



HAL
open science

Étude des dérégulations de l'épissage alternatif du pré-ARN messenger de la troponine T cardiaque humaine associées aux dystrophies myotoniques de types 1 et 2 et des caractéristiques du facteur d'épissage MBNL1 impliqué dans ces pathologies

Audrey Vautrin

► **To cite this version:**

Audrey Vautrin. Étude des dérégulations de l'épissage alternatif du pré-ARN messenger de la troponine T cardiaque humaine associées aux dystrophies myotoniques de types 1 et 2 et des caractéristiques du facteur d'épissage MBNL1 impliqué dans ces pathologies. Médecine humaine et pathologie. Université Henri Poincaré - Nancy 1, 2011. Français. NNT : 2011NAN10141 . tel-01746306

HAL Id: tel-01746306

<https://hal.univ-lorraine.fr/tel-01746306>

Submitted on 29 Mar 2018

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.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

➤ Contact SCD Nancy 1 : theses.sciences@scd.uhp-nancy.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Ecole Doctorale BioSE (Biologie-Santé-Environnement)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de
DOCTEUR DE L'UNIVERSITE HENRI POINCARE

Mention : « Sciences de la Vie et de la Santé »

par **Audrey VAUTRIN**

**Etude des dérégulations de l'épissage alternatif du pré-ARN messenger
de la troponine T cardiaque humaine associées aux dystrophies
myotoniques de types 1 et 2 et des caractéristiques du facteur
d'épissage MBNL1 impliqué dans ces pathologies**

Le 18 Novembre 2011

Membres du Jury :

Rapporteurs :	Mr J. Tazi	Professeur, Université de Montpellier 2
	Mme J. Marie	Directeur de Recherche, INSERM, Gif-sur-Yvette
Examineurs :	Mme M.L. Caillet-Boudin	Chargée de Recherche, INSERM, Lille
	Mr B. Charpentier	Professeur, Université Henry Poincaré, Nancy I
	Mme C. Branlant	Directeur de Recherche, CNRS, Nancy I Directeur de thèse
	Mme I. Behm-Ansmant	Chargée de Recherche, CNRS, Nancy I Co-directeur de thèse

J'adresse mes sincères remerciements à Madame Christiane Branlant, Directeur de l'UMR CNRS-UHP 7214 pour m'avoir accueilli dans son équipe et permis de réaliser ce travail dans d'agréables conditions. Je lui suis particulièrement reconnaissant pour les nombreuses discussions scientifiques que nous avons eues et les conseils qu'elle m'a donnés. Je la remercie aussi pour son aide précieuse dans la rédaction de ce mémoire.

Je tiens à exprimer ma profonde gratitude à Madame Isabelle Behm-Ansmant, ma co-directrice de thèse, pour m'avoir guidée et soutenue au cours de ce travail. Elle a su me communiquer son enthousiasme, sa rigueur et m'a transmis son expertise de la biologie moléculaire et cellulaire. Je la remercie également pour le temps qu'elle m'a accordé tout au long de ma thèse.

Je tiens également à remercier les Docteurs Jamal Tazi et Joëlle Marie pour avoir accepté d'évaluer le travail que j'ai effectué au cours de ma thèse. Je remercie aussi vivement le Docteur Marie-Laure Caillet-Boudin qui a accepté de juger mon travail et pour le travail de collaboration que nous avons réalisé ensemble, ainsi que le Docteur Bruno Charpentier.

Je remercie également l'AFM, la région Lorraine, le CNRS et l'Université Henri Poincaré pour avoir soutenu financièrement notre projet.

J'exprime toute ma reconnaissance à l'équipe de Nicolas Sergeant (Marie-Laure, Dana et Hélène) pour le travail de collaboration que nous avons réalisé ensemble. Je remercie Laetitia Fouillen et Sarah Sanglier du laboratoire d'Alain Van Dorsselaar pour l'énorme travail de spectrométrie de masse qu'elles ont effectué. Je remercie également Denis Furling et son équipe, pour m'avoir accueillie si gentiment dans son laboratoire et fait partager son expertise dans la technique de FISH.

Je remercie tous les membres du laboratoire qui ont participé à l'élaboration de ce travail, en particulier Nathalie Marmier-Gourrier qui a initié ce projet et m'a encadré durant mon DEA ; Florence Oillo pour sa précieuse et efficace aide pour les expériences de retard sur gel, Christelle Aigueperse pour son aide dans l'étude de l'épissage et enfin Jean-Vincent Philippe qui a pris ma relève dans ce merveilleux projet.

Je tiens à remercier tous les membres du consortium DM France, pour toutes ces réunions enrichissantes. Ca a toujours été un plaisir de vous retrouver. Je pense spécialement aux autres thésards: Hélène, FX, Dana, Fred, Fernande mais aussi Virginie et Camille pour les bons moments passés ensemble.

Un énorme MERCI à mes co-locataires de l'ex Box 2 Aileen, Georges, Lilia, Valentin, JV, Flo et Isa. Merci de m'avoir supporté et soutenu toutes ces années aussi bien au labo qu'en dehors. Je n'aurai pas pu « tomber » dans un meilleur Box.

Un autre énorme MERCI à Maryline et Anne, pour toutes ces discussions et fous rires que l'on a pu avoir, merci pour tous les bons moments passés ensemble, et enfin merci pour votre aide salutaire pour la fin de rédaction de ce manuscrit.

Un grand merci aussi à tous les membres de l'équipe Maturation des ARN! Je tiens notamment à remercier Véronique pour les discussions scientifiques, les petits jeunes du Labo (Antoine, Guillaume et Clémence), les zymos (Laure, Raph et Sab), Régis pour votre écoute et joie de vivre.

Enfin, je voudrais terminer en remerciant très affectueusement ma famille pour m'avoir encouragée et soutenue tout au long de ces années....

ABBREVIATIONS

2'O MePS	: 2'-O-Methyl Phosphothioate
A	: adénine
aa	: acide aminé
AAV	: Adeno-Associated Virus
A/mA	: ampère/milliampère
ADN	: acide désoxyribonucléique
AMPC	: adénosine monophosphate cyclique
AMV	: Avian Myeloblastosis Virus
ARN	: acide ribonucléique
ARNm	: ARN messenger
ARNt	: ARN de transfert
ATP	: adenosine 5' triphosphate
BER	: Base Excision Repair
BP	: Binding Protein
BPES	: syndrome du blépharophimosis-Ptois-epicanthus inversus
BSA	: Bovine Serum Albumine
BSE	: Bidirectional Splicing Enhancer
C	: cytosine
°C	: degré Celsius
CaMKII	: Calmoduline Kinase II
CBC	: Cap Binding Complex
CBP	: CREB-Binding Protein
CCD	: dysplasie cleido-cranienne
CCHS	: syndrome d'hypoventilation alvéolaire centrale congénitale
CELF	: CUG-BP and ETR-3 Like Factors
CGRP	: Calcitonin Gene Related Peptide
CIC-1	: canaux chlorique
CNBP	: Cellular Nucleic-acid-Binding Protein
CpG	: Cytosine-phosphate-Guanosine

cpm	: coups par minute
cps	: coups par seconde
CRM1	: Chromosome Region Maintenance 1
CT	: calcitonin
CTCF	: CCCTC-binding factor
CTD	: Carboxy Terminal Domain
CUG-BP	: CUG Binding Protein
D	: A, G ou U
DAPI	: 4',6'-diamidino-2-phénylindole
DM1	: Dystrophie Myotonique de type 1
DM2	: Dystrophie Myotonique de type 2
DMEM	: Dulbecco's Modified Eagle's Medium
DMPK	: Dystrophie Myotonique protéine kinase
DMWD	: Dystrophia Myotonic Gene with WD repeats
DNase	: désoxyribonucléase
dNTP	: désoxyribonucléotide
DO	: densité optique
DRPLA	: Dentatorubral-pallidoluysian atrophy
DSB	: double-strand break
DTT	: dithiothréitol
<i>E.coli</i>	: <i>Escherichia coli</i>
EDEN	: Embryo Deadenylation ElemeNt
EDTA	: éthylène diamine tétraacétate
eIF2	: eukaryotic Initiation Factor 2
ESE	: Exonic Splicing Enhancer
ESS	: Exonic Splicing Silencer
<i>et al.</i>	: <i>et alii</i>
ETR	: Embryonic lethal abnormal vision Type Ribonuclear protein
FEN	: flap endonucléase
FGFR	: Fibroblast Growth Factor Receptor
Fig	: figure
FMR	: Fragile X Mental Retardation
FMRP	: Fragile X Mental Retardation Protein

FRAXA	: Fragile X syndrome
FRDA	: Friedreich Ataxia
FSH	: hormone folliculo-stimulante
FXTAS	: Fragile X associated Tremor Ataxia Syndrome
G	: guanine
g/l	: gramme par litre
g/mg/µg/ng	: gramme/milligramme/microgramme/nanogramme
G/Gly	: glycine
GABA	: acide gamma-aminobutyrique
GRE	: GU-Rich Element
GST	: Glutathion-S-Transferase
h	: heure(s)
H	: histidine
H3	: histone H3
H3K36me3	: lysine en position 36 de l'histone H3 tri-méthylée
hcTNT	: human cardiac Troponin T
HD	: Huntington Disease
HDL2	: Huntington Disease Like 2
HEK	: Human Embryonic Kidney
hnRNP	: heterogeneous RiboNucleoProtein
HPE	: homopros-encéphalie de type 5
IGC	: Interchromatin Granules Cluster
IPTG	: Isopropyl β-D-1-thiogalactopyranoside
IR	: Insulin Receptor
ISE	: Intronic Splicing Enhancer
ISS	: Intronic Splicing Silencer
K	: lysine
Kd	: constante de dissociation
kDa	: kilodalton
KH	: K-homology
l/ml/µl	: litre/millilitre/microlitre
La	: Lupus antigen
LB	: Luria-Bertani
LBA	: Luria-Bertani Ampicilline

LBAC	: Luria-Bertani Ampicilline Chloramphénicol
LSm	: Sm-like
M	: A ou C
M/mM/μM/nM	: molaire/millimolaire/micromolaire/nanomolaire
MA	: maladie d'Alzheimer
MAPT	: Microtubule Associated Protein Tau
Mbl	: Muscleblind
MBNL	: Muscleblind-Like
MBP	: Maltose Binding Protein
MEF2A	: Myocyte Enhancer Factor 2A
MEP	: maladie par expansion de polyglutamine
min	: minute
MMR	: Mismatch Repair
mol/mmol/nmol/pmol/fmol	: mole/millimole/nanomole/picomole/femtomole
MBP	: Maltose-Binding Protein
MSE	: Muscle Specific Enhancer
MTMR1	: Myotubularine-related protein 1
MYPT1	: Myosin Phosphatase Target subunit 1
N	: nucléotide A,C, G,T ou U
NCAM	: Neural Cell Adhesion Molecule
NER	: Nucleotide Excision Repair
NES	: nuclear export signal
NLS	: nuclear localization signal
NMDA	: acide N-méthyl-D-aspartique
NMHC-B	: Nonmuscle Myosin Heavy Chain II-B
NPC	: Nuclear Pore Complex
NPM	: nucleophosmin
NR	: non retenu
nt	: nucléotide
OPMD	: dystrophie musculaire oculo-pharyngée
ORF	: cadre de lecture ouvert
³² P	: phosphore 32
PABP1	: Poly(A) Binding Protein 1
PABPN1	: Poly(A) Binding Protein Nuclear 1

PAGE	: PolyAcrylamide Gel Electrophoresis
PARN	: Poly(A) RiboNuclease
pb	: paire de base
PBS	: Phosphate Buffered Saline
PCBP1	: Poly(rC) Binding Protein 1
PCR	: Polymérisation Chain Reaction
PFA	: paraformaldéhyde
pFAK	: phospho Focal Adhesion Kinase
PKC	: Protein Kinase C
PKR	: Protein Kinase R
PM	: poids moléculaire
PMO	: Phosphorodiamidate Morpholino Oligo
PNK	: PolyNucléotide Kinase
Pol	: polymérase
polyA	: polyadénosine
polyQ	: polyglutamine
PROMM	: PROximal Myotonic Myopathy
PSF	: PTB associated Splicing Factor
PTB	: Polypyrimidine Tract Binding
p-TEFb	: positive Transcription Elongation Factor b
p/v	: poids/volume
qsp	: quantité suffisante pour
R	: purine
RISC	: RNA-Induced Silencing Complex
RMN	: résonance magnétique nucléaire
RNase	: ribonucléase
RNasin	: RNase inhibitor
RNP	: ribonucleoprotein
rNTP	: ribonucléotide
rpm	: rotation par minute
RRM	: RNA recognition motif
RS	: arginin-serin rich domain
RT	: rétro-transcription
RYR	: RYanodin Receptor

s	: seconde(s)
S	: G ou C
SBMA	: amyotrophie spinobulbaire
SCA	: Spino Cerebellar Ataxia
SDS	: sodium dodécyl sulfate
SDSA	: Synthesis Dependant Strand Annealing
SELEX	: Systematic Evolution of Ligands by EXponential enrichment
SERCA	: Sarco/Endoplasmic Reticulum Calcium ATPase
siRNA	: small interfering RNA
shRNA	: short hairpin RNA
SIX5	: SIne oculis homeoboX homolog 5
SKIP	: c-Ski Interacting Protein
SMN	: Survival of Moto Neuron
snRNA	: small nuclear RNA
snRNP	: small nuclear RNP
SPD	: synpolydactylie
SR	: Serin Arginin rich
SRPK	: SR protein kinase
SRrp	: SR protein-related protein
SRSF	: Serin/Arginin-rich Splicing Factor
SWI/SNF	: SWItch/Sucrose NonFermentable
T	: thymine
TBE	: Tris-Borate-EDTA
TCR	: Transcription Coupled Repair
TE	: Tris-EDTA
TF	: Transcription Factor
TNNT3	: troponine T squelettique
U	: uracile
U2AF	: U2 auxiliary factor
UE	: unité enzymatique
UsnRNA	: U small nuclear RNA
UsnRNP	: U small nuclear RNP
UTP	: uridine 5' triphosphate

UTR	: untranslated region
UV	: Ultra-Violet
v/v	: volume/volume
W	: A ou U
W-C	: Watson-Crick
WPRE	: Woodchuck Post-transcriptional Regulatory Element
WT	: Wild-Type
XH	: hypopituitarisme lié à l'X
XLMR	: retard mental lié à l'X
Y	: pyrimidine
Z	: domaine doigt de zinc

Liste des figures**Introduction**

Figure 1 : Localisation dans les gènes des répétitions de séquences impliquées dans des maladies génétiques	p 3
Figure 2 : Mécanismes moléculaires engendrés par les expansions de répétitions	p 4
Figure 3 : Structuration des répétitions ADN	p 10
Figure 4 : Modèles d'expansion des répétitions ADN basés sur des défauts de réplication	p 12
Figure 5 : Modèle d'expansion des répétitions ADN basé sur des événements au cours de la réparation	p 13
Figure 6 : Modèle d'expansion des répétitions ADN basé sur des événements au cours de la recombinaison	p 15
Figure 7 : Structure des régions ARN renfermant des répétitions CUG	p 17
Figure 8 : Les étapes de la réaction d'épissage des introns des pré-ARN messagers nucléaires	p 21
Figure 9 : Séquences essentielles à la réaction d'épissage dans les introns majeurs des pré-ARNm de vertébrés	p 22
Figure 10 : Séquences et structures secondaires des UsnRNA humains majeurs	p 24
Figure 11 : Représentation schématique de la biogénèse des UsnRNP chez les eucaryotes supérieurs	p 26
Figure 12 : L'assemblage séquentiel du spliceosome autour des introns majeurs	p 29
Figure 13 : Les différents modes d'épissage alternatif	p 31
Figure 14 : Organisation structurale des protéines SR	p 36
Figure 15 : Organisation structurale des protéines SR-like ou SRrp (pour « SR protein -related protein »)	p 39
Figure 16 : Rôles des protéines SR dans la sélection des sites d'épissage constitutifs	p 40
Figure 17 : Cas de l'ESE bi-directionnelle du pré-ARNm de l'unité de transcription E1A	p 41
Figure 18 : Rôles des protéines hnRNP dans l'épissage	p 48
Figure 19 : Influence de la vitesse d'élongation sur l'épissage alternatif	p 53
Figure 20 : Couplage entre les différentes machineries et l'épissage	p 54
Figure 21 : Système d'adaptateur entre la chromatine et l'épissage	p 57
Figure 22 : Localisation génomique des expansions de nucléotides responsables des dystrophies myotoniques de type 1 et 2	p 59
Figure 23 : Complexité des répétitions au niveau de l'intron 1 du gène <i>ZNF9</i>	p 64
Figure 24 : Comparaison des symptômes observés chez les patients atteints de DM1 ou DM2	p 65
Figure 25 : Schéma récapitulatif des mécanismes moléculaires impliqués dans la physiopathologie des dystrophies myotoniques	p 66
Figure 26 : Visualisation des <i>foci</i> d'ARN dans le noyau des cellules DM1 par hybridation <i>in situ</i>	p 67
Figure 27 : Visualisation des <i>foci</i> d'ARN dans le noyau des cellules DM2 par hybridation <i>in situ</i>	p 68
Figure 28 : Positionnement des sites de fixation de la protéine CTCF au sein du locus DM1	p 71
Figure 29 : Co-localisation des agrégats nucléaires formés par des expansions CUG et CCUG avec MBNL1	p 72
Figure 30 : Mécanisme physiopathologique de laDM1 médié par les protéines MBNL1 et CUG-BP1	p 74

Figure 31 : Diversité des protéines MBNL1 générées par épissage alternatif	p 81
Figure 32 : Alignement des 4 doigts à zinc des 3 protéines MBNL humaines	p 83
Figure 33 : Structure cristallographique de la paire à doigts de Zn ²⁺ 3 et 4 (TZF34) de MBNL1 en complexe avec un oligonucléotide r(CGCUGU)	p 84
Figure 34 : Structuration des sites de fixation de MBNL1	p 87
Figure 35 : Similarité d'organisation structurale des protéines CELF	p 92
Figure 36 : Régulation de la dégradation et déadénylation des ARNm par CUG-BP1	p 95
Figure 37 : Absence de co-localisation entre la protéine CUG-BP1 et les <i>foci</i> CUG	p 96
Figure 38 : Régulation de l'épissage alternatif de l'exon 5 de la Troponine T cardiaque (cTNT) du poulet	p 100
Figure 39 : Mécanisme de régulation de l'épissage de l'exon 5 du pré-ARNm de la troponine T cardiaque chez l'homme	p 101
Figure 40 : Les différentes cibles testées pour développer des thérapies afin de lutter contre la DM	p 102

Matériels et méthodes

Figure 41 : Protocole de SELEX	p 121
---------------------------------------	-------

Résultats

Figure 42 : Représentation schématique des différentes formes recombinantes de MBNL1 utilisées au cours de mon travail	p 137
Figure 43 : Production et purification des protéines MBNL1 _{ΔCT3} et MBNL1 _{ΔCT}	p 138
Figure 44 : Contrôle de la qualité des échantillons protéiques par expérience de retard sur gel	p 140
Figure 45 : Représentation schématique des liaisons phosphodiester coupées par l'action des sondes enzymatiques au sein des ARN	p 143
Figure 46 : Les ARN (CUG) ₁₆ et (CUG) ₁₇ forment de longues structures en tige-boucle	p 144
Figure 47 : Seule la protéine MBNL1 se fixe directement sur les ARN (CUG) ₁₆ et (CUG) ₁₇	p 146
Figure 48 : Structure secondaire de l'ARN E18	p 151
Figure 49 : Analyse de la protection contre l'action des RNases générée sur l'ARN AC3 par fixation de la protéine MBNL1 _{ΔCT3}	p 152
Figure 50 : Structure de l'ARN E7 de la famille II	p 153
Figure 51 : Structure de l'ARN E14 de famille II	p 154
Figure 52 : Structure de l'ARN E31 de famille II	p 155
Figure 53 : Principe de la méthode triple-hybride	p 156
Figure 54 : Les isoformes naturelles de MBNL1 présentent une forte affinité pour les ARN SELEX de la famille 2	p 157
Figure 55 : Les ARN SELEX de la famille II sont fonctionnels <i>in cellulo</i>	p 159
Figure 56 : Représentation schématique des formes recombinantes de MBNL1, MBNL1 _{Zn12} , MBNL1 _{Zn34} et MBNL1 _{Δ122}	p 163
Figure 57 : Les motifs TZF12 et TZF34 n'interagissent pas avec les ARN cibles de MBNL	p 165
Figure 58 : Identification de 8 sites YGCU dans la région 150-E5-150 du pré-ARNm hcTNT	p169
Figure 59 : La protéine MBNL1 _{ΔCT3} interagit avec les 3 groupes de sites identifiés	p170
Figure 60 : La mutation des résidus G en résidus A des motifs YGCU inhibe la fixation de MBNL1 sur les groupes de sites	p172

