -globine requiert des facteurs d'épissage additionnels, dont la protéine SR SC35, car son intron possède des séquences atypiques suboptimales au niveau de la séquence pyrimidines au site 3' d'épissage. La protéine SC35 est présente dans l'extrait nucléaire mais absente dans la fraction cytoplasmique dite « S100 ». Des études de complémentation de la fraction S100 par différentes fractions de l'extrait nucléaire, ont été réalisées. La fraction nucléaire pouvant restaurer l'épissage a été analysée par des techniques de chromatographie en phase liquide couplée à de la spectrométrie de masse et des protéines candidates ont été découvertes. C'est ainsi que PUF60 a été identifiée comme facteur d'épissage proprement dit.

Des études approfondies de l'analyse de l'épissage *in vitro* ont ensuite montré que PUF60 pouvait remplir en partie la fonction de U2AF⁶⁵, mais que l'épissage était plus efficace en présence des deux protéines. De plus, la réduction du taux de PUF60 et de U2AF⁶⁵ dans des cellules modifiait le profil d'épissage alternatif des transcrits APP, MAPT, UBQLN1 et BIN1, indiquant que les deux protéines pouvaient être impliquées dans la régulation de l'épissage alternatif par l'intermédiaire de la reconnaissance des séquences polypyrimidine en 3'.

En parallèle, l'équipe de M. Sattler a tenté de caractériser les interactions impliquant le domaine UHM de PUF60 [Corsini et al., 2009]. Des analyses *in vitro* d'interaction protéine-protéine par des techniques de GST-pull down, de Titration Calorimétrique Isotherme et de spectroscopie à résonance magnétique nucléaire (RMN) ont montré que le domaine UHM de PUF60 était capable d'interagir avec les séquences ULM des facteurs d'épissage SF1, U2AF⁶⁵ et SF3b155. De plus, ce domaine avait une préférence pour le motif ULM de SF3b155.

Des analyses du transcriptome à haut débit par RNAseq dans des cellules HEK où le taux de PUF60 a été réduit par ARN interférence ont révélé 258 événements d'épissage régulés par la protéine. De plus, les événements d'épissage altérés par la diminution du taux de PUF60 montraient un taux élevé d'exclusion d'exon, indiquant une potentielle fonction activatrice de PUF60 dans la régulation de l'épissage [Královicová et al., 2018].

Des analyses approfondies des séquences introniques et exoniques présentes dans 102 évènements d'épissages régulés négativement par la diminution du taux de PUF60 ont révélé des caractéristiques communes. En effet, les sites d'épissage en 3' présentaient une région riche en polypyrimidine plus longue que la moyenne, une zone d'exclusion en dinucléotide AG (AGEZ pour AG Exclusion Zone) 1.5X plus longue et un point de branchement plus distant de la jonction Intron/exon. Les caractéristiques architecturales de ces séquences pourraient donc indiquer un rôle fonctionnel de PUF60 par l'intermédiaire d'interaction ARN/protéines au niveau de longues séquences introniques en 3' favorisant ainsi l'inclusion de l'exon [Královicová et al., 2018].

2.3.3. Maladies liées à des défauts de PUF60

Des microdélétion du chromosome 8q24.3 incluant le gène PUF60 ont été identifiées chez des patients présentant des retards développementaux, des microcéphalies et des malformations crâniofaciales, rénales et cardiaques [Dauber et al., 2013]. Des phénotypes similaires ont été observés chez des patients possédant des variants de PUF60, suggérant que des mutations au niveau de la protéine sont responsables du syndrome. Les analyses d'exome ont permis de caractériser différentes mutations de PUF60 associées à ces retards développementaux ; la majorité des mutations induisent des changements du cadre de lecture avec arrêt prématuré de la traduction de la protéine. Des mutations non-sens concentrées dans les domaines RRM et UHM de la protéine représentaient un tiers des mutations étudiées [Low et al., 2017]. La présence de ces mutations suggère que la perte de fonction de PUF60 dans ces retards développementaux est résultante de la perte des interactions ARN/protéine par l'intermédiaire du domaine RRM ou des interactions protéine/protéine par l'intermédiaire du domaine UHM. Cette perte de fonction induirait alors des défauts d'épissage ou d'épissage alternatif de gènes exprimés au cours du développement.

2.4. La protéine SPF45

2.4.1. Rôles de SPF45 dans l'épissage

La protéine SPF45 pour *Splicing factor 45* aussi connue sous le nom de RBM17 est un facteur d'épissage à l'origine identifié dans les complexes spliceosomaux [Neubauer et al., 1998] grâce à des analyses de spectrométrie de masse couplée à des analyses de séquences par bio-informatique. Des études d'immunofluorescence ont ensuite mis en évidence une colocalisation

de SPF45 avec la protéine U1A, composante de la snRNP U1 ainsi que les protéines SR dans les speckles. [Sampath et al., 2003].

Dans la séquence protéique de SPF45, on retrouve une région N-terminale non structurée, suivie d'un motif G-patch et d'un domaine UHM en C-terminal. Le motif G-patch adopte une conformation de type hélice α et possède sept résidus glycine hautement conservés. Les domaines à motif G-patch sont impliqués dans des interactions protéine-protéine et protéine/ARN ou ADN.

La fonction de SPF45 dans l'épissage a tout d'abord été étudiée par l'équipe de J. Valcárcel dans la régulation de l'épissage alternatif du gène *sxl* de la drosophile [Lallena et al., 2002]. La protéine SXL est un facteur d'épissage impliqué dans la régulation de l'épissage alternatif de transcrits impliqués dans la différenciation sexuelle et est exprimée de manière différentielle chez les mâles et les femelles. La protéine est également impliquée dans la régulation de l'épissage alternatif de son propre transcrit. En effet, la protéine SXL n'est pas exprimée chez les mâles car l'exon 3 alternatif du pré-ARNm *sxl* est exclu durant le processus d'épissage. Cette exclusion conduit à un décalage du cadre de lecture et l'introduction d'un codon stop prématuré lors de la traduction. La protéine SXL résultante est alors tronquée et non fonctionnelle. Exprimée chez les femelles, la protéine SXL induit l'inclusion de l'exon 3 alternatif du pré-ARNm *sxl* par l'intermédiaire de son interaction avec la séquence de Polypyrimidine en 3' au niveau de l'intron en amont de l'exon 3 alternatif, empêchant ainsi l'accès à la machinerie de l'épissage. Ainsi, la traduction du transcrit contenant l'exon alternatif produit une protéine SXL fonctionnelle.

Des expériences de crosslink par irradiation aux UV d'un extrait nucléaire ont permis d'identifier la protéine SPF45 liée au site d'épissage en 3' de l'intron en amont de l'exon 3 alternatif du pré-ARNm *sxl*. Il a également été démontré que l'interaction de SPF45 avec l'intron était dépendante de l'association du facteur d'épissage U2AF au site d'épissage en 3' de l'intron en aval de l'exon alternatif. Des expériences d'immunodéplétion ont ensuite montré une accumulation de pré-ARNm *sxl* en l'absence de SPF45 mais également une accumulation des produits résultants de la première étape de transesterification, c'est-à-dire le premier exon libre et le deuxième exon lié à l'intron, suggérant que la protéine était impliquée dans la deuxième étape de la réaction de transesterification et ainsi jouerait un rôle dans l'activation du site accepteur AG en 3' de l'intron en amont de l'exon alternatif.

Des études impliquant une surexpression de SPF45 dans le développement de nombreuses tumeurs [Sampath et al., 2003] ont conduit les équipes de M.Sattler et J. Valcarcél [Corsini et al., 2007] à étudier le rôle fonctionnel de SPF45 dans la régulation de l'épissage de gènes impliqués dans la prolifération tumorale. Des expériences de surexpression transitoire de SPF45 dans les cellules HeLa ont permis d'identifier un rôle du facteur d'épissage dans la régulation de l'épissage alternatif de l'exon 6 de FAS. Cet exon code pour le domaine transmembranaire de la protéine. L'inclusion de l'exon induit donc la production d'une protéine Fas pro-apoptotique transmembranaire tandis que l'exclusion de celui-ci induit la production d'une protéine anti-apoptotique soluble. Dans les cellules où SPF45 avait été surexprimé de manière transitoire, une augmentation spécifique de l'exclusion de l'exon 6 de Fas a été observée. Cependant le domaine impliqué dans le rôle fonctionnel de SPF45 dans la régulation de cet épissage restait encore à identifier. Pour cela, les mêmes expériences ont été effectuées en utilisant différents mutants de la protéine. La délétion du domaine G-patch réduisait l'activité d'épissage médiée par SPF45, tandis que la délétion du domaine UHM l'inhibait complètement. Une surexpression du domaine UHM seul dans les cellules était suffisante pour induire l'exclusion de l'exon 6 de Fas. Ces données indiqueraient que SPF45 régule l'épissage alternatif de Fas par l'intermédiaire d'interactions UHM-ULM.

Afin d'identifier les partenaires protéiques de SPF45 impliqués dans les interactions UHM-ULM, des expériences de GST-pull down avec les domaines ULM de SF1, U2AF⁶⁵ et SF3b155 produits *in vitro* ont été réalisées. Les résultats de ces expériences ont montré que SPF45 était capable d'engager des interactions UHM-ULM avec tous les domaines ULM analysés, mais avec une forte préférence pour le motif ULM5 de SF3B155. Des analyses de RMN de l'interaction du domaine UHM de SPF45 avec le domaine ULM5 de SF3b155 ont permis de caractériser cette interaction et de confirmer l'architecture spécifique d'une interaction UHM-ULM. Enfin, des mutations ponctuelles au niveau du domaine UHM de SPF45 ont confirmé que la régulation de l'épissage alternatif de l'exon 6 de Fas était assurée *in vivo* par des interactions UHM-ULM [Corsini et al., 2007].

Récemment, des études ont montré que la délétion de SPF45 dans les neurones de Purkinje, conduisait à une dégénérescence de ces neurones [Tan et al., 2016]. Ces observations ont conduit à réaliser des analyses d'expression différentielles du transcriptome à haut débit dans des cellules de Purkinje déficientes en SPF45. Les analyses bio-informatiques ont mis en évidence 349 gènes différentiellement exprimés dans les cellules mutantes. Parmi ces gènes,

ceux codant pour des protéines pro-apoptotiques étaient sur-représentés et leur expression était augmentée. Les résultats globaux de l'analyse de l'épissage indiquent que SPF45 jouerait un rôle de répresseur de l'épissage de gènes impliqués dans la survie cellulaire.

2.4.2. Maladies liées à des défauts de SPF45

Des expériences d'immunohistochimie ont montré une surexpression de SPF45 dans la majorité des carcinomes de la vessie, du sein, du colon, du poumon, des ovaires, du pancréas et de la prostate. Cette surexpression était exclusivement observée dans les cellules épithéliales et conduisait dans la plupart des cas à des résistances aux médicaments anti-cancéreux [Sampath et al., 2003].

Par ailleurs, des études ont montré que la surexpression de SPF45 dans des cellules FaDu de carcinomes hypopharyngiens induisant de manière significative la prolifération cellulaire, augmentait l'apoptose et bloquait les cellules en phase G2/M. Cela indiquerait que SPF45 serait impliquée dans la régulation de l'expression des gènes impliqués dans la prolifération, la progression du cycle cellulaire et la survie dans les cellules de carcinomes hypopharyngiens [Han et al., 2018].

II. Objectifs

Chez les eucaryotes supérieurs, les transcrits primaires présentent une composition complexe où les exons, séparés par les introns, doivent être joints par un mécanisme d'épissage rigoureux. Les étapes permettant la reconnaissance précise des sites d'épissage sont hautement complexes et restent encore aujourd'hui partiellement caractérisées. Plus particulièrement, les sites 3' d'épissage possèdent des motifs peu conservés et font intervenir, de manière séquentielle ou simultanée, de multiples interactions nécessaires à leur identification. Les travaux de notre équipe visent à caractériser ces interactions, en particulier le rôle des facteurs d'épissage à domaine UHM dans ce processus.

La présence de sept tryptophanes dans des motifs ULM de la séquence de SF3b155, parmi lesquels cinq ont la capacité d'interagir avec des domaines UHM [Cass and Berglund, 2006; Corsini et al., 2007, 2009; Loerch et al., 2014; Thickman et al., 2006], suggère que la protéine pourrait orchestrer des interactions avec des combinaisons de U2AF⁶⁵ ainsi que d'autres facteurs d'épissage à domaine UHM tel que CAPER α , PUF60 et SPF45.

Le gène SF3b1, codant pour la sous unité SF3b155, présente souvent des mutations au niveau de son domaine HEAT repeat, adjacent aux motifs ULM, chez des patients atteints de syndromes myélodysplasiques. Cependant, même si ces mutations ne se situent pas dans les motifs ULM, elles pourraient déstabiliser la structure de la protéine et affecter les interactions impliquant ces domaines. En effet, les motifs ULM de SF3b155 sont concentrés dans une région intrinsèquement désordonnée et dépourvue de structure tridimensionnelle stable en absence d'interactions avec des domaines UHM [Loerch and Kielkopf, 2016].

Les analyses d'interactions UHM-ULM par RMN et Gst- pull down entre les domaines ULM de SF3b155 et les domaines UHM de CAPER α , PUF60 et SPF45 [Loerch et al., 2014; Corsini et al., 2009, 2007] ont mis en évidence une structure d'interaction impliquant un tryptophane contenu dans un ULM de SF3b155 inséré dans une poche hydrophobe présente entre les hélices α d'un domaine UHM. Ces interactions sont renforcées par des interactions électrostatiques intramoléculaires entre les résidus chargés de part et d'autre du tryptophane du motif ULM [Corsini et al., 2007, 2009; Loerch et al., 2014].

D'après des expériences de Titration Calorimétrique Isotherme (ITC), trois domaines UHM de U2AF⁶⁵ peuvent s'associer simultanément à la région multi-ULM de SF3b155 [Thickman et

al., 2006]. Des expériences d'ITC et de chromatographie d'exclusion stérique ont montré que deux domaines UHM de CAPER α pouvaient interagir simultanément avec cette région multi-ULM de SF3b155 [Loerch et al., 2014]. Par ailleurs, chaque UHMSF semble posséder des affinités particulières pour des motifs ULM distincts. En effet, si les domaines UHM de U2AF⁶⁵ et SPF45 semblent interagir préférentiellement avec le motif ULM en C-terminal de SF3b155, l'affinité des domaines UHM de CAPER α et de PUF60 est plus forte pour le domaine en N-terminal de la protéine. Ces préférences de liaison à des motifs ULM particuliers suggèrent que des combinaisons d'interactions des quatre UHMSF avec SF3b155 pourraient exister dans la cellule pendant l'épissage. U2AF⁶⁵, CAPER α , PUF60 et SPF45 pourraient ainsi agir en synergie et réguler l'épissage alternatif de cibles particulières.

Dans ce contexte scientifique, notre objectif était de rechercher de nouvelles preuves de la formation de tels complexes et de proposer un modèle d'implication de ces complexes dans la reconnaissance des sites 3' d'épissage.

Notre premier objectif était de mettre en évidence l'existence d'associations entre les UHMSF U2AF⁶⁵, CAPER α , PUF60 et SPF45 dans les cellules. Pour cela, nous avons utilisé la technique de *Proximity Ligation Assay* qui permet d'analyser la proximité spatiale des protéines dans les cellules. En parallèle, pour étudier les interactions des UHMSF avec SF3b155, nous avons réalisé des expériences de GST-pull down avec la protéine SF3b155 purifiée et incubée avec un extrait de cellules HEK comme source de facteurs UHMSF.

Notre deuxième objectif a été de rechercher un lien fonctionnel des UHMSF dans la régulation de l'épissage. Pour cela, le taux des quatre facteurs d'épissage a été réduit dans les cellules HeLa par transduction virale de shRNA et l'épissage alternatif de 65 exons cassette préalablement sélectionnés par bio-informatique, en fonction de leur séquence introniques en 3', a été analysé par qPCR à moyen débit. Ces expériences ont permis de rechercher des corrélations fonctionnelles entre les UHMSF d'une part et avec des séquences cibles particulières d'autre part.

Les protéines CAPER α et U2AF⁶⁵ présentent des structures très similaires, incluant un domaine RS en N-terminal, deux domaines RRM centraux et un domaine UHM en C-terminal. Notre troisième objectif a alors été de rechercher un lien fonctionnel dans l'épissage entre U2AF⁶⁵ et CAPER α . Pour cela, nous avons réalisé des expériences de sédimentation de U2AF⁶⁵ et de

CAPER α en présence du facteur d'épissage SF3b155 et de séquences cibles d'ARN particulières.

En nous appuyant sur ces résultats, nous avons établi un modèle d'action de ces facteurs d'épissage à domaines UHM sur l'épissage de cibles distinctes en fonction de la composition de leurs séquences introniques en 5' de l'exon cassette.

III. Matériel et Méthodes

1. Matériel

1.1. Souches et plasmides utilisés

a) Souches bactériennes utilisées

La souche DH5 α de la bactérie *E. Coli* (*supE44*, Δ *lacU169* (ϕ 80 *lacZ* Δ M15), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*) a été utilisée pour l'amplification d'ADN plasmidique.

Les souche BL21 (DE3) et C43 (DE3) de la bactérie *E. coli* (*B F- ompT gal dcm lon hsdSB(rB-mB-) λ (DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λ S))*) ont été utilisées pour la production de protéines recombinantes.

b) Cellules d'organismes eucaryotes utilisés

Deux lignées de cellules différentes ont été utilisées pour préparer les extraits cellulaires totaux, les extraits nucléaires, dans les expériences d'immunofluorescence, les préparations de shRNA et les transductions de virus.

- Cellules HeLa S3 (ATCC : CCL-2) : lignées cellulaires humaines issues de tumeurs cancéreuses du col de l'utérus. Ces cellules ont été utilisées dans la préparation d'extraits nucléaires pour les expériences de GST Pull-down, dans la détection des protéines par immunofluorescence et par *Proximity Ligation Assay* et dans les transfections de shRNA et de plasmides pour les préparations d'ARN.
- Cellules HEK 293 (ATCC : CRL-1573TM) : Lignées cellulaires humaines issues de reins embryonnaires. Ces cellules ont été utilisées pour la préparation d'extraits cellulaires totaux pour les expériences de GST Pull-down.

c) Plasmides utilisés

pGEX6p1: ce plasmide a été utilisé pour insérer des séquences codant pour les protéines CAPER α , PUF60, SPF45 et U2AF⁶⁵ précédées de l'étiquette Glutathion S-transférase (GST). L'expression du gène est sous le contrôle du promoteur tac et induite par l'ajout d'IPTG (β -D-thiogalactoside). Dans la séquence du vecteur se trouve un gène codant pour le répresseur Lacq qui se lie à la région opérateur du promoteur tac et inhibe l'expression du gène jusqu'à l'induction par l'IPTG et maintient ainsi le contrôle de l'expression de l'insert.

Le plasmide possède également un site de multicolonnage MCS avec six sites de restrictions afin de faciliter le clonage des inserts. On note également la présence d'un site spécifique de clivage reconnu par la protéase PreScission entre l'étiquette GST et le site de multicolonnage.

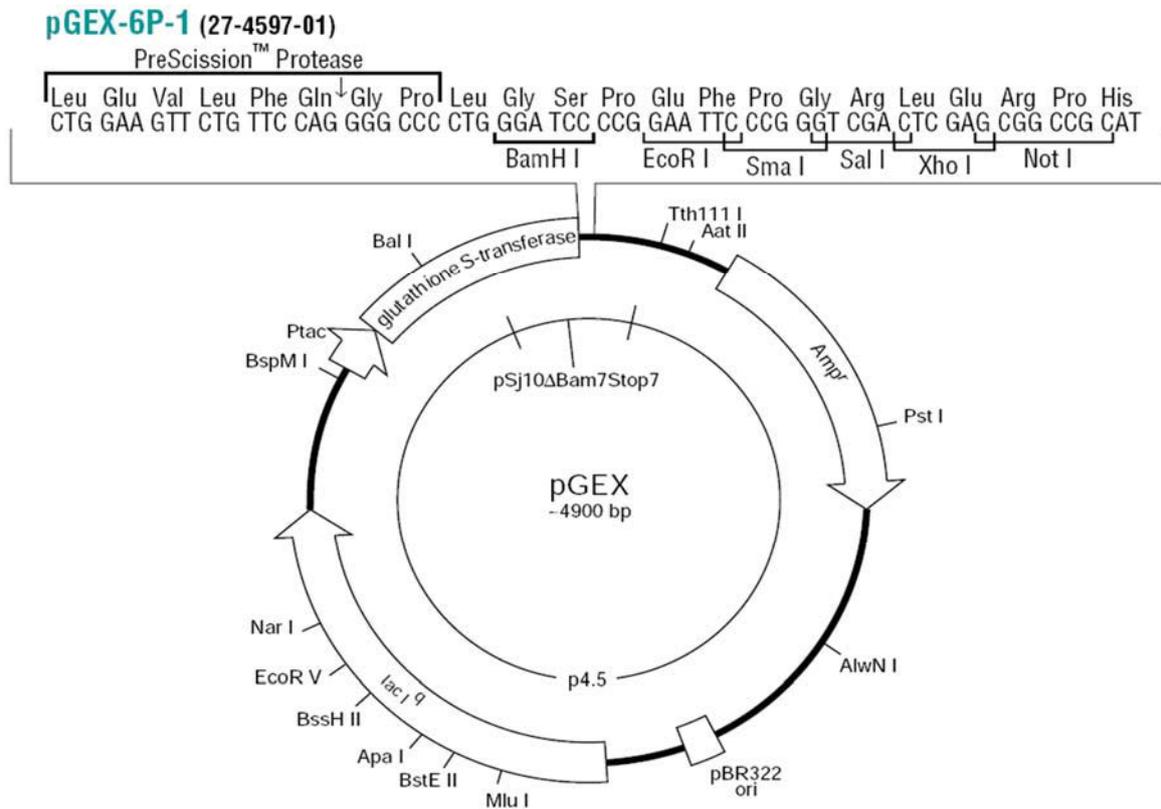


Figure 24. Carte du plasmide pGEX-6P-1 utilisé pour réaliser des constructions de U2AF⁶⁵, CAPERa, PUF60 et SPF45 précédées d'une étiquette GST pour produire les protéines in vitro.

pGL4.16 : Ce plasmide a été utilisé pour insérer les séquences TJP1, DNМ2, TJP1iDNМ2, DNМ2iTJP1, TJP1_SXL et TJP1_9U pour la préparation de constructions utilisées dans les expériences de minigènes rapporteurs. Les séquences sont insérées en amont de la séquence du gène LUC2CP (*Photinus pyralis*) codant pour la luciférase. Le vecteur contient un gène de sélection pour la résistance à l'antibiotique hygromycine dans les cellules de mammifères, ainsi qu'un nombre réduit de séquences régulatrices afin de réduire le bruit de fond et limiter les risques de défauts durant la transcription. Le vecteur pGL4 ne possède pas de promoteur. La séquence du promoteur CMV a été préalablement clonée dans la région située au niveau du site de multicolonnage MCS. Le gène LUC2CP codant pour la luciférase contient deux séquences de déstabilisation de la protéine hCL1 et hPEST, lui permettant de répondre plus rapidement aux variations de l'activité transcriptionnelle et avec une magnitude plus élevée que le gène luc2.

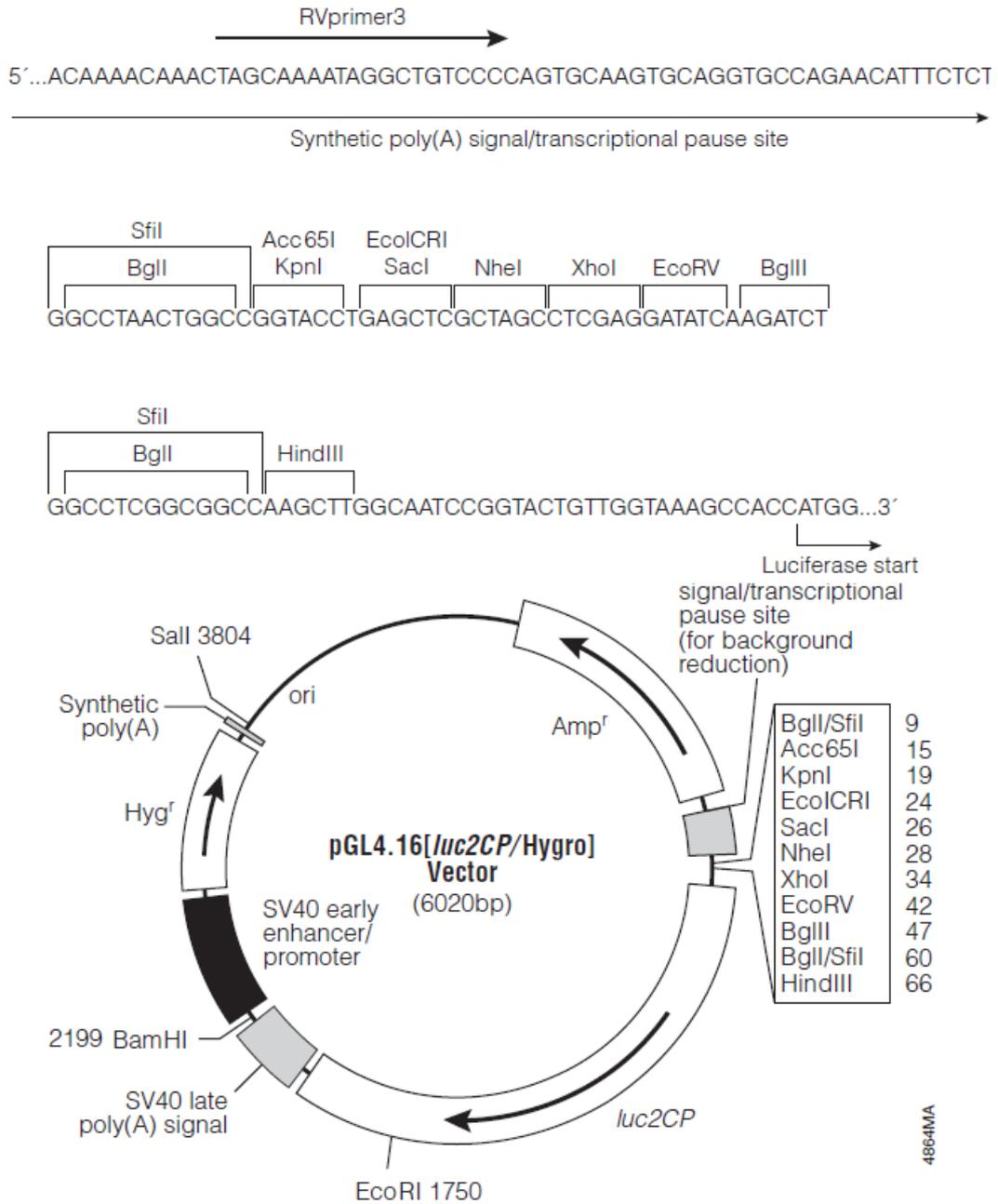


Figure 25. Carte du plasmide pGL4.16 utilisé pour la préparation des constructions TJP1, DNM2, TJP1iDNM2, DNM2iTJP1, TJP1_SXL et TJP1_9U utilisées dans les expériences de minigènes rapporteurs.

Les autres plasmides recombinants utilisés au cours de l'étude sont présentés dans le Tableau 8.

Tableau 8 : Les différents plasmides recombinants utilisés au cours de l'étude.

Plasmides recombinants	Caractéristiques
pGEMT-CAPERα	Plasmides recombinants utilisés pour l'amplification par PCR des séquences codant CAPER α , SPF45, PUF60, U2AF ⁶⁵ , et U2AF ⁶⁵ Δ RS
pGEMT-SPF45	
pGEMT-PUF60	
pGEMT-U2AF⁶⁵	
pGEMT-U2AF⁶⁵ΔRS	
pTRC_CAPERα 1	Vecteur lentiviral utilisé pour la production de shRNA dirigés contre les protéines humaines CAPER α , PUF60, SPF45 ou U2AF ⁶⁵ . Les vecteurs pTRC_NS1 et pTRC_NS2 sont utilisés comme contrôle négatif et ne ciblent aucune protéine dans les cellules.
pTRC_CAPERα 2	
pTRC_U2AF⁶⁵ 1	
pTRC_U2AF⁶⁵ 2	
pTRC_PUF60 1	
pTRC_PUF60 2	
pTRC_SPF45 1	
pTRC_SPF45 2	
pTRC_NS 1	
pTRC_NS 2	
psPAX2	Plasmides utilisés pour exprimer les protéines codant pour l'enveloppe lentivirale
pMD2.G	
pCDNA 3	Plasmide utilisé comme dominant négatif dans les expériences de transfection de minigènes rapporteurs.

1.2. Oligodésoxyribonucléotides amorces et shRNA utilisés

Les séquences des shRNA utilisés pour cibler les séquences des ARNm codant pour CAPER α , PUF60, SPF45 et U2AF⁶⁵ ou shRNA NS sont répertoriées dans le Tableau 9.

Tableau 9 : Séquences des shRNA utilisés au cours de l'étude

Nom	Identification TRC	Séquence
CAPERα 1	TRCN0000021770	RHS3979-9589178 GCTGGACCTATGAGGCTTTAT
CAPERα 2	TRCN0000021772	RHS3979-9589180 CCAACCTACCACAACCTGTTT
U2AF⁶⁵ 1	TRCN0000001162	RHS3979-9569614 CGACGAGGAGTATGAGGAGAT

U2AF⁶⁵ 2	TRCN0000001164	RHS3979-9569616 CGCCTTCTGTGAGTACGTGGA
PUF60 1	TRCN0000001201	RHS3979-9569654 GATGAAGCAGTACGGATATTT
PUF60 2	TRCN0000001203	RHS3979-9569656 AGATGAAGATTATGAGCGAGA
SPF45 1	TRCN0000017248	RHS3979-9584937 CGATGACATCAAGAGCGTGTT
SPF45 2	TRCN0000017252	RHS3979-9584941 TGCAGAAATCATTGTCAAGAT

Les oligosésoxyribonucléotides amorces utilisés pour l'amplification des gènes codant pour les protéines CAPER α , PUF60, SPF45 et U2AF⁶⁵ à partir de plasmides (voir Tableau 8), pour l'amplification des séquences codant pour TJP1 et DNM2, pour la réalisation des constructions hybrides TJP1iDNM2, DNM2iTJP1, TJP1_SXL et TJP1_9U et pour la transcription *in vitro* des ARN TJP1 et DNM2 sont répertoriés dans le Tableau 10.

Tableau 10: Liste des oligosésoxyribonucléotides amorces utilisés pour les expériences de clonage et de transcription *in vitro*.

Nom	Séquence 5' à 3'	Usage
CAPERα_f	CCAGGGGCCCCTGGGATCCATGGCAGACGATA TTGATATTGAAGCAA	Amplification des séquences codant pour les protéines CAPER α , PUF60, SPF45 et U2AF ⁶⁵ .
CAPERα_r	TCGAGTCGACCCGGGAATTCCTATCATCGTCT ACTTGGAAACCAGTAGC	
PUF60_f	CCAGGGGCCCCTGGGATCCATGGCGACGGCGA CCATAGCT	
PUF60_r	CGAGTCGACCCGGGAATTCCTATCACGCAGAG AGGTCACTGTTATCA	
U2AF⁶⁵_f	CCAGGGGCCCCTGGGATCCATGTTCGGACTTCG ACGAGTTCGAGC	
U2AF⁶⁵_r	CGAGTCGACCCGGGAATTCCTATCACCAGAAG TCCCGGCGGTG	
SPF45_f	CCAGGGGCCCCTGGGATCCATGTCCCTGTACG ATGACCTAGG	Amplification d'une partie du pré-ARNm codant pour TJP1 et DNM2 à partir de l'ADNg de cellules HeLa
SPF45_r	GAGTCGACCCGGGAATTCCTATCAAACCTTGTT CTGCCAAATCCAAG	
TJP1_f	ACTGTTGGTAAAGCCACCATGGAAGGATTTAAG CCAGCCTCTCA	
TJP1_r	GGGCCCTTCTTAATGTTTTTGGCATCTTCaccaccG GGATATGGATCCTTCTATACACC	
DNM2_f	TTGGTAAAGCCACCATGGAA AAGAAGAGAGCCATCCCCAA	
DNM2_r	TGGGCCCTTCTTAATGTTTTTGGCATCTTCACCA CCCAGGCTGAT	

TJP1iDNM2_f	TGTGTGTTTGTCTTAACACAGCAATGTACTTTTT TTTTCTTTTGCTGTTTTTTCC	Réalisation de constructions hybrides.
TJP1iDNM2_r	GGGACTGGAGATGAAGCTTCTGCTTTCTTTGAA TATGCAAAGTTAGGGTTG	
DNM2iTJP1_f	AGCTCGGCCTGTGTAAAGCCCCGGATTTTTTTTT TTCCCAAAAAGT	
DNM2iTJP1_r	AGCATGTACTTACCAGGATCTCCCCCTGTGAAG TGTTTAAAATATTTTAA	
TJP1_sxl_f	GTAACAGAAACACTATATTTAAAATATTTTAA ACGUUUUUGUUGUUUUUUUCUACAGAA	
TJP1_sxl_r	ACTGGAGATGAAGCTTCTGCTTTCTGTAGAAAA AAAACAACAAAAACGTTTA	
TJP1_9U_f	AACAGAAACACTATATTTAAAATATTTTAAACA CTTTTTTTTTTACAGAAAGCA	
TJP1_9U_r	TGGAGATGAAGCTTCTGCTTTCTGTGAAAAAAA AAGTGTTTAAAA	
TJP1_if	TAATACGACTCACTATAGGGTTAGGACATGAGG TCATAATTTTGT	Amplification de 130 bases d'une partie de l'intron des pré-ARNm codant pour TJP1 et DN2.
TJP1_ir	CTGTGAAGTGTTTAAAATATTTTAAATAT	
DNM2_if	TAATACGACTCACTATAGGGCCTGCTCTTCTCTT CTCCTGT	
DNM2_ir	CTTTGAATATGCAAAGTTAGGGTTGT	

Les oligosésoxyribonucléotides amorces utilisés pour les expériences de qPCR sont répertoriés dans le Tableau 11.

Tableau 11 : Liste des oligosésoxyribonucléotides amorces utilisés pour les expériences de qPCR

Nom	Séquence 5' à 3'	Usage
CAPERα_f2	ACAGAGCTGCAGCAATGGCAAAC	Oligonucléotides utilisés pour la vérification de l'efficacité des shRNA par qPCR.
CAPERα_r3	CTTTCCAGCAGACTGAAGC	
U2AF⁶⁵_f1	CGGCAGCTCAACGAGAATAAA	
U2AF⁶⁵_r1	GGGAACGAATCAGTCCACCG	
PUF60_h_f1	GACCTCTCAGACGATGACATCA	
PUF60_h_r1	TCTCGTACTCAATGAAGCCGT	
SPF45_f1	CAGAAGGCTGGTCCAAAAAC	
SPF45_r4	GCTTGCTGGTCTTCTCCACT	Oligonucléotides utilisés pour l'amplification du cDNA GAPDH utilisé comme normalisateur.
GAPDH_f1	CCTCCTGCACCACCAACTGCTTA	
GAPDH_r1	GTGATGGCATGGACTGTGGTCAT	

pGL_Luc2_r	ttcCATGGTGGCTTTACCAAC	Oligonucléotides reverse utilisés pour l'amplification des minigènes rapporteurs TJP1, DN2, TJP1iDN2, DN2iTJP1, TJP1_SXL et TJP1_9U par qPCR
------------	-----------------------	--

1.3. Anticorps utilisés

Les différents anticorps primaires utilisés pour les expériences d'immunofluorescence et de Western blot sont listés ci-dessous dans le Tableau 12.

Les anticorps secondaires utilisés en immunofluorescence sont couplés à un fluorochrome Alexa-Fluor-448 ou Alexa Fluor-555 et sont dirigés contre des anticorps primaires de lapin ou de souris (Molecular probes, dilution 1/2000). Les anticorps secondaires utilisés en western blot sont couplés à un IRDye 800CW ou 680CW et sont dirigés contre des anticorps primaires de lapin ou de souris (LI-COR Bioscience, dilution 1/4000).

Tableau 12 : Liste des Anticorps utilisés en western blot et immunofluorescence

Dénomination	Hôte	WB	IF	Origine
Anti-PUF60 A302-818A	Lapin polyclonal	1/2000	1/400	Bethyl
Anti-PUF60 sc-398785	Souris monoclonal		1/400	Santa Cruz
Anti-SPF45 A302-497A	Lapin polyclonal	1/2000	1/400	Bethyl
Anti-CAPERα sc-101103	Souris polyclonal	1/2000	1/200	Santa Cruz
Anti-CAPERα A300-291A	Lapin polyclonal		1/200	Bethyl
Anti-U2AF⁶⁵ MC3	Souris monoclonal	1/10.000	1/10000	Sigma
Anti-U2AF65 A303-667A	Lapin polyclonal		1/500	Bethyl
Anti-SF3b155	Souris monoclonal	1/1000	1/100	abcam
Anti-Tubulin E7	Souris monoclonal	1/10.000		P.Curmi
Anti-Fus NB100-565	Lapin polyclonal		1/150	Novusbio
Anti-SC35	Souris monoclonal		1/10	John Bell

Anti-GST	Souris monoclonal	1/1000	Santa Cruz
----------	----------------------	--------	------------

1.4. Tampons, solutions et milieux de culture utilisés

Les tampons et solutions d'usage utilisés au cours de l'étude sont répertoriés dans le Tableau 13.