- Figure 61** : Chevauchement des sites de fixation de MBNL1 et CUG-BP1 au niveau de l'ARN 3'-hcTNT₀₋₅₀ p173
- Figure 62** : Représentation schématique de la forme recombinante de CUG-BP1 utilisée au cours de mon travail : CUG-BP1_{ΔCT} p174
- Figure 63** : Contrôle de la fixation de la protéine CUG-BP1_{ΔCT} sur l'ARN 3'-hcTNT₀₋₅₀ et l'ARN 3'-hcTNT_{0-50mut4-5-6} p175
- Figure 64** : Absence de compétition entre les protéines MBNL1_{ΔCT3} et CUG-BP1_{ΔCT} pour la fixation sur l'ARN 3'-hcTNT₀₋₅₀ p176
- Figure 65** : Détermination de la taille des régions introniques en amont et en aval de l'exon 5 du pré-ARNm hcTNT requises pour récapituler l'ensemble des régulations de l'inclusion de l'exon 5 p178
- Figure 66** : Schéma représentant la localisation de chaque oligonucléotide amorce utilisé p179
- Figure 67** : Analyse des coupures enzymatiques de la région du pré-ARNm hcTNT requise pour récapituler l'ensemble des régulations et dérégulations de l'inclusion de l'exon 5 p181
- Figure 68** : Analyse de la structure 2D de la région du pré-ARNm hcTNT qui est requise pour récapituler l'ensemble des régulations et dérégulations de l'inclusion de l'exon 5 p183
- Figure 69** : Analyse des protections générées par la fixation de MBNL1_{ΔCT3} sur l'ARNm hcTNT 150-E5-150 p185
- Figure 70** : Protection contre les RNases des résidus G des motifs YGCU de l'ARN 150-E5-150 par la protéine MBNL1_{ΔCT3} p186
- Figure 71** : Analyse par RT-PCR de l'effet des mutations des groupes de sites sur l'inclusion de l'exon 5 dans les ARN produits à partir du minigène IR-hcTNT₁₅₀₋₁₅₀ p188
- Figure 72** : Rôle des différentes régions du minigène IR-hcTNT dans la régulation de l'inclusion de l'exon 5 du pré-ARNm hcTNT p189
- Figure 73** : Analyse par RT-PCR de l'effet de la surexpression de différentes protéines SR sur l'inclusion de l'exon 5 dans les transcrits d'épissage produits à partir du minigène IR-hcTNT₁₅₀₋₁₅₀ p190
- Figure 74** : Analyse par retard sur gel des complexes formés sur l'ARN 150-E5-150 en présence de concentrations croissantes des protéines SRSF1,2,5 et 7 p191
- Figure 75** : La mutation des sites de fixation de MBNL1 augmente la taille de la séquence polypyrimidine p192
- Figure 76** : Modèle de localisation des séquences impliquées dans la régulation de l'inclusion de l'exon 5 et des séquences nécessaires à l'action en *trans* des répétitions CUG p194
- Figure 77** : Principe de la purification des complexes ribonucléoprotéiques par chromatographie d'affinité basée sur l'emploi de la protéine de l'enveloppe du phage MS2 et de la protéine MBP p197
- Figure 78** : Constructions réalisées pour les expériences de purification par chromatographie d'affinité MS2/MBP des RNP formées sur les répétitions CUG et CAG après incubation en présence d'extrait nucléaire de cellules HeLa p198
- Figure 79** : L'ARN (CUG)₅₁ se structure en une longue tige-boucle p200
- Figure 80** : Analyse par retard sur gel des complexes formés sur les ARN CUG₅₁-MS2 (A), MS2 (B) et CUG₅₁ (C) en présence de concentrations croissantes de protéine MS2/MBP p202
- Figure 81** : Purification des protéines contenues dans les complexes ribonucléoprotéiques formés sur les ARN MS2 et CUG₅₁-MS2 dans un extrait nucléaire de cellules HeLa p203

- Figure 82 :** Analyse par électrophorèse en gel SDS-PAGE 10 % des protéines contenues dans les RNP formées sur les ARN (CUG)₅₁-MS2, (CAG)₅₉-MS2 et MS2, purifiées par chromatographie d'affinité MS2/MBP réalisée en conditions « normales » ou « stringentes » **p212**
- Figure 83 :** Un nouveau partenaire potentiel des ARN contenant des répétitions CUG : la nucléophosmine **p217**
- Figure 84 :** La protéine NPM se fixe spécifiquement sur les répétitions CUG et CAG **p219**
- Figure 85 :** La nucléophosmine peut entrer en compétition avec MBNL1 pour la fixation sur les répétitions (CUG)₅₁ **p221**
- Figure 86 :** MBNL1 n'est pas co-immunoprécipitée avec NPM **p222**
- Figure 87 :** Une diminution de la quantité de la protéine NPM est observée dans les lysats cellulaires préparés à partir de myoblastes de patients atteints de DM1 **p223**
- Figure 88 :** Etude de localisation des protéines NPM et NPM-T199 au sein de myoblastes de patients sains ou de patients atteints de DM1 **p224**
- Figure 89 :** Co-localisation de MBNL1 et NPM-T199 au sein de myoblastes de patients atteints de DM1 **p225**
- Figure 90 :** Co-localisation de NPM-T199 et des *foci* nucléaires au sein de myoblastes de patients atteints de DM1 **p226**

Liste des tableaux

Introduction

Tableau I : Dynamique de la composition des complexes A, B et C du spliceosome chez l'Homme	p 28
Tableau II : Les séquences ARN reconnues par les protéines SR	p 37
Tableau III : Les protéines hnRNP : structures, rôles et séquences ARN reconnues	p 44
Tableau IV : Protéines candidates au lien entre transcription et épissage	p 52
Tableau V : Principaux symptômes des formes adultes de la dystrophie myotonique de type 1	p 61
Tableau VI : Principales anomalies d'épissage dans différents tissus de patients DM	p 74
Tableau VII : Pourcentage d'identité de séquence entre les 6 différentes protéines de la famille CELF	p 91

Matériels et méthodes

Tableau VIII : Constructions pSVIRB-hcTNT utilisées au cours de ma thèse	p 112
Tableau IX : Oligonucléotides utilisés pour la construction des différentes protéines recombinantes	p 114
Tableau X : Anticorps utilisés au cours de ma thèse	p 115
Tableau XI : Oligonucléotides utilisés pour les expériences d'épissage <i>in vitro</i> et <i>in cellulo</i>	p 134

Résultats

Tableau XII : Résultats d'identification par spectrométrie de masse des protéines contenues spécifiquement dans les RNP formées sur les répétitions (CUG) ₅₁	p205
Tableau XIII : Protéines identifiées majoritairement présentes dans les complexes formés sur les répétitions (CUG) ₅₁	p209
Tableau XIV : Protéines contenues dans les RNP formées sur les ARN (CUG) ₅₁ -MS2, (CAG) ₅₉ -MS2 ou MS2	p213

Remerciements	
Abréviations	
Liste des Figures et Tableaux	
Sommaire	
Avant-propos	1
<i>Introduction</i>	2
Chapitre I : Généralités sur les maladies génétiques liées à des amplifications de séquences répétées	2
I. Les différents types d'amplifications de courtes séquences dans le génome humain à l'origine de maladies génétiques	2
II. Les mécanismes moléculaires à l'origine des pathologies	3
II.1. Perte de fonction des protéines codées par les gènes mutés	5
II.2. Gain de fonction des protéines codées par les gènes mutés	6
II.3. Gain de fonction des ARN entraînant des dérégulations de l'épissage alternatif	7
III. Mécanismes proposés pour expliquer l'expansion de répétitions de séquence	9
III.1. Structuration particulière des répétitions ADN	9
III.2. Défauts de réplication	11
III.3. Evénements de réparation	12
III.4. Amplifications résultant d'événements de recombinaison	14
III.6. Modulation de l'instabilité des répétitions par éléments <i>cis</i> et changements épigénétiques	16
IV. Les structures ARN particulières générées par des répétitions de triplets	16
Chapitre II : L'épissage	19
I. Description générale de l'épissage des ARN	20
I.1. La réaction chimique d'épissage	20
I.2. Les séquences définissant les introns	21
I.2.a. Le site 5' donneur d'épissage	22
I.2.b. Le site 3' donneur d'épissage	22
I.3. Les composants et l'assemblage du spliceosome	23
I.3.a. Biogénèse des UsnRNP	23
I.3.b. Protéines associées aux complexes spliceosomaux	27
I.3.c. Assemblage du spliceosome et catalyse de la réaction	29
II. L'épissage alternatif	30
II.1. Les différents types d'épissage alternatif	30
II.1.a. Les exons dits « cassettes »	31
II.1.b. Les exons dits mutuellement exclusifs	32
II.1.c. Les exons variables du fait de la présence de sites d'épissage 5' et 3' alternatifs	32
II.1.d. Les exons résultant d'une rétention d'intron	32

II.1.e. Les exons terminaux alternatifs du fait de la présence de promoteurs ou de sites de polyadénylation alternatifs	33
II.2. Régulation de l'épissage alternatif	33
II.2.a. Les sites d'épissage alternatifs sont souvent des sites sous-optimaux	33
II.2.b. Les éléments cis régulateurs	34
II.2.c. Autres paramètres influençant les régulations de l'épissage alternatif	34
III. Régulateurs de l'épissage constitutif et alternatif	35
III.1. Protéines nucléaires jouant un rôle majeur dans la régulation de l'épissage constitutif et <i>alternatif</i>	35
III.1.a. Les protéines SR	35
III.1.b. Les protéines hnRNP	42
III.2. Autres protéines et paramètres intervenant dans la régulation de l'épissage.	50
III.2.a. La famille de protéines Fox	50
III.2.b. La famille de protéines Nova	50
IV. Lien entre transcription et épissage	51
IV.1. Lien entre l'ARN Pol II et le processus d'épissage	51
IV.2. Rôle du promoteur et de la vitesse d'élongation de l'ARN Pol II sur le processus d'épissage	52
IV.3. Influence de l'épissage sur la transcription	53
IV.4. Régulation épigénétique de l'épissage alternatif	54
IV.4.a. Rôle de la structure de la chromatine	55
IV.4.b. Implication des modifications épigénétiques des histones	56
Chapitre III : Les dystrophies myotoniques	58
I. Historique	58
II. Les Dystrophies myotoniques	60
II.1. DM1	60
II.1.a. Aspects génétiques	60
II.1.b. Les phénomènes de mosaïcisme et d'anticipation du nombre de répétitions	60
II.1.c. Les aspects cliniques	61
II.2. DM2	63
II.2.a. Aspects génétiques	63
II.2.b. Un phénomène d'anticipation ?	64
II.2.c. Les aspects cliniques	64
III. Bases moléculaires des DM1 et DM2	65
III.1. Mécanismes de pathogénicité des répétitions CUG et CCUG	65
III.1.a. Conséquences de la perte d'expression des protéines	67
III.1.b. Conséquences de la présence de répétitions sur l'expression des gènes voisins	69
III.1.c. Gain de fonction <i>trans</i> -dominant du transcrit muté	71
III.2. Lien entre dérégulation de l'épissage et symptômes	74
III.3. Autres mécanismes de pathogénicité observés	76

III.3.a. Dérégulation des micro-ARNs	76
III.3.b. Expression d'un ARN antisens	77
III.3.c. Dérégulation de la transcription	78
III.3.d. Dérégulation de la traduction	78
III.3.e. Mécanisme d'ARN interférence	79
IV. Les protéines de la famille Muscleblind	79
IV.1. Diversité des protéines Muscleblind	79
IV.1.a. Les protéines Mbl	80
IV.1.b. Les protéines MBNL humaines	81
IV.2. La protéine MBNL1 : relation structure-fonction.	82
IV.2.a. Architecture générale de la protéine MBNL1	82
IV.2.b. Le domaine N-terminal de MBNL1	82
IV.2.c. Le domaine C-terminal de MBNL1	85
IV.3. Les motifs ARN reconnus par MBNL1	86
IV.4. Fonctions des protéines MBNL	87
IV.4.a. Régulateur de l'épissage alternatif	87
IV.4.b. Les autres fonctions des protéines MBNL	88
IV.5. Implication de MBNL dans la DM	89
V. Les protéines de la famille CELF	90
V.1. Généralités	90
V.2. Structure des protéines	92
V.3. Les différents rôles des protéines CELF	93
V.3.a. Dans l'épissage alternatif	93
V.3.b. Dans la régulation post-transcriptionnelle	94
V.4. Implication des protéines CELF dans la DM	95
VI. Exemple de la dérégulation de l'épissage chez les patients atteints de DM1 : l'exon 5 de la troponine T cardiaque (cTNT)	97
VI.1. Rôle de la troponine T cardiaque (cTNT) dans la contraction musculaire	97
VI.2. Généralités	98
VI.3. Régulation et dérégulation de l'épissage alternatif de l'exon 5 de cTNT dans la DM	98
VII. Les différentes cibles et approches possibles pour une thérapie globale	102
VII.1. Elimination des transcrits mutants	103
VII.1.a. Par un ribozyme à tête de marteau	103
VII.1.b. Par ARN interférence	103
VII.1.c. Par oligonucléotide ou ARN anti-sens	104
VII.1.d. Par export des transcrits mutants	104
VII.2. Empêcher l'interaction entre la protéine MBNL1 et les expansions de répétitions CUG ou CCUG	104
VII.2.a. Morpholino-oligonucléotides	105
VII.2.b. Par cribles de petites molécules	105
VII.2.c. Par polymérisation de molécules	106

VII.3. Correction des défauts d'épissage	107
VII.3.a. Correction de l'épissage par utilisation de morpholino-oligonucléotides	107
VII.3.b. Augmenter l'activité de MBNL1	107
VII.3.c. Diminuer l'activité de CUG-BP1	108
Objectifs	109
Matériel et Méthodes	112
I Matériels	112
I. Souches d'E. coli utilisées	112
II. Lignées de cellules eucaryotes	112
III. Plasmides utilisés	112
III.1. Plasmides recombinants utilisés pour la production in vitro d'ARN	112
III.2. Vecteurs d'expression eucaryotes	113
III.3. Vecteurs utilisés pour la production des protéines recombinantes	114
IV. Les anticorps	115
V. Milieux de culture, tampons et solutions d'usage courant	115
II. Méthodes	116
<hr/>	
Chapitre de Livre : Alternative pre-mRNA splicing Analysis of Site-Specific RNA-Protein Interactions N. Marmier-Gourrier, A. Vautrin, C. Branlant and I. Behm-Ansmant	
<hr/>	
I. Autres techniques d'analyse des interactions ARN/protéines	116
I.1. Analyses de complexes [ARN/protéine] par gel-retard	116
I.2. Prédiction de structure secondaire par ordinateur	119
II. SELEX (systematic evolution of ligands by exponential enrichment)	120
III. Production et purification de protéines recombinantes	124
III.1. Production et purification des différentes isoformes des protéines MBNL, CUG-BP1 et NPM	124
III.2. Production et purification de la protéine de fusion MBP-MS2	126
IV. Purification des complexes ribonucléoprotéiques par chromatographie d'affinité avec la technique MS2	127
V. Méthodes utilisées pour la culture cellulaire	129
V.1 Culture cellulaire	129
V.1.a Lignée HEK293	129
V.1.b. Lignées cellulaires de myoblastes	129
V.1.c Technique de transfection	129
V.2. Technique d'immunofluorescence	130
V.2.a. Culture sur lames, fixation et perméabilisation des cellules	130
V.2.b. RNA-FISH et immunodétection de protéines	130
V.2.c. Montage des lames et acquisition des images	131

V.3. Préparation d'extrait nucléaire	131
V.4. Analyse d'extrait nucléaire par Western Blot	131
VI. Techniques utilisées pour analyser l'épissage	132
VI.1. Méthodes spécifiques de l'étude de l'épissage <i>in vitro</i>	132
VI.1.a Préparation de la matrice de transcription	132
VI.1.b Transcription <i>in vitro</i>	132
VI.1.c Epissage <i>in vitro</i>	133
VI.1.d Conversion des ARN en ADN complémentaires par RT-PCR	133
VI.2. Méthodes spécifiques de l'étude de l'épissage <i>in vivo</i>	134
VI.2.a Extraction des ARN totaux	134
VI.2.b. RT-PCR	135
Résultats et Discussion	136
Chapitre I : Etude des propriétés de fixation à l'ARN de MBNL1	136
I. Matériel utilisé	136
I.1. Production de protéines MBNL1 recombinantes	136
I.2. Contrôle de la qualité des échantillons protéiques par expérience de retard sur gel	138
I.3. Production des ARN (CUG) ₁₆ et (CUG) ₁₇	141
II. Etude de la fixation de MBNL1 sur des répétitions CUG	141
II.1. Etude de la structure secondaire des ARN (CUG) ₁₆ et (CUG) ₁₇	142
II.2. Etude de la fixation des protéines MBNL1 et CUG-BP1 sur les ARN (CUG) ₁₆ et (CUG) ₁₇	145
III. Recherche des séquences ARN ayant une forte affinité pour la protéine MBNL1 par la méthode SELEX	147
III.1. Présentation de l'article n°1	147
<hr/>	
A peculiar stem-loop structure with strong affinity for the MBNL1 splicing factor A. Vautrin, C. Aigueperse, I. Behm-Ansmant and C. Branlant Manuscrit en préparation	
<hr/>	
III.2. Résultats complémentaires	151
III.2.a. Figures supplémentaires concernant les analyses structurales	151
III.2.b. Etude par triple-hybride de l'affinité des différentes isoformes de MBNL1 pour les ARN sélectionnés	155
III.2.c. Etude de la fonctionnalité des séquences SELEX <i>in cellulo</i>	157
IV. Rôle des différents domaines de MBNL1 dans la reconnaissance des ARN	160
IV.1. Présentation de l'article n°2	160

Analysis of exonic regions involved in nuclear localization, splicing activity, and dimerization of Muscleblind-like-1 isoforms

H. Tran, N. Gourrier, C. Lemercier-Neuillet, C-M. Dhaenens, A. Vautrin, *et al.*

The Journal of Biological Chemistry, Vol. 286, no. 18, pp. 16435-16446, May 6, 2011

IV.2. Rôle des paires de doigts à zinc dans la reconnaissance des ARN	163
V. Conclusions et perspectives	166
Chapitre II : Etude des mécanismes régulant le taux d'inclusion de l'exon 5 du pré-ARNm de la troponine T cardiaque humaine.	168
I. Mise en évidence par examen des séquences de nombreux sites potentiels de fixation de MBNL1 situés en amont et en aval de l'exon 5 du pré-ARNm de la Troponine T cardiaque humaine.	168
I.1. Test de fixation de MBNL1 sur ses différents sites potentiels de fixation	169
I.2. Abolition de la fixation de MBNL1 suite à la mutation de ces sites de fixation	171
II. Test d'une compétition éventuelle entre CUG-BP1 et MBNL1 pour leur fixation aux séquences introniques bordant l'exon 5	173
II.2. Test de fixation des protéines par expériences de retard sur gel	174
II.3. Expérience de compétition par pontage covalent aux UV entre la protéine MBNL1 et CUG-BP1	175
III. Caractérisation des séquences nucléotidiques requises pour la régulation de l'inclusion de l'exon 5 dans l'ARNm hcTNT	176
II.1. Identification des séquences nucléotidiques requises pour la régulation de l'inclusion de l'exon 5 dans l'ARNm hcTNT par des études <i>in cellulo</i>	176
II.2. Analyse de la structure 2D de la région du pré-ARNm hcTNT requise pour la régulation de l'inclusion de l'exon 5 et test de la fixation de MBNL1 par expériences	179
II.2.a. Analyse de la structure 2D	179
II.2.b. Protection de l'ARN 150-E5-150 par la protéine MBNL1 _{ΔCT3}	184
IV. Test de l'importance des différents groupes de sites de fixation de MBNL1 mis en évidence dans la régulation de l'inclusion de l'exon 5 <i>in cellulo</i>	186
V. Recherche des autres facteurs d'épissage qui pourraient être impliqués dans la régulation de l'épissage alternatif de l'exon 5 du pré-ARNm hcTNT	188
VI. Conclusions et perspectives	191
Chapitre III : Recherche de nouvelles protéines associées aux répétitions CUG : identification de la nucléophosmine	196
I. Recherche de nouvelles protéines associées aux répétitions CUG	196
I.1. Principe de la méthode de chromatographie d'affinité basée sur le système MS2/MBP	196
I.2. Constructions utilisées	198
I.3. Etude de la structuration de l'ARN CUG ₅₁	199

I.4. Mise au point des conditions de purification des RNP	200
I.5. Purification des RNP formées avec l'ARN (CUG) ₅₁ -MS2 dans un extrait nucléaire de cellules HeLa	202
I.7. Amélioration de la technique de purification de complexes	210
II. Identification d'une nouvelle protéine capable de s'associer à un ARN contenant des répétitions CUG : la nucléophosmine.	215
II.1. Présentation de la nucléophosmine	215
II.2. La nucléophosmine interagit directement avec l'ARN (CUG) ₅₁	218
II.3. La nucléophosmine pourrait entrer en compétition avec la fixation de MBNL1 sur les répétitions (CUG) ₅₁	220
II.4. La protéine NPM n'interagit pas avec la protéine MBNL1	221
II.5. La quantité de NPM libre dans le noyau diminue dans les cellules DM1	223
II.6. Etude par immunofluorescence de la localisation cellulaire de la protéine NPM dans des myoblastes de patients atteints de DM1	224
III. Conclusion et perspectives	226
 <i>Conclusion Générale et Perspectives</i>	 234

Avant-propos

L'épissage des pré-ARNm est une étape clé de la régulation post-transcriptionnelle de l'expression des gènes eucaryotes. L'épissage alternatif correspond à la sélection de différentes combinaisons de sites 5' et 3' d'épissage. Il permet l'expression de multiples ARNm et donc de multiples variants protéiques à partir d'un gène unique et contribue de ce fait grandement à la diversité du protéome. La régulation de l'épissage alternatif joue un rôle majeur dans la différenciation cellulaire et dans le développement. Elle dépend de l'expression et de l'activité de nombreux facteurs régulateurs de l'épissage qui présentent des différences d'expression au cours du développement et en fonction du type cellulaire. Des dérégulations de l'épissage alternatif sont impliquées dans de nombreuses maladies génétiques et peuvent, comme dans le cas des maladies liées à des amplifications de répétitions telles que les dystrophies myotoniques, être à l'origine de la majorité des symptômes cliniques observés chez les patients.

Les dystrophies myotoniques (DM) sont les formes les plus courantes de dystrophies musculaires observées chez l'adulte. Ces maladies sont caractérisées par divers symptômes dont des défauts de la conduction et du rythme cardiaques qui sont une cause fréquente de décès des patients. Les DM sont dues à l'expression d'ARN mutés contenant des expansions de répétitions des nucléotides CUG (DM1) ou CCUG (DM2). Ces expansions de répétitions séquestrent le facteur d'épissage MBNL1 au sein de *foci* nucléaires et stabilisent le facteur d'épissage CUG-BP1, qui est un membre de la famille CELF et un antagoniste de MBNL1. Il est largement admis que la majorité des défauts d'épissage observés chez les patients DM résulte des altérations des facteurs régulateurs d'épissage des familles CELF et MBNL.

Aucun remède n'existe à ce jour pour la DM et les traitements utilisés se limitent à atténuer les signes et les symptômes de la maladie. Le but ultime des approches thérapeutiques développées actuellement est l'inversion du phénotype de la maladie, notamment en empêchant la séquestration de MBNL1 par les répétitions CUG, sans affecter la régulation de l'épissage des pré-ARNm cibles de MBNL1. Lorsque j'ai débuté ma thèse, peu de données étaient disponibles sur les modalités de fixation à l'ARN de MBNL1. La compréhension des bases moléculaires à l'origine de l'interaction de MBNL1 avec ses différents ARNs partenaires a constitué la majeure partie de mon travail de thèse.

Introduction

Introduction

L'introduction bibliographique qui va suivre est orientée autour de trois grands axes. Dans un premier temps je présenterai des généralités sur les maladies génétiques liées à des amplifications de répétitions. J'introduirai ensuite les différents mécanismes moléculaires proposés pour expliquer les symptômes liés à ces maladies et je finirai par une revue des mécanismes moléculaires pouvant expliquer l'apparition de ces expansions de répétitions dans le génome. Le second chapitre de cette introduction récapitulera les bases moléculaires de l'épissage et notamment de la régulation de l'épissage alternatif. Le troisième et dernier chapitre sera dédié à la présentation des dystrophies myotoniques. Il inclura une description des protéines des familles MBNL et CELF et de notre modèle d'étude de l'épissage alternatif : le pré-ARNm hcTNT. Les pistes thérapeutiques envisagées actuellement seront également présentées dans cette partie.

Chapitre I : Généralités sur les maladies génétiques liées à des amplifications de séquences répétées

I. Les différents types d'amplifications de courtes séquences dans le génome humain à l'origine de maladies génétiques

Le séquençage du génome humain a révélé que ce dernier contenait 3% de séquences microsatellites, qui sont de courtes séquences (généralement de 2 à 6 paires de bases) répétitives polymorphiques (Lander *et al.*, 2001). Ces répétitions sont souvent instables et hypervariables à cause d'erreurs au cours de la réplication, de la réparation et de la recombinaison. Dès 1991, des expansions de ces séquences microsatellites ont été identifiées comme étant à l'origine de maladies génétiques héréditaires, telles que le syndrome de l'X fragile ou l'amyotrophie bulbospinale liée à l'X (Verkerk *et al.*, 1991 ; La Spada *et al.*, 1991). A l'heure actuelle, une trentaine de pathologies liées à des expansions de ce type ont été clairement identifiées, cependant la liste est loin d'être exhaustive. Ces pathologies résultent d'amplifications de triplets (CTG, CAG, CGG ou GAA), de quadruplets (CCTG), de pentanucléotides (ATTCT), voire même de répétitions allant jusqu'à 12 nucléotides (C₄GC₄GCG) (pour revue, Pearson *et al.*, 2005 ; Shin *et al.*, 2009). Ces répétitions peuvent être localisées au niveau des régions codantes des gènes (les exons) ou au niveau des régions

non codantes (les introns et les régions 5'- et 3'- non traduites (5'-UTR et 3'-UTR) (Fig.1) (pour revue, Mirkin, 2007).

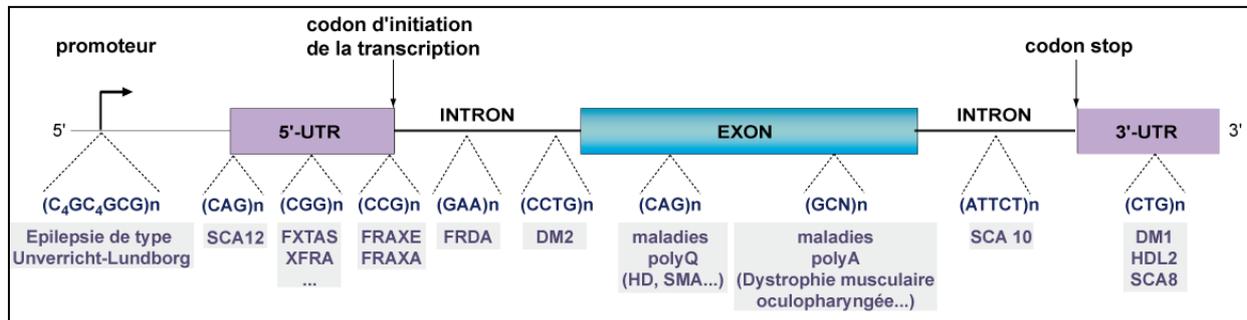


Figure 1 : Localisation dans les gènes des répétitions de séquences impliquées dans des maladies génétiques.

Les répétitions peuvent se trouver dans les différentes régions du gène : les régions traduites (exon) et les régions non-traduites (promoteur, 5'-UTR, intron et 3'-UTR). La localisation et la séquence de ces répétitions sont à l'origine de différentes pathologies.

Ces maladies génétiques autosomales dominantes ont la particularité d'avoir un mode de transmission dit « par anticipation », qui consiste en l'augmentation de la longueur des répétitions au cours des générations. De ce fait, les générations successives sont atteintes de façon plus précoce et présentent des symptômes plus sévères de la maladie (pour revues, Cummings et Zoghbi, 2000 ; Usdin et Grabczyk, 2000).

II. Les mécanismes moléculaires à l'origine des pathologies

Selon la localisation et le type d'expansion de répétitions, les mécanismes moléculaires à l'origine des pathologies varient et certains ne sont toujours pas compris. Nous allons à présent passer en revue les mécanismes moléculaires les plus fréquents (Fig.2).

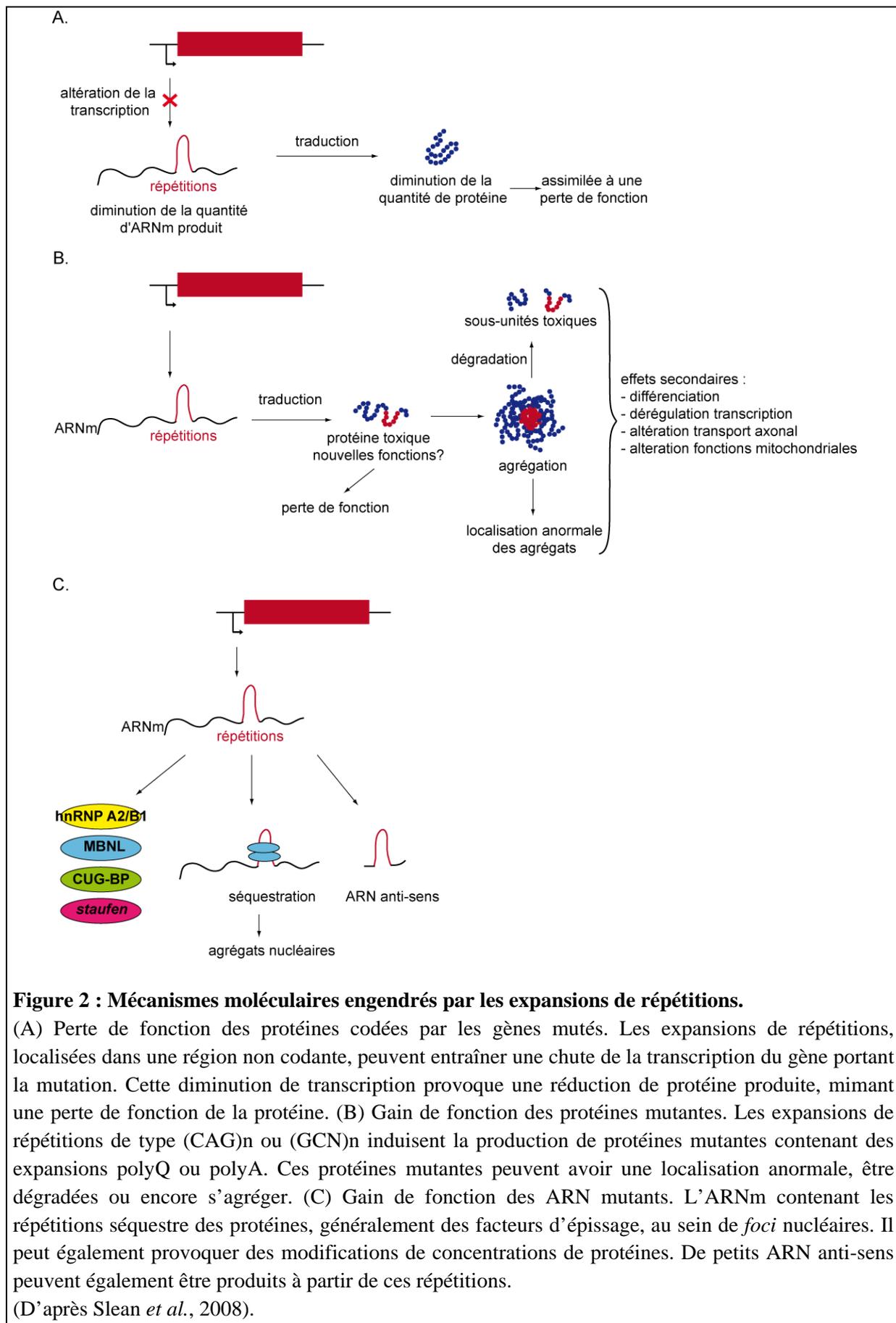


Figure 2 : Mécanismes moléculaires engendrés par les expansions de répétitions.

(A) Perte de fonction des protéines codées par les gènes mutés. Les expansions de répétitions, localisées dans une région non codante, peuvent entraîner une chute de la transcription du gène portant la mutation. Cette diminution de transcription provoque une réduction de protéine produite, mimant une perte de fonction de la protéine. (B) Gain de fonction des protéines mutantes. Les expansions de répétitions de type (CAG) $_n$ ou (GCN) $_n$ induisent la production de protéines mutantes contenant des expansions polyQ ou polyA. Ces protéines mutantes peuvent avoir une localisation anormale, être dégradées ou encore s'agréger. (C) Gain de fonction des ARN mutants. L'ARNm contenant les répétitions séquestre des protéines, généralement des facteurs d'épissage, au sein de *foci* nucléaires. Il peut également provoquer des modifications de concentrations de protéines. De petits ARN anti-sens peuvent également être produits à partir de ces répétitions.

(D'après Slean *et al.*, 2008).

II.1. Perte de fonction des protéines codées par les gènes mutés (Fig.2A)

Les expansions de répétitions au niveau de l'ADN peuvent entraîner une diminution de la transcription du gène hôte. *In fine*, cette diminution s'apparente à une perte de fonction des protéines codées par les gènes mutés. Ceci a été notamment démontré par l'étude de souris invalidées pour ces gènes hôtes et/ou par l'étude de cas cliniques aux symptômes identiques, ne présentant pas d'expansion de répétitions, mais dus à des mutations dans la séquence codante des gènes entraînant une perte de fonction (REF).

Le syndrome le plus commun, dans lequel la fonction de la protéine est perdue, est le syndrome du X fragile (FRAXA). Ce syndrome intervient lorsque le nombre de répétitions (CGG) localisées dans la région 5'-UTR du gène *FMR1* (pour « fragile X mental retardation ») est supérieur à 200. La mutation cause alors une méthylation anormale de l'îlot CpG localisé dans la région du promoteur de *FMR1* (Verkerk *et al.*, 1991). Cette augmentation de la méthylation est couplée à une diminution de l'acétylation des histones, entraînant une perte d'expression de *FMR1* et de fonction de la protéine neuronale FMRP (Coffee *et al.*, 1999). Cette protéine régule la traduction des ARNm dans les dendrites et module ainsi les fonctions synaptiques. La protéine FMRP pourrait également être impliquée dans le métabolisme de l'ARN et le transport d'ARN messagers (ARNm) notamment vers les dendrites (Dictenberg *et al.*, 2008 ; Estes *et al.*, 2008). Son absence est à l'origine du retard mental observé chez les patients souffrant de FRAXA (Willemsen *et al.*, 2004).

Le syndrome du X fragile (FRAXE) résulte de la perte de fonction de la protéine FMR2 due à la présence de répétitions (CCG) dans la région 5'-UTR du gène *FMR2* (Gecz *et al.*, 1997 ; Gu *et al.*, 1996). La protéine FMR2 est une protéine nucléaire dont la fonction n'est pas encore clairement établie. Du fait de sa capacité à activer la transcription, il a été proposé qu'elle soit un facteur de transcription (Nilson *et al.*, 1997). Cependant des données récentes montrent i) qu'elle est localisée dans les speckles, ii) qu'elle interagit spécifiquement avec des ARN contenant des quartets de G et iii) qu'elle peut réguler *in vivo* l'épissage du pré-ARNm FMR1 via une interaction directe avec un quartet de G, suggérant que FMR2 pourrait être impliquée dans la régulation de l'épissage alternatif (Bensaid *et al.*, 2009).

Une autre pathologie liée à une perte de fonction d'une protéine est l'Ataxie de Friedreich (FRDA). Ce trouble récessif autosomal est causé par une expansion de répétitions (GAA) dans le premier intron du gène *FRDA*, également appelé *FXN*, qui code la frataxine (Campuzano *et al.*, 1996). La frataxine est une protéine de la matrice mitochondriale qui intervient de manière directe ou indirecte dans le contrôle de la disponibilité du fer et dans la biogenèse des clusters Fer-Soufre et des hèmes (Pandolfo, 2002). Les répétitions (GAA)

inhibent la transcription du gène de la frataxine. En effet, ces répétitions modifient la structure de l'ADN en induisant la formation d'une triple hélice d'ADN qui bloque la progression de l'ARN polymérase (Grabczyk et Usdin, 2000). La perte d'expression de la frataxine entraîne une diminution de la production d'énergie par les mitochondries et une augmentation de la production de radicaux libres par oxydation du fer (Karthikeyan *et al.*, 2003 ; Yoon et Cowan, 2003).

II.2. Gain de fonction des protéines codées par les gènes mutés (Fig.2B)

Les répétitions (CAG) dans la région codante des gènes correspondent au type d'expansion de triplets le plus répandu, entraînant une altération de la physiologie des protéines. En effet, la traduction de ces répétitions engendre une expansion de polyglutamines (polyQ). Les maladies par expansion de polyglutamines (MEP) constituent un groupe de neuf maladies neurodégénératives héréditaires incluant la maladie de Huntington (HD), l'atrophie dentatorubral et pallidolusienne (DRPLA), l'amyotrophie spinobulbaire (SBMA ou maladie de Kennedy) et les atrophies spinocérébelleuses 1, 2, 3, 6, 7 et 17 (SCA1/2/3/6/7/17) (Zoghbi et Orr, 2000 ; Nakamura *et al.*, 2001 ; pour revue, Ravache *et al.*, 2010). Bien que la taille, la structure, la localisation cellulaire et la fonction des protéines résultantes diffèrent, toutes les pathologies induites par les répétitions (CAG) correspondent à des troubles neurodégénératifs. Les expansions polyQ ont la capacité de former des agrégats hautement structurés s'apparentant aux fibres amyloïdes retrouvées dans d'autres maladies neurodégénératives comme la maladie d'Alzheimer (MA), la maladie de Parkinson et la sclérose latérale amyotrophique (pour revue, Ross et Poirier, 2004). Outre les protéines mutantes, ces agrégats séquestrent de nombreuses protéines, tel des facteurs de transcription et des constituants des protéasomes, les détournant vraisemblablement de leur fonction normale (pour revue, Usdin et Grabczyk, 2000). Cependant, des études réalisées chez l'Homme et la souris montrent bien que ces pathologies résultent d'un gain de fonction toxique des protéines et non juste d'une perte d'expression. En effet, des modèles de souris transgéniques exprimant les protéines mutées reproduisent les phénotypes, alors que l'inactivation des gènes codant ces protéines ne permet pas de reproduire les symptômes des pathologies (Watase *et al.*, 2002 ; Yoo *et al.*, 2003 ; Matilla *et al.*, 1998). Cependant, les bases moléculaires à l'origine de ces différentes pathologies ne sont pas encore totalement comprises. Les points communs entre ces neuf MEP sont (i) un changement de conformation des protéines mutantes, (ii) leur accumulation au sein d'agrégats et (iii) la séquestration au sein des agrégats de nombreuses protéines impliquées dans des processus aussi divers que la transcription, la dégradation des protéines et

le contrôle qualité des ARNm (Jiang *et al.*, 2003). Les protéines mutantes pourraient également être clivées par des protéases, entraînant la formation de fragments ayant des effets pathogènes. D'autres processus cellulaires peuvent être affectés tels que les fonctions mitochondriales, l'homéostasie du calcium et le transport axonal menant à des dysfonctionnements neuronaux et à la mort (pour revue, Gatchel et Zoghbi, 2005).

Un autre type de répétitions peut être présent dans la région codante des gènes : les répétitions (GCN) engendrant après traduction une expansion de polyalanines (polyA). Les pathologies issues de ces expansions de polyalanines constituent un groupe de neuf maladies héréditaires incluant la dystrophie musculaire oculo-pharyngée (OPMD), la syndactylie de type 2 ou synpolydactylie (SPD), dysplasie cleido-cranienne (CCD), l'homopros-encephalie de type 5 (HPE), le syndrome main-pied-génital (HFGS), le syndrome du blépharophimosis-Ptoisis-epicanthus inversus (BPES), le syndrome d'hypoventilation alvéolaire centrale congénitale (CCHS ou syndrome d'Ondine), l'hypopituitarisme lié à l'X (XH) et le retard mental lié à l'X (XLMR) (pour revue, Messaed et Rouleau, 2009). A l'exception du gène PABPN1 muté dans le cas de l'OPMD, l'ensemble des gènes affecté par des expansions de polyalanines code pour des facteurs de transcription. PABPN1 est un facteur de polyadénylation qui est également impliqué dans l'export nucléocytoplasmique des ARN poly(A) et dans la régulation de la transcription des muscles squelettiques (Perie *et al.*, 2006 ; Wang et Bag, 2006). Les altérations moléculaires issues de ces répétitions sont variables : elles peuvent conduire à un gain de fonction des protéines par un mécanisme proche de celui des MEP, mais aussi et ceci dans la majorité des cas, à une perte totale ou partielle de fonction ou exercer un effet négatif dominant. En effet, les expansions polyA peuvent entraîner un mauvais repliement protéique menant à une protéine anormale, une dégradation de la protéine, un problème de localisation et/ou une accumulation au sein d'agrégats cytoplasmiques ou nucléaires (pour revue, Messaed et Rouleau, 2009).