Tableau 13 : Tampons et solutions utilisés

Condition d'utilisation	Dénomination	Composition
Tampon SDS-PAGE	Tampon de migration 10X	Tris base 0,25M 30,25g Glycine 1,92M 144g SDS 10% 50ml (20%) H ₂ O qsp 1L
	Bleu de dépôt 5X	SDS 4%(p/v) ; glycérol 17,4% (v/v) ; β-mercaptoéthanol 3,7M ; Bleu de bromophénol ; Tris- HCL125mm ; pH 6,8
Western blot	Tampon de transfert 10X	Tris base 0,25M 30,25g Glycine 1,92M 144g H ₂ O qsp 1L
	Solution de transfert 1X	Blotting buffer 10X 100mL EtOH 96% 200mL H ₂ O 700mL
	Rouge Ponceau	Acide acétique 5%(p/v) ; Ponceau S 0.1% (p/v)
	TBS 20X	Tris base 193.6 g NaCl 640 g H ₂ O qsp 4L
	Solution de lavage	TBS 20X 50mL Tween 1% 10mL (0.1%) H ₂ O qsp 1L
	Solution de blocage	Solution de lavage 50 mL Lait demi-écrémé 2,5g (5%)
	Gel de polyacrylamide SDS-PAGE 10%	Gel de concentration
Gel de Séparation		4ml H ₂ O; 3,2ml Acrylamide/bisacrylamide 30% ; 2,6mlTris-HCL 1M pH 8,8; 100µl SDS 10%; 100µl APS 10%; 10µl TEMED

Electrophorèse d'ADN en gel d'agarose	TAE 50X	Tris base Acide acétique EDTA 0,5M H ₂ O	242 g 57,1 mL 100 mL qsp 1L
Préparation cellulaires	d'extraits Buffer « 3 »	50mM Tris-HCL 100mM NaCl ; 1% NP40 ; 1X Protéase inhibiteur ; 10µg/ml RNaseI ; 1mM DTT	pH 8 ; 1mM EDTA ;
Préparation nucléaires	d'extraits Buffer A	10mM HEPES-KCL 10mM KCL ; 0,1mM EDTA ; 0,1mM EGTA ; 1mM DTT ; 1X Protéase inhibiteur.	pH7,9 ;
Tampon d'interaction pull down	GST- Buffer I	10mM Tris-HCL NaCl ; 1mM EDTA ; 1% NP40 ; 1X Protéase inhibiteur ; 10% Glycérol ; 1mM DTT	pH 8 ; 50mM

Les milieux de culture utilisés lors de l'étude sont répertoriés dans le Tableau 14.

Tableau 14 : Milieux de culture utilisés

	Milieu de culture de cellules de mammifères	Milieu de culture de bactéries E.coli
Composition	Milieu DMEM <i>Dulbecco's Modified Eagle Medium</i> (gibco), high glucose (4,5 g/L, L-Glutamine (45g/L) + Serum de veau foetal 10% (v/v) (gibco)	Milieu de culture liquide riche LB-broth (SIGMA). Milieu de culture solide riche LB-Agar (SIGMA).
Antibiotiques	<i>Penicillin-Streptomycin</i> (10,000 U/mL) (gibco) 0.5% (v/v) <i>Puromycine</i> (1µg/ml) pour la sélection des cellules transduites par shRNA	<i>Ampicilline</i> 100mg/ml Pour la sélection des vecteurs recombinants portant le gène de résistance à l'antibiotique.
Condition de stérilisation	Les produits sont fournis stériles et doivent être mélangés sous hotte à flux laminaire.	Le milieu est autoclavé 20 min à 120 °C sous une pression de 1 bar.

2. Méthodes

2.1. Méthodes de biologie moléculaire

2.1.1. Préparation de constructions de plasmides par la méthode de RF-cloning

a) Principe de la méthode

La méthode classique de clonage de fragments d'ADN implique généralement la coupure du plasmide de destination et du fragment d'ADN à insérer par des enzymes de restrictions, suivi d'une ligation par une ligase des bouts cohésifs générés en 5' et en 3'. Cette technique requiert la présence de mêmes sites de restriction dans le plasmide et dans le fragment d'ADN, ce qui est souvent limitant. Plusieurs méthodes de clonages indépendantes d'enzymes de restrictions et de ligase ont été développées en parallèle. On peut citer les techniques de clonage Gateway® (Invitrogen) ou In-Fusion™ (Clontech) qui requièrent la présence de séquences de recombinaison spécifiques et/ou l'utilisation de recombinases.

La méthode de RF-cloning pour *Restriction Free cloning* [Bond and Naus, 2012] ou clonage sans enzymes de restriction, permet de dépasser les limites de la présence impérative de sites de restrictions compatibles entre l'insert et le plasmide de destination. La technique est basée sur le principe de mutagenèse par extension de séquences chevauchantes [Ho et al, 1989] (Figure 26). Une paire d'oligonucléotides amorces est synthétisée de manière à posséder des séquences complémentaires à la fois à la séquence du plasmide et au fragment d'ADN à insérer. La séquence d'ADN est ensuite amplifiée par PCR en utilisant cette paire d'amorces et le produit PCR purifié est ensuite utilisé comme « megaprimer » dans une deuxième réaction de PCR en présence du plasmide de destination (Figure 26). Durant cette deuxième réaction de PCR, le plasmide est amplifié avec comme amorce le « megaprimer » préalablement synthétisé,

insérant ainsi le fragment d'ADN d'intérêt. Le plasmide de départ est ensuite éliminé par un traitement à l'enzyme de restriction DpnI qui dégrade les fragments d'ADN méthylés dans les bactéries, laissant ainsi les plasmides néosynthétisés.

Durant ma thèse, j'ai utilisé cette méthode pour réaliser des constructions plasmidiques pour deux études différentes. La première a été pour insérer les fragments d'ADN codants pour les protéines CAPER α , U2AF⁶⁵, PUF60 et SPF45 dans le vecteur pGEX-6P-1 afin de produire les protéines *in vitro* pour réaliser les expériences de GST-pull Down. La deuxième a été pour

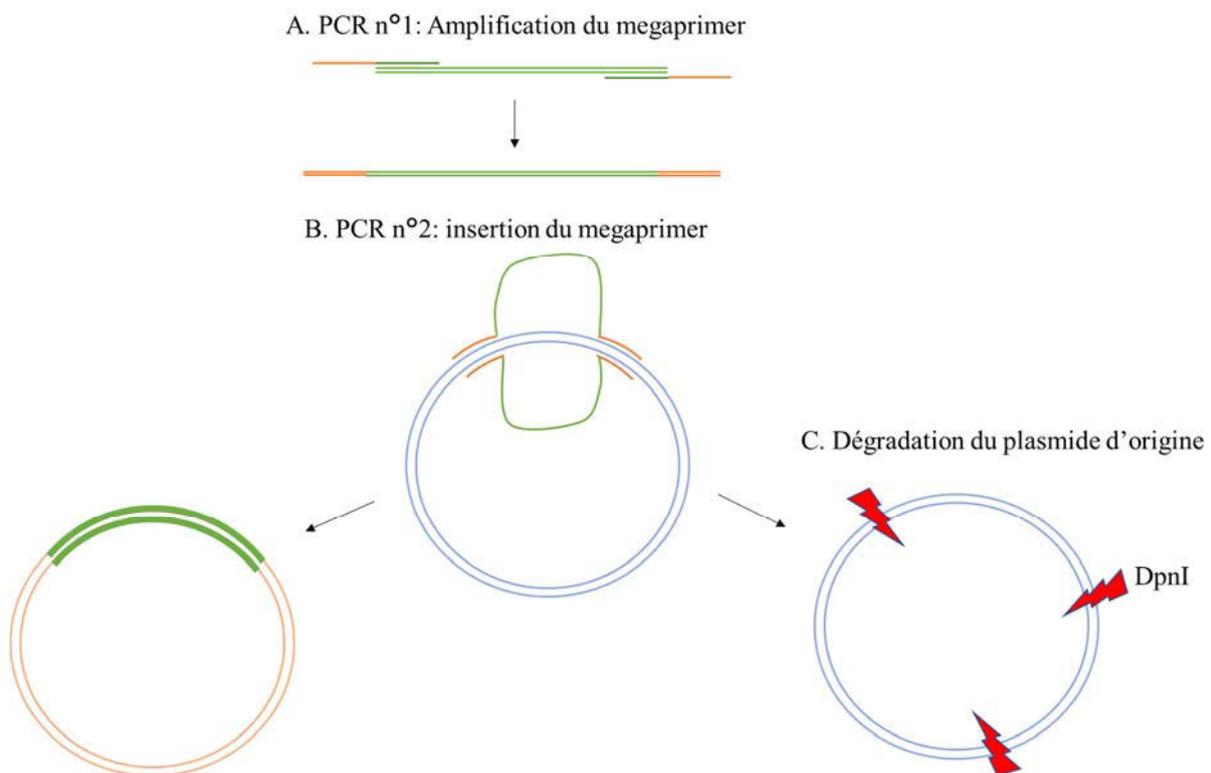


Figure 26. Schéma représentant l'insertion d'un fragment d'ADN par RF-cloning. Les oligonucléotides hybrides sont dessinés de manière à être complémentaires à l'insert (vert) et au plasmide (orange). A : Une première PCR est réalisée pour générer le « megaprimer » contenant l'insert flanqué de séquences complémentaires au plasmide de destination ; B : Une deuxième PCR est ensuite réalisée en utilisant le megaprimer comme amorce pour l'amplification du plasmide de destination (bleu) ; C : Le plasmide d'origine est dégradé par l'enzyme DpnI qui dégrade les ADN méthylés dans les bactéries.

insérer les fragments TJP1, DN2M2 dans le vecteur Luciférase pGL4.16 et réaliser les constructions hybrides TJPIiDN2M2, DN2M2iTJP1 TJP1_9U et TJP1_sxl pour les études de minigènes rapporteurs d'épissage par qPCR.

b) Protocole de clonage

Les paires d'oligonucléotides utilisées pour le clonage sont générées par le site <http://rf-cloning.org/>. En introduisant la séquence du plasmide de destination, la position du site

d'insertion et la séquence du fragment d'ADN à insérer, une paire d'oligonucléotides complémentaires au plasmide et au fragment d'ADN est dessinée afin de remplir toutes les conditions expérimentales.

Les fragments d'ADN d'intérêt sont amplifiés par PCR selon les conditions décrites dans le Tableau 15 à partir de plasmides les contenant (10 ng) pour les séquences de CAPER α , U2AF⁶⁵, PUF60 et SPF45, à partir d'ADN génomique de cellules HeLa (1 μ g) pour la séquence de TJP1 ou d'une matrice synthétique d'ADN pour la séquence de DNMT2 dans un mélange contenant l'ADN polymérase phusion (Biolabs) (2000 U/ml), un couple d'oligonucléotides amorces spécifiques à la séquence (5 μ M) et un mélange de dNTP (2mM) le tout dans un tampon spécifique de l'enzyme.

Tableau 15 : Cycles d'amplification des fragments d'ADN par PCR

Conditions de PCR			
Dénaturation	98°C	30 sec	X1
Dénaturation	98°C	10 sec	X35
Annealing	50°C – 55°C	30 sec	
Extension	72°C	30-45sec/Kb	
Extension Finale	72°C	5 min	X1

Les ADN obtenus par PCR sont purifiés sur gel d'agarose avec le kit NucleoSpin[®] Gel and PCR Clean-up (Macherey-Nagel) selon les recommandations du fournisseur, puis quantifiés par mesure d'absorbance à 260 nm (NanoDrop, Thermo Fisher Scientific).

La séquence purifiée est ensuite utilisée comme megaprimer dans une deuxième réaction de PCR dont les conditions sont décrites dans le Tableau 16 avec le vecteur de destination utilisé comme matrice. Pendant l'amplification du plasmide, le megaprimer se lie comme un ADN simple brin, et la séquence complémentaire est synthétisée permettant l'intégration de la séquence d'intérêt dans le vecteur. La quantité de megaprimer et de vecteur à utiliser pour chaque réaction de PCR est déterminée par le logiciel <http://rf-cloning.org/> lors du choix des paires d'oligonucléotides. Le plasmide d'origine est ensuite dégradé avec l'enzyme de restriction Dpn1 en incubant les échantillons pendant 1h30 à 37°C.

Tableau 16 : Cycles d'amplification du plasmide et insertion du megaprimer par PCR.

Conditions de PCR			
Dénaturation	95°C	30sec	X1
Dénaturation	95°C	30sec	
Annealing	55°C	45sec/Kb	X18
Extension	68°C	10min	

Les cellules E.coli DH5 α (50 μ l) super-compétentes sont ensuite transformées avec 1 μ l de produit PCR, en effectuant un choc thermique (30 min sur la glace puis 45 sec à 42°C). Les bactéries transformées sont ensuite mises en culture sur milieu LB solide une nuit à 37°C et sélectionnées par leur résistance à l'ampicilline (voir tableau 14). Le lendemain, une PCR est effectuée selon les conditions décrites dans le Tableau 17 en utilisant comme matrice d'ADN les colonies bactériennes résistantes à l'antibiotique dans un mélange contenant L'ADN polymérase Dream Taq (MBI Fermentas) et un couple d'oligonucléotides amorces sens complémentaire à la séquence d'ADN et antisens complémentaire au plasmide, ainsi qu'un mélange de dNTP dans un tampon spécifique de l'enzyme.

Tableau 17 : Cycles d'amplification d'ADN provenant de colonies bactériennes par PCR

Conditions de PCR			
Dénaturation	95°C	1 min	X1
Dénaturation	95°C	30 sec	X30
Annealing	50°C – 55°C	1 min	
Extension	72°C	30-45sec/Kb	
Extension	72°C	7 min	X1

Les produits PCR sont ensuite déposés sur gel d'agarose 1%. Les clones sont sélectionnés selon la taille de migration du produit d'amplification correspondant à celle attendue puis mis en culture sur milieu LB liquide en présence d'Ampicilline sous agitation pendant une nuit à 37°C. Les plasmides sont ensuite purifiés en utilisant le protocole de mini-préparation (plasmid DNA purification, Macherey-Nagel) selon les recommandations du fournisseur puis quantifiés par

mesure d'absorbance à 260 nm (NanoDrop, Thermo Fisher Scientific). La bonne intégration et l'intégrité de la séquence sont ensuite vérifiées par séquençage.

Afin de synthétiser les minigènes rapporteurs hybrides TJP1iDNM2 et DN2iTJP1, des séquences de 150 pb en 3' des introns de TJP1 et DN2 sont amplifiées par PCR à partir des plasmides pGL4.16_TJP1 et pGL4.16_DN2 construits précédemment. Après purification, les séquences sont ensuite utilisées comme megaprimers dans une deuxième réaction de PCR comme décrit ci-dessus.

En ce qui concerne les minigènes rapporteurs pGL4.16_TJP1_9U et pGL4.16_TJP1_sxl de courtes séquences de respectivement 9 et 22 paires de bases sont insérées en amplifiant le vecteur pGL4.16_TJP1 par PCR avec une paire d'oligonucléotides contenant la séquence à insérer.

2.1.2. Purification de protéines recombinantes

a) Production des protéines dans la bactérie *E. coli*

Toutes les protéines purifiées au cours de cette étude ont été produites dans la bactérie *Escherichia coli* (*E. coli*) par Jean SALONE et Alexandre MAUCUER. La bactérie *E. coli* est une entérobactérie à gram négative retrouvée à l'origine dans la flore intestinale. Elle est couramment utilisée en biologie moléculaire et plus particulièrement dans la purification de protéines car elle a la particularité de croître très rapidement et donc d'obtenir une quantité importante de protéines en un temps record.

Les souches BL21 (DE3) et C43 (DE3) ont été utilisées pour la production des protéines. Elles contiennent toutes les deux le lysogène DE3 qui porte le gène codant pour l'ARN polymérase T7 sous le contrôle du promoteur *lacUV5*, permettant d'induire la production des protéines en ajoutant l'IPTG pour *isopropyl β-d-thiogalactopyranoside*, un analogue du lactose. Une mutation au niveau du gène RNases (*me131*) réduit le niveau de RNases endogènes, induisant ainsi une meilleure stabilité du transcrit et par conséquent un meilleur rendement en protéines.

Elle exprime également le lysosyme T7, un inhibiteur de l'ARN polymérase T7 afin de prévenir de l'expression basale de l'ARN polymérase T7 avant l'induction.

La souche CE43 (DE3) est un dérivé de la souche BL21 (DE3) qui permet d'exprimer des protéines qui sont toxiques pour la bactérie. Cette souche contient une mutation génétique lui

permettant d'éviter la mort cellulaire associée à l'expression de protéines recombinantes toxiques pour la bactérie.

Les bactéries BL21 (DE3) ou C43 (DE3) sont transformées avec les constructions pGEX-6P-1 CAPER α , U2AF⁶⁵, PUF60 ou SPF45 et sont mises en culture sur milieu LB solide supplémenté de 100 μ g/ml d'Ampicilline une nuit à 37°C. Les colonies transformées sont ensuite mises en culture une nuit à 37°C dans 5ml de milieu LB supplémenté en Ampicilline. Ces pré-cultures sont ajoutées à 1L de milieu LB et mises en culture sous agitation jusqu'à obtention d'une DO600 nm comprise entre 0,4 et 0,6. L'induction de la production des protéines est ensuite initiée en ajoutant 0,4 mM d'IPTG et les bactéries sont mises en culture pendant 3h à 30°C sous agitation.

Les bactéries sont ensuite centrifugées pendant 30 min à 4000 rpm. Le surnageant est éliminé et le culot est conservé à -80°C.

b) Purification des protéines

Afin d'extraire les protéines produites dans la bactérie, la membrane bactérienne doit tout d'abord être lysée. Après décongélation, le culot bactérien est resuspendu dans un tampon de lyse contenant 25 mM de Hepes pH 7.5, 1 M de NaCl, 1 mM d'EDTA, 10% de glycérol, 5 mM de DTT, 0.1 mM de PMSF (phenylmethylsulfonylfluoride) et un mélange d'inhibiteurs de protéases (Complete protease inhibitor, Roche). Les bactéries sont ensuite placées sur la glace et les membranes sont lysées par sonication à une amplitude de 30% pendant 10 min (5 sec d'impulsion, 10 sec de pause). Le lysat obtenu est ensuite centrifugé pendant 40 min à 15 000 x g à 4°C afin d'éliminer les débris cellulaires et les fractions insolubles.

Toutes les protéines produites sont précédées d'une étiquette GST en N-terminal. Afin de les extraire, le surnageant soluble est chargé sur une colonne GST-trap (GE Healthcare) en utilisant le système de chromatographie d'affinité äkta (GE Healthcare). Les protéines possédant l'étiquette GST sont liées au milieu d'affinité et les impuretés sont éliminées par lavage. Les protéines étiquetées sont ensuite éluées dans des conditions non dénaturantes en utilisant 40mM de glutathion réduit dans un tampon contenant 25 mM de Hepes pH 7.5 et 1 M de NaCl.

Le produit d'éluion est ensuite chargé sur une colonne Superdex 200 increase 10/300 (GE Healthcare) pour améliorer la purification par filtration sur gel. Les protéines purifiées sont ensuite concentrées par filtration par centrifugation et la concentration en sel présent dans le tampon est réduite à 200Mm. La concentration des protéines est ensuite déterminée par

absorbance UV à 280 nm et par gel de polyacrylamide SDS-PAGE suivi d'une coloration au bleu de Coomassie.

2.1.3. Transcription des ARN *in vitro*

Les fragments d'ADN d'intérêt sont amplifiés par PCR selon les conditions décrites dans le Tableau 14 à partir d'ADN génomique de cellules HeLa (1 µg) dans un mélange contenant l'ADN polymérase phusion (Biolabs) (2000 U/ml), un couple d'oligonucléotides amorces spécifiques à la séquence (5 µM) et un mélange de dNTP (2mM), le tout dans un tampon spécifique de l'enzyme.

Les produits d'amplification sont purifiés sur gel d'agarose avec le kit NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel) selon les recommandations du fournisseur, puis quantifiés par mesure d'absorbance à 260 nm (NanoDrop, Thermo Fisher Scientific).

Les ADN purifiés sont ensuite utilisés comme matrice dans une réaction de transcription *in vitro* avec le kit T7 Ribomax (Promega) selon les recommandations du fournisseur. La réaction de transcription *in vitro* est réalisée dans un volume de 20 µl contenant 1 µg de matrice d'ADN, 7 µl d'un mélange d'enzymes T7 Express (Promega) et 10 µl de tampon spécifique de l'enzyme (Express T7 2X Buffer, Promega). Les échantillons sont ensuite incubés pendant 30 min à 37°C.

2.2. Méthodes de biologie cellulaire

2.2.1. Culture de cellules

Les cellules HeLa et HEK 293 sont cultivées dans le milieu DMEM (pour Dulbecco's Modified Eagle Medium) supplémenté avec 10% de Sérum de Veau Foetal (FBS, Invitrogen), 100 U/ml de pénicilline et 100 µg/ml de streptomycine. Les cellules sont cultivées dans des boîtes de 10 cm de diamètre et placées dans un incubateur à 37°C, sous une atmosphère humide et à 5% CO₂. La confluence des cellules est vérifiée par microscopie. Une fois arrivées à 70-90%, les cellules sont rincées avec du PBS puis incubées avec 2 ml de trypsine 0.25% (gibco) à 37°C environ 20 min afin de décrocher les cellules. L'activité enzymatique de la trypsine est arrêtée par ajout d'un volume de milieu de culture en fonction de la dilution, du nombre de boîtes souhaitées et du type cellulaire. Pour une dilution au 10^{ème}, une nouvelle boîte de 10 cm contenant 10 ml de milieu de culture estensemencée avec 1 ml de cellules resuspendues dans 8 ml de milieu. Les cellules sont ensuite gardées en culture jusqu'à confluence.

2.2.2. Préparation d'extraits à partir de cellules

a) Préparation d'extraits totaux pour l'analyse par western blot

Les cellules sont lavées deux fois au PBS 1X et reprises dans 50 µl de Tris-HCL 0,25 M. Les extraits cellulaires sont ensuite réalisés par 4 cycles de congélation rapide à l'azote liquide suivi d'une décongélation à 37°C au bain marie. Les débris cellulaires sont éliminés par centrifugation pendant 15 min à 10 000 rpm à 4°C. La quantité de protéines totales présentes dans l'échantillon est déterminée par mesure de l'absorbance UV à 280 nm.

b) Préparation d'extraits totaux pour la réalisation d'expériences de GST-pull down

Des cellules HEK sontensemencées dans des boîtes de 15 cm de diamètre et mises en cultures jusqu'à 90% de confluence. Le milieu de culture est éliminé et les cellules sont récoltées par grattage en ajoutant 5 ml de PBS 1X par boîte. Les cellules grattées sont ensuite transférées dans des tubes de 50 ml et centrifugées pendant 5 min à 800 xg à 4°C. Après élimination du PBS, le culot est resuspendu avec 1 ml de buffer « 3 » (voir Tableau 12) par boîte et vortexé à vitesse maximale pendant 10 sec à 3 reprises. Les cellules sont ensuite transférées dans un tube de 15 ml et centrifugées 10 min à 4000 rpm à 4°C afin d'éliminer les débris cellulaires. Le surnageant est aliquoté dans des tubes de 1,5 ml et conservé à -80°C pour des utilisations ultérieures.

c) Préparation d'extraits nucléaires pour la réalisation d'expériences de GST-pull down

Des cellules HeLa sontensemencées et mises en culture dans des boîtes de 10 cm de diamètre jusqu'à ce qu'elles atteignent 90% de confluence. Le milieu de culture est éliminé et les cellules sont lavées deux fois au PBS. Les boîtes sont disposées sur la glace et mises sous agitation pendant 5 min en présence de 400 µl de buffer A (voir Tableau 12). Les cellules sont ensuite grattées et récupérées dans des tubes Eppendorf de 1,5 ml. Après ajout de 24 µl de NP40 10% (0,6% du volume final), les échantillons sont immédiatement vortexés à vitesse maximale pendant 9 sec, puis centrifugés pendant 30 sec à 20 000 xg. Le surnageant représentant la fraction cytosolique est immédiatement prélevé. Le culot composé de la fraction nucléaire est resuspendu avec 100 µl de buffer A supplémenté de 0,5% de NP40 et la membrane nucléaire est lysée par sonication à une amplitude de 40% pendant 1 min (2 sec d'impulsion, 10 sec de pause). Les échantillons sont ensuite centrifugés pendant 5 min à 10 000 xg et le surnageant représentant l'extrait nucléaire total est prélevé et conservé à -80°C pour des utilisations ultérieures.

2.2.3. Préparation de surnageants viraux de shRNA

Le système d'interférence ARN est un outil de recherche répandu utilisé pour contrôler l'expression de gènes spécifiques dans des conditions expérimentales. Les shRNA pour *small hairpin RNA* sont des petits ARN double brin avec une structure à épingle à cheveux. Ils sont utilisés pour diminuer le niveau d'expression des gènes de manière stable.

Les shRNA peuvent être introduits dans la cellule par des vecteurs lentiviraux. Ce système est un outil intéressant car il permet d'intégrer les shRNA dans le génome des cellules, quel que soit le stade de division. De plus la transduction virale présente moins de toxicité cellulaire et permet un rendement d'intégration plus élevé qu'une transfection classique.

Afin de produire des vecteurs lentiviraux portant les shRNA, les plasmides codant pour les protéines de la capsule et la reverse transcriptase ainsi que le plasmide portant la séquence du shRNA sont d'abord transfectés dans des cellules. Les particules virales sont alors produites dans la cellule et libérées dans le milieu de culture.

Afin de produire les particules virales utilisées pour la diminution du taux de CAPER α , U2AF⁶⁵, PUF60 et SPF45 ou des particules virales contrôles NS, des cellules HEK 293 sontensemencées à 1×10^6 cellules par puits et mises en culture dans un puits de 3 cm d'une plaque 6 puits. 24h après la mise en culture, les cellules sont co-transfectées avec un mélange contenant 1 μ g de plasmide psPAX2, 0,5 μ g de plasmide pMD2.G et 1 μ g de plasmide TRC portant la séquence du shRNA en utilisant le kit *Effectene Transfection Reagent* (QIAGEN) selon les recommandations du fournisseur.

Le lendemain, le milieu de culture est aspiré et remplacé par 2,5 ml de milieu frais. Entre 48 et 56 heures après la transfection, le milieu de culture est récupéré à l'aide d'une seringue et passé dans un filtre de 0.45 μ . Les surnageants viraux sont récupérés dans des tubes de 5 ml puis aliquotés dans des tubes Eppendorf de 1,5 ml et conservés à -80°C pour des utilisations ultérieures.

2.2.4. Transduction de cellules par les surnageants viraux de shRNA

a) Mécanisme d'action des shRNA

Une fois le vecteur du shRNA introduit dans la cellule par transduction virale et intégré dans le génome, la séquence du shRNA est transcrite par l'ARN polymérase II. Comme illustré dans la Figure 27, le pré-shRNA est pris en charge par la protéine Drosha et maturé en shRNA puis exporté dans le cytoplasme à l'aide de l'exportine 5. Le shRNA forme alors un complexe avec Dicer, une RNase spécifique des ARN double brin qui génère des siRNA de 21 à 23 nucléotides. Le produit résultant du clivage est ensuite incorporé dans le complexe RISC, composé des protéines Argonaute-2, Dicer et TRBP, et un des deux brins d'ARN est éliminé. Le brin d'ARN restant dirige le complexe RISC vers l'ARNm possédant une séquence complémentaire. La séquence du shRNA s'apparie avec la séquence complémentaire de l'ARNm et celui-ci est clivé ou séquestré par le complexe RISC. Il en résulte la diminution de l'expression du gène cible.

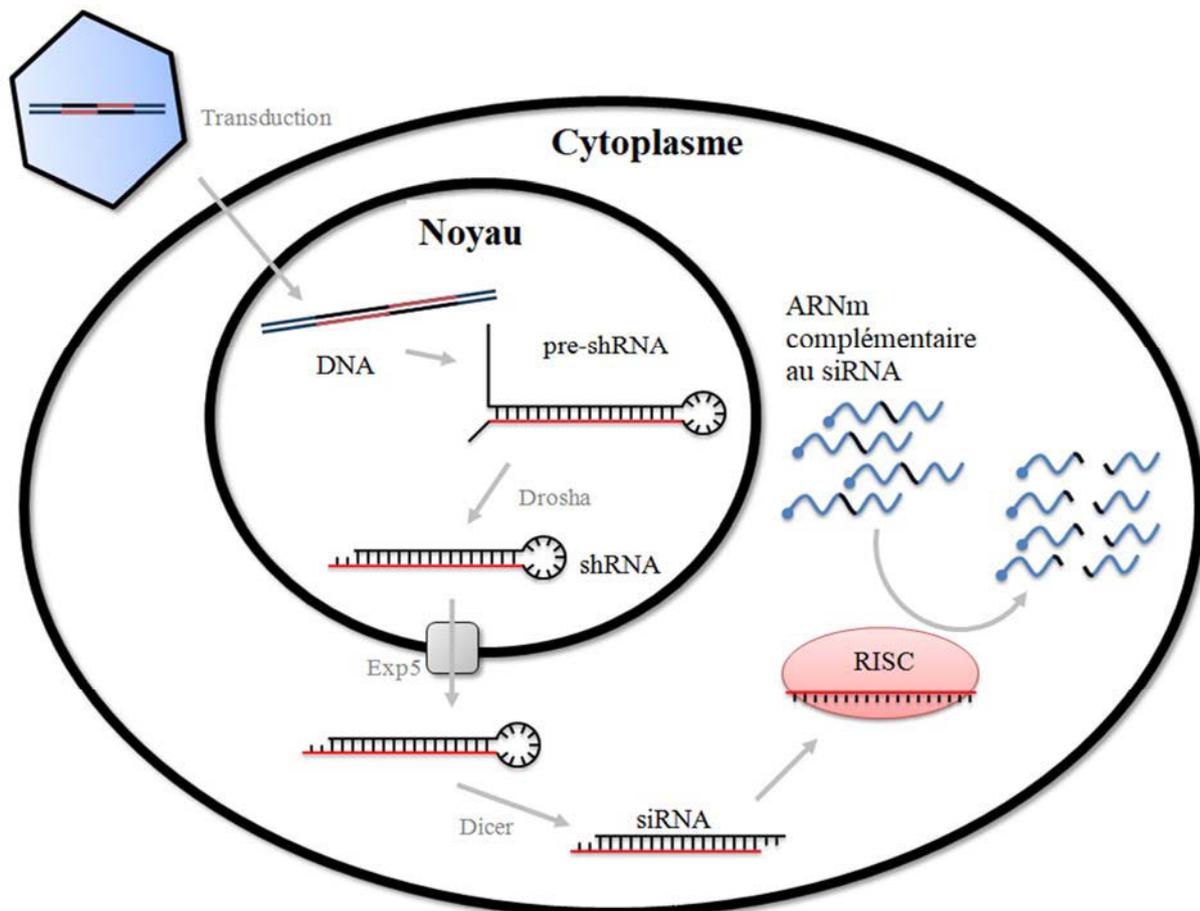


Figure 27. Mécanisme d'action du shRNA (D'après Cecconi et al, 2010)

b) Protocole de transduction de cellules par shRNA

Des cellules HeLa sont maintenues en culture à 37°C sous une atmosphère à 5% de CO₂. La veille de la transduction, 100.000 cellules par puits sontensemencées dans un puits d'une plaque 6 puits de 3 cm de diamètre contenant 3 ml de milieu de culture (DMEM, SVF 10%, Pen/strep 2,5%). Afin de diminuer la quantité de protéines CAPER α , U2AF⁶⁵, PUF60 ou SPF45, les cellules sont infectées avec 40 μ l de surnageants viraux contenant les shRNA dirigés contre les ARN cibles, dilués dans 1 ml de milieu de culture contenant du polybrène (8 μ g/ml). Deux shRNA différents dirigés contre chaque ARN cible ont été utilisés. Des cellules sont également infectées avec des shRNA n'ayant aucune cible dans les cellules HeLa, comme contrôle.

Le lendemain, la puromycine est ajoutée au milieu de culture à une concentration de 1 μ g/ml. Seules les cellules infectées se développent. La sélection est maintenue pendant 72 heures.

2.2.5. Transfection de cellules

Les transfections de plasmides recombinants contenant les minigènes rapporteurs TJP1, DNМ2 TJPIiDNМ2, DNМ2iTJP1, TJP1_9U et TJP1_sxl ont été réalisées dans des cellules HeLa préalablement infectées par des shRNA dirigés contre CAPER α , U2AF⁶⁵ ou un shRNA contrôle (NS) et maintenues en culture 48h après sélection à la puromycine. Les transfections ont été réalisées à l'aide de la lipofectamine 2000 (Invitrogen) selon les recommandations du fournisseur. Les plasmides recombinants sont incorporés dans un liposome qui est endocyté par la membrane cellulaire, induisant l'introduction du plasmide dans le cytoplasme. Le plasmide est ensuite dirigé dans le noyau où il est transcrit par L'ARN polymérase II.

Les cellules sont transfectées avec 5 ng de plasmide et 6 μ l de Lipofectamine 2000 contenu dans du milieu OPTI-MEM (invitrogen) dans un volume final de 500 μ l, puis maintenues en culture pendant 24 heures.

2.2.6. Préparation d'ARN totaux

Les cellules mises en cultures sont lavées deux fois au PBS 1X. Le réactif TriPure (1 ml) (Euromedex) est ajouté directement aux cellules dans la boîte de culture. Les cellules se décrochent instantanément et sont passées 8 fois dans une seringue de 1 ml puis transférées dans un tube de 1,5 ml et incubées 5 min à température ambiante afin de dissoudre les complexes.

200 µl de chloroforme sont ajoutés dans chaque tube. Le mélange est vigoureusement agité et incubé 3 min à température ambiante, puis centrifugé à 12000 xg à 4°C pendant 15 min. La phase aqueuse est récupérée et les ARN sont précipités par ajout de 0,5 ml d'isopropanol incubés à température ambiante pendant 10 min. Les ARN sont précipités par centrifugation à 12000 xg à 4°C pendant 10 min. Le culot est lavé avec 1 ml d'éthanol 75% et centrifugé à nouveau pendant 7 min à 12 000 xg à 4°C. Après élimination de l'éthanol, les culots sont séchés pendant 5 min à température ambiante puis repris dans 50 µl de H₂O stérile. Les culots sont solubilisés en les incubant pendant 10 min sur la glace suivi de 20 min à 70°C. La concentration des ARN resuspendus est déterminée par mesure de l'absorbance UV à 260 nm.

2.3. Etude des complexes formés par les facteurs d'épissage à domaines UHM et avec SF3b155.

2.3.1. Etude des associations entre les facteurs d'épissage à domaines UHM et avec SF3b155 dans la cellule par Proximity Ligation Assay

a) Visualisation des protéines par immunofluorescence

Des cellules HeLa sontensemencées à 10 000 cellules par puits et mises en culture sur des lamelles de 12 mm dans des puits de 14 mm de plaques 24 puits, afin de permettre leur visualisation ultérieure au microscope. Après 24 heures, le milieu de culture est retiré et les cellules sont lavées avec du PBS 1X. Les cellules sont fixées par pontage covalent en incubant les lamelles dans une solution de paraformaldéhyde 4% - PBS 1X (v/v) pendant 10 minutes à 37°C. La membrane cellulaire est ensuite perméabilisée et bloquée à l'aide d'une solution de PBS 1X/Triton 0,1%/BSA 3% pendant 30 minutes à 37°C.

Après fixation, perméabilisation et blocage des cellules, 3 lavages au PBS 1X sont réalisés afin d'éliminer les détergents. Les cellules sont ensuite incubées pendant 1h30 à température ambiante avec l'anticorps primaire dilué dans une solution de PBS 1X/Triton 0,1%/BSA 3% selon le facteur de dilution indiqué dans le tableau 11.

Après 5 lavages de 5 minutes au PBS-Triton, les cellules sont incubées avec un anticorps secondaire couplé à un fluorophore pendant 45 min à température ambiante puis de nouveau lavées 5 fois avec une solution de PBS-Triton.

Les lamelles sont ensuite montées sur lame avec 6 µL de Mowiol et conservées à 4°C jusqu'à observation au microscope à fluorescence.

b) Principe de la technique de Proximity Ligation Assay

La technique de Proximity Ligation Assay permet la détection d'une proximité spatiale entre deux protéines dans les cellules. Pour cela, des cellules fixées sur lamelles et perméabilisées sont incubées avec un mélange de deux anticorps primaires produits chez des hôtes différents et spécifiques des deux protéines étudiées (Figure 28). Des anticorps secondaires fusionnés à une courte séquence oligonucléotidique d'ADN appelés « sondes PLA » sont ensuite ajoutés et se fixent aux anticorps primaires. Des oligonucléotides complémentaires sont ajoutés et s'hybrident aux sondes. Si celles-ci sont suffisamment proches l'une de l'autre dans la cellule (entre 10 et 40 nm), elles sont liguées par une ligase et forment un ADN circulaire. L'un des oligonucléotides sert alors d'amorce et une ADN polymérase initie une réaction d'amplification en utilisant le cercle comme matrice. Lors de l'amplification, des nucléotides fluorescents s'intègrent au produit d'amplification. Le signal fluorescent correspondant à la co-localisation des deux protéines peut alors être observé en microscopie à fluorescence.

c) Protocole expérimental de Proximity Ligation Assay

Des cellules HeLa sontensemencées à 10.000 cellules par puits sur une lamelle de 12 x 12 mm déposée dans un puits d'une plaque 24 puits et mises en culture pendant 48 heures.