II.3. Gain de fonction des ARN entraînant des dérégulations de l'épissage alternatif (Fig.2C)

Actuellement, quatre pathologies humaines importantes d'origine génétique sont clairement identifiées comme étant dues à des dérégulations de l'épissage, consécutives d'amplifications de motifs CTG, CCTG, CGG ou ATTCT dans le génome humain. Il s'agit de la Dystrophie Myotonique de type 1 (ou DM1), de la Dystrophie Myotonique de type 2 (ou DM2), du syndrome X-associé fragile de tremblement/ataxie (FXTAS) et de l'ataxie spinocérébelleuse de type 10 (SCA10). Les DM1 et DM2, caractérisées respectivement par

une expansion de répétitions de triplets CTG dans la région 3'-UTR du gène de la myotonine protéine kinase (*DMPK*) et par une expansion de quadruplets CCTG dans l'intron 1 du gène *ZNF9*, sont l'objet de mon travail de thèse et seront décrites dans le chapitre III de cette introduction.

Le syndrome FXTAS (Fragile X associated Tremor Ataxia Syndrome) est un syndrome neurodégénératif spécifique des sujets porteurs d'une prémutation (60-200 CGG) au niveau de la 5'-UTR du gène *FMR1*. Il touche essentiellement les hommes prémutés âgés de plus de 50 ans. Bien que le gène muté soit le même, cette pathologie est distincte du syndrome de l'X fragile. La présence des répétitions CGG entraîne une augmentation de la quantité d'ARNm *FMR1*, mais aussi une diminution de la protéine codée par cet ARNm, la protéine FMRP (pour revue, Garcia-Arocena et Hagerman,). Le mécanisme exact de la pathologie n'est pas encore clairement identifié, cependant il a été observé que les ARNm *FMR1* mutants sont séquestrés au sein d'inclusions neuronales intranucléaires (Tassone *et al.*, 2004) et que différentes protéines de fixation à l'ARN, tels que les facteurs de régulation de l'épissage MBNL1 et hnRNP A2, sont également séquestrés au sein de ces inclusions (Iwahashi *et al.*, 2006). Il est donc probable que le syndrome FXTAS soit dû à une altération de la fonction des protéines séquestrées au sein des inclusions.

La SCA10 est une pathologie rare, due à une expansion de répétitions ATTCT (800-4500) dans l'intron 9 du gène *ATXN10* (Matsuura *et al.*, 2000). Le mécanisme moléculaire de la pathologie n'a pas encore été clairement défini, cependant les répétitions ATTCT ont la capacité de séquestrer le facteur d'épissage hnRNP K, ce qui conduirait à l'apoptose des cellules (White *et al.*, 2010).

Cependant, il est important de noter que le mécanisme moléculaire de la majorité des pathologies liées à des amplifications de répétitions correspond en réalité à une combinaison des mécanismes de gain de fonction des ARN mutants et de gain de fonction des protéines mutées. Par exemple, l'ataxie spino-cérébrale de type 8 (SCA8) est causée par une expansion de répétitions CTG au niveau de la région 3'-UTR du gène *ATXN8OS* (Koob *et al.*, 1999). Or, la transcription bidirectionnelle de ce gène produit à la fois des répétitions CUG non traduites et des répétitions CAG traduites en polyglutamines (Moseley *et al.*, 2006). Ainsi, la pathologie pourrait être due à la fois à un gain de fonction de l'ARN mutant et à un gain de fonction de la protéine mutée (Daughters *et al.*, 2009 ; Koob *et al.*, 1999 ; Nemes *et al.*, 2000). De même, dans le cas de la HDL2 (pour Huntington Disease Like 2), résultant de l'expansion de répétitions CTG dans la région codante du gène *JPH3*, les ARNm issus des

différents événements d'épissage alternatif du transcrit de ce gène contiennent des répétitions CUG codant pour des polyalanines ou des polyleucines selon l'ATG choisi, ou des répétitions CUG non codantes entraînant la formation d'agrégats toxiques (Holmes *et al.*, 2001 ; Rudnicki *et al.*, 2007). Enfin, des études récentes montrent que des répétitions CAG situées dans une région codante peuvent former des agrégats d'ARN qui recrutent MBNL1, comme dans le cas de la chorée de Huntington (de Mezer *et al.*, 2011) et être partiellement toxiques au niveau ARN chez la drosophile et chez la souris (Hsu *et al.*, 2011 ; Li *et al.*, 2008). Inversement, une étude récente montre que des répétitions CAG situées dans une région non traduite peuvent, en absence de codon ATG, être traduites en polyglutamines, polyalanines ou polysérines (Zu *et al.*, 2011). L'ensemble de ces données montre la multitude de conséquences que la présence d'expansion de répétitions peut avoir.

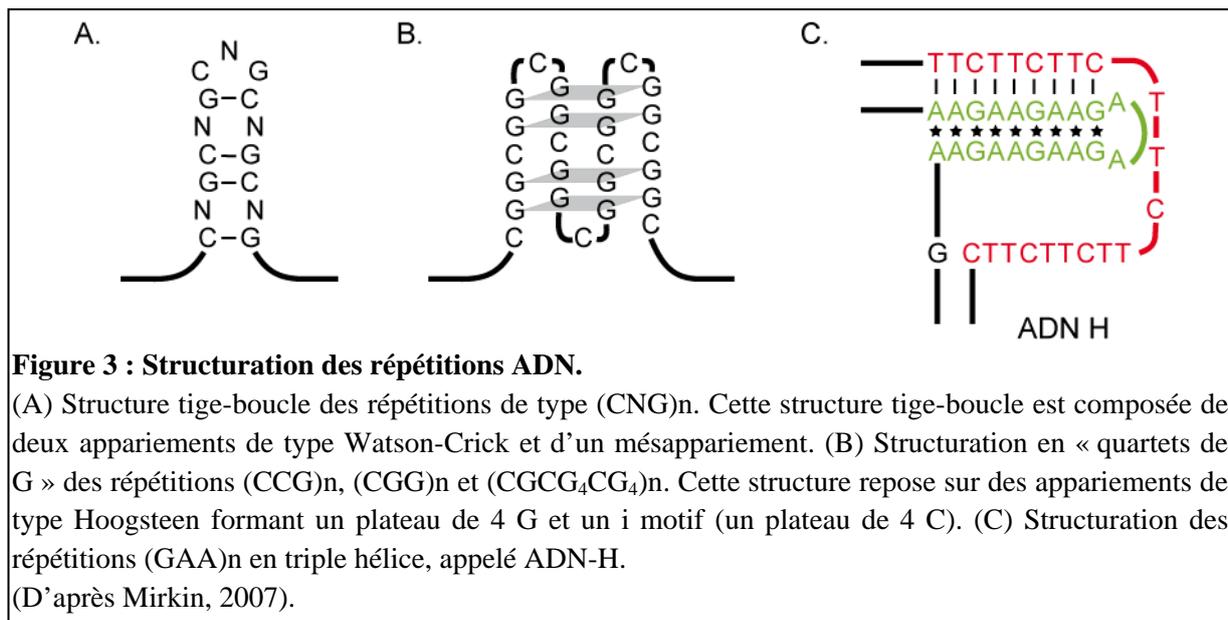
III. Mécanismes proposés pour expliquer l'expansion de répétitions de séquence

Les expansions de répétitions semblent sensibles à de nombreux paramètres tels que le type et la localisation des répétitions, le sexe du parent transmettant les répétitions, le type cellulaire ou encore le stade de développement cellulaire (pour revues, McMurray, 2010 ; Kovtun et McMurray, 2008). Ainsi, la complexité de l'instabilité des répétitions chez l'Homme implique la contribution de multiples mécanismes, ceux-ci pouvant agir individuellement ou en combinaison en fonction des paramètres. A l'heure actuelle deux sources majeures ont été mises en évidence, la réparation et la réplication. Deux autres mécanismes plus controversés ont également été proposés, la recombinaison et la transcription. Il est également important de noter que ces différents mécanismes peuvent expliquer les phénomènes de contraction de répétitions qui peuvent être observés.

Malgré la diversité des mécanismes proposés, l'ensemble de ces modèles présente une étape commune : la formation d'un ADN simple-brin permettant aux répétitions de former des structures secondaires. Comme nous le verrons par la suite, ces structures ADN particulières varient en fonction du type de répétition et sont indispensables au phénomène d'instabilité des répétitions.

III.1. Structuration particulière des répétitions ADN

L'ensemble des modèles permettant d'expliquer les expansions de répétitions est basé sur un principe commun : lors de la séparation des deux brins d'ADN au cours des processus de réplication, réparation ou recombinaison, les répétitions alors en simple brin forment des structures inhabituelles (Fig.3) (pour revues, Wells *et al.*, 2005 ; McMurray, 1999).



Les répétitions simple-brin de type (CNG)_n, auxquelles appartiennent les répétitions CTG observées dans la DM1, forment des structures tige-boucles composées de deux appariements de type Watson-Crick (W-C) et d'un mésappariement (N-N) (Gacy *et al.*, 1995). En conditions physiologiques, la stabilité de ces structures tige-boucles imparfaites diminue en fonction de leur séquence et plus particulièrement de la contribution énergétique du mésappariement, ainsi on obtient CGG > CTG > CAG = CCG (pour revue, Mirkin, 2007). Les répétitions de quadruplets (CAGG)_n et (CCTG)_n, présents chez les patients atteints de DM2, se structurent également en structures tige-boucles, avec deux appariements de type W-C et deux mésappariements (N-N) (Lam *et al.*, 2011 ; Dere *et al.*, 2004). Les répétitions de type (CCG)_n, (CGG)_n et (CGCG₄CG₄)_n adoptent également une structuration particulière dite en « quartets de G ». Cette structure repose sur des appariements de type Hoogsteen formant un plateau de 4 G et un i motif (un plateau de 4 C) (Usdin, 1998).

La dénaturation et renaturation de fragments d'ADN double-brin contenant des répétitions conduit à des glissements de brin, provoquant la formation de structures particulières sur les deux brins d'ADN localisées de manière asymétrique, l'un des deux brins étant toujours plus fortement structuré que l'autre. Par exemple, dans le cas des répétitions (CTG)_n/(CAG)_n, les répétitions (CTG) forment une structure tige-boucle stable alors que les répétitions (CAG) sont principalement dans l'état d'une bobine aléatoire (Pearson et Sinden, 1996 ; Pearson *et al.*, 2002).

Les structures dues à des glissements de brin ne sont pas les seules structures inhabituelles observées au niveau de l'ADN. Les répétitions (GAA)_n • (TTC)_n peuvent être

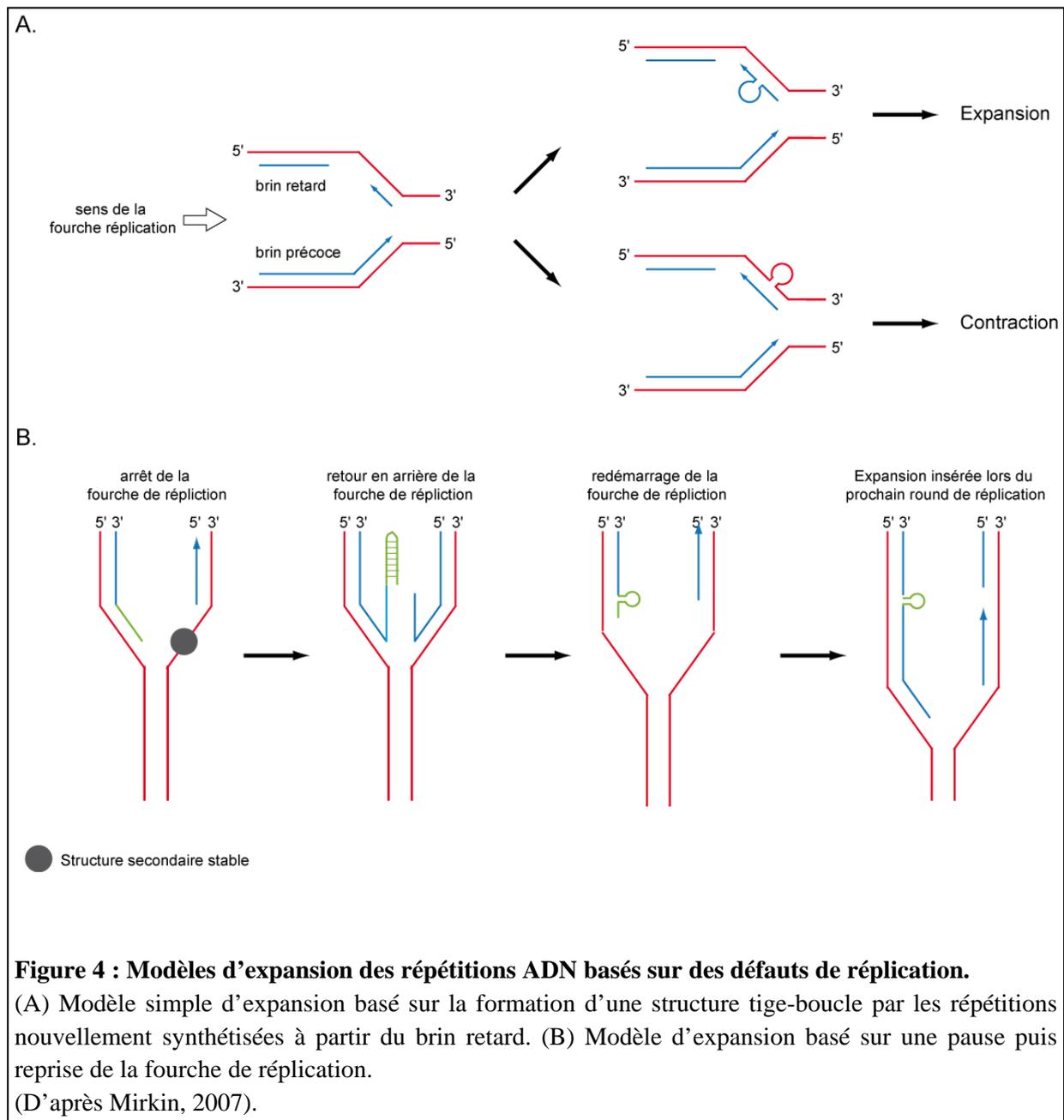
converties en triple hélice, appelé ADN-H, avec l'obtention d'un brin triple d'ADN et d'un simple brin (Gacy et McMurray, 1998).

III.2. Défauts de réplication

Le premier modèle proposé, afin d'expliquer les expansions de répétitions, était basé sur la réplication de l'ADN. En effet, l'accumulation rapide de séquences répétées ne peut se faire sans synthèse d'ADN. Cette hypothèse était favorisée par la présence d'origine de réplication à proximité des répétitions, tels que dans les loci FRAXA, HD, FRAXE, SCA7, SCA10 et DM1. Ainsi, différentes études sur de nombreux systèmes modèles tels que la levure, des souris transgéniques ou des lignées de cellules humaines, ont permis de confirmer le rôle important de la réplication sur l'instabilité des répétitions (pour revue, Lopez Castel *et al.*, 2010 ; Cleary *et al.*, 2010 ; Farrell et Lahue, 2006 ; Cleary et Pearson, 2005 ; Pelletier *et al.*, 2005 ; Yang *et al.*, 2003).

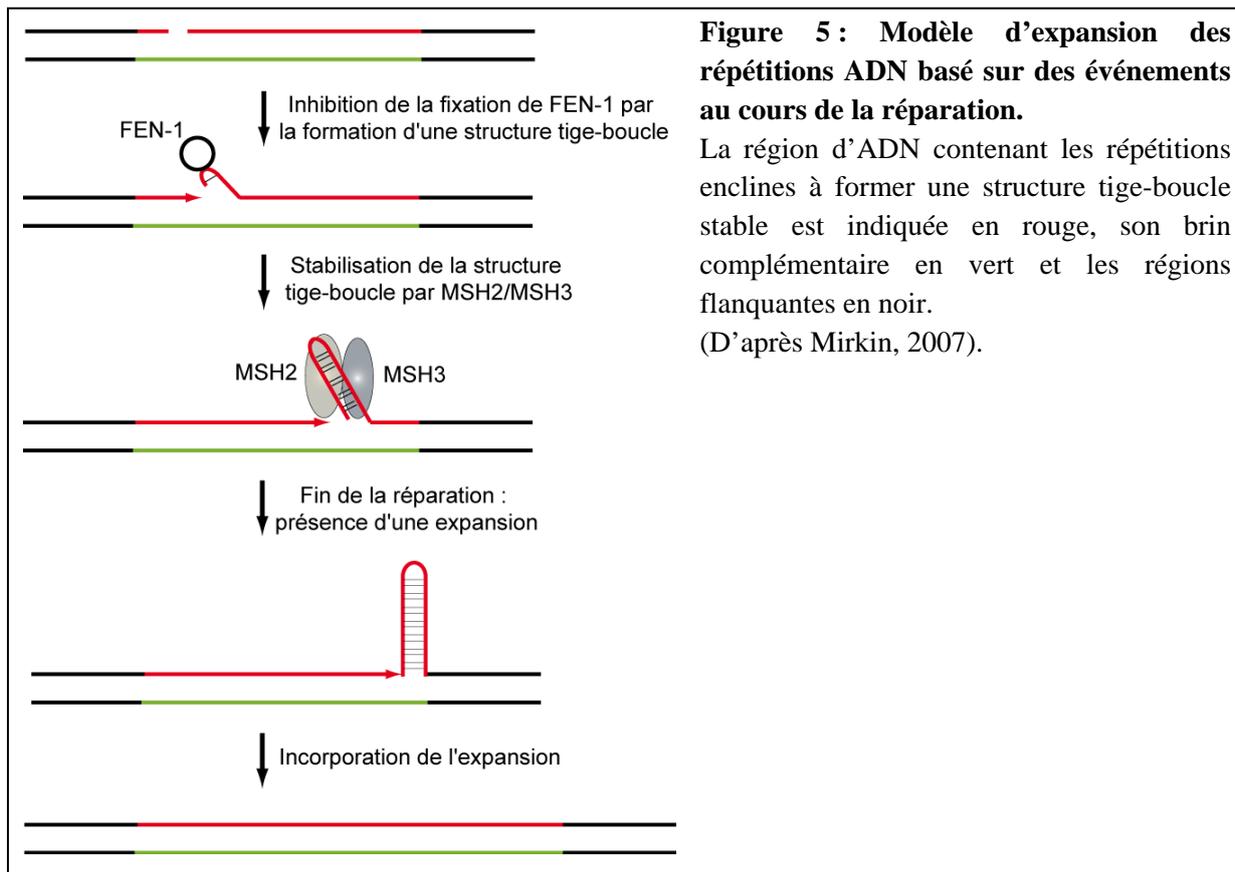
Le modèle d'expansion est basé sur la formation d'une structure tige-boucle par les répétitions nouvellement synthétisées à partir du brin retard lors de la fixation ou de l'élongation d'un fragment d'Okazaki. L'expansion des répétitions se produit lors d'une nouvelle phase de réplication (Fig.4A). Cependant, ce mécanisme n'explique que de courtes expansions (pour revue, McMurray, 2010). De plus, la stabilité des répétitions au cours de la réplication dépend alors en grande partie de leur orientation et de leur distance par rapport à l'origine de réplication (Cleary *et al.*, 2010 ; Delagoutte *et al.*, 2008 ; Cleary et Pearson, 2005 ; Mirkin et Smirnova, 2002).

Un second modèle a été mis en évidence basé sur la pause puis la reprise de la fourche de réplication (Fig.4B). D'après ce modèle, la formation d'une structure secondaire stable par les répétitions sur le brin retard bloque la synthèse de ce brin et déclenche un retour en arrière de la fourche de réplication. Cela crée une structure particulière dite en « pieds de poulet » dans laquelle les répétitions nouvellement synthétisées à partir du brin matrice se retrouvent sans leur séquence complémentaire et forment une structure tige-boucle. Lors de la reprise de la fourche de réplication revenue en arrière, des répétitions supplémentaires peuvent être ajoutées à celles précédemment synthétisées (pour revue, McMurray, 2010 ; Shishkin *et al.*, 2009 ; Fouche *et al.*, 2006). Dans ce cas de figure, la taille des expansions ne peut excéder la longueur du fragment d'Okazaki, c'est-à-dire entre 100 et 200 nucléotides chez les eucaryotes, ce qui est cohérent avec la taille des expansions observées lors de la spermatogénèse chez les patients atteints de DM1 (Martorell *et al.*, 2004).



III.3. Événements de réparation (Fig.5)

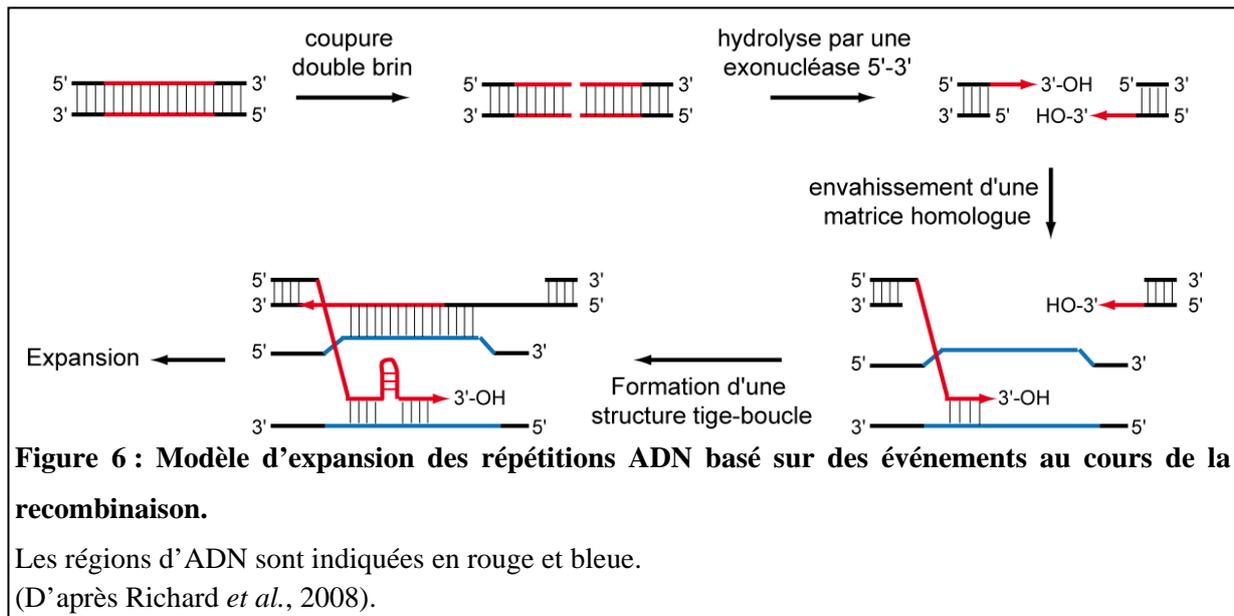
Différentes observations sur l'amplification des répétitions ne pouvaient être expliquées par un mécanisme d'expansion reposant uniquement sur la répliation. En effet, l'instabilité des répétitions augmente avec l'âge et ceci même dans les cellules ne se divisant plus, comme les neurones. De manière intéressante, de nombreuses protéines communes au système de répliation et réparation, telles que la ligase I ou la protéine FEN-1, jouent un rôle dans l'instabilité des répétitions (Yang et Freudenreich, 2007).



Suite à ces observations, une contribution des systèmes de réparation dans l'expansion des répétitions a été proposée. Les systèmes de réparation des cassures double brins d'ADN également appelés DSB (pour « double-strand breaks), tels que les systèmes de recombinaison homologue ou d'enjambement, ne semblent avoir aucun effet sur la stabilité des répétitions chez l'Homme ou chez la souris (Wolff, 1989 ; Savouret *et al.*, 2003). A l'inverse, des modèles de souris transgéniques HD ou DM déplétées de protéines impliquées dans les différents systèmes de réparation de cassure simple-brin (MSH2, MSH3, OGG1 ou CSB par exemple) indiquaient un rôle de ces systèmes dans l'instabilité des répétitions (Kovtun *et al.*, 2007 ; Rolseth *et al.*, 2008 ; Slean *et al.*, 2008 ; Goula *et al.*, 2009 ; Tome *et al.*, 2009). Ces systèmes de réparation sont les systèmes BER (pour « base excision repair), NER (« nucleotide excision repair ») et plus particulièrement le système de réparation couplé à la transcription TCR (« transcription coupled repair »), et enfin le système MMR (pour « Mismatch Repair »). Dans l'ensemble de ces mécanismes, l'ADN simple brin forme une structure tige-boucle stabilisée par le complexe Msh2-Msh3 qui est ensuite incorporé au sein de l'ADN par un mécanisme encore indéterminé (pour revue, McMurray, 2010).

III.4. Amplifications résultant d'événements de recombinaison (Fig.6)

Un troisième mécanisme basé sur la recombinaison d'ADN a été proposé. Le mécanisme le plus simple serait un croisement inégal entre des répétitions localisées sur des chromosomes homologues durant la méiose, aboutissant à des expansions ou contractions réciproques. Ce processus a été proposé afin d'expliquer les expansions de répétitions (GCN)_n, qui codent des séquences polyA. En effet, des variations dans la troisième position de la répétition ont été détectées, indiquant que l'expansion de ces répétitions serait due à des croisements inégaux entre des blocs de répétitions (GCN)_n (Warren, 1997 ; Cocquempot *et al.*, 2009). Cependant, la recombinaison méiotique a été écartée en tant que source d'instabilité des répétitions en raison de l'absence de changement des régions flanquantes (Richards et Sutherland, 1992). A l'inverse, la recombinaison mitotique est une source d'instabilité plausible. En effet, l'expansion de répétitions stimule la recombinaison mitotique dans des cellules en division et l'augmentation de l'activité de recombinaison est proportionnelle à la longueur des répétitions (Jakupciak et Wells, 1999 ; Dere et Wells, 2006). Des études chez la bactérie et la levure montrent une relation entre la recombinaison et la réplication. Le mécanisme proposé est basé sur le modèle SDSA (pour « Synthesis Dependant Strand Annealing »), démontré chez les microorganismes (*Escherichia coli* et *Saccharomyces cerevisiae*). Ce modèle semble pouvoir s'appliquer au génome humain. Suite à une coupure double-brin, chaque brin d'ADN peut être hydrolysé par une exonucléase 5'→3', ce qui conduit à la formation d'extrémités 3'-OH protubérantes. La réparation de la coupure se fait par envahissement d'une matrice homologe, telle que l'allèle homologe, par les extrémités 3'-OH protubérantes. Après appariement des brins, les deux extrémités 3'-OH générées par la cassure servent d'amorces pour la synthèse réparatrice d'ADN. Si la cassure double-brin a eu lieu au niveau de répétitions, d'une part l'appariement du brin 3' protubérant peut se faire à différentes positions de la séquence répétée et d'autre part ces répétitions peuvent former des structures en tige-boucle lors de la synthèse d'ADN, entraînant une expansion des répétitions (Richard et Paques, 2000).



III.5. Instabilité des répétitions au cours de la transcription

L'implication du système TCR dans l'instabilité des répétitions indique que la transcription pourrait également y jouer un rôle bien que le taux de transcription ne soit pas toujours corrélé avec l'instabilité des répétitions (Pearson *et al.*, 2005). De nombreux modèles de souris transgéniques, de lignées cellulaires humaines transfectées par un minigène inducible contenant des répétitions CAG ou des expériences de siRNA montrent que la transcription augmente la fréquence des contractions mais également des expansions des répétitions CAG (Nakamori *et al.*, 2011 ; Jung et Bonini, 2007 ; Lin *et al.*, 2006, Mangiarini *et al.*, 1997). Contrairement aux autres mécanismes, l'effet de la transcription n'est pas direct mais se fait via une interaction avec les systèmes NER ou MMR de réparation de l'ADN chez les mammifères, elle peut également se faire via la réplication mais uniquement dans les bactéries et les levures (Salinas-Rios *et al.*, 2011 ; pour revue, Lin *et al.*, 2009). En effet, la synthèse d'ARN ne peut pas changer la longueur d'une matrice ADN mais elle peut participer à la formation d'une structure hybride entre l'ARN et l'ADN au niveau des répétitions. Cette structure particulière, également appelée « R-loop », prolonge la demi-vie de la forme simple brin des répétitions ADN par ralentissement de l'ARN polymérase, permettant ainsi la formation accrue de structures secondaires de l'ADN qui engagent à leur tour un ou plusieurs systèmes de réparation de l'ADN (Reddy *et al.*, 2011 ; Lin *et al.*, 2010 ; Li et Manley, 2006).

III.6. Modulation de l'instabilité des répétitions par éléments *cis* et changements épigénétiques

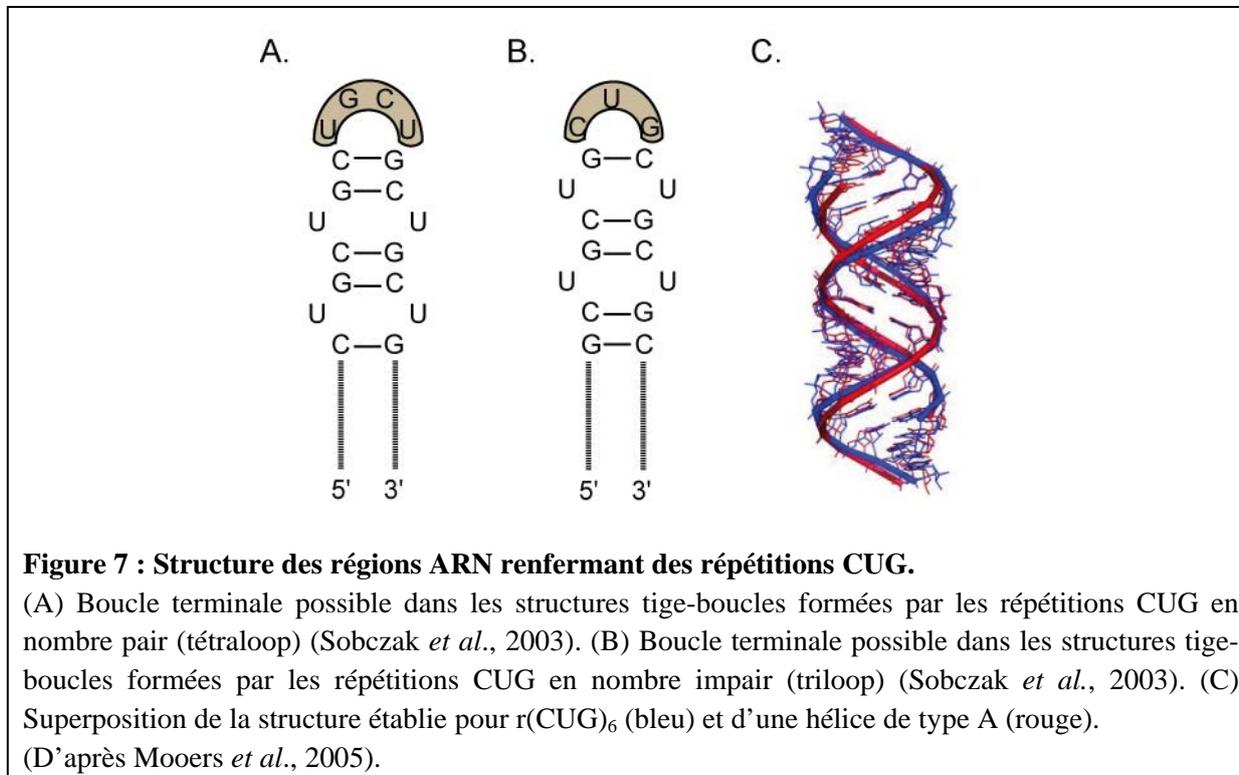
Le mosaïcisme du nombre de répétitions observé dans les différents tissus d'un même organisme, ainsi qu'entre différentes pathologies présentant le même type de répétitions, peut s'expliquer par la présence d'éléments *cis* régulateurs environnant les répétitions et par la présence de modifications épigénétiques de l'ADN. Ainsi, l'analyse des régions flanquant les répétitions de nombreux loci révèle la présence d'îlots CpG entourant ces répétitions. La méthylation de la cytosine de ces îlots inhibe la fixation de la protéine CTCF, perturbant alors la condensation de la chromatine. De nombreuses expériences ont pu mettre en évidence un lien entre la méthylation de l'ADN et la stabilité des répétitions. Ainsi, le traitement de cellules mammifères avec des inhibiteurs de méthyltransferase entraîne une déstabilisation des répétitions CTG/CAG au niveau du locus DM. La méthylation de cet élément est observée dans la forme congénitale de DM, forme qui présente le plus grand nombre de répétitions CTG/CAG, suggérant que le contexte de condensation de la chromatine a un impact sur la stabilité des répétitions de triplets.

IV. Les structures ARN particulières générées par des répétitions de triplets

Les séquences amplifiées, contenues dans la partie transcrite des gènes, se retrouvent au niveau du produit de transcription, et peuvent donc modifier sa structure. Récemment, des études de dichroïsme circulaire ont permis de classer en 4 familles les différents types de répétitions ARN en fonction de leur structuration (Sobczak *et al.*, 2010). Ainsi, les répétitions de motif CAA, UUG, AAG, CUU, CCU, CCA et UAA se structurent faiblement, alors que les répétitions CAU, CUA, UUA, AUG et UAG ont tendance à former des structures tige-boucles semi-stables. L'ensemble des répétitions CGA, CGU et tous les motifs CNG forment des structures en épingles à cheveux relativement stables, tandis que les motifs AGG et UGG se structurent en quadruplexes de G stables.

Les structures formées par les différents types de répétitions CNG dans les ARN ont été étudiées en solution par emploi de sondes enzymatiques et chimiques, ainsi que par approche RMN, par microscopie électronique et très récemment par radiocristallographie (Kiliszek *et al.*, 2011 ; Kiliszek *et al.*, 2010 ; Zumwalt *et al.*, 2007 ; Mooers *et al.*, 2005 ; Sobczak *et al.*, 2003 ; Leppert *et al.*, 2004 ; Michalowski *et al.*, 1999). Les différentes données RMN suggèrent que ces séquences forment une structure double-brin au niveau de la tige, où 2 paires Watson-Crick G-C et C-G sont séparées de façon régulière par un

mésappariement N:N. Ces structures en tige-boucle sont stables à partir de 10 répétitions. Les structures formées sont hétérogènes car différentes structures en tige-boucle décalées peuvent être formées (Napierala et Krzyzosiak, 1997). Selon la structure formée, les séquences en simple-brin situées aux extrémités de la molécule sont de tailles variables. Lorsque le nombre de répétitions est pair ($2n$) et que la tige-boucle la plus longue possible est formée, la boucle terminale est une boucle à 4 nucléotides (“tetraloop”). Lorsque le nombre de répétitions est impair ($2n+1$) et que la tige-boucle la plus longue possible est formée, la boucle terminale est une boucle à 3 nucléotides (“triloop”) (Fig.7) (Sobczak *et al.*, 2003). La stabilité de ces structures augmente de façon croissante avec la longueur des répétitions (Napierala et Krzyzosiak, 1997). Il faut aussi noter que les séquences environnant les régions répétées peuvent participer à la stabilisation des structures en tige-boucle (CNG) $_n$. En effet, les régions d’ARN bordantes peuvent former une structure en double-brin qui prolonge l’hélice formée par les répétitions (Jasinska *et al.*, 2003). Des études réalisées par diffraction des rayons X sur des cristaux obtenus avec un ARN renfermant 6 répétitions CUG ou 3 répétitions CGG montrent que les hybrides formés par appariement de deux molécules adoptent une structure hélicoïdale de type A (Fig.7) (Kiliszek *et al.*, 2011 ; Mooers *et al.*, 2005).



La structuration particulière de ces ARN joue en elle même un rôle primordial dans les différentes pathologies. En effet, ces structures peuvent être reconnues par différentes

protéines, entraînant comme nous l'avons vu précédemment la formation d'agrégats et la séquestration de ces protéines. La stabilité et le type de mésappariement N:N semblent permettre de sélectionner différentes protéines. Alors que la protéine MBNL1 est détectée dans les *foci* CUG, CCUG, CAG et CGG, les protéines Pur α et Sam68 ont été trouvées uniquement avec les répétitions de type CGG et enfin la protéine hnRNP K est associée *in situ* avec les répétitions AUUCU (pour revue, Wojciechowska et Krzyzosiak, 2011 ; Sellier *et al.*, 2010).

Ce descriptif montre la complexité des mécanismes moléculaires mis en jeu, tant d'un point de vue de la mise en place des répétitions que des effets physiologiques qui en découlent. Cela met également en avant la difficulté de travailler sur la compréhension des mécanismes à l'origine des maladies génétiques.

Chapitre II : L'épissage

Chez les eucaryotes, les gènes nucléaires codant les protéines comportent pour la plupart des séquences codantes (exons) interrompues par des séquences non codantes (introns). Des séquences introniques sont également présentes dans des gènes codant des ARN stables (ARN de transfert, ARN ribosomiques), dans des gènes d'organites cellulaires (chloroplaste, mitochondrie), ainsi que dans certains gènes bactériens et d'Archaea, mais le mécanisme d'élimination de ces introns est différent de celui utilisé pour les gènes nucléaires eucaryotes.

Les transcrits des gènes nucléaires sont produits par l'ARN polymérase II et sont soumis à des processus complexes de maturation, ceci pour convertir les pré-ARN messagers (pré-ARNm) en ARN messagers (ARNm) fonctionnels. Avant qu'un ARNm puisse être transporté en dehors du noyau pour être traduit, il doit être coiffé au niveau de son extrémité 5', épissé de ses introns et modifié au niveau de son extrémité 3' par un clivage et l'ajout d'une queue Poly(A). L'épissage correspond à l'excision des introns du transcrit primaire puis à la jonction des exons. Il est dit constitutif lorsqu'un exon est systématiquement inséré dans le transcrit mature. A l'inverse, l'épissage alternatif permet de produire des protéines différentes à partir d'un seul gène, ceci du fait de l'existence de plusieurs sites potentiels d'épissage en compétition les uns avec les autres au sein du pré-ARNm. Certains exons seront choisis et inclus dans l'ARNm, alors que d'autres ne le seront pas. Environ 75 % des événements d'épissage alternatif ont pour conséquence de modifier la séquence codante des protéines, ce qui conduit à la production de protéines distinctes et seulement 19 % conduiraient à la production de protéines tronquées (Modrek et Lee, 2002, Zavolan *et al.*, 2003). Ce processus contribue grandement à la diversité du protéome et fournit une part d'explication au différentiel entre le nombre de gènes et celui des protéines., l'épissage alternatif constitue un mode particulier de régulation de l'expression des gènes. Ainsi, un gène donné est en général capable de coder différentes protéines, en fonction du stade du développement, l'état de différenciation cellulaire ou l'état physiologique (Lander *et al.*, 2001 ; Modrek et Lee, 2002 ; Resch *et al.*, 2004 ; pour revue, Beyersmann, 2000). Pour atteindre cette flexibilité, la sélection des sites d'épissage chez les eucaryotes supérieurs a évolué pour pouvoir dépendre de multiples paramètres comme la force du site d'épissage, la présence ou l'absence de régulateurs d'épissage, la structure secondaire de l'ARN, l'architecture exon/intron et le processus de synthèse du pré-ARNm. La participation relative

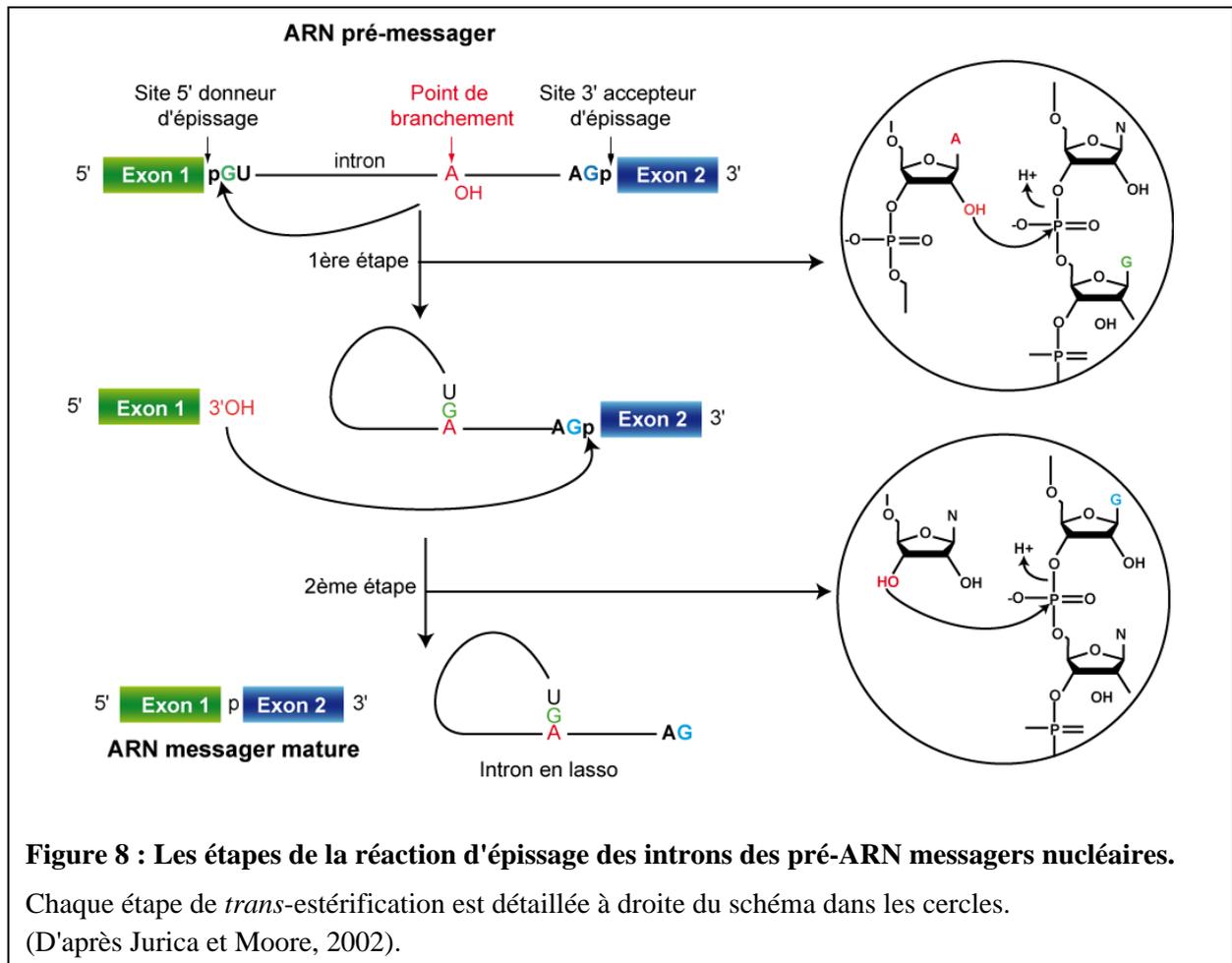
de chacun de ces paramètres contrôle l'efficacité de la reconnaissance des sites d'épissage et le retrait des introns flanquants.