Après fixation et perméabilisation, comme décrit dans la partie 2.3.1.a, une solution de Blocage (Blocking solution, Duolink II Fluorescence, OLINK Bioscience) est ajoutée et les lamelles sont incubées pendant 20 min à température ambiante en chambre humide. Chaque lamelle est ensuite incubée pendant une nuit à 4°C en chambre humide dans 30 µl de mélange des deux anticorps primaires produits dans deux hôtes différents reconnaissant les deux protéines

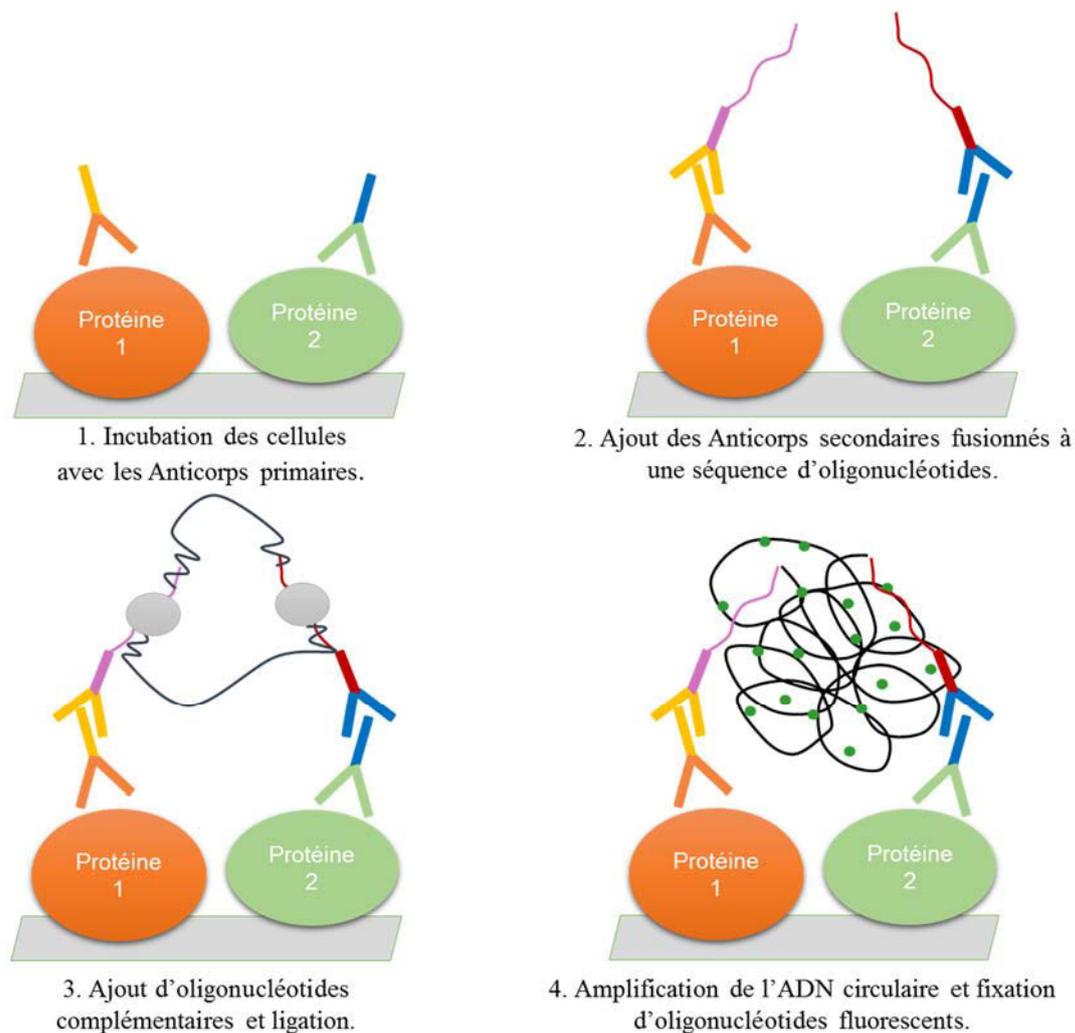


Figure 28. Principe de la méthode de Proximity Ligation Assay.

d'intérêt dilués dans une solution de PBS 1X, Triton 0,1%, BSA 3%. Après deux lavages au PBS 1X, les lamelles sont incubées pendant une heure à 37°C en chambre humide avec un mélange de sondes couplées aux anticorps secondaires PLA PLUS souris et PLA MINUS Lapin. Les lamelles sont ensuite incubées avec un mélange composé d'une solution de ligation (Ligation solution, Duolink II Fluorescence, OLINK Bioscience) diluée au 1/5^{ème} et d'une ligase diluée au 1/40^{ème} pendant 30 min à 37°C en chambre humide. Après deux lavages aux PBS 1X pendant 2 min sous agitation, une solution d'amplification (Amplification solution, Duolink II Fluorescence, OLINK Bioscience) diluée au 1/5^{ème} et une polymérase diluée au 1/80^{ème} est ajoutée aux lamelles pendant 100 min à 37°C en chambre humide. Afin de marquer le noyau, les lamelles sont incubées pendant 30 sec dans 1 ml de PBS 1X contenant du DAPI. Les lamelles sont ensuite montées sur lames avec 6 µl de Mowiol et conservées à -20°C. Les lamelles sont analysées au microscope à fluorescence Leica DM4 B en utilisant le logiciel leica Las X, puis traitées avec le logiciel ImageJ.

2.3.2. Etude des interactions directes entre les facteurs d'épissage à domaines UHM et SF3B155 par GST-Pull Down

a) Principe de la méthode

La technique de GST-pull Down est une méthode d'analyse d'interactions directes entre les protéines. Elle repose sur l'utilisation de protéines produites *in vitro* précédées d'une étiquette, immobilisées sur un ligand spécifique de haute affinité. Dans notre cas, les protéines produites au laboratoire ont été fusionnées à une étiquette GST. Durant les expériences de Pull Down, les protéines étiquetées GST sont immobilisées sur des billes d'agarose portant du glutathion réduit (Figure 29). Cette protéine appelée « appât » est mise en contact avec une source de protéines, dans notre cas un extrait cellulaire total ou un extrait nucléaire. Les partenaires protéiques présents dans l'extrait interagissent avec la protéine appât, tandis que les autres sont éliminées lors des lavages successifs. Les complexes protéiques isolés sont ensuite fractionnés sur gel SDS-PAGE et la composition de ces complexes est révélée par Western blot.

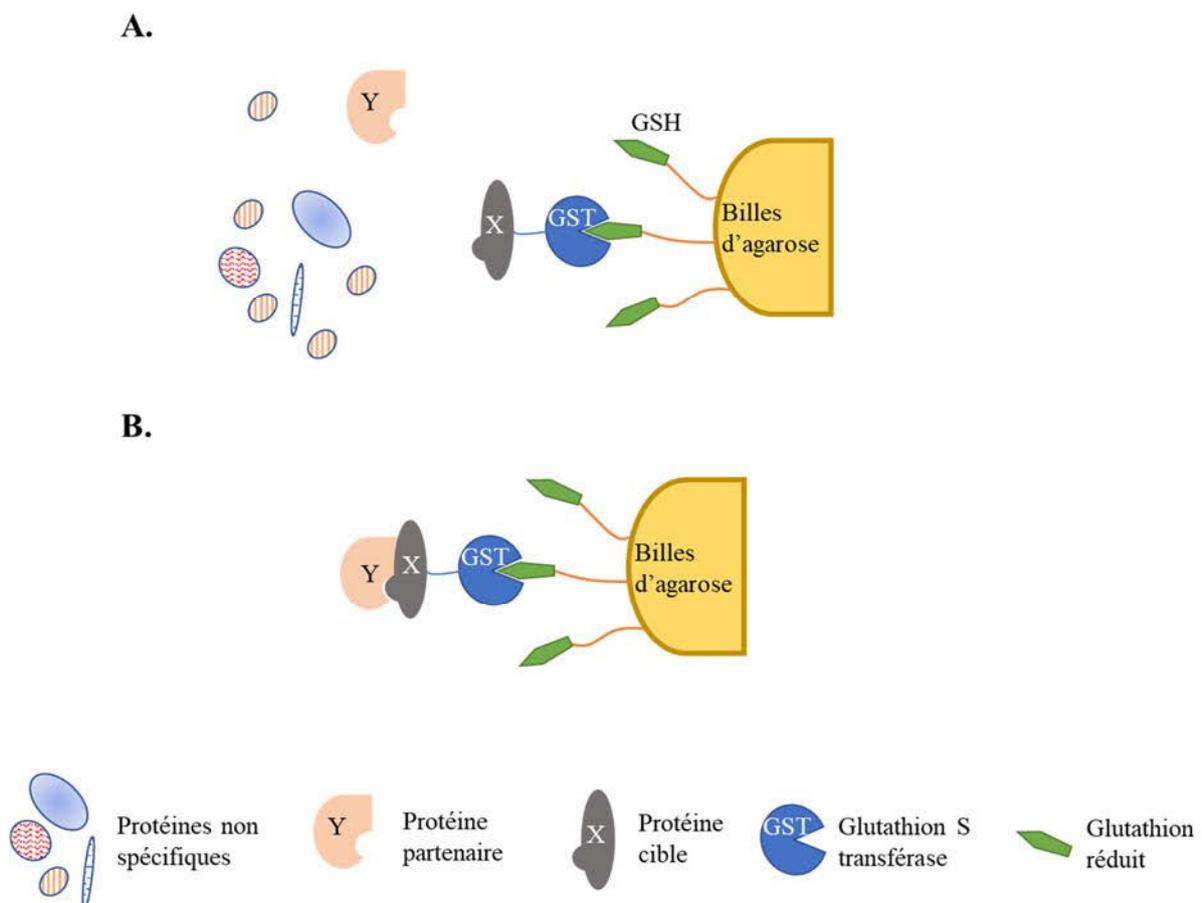


Figure 29. Représentation schématique de la méthode de GST-Pull Down : A. fixation de la protéine appât fusionnée à une étiquette GST sur les billes d'agarose et mise en contact avec une source de protéines ; B. Interaction de la protéine appât avec un partenaire protéique et élimination des protéines non spécifiques par lavage.

Des extraits cellulaires totaux de cellules HEK, préparés comme indiqué dans le chapitre 2.2.2.b, ont été utilisés pour analyser les interactions entre la protéine GST-SF3b155 produite *in vitro* et les facteurs d'épissage à domaines UHM endogènes. Des extraits nucléaires totaux de cellules HeLa, préparés comme indiqué dans le chapitre 2.2.2.c, ont été utilisés pour analyser les interactions entre les facteurs d'épissage à domaines UHM GST-CAPER α , GST-U2AF⁶⁵, GST-PUF60 et GST-SPF45 produits *in vitro* et la protéine SF3b155 endogène.

Pour chaque réaction, 40 pMoles de protéines appât fusionnées GST sont incubées pendant 1h30 sous agitation à 4°C avec 90 μ l d'extrait cellulaire ou nucléaire total dans 90 μ l de tampon d'interaction « I » (Tableau 12).

Des billes d'agarose fusionnées à du glutathion réduit sont lavées 4 fois avec 1 ml de tampon d'interaction « I » puis 40 μ l de billes sont ajoutées à chaque réaction et incubées pendant 30 min à 4°C sous agitation afin de fixer les protéines fusionnées à une étiquette GST.

Après centrifugation, les surnageants sont éliminés et les billes sont lavées 4 fois avec le tampon d'interaction « I » afin d'éliminer les protéines non spécifiques. Sont ajoutées aux billes, 14 μ l de bleu de dépôt 2 X (Tableau 12). Les protéines interagissant avec la protéine appât sont fractionnées sur gel de polyacrylamide SDS-PAGE 10% (Tableau 12) dans un tampon de migration 1 X (Tableau 12) pendant 1h30 à 120V. Un ensemble de marqueurs de taille est fractionné sur gel simultanément aux échantillons.

Après migration, les protéines sont transférées à l'aide d'un champ électrique (2h à 200mA) sur une membrane de nitrocellulose (0.45 μ m, GE Healthcare) dans une solution de transfert (Tableau 12). Les protéines transférées sont ensuite révélées par incubation de la membrane pendant 10 min dans une solution de rouge ponceau, puis fixées par incubation pendant 10 minutes dans une solution d'acide acétique 1%. Après plusieurs lavages au TBS-tween jusqu'à élimination complète du rouge ponceau, la membrane est incubée dans une solution de lait demi-écrémé 5% dans le tampon TBS-tween pendant 40 min sous agitation à température ambiante afin de bloquer les éventuels sites non spécifiques.

Afin de révéler les protéines présentes sur la membrane, celle-ci est incubée avec un anticorps primaire dilué dans une solution de TBS-Tween/Lait 1% selon le facteur de dilution indiqué dans le Tableau 11 sous agitation pendant 1h à température ambiante, ou une nuit à 4°C. Après 5 lavages de 5 min au TBS-Tween, la membrane est incubée pendant 45 min sous agitation à température ambiante et à l'abri de la lumière avec un anticorps secondaire fluorescent dans une solution de TBS-Tween/Lait 1%.

Le signal fluorescent correspondant aux protéines cibles est ensuite révélé en scannant la membrane à l'aide de l'appareil LI-COR Odyssey Scanner. Les images obtenues sont traitées et l'intensité du signal correspondant est mesurée par le logiciel ImageJ.

2.4. Etude du rôle des facteurs d'épissage à domaines UHM dans l'épissage alternatif

2.4.1. Etude de l'effet de la réduction du taux des facteurs d'épissage à domaines UHM sur 65 exons cassette par le système BioMark HD

a) Principe de la méthode

La méthode de BioMark HD développée par Fluidigm est une technique de qPCR à moyen débit permettant de réaliser simultanément un grand nombre de PCR (de 2 304 à 36 960) grâce à la technologie de puces microfluidiques.

Les réactions de PCR sont réalisées dans les puces dédiées appelées Dynamic™ Arrays des circuits IFC pour *Integrated Fluidic Circuit*. Pour ce faire, les échantillons à analyser sont déposés dans les puits correspondants de la plaque microfluidique appelés « sample inlets » et les couples d'amorces correspondants aux cibles à analyser sont déposés dans les puits correspondants appelés « assay inlets » (voir Figure 30).

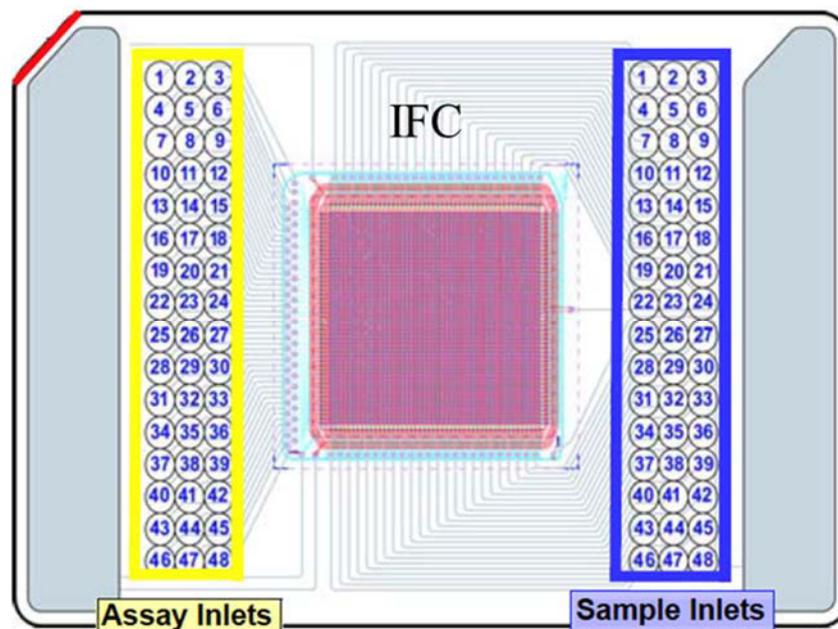


Figure 30. Exemple d'une plaque microfluidique. IFC : Integrated Fluidic Circuit

Les échantillons et les amorces sont ensuite chargés dans le circuit microfluidique permettant d'analyser 48X48 ou 96X96 combinaisons selon la taille de la plaque choisie. Ce chargement s'effectue de manière automatique sur l'appareil IFC controler.

La technique de Biomark HD présente l'avantage d'utiliser une faible quantité d'échantillon de départ pour réaliser un grand nombre d'analyse de cibles. En effet seulement 9 à 6 nl de volume final est nécessaire à une réaction de PCR. Cependant, ce faible volume requiert d'augmenter la concentration des ADNc de départ présents dans chaque échantillon par une pré-amplification en utilisant la totalité des couples d'amorces présents dans l'étude.

b) Protocole expérimental

- Transcription inverse

Avant la transcription inverse, 1 µg d'ARN totaux sont préalablement dénaturés à 65°C pendant 5 min puis refroidis rapidement dans la glace, en présence de 1 µg d'oligonucléotides hexamériques de séquences aléatoires.

La réaction de transcription inverse est réalisée dans un volume de 20 µl contenant 200U de l'enzyme ImProm-II™ reverse transcriptase (Promega), 20U de l'enzyme RNasin (Euromedex), 10 mM d'un mélange de nucléotides ATP, TTP, CTP et GTP dans le tampon spécifique de l'enzyme. Les échantillons sont ensuite incubés pendant 10 min à température ambiante. L'élongation est effectuée à 42°C pendant 50 minutes puis la réaction de transcription inverse est arrêtée par incubation des échantillons pendant 15 min à 70°C.

- Pré-amplification des ADNc

Dans un microtube de 1,5 ml, un mélange de tous les oligonucléotides amorces inclus dans l'étude à un volume de 1µl par oligonucléotide à une concentration de 100 µM est préparé dans un tampon de suspension (10 mM tris, pH8, 0,1 mM EDTA) de façon à obtenir une concentration finale de 500 nM.

La réaction de pré-amplification est réalisée dans un volume de 5 µl contenant 1,25 µl d'ADNc, 1 µl de mix de préamplification PREAMP Mastermix (Fluidigm PN 100-5580), 0,5 µl de mélange d'oligonucléotides amorces dans 2,25 µl d'eau stérile. 10 cycles de PCR sont effectués afin d'amplifier les fragments d'ADNc (15 sec à 95°C, 4 min à 60°C).

La polymérase est ensuite inactivée en ajoutant 2 μl d'enzyme Exonucléase I diluée à 4U/ μl dans son tampon correspondant (New England bioLabs, PN MO0293S) et les échantillons sont incubés 30 min à 37°C puis 15 min à 80°C afin d'inactiver l'enzyme.

Le produit final de 7 μl est ensuite dilué dans 43 μl de tampon de suspension de l'ADN (10 mM tris, pH8, 0,1 mM EDTA) de façon à obtenir un volume final de 50 μl .

- Préparation de la plaque microfluidique

La conformité de la puce microfluidique 48.48 IFC (Fluidigm) est vérifiée en injectant de l'huile dans chacun des accumulateurs. La puce est ensuite mise sous pression dans l'appareil Fluidigm IFC controller.

Un volume total de 5 μl par échantillon contenant 2,25 μl d'ADNc pré-amplifié, 2,5 μl de supermix (BioRad PN 172-5211) et 0,25 μl de DNA binding dye (Fluidigm PN 100-7609) est déposé dans les puits correspondants de la puce microfluidique. Un volume total de 5 μl par couple d'oligonucléotides contenant 0,25 μl d'un mélange d'oligonucléotides Forward et reverse à 100 μM , 2,5 μl de Assay loading reagent (Fluidigm PN 100-7611) et 2,5 μl de tampon de suspension de l'ADN est déposé dans les puits correspondants de la puce microfluidique. Les échantillons et les oligonucléotides sont ensuite chargés dans les capillaires de la puce microfluidique dans l'appareil Fluidigm IFC controller.

- Mesure de l'expression de gènes par qPCR et traitement des données

La PCR quantitative est réalisée sur l'appareil Biomark sur la base de l'acquisition de données de fluorescence en utilisant le logiciel biomark data collection.

Les valeurs de ct sont calculées à partir du logiciel Fluidigm real-time PCR analysis. Les valeurs de ct brutes ont été normalisées en fonction du ct de l'ARN contrôle GAPDH et les analyses statistiques ont ensuite été effectuées sur le logiciel Microsoft Excel.

2.4.1. Etude de l'effet de la réduction du taux de CAPER α et U2AF⁶⁵ sur des minigènes rapporteurs d'épissage par qRT-PCR

a) Principe de la méthode

La qRT PCR est une technique de PCR permettant de quantifier l'expression des ARN dans un échantillon donné. Elle est précédée par une étape de rétrotranscription de l'ARN total par une transcriptase inverse afin d'obtenir une banque d'ADNc. La quantité initiale d'ARN est ensuite mesurée par qPCR à partir de la quantité initiale d'ADNc obtenue après rétrotranscription.

La quantité d'ADNc totale est mesurée à chaque cycle de PCR à l'aide d'un marqueur fluorescent appelé SYBR Green qui se fixe sur l'ADN double brin pendant l'amplification. Ces sondes n'émettent de la fluorescence qu'une fois fixées à l'ADN. Les données d'amplifications sont recueillies par le logiciel de l'appareil qui détermine un cycle seuil, ou *ct* pour *cycle threshold* correspondant à la fin de la phase de bruit de fond où la quantité de produit PCR dépasse la valeur seuil de détection de la technique. C'est donc cette valeur de *ct* qui sera à la base de la mesure de la quantité d'ADNc de départ.

b) Protocole expérimental

Dans une plaque de 384 puits, 10 µl de réaction sont préparés, contenant 4 µl d'un master mix SYBR Green 5 X (GoTaq® qPCR Master Mix, Promega), 0,25 µM d'un mix d'oligonucléotides amorces spécifique de l'ADNc à analyser et 5 µl d'un mélange d'ADNc dilué. La réaction s'effectue dans le thermocycler CFX384® (Biorad). Les valeurs de *ct* sont recueillies par le logiciel Bio-rad CFX Maestro® et l'expression des cibles est calculée à partir du *ct* par Microsoft Office Excel en normalisant sur les valeurs d'expression de la GAPDH. Les analyses statistiques incluant le calcul des moyennes, de l'écart-type et de l'index d'épissage ont été effectuées par Microsoft Office Excel.

2.5. Etude des assemblages macromoléculaires formés par U2AF⁶⁵ et CAPER α avec l'ARN et le domaine multi-ULM de SF3b155

2.5.1. Etude des assemblages macromoléculaires formés par U2AF⁶⁵ et CAPER α avec l'ARN et le domaine multi-ULM de SF3b155 par des expériences de sédimentation

a) Principe de la méthode

Le principe de l'expérience de sédimentation est de détecter la présence différentielle d'assemblages macromoléculaires formés par U2AF⁶⁵ et CAPER α en fonction de la nature des séquences d'ARN. Pour cela, chaque protéine recombinante produite *in vitro* est incubée avec des ARN de séquences différentes. La formation d'assemblages macromoléculaires est ensuite détectée après centrifugation des échantillons en analysant la composition en protéines des culots et des surnageants. En effet, s'il y a formation d'assemblages macromoléculaires, une forte concentration de protéines sera observée dans le culot.

b) Protocole expérimental

Les aliquots de protéines recombinantes produites *in vitro* sont centrifugés à 20 000 xg pendant 10 min à 4°C afin de ne garder que les protéines solubles.

L'expérience de sédimentation est réalisée dans des tubes eppendorf de 1,5 ml préalablement incubés pendant 5 min avec une solution de lait écrémé 1% contenu dans un tampon 20 mM Tris-HCl pH 7,5 ; 100mM NaCl afin d'éviter d'éventuelles interactions non spécifiques des protéines avec la surface du tube. Dans un volume final de 10 µl, les protéines recombinantes sont incubées dans un tampon 20mM Hepes en présence ou non d'ARN pendant 1h à 4°C. Les échantillons sont ensuite centrifugés pendant 10 min à 20 000 xg à 4°C. Les surnageant sont immédiatement transférés dans des nouveaux tubes eppendorf de 1,5 ml et les culots sont resuspendus dans un tampon 20mM Hepes ; 100 mM NaCl ; 0,1% SDS ; 3% Glycérol ; Xylène cyanol. La composition du surnageant est ajustée en ajoutant 0,1% SDS ; 3% Glycérol ; Xylène cyanol.

Les culots et surnageant correspondants à chaque échantillon sont ensuite fractionnés sur gel d'agarose 1% et sur gel de polyacrylamide SDS-PAGE 10%. La présence de protéines dans les échantillons est ensuite révélée par coloration du gel de polyacrylamide au bleu de coomassie. Le signal correspondant aux protéines cibles est ensuite révélé en scannant le gel à l'aide de l'appareil LI-COR Odyssey Scanner à une longueur d'ondes de 685 nm. Les images obtenues sont traitées et l'intensité du signal correspondant est mesurée par le logiciel ImageJ.

IV. Résultats

1. Etude biochimique et en immunofluoromicroscopie des complexes formés par les facteurs d'épissage à domaines UHM avec SF3b155

1.1. Constat initial

La recherche systématique de partenaires protéiques des protéines à domaines UHM ont permis d'identifier cinq motifs ULM de SF3b155 capables d'interagir avec des domaines UHM [Cass and Berglund, 2006; Corsini et al., 2007, 2009; Loerch et al., 2014]. Cet enchainement de motifs ULM refléterait de la présence d'un mécanisme complexe se déroulant durant l'épissage, dans lequel SF3b155 jouerait le rôle de protéine échafaudage afin de former un complexe multiprotéique. Ce complexe multiprotéique interviendrait dans le mécanisme de reconnaissance des sites 3' d'épissage. Ce mécanisme impliquerait d'une part des interactions protéine-protéine de type UHM-ULM et d'autre par des interactions protéine-ARN avec le pré-ARNm par l'intermédiaire de leurs domaines RRM. Des facteurs d'épissage à domaines UHM pourraient alors interagir avec SF3b155 et former des combinaisons d'assemblages macromoléculaires afin de reconnaître les sites 3' d'épissage.

Des analyses *in vitro* par RMN et par Gst-pull down des interactions UHM-ULM entre les domaines ULM de SF3b155 et les domaines UHM de U2AF⁶⁵, CAPER α , PUF60 et SPF45 [Corsini et al., 2007; Loerch et al., 2014; Kielkopf et al., 2001; Corsini et al., 2009] avaient montré que ces quatre protéines avaient la capacité d'interagir avec SF3b155 par l'intermédiaire d'interaction UHM-ULM.

Notre objectif lors de cette première partie a alors été de mettre en évidence la présence de ces interactions dans un contexte cellulaire et de tenter d'identifier la présence de complexes de protéines à domaine UHM en interaction avec SF3b155.

Pour répondre à cet objectif, nous avons dans un premier temps tenté de mettre en évidence la proximité spatiale des UHMSF U2AF⁶⁵, CAPER α , PUF60 et SPF45 dans des cellules humaines par une technique de fluorescence appelée *Proximity Ligation Assay*. Par la même méthode, afin de mettre en évidence la présence de ces protéines durant l'épissage, nous avons également tenté de visualiser la proximité spatiale de ces quatre protéines avec SF3b155.

Dans un deuxième temps, afin d'étudier la formation de complexes entre les protéines UHMSF et SF3b155, nous avons réalisé des expériences de GST-pull down tout d'abord avec la protéine GST-SF3b155 produite *in vitro*, incubée avec un extrait de cellules HEK comme source de UHMSF endogènes, puis avec les facteurs UHMSF produits *in vitro* précédés d'une étiquette GST incubés avec des extraits nucléaires de cellules HeLa comme source de SF3b155 endogène.

Par ailleurs, les données de la littérature ont montré que chaque UHMSF semblait posséder des affinités particulières pour des motifs ULM distincts. Ces préférences de liaison à des motifs ULM particuliers suggèreraient des interactions simultanées des quatre UHMSF avec SF3b155 ainsi que la présence d'un mécanisme de coopérativité dans la formation de ces complexes. C'est ce que nous avons tenté de mettre en évidence par des expériences de GST-pull down en utilisant la protéine GST-SF3b155 produite *in vitro* incubée avec un extrait cellulaire de cellules HEK comme source de UHMSF endogènes et en ajoutant une quantité croissante de U2AF⁶⁵ produite *in vitro*.

1.2. Les facteurs d'épissage à domaines UHM sont colocalisés dans les noyaux de cellules de mammifères

La technique de *Proximity Ligation Assay* permet de détecter la proximité spatiale de deux protéines dans un contexte cellulaire. Pour cela, les cellules doivent être incubées avec un mélange de deux anticorps primaires produits dans deux hôtes différents et spécifiques des deux protéines étudiées. Comme décrit dans le tableau 17, afin d'observer d'éventuelles proximités des protéines à domaines UHM entre elles et avec SF3b155, nous avons utilisé des anticorps dirigés contre U2AF⁶⁵, CAPER α , PUF60 et SF3b155 produits chez la souris et des anticorps dirigés contre CAPER α , PUF60 et SPF45 produits chez le lapin. Afin d'identifier d'éventuels signaux non spécifiques, nous avons utilisé comme contrôles négatifs un anticorps de souris dirigé contre la protéine SR SC35 avec les anticorps de lapin qui reconnaissent les facteurs d'épissage à domaines UHM et un anticorps de lapin qui reconnaît la protéine hnRNP FUS avec les anticorps de souris qui reconnaissent les facteurs d'épissage à domaines UHM. Comme contrôle positif de l'expérience, nous avons analysé la colocalisation entre les protéines U2AF⁶⁵ et SF1.

Tableau 18 : Colocalisations protéiques analysées au cours de l'étude.

		Anticorps produits chez la Souris				
		U2AF65	CAPER α	PUF60	SF3b155	SC35
Anticorps produits chez le lapin	CAPER α					Contrôles négatifs lapin
	PUF60					
	SPF45					
	Fus	Contrôles négatifs souris				
	SF1	Contrôle +				

Afin d'observer l'expression basale de chaque protéine dans les cellules HeLa, nous avons d'abord visualisé les protéines étudiées par immunofluorescence. Comme on peut l'observer dans la figure 31, des signaux fluorescents correspondant à la reconnaissance de chaque protéine par un anticorps spécifique sont principalement localisés dans le noyau.

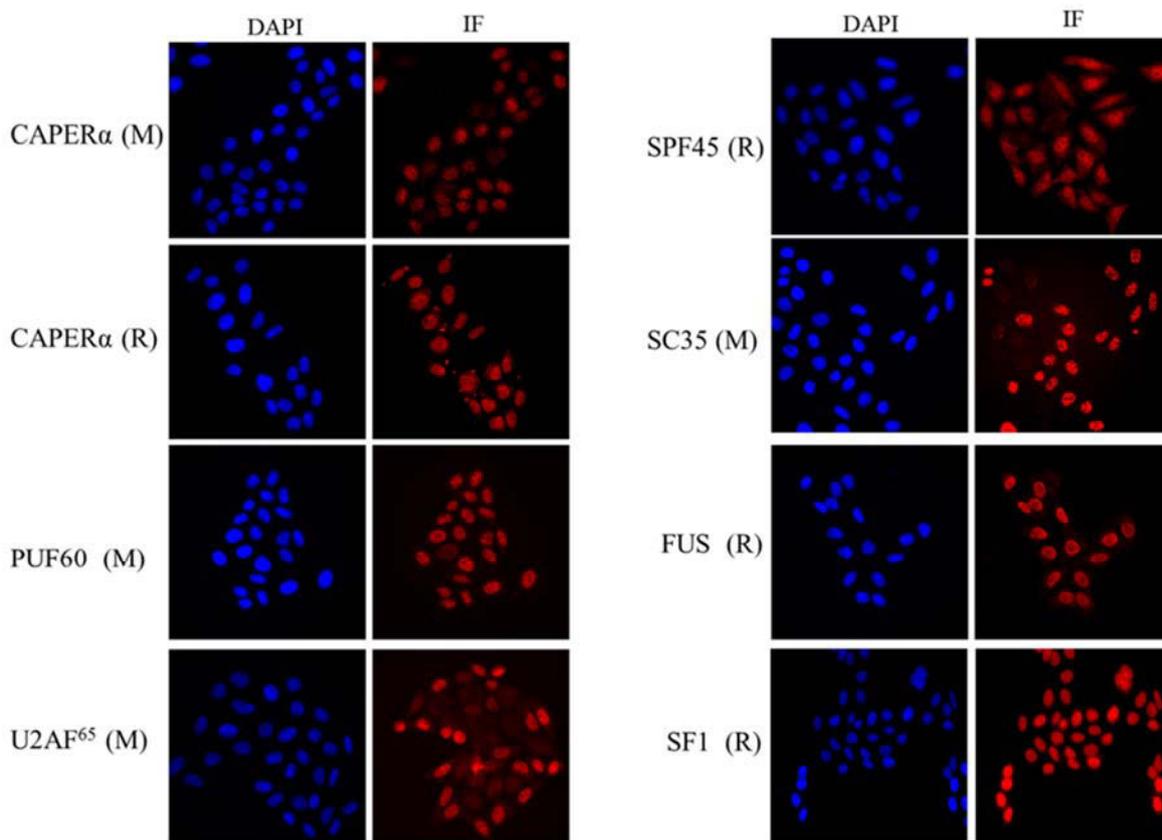


Figure 31. Contrôle de l'expression basale des protéines par immunofluorescence. La première colonne correspond au marquage du noyau au DAPI, la deuxième colonne correspond au marquage de la protéine en immunofluorescence.

Comme on peut l'observer dans la figure 32, un signal fluorescent est observé, montrant la proximité des protéines U2AF⁶⁵ et SF1. Ce signal, correspondant à l'amplification des sondes PLA en conséquence de la proximité des protéines, indique que U2AF⁶⁵ et SF1 sont exprimées dans la cellule en même temps et sont localisées au même moment dans le noyau, comme on peut l'observer dans l'image correspondant à la superposition du marquage du noyau et du signal fluorescent (Merge).

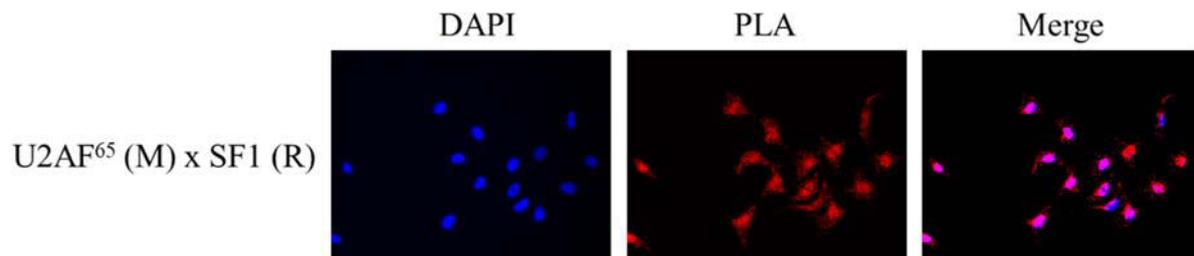


Figure 32. Contrôle positif de l'expérience de Proximity Ligation Assay. La première colonne correspond au marquage du noyau par le DAPI. La deuxième colonne correspond au signal de PLA. La troisième colonne correspond à la superposition des deux premières. La proximité de U2AF⁶⁵ et SF1 est bien retrouvée comme attendu. Des signaux fluorescents sont présents principalement dans le noyau.

Les protéines FUS et SC35 sont des protéines liées au pré-ARNm. La protéine SC35 est une protéine SR qui joue un rôle dans la stabilisation de la snRNP U1 pendant la reconnaissance du site 5' d'épissage. La protéine FUS quant à elle, est une hnRNP qui intervient, comme la protéine SC35, dans la définition des sites d'épissage en 5'. Ces protéines ont été choisies comme contrôles négatifs de l'expérience car elles sont retrouvées dans les complexes d'épissage et interviennent, comme les facteurs d'épissage à domaines UHM, durant les étapes précoces de reconnaissance des sites d'épissage. Les expériences de PLA devraient cependant donner un signal négatif étant donné que les facteurs d'épissage à domaines UHM se localiseraient en 3', alors que FUS et SC35 agissent en 5'.

Comme on peut l'observer dans la figure 33, lors de l'analyse de la proximité de U2AF⁶⁵, CAPER α et PUF60 avec FUS et de CAPER α , PUF60 et SPF45 avec SC35, aucun signal de fluorescence correspondant à l'amplification des sondes PLA n'a été observé. Ce résultat négatif indiquerait que les protéines ne sont pas assez proches dans la cellule (entre 10 et 40 nm) pour induire la réaction d'amplification. Cependant, on ne peut pas exclure que l'absence de signal pourrait être due à la difficulté d'accessibilité des sondes PLA, rendant la réaction d'amplification impossible.

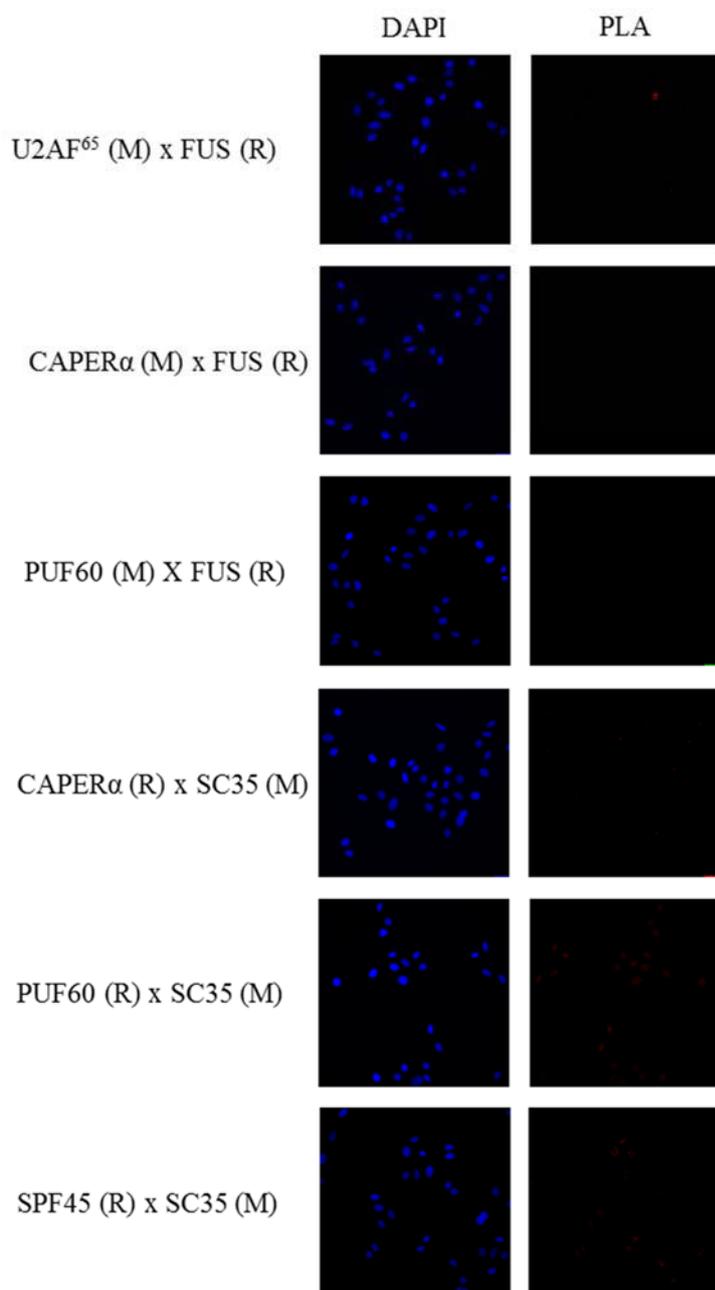


Figure 33. Contrôles négatifs de l'expérience de PLA. La colonne de gauche correspond au marquage du noyau au DAPI. La colonne de droite correspond au résultat de PLA. Afin de détecter d'éventuels signaux de colocalisation non spécifique, une colocalisation avec les deux facteurs d'épissage FUS et SC35 a été testée. Aucun signal de PLA n'a été observé, indiquant qu'il n'y a pas de signal de colocalisation non spécifique.

La proximité des facteurs d'épissage à domaines UHM a ensuite été étudiée. Après avoir réalisé les expériences de PLA, les lames ont été observées au microscope à fluorescence et quatre images de quatre champs de la lamelle ont été photographiées. Après traitement des images par le logiciel imageJ, l'intensité intégrée de chaque signal fluorescent présent dans le noyau a été mesurée à l'aide du logiciel « Cell profiler ». La moyenne de l'intensité de fluorescence intégrée observée dans tous les noyaux ainsi que l'écart-type correspondant ont ensuite été calculés avec

le logiciel Microsoft Excel. La moyenne de l'intensité de fluorescence intégrée de chaque proximité étudiée a ensuite été comparée à la moyenne de fluorescence intégrée calculée dans le contrôle négatif correspondant et la significativité a été déterminée par un test de student unilatéral au risque $\alpha = 0,5\%$ et aux variances égales.

Comme on peut l'observer dans la figure 34, des signaux fluorescents correspondants à l'amplification des sondes PLA résultant de la proximité des protéines sont observés dans les analyses de proximité des couples [U2AF⁶⁵/CAPER α], [U2AF⁶⁵/PUF60] et [U2AF⁶⁵/SPF45]. Ce signal est significatif comparé aux signaux observés dans le contrôle négatif correspondant.

Des signaux significatifs ont également été observés pour l'analyse de colocalisation et de co-expression des couples [CAPER α /PUF60], [CAPER α /SPF45] et [PUF60/SPF45] (Figure 35).

Par cette technique, nous avons détecté une proximité dans le noyau de toutes les combinaisons des facteurs d'épissage à domaines UHM étudiés CAPER α , PUF60, SPF45 et U2AF⁶⁵.

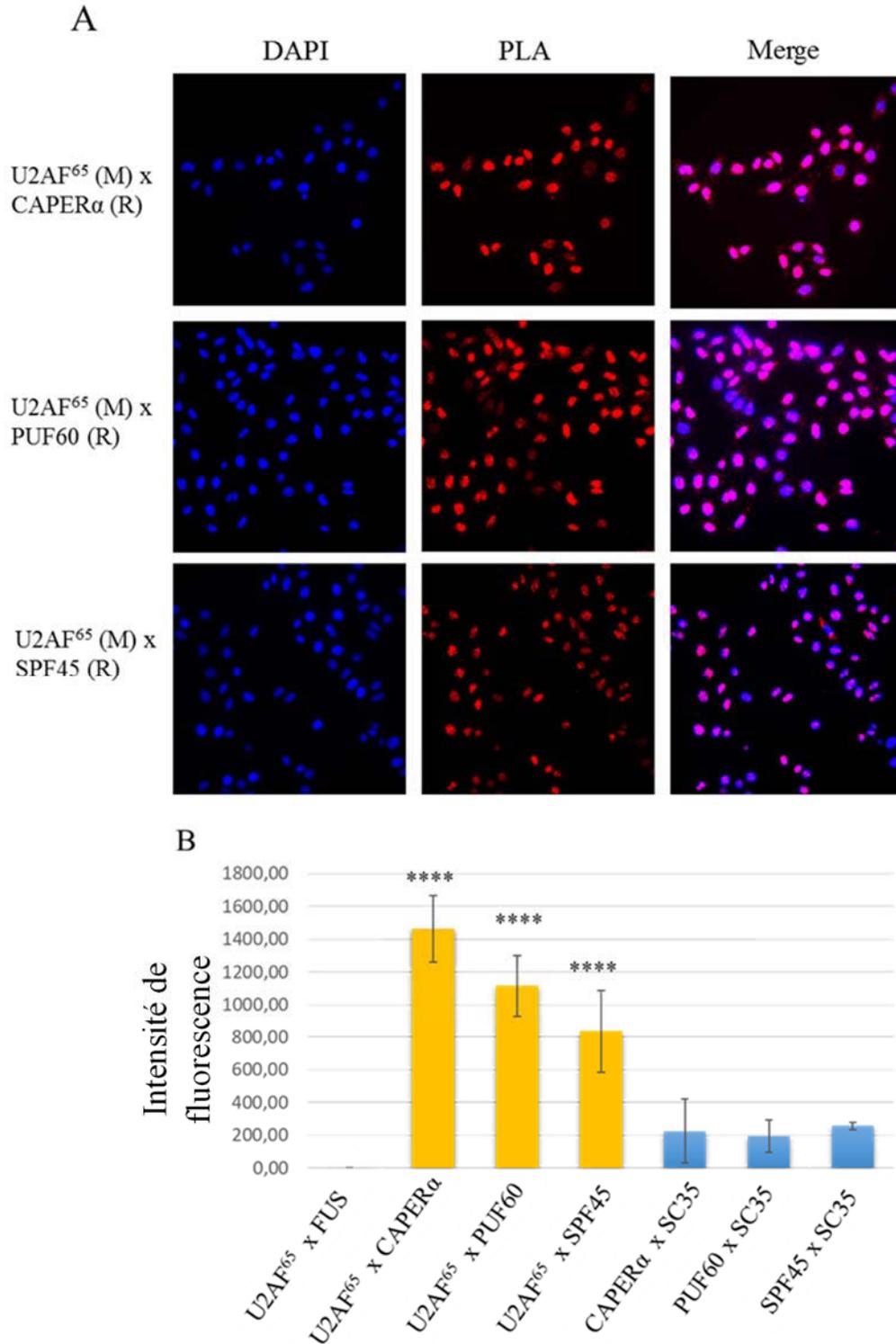


Figure 34. Résultats obtenus lors de l'étude de la proximité de U2AF⁶⁵ avec CAPER α , PUF60 et SPF45. A. Les colonnes sont disposées comme indiqué dans la figure 32. Des signaux de fluorescence ont été observés lors de l'analyse des colocalisations entre U2AF⁶⁵ et CAPER α , U2AF⁶⁵ et PUF60 et U2AF⁶⁵ et SPF45. B. L'intensité de fluorescence dans le noyau de chaque proximité étudiée (jaune) a été mesurée par le logiciel cell profiler et comparée à l'intensité de fluorescence dans le noyau mesurée pour chaque contrôle négatif correspondant (bleu). **** $p < 0,0001$

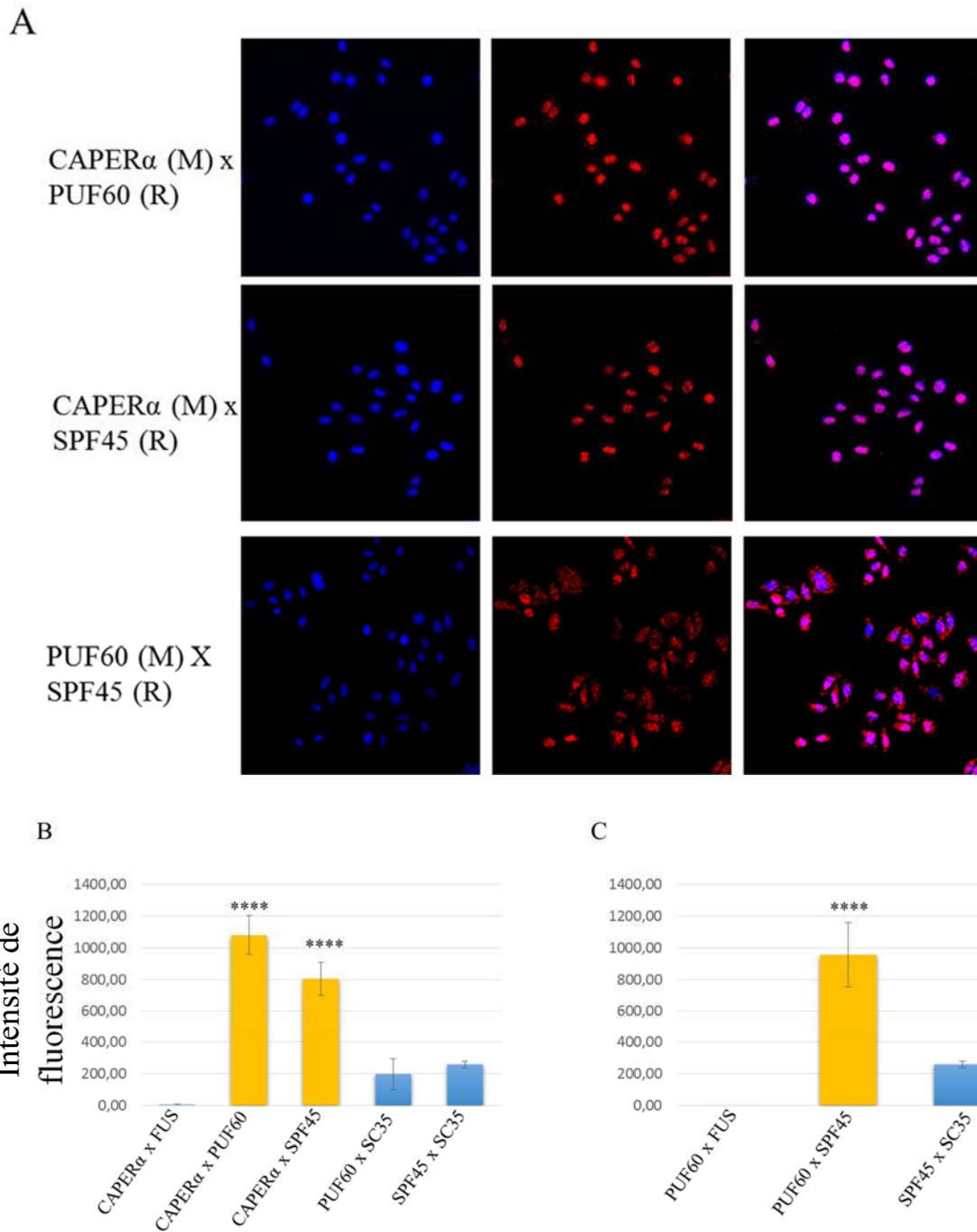


Figure 35. Résultats obtenus lors de l'étude de la proximité de CAPER α avec PUF60 et SPF45 et de PUF60 avec SPF45. A. Les colonnes sont disposées comme indiqué dans la figure 32. Des signaux fluorescents ont été observés lors des analyses de colocalisation entre CAPER α et PUF60, CAPER α et SPF45 et PUF60 et SPF45. B. L'intensité de fluorescence dans le noyau de chaque proximité étudiée (jaune) a été mesurée par le logiciel cell profiler et comparée à l'intensité de fluorescence dans le noyau mesurée pour chaque contrôle négatif correspondant (bleu). **** $p < 0,0001$

Afin de déterminer si cette proximité nucléaire des quatre protéines à domaines UHM se déroulait pendant l'épissage, des analyses de proximité par PLA ont été réalisées entre chaque protéine étudiée et le facteur d'épissage SF3b155 (Figure 36). Comme on peut l'observer, aucun signal fluorescent correspondant à l'amplification des sondes PLA n'a pu être observé dans les conditions expérimentales utilisées, indiquant l'absence de détection de colocalisation de ces protéines par cette méthode.

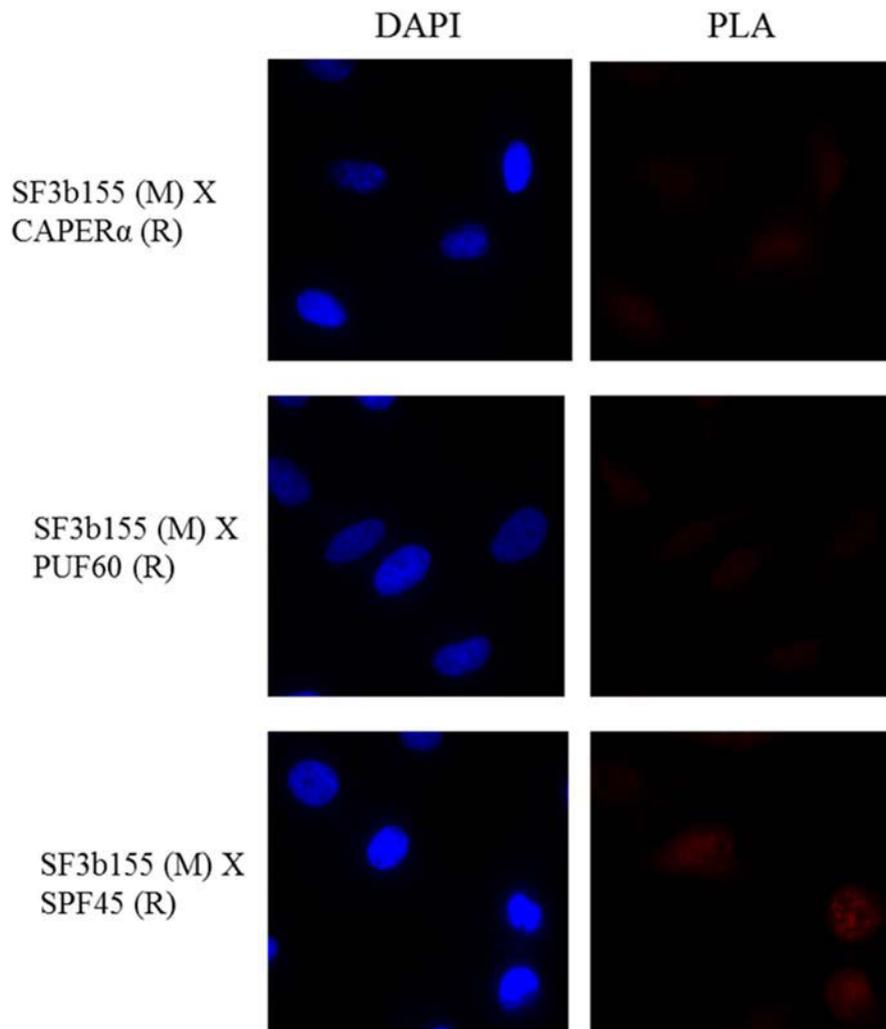


Figure 36. Résultats obtenus lors de l'étude de la proximité de SF3b155 avec CAPER α , PUF60 et SPF45. Les colonnes sont disposées comme indiqué dans la figure 32. Aucun signal de fluorescence n'a été observé lors des analyses de colocalisation entre SF3b155 et CAPER α , SF3b155 et PUF60, et SF3b155 et SPF45.

1.3. Les facteurs d'épissage à domaines UHM interagissent de manière directe avec SF3b155

Après avoir observé la proximité des facteurs d'épissage à domaines UHM U2AF⁶⁵, CAPER α , PUF60 et SPF45 par Proximity Ligation Assay, nous avons tenté de visualiser de potentielles interactions de ces protéines avec des facteurs d'épissage possédant des domaines ULM.

Par des techniques de GST-pull down, nous avons recherché de potentielles interactions entre les facteurs d'épissages SF3b155 et SF1 et les protéines U2AF⁶⁵, CAPER α , PUF60 et SPF45. Pour cela, 25 pmoles de protéines tronquées SF3b155 (GST-SF3b155c) et SF1 (GST-SF1c) produites *in vitro* fusionnées avec une étiquette GST (Figure 37.A) ont été incubées avec une quantité croissante d'extraits totaux de cellules HEK comme source de protéines UHM. Afin de certifier que ces interactions sont indépendantes de l'ARN, les extraits nucléaires ont été traités à la RNase.

Les échantillons ont ensuite été fractionnés sur gel SDS-PAGE et les interactions ont été analysées par western blot (Figure 37) en utilisant les anticorps spécifiques reconnaissant chaque protéine endogène. La membrane a également été incubée avec un anticorps qui détecte l'étiquette GST afin de confirmer que chaque extrait cellulaire a été incubé avec la même quantité de protéines purifiées.

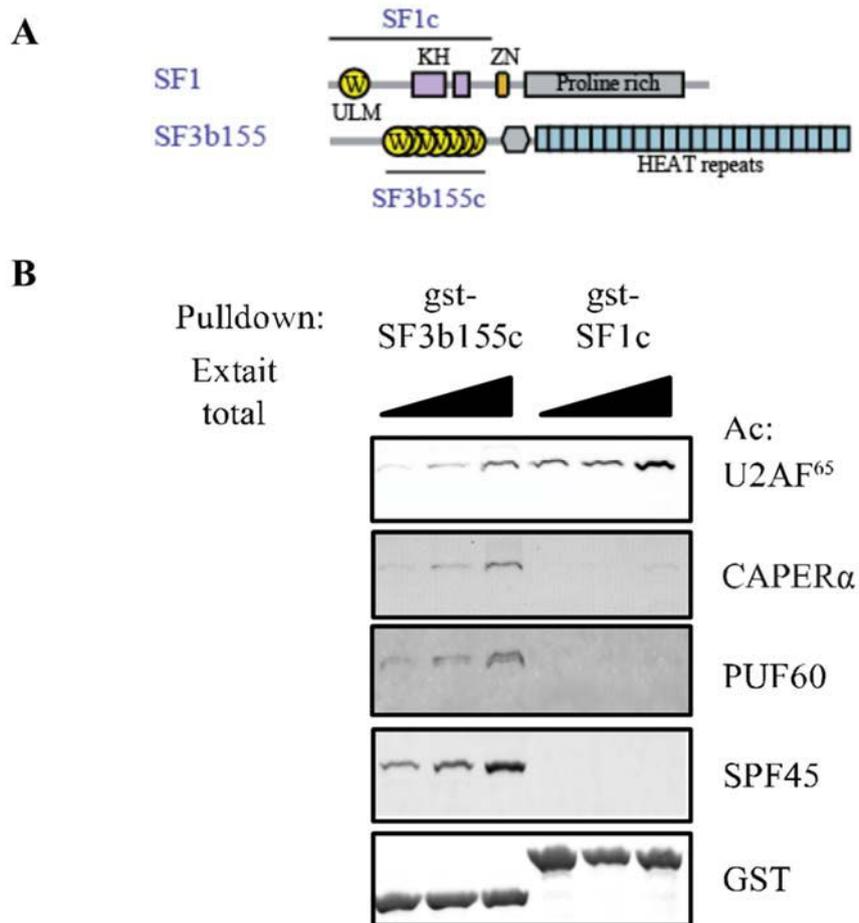


Figure 37. Interaction des protéines UHM avec SF3b155. *A.* Séquence des protéines SF1c et SF3b155c utilisées au cours de l'étude. *B.* Expérience de Gst-pull down : une quantité croissante d'extraits cellulaires totaux de cellules HEK traitées aux RNases a été incubée avec les protéines GST-SF3b155 et GST-SF1. Les échantillons ont ensuite été fractionnés sur gel SDS-PAGE et les interactions ont été analysées par western blot (Expérience réalisée par Valérie Manceau).

Nous observons également une interaction de U2AF⁶⁵ avec SF3b155. Cette interaction est moins forte qu'avec SF1. Par ailleurs, CAPER α , SPF45 et PUF60 semblent interagir de manière préférentielle avec SF3b155. Aucune interaction de ces protéines n'a été observée avec SF1 indiquant une spécificité d'interaction de CAPER α , SPF45 et PUF60 avec les domaines ULM de SF3b155.

Afin de tester si la protéine SF3b155 endogène est également capable d'interagir avec les protéines UHM, des expériences de GST-pull down ont été réalisées en incubant 40 pMol de protéines GST-U2AF⁶⁵, GST-CAPER α , GST-PUF60 et GST-SPF45 produites *in vitro* au laboratoire avec un extrait nucléaire de cellules HeLa comme source de protéines SF3b155 endogènes.

Comme contrôle négatif de l'expérience, les extraits nucléaires ont également été incubés avec une étiquette GST (GST) afin d'observer d'éventuelles interactions non spécifiques de la protéine.

Les protéines issues d'un aliquot des extraits nucléaires de cellules HeLa (10%) ainsi que les échantillons ont été fractionnés sur gel SDS-PAGE et les interactions ont été analysées par western blot (Figure 38) en utilisant les anticorps spécifiques reconnaissant la protéine SF3b155 endogène.

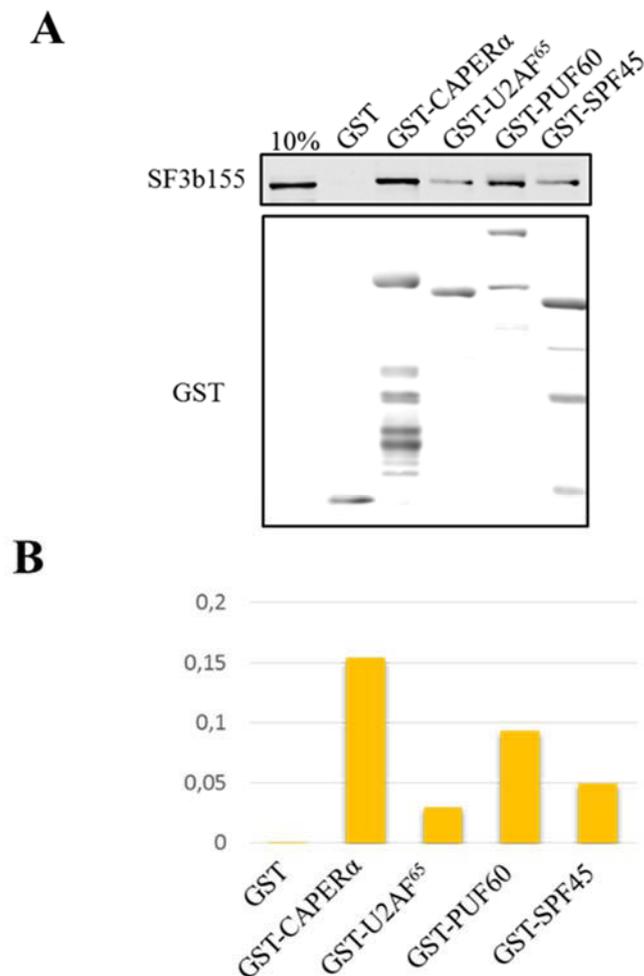


Figure 38. Interaction de SF3b155 avec les protéines UHM. A. Les extraits nucléaires de cellules HeLa traités à la RNase A ont été incubés avec les protéines GST-U2AF⁶⁵, GST-CAPER α , GST-PUF60 et GST-SPF45 purifiée. Pour déterminer d'éventuelles interactions non spécifiques, la même quantité d'extrait nucléaire a été incubée avec la GST. Les protéines des extraits nucléaires de départ (10%) ainsi que les échantillons ont ensuite été fractionnés sur gel SDS-PAGE et les interactions ont été analysées par western blot; B. L'intensité de chaque interaction a été mesurée et l'intensité du signal GST correspondant a été utilisée comme normalisateur.

Comme on peut l'observer dans la figure 38.A, la protéine SF3b155 a été retrouvée dans l'extrait nucléaire (10%). Aucun signal n'a été observé en incubant les extraits nucléaires avec la GST, indiquant qu'il n'y a pas d'interactions non spécifiques de la protéine SF3b155 avec l'étiquette GST. La membrane a également été incubée avec un anticorps qui détecte l'étiquette GST afin de détecter la présence des protéines purifiées.

Nous observons une interaction de SF3b155 avec CAPER α , U2AF⁶⁵, PUF60 et SPF45. L'intensité de chaque interaction a été quantifiée avec le logiciel ImageJ et l'intensité du signal GST de chaque protéine correspondante a été utilisée comme normalisateur. Comme on peut l'observer dans la figure 38.B, SF3b155 semble interagir de manière préférentielle avec CAPER α , tandis qu'un signal d'interaction moins fort a été observé avec PUF60, SPF45 et U2AF⁶⁵.

1.4. Coopérativité de CAPER α et U2AF⁶⁵ pour les interactions avec SF3b155 dépendante des interactions UHM-ULM

Après avoir observé l'interaction des facteurs d'épissage à domaines UHM U2AF⁶⁵, CAPER α , PUF60 et SPF45 avec SF3b155 par GST-pull down, nous avons tenté d'observer par la même méthode si ces interactions se produisaient de manière coopérative. Pour cela, des extraits totaux de cellules HEK ont été utilisés comme source de protéines UHM et ces extraits ont ensuite été incubés avec la protéine SF3b155 produite *in vitro* fusionnée avec une étiquette GST (GST-SF3b155c) en présence d'une gamme de 0 à 40 pMoles de U2AF⁶⁵ produite *in vitro*. Afin d'exclure l'hypothèse d'une interaction de type UHM-ULM entre le domaine UHM de CAPER α , PUF60 et SPF45 et le tryptophane du domaine ULM de U2AF⁶⁵, des expériences similaires ont été réalisées dans les mêmes conditions en présence d'une gamme de U2AF⁶⁵ complexé au domaine UHM de U2AF³⁵.

Les échantillons ont ensuite été fractionnés sur gel SDS-PAGE et les interactions ont été analysées par western blot (Figure 39) en utilisant les anticorps spécifiques reconnaissant chaque protéine endogène.

Comme on peut l'observer dans la figure 39 et comme nous l'avons déjà vu précédemment dans les figures 37 et 38, U2AF⁶⁵, CAPER α , PUF60 et SPF45 interagissent avec SF3b155. Cette interaction est observée dans toutes les conditions étudiées. Plus particulièrement, nous observons une augmentation de l'interaction de CAPER α avec SF3b155 dépendante de la

quantité d'U2AF⁶⁵ ou du dimère U2AF⁶⁵/U2AF³⁵ présente dans l'extrait. Cependant, les interactions de PUF60 et SPF45 avec SF3b155 restent d'intensité similaire quel que soit les conditions.

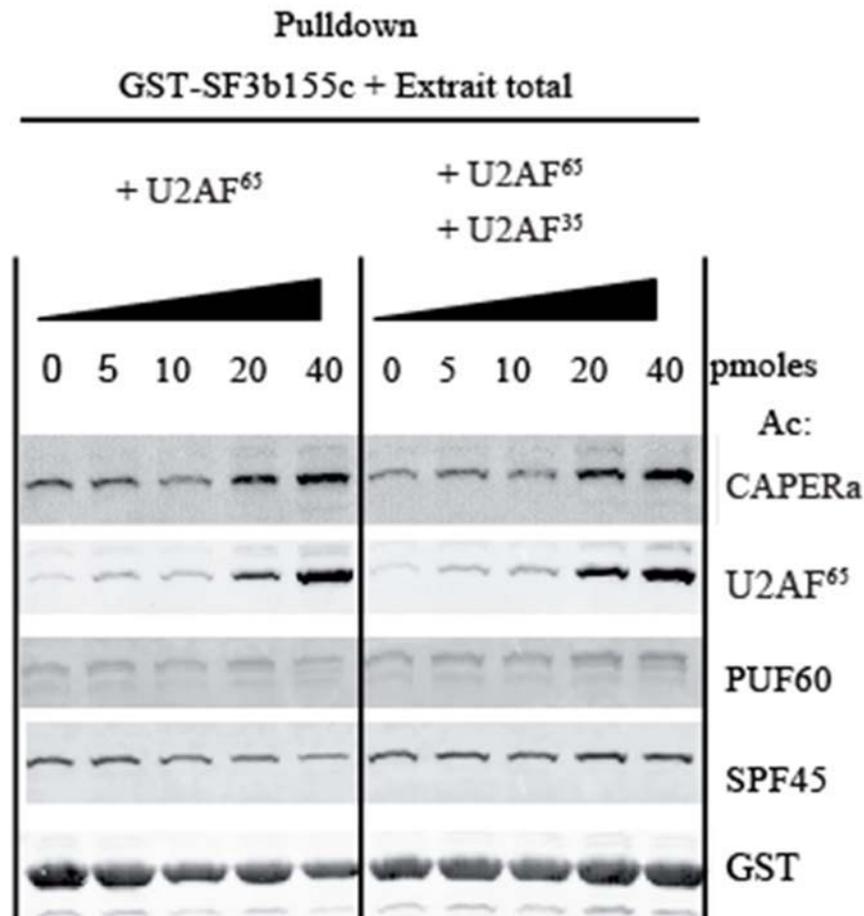


Figure 39. U2AF⁶⁵ renforce le recrutement de CAPERα lors de l'interaction avec SF3b155. Des extraits cellulaires totaux de cellules HEK ont été incubés avec la protéine GST-SF3b155 purifiée en présence d'une gamme de 0 à 40 pMoles de U2AF⁶⁵ produite *in vitro* ou d'une gamme de U2AF⁶⁵ en présence du domaine UHM de U2AF³⁵ les échantillons ont ensuite été fractionnés sur gel SDS-PAGE et les interactions ont été analysées par western blot (expérience réalisée par Valérie Manceau).

Ces résultats indiquent que U2AF⁶⁵ renforcerait le recrutement de CAPERα lors de l'interaction avec SF3b155.

Afin de déterminer si le recrutement de CAPERα par U2AF⁶⁵ était spécifique d'une interaction avec le domaine multi-ULM de SF3b155, des expériences de GST pull down ont été réalisées en incubant des extraits totaux de cellules HEK avec les protéines SF3b155 ou SF1 produites *in vitro* fusionnée avec une étiquette GST (GST-SF3b155c) en présence ou non de U2AF⁶⁵

produite *in vitro*. Les échantillons ont ensuite été fractionnés sur gel SDS-PAGE et l'interaction de CAPER α dans chaque condition étudiée a ensuite été analysée par western blot (Figure 40).

La protéine CAPER α est bien détectée dans l'extrait cellulaire total (1%). On observe des intensités de signaux similaires lors de la détection de l'étiquette GST par un anticorps spécifique, indiquant que la même quantité de protéines purifiées a été ajoutée dans chaque condition. Comme nous l'avons déjà observé dans la figure 39, le recrutement de CAPER α par SF3b155 est fortement renforcé par l'ajout de U2AF⁶⁵. Ceci n'est pas observé dans le cas de l'interaction de CAPER α avec SF1.

L'augmentation du recrutement de CAPER α par U2AF⁶⁵ serait alors dépendante de SF3b155. Ceci pourrait être dû à l'enchaînement de motifs ULM présents dans la protéine.

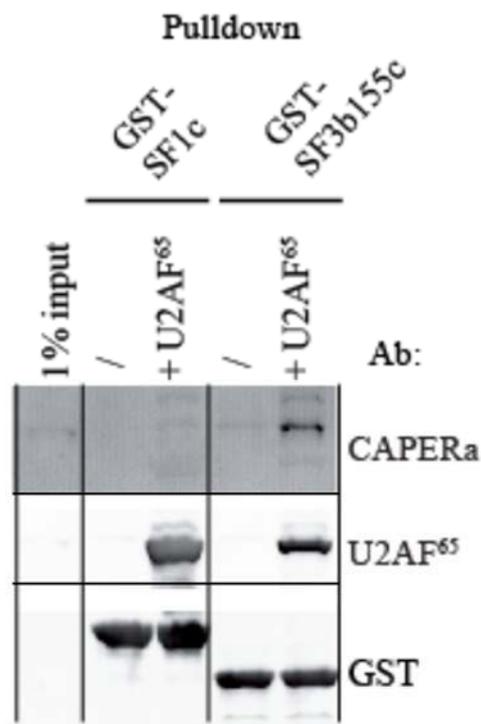


Figure 40. Le renforcement du recrutement de CAPER α par U2AF⁶⁵ est dépendant de SF3b155. Des expériences de GST pull down ont été réalisées en incubant des extraits totaux de cellules HEK avec les protéines SF3b155 ou SF1 produite *in vitro* fusionnée avec une étiquette GST (GST-SF3b155c) en présence ou non de U2AF⁶⁵. Les échantillons ont ensuite été fractionnés sur gel SDS-PAGE et les interactions ont été analysées par western blot (expérience réalisée par Valérie Manceau).

Nous avons ensuite cherché à identifier les domaines de U2AF⁶⁵ impliqués dans le renforcement de l'interaction de CAPER α avec SF3b155. Pour cela, trois différents mutants de

U2AF⁶⁵ ont été obtenus du laboratoire de Clara Kielkopf (Figure 41.A). Le premier mutant « RRM12U » correspondant aux résidus 148 à 475 contient les deux domaines RRM et le domaine UHM de U2AF⁶⁵. Le deuxième mutant « RRM2U » correspondant aux résidus 285 à 475 est composé du deuxième domaine RRM et du domaine UHM de U2AF⁶⁵, tandis que le troisième mutant « UHM » correspondant aux résidus 375 à 475 représente le domaine UHM seul. Ces protéines ont été analysées par SDS-PAGE afin de s'assurer de leur intégrité et de leur concentration (Figure 41.A).

Des expériences de GST-pull down ont ensuite été réalisées en incubant la protéine SF3b155 avec un extrait total de cellules HEK en présence ou non de U2AF⁶⁵ ou des mutants RRM12U, RRM2U ou UHM. Les échantillons ont ensuite été fractionnés sur gel SDS-PAGE et l'interaction de CAPER α dans chaque condition étudiée a ensuite été analysée par western blot (Figure 41.B).

On observe que l'interaction de CAPER α avec SF3b155 n'est augmentée qu'en présence de la protéine U2AF⁶⁵ entière. Ces résultats indiqueraient que les domaines RRM et UHM de U2AF⁶⁵ ne seraient pas suffisants pour renforcer le recrutement de CAPER α et qu'un domaine situé en N-terminal, potentiellement le domaine RS de la protéine serait nécessaire à cette fonction.

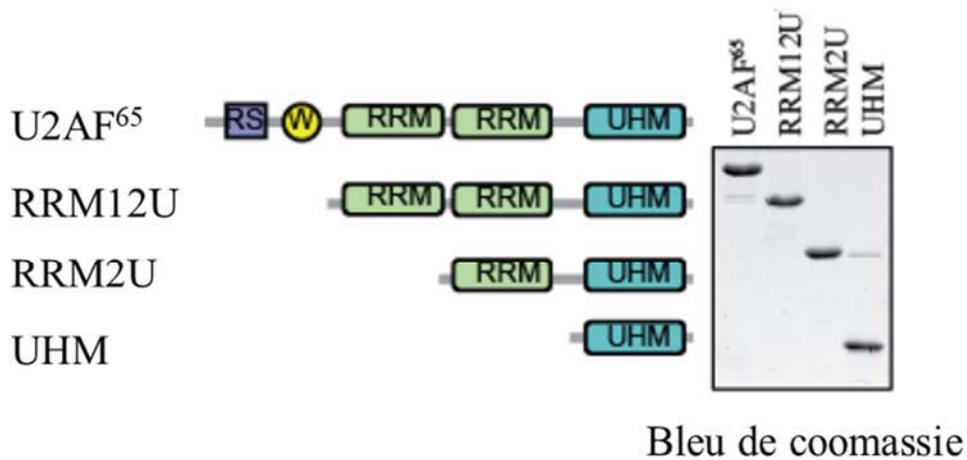
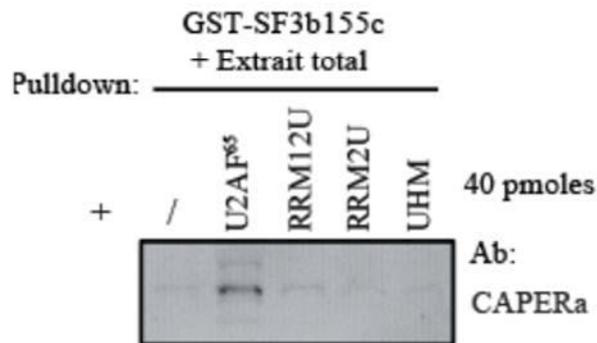
A**B**

Figure 41. L'augmentation du recrutement de CAPER α sur SF3b155 est dépendant d'un domaine en N-terminal de U2AF⁶⁵. A. représentation schématique des mutants de U2AF⁶⁵ produits *in vitro* (gauche) et profil de migration de chacune sur gel SDS-PAGE 10% (droite) ; B. Des expériences de GST pull down ont été réalisées en incubant des extraits totaux de cellules HEK avec la protéines SF3b155 *in vitro* fusionnée avec une étiquette GST (GST-SF3b155c) en présence ou non de U2AF⁶⁵ ou des formes tronquées de la protéine. Les échantillons ont ensuite été fractionnés sur gel SDS-PAGE et les interactions ont été analysées par western blot (expérience réalisée par Valérie Manceau).

Ces expériences suggéraient l'importance du domaine multi-ULM de SF3b155 pour l'interaction avec CAPER α et le renforcement de l'interaction entre CAPER α par U2AF⁶⁵, nous avons ensuite cherché à déterminer l'importance de cet enchainement de motifs ULM pour ces interactions.

Pour cela, des mutants de la protéine SF3b155 avec 1, 2 ou 3 domaines ULM ont été purifiés et produits *in vitro* au laboratoire. Les motifs ULM conservés et les combinaisons de motifs

ULM ont été sélectionnés selon leur affinité d'interaction avec les Protéines U2AF⁶⁵ et CAPER α . Les motifs ULM ayant la plus forte affinité d'interaction avec CAPER α et U2AF⁶⁵ (W200, W293 et W338) ont été conservés.