L'élimination des introns est un processus très complexe qui nécessite l'assemblage d'un macrocomplexe, le spliceosome, composé de 5 petits ARN nucléaires snRNA (U1, U2, U4, U5 et U6) et d'un très grand nombre de protéines (plus de 100 protéines différentes) autour de l'intron à éliminer (pour revue, Will et Luhrmann, 2001). Il existe en fait 2 types d'introns chez les eucaryotes supérieurs : les introns dits majeurs, éliminés par le spliceosome que je décrirai par la suite et les introns dits mineurs, rencontrés moins fréquemment. Ces introns sont éliminés par un autre type de spliceosome, renfermant les snRNA U11 et U12 à la place des snRNA U1 et U2, les snRNA U4atac et U6atac à la place des snRNA U4 et U6, et le snRNA U5 (pour revue, Tarn et Steitz, 1997). Ici, nous allons uniquement nous intéresser au spliceosome majeur.

I. Description générale de l'épissage des ARN

I.1. La réaction chimique d'épissage

La réaction d'épissage se déroule en deux étapes successives de trans-estérification (Fig.8). La première étape consiste en l'attaque nucléophile du phosphate en 5' de l'intron (site 5' donneur) par l'hydroxyle en 2' d'un résidu adénosine situé dans l'intron au niveau d'un élément de séquence conservé, la boîte de branchement. Deux intermédiaires réactionnels sont alors produits, d'une part l'exon 1, avec une extrémité 3'OH libre, et d'autre part un intermédiaire contenant l'exon 2 et l'intron structuré en lasso, dont la guanosine 5' terminale est covalamment liée au point de branchement par une liaison 2'-5' phosphodiester. Dans la seconde étape de trans-estérification, l'extrémité 3'OH de l'exon 1 réalise une attaque nucléophile du phosphate situé à l'extrémité 3' de l'intron (site 3' accepteur d'épissage). Ceci conduit à la ligation des deux exons et à la libération de l'intron sous forme de lasso. Ce sont les produits finaux de la réaction d'épissage. L'intron est alors linéarisé et dégradé (Jurica *et al.*, 2002).



I.2. Les séquences définissant les introns

L'élimination des introns nécessite un ensemble de séquences conservées permettant leur reconnaissance par la machinerie d'épissage. Malgré la grande variabilité des séquences introniques des pré-ARNm des eucaryotes supérieurs, la comparaison des séquences situées aux extrémités des introns a permis de définir des séquences consensus pour les sites 5' et 3' d'épissage (Fig.9).

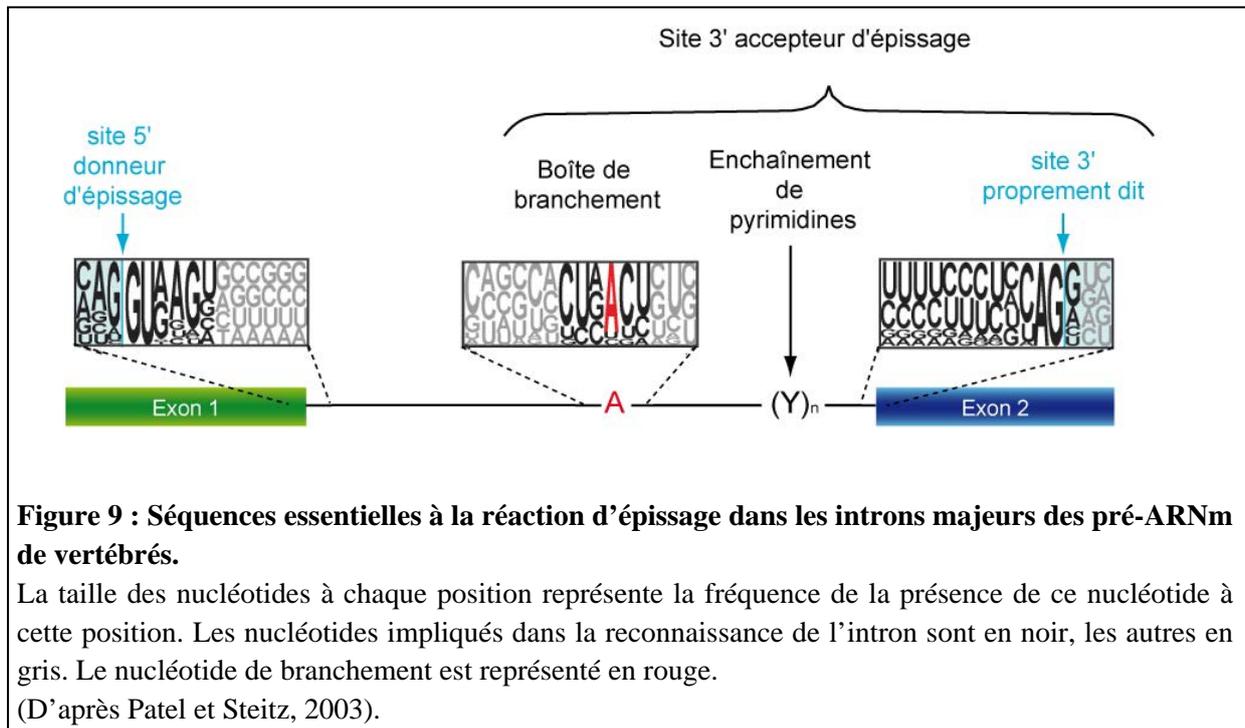


Figure 9 : Séquences essentielles à la réaction d'épissage dans les introns majeurs des pré-ARNm de vertébrés.

La taille des nucléotides à chaque position représente la fréquence de la présence de ce nucléotide à cette position. Les nucléotides impliqués dans la reconnaissance de l'intron sont en noir, les autres en gris. Le nucléotide de branchement est représenté en rouge.

(D'après Patel et Steitz, 2003).

I.2.a. Le site 5' donneur d'épissage

La bordure exon/intron est appelée site 5' donneur d'épissage. La comparaison de ces bordures exon/intron a permis d'établir la séquence consensus: AG/**G**URAGU (R = purine et / = jonction exon/intron), dans laquelle le dinucléotide **GU** en 5' de l'intron est le motif le plus conservé (Moore et Sharp, 1993) (Fig.9). En effet, des mutations ponctuelles du site 5' d'épissage, notamment au niveau du dinucléotide GU très conservé, diminuent ou abolissent l'épissage normal. Ces mutations entraînent généralement l'utilisation d'un site 5' cryptique qui n'est pas utilisé normalement et qui est généralement proche du site 5' naturel. La reconnaissance d'un site 5' cryptique par le spliceosome n'est pas dépendante simplement de sa séquence mais plus d'un contexte global favorable à l'utilisation de ce site, ce qui reflète une grande flexibilité dans les pré-ARNm de mammifères.

I.2.b. Le site 3' donneur d'épissage

La définition des sites 3' accepteurs d'épissage est plus complexe. En effet, chez les vertébrés, ils renferment 3 éléments distincts : la boîte de branchement, une séquence polypyrimidine et le site 3' proprement dit (Fig.9).

La boîte de branchement est située 18 à 40 nts en amont du site 3' d'épissage. Sa séquence consensus établie est très dégénérée : YNYURAC (Y = pyrimidine). Le résidu adénosine conservé (en gras), correspondant au point de branchement, est responsable de la

première réaction de trans-estérification. Dans quelques cas, ce résidu peut être une cytosine ou une uridine. La séquence polypyrimidine correspond à un enchaînement de 10 à 20 résidus pyrimidines, les résidus U étant plus favorables à sa fonction. Elle est localisée en amont du site 3' accepteur d'épissage. Enfin, le site 3' d'épissage, proprement dit, se situe à la bordure intron/exon et est caractérisé par une courte séquence conservée YAG/G.

Les sites 5' et 3' d'épissage sont définis par des séquences très courtes qui, à l'exception de quelques nucléotides consensus, sont relativement dégénérées. La dégénérescence de ces résidus permet une reconnaissance moins efficace de l'exon alternatif, laissant la place à une régulation plus fine pour favoriser l'inclusion ou l'exclusion de l'exon (Stamm *et al.*, 1994 ; Carmel *et al.*, 2004 ; Sorek *et al.*, 2004 ; Lear *et al.*, 1990).

I.3. Les composants et l'assemblage du spliceosome

Le spliceosome est un macrocomplexe ribonucléoprotéique (RNP) architecturé par l'assemblage de 5 particules ribonucléoprotéiques appelées UsnRNP (snRNP U1, U2, U4, U5, U6). Chacune de ces UsnRNP s'architecture elle-même autour d'un ou plusieurs petits ARN nucléaires riches en uridine (ou UsnRNA), et comprend un ensemble de 7 protéines Sm ou Sm-like (Lsm) ainsi que plusieurs protéines spécifiques à chaque UsnRNP. Comme décrit par la suite, les UsnRNP et les protéines associées au spliceosome reconnaissent les séquences nécessaires à l'épissage décrites dans le paragraphe précédent. Des interactions ARN-ARN et ARN-protéines se forment alors de manière ordonnée et transitoire autour des sites d'épissage, conduisant à la formation du spliceosome au sein duquel ont lieu les 2 réactions de trans-estérification (pour revues, Hastings et Krainer, 2001 ; Will et Luhrmann, 2001).

I.3.a. Biogénèse des UsnRNP

Chez les mammifères, les UsnRNA (U1, U2, U4, U5 et U6) sont de petits ARN (de 100 à 300 nts) riches en résidus uridine dont la séquence et la structure 2D sont très conservées (Fig.10) (pour revue, Will et Luhrmann, 1997; Branlant *et al.*, 1980 ; Krol *et al.*, 1981 ; Branlant *et al.*, 1982 ; Branlant *et al.*, 1983 ; Mougin *et al.*, 2002)

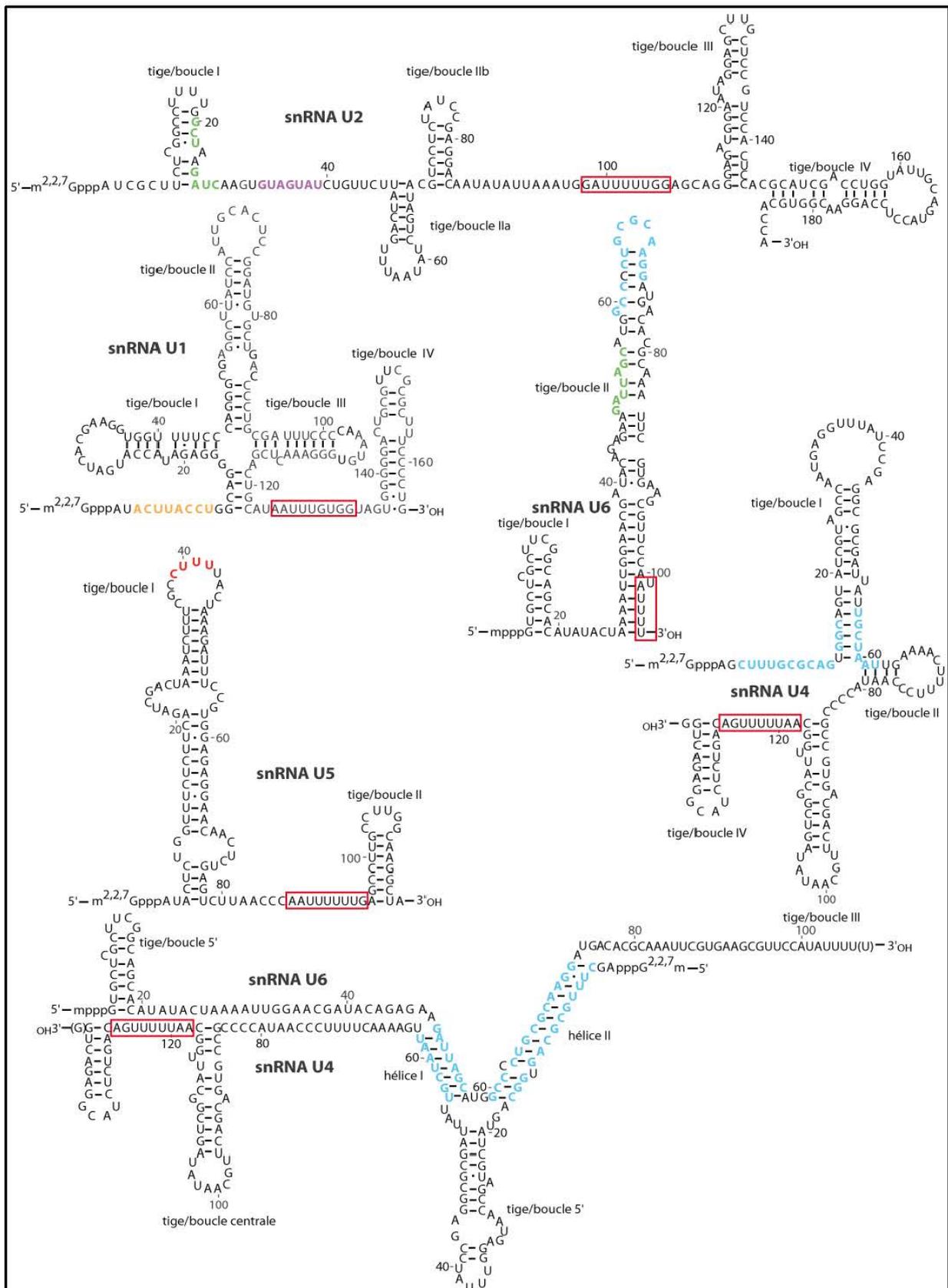
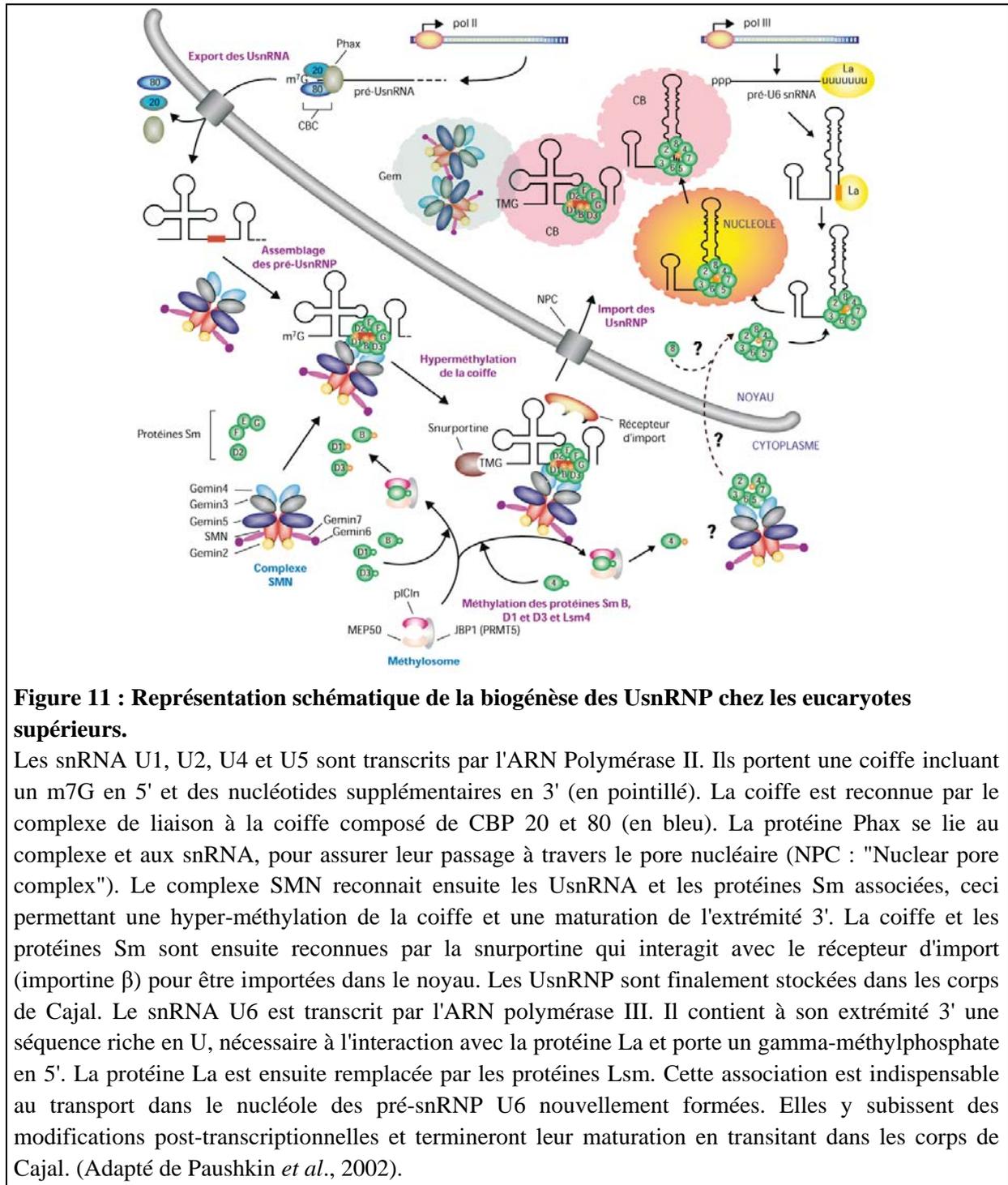


Figure 10 : Séquences et structures secondaires des UsnRNA humains majeurs.

Les structures secondaires représentées ci-dessus ont été établies à partir de celles déterminées expérimentalement chez le rat par : Branlant *et al.*, 1980, Mount et Steitz, 1981 pour le snRNA U1, Branlant *et al.*, 1982 pour le snRNA U2, Krol *et al.*, 1981 pour le snRNA U4 et Branlant *et al.*, 1983 pour le snRNA U5. La structure secondaire du snRNA U6 humain a été établie par Mougouin *et al.*,

2002. Enfin, la structure secondaire du duplex U4/U6 a été établie par Brow et Guthrie, 1988 et Mougin *et al.*, 2002. Le site Sm permettant la fixation des protéines Sm (Branlant *et al.*, 1982) ou Lsm (Achsel *et al.*, 1999) est encadré en rouge dans les différents snRNA. Les nucléotides en orange sont impliqués dans des appariements avec le site 5' (snRNA U1), ceux en violet avec la boîte de branchement (snRNA U2) et ceux en rouge avec les deux exons (snRNA U5). Les nucléotides en vert sont impliqués dans les appariements entre les snRNA U2 et U6, ceux en bleu sont impliqués dans les appariements entre les snRNA U4 et U6. Les modifications post-transcriptionnelles des UsnRNA ne sont pas indiquées sur cette figure.