La figure 42 résume les différents mutants produits au cours de l'étude et utilisés dans ces expériences. Chaque protéine mutante de SF3b155 a été nommée selon la position du tryptophane présent dans la séquence de SF3b155. Par exemple, le mutant SF3b155W200 contient un tryptophane en position 200 de la séquence de SF3b155 tandis que les 6 autres tryptophanes ont été mutés en alanine. De même, le mutant SF3b155W200-293-338 contient trois tryptophanes dans la séquence ULM de SF3b155 en position 200, 293 et 338 tandis que les quatre autres ont été mutés en alanine.

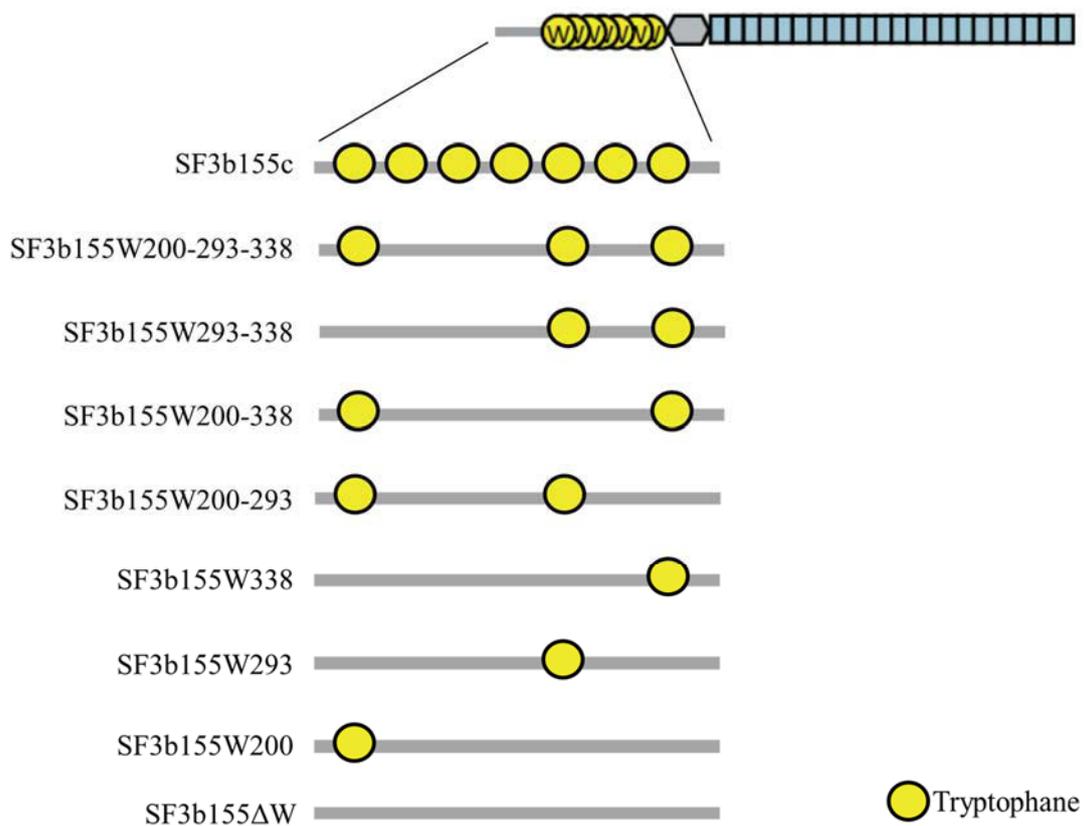


Figure 42. Représentation schématique des différents mutants de SF3b155 utilisés au cours de l'étude. Chaque protéine mutante de SF3b155 a été nommée selon la position du ou des tryptophane(s) présent(s) dans la séquence de SF3b155. W : tryptophane.

Nous avons ensuite testé l'interaction de chaque mutant de SF3b155 avec CAPER α et U2AF⁶⁵ par GST-pull down. Pour cela, des extraits totaux de cellules HEK ont été incubés avec la protéine gst-SF3b155c ou avec des formes de SF3b155c avec 1 (SF3b155W200,

SF3b155W298, SF3b155W338), 2 (SF3B155W200-293, SF3b155W200-338, SF3B155W293-338), 3 (SF3b155W200-293-338), ou aucun tryptophane(SF3b155ΔW)

Les coprecipitats ont ensuite été fractionnés sur gel SDS-PAGE et les interactions ont été analysées par western blot (Figure 43).

L'interaction de CAPER α et U2AF⁶⁵ avec SF3b155 et ses différents mutants a été quantifiée avec le logiciel ImageJ et l'intensité du signal GST correspondant a été utilisé comme normalisateur.

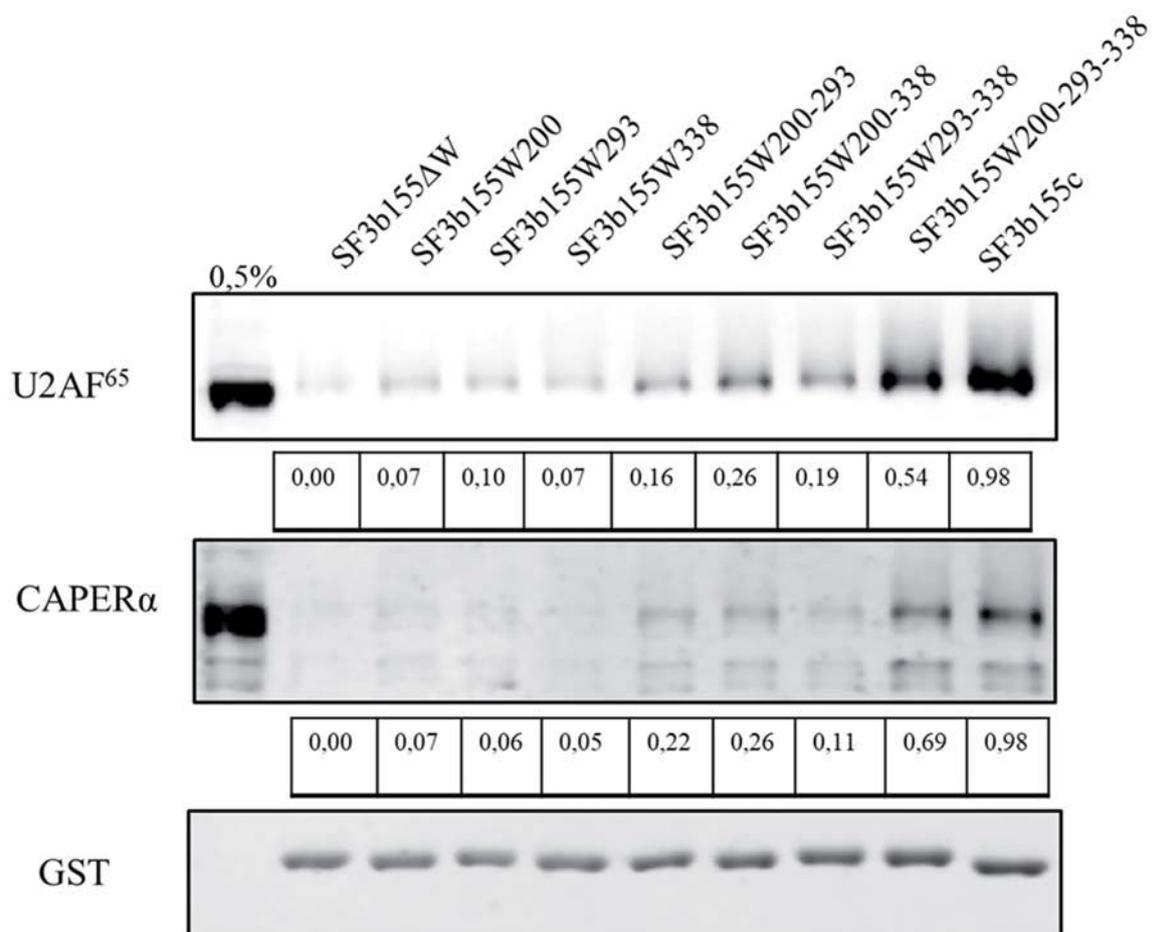


Figure 43. Interactions de CAPER α et U2AF⁶⁵ avec SF3b155 dépendante du nombre de tryptophanes dans les motifs ULM. Des extraits totaux de cellules HEK ont été incubés avec la protéine SF3b155 produite in vitro fusionnée avec une étiquette GST (SF3b155c), avec les différentes protéines SF3b155 possédant 1 (SF3b155W200, SF3b155W298, SF3b155W338), 2 (SF3B155W200-293, SF3b155W200-338, SF3B155W293-338), ou 3 (SF3b155W200-293-338) tryptophanes ou avec la protéine SF3b155 dont tous les tryptophanes ont été mutés en alanine (SF3b155ΔW). L'expérience a été réalisée quatre fois. L'intensité de chaque interaction a été quantifiée avec le logiciel ImageJ et l'intensité du signal GST correspondant a été utilisé comme normalisateur.

Comme on peut l'observer dans la figure 43, l'intensité d'interaction de U2AF⁶⁵ et de CAPER α avec SF3b155 augmente avec le nombre de tryptophanes présents dans la séquence de SF3b155.

La quantification de l'intensité de chaque signal montre que la fraction de U2AF⁶⁵ interagissant avec le mutant SF3b155W200-293-338 est deux fois plus importante que la somme des fractions de U2AF⁶⁵ interagissant avec les mutants de SF3b155 possédant deux tryptophanes et le mutant complémentaire avec un seul tryptophane. Une observation similaire est obtenue pour CAPER α .

Ces observations indiqueraient une coopérativité entre CAPER α et U2AF⁶⁵ pour les interactions avec les domaines ULM de SF3b155

Afin de caractériser la coopérativité de CAPER α et U2AF⁶⁵ dépendante des motifs ULM de SF3b155, nous avons réalisé des expériences de GST-pull down de CAPER α avec les différents mutants de SF3b155 en présence ou non de U2AF⁶⁵ produite *in vitro* et j'ai ensuite observé l'effet de l'ajout de U2AF⁶⁵ sur l'interaction de CAPER α avec SF3b155.

Des extraits totaux de cellules HEK ont été incubés avec la protéine SF3b155 (SF3b155c), avec les mutants SF3b155 possédant 1 (SF3b155W338), 2 (SF3b155W200-338), ou 3 (SF3b155W200-293-338) tryptophanes ou avec la protéine SF3b155 dont tous les tryptophanes présents dans les motifs ULM ont été mutés en alanine (SF3b155 Δ W) (Figure 44).

La fraction de CAPER α interagissant avec SF3b155 et ses différents mutants a été quantifiée avec le logiciel ImageJ et le pourcentage d'interaction de CAPER α avec SF3b155 a été calculé à partir de l'intensité du signal de CAPER α présent dans 0,5% de l'extrait cellulaire total.

Comme on peut l'observer dans la figure 44, l'enrichissement de signal de CAPER α avec le mutant SF3b155 possédant trois tryptophanes augmente fortement en ajoutant U2AF⁶⁵ à l'extrait cellulaire. En effet, le renforcement du recrutement de CAPER α avec cette forme de SF3b155 est presque égal à l'intensité de l'interaction avec la protéine SF3b155 totale. Ces résultats confortent l'hypothèse d'une coopérativité entre CAPER α et U2AF⁶⁵ pour les interactions avec la région multi-ULM de SF3b155.

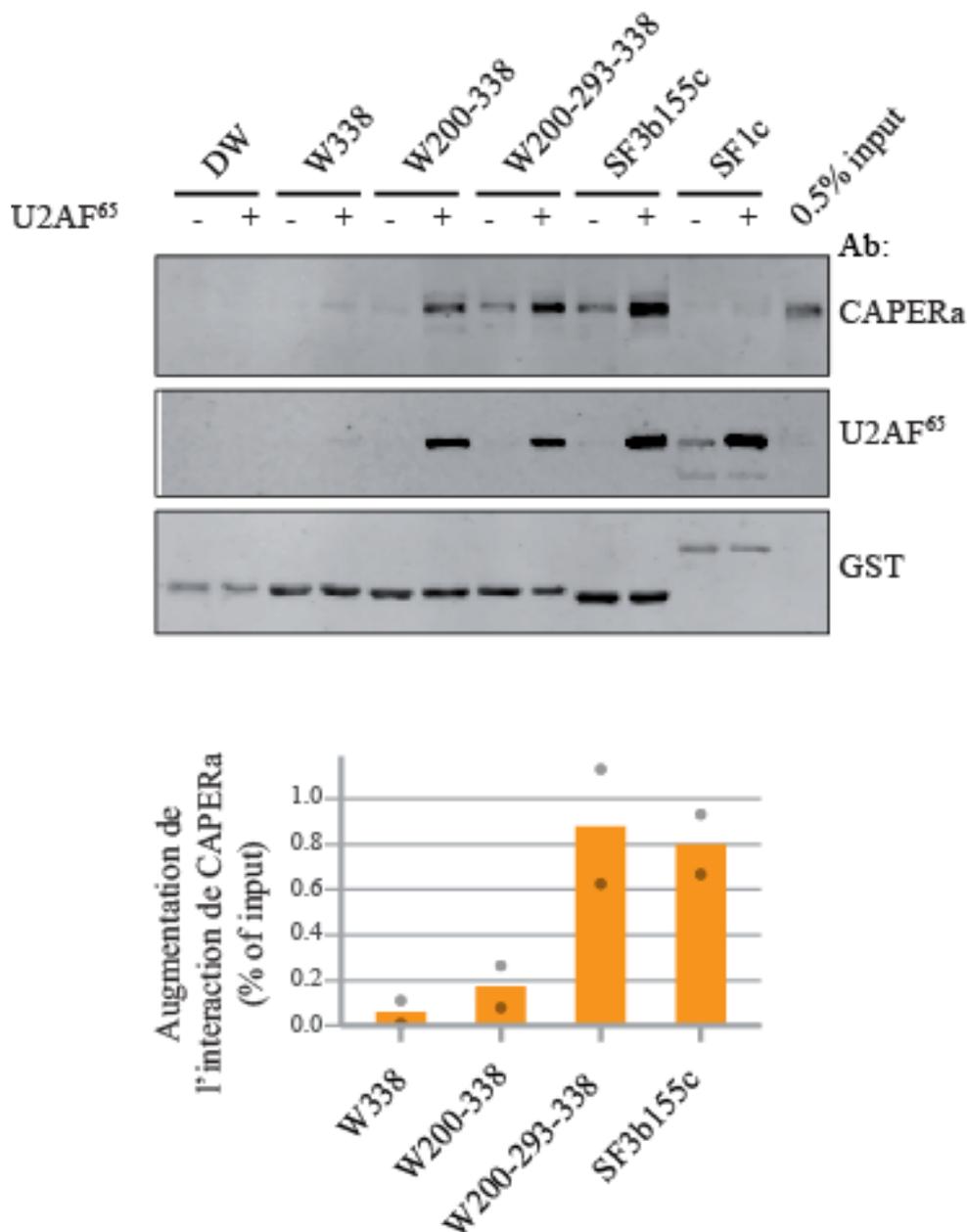


Figure 44. Coopérativité de CAPERa et U2AF⁶⁵ pour les interactions avec SF3b155 dépendante du nombre de tryptophanes dans les motifs ULM. **A.** Des extraits totaux de cellules HEK ont été incubés avec la protéine SF3b155 produite *in vitro* fusionnée avec une étiquette GST (SF3b155c), avec les mutants SF3b155 possédant 1 (SF3b155W338), 2 (SF3b155W200-338), ou 3 (SF3b155W200-293-338) tryptophanes, avec la protéine SF3b155 dont tous les tryptophanes présents dans les motifs ULM ont été mutés en alanine (SF3b155ΔW) ou avec la protéine SF1 en présence ou non de U2AF⁶⁵ ; **B.** L'expérience a été réalisée deux fois. L'intensité d'interaction de CAPERa avec SF3b155 et ses différents mutants été quantifiée avec le logiciel ImageJ et le pourcentage d'interaction de CAPERa avec SF3b155 a été calculé à partir de l'intensité du signal de CAPERa présent dans 0,5% de l'extrait cellulaire total.

2. Etude fonctionnelle du rôle des facteurs d'épissage à domaines UHM dans l'épissage alternatif.

2.1. Constat initial

De nombreuses études ont été menées afin d'analyser les rôles des facteurs d'épissage à domaines UHM U2AF⁶⁵, CAPER α , PUF60 et SPF45 dans la régulation de l'épissage alternatif. Ces études ont permis d'identifier des cibles particulières dont l'épissage alternatif était dépendant d'un de ces quatre facteurs d'épissage. C'est le cas par exemple de l'épissage alternatif du gène c-src, dépendant du facteur U2AF [Sharma et al., 2005], de l'épissage alternatif de l'exon 6 de Fas régulé par SPF45 [Corsini et al., 2007], ou encore de la régulation spécifique de l'épissage alternatif des deux isoformes VEGF165 et VEGF189 par CAPER α [Huang et al., 2012].

D'autres études à plus grande échelle, notamment des analyses à haut débit du transcriptome par RNAseq ont pu mettre en évidence un grand nombre d'évènements d'épissage dépendant des facteurs d'épissage à domaines UHM. C'est le cas notamment de la découverte de 258 évènements d'épissage alternatif régulés par PUF60 et 198 exons régulés par CAPER α dans les cellules HEK [Královicová et al., 2018].

D'autres analyses approfondies des séquences introniques ont permis d'identifier certaines séquences caractéristiques spécifiques des facteurs d'épissage à domaines UHM, comme par exemple les études par CLIP-seq des interactions de CAPER α avec l'ARN, qui ont révélé que la protéine se liait spécifiquement aux sites d'épissage en 3' et en 5' [Mai et al., 2016] ou les analyses des séquences introniques présentes dans 102 évènements d'épissage régulés par PUF60 qui ont mis en évidence une composition particulière des introns, caractérisée par une séquence Polypyrimidine plus longue que la moyenne [Královicová et al., 2018].

Cependant, malgré leur caractère informatif, ces études ont visé à caractériser de manière indépendante l'effet de chaque facteur d'épissage à domaine UHM dans l'épissage alternatif et à identifier des cibles et des séquences cibles spécifiques.

Toutefois, appuyés par les résultats obtenus lors des analyses d'interactions de U2AF⁶⁵, CAPER α , PUF60 et SPF45 avec SF3b155 par GST-pull down, nous avons choisi de réaliser des études comparatives de l'effet des différents facteurs d'épissage à domaines UHM sur la régulation de l'épissage alternatif.

Afin d'étudier et de comprendre le rôle de ces protéines dans la régulation de l'épissage et d'identifier les séquences qu'elles ciblent dans les introns, nous avons comparé l'effet de la réduction du taux de U2AF⁶⁵, CAPER α , PUF60 et SPF45 sur des événements d'épissage alternatif.

Pour cela, le taux des quatre facteurs d'épissage a été réduit dans les cellules Hela par shRNA et l'épissage alternatif de 65 exons cassette préalablement sélectionnés par bio-informatique, a été analysé par qPCR à moyen débit.

Afin de décrypter les mécanismes impliqués dans la régulation de l'épissage alternatif par les facteurs d'épissage à domaines UHM et de caractériser les séquences cibles impliquées dans cette régulation, nous avons mis au point et utilisé un modèle d'étude de l'épissage alternatif par minigènes rapporteurs dans les cellules.

2.2. Sélection des événements d'épissage alternatif étudiés

Afin d'étudier le rôle fonctionnel des facteurs d'épissage à domaines UHM U2AF⁶⁵, CAPER α , PUF60 et SPF45 dans la régulation de l'épissage alternatif, 65 événements d'épissage alternatif ont été sélectionnés par bio-informatique (Figure 45).

En utilisant le serveur Galaxy (<http://galaxy.org>), les événements d'épissage ont été sélectionnés à partir des exomes humain (hg38) et murin (mm10) issus de la collection RefSeq en utilisant le pipeline suivant :

A partir de 610 000 exons humains et 310 000 exons murins, un premier filtre a été appliqué afin d'exclure tous les exons possédant des sites 5' et 3' alternatifs, ainsi que ceux dont les exons en amont et en aval possédaient des sites respectivement 3' et 5' alternatifs. A partir des 170 000 exons humains, et 150 000 exons murins ainsi sélectionnés, les exons cassette ont été identifiés et retenus (5642 exons cassette humains et 2077 exons cassette murins). Ces exons cassette ont ensuite été filtrés sur la base de leur conservation chez l'homme et la souris, afin de ne garder que les événements d'épissage maintenus dans les deux espèces. Enfin, parmi ces 475 exons cassette conservés chez l'homme et la souris, seuls ceux qui s'exprimaient de manière homogène dans les cellules HeLa ont été sélectionnés en effectuant un alignement de données RNAseq publiées sur notre banque de jonction exon-exon.

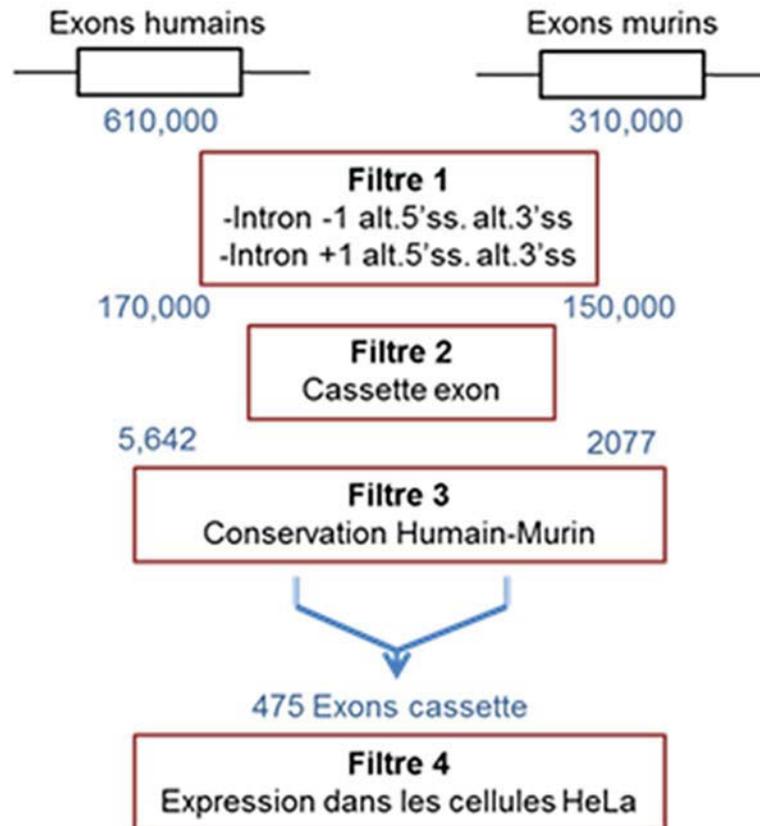


Figure 45. Méthode de sélection des évènements d'épissage alternatif analysés au cours de l'étude. 65 exons cassettes ont été sélectionnés par des méthodes de bio-informatique en fonction de leurs séquences en 5' et 3', de leur conservation entre l'Homme et la souris et de leur expression dans les cellules HeLa.

Pour chaque exon cassette, un score a ensuite été calculé prenant compte de l'enrichissement de la région 3' des introns en séquences polypyrimidine, qui constituent des sites de liaison potentiels pour U2AF⁶⁵. Ce score prend en compte le nombre et la nature des séquences polypyrimidine (SPY, Small Pyrimidine stretch) présentes dans une région de 100 nucléotides en amont du site 3'.

La méthode de détection des séquences polypyrimidine SPY et le calcul du score correspondant est dérivée de la méthode de *Corvelo et al, 2010* [Corvelo et al., 2010] et prend en compte la fréquence et la nature des pyrimidines sur une étendue de 8 ou 9 nucléotides.

Ainsi, les SPY ont été identifiées dans la région 3' intronique en recherchant des séquences de 8 et 9 bases présentant les caractéristiques suivantes :

- Présence d'une pyrimidine en 5' et 3' de la séquence.
- Pas de purines contiguës.
- Chaque purine doit être flanquée d'au moins 4 pyrimidines
- Pas de chevauchement de séquence SPY

Pour chaque SPY, un score a été calculé en fonction de la taille de la séquence et de sa composition selon la formule suivante :

$$score(SPY) = \sum_{x \in \{G,A,T,C\}} f(x).s(x)$$

$f(x)$: fréquence absolue du nucléotide x dans le SPY

$s(T)=3$, $s(C)=2$, $s(A)=-2$, $s(G)=-2$

Ensuite, pour chaque séquence de 100 nucléotides en amont des sites 3', un score global de SPY a été calculé, correspondant à la somme des scores de l'ensemble des SPY détectés dans la séquence.

La différence entre le score SPY global de l'intron en 5' de l'exon cassette et le score SPY global de l'intron en aval a également été calculée afin de prendre en compte la compétition entre les sites d'épissage (ΔSPY_score).

Sur la base du score ΔSPY des exons cassette exprimés dans les cellules HeLa, 65 exons cassette ont été sélectionnés, dont 32 possédaient un score < 0 et présentaient donc une région relativement pauvre en répétitions de polypyrimidine, et 33 possédaient un score > 0 et présentaient une région relativement riche en répétitions de polypyrimidine (Figure 46).

Des sondes d'oligonucléotides spécifiques ont ensuite été dessinées de manière à pouvoir quantifier par qPCR les deux formes d'ARN pour les 65 événements d'épissage alternatif.

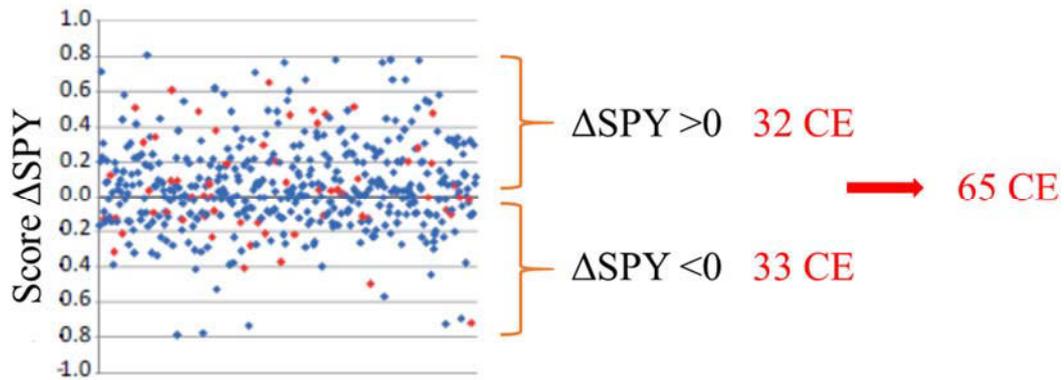


Figure 46. Sélection de 65 exons cassette sur la base de leur score ΔSPY : 32 possèdent un score < 0 et présentent une région pauvre en répétitions de polypyrimidines, et 33 possèdent un score > 0 et présentent une région riche en répétitions de polypyrimidines

2.3. Etude de l'influence d'un déficit en facteurs d'épissages à domaines UHM sur l'épissage de 65 exons cassette

Le taux des protéines U2AF⁶⁵, CAPER α , PUF60 et SPF45 a été diminué dans des cellules HeLa par transduction virale de shRNA. Pour chacune des protéines, deux shRNA ont été utilisés et l'expérience a été réalisée trois fois. Des cellules ont également été infectées avec des virus exprimant des shRNA n'ayant aucune cible dans les cellules HeLa, comme contrôle (NS). Après sélection des cellules sur la base de leur résistance à la puromycine elles sont maintenues en culture pendant trois jours (Figure 47.A). Des extraits totaux de protéines et d'ARN sont alors préparés. Afin de vérifier l'efficacité de la réduction du taux de U2AF⁶⁵, CAPER α , PUF60 et SPF45, des quantités égales d'extraits protéiques totaux ont été fractionnées par électrophorèse et les protéines indiquées ont été révélées par western blot (Figure 47.B). La quantité de chaque protéine présente dans l'extrait a été quantifiée en utilisant le logiciel ImageJ et l'intensité du signal de la tubuline a été utilisée comme normalisateur.

La même quantité de Tubuline est détectée dans tous les échantillons fractionnés, ce qui confirme que la même quantité de protéines totales a bien été déposée sur gel. Une diminution des protéines est observée dans les cellules infectées par les shRNA correspondant par comparaison au shRNA contrôle (NS). Les taux des autres protéines testées restent stables, sauf dans le cas des échantillons où le taux de CAPER α a été diminué, qui présentent une augmentation de l'expression de U2AF⁶⁵.

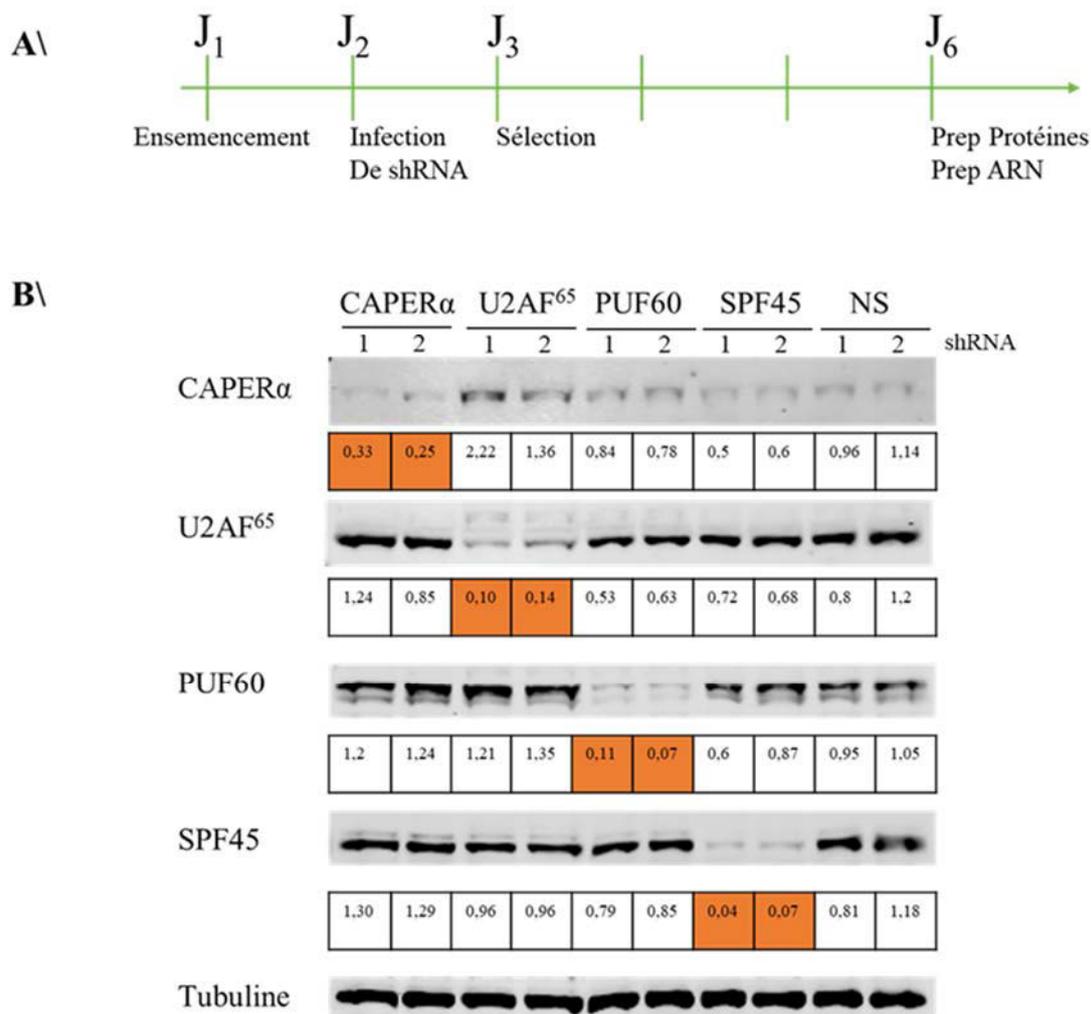


Figure 47. Niveau d'expression des protéines U2AF⁶⁵, CAPER α , PUF60 et SPF45 par transduction virale de shRNA. A. Les cellules HeLa ont été infectés avec des virus contenant des shRNA spécifiques des protéines étudiées. Après sélection, les cellules sont maintenues en culture pendant 72 heures et des extraits totaux de protéines et d'ARN sont préparés ; B. Des quantités identiques de protéines totales contenues dans les extraits ont été fractionnées sur gel SDS-PAGE et les protéines indiquées ont été révélées par western blot. La quantité de chaque protéine présente dans l'extrait a été quantifiée en utilisant le logiciel ImageJ et l'intensité du signal de la tubuline a été utilisée comme normalisateur.

Après avoir confirmé que le taux des protéines U2AF⁶⁵, CAPER α , PUF60 et SPF45 a bien été réduit, des extraits d'ARN totaux sont préparés et une transcription inverse est réalisée. Des expériences de qPCR ont été effectuées pour déterminer le taux des ARNm U2AF⁶⁵, CAPER α , PUF60 et SPF45. Les taux relatifs de chacun des ARNm dans les cellules HeLa traitées avec des shRNA spécifiques des protéines étudiées ont été comparés aux taux relatifs de chacun des ARNm dans les cellules HeLa traitées avec des shRNA contrôles (NS). L'ARNm codant la protéine GAPDH a été utilisé comme normalisateur.

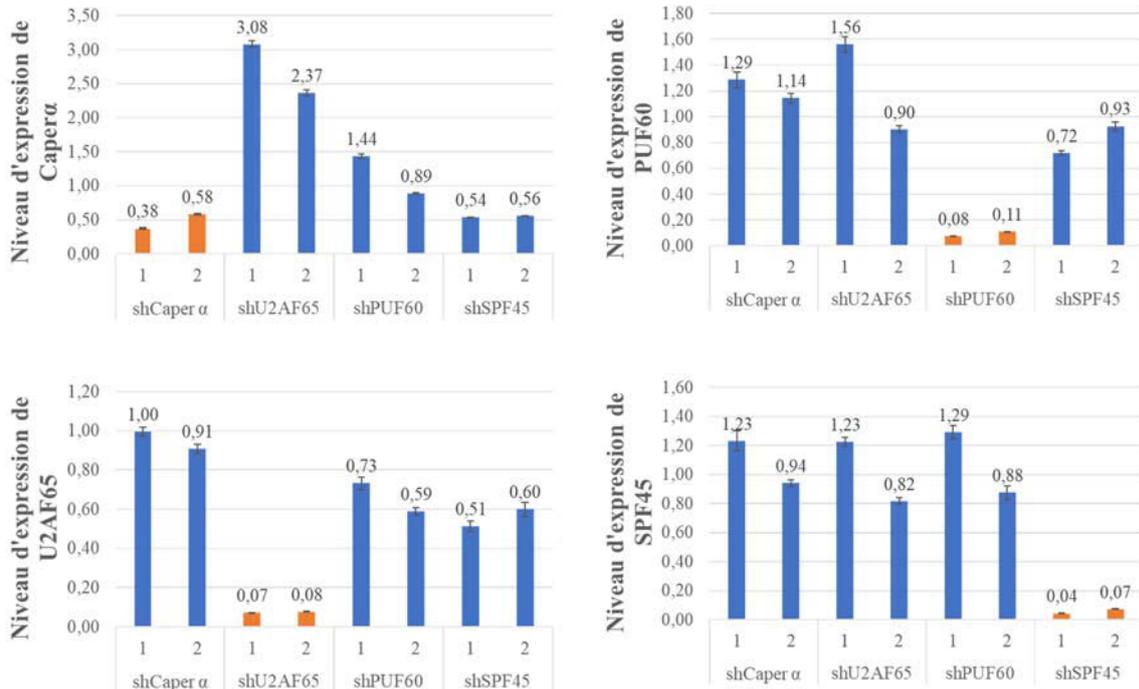


Figure 48. Niveaux d'expression des ARNm CAPERα, U2AF⁶⁵, PUF60 et SPF45 par qPCR. Les taux relatifs de chacun des ARNm dans les cellules HeLa traitées avec des shRNA spécifiques des protéines étudiées (en orange) ont été comparés aux taux relatifs de chacun des ARNm dans les cellules HeLa traitées avec des shRNA contrôles. L'ARNm codant la protéine GAPDH a été utilisé comme normalisateur.

L'écart-type, représenté par les barres d'erreur, correspond au triplicat technique de chaque mesure.

Une diminution du niveau d'expression des ARN est observée dans les cellules infectées par les shRNA correspondant (Figure 48). Comme il avait été précédemment observé pour les protéines, les taux des autres ARN testés restent stables, sauf dans le cas des échantillons où le taux de CAPERα a été réduit, qui présentent une augmentation de l'expression de U2AF⁶⁵. En accord avec les résultats publiés par Vorechovski et collaborateurs [Královicová et al., 2018].

Après avoir confirmé par western blot et qPCR que l'expression des quatre facteurs d'épissage à domaines UHM a bien été diminué par les deux shRNA spécifiques, les échantillons d'ADNc ont été préamplifiés avec un mélange de tous les oligonucléotides amorces inclus dans l'étude, c'est-à-dire les oligonucléotides amorces nécessaires à l'amplification des 65 exons cassette préalablement sélectionnés, ainsi que les oligonucléotides amorce nécessaires à l'amplification de la GAPDH, utilisée comme normalisateur.

Des expériences de qPCR à haut débit par puces microfluidiques Biomark ont ensuite été réalisées sur la plateforme CRT de l'Institut Pasteur. Les niveaux relatifs de chaque cible

normalisée par rapport à celui de la GAPDH ont été déterminées par le logiciel « Fluidigm real-time PCR analysis » par la méthode Delta Ct.

Pour chacun des exons cassette et pour chaque condition, un index d'épissage (SI, pour splicing index) a été calculé, en fonction des niveaux d'expression de la forme avec exon (Sin) Et de la forme sans exon (Sout).

$$SI = \frac{Sin}{(Sin + Sout)}$$

Le changement d'épissage (SC, pour Splicing Change) a ensuite été calculé à partir de l'index d'épissage de chaque exon cassette et pour chaque condition. Ce changement d'épissage représente le ratio de la différence de l'index d'épissage observée sur la différence maximale théorique.

$$SC = \frac{(SI - SIc)}{(1 - SIc)} \text{ si } SI \geq SIc$$

$$SC = \frac{(SI - SIc)}{SIc} \text{ si } SI \leq SIc$$

SC : Changement d'épissage

SI : Index d'épissage

SIc : Index d'épissage dans l'échantillon contrôle

Un test de student est ensuite appliqué aux données afin de déterminer les valeurs significatives de changements d'épissage dans les conditions étudiées.