A l'exception du snRNA U6, tous les précurseurs de snRNA U1, U2, U4 et U5 sont transcrits par l'ARN polymérase II dans le noyau (Fig.11). Ils portent une coiffe monométhylée (m7G) en 5' et des nucléotides additionnels à leur extrémité 3'. Une fois exportés dans le cytoplasme, chaque UsnRNA s'associe de manière ordonnée avec sept protéines Sm (B/B', D3, D2, D1, E, F, et G) pour former les pré-UsnRNP grâce à l'action du complexe SMN (pour Survival of Moto Neuron). Cette étape s'accompagne pour chaque UsnRNA, de la conversion de la coiffe monométhylée m7G en une coiffe triméthylée m2,2,7G, d'une maturation de l'extrémité 3' et de l'import des pré-UsnRNP dans le noyau (pour revue, {Paushkin, 2002 #228}). Une fois dans le noyau, les pré-UsnRNP s'accumulent de manière transitoire dans les corps de Cajal, où elles subissent des modifications post-transcriptionnelles à différentes positions. Ces modifications, qui sont majoritairement des pseudouridylations et des 2'-O-méthylations, seraient ciblées par des scaRNA, de petits ARN guides à boîte H/ACA ou à boîte C/D ({Massenet, 1999 #229} ; pour revue, {Kiss, 2004 #231}). Cette étape de modifications semble essentielle à l'assemblage et au fonctionnement du spliceosome ({Yu, 1998 #233} ; {Jady, 2001 #234}). L'assemblage des UsnRNP se termine avec l'association des protéines spécifiques de chaque particule avec chaque UsnRNA. Les particules UsnRNP matures seront alors, soit immédiatement assemblées en spliceosome, soit accumulées dans des structures de la région interchromatinienne appelées IGC (Interchromatin Granules Cluster) ou « speckles ». Les speckles sont généralement localisés à proximité des sites actifs de transcription et ont plusieurs fonctions dont le stockage/relargage de facteurs ou protéines impliqués dans les processus de transcription et d'épissage (pour revue, {Lamond, 2003 #235}).



Contrairement aux autres snRNA, le snRNA U6 est transcrit dans le noyau par l'ARN polymérase III et sa coiffe est γ -méthylée (Fig.11) (Kunkel *et al.*, 1986). L'assemblage de la snRNP U6 se déroule entièrement dans le noyau (pour revue, Will et Luhrmann, 2001). Le snRNA U6 possède en 3' une séquence courte et riche en résidus uridine sur laquelle vient se fixer la protéine La (pour « Lupus antigen ») (Rinke et Steitz, 1985). Ceci permet la stabilisation du UsnRNA et facilite l'assemblage de la snRNP (pour revue, Wolin et

Cedervall, 2002). Après fixation d'un phosphate 2'-3' cyclique sur le ribose en 3' du snRNA U6 (Lund et Dahlberg, 1992), la protéine La est remplacée par les protéines Sm-like (LSm2 à 8) (Seraphin, 1995 ; Mayes *et al.*, 1999 ; Salgado-Garrido *et al.*, 1999). Contrairement aux autres UsnRNA, les modifications post-transcriptionnelles du snRNA U6 ne seraient pas réalisées dans les corps de Cajal, mais dans le nucléole. Le snRNA U6 n'a pas de site Sm et la formation de la snRNP U6 implique l'association de 7 protéines Sm-like (LSm2, LSm3, LSm4, LSm5, LSm6, LSm7 et LSm8) avec la séquence riche en U localisée à l'extrémité 3' du snRNA U6 (Vidal *et al.*, 1999)Vidal *et al.*, 1999. Le mode d'association des protéines LSm au snRNA U6, ou encore le rôle du complexe SMN dans la biogenèse de la snRNP U6 n'est pas encore bien connu. Enfin, la particule snRNP U6 interagit avec la snRNP U4 par la formation de deux hélices hétérologues entre les deux ARN pour donner la di-snRNP U4/U6 (Hashimoto et Steitz, 1984 ; Mougin *et al.*, 2002).

I.3.b. Protéines associées aux complexes spliceosomaux

Un grand nombre de protéines est requis pour le mécanisme de base de l'épissage. En effet, la fixation des UsnRNP sur le pré-ARNm fait intervenir un réseau très complexe d'interactions ARN-ARN, ARN-protéine et protéine-protéine, impliquant des protéines appartenant ou non aux UsnRNP. Elles participent à des fonctions indispensables, telles que la reconnaissance des sites 5' et 3' d'épissage, la mise en place et l'architecture des complexes spliceosomaux, ainsi que leurs transitions structurales. La formation des complexes spliceosomaux et leur activité mettent en œuvre une dynamique exceptionnelle d'associations et de dissociations de complexes protéiques et ribonucléoprotéiques (Tableau I) (pour revue, Wahl *et al.*, 2009).

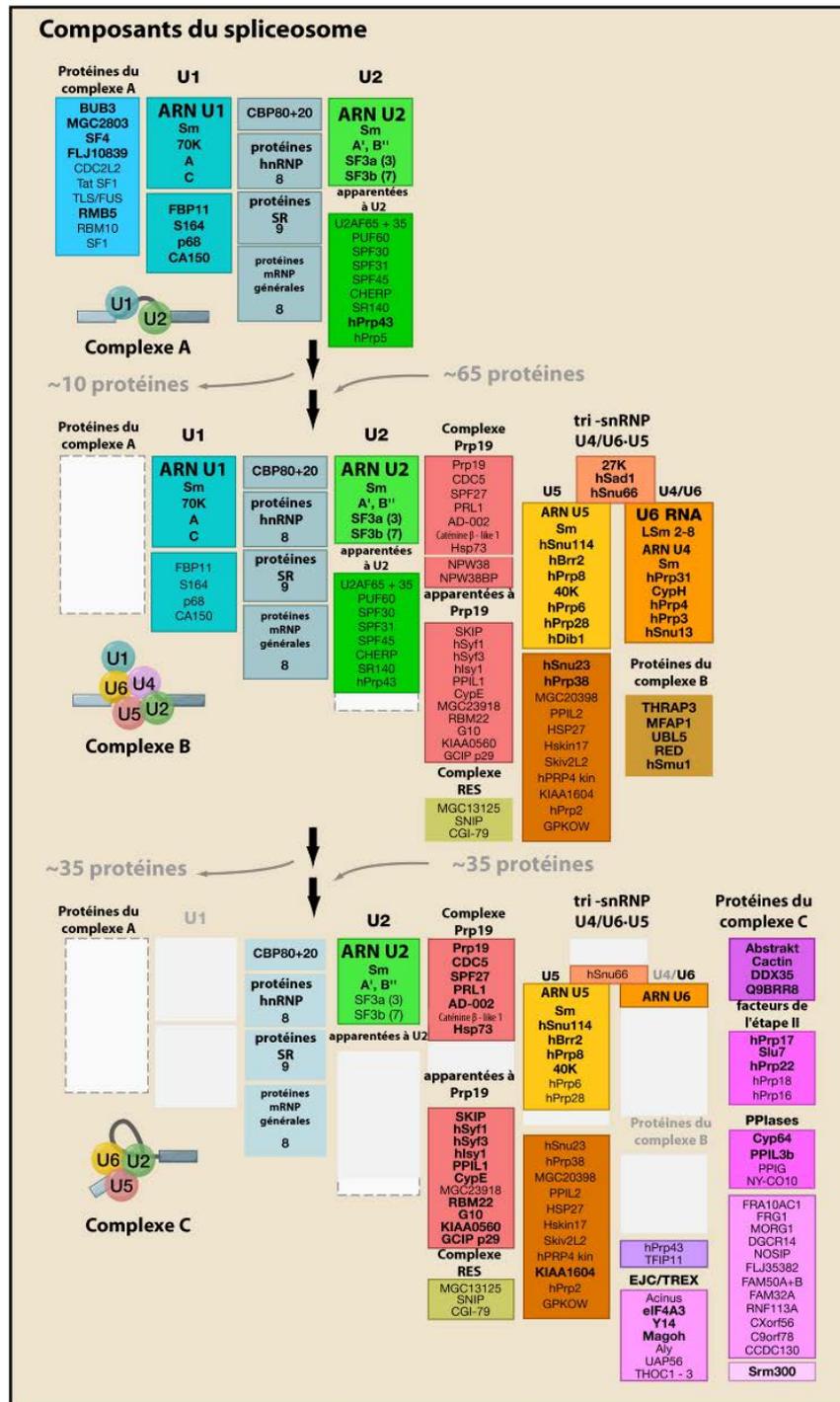


Tableau I : Dynamique de la composition des complexes A, B et C du spliceosome chez l'Homme.

Les protéines (nomenclature chez l'Homme) sont groupées selon leur fonction, leur association stable avec les snRNP, ou leur présence dans un autre complexe spliceosomal, comme indiqué. La composition en protéines des complexes a été déterminée par spectrométrie de masse (voir Wahl *et al.*, 2009). Les protéines présentes en quantités sub-stœchiométriques, conformément au nombre de peptides détecté par spectrométrie de masse, sont grisées. Les protéines SR identifiées sont ASF/SF2, 9G8, SRp20, SRp30c, SRp38, SRp40, SRp55 et SRp75 et la protéine SR-like Tra2- α . Les protéines hnRNP identifiées sont les hnRNP A1, A3, C, G, K, M, U, et PCBP1. (Tiré de Wahl *et al.*, 2009).

I.3.c. Assemblage du spliceosome et catalyse de la réaction

L'assemblage du spliceosome autour de l'intron se fait de manière séquentielle. Ainsi chez les vertébrés, les premières étapes de l'épissage consistent à reconnaître les sites donneur et accepteur d'épissage. S'assemble ensuite le reste de la machinerie d'épissage qui confère au pré-ARNm une configuration spatiale adéquate pour permettre la réaction d'épissage. Les études *in vitro* de l'assemblage du spliceosome sur le pré-ARNm ont permis de mettre en évidence la formation successive de 5 complexes ayant une mobilité électrophorétique différente lorsqu'ils sont fractionnés en gel non dénaturant (Fig.12) (Jurica *et al.*, 2002).

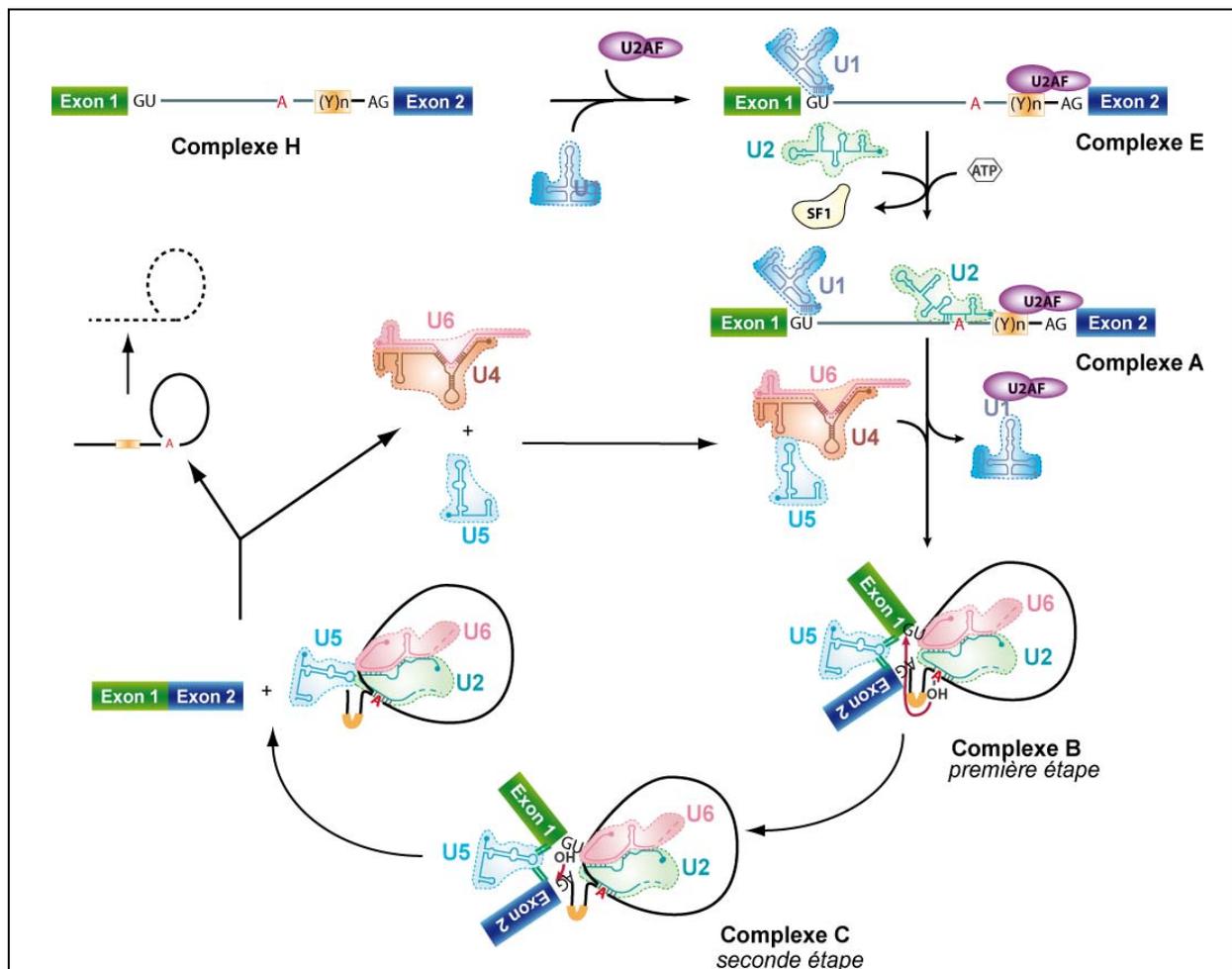


Figure 12 : L'assemblage séquentiel du spliceosome autour des introns majeurs.

L'assemblage du spliceosome commence par l'association de la UsnRNP U1 avec le site 5' d'épissage et du facteur U2AF avec la séquence polypyrimidine située au site 3'. Ce facteur facilite ensuite la fixation de la UsnRNP U2, qui a lieu en présence d'ATP. Toutes les étapes suivantes nécessitent de l'ATP. Après association de la triparticule [U4/U6-U5], la UsnRNP U1 est dissociée ainsi que le UsnRNP U4 (formation du complexe B, dans lequel a lieu la première étape de la réaction d'épissage). Les UsnRNA U2 et U6 interagissent entre eux et avec l'intron (au niveau de la boîte de branchement et du site 5' d'épissage). Après la première réaction d'épissage, des réarrangements structuraux ont lieu conduisant au complexe C dans lequel 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 formé et l'intron sous forme de lasso sont libérés. Y : Pyrimidine.

Ces complexes H, E, A, B et C correspondent à l'assemblage ordonné des différentes UsnRNP et protéines associées sur le pré-ARNm. Ainsi, dès leur synthèse, les transcrits Pol II sont associés à un ensemble de protéines formant un complexe ribonucléoprotéique appelé complexe H. Le processus dynamique est ensuite initié par la reconnaissance et l'interaction de la snRNP U1 avec le site 5' donneur. Le complexe ainsi formé est appelé complexe d'engagement ou complexe E. En présence d'ATP, la snRNP U2 interagit avec la boîte de branchement pour former avec la snRNP U1 le pré-spliceosome (complexe A). La tri-snRNP [U4/U6-U5] s'associe alors au complexe A pour former le complexe B* qui subit des réarrangements structuraux nécessaires à la formation du complexe B capable de catalyser la 1^{ère} étape de la réaction d'épissage (pour revue, Wahl *et al.*, 2009). Lors de l'association de la tri-snRNP [U4/U6-U5], la particule snRNP U1 se dissocie, ce qui permet l'interaction du snRNA U6 avec le site 5' d'épissage. L'interaction entre les snRNA U4 et U6 est dissociée au profit d'une interaction entre les snRNA U2 et U6. Ces transitions structurales s'accompagnent d'un remaniement des protéines. La 1^{ère} réaction a lieu. Elle est suivie de nouvelles transitions structurales. Enfin, la boucle terminale du snRNA U5 interagit avec les 2 extrémités des exons, ce qui permet de les rapprocher afin de les lier.

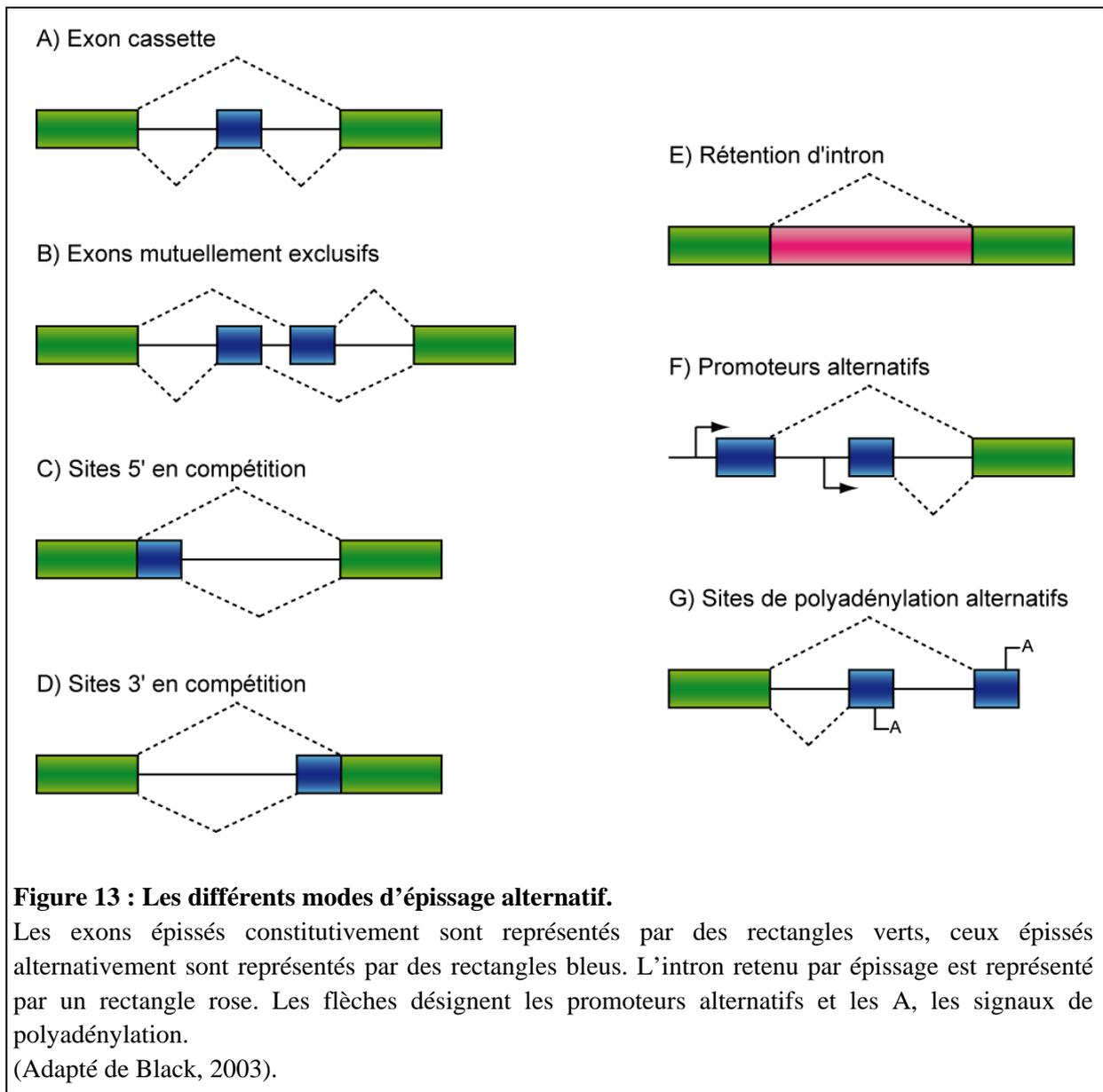
La fixation des UsnRNP sur le pré-ARNm fait intervenir un réseau complexe d'interactions ARN-ARN, ARN-protéine et protéine-protéine, impliquant des protéines appartenant ou non aux UsnRNP. Ainsi, au total, près de 300 protéines s'assemblent sur le pré-ARNm pour former le spliceosome (pour revue, Wahl *et al.*, 2009). Deux familles de protéines, les protéines SR et hnRNP jouent un rôle central dans l'ensemble du processus d'épissage, elles feront l'objet d'un paragraphe particulier plus loin dans cette introduction.