Les valeurs des moyennes des index d'épissage SI ou changement d'épissage SC pour chaque condition sont représentées sous forme de graphiques. La figure 49 représente trois exemples d'index d'épissage et de changements d'épissage induits par la diminution du taux de U2AF⁶⁵, CAPER α , PUF60 ou SPF45.

Dans chaque cas, l'épissage est augmenté (orange), diminué (vert) ou inchangé (gris). Par exemple, comme on peut l'observer dans la figure 49, l'index d'épissage de L'ARN DNM2 est augmenté par l'inhibition de CAPER α , U2AF⁶⁵ et SPF45, c'est-à-dire que l'inclusion de l'exon

cassette de DNM2 est augmentée lors de l'inhibition de CAPER α , U2AF⁶⁵ et SPF45. Par ailleurs, l'index d'épissage de l'ARN TJP1 est diminué par l'inhibition de CAPER α et U2AF⁶⁵. On peut observer également que l'épissage de certains ARN reste inchangé. C'est le cas par exemple de L'ARN PKP4.

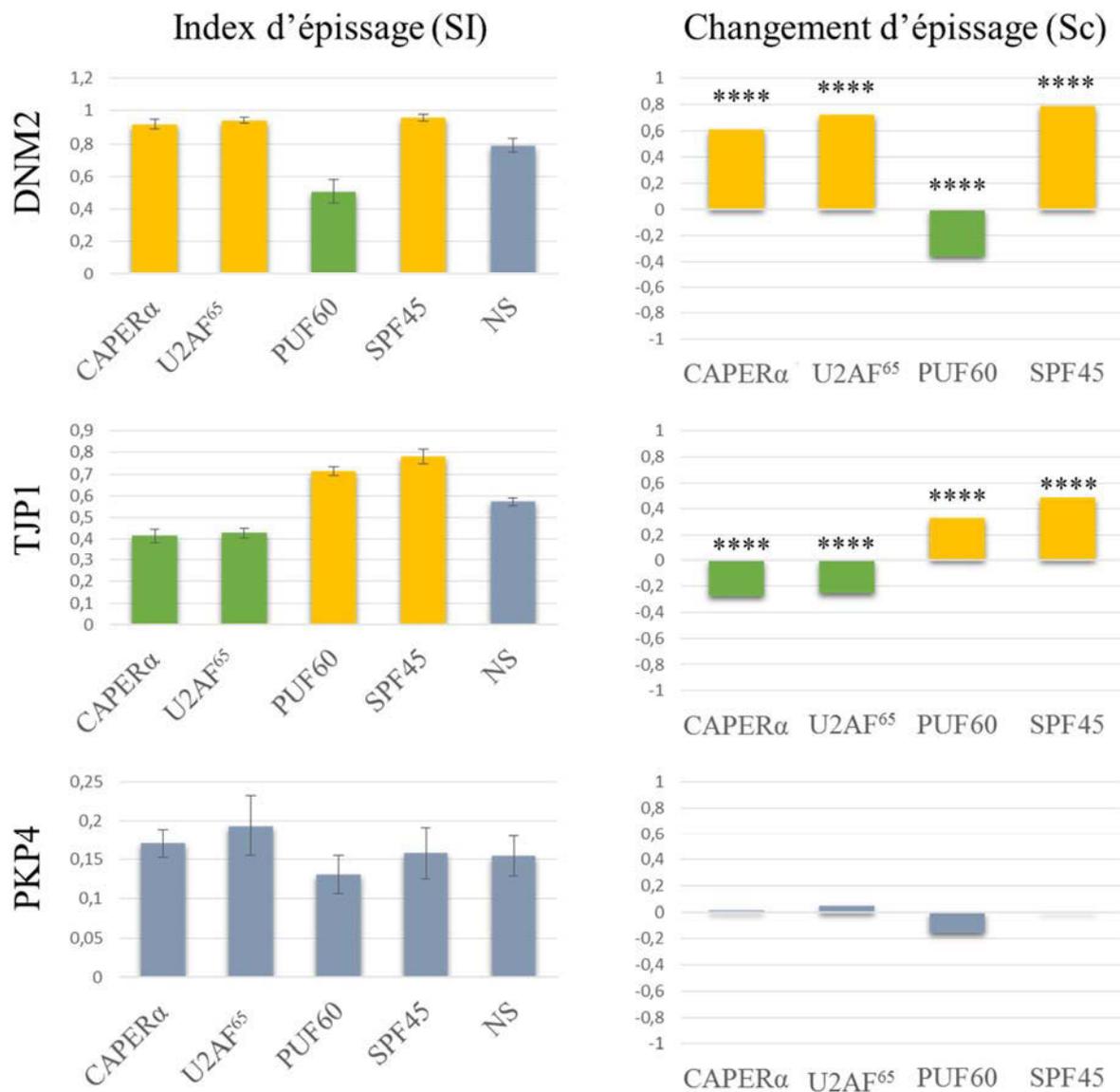


Figure 49. Représentation graphique de trois exemples d'index d'épissage et de changements d'épissage induits par la diminution du taux de U2AF⁶⁵, CAPER α , PUF60 ou SPF45. **** $p < 0.0005$ **; $p < 0.01$ NS: contrôle; orange: augmenté; vert: diminué; gris: inchangé.

L'épissage des cibles étudiées peut parfois être affecté de différente manière selon les conditions. Par exemple, l'index de L'ARN TJP1 est diminué par l'inhibition de CAPER α et U2AF⁶⁵, tandis qu'il est augmenté par l'inhibition de PUF60 et SPF45. On peut également observer le cas inverse dans le cas de l'ARN de DNM2 dont l'index d'épissage est augmenté par l'inhibition de CAPER α , U2AF⁶⁵ et SPF45 mais qui est diminué par l'inhibition de PUF60.

Lorsque l'on considère l'effet global de l'inhibition de chaque facteur d'épissage à domaines UHM sur les événements étudiés (Figure 50), on observe tout d'abord que les quatre protéines pourraient jouer un rôle activateur ou inhibiteur de l'épissage, en fonction de la cible. Cependant, l'inhibition de CAPER α , U2AF⁶⁵ et PUF60 semble en majorité diminuer l'index d'épissage, tandis que l'inhibition de SPF45 induirait une augmentation de l'index d'épissage. CAPER α , U2AF⁶⁵ et PUF60 auraient donc généralement un rôle activateur de l'épissage, contrairement à SPF45 qui serait plutôt inhibiteur.

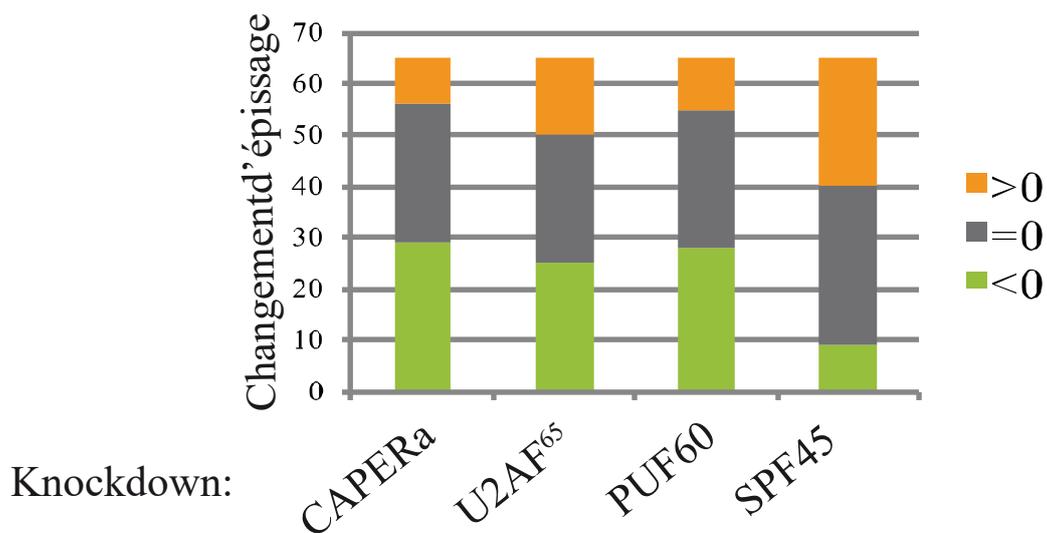


Figure 50. Représentation graphique des changements d'épissage induits par l'inhibition de CAPER α , U2AF⁶⁵, PUF60 et SPF45

On observe également un grand nombre de cibles dont l'index d'épissage est inchangé par l'inhibition d'au moins une protéine, indiquant peut-être une spécificité de mécanisme des protéines étudiées dépendante de la nature des cibles à épisser (Figure 51).

Nous avons ensuite tenté d'observer s'il existait des cibles communes aux quatre protéines étudiées. La figure 51 représente les changements d'épissage communs entre CAPER α , U2AF⁶⁵, PUF60 et SPF45. A gauche sont représentés les changements d'épissage diminués lors de l'inhibition des protéines, et à droite les changements d'épissage augmentés par l'inhibition des protéines. Aucune cible commune aux quatre facteurs d'épissage ne semble être observée.

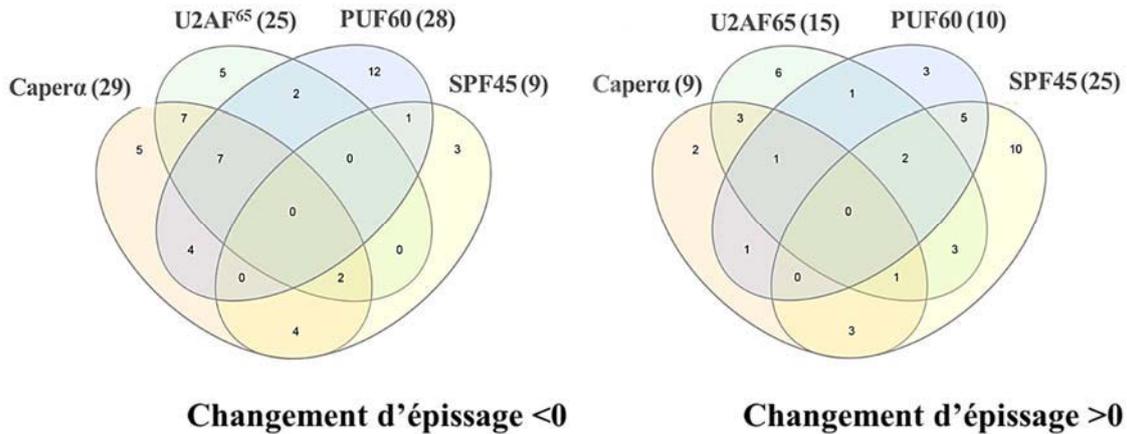


Figure 51. Diagramme de Venne représentant les changements d'épissage diminués (gauche) et augmentés (droite) communs aux quatre facteurs d'épissage à domaines UHM.

Cependant, lorsque l'on étudie en détail les changements d'épissage commun entre deux protéines, on observe un nombre élevé significatif de cibles communes dont l'épissage est diminué par l'inhibition de CAPER α et U2AF⁶⁵ (Figure 52).

Changements d'épissage <0				Changements d'épissage >0				
	U2AF ⁶⁵ (25)	PUF60 (28)	SPF45 (9)		U2AF ⁶⁵ (15)	PUF60 (10)	SPF45 (25)	
CAPER α (29)	16	**	12	6	CAPER α (9)	4	2	4
U2AF ⁶⁵ (25)		10	2		U2AF ⁶⁵ (15)		4	6
PUF60 (28)			2		PUF60 (10)			7

Figure 52. Tableau représentant les changements d'épissage communs à deux facteurs d'épissage à domaines UHM. **: $p < 0.01$

Nous avons ensuite recherché des corrélations pour l'ensemble des cibles étudiées. La figure 53 est une représentation graphique en nuage de points des valeurs de changements d'épissage pour deux facteurs d'épissage à domaines UHM. Après avoir mesuré la dépendance statistique en appliquant le test de Spearman, on observe une corrélation positive entre CAPER α et U2AF⁶⁵ avec un coefficient de corrélation de Spearman $\rho = 0.26$ et une probabilité $p = 0.03$ pour les changements d'épissage des cibles étudiées. Aucune autre corrélation significative n'a été observée pour les autres protéines.

Ces résultats, et les résultats précédents, révèlent un lien fonctionnel entre CAPER α et U2AF⁶⁵ dépendant du choix des cibles d'épissage.

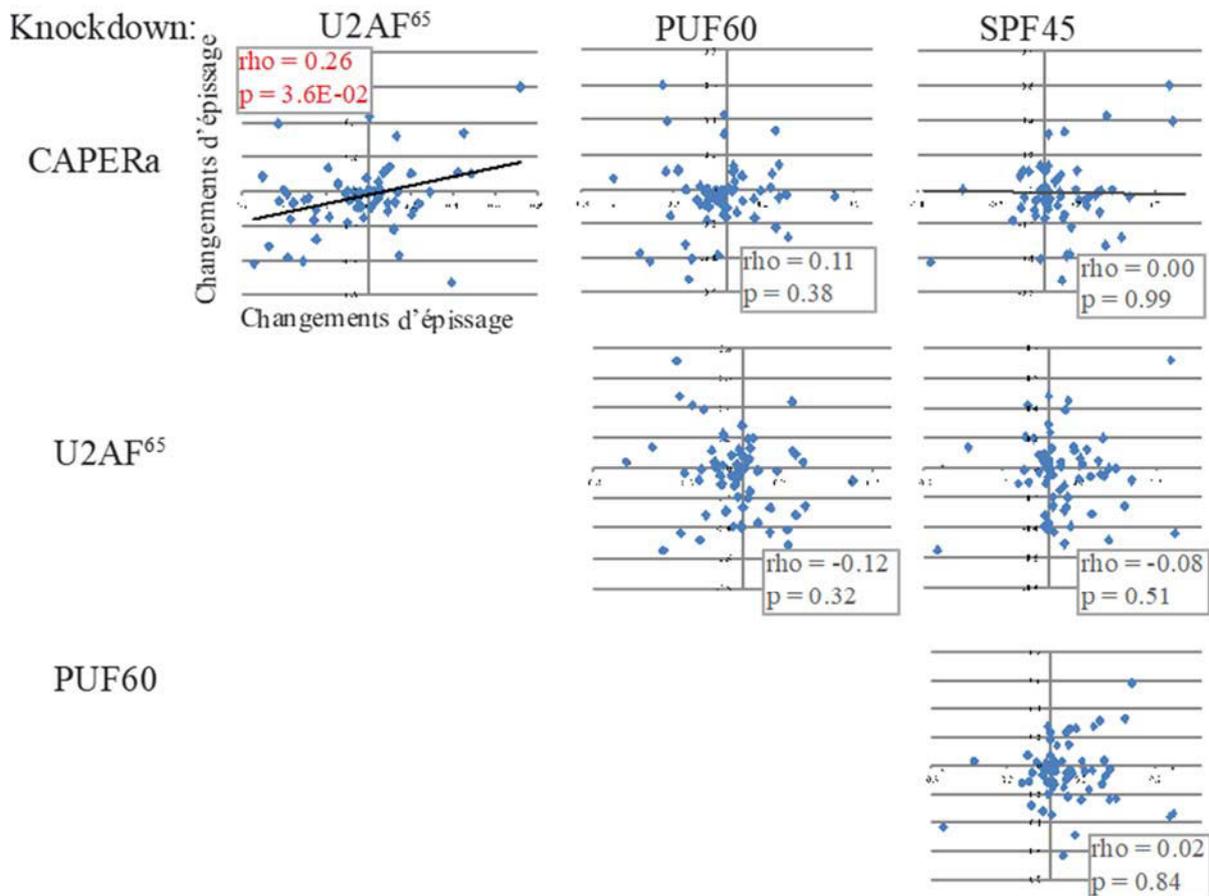


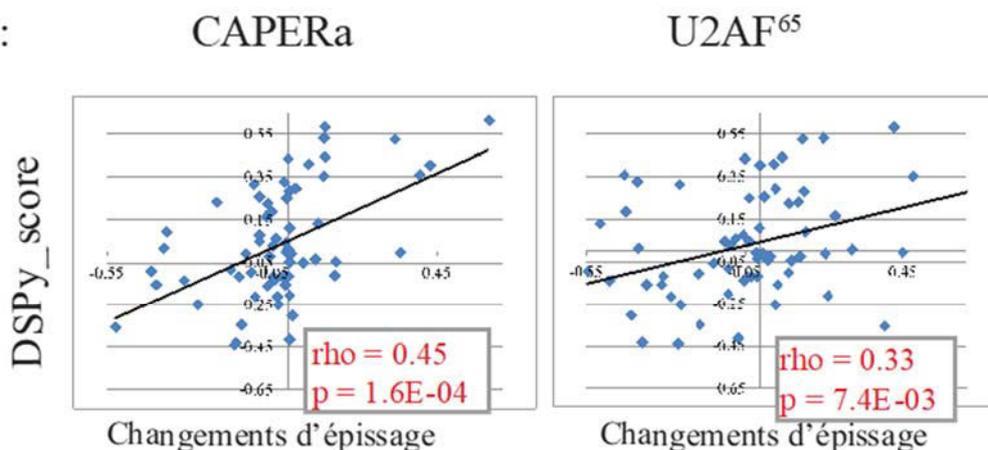
Figure 53. Représentation graphique en nuage de points des valeurs de changements d'épissage observées entre deux facteurs d'épissage à domaines UHM.

On peut alors se demander si ces cibles d'épissage sensibles de manière différentielle à chaque facteur d'épissage possèdent des séquences similaires. Afin d'analyser une potentielle corrélation entre l'effet de chaque facteur d'épissage et la nature de la séquence de l'intron en 5' de l'exon cassette, nous avons analysé le lien entre la valeur du changement d'épissage lors de l'inhibition de chaque protéine et l'enrichissement de la séquence en motifs polypyrimidine. Pour cela, nous avons recherché une corrélation entre le score Δ SPY calculé pour chaque cible d'épissage et le changement d'épissage observé lors de l'inhibition de chaque protéine. La figure 54 est une représentation graphique en nuage de points des valeurs de changements d'épissage par rapport au score Δ SPY. Nous observons une corrélation positive entre le score Δ SPY et les changements d'épissage observés lors de l'inhibition de CAPER α et U2AF⁶⁵ avec un coefficient de corrélation de Spearman respectivement de rho = 0,46, p = 0,0001 et rho = 0,33, p = 0,007.

Aucune corrélation significative entre le score Δ SPY et le changement d'épissage n'a été observée pour les protéines PUF60 et SPF45.

Ces résultats indiquent un lien fonctionnel entre CAPER α et U2AF⁶⁵ et la composition en pyrimidines des séquences de l'intron en amont de l'exon cassette.

Knockdown:



Knockdown:

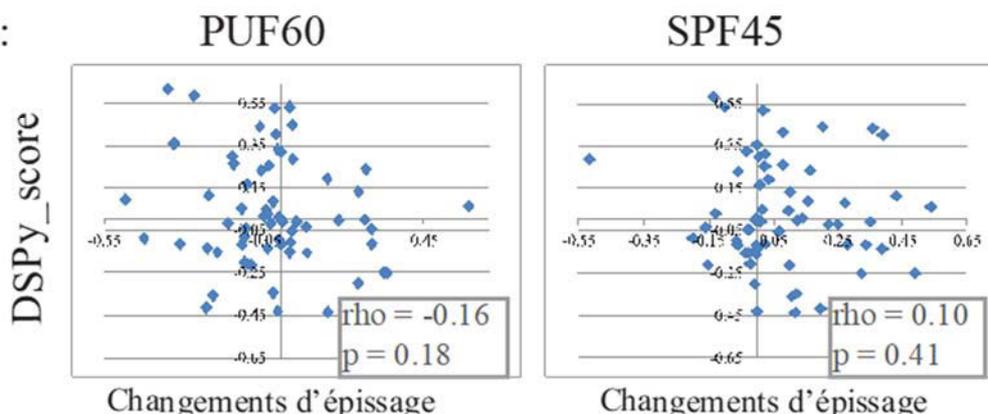


Figure 54. Représentation graphique en nuage de points des valeurs de changement d'épissage lors de l'inhibition de CAPER α , U2AF⁶⁵, PUF60 et SPF45 par rapport au score Δ SPY.

Ainsi, l'inclusion de l'exon est généralement augmentée par la réduction du taux de CAPER α et U2AF⁶⁵ lorsque la région intronique en 5' de l'exon cassette est riche en répétitions de polypyrimidine. Ce phénomène est inversé, c'est-à-dire que l'inclusion de l'exon cassette est diminuée, lors de l'inhibition de CAPER α et U2AF⁶⁵ lorsque les événements d'épissage possèdent une région intronique en 5' de l'exon cassette pauvre en répétitions de polypyrimidine.

Ces résultats permettent de proposer un modèle coopératif d'action des facteurs U2AF⁶⁵ et CAPER α dans la reconnaissance des séquences 3' en fonction de l'enrichissement en répétitions de polypyrimidine.

Afin d'étudier ce modèle, nous avons mis en place une technique d'analyse de l'épissage par des constructions de minigènes rapporteurs. Nous avons sélectionné deux événements d'épissage pour lesquels les changements d'épissage dus aux inhibitions de CAPER α et U2AF⁶⁵ étaient opposés. Ces deux événements d'épissage correspondent à l'exon 19 de l'ARNm TJP1 et à l'exon 13 de l'ARNm DNM2 (Figure 55.A).

L'exon cassette de L'ARNm TJP1 présente une valeur de changement d'épissage respectivement de -0,27 et -0,25 lors de l'inhibition de CAPER α et U2AF⁶⁵ et possède un score Δ SPY de -0,24 reflétant une pauvreté relative en répétitions de polypyrimidine.

L'exon cassette de l'ARNm DNM2 présente quant à lui une valeur de changement d'épissage respectivement de 0,60 et 0,70 lors de l'inhibition de CAPER α et U2AF⁶⁵ et possède un score Δ SPY de 0,61 reflétant une richesse relative en répétitions de polypyrimidine.

Nous avons réalisé des constructions de minigènes correspondants à ces deux ARNm (Figure 55). Pour le gène TJP1, la séquence du minigène correspond à la totalité de la séquence génomique comprise entre les 24 dernières bases de l'exon en amont de l'exon cassette (exon -1) et les 24 premières bases de l'exon en aval de l'exon cassette (exon +1). Etant donné la longueur des séquences introniques du gène DNM2 et les difficultés rencontrées au cours du clonage, la séquence du minigène a été raccourcie et est composée des 24 dernières bases de l'exon en amont de l'exon cassette (exon -1), 150 bases des séquences introniques de part et d'autre de l'exon cassette, la totalité de l'exon cassette et les 24 premières bases de l'exon en aval de l'exon cassette (exon +1) (Figure 55).

Afin d'étudier le lien fonctionnel entre la composition de la séquence intronique en 5' de l'exon cassette et la régulation de l'épissage de cet exon par CAPER α et U2AF⁶⁵, nous avons réalisé des constructions hybrides de ces minigènes (Figure 55). Pour cela, des séquences de 150 bases en 5' des exons cassette ont été interverties afin d'obtenir le minigène TJP1 avec une séquence intronique du gène DNM2 (TJP1iDNM2) et le minigène DNM2 avec une séquence intronique du gène TJP1 (DNM2iTJP1).

Afin d'identifier l'effet spécifique de ces séquences, des constructions hybrides de minigènes ont également été réalisées en introduisant dans l'intron du minigène TJP1 une séquence composée de répétitions de 9 uridines (TJP1_9U) ou une séquence de 22 bases de forte affinité pour U2AF⁶⁵ présente dans l'intron du gène *sxl* (TJP1_sxl).

Afin d'étudier l'effet de l'inhibition de CAPER α et U2AF⁶⁵ sur l'épissage des minigènes, le taux des protéines U2AF⁶⁵ et CAPER α a été diminué dans des cellules HeLa par transduction virale de shRNA. Pour chacune des protéines, deux shRNA spécifiques ont été utilisés et l'expérience a été réalisée deux fois. Des cellules sont également infectées avec des virus contrôle (NS). Après sélection des cellules sur la base de leur résistance à la puromycine, elles sont maintenues en culture pendant deux jours. Après 48 heures de culture, les cellules sont transfectées avec chaque construction de minigène et maintenues en culture pendant 24 heures.

Après extraction des ARN totaux et transcription inverse, des analyses par qPCR ont été réalisées en utilisant deux paires d'amorces pour chaque construction afin de mesurer le taux d'expression de chaque forme avec (Sin) et sans (Sout) exon cassette.

Pour chacun des exons cassette et pour chaque condition de déplétion, l'index d'épissage (SI) et le changement d'épissage (SC) ont été calculés (Figure 56, Figure 57 et Figure 58). Les valeurs des moyennes des index d'épissage SI ou changement d'épissage SC pour chaque condition sont représentées sous forme de graphiques. La figure 56 représente les index et changements d'épissage des minigènes TJP1 et DNM2. On observe une diminution de l'épissage de TJP1 et une augmentation de l'épissage de DNM2 lors de l'inhibition de CAPER α et U2AF⁶⁵ par rapport au contrôle NS. Ces résultats concordent avec les changements d'épissage observés lors de l'analyse des changements d'épissage des gènes endogènes (Figure 49) et confirment la présence d'un lien fonctionnel entre la composition de la séquence intronique en 5' de l'exon cassette et la régulation de l'épissage de cet exon par CAPER α et U2AF⁶⁵.

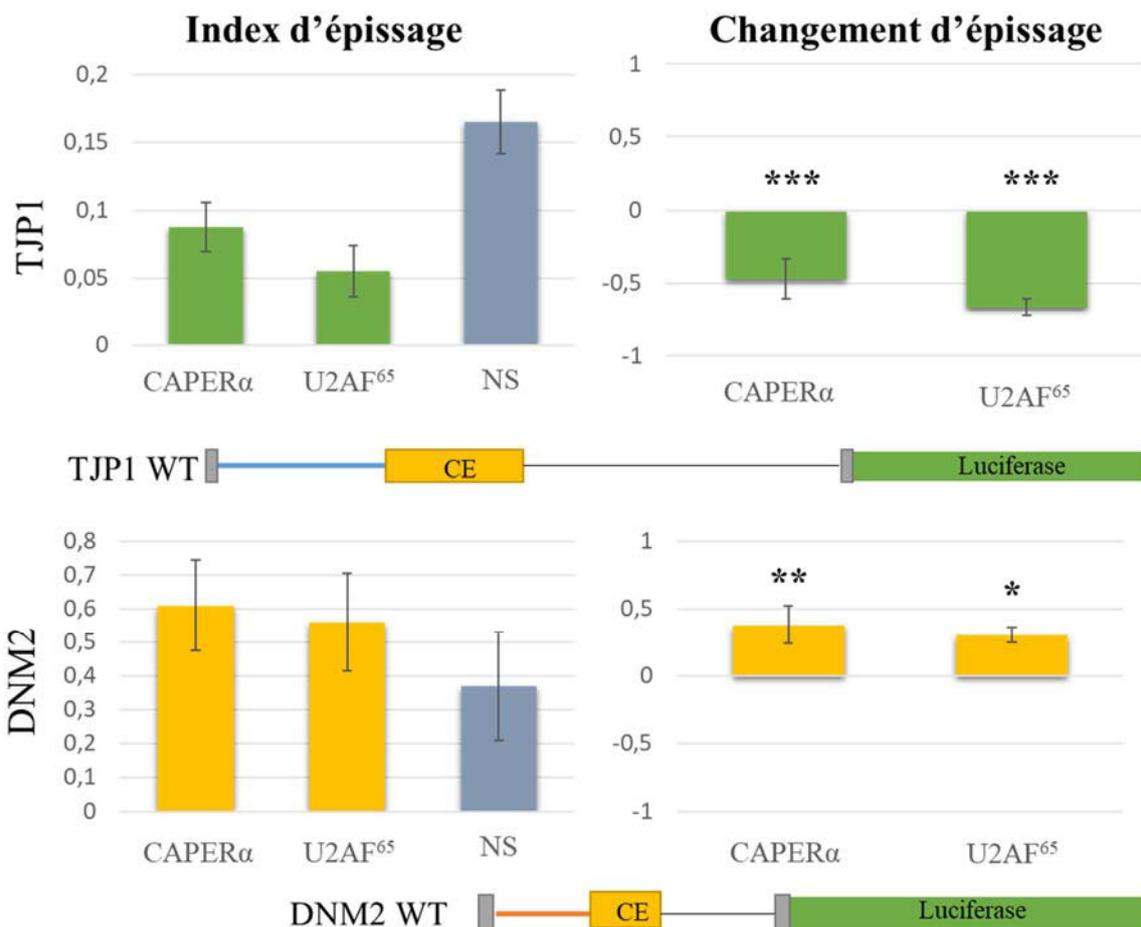


Figure 56. Représentation graphique des index d'épissage et des changements d'épissage des minigènes rapporteurs TJP1 ET DNM2 observés lors de l'inhibition de U2AF⁶⁵ et CAPER α . **** $p < 0.0005$ *** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$

Lors de l'analyse des index et changements d'épissage des minigènes hybrides dans lesquels les 150 bases en 5' de l'exon cassette ont été interverties (TJP1iDNM2 et DNM2iTJP1) (Figure 57), on observe une modification de l'index et du changement d'épissage lors de l'inhibition de CAPER α et U2AF⁶⁵. En effet, en introduisant la séquence intronique riche en répétitions de pyrimidines de DNM2 en amont de l'exon cassette de TJP1, on observe une augmentation de l'index et du changement d'épissage du minigène lorsque CAPER α et U2AF⁶⁵ sont inhibées. L'effet inverse est également observé pour le minigène DNM2 ; on note une diminution de l'index et du changement d'épissage lors de l'introduction de la séquence intronique pauvre en répétition de pyrimidines de TJP1.

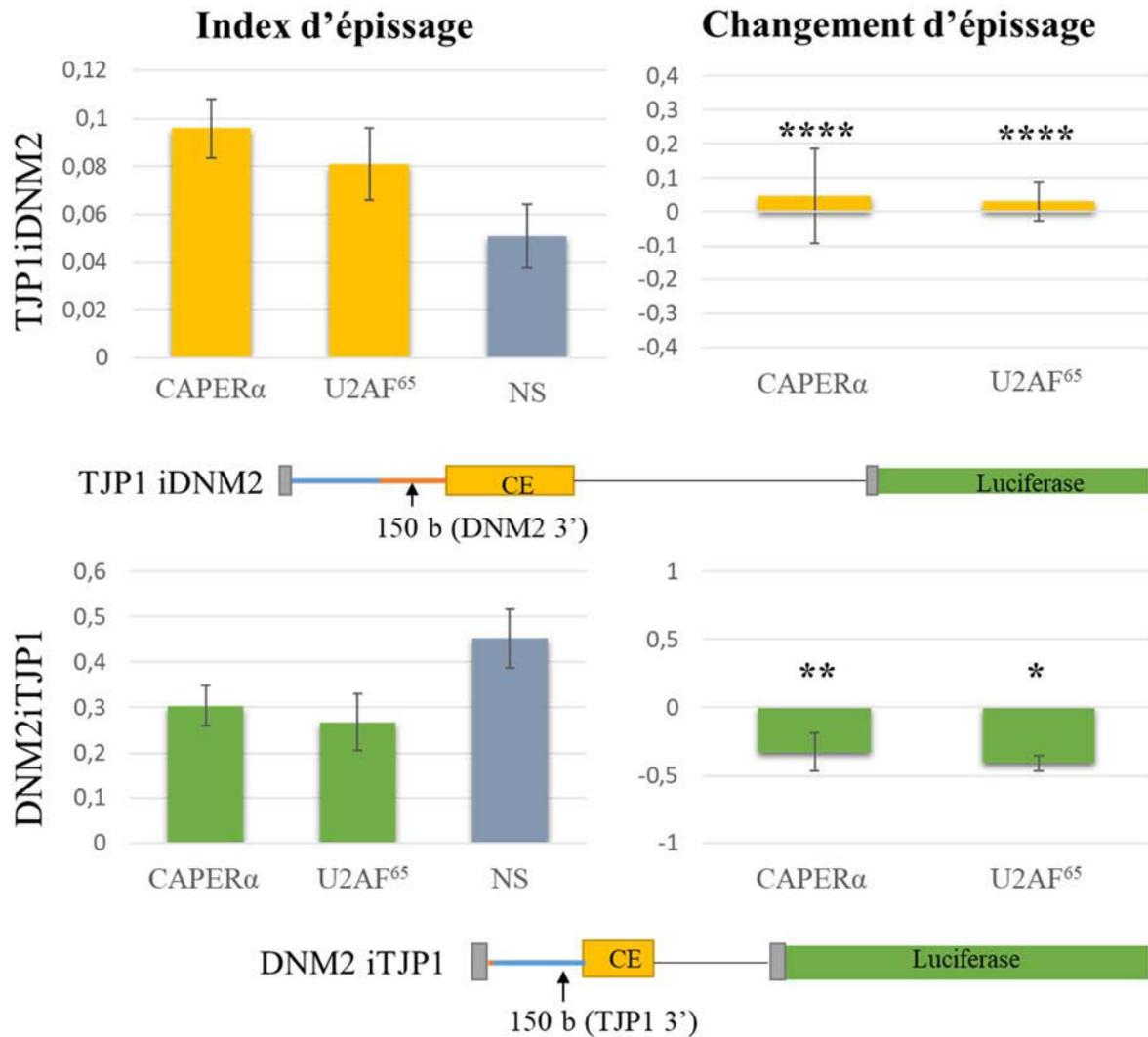


Figure 57. Représentation graphique des index d'épissage et des changements d'épissage des minigènes rapporteurs TJP1iDNM2 et DNM2iTJP1 observés lors de l'inhibition de U2AF⁶⁵ et CAPER α . **** $p < 0.0005$ *** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$

Lorsque la séquence de neuf uridines ou la séquence de forte affinité pour U2AF⁶⁵ est introduite dans l'intron en amont de l'exon cassette TJP1 (TJP1_9U et (TJP1_sxl)) (Figure 58), on observe une inclusion quasi-totale de l'exon cassette dans toutes les conditions étudiées quel que soit le taux d'expression de CAPER α ou U2AF⁶⁵.

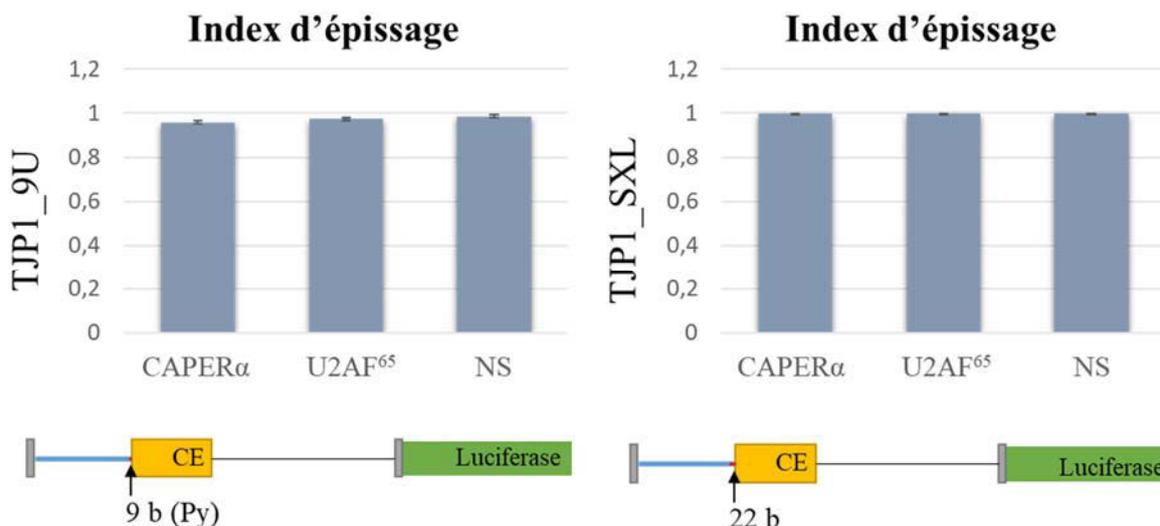


Figure 58. Représentation graphique des index d'épissage des constructions de minigènes rapporteurs TJP1_9U et TJP1_sxl observés lors de l'inhibition de U2AF⁶⁵ et CAPER α .

3. Etude des assemblages macromoléculaires formés par U2AF⁶⁵ et CAPER α autour du domaine multi-ULM de SF3b155

3.1. Constat initial

Les protéines CAPER α et U2AF⁶⁵ possèdent des structures protéiques similaires. Les deux protéines possèdent un domaine RS en N-terminal, deux domaines RRM centraux et un domaine UHM en C-terminal. Les résultats obtenus par Gst-pull down et les analyses des cibles d'épissage par qPCR ont révélé un lien particulier entre ces deux protéines. L'augmentation du recrutement de CAPER α lors de l'interaction avec SF3b155 en présence de U2AF⁶⁵ pourrait prédire la formation d'assemblages macromoléculaires par ces deux protéines pour la reconnaissance des sites d'épissage.

Des expériences de sédimentation en centrifugeant les échantillons à 20 000 xg pendant 10 min suivi d'un fractionnement des culots et des surnageants sur gel de polyacrylamide SDS-PAGE nous a permis de constater qu'une fraction importante de U2AF⁶⁵ et CAPER α sédimentait dans le culot. Ceci n'était pas le cas de PUF60, SPF45, SF1 ou SF3b155 (Figure 59). Plus particulièrement, la protéine U2AF⁶⁵ dont le domaine RS a été tronqué (U2AF⁶⁵ Δ RS) ne présentait plus cette propriété de sédimentation.

Ces observations suggèreraient que les propriétés de sédimentation particulières à U2AF⁶⁵ et CAPER α témoigneraient d'un mécanisme de formation d'assemblages macromoléculaires par ces protéines. Ces assemblages macromoléculaires seraient orchestrés par leur domaine RS.

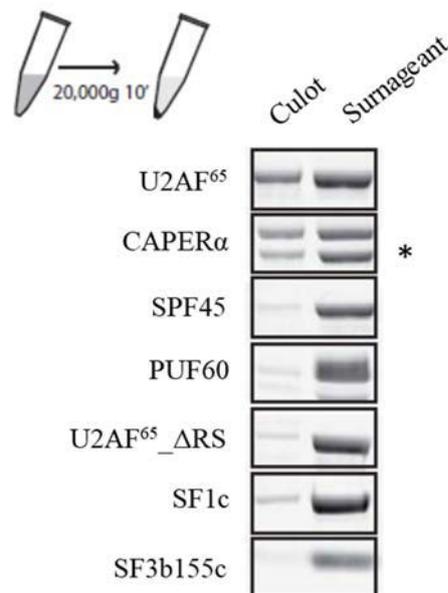


Figure 59. Sédimentation *in vitro* de CAPER α et U2AF⁶⁵. Les culots et surnageants correspondants à chaque échantillon ont été fractionnés sur gel de polyacrylamide SDS-PAGE. * produit de dégradation de CAPER α

3.2. Formation des assemblages macromoléculaires de CAPER α et U2AF⁶⁵ avec le domaine multi-ULM de SF3b155 dépendant de la composition de l'ARN en enchainement de polypyrimidine

Afin d'étudier le rôle fonctionnel dans l'épissage des assemblages macromoléculaires formés par CAPER α et U2AF⁶⁵, nous avons réalisé des expériences de sédimentation dans différentes conditions. Lors de l'étude du rôle des facteurs d'épissage à domaines UHM dans la régulation de l'épissage alternatif de 65 exons cassettes, nous avons observé un effet commun de CAPER α et U2AF⁶⁵ sur la régulation de l'épissage dépendant de la composition en enchainements de polypyrimidine de l'intron en amont de l'exon cassette.

Sur la base de ces résultats, nous avons tenté d'observer si cette composition différentielle des introns en enchainements de polypyrimidine permettait de moduler la formation des assemblages macromoléculaires par CAPER α et U2AF⁶⁵.

Afin de tester cette hypothèse, des ARNs de 130 bases correspondant aux régions introniques en 5' des exons cassettes des pré-ARNm de TJP1 et DNM2 ont été transcrits *in vitro*. Comme nous l'avons vu précédemment, l'intron en 5' de l'exon cassette de TJP1 est pauvre en polypyrimidine, tandis que celle de l'intron en 5' de l'exon cassette de DNM2 est très riche en enchainements en polypyrimidine.

Des expériences de sédimentation ont ensuite été réalisées en incubant la protéine U2AF⁶⁵ seule ou en présence des ARN transcrits *in vitro*. Afin d'analyser l'effet de SF3b155 sur la formation d'assemblage macromoléculaires par U2AF⁶⁵, des expériences ont également été réalisées en ajoutant la protéine SF3b155 (SF3b155c) ou la protéine SF3b155 dont les tryptophanes des domaines ULM ont été mutés en alanine (SF3b155_ΔW).

Après centrifugation, les fractions de chaque échantillon représentant le culot (C) et le surnageant (S) ont été fractionnées sur gel de polyacrylamide et les protéines ont été révélées par coloration du gel au bleu de Coomassie (Figure 60).

L'intensité du signal U2AF⁶⁵ représentant les fractions insolubles (C) et solubles (S) de la protéine ont été quantifiées avec le logiciel ImageJ. Dans la figure 60, nous observons une augmentation de la fraction insoluble de U2AF⁶⁵ en présence de l'ARN DNM2, tandis que celle-ci reste inchangée en présence de l'ARN TJP1. La fraction insoluble de U2AF⁶⁵ est encore augmentée en présence de DNM2 lorsque la protéine SF3b155 est ajoutée à la réaction alors que la protéine SF3b155 avec des domaines ULM non fonctionnels n'a pas d'effet sur la sédimentation de U2AF⁶⁵. En effet, la fraction insoluble de U2AF⁶⁵ en présence de DNM2 et de SF3b155c est deux fois plus importante qu'en présence de DNM2 seul ou en présence de DNM2 et SF3b155_ΔW. Par ailleurs, l'ajout de SF3b155 en présence de TJP1 n'augmente pas la fraction insoluble de U2AF⁶⁵.

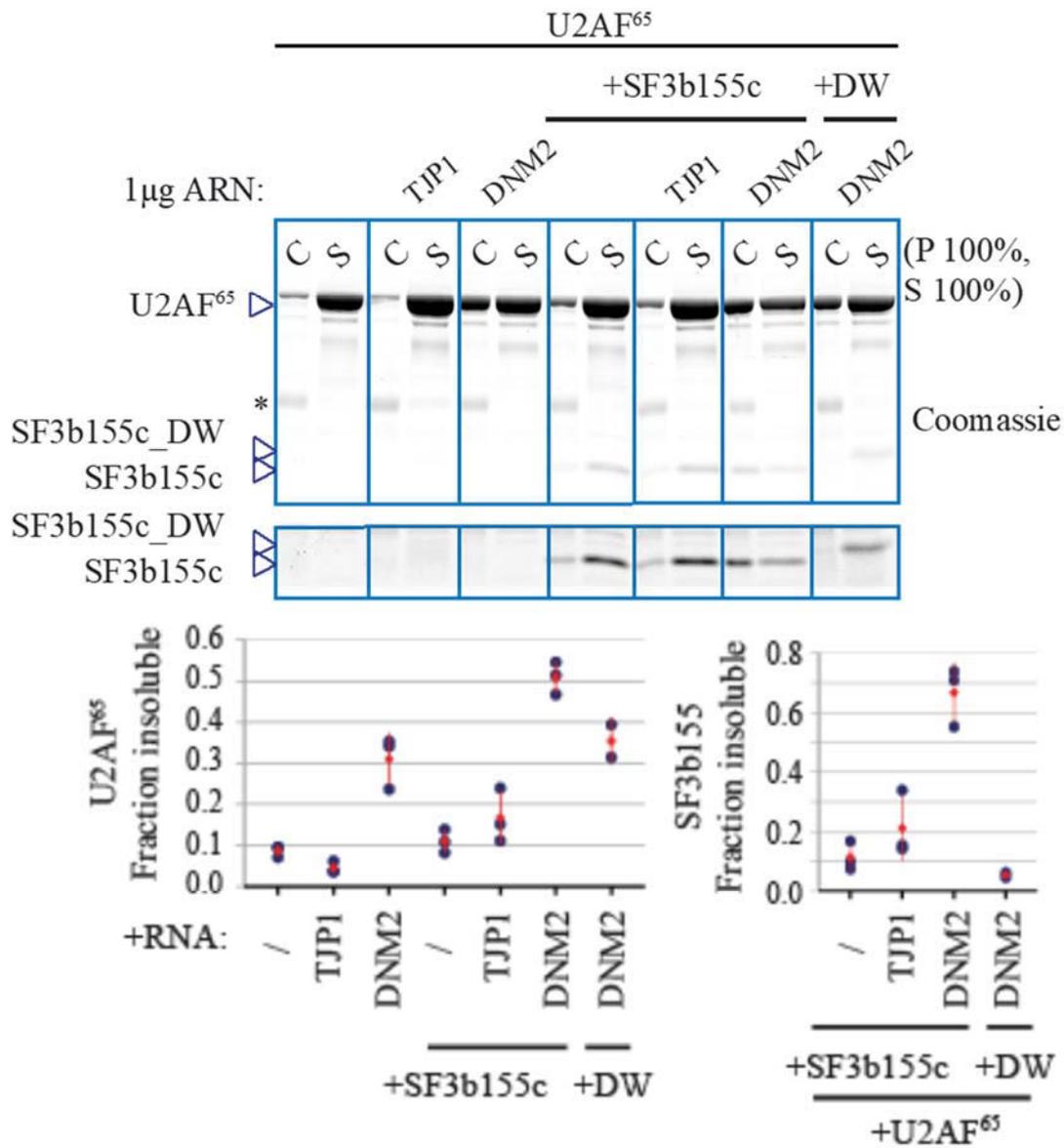


Figure 60. Sédimentation différentielle de U2AF⁶⁵ avec les ARN TJP1 et DNM2 en présence de SF3b155. Les fractions de chaque échantillon représentant le culot (C) et le surnageant (S) ont été déposées sur gel de polyacrylamide et les protéines ont été révélées par coloration du gel au bleu de Coomassie. L'intensité du signal U2AF⁶⁵ et SF3b155 représentant les fractions insolubles (C) et fractions solubles (S) ont été quantifiées avec le logiciel ImageJ. *Protéines de lait.

Afin de suivre le comportement des ARN TJP1 et DNM2 dans différentes conditions, les mêmes expériences de sédimentation ont été réalisées et les fractions correspondant au culot (C) et au surnageant (S) de chaque échantillon ont été déposés sur gel d'agarose 1% et la présence d'ARN a été révélée par exposition aux UV (Figure 61).

En l'absence de U2AF⁶⁵, les ARN DNM2 et TJP1 sont retrouvés exclusivement dans la partie soluble de l'échantillon. Lorsque la protéine U2AF⁶⁵ est ajoutée, on observe une partie de l'ARN DNM2 dans la fraction insoluble. De plus, comme nous l'avons observé précédemment pour U2AF⁶⁵, la quantité d'ARN DNM2 présente dans la fraction insoluble augmente en présence de SF3b155. Ceci n'est pas le cas pour l'ARN TJP1 qui reste dans la fraction soluble dans toutes les conditions.

L'enrichissement de U2AF⁶⁵ et de l'ARN DNM2 dans la fraction insoluble en présence de SF3b155 suggère que SF3b155 intervient dans le recrutement de U2AF⁶⁵ par l'intermédiaire d'interactions UHM-ULM. U2AF⁶⁵ formerait alors des assemblages macromoléculaires par l'intermédiaire d'interactions protéine-protéine grâce à son domaine RS en présence d'ARN possédant des séquences riches en enchainements de polypyrimidine.

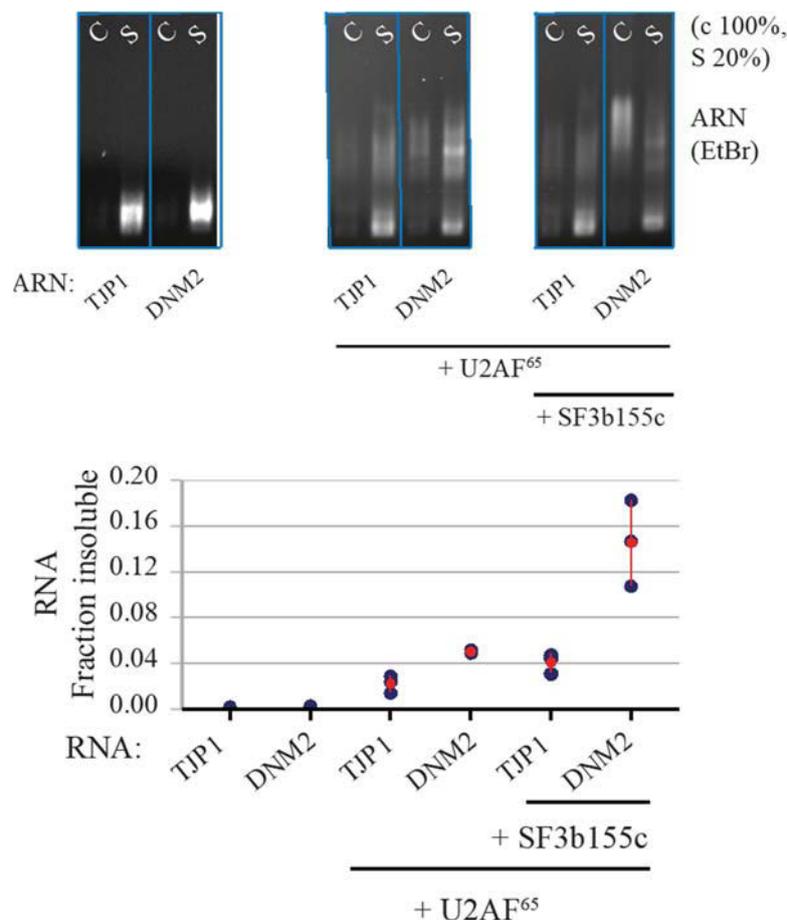


Figure 61. Sédimentation différentielle des ARN TJP1 et DNM2 avec U2AF⁶⁵ en présence de SF3b155. Les fractions de chaque échantillon représentant le culot (C) et le surnageant (S) ont été déposées sur gel d'agarose et les ARN ont été révélés par exposition aux UV. L'intensité de chaque signal été quantifiée avec le logiciel ImageJ.

Afin de rechercher l'implication de CAPERα dans la formation d'assemblages macromoléculaires dans l'épissage et d'étudier l'influence mutuelle de CAPERα et U2AF⁶⁵ dans la reconnaissance des sites d'épissage en 3', des expériences de sédimentation similaires ont été réalisées.

Pour cela, 0,5 µg d'ARN DNM2 produits *in vitro* ont été incubés avec les protéines CAPERα, U2AF⁶⁵ ou les deux protéines à la fois en présence ou non de la protéine SF3b155.

Après centrifugation, les fractions de chaque échantillon représentant le culot (C) et le surnageant (S) ont été déposés sur gel de polyacrylamide et les protéines ont été révélées par western blot en utilisant des anticorps spécifiques reconnaissant chaque protéine (Figure 62).

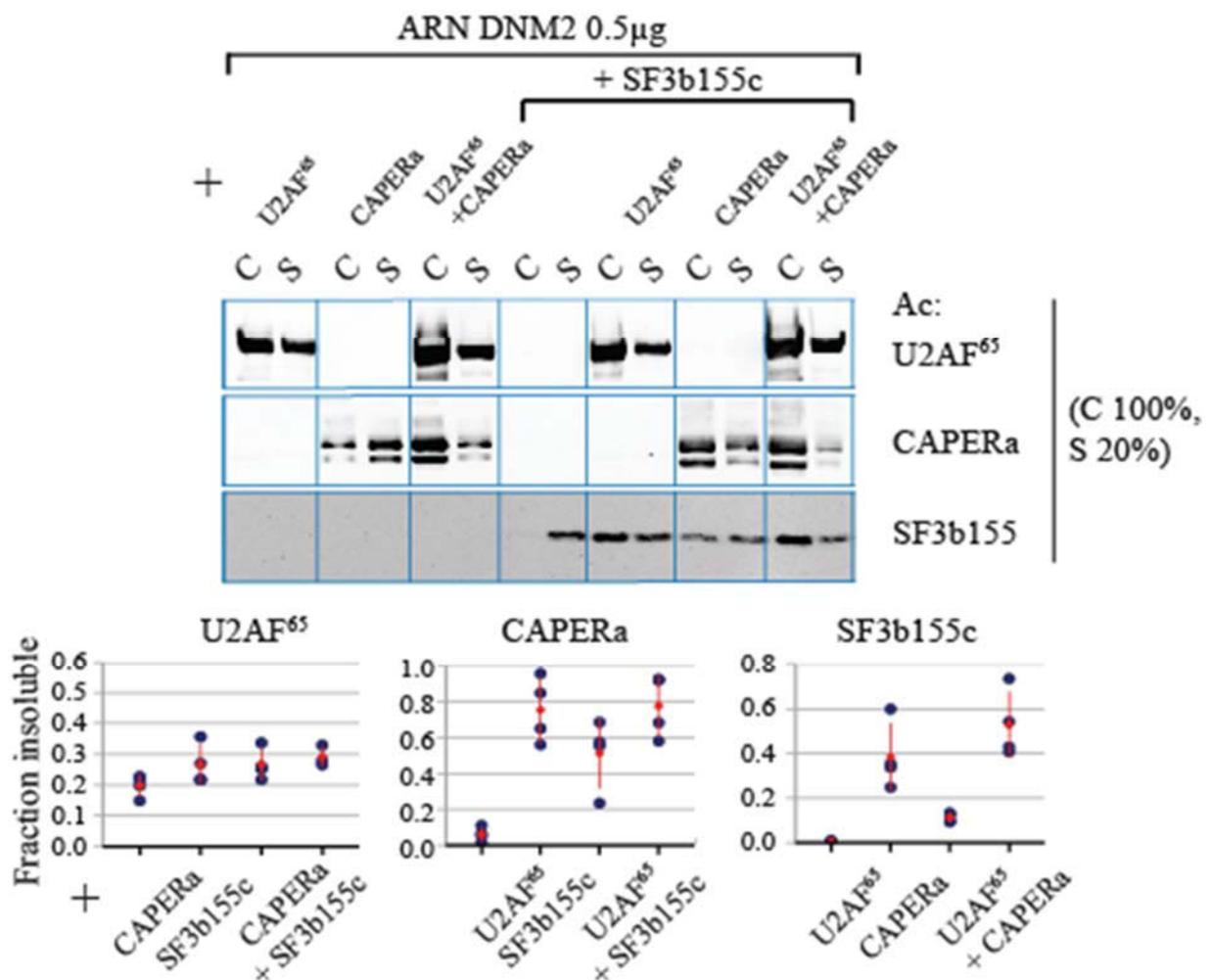


Figure 62. Formation d'assemblages macromoléculaires par U2AF⁶⁵ et CAPERα en présence de SF3b155 et d'ARN riche en enchainements de pyrimidines. Les fractions de chaque échantillon représentant le culot (C) et le surnageant (S) ont été déposées sur gel de polyacrylamide et les protéines ont été révélées par western blot. L'intensité de chaque signal a été quantifiée avec le logiciel ImageJ

Afin de suivre le comportement de l'ARN DNM2, les mêmes expériences de sédimentation ont été réalisées et les fractions de chaque échantillon représentant le culot (C) et le surnageant (S) ont été fractionnées sur gel d'agarose 1%.

Dans la figure 62, nous observons une augmentation de la fraction insoluble de U2AF⁶⁵ lorsque CAPER α est présente. De même, la présence de U2AF⁶⁵ renforce la quantité de CAPER α dans la fraction insoluble.

Les quantités de CAPER α et U2AF⁶⁵ dans la fraction insoluble sont plus importantes en présence de SF3b155. De même, SF3b155 favorise la sédimentation de DNM2 (Figure 63). En effet, nous observons une forte augmentation de DNM2 dans la fraction insoluble lorsque les deux protéines CAPER α et U2AF⁶⁵ sont présentes dans le mélange. Cette augmentation est également observée lorsqu'une des deux protéines est présente et SF3b155 est ajoutée. La plus forte fraction insoluble de l'ARN DNM2 est observée lorsque les trois protéines sont présentes simultanément dans le mélange.

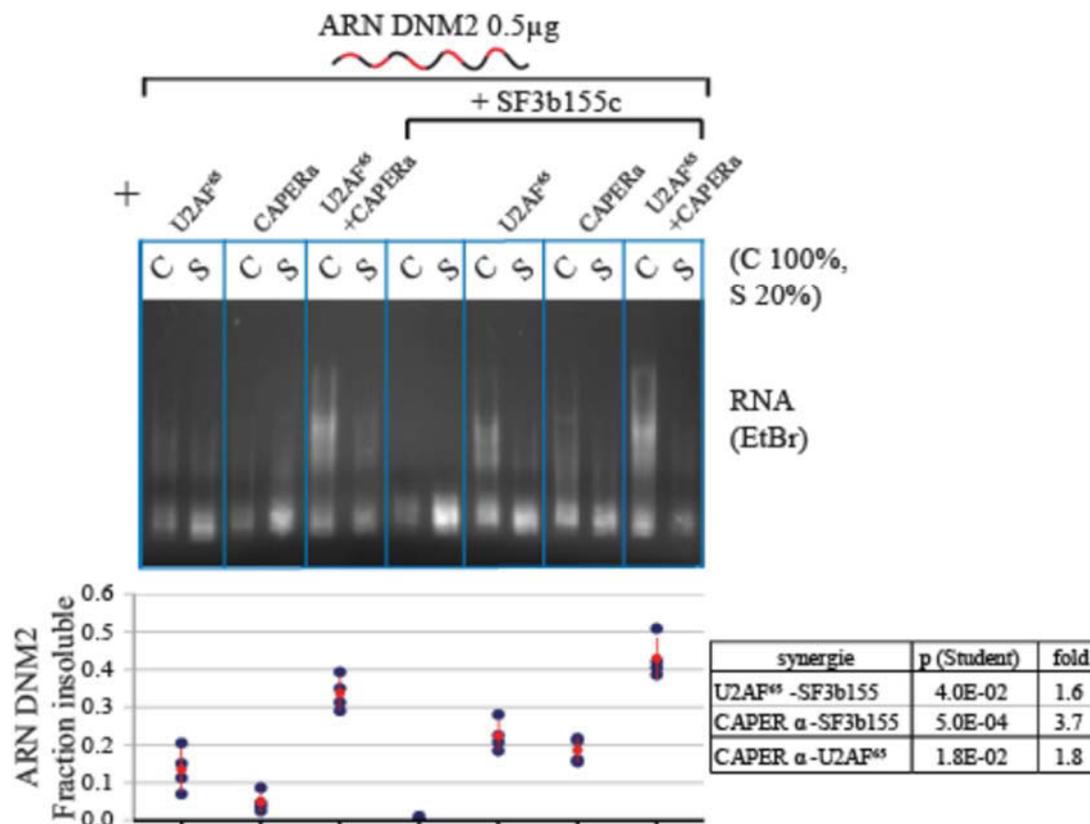


Figure 63. Formation d'assemblages macromoléculaires par U2AF⁶⁵ et CAPER α en présence de SF3b155 et d'ARN riche en enchainements de pyrimidines. Les fractions de chaque échantillon représentant le culot (C) et le surnageant (S) ont été déposées sur gel d'agarose et les ARN ont été révélés par exposition aux UV. L'intensité de chaque signal a été quantifiée avec le logiciel ImageJ

L'augmentation des assemblages macromoléculaires de U2AF⁶⁵ et CAPER α en présence de l'ARN DNM2 et l'augmentation de l'ARN DNM2 dans la fraction insoluble en présence de SF3b155 suggère que CAPER α et U2AF⁶⁵ interviennent simultanément dans le recrutement des ARN possédant des séquences riches en polypyrimidine durant l'épissage. Ce recrutement serait favorisé par des interactions de type UHM-ULM avec SF3b155.

V. Discussion, perspectives et conclusion

Le mécanisme d'épissage est une dynamique hautement complexe qui fait intervenir des centaines de protéines et de complexes ribonucléoprotéiques. Afin de générer des transcrits matures, les exons doivent être identifiés et joints de manière précise. Ainsi, un dérèglement ou un défaut d'assemblage de la machinerie d'épissage, une surexpression ou une inhibition de l'expression d'un facteur de régulation de l'épissage conduit à une mauvaise maturation du transcrit primaire. Cette altération de la machinerie d'épissage peut induire une reconnaissance incorrecte des sites qui définissent les jonctions entre les exons et les introns, induisant un processus d'épissage défectueux, générant des ARNm instables. Ces transcrits anormaux seront soit dégradés, soit traduits en protéines. Ces dernières seront tronquées, défectueuses, et ne pourront plus assurer leur rôle physiologique dans la cellule.

Chez l'Homme, les séquences qui définissent les sites d'épissage sont peu conservées. Ces séquences sont nécessaires, mais insuffisantes pour définir les frontières entre les exons et les introns. Le processus d'épissage fait alors intervenir des protéines régulatrices pour s'assurer de l'intégrité de la définition des jonctions entre les exons et les introns.

Afin de comprendre le processus complexe de régulation de l'épissage, il est important de connaître et de savoir comment se déroule le processus initial de reconnaissance et de sélection des sites d'épissage en compétition durant l'assemblage du spliceosome.

La production de plusieurs ARNm matures à partir d'un seul transcrit primaire est un mécanisme répandu chez l'Homme. Ce mécanisme dépend d'une coordination entre des processus cellulaires qui se déroulent de manière simultanée.

L'épissage alternatif dans les cellules neuronales a pendant longtemps été le mécanisme le plus étudié. Non seulement le défaut d'épissage dans ces cellules induit des phénotypes observables et identifiables, mais est également causé par des mutations de facteurs d'épissages spécifiques. Ces informations permettent aujourd'hui de mieux identifier les processus impliqués dans la modulation de l'épissage de ces transcrits caractéristiques des cellules neuronales.

Les études génomiques à grande échelle qui émergent ces dernières années ont permis d'augmenter le répertoire et d'identifier de nouvelles protéines régulatrices de l'épissage dans différents tissus. Ces études soulignent la complexité de ce processus et l'importance de caractériser le rôle de ces nouveaux facteurs dans l'épissage.

Les travaux de notre équipe visent à caractériser les interactions nécessaires à l'identification des sites 3' d'épissage. Ces sites possèdent des séquences peu conservées et requièrent de multiples interactions.

1. Interactions des facteurs d'épissage à domaines UHM avec SF3b155

Depuis leur découverte et leur caractérisation au début des années 2000, plusieurs équipes ont étudié les interactions biochimiques et les rôles fonctionnels des facteurs d'épissage à domaines UHM dans la régulation de l'épissage alternatif.

Une localisation des protéines CAPER α , PUF60 et SPF45 dans les complexes spliceosomaux [Jurica and Moore, 2003; Neubauer et al., 1998] et de CAPER α , U2AF⁶⁵ et SPF45 dans les speckles [Anderson et al., 2016; Sampath et al., 2003; Zamore and Green, 1991], suggéré un potentiel de ces protéines dans l'épissage.

Des interactions entre CAPER α et U2AF⁶⁵ avaient déjà été observées par des techniques de transfert d'énergie entre molécules fluorescentes (FRET) couplées à une microscopie à imagerie de durée de vie de fluorescence (FLIM) [Ellis et al., 2008]. Une présence simultanée de PUF60 et U2AF⁶⁵ [Hastings et al., 2007] au niveau des sites 3' d'épissage avait également été préalablement identifiée.

Par des expériences de *Proximity Ligation assay*, nous avons pu mettre en évidence une proximité de toutes les combinaisons deux à deux des facteurs d'épissage à domaines UHM CAPER α , PUF60, SPF45 et U2AF⁶⁵ dans des cellules humaines.

Afin d'examiner si les facteurs d'épissage à domaines UHM CAPER α , PUF60 et SPF45 intervenaient comme U2AF⁶⁵ pendant les étapes précoces de la reconnaissance des sites 3' d'épissage, nous avons tenté de visualiser une proximité de ces protéines avec SF3b155 par *Proximity Ligation Assay*. Cependant, nous n'avons pas pu observer de signaux de fluorescence témoignant d'une proximité spatiale des protéines par cette technique. Il se pourrait que cette absence de signaux soit due à des encombrements stériques empêchant la proximité des deux anticorps primaires ou la réaction d'amplification. Il ne serait alors possible d'observer un signal PLA que lorsque les deux protéines ne font pas partie d'un complexe. Dans les cellules, la protéine SF3b155 est généralement associée à la snRNP U2, un grand complexe

ribonucléoprotéique d'une taille de 18 nm [Kastner et al., 1990]. Il se pourrait donc que l'on soit dans les limites de la sensibilité de la technique.

Afin d'observer dans la cellule si les proximités de CAPER α , PUF60, SPF45 et U2AF⁶⁵ sont dépendantes de l'épissage, il serait intéressant d'inhiber l'expression de SF3b155 par shRNA ou par la technologie CRISPR-Cas9 ou de bloquer totalement l'épissage en traitant les cellules à la spliceostatine A, une molécule qui cible la snRNP U2. Nous pourrions alors observer par *Proximity Ligation Assay* si les proximités spatiales des protéines précédemment observées sont maintenues. Ces expériences nous donneraient plus d'information quant à la compréhension et le mécanisme de recrutement des facteurs d'épissage à domaines UHM lors des étapes précoces de reconnaissance des sites 3' d'épissage. Un maintien de la proximité après inhibition du facteur SF3b155 indiquerait un recrutement des protéines UHM au site 3' indépendant de SF3b155, comme c'est le cas du facteur U2AF. Cependant, si la colocalisation n'est plus observée, cela pourrait peut-être révéler un rôle de SF3b155 dans le recrutement des facteurs d'épissage à domaines UHM autour des sites 3' d'épissage.

A l'aide des expériences de GST pull down à partir d'extraits cellulaires et nucléaires, nous avons montré une préférence d'interaction de CAPER α , PUF60 et SPF45 avec SF3b155 par l'intermédiaire d'interactions UHM-ULM.

Ces résultats complètent les expériences par Gst-pull down et par RMN des équipes de Michael Sattler [Corsini et al., 2007, 2009] et Clara Kielkopf [Loerch et al., 2014] qui avaient montré que les domaines UHM isolés de SPF45, PUF60 et CAPER α étaient capables d'interagir avec les séquences ULM de SF3b155.

2. U2AF⁶⁵ renforce le recrutement de CAPER α lors des interactions avec SF3b155

Les expériences de Gst-pull down que nous avons réalisées au cours de cette étude ont révélé un rapport particulier entre CAPER α et U2AF⁶⁵. En effet, nous avons tout d'abord pu observer que l'ajout de U2AF⁶⁵ améliorerait le recrutement de CAPER α sur SF3b155, ce qui n'est pas le cas pour PUF60 et SPF45. Nous avons ensuite pu constater que le domaine RS situé en région N-terminale de U2AF⁶⁵ était nécessaire pour cette action de U2AF⁶⁵.

La similitude d'organisation structurale de CAPER α et U2AF⁶⁵, incluant un domaine RS en N-terminal, deux domaines RRM et un domaine UHM en C-terminal témoigne également d'une relation particulière entre les deux protéines. Par des expériences de sédimentation, nous avons

pu observer que les deux protéines avaient la capacité de former des assemblages macromoléculaires *in vitro* grâce à leurs domaines RS. L'augmentation de la sédimentation de CAPER α et U2AF⁶⁵ en présence de SF3b155 suggère que ces assemblages macromoléculaires se déroulent pendant l'épissage autour du domaine multi-ULM de SF3b155.

Ceci est d'autant plus convaincant lorsque l'on observe que l'interaction de CAPER α et U2AF⁶⁵ avec SF3b155 est dépendante et renforcée par le nombre de tryptophanes présents dans les motifs ULM de SF3b155.

3. Rôle des facteurs d'épissage à domaines UHM dans l'épissage des exons cassette

Les résultats d'analyses qPCR à moyen débit de 65 exons cassette ont montré que CAPER α , U2AF⁶⁵ et PUF60 apparaissaient plutôt comme des facteurs activateurs de l'épissage. Ces observations concordent avec les récentes études des analyses d'expression du transcriptome par RNAseq dans des cellules MCF-7 dans lesquelles le taux de CAPER α a été réduit [Mai et al., 2016] ainsi que les analyses du transcriptome par RNAseq dans des cellules HEK où le taux de PUF60 a été diminué [Královicová et al., 2018].

La protéine SPF45 quant à elle, apparaît généralement comme un inhibiteur de l'épissage. Des observations similaires ont été faites lors de l'analyse de l'expression différentielle du transcriptome dans des cellules de Purkinje déficientes en SPF45 [Tan et al., 2016].

Plus particulièrement, notre étude a montré une corrélation entre les exons cassettes régulés par CAPER α et U2AF⁶⁵, en concordance avec la publication de *Sanyue Mai et al.*, [Mai et al., 2016] qui avait révélé qu'approximativement 20% des événements d'épissage alternatif régulés par CAPER α étaient régulés de manière similaire par U2AF⁶⁵.

Cependant, aucune corrélation de cibles communes avec PUF60 ou SPF45 n'a pu être observée dans nos conditions expérimentales. Afin d'augmenter le nombre d'évènements d'épissage à analyser, il serait intéressant de réaliser des études comparatives par RNAseq des transcriptomes des cellules dont l'expression de chacun des facteurs d'épissage CAPER α , PUF60, SPF45 et U2AF⁶⁵ aura été préalablement inhibée. Ces analyses permettront de confirmer la corrélation entre CAPER α et U2AF⁶⁵, mais surtout d'identifier les cibles de PUF60 et SPF45 dans l'épissage.

Nos analyses de l'épissage ont mis en évidence une corrélation entre un type de séquences en 3' de l'intron et la régulation de l'inclusion de l'exon par CAPER α et U2AF⁶⁵. En effet, nous avons observé une corrélation statistiquement significative entre l'inclusion de l'exon par CAPER α et U2AF⁶⁵ et la richesse de l'intron en 5' de l'exon cassette en enchainements de polypyrimidine. En utilisant des minigènes rapporteurs de l'épissage, nous avons confirmé que CAPER α et U2AF⁶⁵ activaient l'épissage des exons cassette précédés par des séquences riches en enchainements de polypyrimidine.

Afin d'identifier l'arrangement précis des séquences introniques en 5' de l'exon cassette reconnues par CAPER α et U2AF⁶⁵, il serait intéressant de réaliser des expériences de mutagenèse dirigée sur les introns des minigènes rapporteurs et d'analyser, par qPCR, les changements d'épissage résultants des différentes mutations.

Il serait également intéressant de réaliser des expériences de CLIPseq afin d'identifier les sites de fixations de chacune des protéines sur le pré-ARNm et d'étudier de possibles similitudes de séquences.

Par ailleurs, aucune corrélation statistiquement significative entre la composition des séquences de l'intron en enchainements de polypyrimidine et l'effet de PUF60 ou de SPF45 sur l'épissage de l'exon cassette n'a pu être observée dans nos conditions expérimentales. Néanmoins, les récentes études de Vorechvosky et collaborateurs, [Královicová et al., 2018] ont montré que les sites d'épissage 3' des exons cassettes dont l'inclusion était favorisée par PUF60 présentaient une séquence polypyrimidine plus longue que la moyenne. Ces résultats font écho à nos observations concernant U2AF⁶⁵ et CAPER α .

Dans nos expériences de sédimentation, nous avons également observé qu'une séquence riche en polypyrimidine favorisait la formation d'assemblages macromoléculaires de U2AF⁶⁵ et CAPER α . Il serait intéressant de réaliser des expériences de sédimentation en présence de PUF60 et de SPF45 et ainsi observer l'effet de ces protéines sur les assemblages macromoléculaires formés par U2AF⁶⁵ et CAPER α .

Dans nos expériences de sédimentation, nous observons que le domaine multi-ULM de SF3b155 favorise la formation d'assemblages macromoléculaires de U2AF⁶⁵ et CAPER α en présence de séquences riches en enchainements de polypyrimidine. En concordance avec nos expériences de Gst-pull down avec différents mutants de SF3b155 possédants 1, 2 ou 3 tryptophanes présents dans des motifs ULM, nos observations démontrent l'importance du

domaine multi-ULM de SF3b155 dans la formation d'assemblages macromoléculaires par CAPER α et U2AF⁶⁵.

4. Modèle d'assemblage des facteurs d'épissage à domaines UHM autour de SF3b155 lors de la reconnaissance des sites 3' d'épissage

Soutenus par les résultats obtenus au cours de cette étude et des données de la littérature, nous proposons un modèle d'assemblage des facteurs d'épissage à domaines UHM au cours des étapes précoces de l'épissage (Figure 64). Ce modèle est soutenu par l'observation d'un effet différentiel de la régulation de l'épissage par CAPER α et U2AF⁶⁵ en fonction de la composition des introns en enchainements de polypyrimidine, et la confirmation de cet effet par des minigènes rapporteurs de l'épissage. Ces observations sont concordantes avec la formation d'assemblages macromoléculaires par CAPER α et U2AF⁶⁵ orchestrée par leurs domaines RS en présence de séquences riches en enchainements de polypyrimidine. De plus, le domaine multi-ULM de SF3b155 améliore la formation d'assemblages macromoléculaires de CAPER α et U2AF⁶⁵.

En s'appuyant sur ces résultats, nous proposons que des assemblages macromoléculaires des facteurs d'épissage à domaines UHM sont formés en contact avec les séquences introniques en 5' de l'exon cassette présentant de multiple motifs polypyrimidine. Plusieurs molécules de U2AF⁶⁵ et de CAPER α interagiraient avec ces séquences ainsi que le domaine multi-ULM de SF3b155.

La formation d'assemblages macromoléculaires et les combinaisons d'interactions de CAPER α et U2AF⁶⁵ avec les polypyrimidines permettraient alors la reconnaissance des sites 3' d'épissage et le recrutement de la snRNP U2 au point de branchement par l'intermédiaire d'interactions avec le domaine multi-ULM de SF3b155.

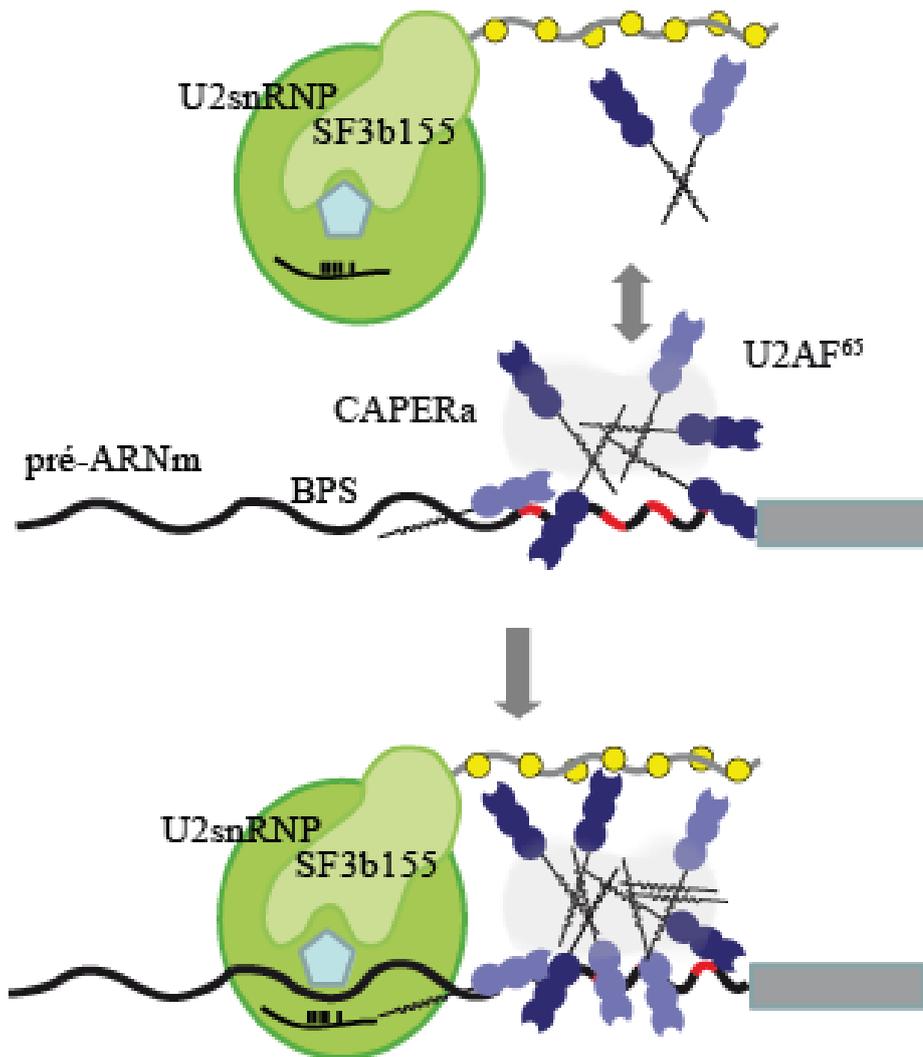


Figure 64. Modèle de formation d'assemblages macromoléculaires de CAPER α et U2AF⁶⁵ pour la reconnaissance des séquences introniques riches en enchainements de pyrimidines et le recrutement de la snRNP U2.

5. Les facteurs d'épissage à domaines UHM : enjeux et stratégies thérapeutiques ?

Plusieurs mutations dans les séquences de U2AF³⁵ et SF3b155 ont été associées à des syndromes myélodysplasiques [Yoshida et al., 2011].

En fonction du niveau d'expression et du type cellulaire, CAPER α peut favoriser la progression tumorale. Par exemple, CAPER α est surexprimée dans des cas de cancers du poumon, certains cancers du sein et du colon [Mercier et al., 2014]. Ces observations sont peut-être liées au rôle de CAPER α dans la régulation de l'épissage. En effet, des analyses d'expression différentielles

du dans des cellules MCF-7 où le taux de CAPER α a été diminué, avaient révélé que la protéine était impliquée dans la régulation de l'épissage de plusieurs gènes impliqués dans la réponse aux dommages liés à l'ADN, ou la régulation du cycle cellulaire [Mai et al., 2016].

De même, la régulation spécifique de l'épissage alternatif du gène FAS par SPF45 [Corsini et al., 2007] et la surexpression de SPF45 dans plusieurs cancers [Han et al., 2018] suggère que la protéine jouerait un rôle dans la régulation de l'expression des gènes impliqués dans la prolifération, la survie et la progression du cycle cellulaire.

Des délétions et des mutations engendrant des pertes de fonction de PUF60 ont été retrouvées chez des patients présentant des retards développementaux, des microcéphalies et des malformations craniofaciales, rénales et cardiaques [Dauber et al., 2013]. Ces pertes de fonction induiraient alors des défauts d'épissage ou d'épissage alternatif de gènes exprimés au cours du développement.

L'ensemble des mutations retrouvées dans les facteurs d'épissage à domaines UHM, induisant des pertes de fonctions, ainsi que les sous- ou surexpressions, engendreraient des déséquilibres dans la régulation de l'épissage de cibles caractéristiques à chacune des protéines, conduisant à l'apparitions de syndromes myélodysplasiques, de cancers et de retards développementaux.

Dans leur globalité, les facteurs d'épissage à domaines UHM représentent des cibles thérapeutiques intéressantes. Après avoir identifié et caractérisé le profil d'épissage des pré-ARNm affectés par un déficit ou une surexpression de ces facteurs d'épissage, il serait possible de restaurer un profil d'épissage typique.

Afin d'étudier le rôle fonctionnel des facteurs d'épissage à domaines UHM et identifier leurs cibles durant l'épissage, l'équipe de M.Sattler [Jagtap et al, 2016] a mis en place une technique basée sur un peptide inhibiteur des domaines UHM. Ce peptide cyclique d'une dizaine d'acides aminés se lie spécifiquement au domaine UHM de SPF45. Des analyses *in vitro* des cibles d'épissage de SPF45 ont révélé que ce peptide inhibiteur était capable de moduler les interactions UHM-ULM. Cependant, cette stratégie doit encore être optimisée *in vivo*, puisque l'affinité du peptide pour le domaine UHM de SPF45 était 500 à 1000 fois plus faible dans les cellules. Cette faible affinité serait peut-être due à des interactions non spécifiques du peptide avec des protéines cellulaires ou a un faible recrutement dans le noyau.

Au laboratoire, Asaki Kobayashi et Alexandre Maucuer en collaboration avec la société SYNSIGHT, ont identifié par criblage de banques de données, grâce à des outils de dynamique moléculaire et en s'appuyant sur les données structurales obtenues par RMN, de petites molécules candidates capables d'interagir avec les domaines UHM des UHMSF. Par des expériences de RMN, ils chercheront à identifier les molécules présentant les meilleures affinités pour les domaines UHM et analyseront par qPCR, les changements d'épissage dans les cellules résultants de l'interaction des UHMSF avec la molécule.

Ces études présentent un intérêt non seulement dans le développement d'outils d'analyse des mécanismes d'épissage régulés par les UHMSF, mais procurent également à long terme des stratégies thérapeutiques afin de moduler l'action des facteurs d'épissage lorsqu'ils sont surexprimés dans de nombreux cancers et syndromes myélodysplasiques.

VI. Références bibliographiques

- Amara SG, Jonas V, Rosenfeld MG, Ong ES, Evans RM. 1982. Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature* 298(5871):240–244.
- Anderson DM, Cannavino J, Li H, Anderson KM, Nelson BR, McAnally J, Bezprozvannaya S, Liu Y, Lin W, Liu N, Bassel-Duby R, Olson EN. 2016. Severe muscle wasting and denervation in mice lacking the RNA-binding protein ZFP106. *Proc Natl Acad Sci U S A* 113(31):E4494–E4503.
- Ankö M-L, Neugebauer KM. 2012. RNA-protein interactions in vivo: global gets specific. *Trends Biochem Sci* 37(7):255–262.
- Bandziulis RJ, Swanson MS, Dreyfuss G. 1989. RNA-binding proteins as developmental regulators. *Genes Dev* 3(4):431–437.
- Batsché E, Yaniv M, Muchardt C. 2006. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol* 13(1):22–29.
- Bauer WJ, Heath J, Jenkins JL, Kielkopf CL. 2012. Three RNA recognition motifs participate in RNA recognition and structural organization by the pro-apoptotic factor TIA-1. *Journal of molecular biology* 415(4):727–740.
- Bennett M, Michaud S, Kingston J, Reed R. 1992. Protein components specifically associated with prespliceosome and spliceosome complexes. *Genes Dev* 6(10):1986–2000.
- Berget S, Moore C, Sharp P. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* 74(8):3171–3175.
- Berglund J, Abovich N, Rosbash M. 1998a. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes & development* 12(6):858–867.
- Berglund J, Chua K, Abovich N, Reed R, Rosbash M. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89(5):781–787.
- Berglund J, Fleming M, Rosbash M. 1998b. The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA (New York, NY)* 4(8):998–1006.
- Berk AJ, Sharp PA. 1978. Spliced early mRNAs of simian virus 40. *Proc Natl Acad Sci USA* 75(3):1274–1278.
- Birney E, Kumar S, Krainer AR. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* 21(25):5803–5816.
- Blencowe BJ, Baurén G, Eldridge AG, Issner R, Nickerson JA, Rosonina E, Sharp PA. 2000. The SRm160/300 splicing coactivator subunits. *RNA* 6(1):111–120.
- Boise LH, González-García M, Postema CE, Ding L, Lindsten T, Turka LA, Mao X, Nuñez G, Thompson CB. 1993. *bcl-x*, a *bcl-2*-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* 74(4):597–608.
- Bond SR, Naus CC. 2012. RF-Cloning.org: an online tool for the design of restriction-free cloning projects. *Nucleic Acids Res* 40(Web Server issue):W209-213.

- Branlant C, Krol A, Ebel JP, Lazar E, Gallinaro H, Jacob M, Sri-Widada J, Jeanteur P. 1980. Nucleotide sequences of nuclear U1A RNAs from chicken, rat and man. *Nucleic Acids Res* 8(18):4143–4154.
- Brown MT, Cooper JA. 1996. Regulation, substrates and functions of src. *Biochim Biophys Acta* 1287(2–3):121–149.
- Burd CG, Dreyfuss G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* 265(5172):615–621.
- Cáceres JF, Misteli T, Sreaton GR, Spector DL, Krainer AR. 1997. Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol* 138(2):225–238.
- Cao W, Jamison SF, Garcia-Blanco MA. 1997. Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. *RNA* 3(12):1456–1467.
- Cartegni L, Maconi M, Morandi E, Cobianchi F, Riva S, Biamonti G. 1996. hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J Mol Biol* 259(3):337–348.
- Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31(13):3568–3571.
- Cass DM, Berglund J. 2006. The SF3b155 N-terminal domain is a scaffold important for splicing. *Biochemistry* 45(33):10092–10101.
- Castro E de, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* 34(Web Server issue):W362–365.
- Cavaloc Y, Bourgeois CF, Kister L, Stévenin J. 1999. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA* 5(3):468–483.
- Cazalla D, Newton K, Cáceres JF. 2005. A novel SR-related protein is required for the second step of Pre-mRNA splicing. *Molecular and cellular biology* 25(8):2969–2980.
- Cazzola M, Rossi M, Malcovati L, Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie Mieloproliferative. 2013. Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. *Blood* 121(2):260–269.
- Chai Y, Liu X, Dai L, Li Y, Liu M, Zhang J-Y. 2014. Overexpression of HCC1/CAPER α may play a role in lung cancer carcinogenesis. *Tumour Biol* 35(7):6311–6317.
- Champion-Arnaud P, Reed R. 1994. The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes Dev* 8(16):1974–1983.
- Chandler SD, Mayeda A, Yeakley JM, Krainer AR, Fu XD. 1997. RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. *Proc Natl Acad Sci USA* 94(8):3596–3601.
- Chatrikhi R, Wang W, Gupta A, Loerch S, Maucuer A, Kielkopf CL. 2016. SF1 Phosphorylation Enhances Specific Binding to U2AF(65) and Reduces Binding to 3'-Splice-Site RNA. *Biophysical journal* 111(12):2570–2586.

- Chen CD, Kobayashi R, Helfman DM. 1999. Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev* 13(5):593–606.
- Chen M, Manley JL. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* 10(11):741–754.
- Chiara C de, Menon RP, Strom M, Gibson TJ, Pastore A. 2009. Phosphorylation of S776 and 14-3-3 binding modulate ataxin-1 interaction with splicing factors. *PLoS one* 4(12).
- Cho S, Moon H, Loh TJ, Jang HN, Liu Y, Zhou J, Ohn T, Zheng X, Shen H. 2015. Splicing inhibition of U2AF65 leads to alternative exon skipping. *Proceedings of the National Academy of Sciences of the United States of America* 112(32):9926–9931.
- Chodavarapu RK, Feng S, Bernatavichute YV, Chen P-Y, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, Clark AT, Krämer U, Merchant SS, Zhang X, Jacobsen SE, Pellegrini M. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature* 466(7304):388–392.
- Chow L, Gelinis R, Broker T, Roberts R. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12(1):1–8.
- Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, Bell JC, Duncan PI. 1996. The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J* 15(2):265–275.
- Corsini L, Bonnal S, Bonna S, Basquin J, Hothorn M, Scheffzek K, Valcárcel J, Sattler M. 2007. U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nature structural & molecular biology* 14(7):620–629.
- Corsini L, Hothorn M, Stier G, Rybin V, Scheffzek K, Gibson TJ, Sattler M. 2009. Dimerization and protein binding specificity of the U2AF homology motif of the splicing factor Puf60. *The Journal of biological chemistry* 284(1):630–639.
- Corvelo A, Hallegger M, Smith CW, Eyras E. 2010. Genome-wide association between branch point properties and alternative splicing. *PLoS computational biology* 6(11).
- Das B, Xia L, Palandjian L, Gozani O, Chyung Y, Reed R. 1999. Characterization of a protein complex containing spliceosomal proteins SAPs 49, 130, 145, and 155. *Molecular and cellular biology* 19(10):6796–6802.
- Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, Gygi SP, Reed R. 2007. SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell* 26(6):867–881.
- Dauber A, Golzio C, Guenot C, Jodelka FM, Kibaek M, Kjaergaard S, Leheup B, Martinet D, Nowaczyk MJ, Rosenfeld JA, Zeesman S, Zunic J, Beckmann JS, Hirschhorn JN, Hastings ML, Jacquemont S, Katsanis N. 2013. SCRIB and PUF60 are primary drivers of the multisystemic phenotypes of the 8q24.3 copy-number variant. *American journal of human genetics* 93(5):798–811.
- Daubner GM, Cléry A, Allain FH-T. 2013. RRM-RNA recognition: NMR or crystallography...and new findings. *Curr Opin Struct Biol* 23(1):100–108.
- Dembowski JA, An P, Scoulos-Hanson M, Yeo G, Han J, Fu X-D, Grabowski PJ. 2012. Alternative Splicing of a Novel Inducible Exon Diversifies the CASK Guanylate Kinase Domain. *J Nucleic Acids* 2012:816237.

Dowhan DH, Hong EP, Auboeuf D, Dennis AP, Wilson MM, Berget SM, O'Malley BW. 2005. Steroid hormone receptor coactivation and alternative RNA splicing by U2AF65-related proteins CAPERalpha and CAPERbeta. *Molecular cell* 17(3):429–439.

Dreyfuss G, Kim V, Kataoka N. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nature reviews Molecular cell biology* 3(3):195–205.

Dutta J, Fan G, Gélinas C. 2008. CAPERalpha is a novel Rel-TAD-interacting factor that inhibits lymphocyte transformation by the potent Rel/NF-kappaB oncoprotein v-Rel. *Journal of virology* 82(21):10792–10802.

Dvinge H. 2018. Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Lett* 592(17):2987–3006.

Eekelen CA van, Riemen T, Venrooij WJ van. 1981. Specificity in the interaction of hnRNA and mRNA with proteins as revealed by in vivo cross linking. *FEBS Lett* 130(2):223–226.

Eldridge AG, Li Y, Sharp PA, Blencowe BJ. 1999. The SRm160/300 splicing coactivator is required for exon-enhancer function. *Proc Natl Acad Sci USA* 96(11):6125–6130.

Ellis JD, Llères D, Denegri M, Lamond AI, Cáceres JF. 2008. Spatial mapping of splicing factor complexes involved in exon and intron definition. *The Journal of cell biology* 181(6):921–934.

Eperon IC, Makarova OV, Mayeda A, Munroe SH, Cáceres JF, Hayward DG, Krainer AR. 2000. Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol* 20(22):8303–8318.

Eperon LP, Graham IR, Griffiths AD, Eperon IC. 1988. Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell* 54(3):393–401.

Erkelenz S, Mueller WF, Evans MS, Busch A, Schöneweis K, Hertel KJ, Schaal H. 2013. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA* 19(1):96–102.

Estes PA, Cooke NE, Liebhaber SA. 1992. A native RNA secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms. *J Biol Chem* 267(21):14902–14908.

Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, Lührmann R. 2009. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* 36(4):593–608.

Fairbrother WG, Chasin LA. 2000. Human genomic sequences that inhibit splicing. *Mol Cell Biol* 20(18):6816–6825.

Fetzer S, Lauber J, Will CL, Lührmann R. 1997. The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated protein kinase. *RNA* 3(4):344–355.

Görlach M, Burd CG, Dreyfuss G. 1994. The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. *J Biol Chem* 269(37):23074–23078.

Görlach M, Burd CG, Portman DS, Dreyfuss G. 1993. The hnRNP proteins. *Mol Biol Rep* 18(2):73–78.

- Gozani O, Feld R, Reed R. 1996. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev* 10(2):233–243.
- Grosso AR, Gomes AQ, Barbosa-Morais NL, Caldeira S, Thorne NP, Grech G, Lindern M von, Carmo-Fonseca M. 2008. Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Res* 36(15):4823–4832.
- Guiro J, O'Reilly D. 2015. Insights into the U1 small nuclear ribonucleoprotein complex superfamily. *Wiley Interdiscip Rev RNA* 6(1):79–92.
- Guth S, Valcárcel J. 2000. Kinetic role for mammalian SF1/BBP in spliceosome assembly and function after polypyrimidine tract recognition by U2AF. *The Journal of biological chemistry* 275(48):38059–38066.
- Hallay H, Locker N, Ayadi L, Ropers D, Guittet E, Branlant C. 2006. Biochemical and NMR study on the competition between proteins SC35, SRp40, and heterogeneous nuclear ribonucleoprotein A1 at the HIV-1 Tat exon 2 splicing site. *J Biol Chem* 281(48):37159–37174.
- Han SP, Tang YH, Smith R. 2010. Functional diversity of the hnRNPs: past, present and perspectives. *Biochem J* 430(3):379–392.
- Han Y, Zhang M, Chen D, Li H, Wang X, Ma S. 2018. Downregulation of RNA binding motif protein 17 expression inhibits proliferation of hypopharyngeal carcinoma FaDu cells. *Oncol Lett* 15(4):5680–5684.
- Hastings ML, Allemand E, Duelli DM, Myers MP, Krainer AR. 2007. Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF(65). *PloS one* 2(6).
- He W, Parker R. 2000. Functions of Lsm proteins in mRNA degradation and splicing. *Curr Opin Cell Biol* 12(3):346–350.
- Hnilicová J, Hozeifi S, Dušková E, Icha J, Tománková T, Staněk D. 2011. Histone deacetylase activity modulates alternative splicing. *PLoS ONE* 6(2):e16727.
- Hsin J-P, Manley JL. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* 26(19):2119–2137.
- Huang G, Zhou Z, Wang H, Kleinerman ES. 2012. CAPER- α alternative splicing regulates the expression of vascular endothelial growth factor₁₆₅ in Ewing sarcoma cells. *Cancer* 118(8):2106–2116.
- Hui J, Hung L-H, Heiner M, Schreiner S, Neumüller N, Reither G, Haas SA, Bindereif A. 2005. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J* 24(11):1988–1998.
- Hui J, Stangl K, Lane WS, Bindereif A. 2003. HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat Struct Biol* 10(1):33–37.
- Jung D-JJ, Na S-YY, Na DS, Lee JW. 2002. Molecular cloning and characterization of CAPER, a novel coactivator of activating protein-1 and estrogen receptors. *The Journal of biological chemistry* 277(2):1229–1234.
- Jurica MS, Moore MJ. 2003. Pre-mRNA splicing: awash in a sea of proteins. *Molecular cell* 12(1):5–14.

- Kastner B, Bach M, Lührmann R. 1990. Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common structure-determining principle in the major U snRNP family. *Proc Natl Acad Sci USA* 87(5):1710–1714.
- Kielkopf C, Rodionova N, Green M, Burley S. 2001. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* 106(5):595–605.
- Kielkopf CL, Lücke S, Green MR. 2004. U2AF homology motifs: protein recognition in the RRM world. *Genes & development* 18(13):1513–1526.
- Kiledjian M, Dreyfuss G. 1992. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J* 11(7):2655–2664.
- Kondé E, Bourgeois B, Tellier-Lebegue C, Wu W, Pérez J, Caputo S, Attanda W, Gasparini S, Charbonnier J-BB, Gilquin B, Worman HJ, Zinn-Justin S. 2010. Structural analysis of the Smad2-MAN1 interaction that regulates transforming growth factor- β signaling at the inner nuclear membrane. *Biochemistry* 49(37):8020–8032.
- Konieczny P, Stepniak-Konieczna E, Sobczak K. 2014. MBNL proteins and their target RNAs, interaction and splicing regulation. *Nucleic acids research* 42(17):10873–10887.
- Královicová J, Ševčíková I, Stejskalová E, Obuca M, Hiller M, Stanek D, Vorechovský I. 2018. PUF60-activated exons uncover altered 3' splice-site selection by germline missense mutations in a single RRM. *Nucleic Acids Res* 46(12):6166–6187.
- Lallena MJJ, Chalmers KJ, Llamazares S, Lamond AI, Valcárcel J. 2002. Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* 109(3):285–296.
- Lambowitz AM, Zimmerly S. 2011. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harbor perspectives in biology* 3(8).
- Laski FA, Rio DC, Rubin GM. 1986. Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing. *Cell* 44(1):7–19.
- Lewis HA, Musunuru K, Jensen KB, Edo C, Chen H, Darnell RB, Burley SK. 2000. Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* 100(3):323–332.
- Libri D, Lemonnier M, Meinnel T, Fiszman MY. 1989. A single gene codes for the beta subunits of smooth and skeletal muscle tropomyosin in the chicken. *J Biol Chem* 264(5):2935–2944.
- Lim J, Crespo-Barreto J, Jafar-Nejad P, Bowman AB, Richman R, Hill DE, Orr HT, Zoghbi HY. 2008. Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. *Nature* 452(7188):713–718.
- Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. 2011. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci USA* 108(27):11093–11098.
- Liu J, He L, Collins I, Ge H, Libutti D, Li J, Egly J, Levens D. 2000. The FBP interacting repressor targets TFIIH to inhibit activated transcription. *Molecular cell* 5(2):331–341.

- Loerch S, Kielkopf CL. 2016. Unmasking the U2AF homology motif family: a bona fide protein-protein interaction motif in disguise. *RNA (New York, NY)* 22(12):1795–1807.
- Loerch S, Maucuer A, Manceau V, Green MR, Kielkopf CL. 2014. Cancer-relevant splicing factor CAPER α engages the essential splicing factor SF3b155 in a specific ternary complex. *The Journal of biological chemistry* 289(25):17325–17337.
- Long JC, Caceres JF. 2009. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* 417(1):15–27.
- Low KJ, Ansari M, Abou Jamra R, Clarke A, El Chehadeh S, FitzPatrick DR, Greenslade M, Henderson A, Hurst J, Keller K, Kuentz P, Prescott T, Roessler F, Selmer KK, Schneider MC, Stewart F, Tatton-Brown K, Thevenon J, Vigeland MD, Vogt J, Willems M, Zonana J, Study DDD, Smithson SF. 2017. PUF60 variants cause a syndrome of ID, short stature, microcephaly, coloboma, craniofacial, cardiac, renal and spinal features. *Eur J Hum Genet* 25(5):552–559.
- Mai S, Qu X, Li P, Ma Q, Cao C, Liu X. 2016. Global regulation of alternative RNA splicing by the SR-rich protein RBM39. *Biochimica et biophysica acta* 1859(8):1014–1024.
- Manceau V, Kielkopf CL, Sobel A, Maucuer A. 2008. Different requirements of the kinase and UHM domains of KIS for its nuclear localization and binding to splicing factors. *Journal of molecular biology* 381(3):748–762.
- Manceau V, Swenson M, Le Caer J-PP, Sobel A, Kielkopf CL, Maucuer A. 2006. Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF65. *The FEBS journal* 273(3):577–587.
- Manley JL, Krainer AR. 2010. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev* 24(11):1073–1074.
- Markovtsov V, Nikolic JM, Goldman JA, Turck CW, Chou MY, Black DL. 2000. Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol Cell Biol* 20(20):7463–7479.
- Martinez-Contreras R, Fiset J-F, Nasim FH, Madden R, Cordeau M, Chabot B. 2006. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol* 4(2):e21.
- Massenet S, Branlant C. 1999. A limited number of pseudouridine residues in the human ataxic spliceosomal UsnRNAs as compared to human major spliceosomal UsnRNAs. *RNA* 5(11):1495–1503.
- Mata M de la, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR. 2003. A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell* 12(2):525–532.
- Maucuer A, Ozon S, Manceau V, Gavet O, Lawler S, Curmi P, Sobel A. 1997. KIS is a protein kinase with an RNA recognition motif. *The Journal of biological chemistry* 272(37):23151–23156.
- Maunakea AK, Chepelev I, Cui K, Zhao K. 2013. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res* 23(11):1256–1269.
- McLaughlin KJ, Jenkins JL, Kielkopf CL. 2011. Large favorable enthalpy changes drive specific RNA recognition by RNA recognition motif proteins. *Biochemistry* 50(9):1429–1431.

- Mercier I, Gonzales DM, Quann K, Pestell TG, Molchansky A, Sotgia F, Hult J, Gandara R, Wang C, Pestell RG, Lisanti MP, Jasmin J-FF. 2014. CAPER, a novel regulator of human breast cancer progression. *Cell cycle (Georgetown, Tex)* 13(8):1256–1264.
- Muto Y, Pomeranz Krummel D, Oubridge C, Hernandez H, Robinson CV, Neuhaus D, Nagai K. 2004. The structure and biochemical properties of the human spliceosomal protein U1C. *J Mol Biol* 341(1):185–198.
- Neubauer G, King A, Rappsilber J, Calvio C, Watson M, Ajuh P, Sleeman J, Lamond A, Mann M. 1998. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet* 20(1):46–50.
- Page-McCaw P, Amonlirdviman K, Sharp P. 1999. PUF60: a novel U2AF65-related splicing activity. *Rna*.
- Pandit S, Zhou Y, Shiue L, Coutinho-Mansfield G, Li H, Qiu J, Huang J, Yeo GW, Ares M, Fu X-D. 2013a. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol Cell* 50(2):223–235.
- Pandit S, Zhou Y, Shiue L, Coutinho-Mansfield G, Li H, Qiu J, Huang J, Yeo GW, Ares M, Fu X-DD. 2013b. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Molecular cell* 50(2):223–235.
- Papasaikas P, Tejedor J, Vigevani L, Valcárcel J. 2015. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Molecular cell* 57(1):7–22.
- Patel AA, Steitz JA. 2003. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* 4(12):960–970.
- Patel NA, Kaneko S, Apostolatos HS, Bae SS, Watson JE, Davidowitz K, Chappell DS, Birnbaum MJ, Cheng JQ, Cooper DR. 2005. Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase Cbeta11 alternative splicing via phosphorylation of serine/arginine-rich splicing factor SRp40. *J Biol Chem* 280(14):14302–14309.
- Peled-Zehavi H, Berglund J, Rosbash M, Frankel A. 2001. Recognition of RNA branch point sequences by the KH domain of splicing factor 1 (mammalian branch point binding protein) in a splicing factor complex. *Molecular and cellular biology* 21(15):5232–5241.
- Pellizzoni L, Kataoka N, Charroux B, Dreyfuss G. 1998. A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell* 95(5):615–624.
- Piñol-Roma S, Dreyfuss G. 1993. hnRNP proteins: localization and transport between the nucleus and the cytoplasm. *Trends Cell Biol* 3(5):151–155.
- Prasad J, Colwill K, Pawson T, Manley JL. 1999. The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. *Mol Cell Biol* 19(10):6991–7000.
- Prigge JR, Iverson SV, Siders AM, Schmidt EE. 2009. Interactome for auxiliary splicing factor U2AF(65) suggests diverse roles. *Biochimica et biophysica acta* 1789(6–8):487–492.
- Rain J, Rafi Z, Rhani Z, Legrain P, Krämer A. 1998. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA (New York, NY)* 4(5):551–565.

- Rappsilber J, Ryder U, Lamond AI, Mann M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res* 12(8):1231–1245.
- Reed R, Griffith J, Maniatis T. 1988. Purification and visualization of native spliceosomes. *Cell* 53(6):949–961.
- Ropers D, Ayadi L, Gattoni R, Jacquenet S, Damier L, Branlant C, Stévenin J. 2004. Differential effects of the SR proteins 9G8, SC35, ASF/SF2, and SRp40 on the utilization of the A1 to A5 splicing sites of HIV-1 RNA. *J Biol Chem* 279(29):29963–29973.
- Rossi F, Labourier E, Forné T, Divita G, Derancourt J, Riou JF, Antoine E, Cathala G, Brunel C, Tazi J. 1996. Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature* 381(6577):80–82.
- Rothrock CR, House AE, Lynch KW. 2005. HnRNP L represses exon splicing via a regulated exonic splicing silencer. *EMBO J* 24(15):2792–2802.
- Ruskin B, Zamore P, Green M. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* 52(2):207–219.
- Sampath J, Long PR, Shepard RL, Xia X, Devanarayan V, Sandusky GE, Perry WL, Dantzig AH, Williamson M, Rolfe M, Moore RE. 2003. Human SPF45, a splicing factor, has limited expression in normal tissues, is overexpressed in many tumors, and can confer a multidrug-resistant phenotype to cells. *The American journal of pathology* 163(5):1781–1790.
- Sanford JR, Bruzik JP. 2001. Regulation of SR protein localization during development. *Proc Natl Acad Sci USA* 98(18):10184–10189.
- Sanford JR, Wang X, Mort M, Vanduy N, Cooper DN, Mooney SD, Edenberg HJ, Liu Y. 2009. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res* 19(3):381–394.
- Saulière J, Sureau A, Expert-Bezançon A, Marie J. 2006. The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. *Mol Cell Biol* 26(23):8755–8769.
- Schaal TD, Maniatis T. 1999. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol Cell Biol* 19(3):1705–1719.
- Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Krämer A, Sattler M. 2003. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Molecular cell* 11(4):965–976.
- Shao C, Yang B, Wu T, Huang J, Tang P, Zhou Y, Zhou J, Qiu J, Jiang L, Li H, Chen G, Sun H, Zhang Y, Denise A, Zhang D-EE, Fu X-DD. 2014. Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. *Nature structural & molecular biology* 21(11):997–1005.
- Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 15(17):7155–7174.
- Sharma S, Falick AM, Black DL. 2005. Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol Cell* 19(4):485–496.

- Sickmier E, Frato KE, Shen H, Paranawithana SR, Green MR, Kielkopf CL. 2006. Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Molecular cell* 23(1):49–59.
- Sillars-Hardebol AH, Carvalho B, Beliën JAM, Wit M de, Delis-van Diemen PM, Tijssen M, Wiel MA van de, Pontén F, Meijer GA, Fijneman RJA. 2012. CSE1L, DIDO1 and RBM39 in colorectal adenoma to carcinoma progression. *Cell Oncol (Dordr)* 35(4):293–300.
- Sims RJ, Millhouse S, Chen C-F, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL, Reinberg D. 2007. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol Cell* 28(4):665–676.
- Smith C, Valcárcel J. 2000. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends in biochemical sciences* 25(8):381–388.
- Sridharan V, Singh R. 2007. A conditional role of U2AF in splicing of introns with unconventional polypyrimidine tracts. *Molecular and cellular biology* 27(20):7334–7344.
- Sugnet CW, Kent WJ, Ares M, Haussler D. 2004. Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac Symp Biocomput* 66–77.
- Tan Q, Yalamanchili HK, Park J, De Maio A, Lu H-CC, Wan Y-WW, White JJ, Bondar VV, Sayegh LS, Liu X, Gao Y, Sillitoe RV, Orr HT, Liu Z, Zoghbi HY. 2016. Extensive cryptic splicing upon loss of RBM17 and TDP43 in neurodegeneration models. *Human molecular genetics*. <https://doi.org/10.1093/hmg/ddw337>.
- Thickman KR, Swenson MC, Kabogo JM, Gryczynski Z, Kielkopf CL. 2006. Multiple U2AF65 binding sites within SF3b155: thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors. *Journal of molecular biology* 356(3):664–683.
- Turunen JJ, Niemelä EH, Verma B, Frilander MJ. 2013. The significant other: splicing by the minor spliceosome. *Wiley interdisciplinary reviews RNA* 4(1):61–76.
- Ule J, Ule A, Spencer J, Williams A, Hu J-SS, Cline M, Wang H, Clark T, Fraser C, Ruggiu M, Zeeberg BR, Kane D, Weinstein JN, Blume J, Darnell RB. 2005. Nova regulates brain-specific splicing to shape the synapse. *Nature genetics* 37(8):844–852.
- Valcárcel J, Gaur R, Singh R, Green M. 1996. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science (New York, NY)* 273(5282):1706–1709.
- Valcárcel J, Singh R, Zamore PD, Green MR. 1993. The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* 362(6416):171–175.
- Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* 136(4):701–718.
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456(7221):470–476.
- Wang W, Maucuer A, Gupta A, Manceau V, Thickman KR, Bauer WJ, Kennedy SD, Wedekind JE, Green MR, Kielkopf CL. 2013. Structure of phosphorylated SF1 bound to U2AF⁶⁵ in an essential splicing factor complex. *Structure (London, England : 1993)* 21(2):197–208.

- Wang Z, Kayikci M, Briese M, Zarnack K, Luscombe NM, Rot G, Zupan B, Curk T, Ule J. 2010. iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol* 8(10):e1000530.
- Warf MB, Berglund JA. 2010. Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci* 35(3):169–178.
- Wei W-JJ, Mu S-RR, Heiner M, Fu X, Cao L-JJ, Gong X-FF, Bindereif A, Hui J. 2012. YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts. *Nucleic acids research* 40(17):8622–8636.
- Will C, Schneider C, MacMillan A, Katopodis N, Neubauer G, Wilm M, Lührmann R, Query C. 2001. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *The EMBO journal* 20(16):4536–4546.
- Will CL, Lührmann R. 2011. Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* 3(7).
- Wollerton MC, Gooding C, Robinson F, Brown EC, Jackson RJ, Smith CW. 2001. Differential alternative splicing activity of isoforms of polypyrimidine tract binding protein (PTB). *RNA* 7(6):819–832.
- Wu T, Fu X-DD. 2015. Genomic functions of U2AF in constitutive and regulated splicing. *RNA biology* 12(5):479–485.
- Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, Shiosaka M, Kawahata R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Ishiyama K, Mori H, Nolte F, Hofmann W-KK, Miyawaki S, Sugano S, Haferlach C, Koeffler H, Shih L-YY, Haferlach T, Chiba S, Nakauchi H, Miyano S, Ogawa S. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478(7367):64–69.
- Zahler AM, Lane WS, Stolk JA, Roth MB. 1992. SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev* 6(5):837–847.
- Zamore P, Green M. 1991. Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *The EMBO journal* 10(1):207–214.
- Zamore PD, Green MR. 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci USA* 86(23):9243–9247.
- Zarnack K, König J, Tajnik M, Martincorena I, Eustermann S, Stévant I, Reyes A, Anders S, Luscombe NM, Ule J. 2013. Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* 152(3):453–466.
- Zhou X, Wu W, Li H, Cheng Y, Wei N, Zong J, Feng X, Xie Z, Chen D, Manley JL, Wang H, Feng Y. 2014. Transcriptome analysis of alternative splicing events regulated by SRSF10 reveals position-dependent splicing modulation. *Nucleic acids research* 42(6):4019–4030.
- Zhou Z, Fu X-D. 2013. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* 122(3):191–207.

Titre : Etude du rôle des facteurs d'épissage à domaines UHM dans la régulation de l'épissage alternatif

Mots clés : Epissage, Epissage Alternatif, UHM, ULM, Facteurs d'épissage, U2AF⁶⁵

Résumé : Les protéines U2AF⁶⁵, CAPER α , PUF60 et SPF45 sont des facteurs d'épissage qui possèdent des domaines similaires appelés UHM et qui interagissent pendant les étapes précoces de l'épissage avec des protéines qui possèdent des domaines ULM, comme SF3b155. Par des approches biochimiques, nous avons mis en évidence la formation d'assemblages macromoléculaires par U2AF⁶⁵ et CAPER α au contact du domaine multi-ULM de SF3b155. En diminuant les taux d'expression des facteurs d'épissage à domaines UHM avec des shRNA et en analysant par qPCR l'épissage de 65 exons cassette, nous avons identifié un rôle activateur de CAPER α , U2AF⁶⁵ et PUF60 et un rôle répresseur de SPF45 dans l'épissage. Plus particulièrement, CAPER α et U2AF⁶⁵ activent l'épissage des exons cassette présentant des séquences flanquantes en 5' riches en motifs polypyrimidine. De plus, ces séquences favorisent la formation des assemblages macromoléculaires de U2AF⁶⁵ et CAPER α . Appuyés par ces résultats, nous proposons un modèle dans lequel des interactions multivalentes conduisent à la formation d'assemblages macromoléculaires par CAPER α et U2AF⁶⁵; ces complexes ont une affinité particulière d'une part pour les séquences introniques riches en motifs polypyrimidine et d'autre part avec le domaine multi-ULM de SF3b155. L'ensemble de ces interactions favorise la reconnaissance des sites 3' d'épissage.

Title : Study of the role of UHM splicing factors in the regulation of alternative splicing

Keywords : Splicing, Alternative splicing, ULM, ULM, Splicing factors, U2AF⁶⁵

Abstract : U2AF⁶⁵, CAPER α , PUF60 and SPF45 are splicing factors that hold similar domains called UHM that interact during the early splicing steps with ULM domains proteins, such as SF3b155. Using biochemical approaches, we highlighted the formation of macromolecular assemblies by U2AF⁶⁵ and CAPER α in contact with the multi-ULM domain of SF3b155. The inhibition of the expression of the UHM splicing factors by shRNA, followed by a qPCR analysis of 65 cassette exons led us to identify an activating role of CAPER α , U2AF⁶⁵ and PUF60 and a repressing role of SPF45 in splicing. Particularly, CAPER α and U2AF⁶⁵ activate splicing of cassette exons presenting long pyrimidine-rich 5' flanking regions. Moreover, these regions favor the formation of macromolecular assemblies of U2AF⁶⁵ and CAPER α . On the basis of these results, we propose a model in which multivalent interactions lead to CAPER α and U2AF⁶⁵ macromolecular assemblies; these assemblies present a particular affinity on one hand for long pyrimidine-rich introns and on the other one for the multi-ULM domain of SF3b155. All these interactions promote 3' splice sites recognition.