II. L'épissage alternatif

Comme nous l'avons précisé précédemment, les eucaryotes supérieurs et les virus utilisent l'épissage alternatif pour accroître la capacité codante de leur génome. Chez l'Homme, l'épissage alternatif est un phénomène majeur, car il concerne plus de 60% des 30 000 gènes (Modrek et Lee, 2002). L'altération de cet épissage peut conduire à de graves maladies génétiques dont nous verrons quelques exemples.

II.1. Les différents types d'épissage alternatif

L'épissage alternatif s'appuie sur l'utilisation de sites 5' donneurs et 3' accepteurs d'épissage. Les profils d'utilisation de ces sites donnent lieu à plusieurs types d'épissage (Fig.13) (Black, 2003).



II.1.a. Les exons dits « cassettes » (Fig.13A)

C'est le cas le plus fréquemment rencontré. Ces exons sont soit inclus, soit exclus de l'ARN messager mature. Ils sont en général de petite taille, exprimés de manière tissu-spécifique et codent un domaine supplémentaire permettant d'obtenir une isoforme de la protéine.

Par exemple, chez l'Homme, la protéine neurone-spécifique Tau possède (selon le stade de développement) 3 ou 4 domaines de fixation aux microtubules, ce qui lui confère une capacité différente à se lier aux microtubules en fonction de l'inclusion ou non de l'exon 10. Ainsi, lors de l'embryogenèse, l'isoforme fœtale sans exon 10 conduit à une flexibilité et un dynamisme des microtubules facilitant la migration cellulaire et la mise en place de connexions nerveuses. Après la naissance, le passage à l'isoforme contenant l'exon 10 est

corrélé avec une stabilité plus importante du réseau microtubulaire et une diminution de la plasticité du cytosquelette (Goedert *et al.*, 1989 ; Kosik *et al.*, 1989). Cette propriété est nécessaire pour le maintien des structures déjà mises en place. Une dérégulation de la balance normale entre ces isoformes peut causer différentes tautopathies (Sergeant *et al.*, 2005).

II.1.b. Les exons dits mutuellement exclusifs (Fig.13B)

Ces exons ne sont jamais présents ensemble dans les ARNm matures, bien qu'ils aient des sites 5' et 3' fonctionnels.

Par exemple, le précurseur du gène de la β -tropomyosine de poulet subit un épissage alternatif au niveau des exons 6A et 6B selon les tissus pour donner des messagers codant pour différentes isoformes de cette protéine. L'exon 6B est inclus dans les cellules musculaires squelettiques tandis que l'exon 6A est présent dans tous les autres tissus (pour revue, Gallego *et al.*, 1996).

II.1.c. Les exons variables du fait de la présence de sites d'épissage 5' et 3' alternatifs (Fig.13C et D)

Par l'utilisation de différents sites 5' et 3' d'épissage, les exons peuvent être raccourcis ou rallongés en taille. Des sites d'épissage alternatifs en compétition les uns avec les autres peuvent être présents dans une séquence exonique et la sélection alternative de ces sites génère des isoformes où une ou plusieurs parties de l'exon sont incluses ou exclues de l'ARNm.

Par exemple, l'unité E1A de l'adénovirus contient 3 sites 5' en compétition les uns avec les autres et un seul site 3'. Leur utilisation permet donc la production de trois ARNm 9S, 12S et 13S (pour revue, Akusjarvi et Stevenin, 2003).

II.1.d. Les exons résultant d'une rétention d'intron (Fig.13E)

Parfois, l'intron peut ne pas être éliminé du transcrit mature. Dans ce cas, il fait partie de l'ARN messager qui est transporté dans le cytoplasme pour la traduction. Ce phénomène, généralement tissu-spécifique, résulte de l'échec de la machinerie d'épissage à reconnaître des sites d'épissage faibles flanquant les introns (Kim, 2008). La rétention d'intron est le type d'épissage alternatif le moins connu et le moins étudié car l'ARNm obtenu est confondu avec des pré-ARNm partiellement ou non épissés. Seuls 2 ou 3 cas d'événements de rétention d'introns sont connus pour avoir des conséquences biologiques avérées.

Ainsi, une rétention anormale de l'intron 2 du pré-ARNm du canal à chlore (Clc-1), a été observée chez les patients atteints de DM. Cette rétention d'intron entraîne une dégradation de l'ARNm et donc à une perte de fonction de Clc-1, cette perte de fonction étant responsable de l'un des symptômes les plus courants de la DM, la myotonie (Charlet *et al.*, 2002).

II.1.e. Les exons terminaux alternatifs du fait de la présence de promoteurs ou de sites de polyadénylation alternatifs (Fig.13F et G)

L'initiation de la transcription au niveau de promoteurs différents, dits alternatifs, donne lieu à différents ARN messagers, de même l'épissage peut également conduire à l'utilisation de sites de polyadénylation différents.

C'est le cas du transcrit primaire *CT/CGRP* commun à la calcitonine et au neuropeptide CGRP. Dans la thyroïde, l'utilisation du site de polyadénylation de l'exon 4 conduit à la production de la calcitonine, tandis que dans les neurones, le neuropeptide CGRP est produit par utilisation du site de polyadénylation de l'exon 6 (pour revue, Lou et Gagel, 1998).

Plusieurs types d'épissage alternatif peuvent être combinés au sein d'un même pré-ARNm. Comme nous allons le voir par la suite, la production des différentes isoformes d'ARNm est régulée par la nature des sites d'épissage, la présence ou non de séquences régulatrices en *cis* et la concentration cellulaire des facteurs d'épissage (facteurs en *trans*) qui se lient à ces séquences régulatrices.

II.2. Régulation de l'épissage alternatif

II.2.a. Les sites d'épissage alternatifs sont souvent des sites sous-optimaux

Chez l'Homme, la taille moyenne d'un exon est de 140 nt, tandis que les introns font en moyenne 3500 nt (Wang et Burge, 2008 ; Deutsch et Long, 1999) et cette taille peut aller jusqu'à 500 000 nt (Rowen *et al.*, 2002). On peut donc imaginer le défi que représente pour la machinerie d'épissage l'identification des exons au sein du génome. Ceci d'autant plus que la présence de séquences 5', 3' et de points de branchement dits sous-optimaux, rend moins efficace l'interaction des composants spliceosomaux avec les jonctions exon/intron et intron/exon. Il est aussi possible que de multiples copies de ces sites apparaissent par hasard dans les introns de grande taille. D'autres informations et interactions sont donc nécessaires pour activer ces sites et pour définir les exons. Généralement, des protéines du spliceosome

interagissent avec des séquences introniques ou exoniques, les éléments *cis* régulateurs, afin de stimuler ou inhiber l'excision des introns.

II.2.b. Les éléments *cis* régulateurs

Les pré-ARNm présentant des sites d'épissage faibles, car très dégénérés, contiennent des éléments de séquences introniques et/ou exoniques localisés en amont ou en aval de l'exon régulé et qui agissent en *cis* pour stimuler ou au contraire, réprimer l'épissage. Ces séquences interagissent en général avec des facteurs protéiques *trans*-régulateurs dont nous présenterons le mécanisme d'action dans le Chapitre III.

Les séquences activatrices d'épissage peuvent être exoniques, ce sont alors des ESE (pour « Exonic Splicing Enhancer »), ou introniques, on parle alors d'ISE (pour « Intronic Splicing Enhancer »). A l'inverse, il existe aussi des séquences inhibitrices exoniques, appelées ESS (pour « Exonic Splicing Silencer »), ou introniques, appelées ISS (pour « Intronic Splicing Silencer »). Ces éléments régulateurs agissent par l'intermédiaire de leurs protéines partenaires. Les protéines hnRNP sont généralement associées aux éléments répresseurs, tandis que les protéines SR sont généralement associées aux éléments activateurs, mais nous verrons par la suite qu'il existe quelques exemples contraires.

Ces éléments agissant en *cis* ne sont pas spécifiquement associés à des sites utilisés de manière alternative. D'une manière générale, la distribution des séquences activatrices et inhibitrices varie en fonction des exons constitutifs, des exons alternatifs et des introns (pour revue, Blencowe, 2006). En effet, dans les introns on trouve une forte proportion de séquences inhibitrices tandis que dans les exons constitutifs, on trouve une forte proportion de séquences activatrices. Dans les exons alternatifs, les séquences activatrices et les séquences inhibitrices sont représentées de manière équivalente (pour revue, Blencowe, 2006).

II.2.c. Autres paramètres influençant les régulations de l'épissage alternatif

D'autres paramètres peuvent également jouer un rôle déterminant dans le choix des sites d'épissage. Ainsi, la présence de structures secondaires peut limiter l'accessibilité des sites d'épissage, comme c'est le cas pour l'exon 6B du gène de la β -tropomyosine aviaire. Dans les cellules non musculaires ou musculaires lisses, la structuration considérable du pré-ARNm de part et d'autre de l'exon 6B interfère avec la fixation de la machinerie d'épissage, ce qui provoque l'exclusion de l'exon de l'ARNm mature (Sirand-Pugnet *et al.*, 1995). De plus, certaines protéines régulatrices peuvent influencer la structuration des sites d'épissage et provoquer de ce fait l'inhibition ou l'activation de leur utilisation (Hiller *et al.*, 2007). La

taille des exons semble également jouer un rôle dans la régulation de l'épissage. En effet, les exons internes chez l'Homme ont une taille d'environ 140 nucléotides et moins de 5 % dépassent 250 nucléotides. D'après des analyses statistiques réalisées sur les séquences d'exons alternatifs en cassette, il a été montré que les exons alternatifs sont significativement plus courts que les exons constitutifs et que l'épissage alternatif est beaucoup plus fréquent dans les micro-exons (exon inférieur à 30 bases) (Sorek *et al.*, 2004).

III. Régulateurs de l'épissage constitutif et alternatif

Certains facteurs d'épissage peuvent être impliqués aussi bien dans l'épissage constitutif qu'alternatif. D'autres, par contre, ne sont impliqués que dans l'épissage alternatif et sont nommés « facteurs régulateurs ». Certains facteurs d'épissage sont ubiquitaires, tandis que d'autres, et notamment les facteurs régulateurs, sont tissus spécifiques ou encore régulés au cours du développement. Dans tous les cas, il s'agit d'actions pour empêcher l'inclusion d'exons ou la promouvoir. On parle alors respectivement de facteurs répresseurs et activateurs.

III.1. Protéines nucléaires jouant un rôle majeur dans la régulation de l'épissage constitutif et alternatif

III.1.a. Les protéines SR

Les protéines SR (Serine Arginine rich) constituent la famille la plus étudiée des facteurs d'épissage car elles sont essentielles à l'épissage constitutif. Ces protéines sont très conservées phylogénétiquement particulièrement chez les métazoaires et chez les plantes. Néanmoins, certains eucaryotes n'expriment pas de protéine SR, comme *S. cerevisiae* (pour revue, Graveley, 2000). Certains de ses membres sont impliqués aussi bien dans l'épissage constitutif qu'alternatif. D'autres sont considérés comme des facteurs régulateurs jouant des rôles spécifiques dans l'épissage alternatif (Long et Caceres, 2009). Cette famille comprend douze membres appelés SRSF1 à 12 (Zahler *et al.*, 1992 ; Long et Caceres, 2009).

Les premières études présentaient les protéines SR comme ayant une fonction redondante *in vitro* vis-à-vis de la capacité à épisser un pré-ARNm, puisqu'elles peuvent toutes compléter un extrait cytoplasmique S100. Pourtant, *in vivo*, la forte conservation entre les espèces qui possèdent ces protéines, des localisations cellulaires spécifiques à chaque protéine SR et de leur spécificité de substrat semblent aller dans le sens d'un rôle spécifique à chacune. Le fait que l'ajout de protéines SR dans un extrait S100 permette

l'épissage montre que les protéines SR ont en tout cas un rôle commun dans l'épissage (Fu, 1993).

Des études récentes tendent à montrer l'implication de plusieurs de ces protéines dans des processus autres que l'épissage, tels que la transcription (Lin *et al.*, 2008 , Fededa et Kornblihtt, 2008), la traduction (Sanford *et al.*, 2004 , Michlewski *et al.*, 2008 , Bedard *et al.*, 2007 , Swartz *et al.*, 2007), le transport (Huang *et al.*, 2003 , Huang et Steitz, 2001) et la dégradation des ARN (Zhang et Krainer, 2004).

- Caractéristiques communes des protéines SR

La caractéristique principale des protéines SR est leur ressemblance structurale (pour revue, Graveley, 2000). En effet, toutes sont composées d'un ou deux domaines de liaison à l'ARN de type RRM (pour « RNA Recognition Motif) en N-terminal, suivi d'un domaine riche en dipeptides arginine-serine, appelé domaine RS en C-terminal (Fig.14). Les protéines SR peuvent être divisées en deux sous-familles selon leur nombre de domaines RRM (Mayeda *et al.*, 1999). Le premier regroupe les protéines ayant un motif RRM classique avec RNP-1 et RNP-2 le second plus atypique incluant un heptapeptide caractéristique (SRSF1 et SRSF5 par exemple). L'autre groupe ne contient que les protéines à un seul domaine RRM (SRSF2 et SRSF7 entre autres).

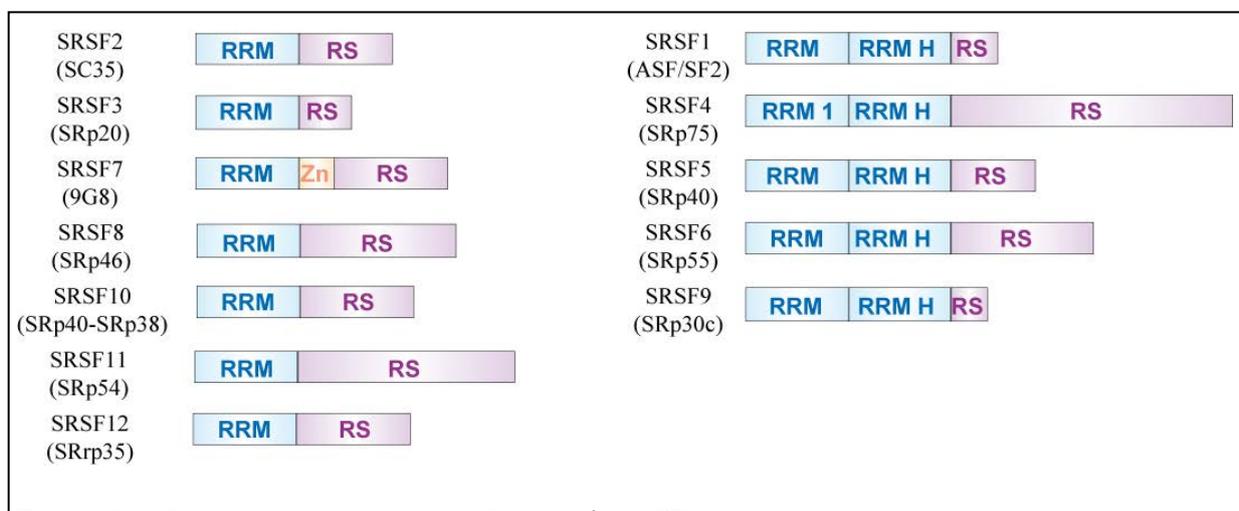


Figure 14 : Organisation structurale des protéines SR.

Représentation schématique de l'organisation modulaire des protéines SR humaines. RRM : motif de reconnaissance de l'ARN ; RRM H : domaine RRM homologue ; Z : domaine doigt de zinc ; RS : domaine riche en dipeptide arginine/sérine. La taille du domaine RS représenté est proportionnelle à la longueur du domaine donné.

(Adapté de Graveley, 2000).

Les protéines SR sont généralement associées à des séquences ESE. En effet les protéines SR ont majoritairement une influence positive sur l'épissage et se fixent essentiellement dans les exons. Afin de comprendre les fonctions des protéines SR, plusieurs études, en particulier par la méthode SELEX (pour Systematic Evolution of Ligands by EXponential enrichment), ont tenté d'élucider les spécificités de liaison de chaque protéine SR. Le résultat global de ces expériences est que les protéines SR reconnaissent un vaste profil de séquences ARN (Tableau II). Les consensus des sites de fixation sont souvent dégénérés et dans beaucoup de cas, le motif reconnu par une protéine SR donnée est également reconnu par d'autres protéines SR (Liu *et al.*, 1998; Tacke et Manley, 1995).

Protéine SR	Site de fixation	Méthode d'identification
SRSF1	RGAAGAAC AGGACRRAGC SRSASGA UGRWG	SELEX SELEX SELEX fonctionnel CLIP
SRSF2	AGSAGAGUA GUUCGAGUA GRYYCSYR UGUUCSAGWU GWUWCCUGCUA GGGUAUGCUG GAGCAGUAGKS AGGAGAU UGCNGYY	SELEX SELEX SELEX fonctionnel SELEX SELEX SELEX SELEX SELEX SELEX fonctionnel
SRSF3	GGUCCUCUUC WCWWC CUCKUCY	Gel shift Test d'épissage Affinité à l'ARN
SRSF4	GAAGGA	UV cross-linking
SRSF5	GAGCAGUCGGCUC ACDGS	SELEX SELEX fonctionnel
SRSF6	USCGKM UCAACCAGGCGAC	SELEX fonctionnel SELEX
SRSF7	UCAACA ACGAGAGAY GGACGACGAG	UV cross-linking SELEX SELEX fonctionnel
SRSF9	GACGAC AAAGAGCUCGG CUGGAUU	SELEX fonctionnel SELEX fonctionnel Gel shift
SRSF11	Riche en C	UV cross-linking

N: tout nucléotide; R: purine; Y: pyrimidine; D: A, G ou U; K: U ou G; M: A ou C; S: G ou C; W: A ou U

Tableau II : Les séquences ARN reconnues par les protéines SR.

(D'après Long et Caceres, 2009).

Le domaine RS riche en enchaînements de sérines et d'arginines permet d'établir des interactions protéine/protéine par lesquelles les protéines SR s'associent entre elles ou avec les différents composants de la machinerie d'épissage contenant aussi un domaine RS (Wu et Maniatis, 1993). Néanmoins, il est apparu récemment que ce domaine pourrait aussi permettre une interaction avec l'ARN afin de définir les différents signaux d'épissage et faciliter leur reconnaissance par les composants du spliceosome (Shen et Green, 2004). De plus, le domaine RS contient généralement le signal de localisation de la protéine, mais celui-ci peut parfois s'étendre jusqu'au domaine de fixation à l'ARN. Le signal de localisation nucléaire permet aux protéines de rejoindre les speckles Lamond et Spector, 2003.

Le domaine RS est sujet à de nombreux événements de phosphorylation qui permettent de réguler l'activité des protéines, mais aussi parfois leur localisation (Sanford et Bruzik, 2001). La phosphorylation de ces protéines est indispensable à l'assemblage du spliceosome (Roscigno et Garcia-Blanco, 1995 ; Xiao et Manley, 1997). Par contre, l'addition d'inhibiteurs de phosphatases bloque la réaction d'épissage, indiquant que les protéines SR doivent être déphosphorylées pour que la réaction biochimique ait lieu (Mermoud *et al.*, 1994). Ainsi, les protéines SR subissent de multiples cycles de phosphorylation/déphosphorylation indispensables à l'épissage. Aujourd'hui, plusieurs familles de protéines kinases sont connues pour phosphoryler les protéines SR, telles que la SRPK1 et 2 (pour « SR protein kinase »), Clk/Sty ou encore la kinase Akt (Gui *et al.*, 1994; Wang *et al.*, 1998 , Prasad *et al.*, 1999 ; Colwill *et al.*, 1996, Patel *et al.*, 2005) et de façon inattendue, l'ADN topoisomérase I (Rossi *et al.*, 1996).

- Les protéines « SR-like »

Plusieurs autres facteurs, distincts des protéines SR et impliqués dans différents aspects du métabolisme de l'ARN pré-messager tels que l'épissage mais aussi le remodelage de la chromatine, la transcription par l'ARN Pol II ou encore la progression du cycle cellulaire, possèdent aussi un domaine RS (Boucher *et al.*, 2001 ; pour revue, Long et Caceres, 2009). Ces protéines ne contiennent pas forcément un domaine de liaison à l'ARN de type RRM mais peuvent posséder par exemple un motif PIWI comme SRm160 (Blencowe *et al.*, 1998 ; Szymczyna *et al.*, 2003). De plus, leur domaine RS est généralement localisé à l'extrémité N-terminale (Fig.15). Ces protéines, classées dans la famille des protéines SR-like ou SRrp (pour « SR protein-related protein »), regroupent par exemple les deux sous-unités du facteur U2AF (U2AF35 et U2AF65), des composants des snRNP (U1-70K, U5-100K et U4/U6.U5-

27K), des co-activateurs d'épissage SRm 160 et 300, des régulateurs de l'épissage comme Tra et Tra2 ou encore des hélicases à ARN telles que hPrp16 (pour revue, Long et Caceres, 2009).

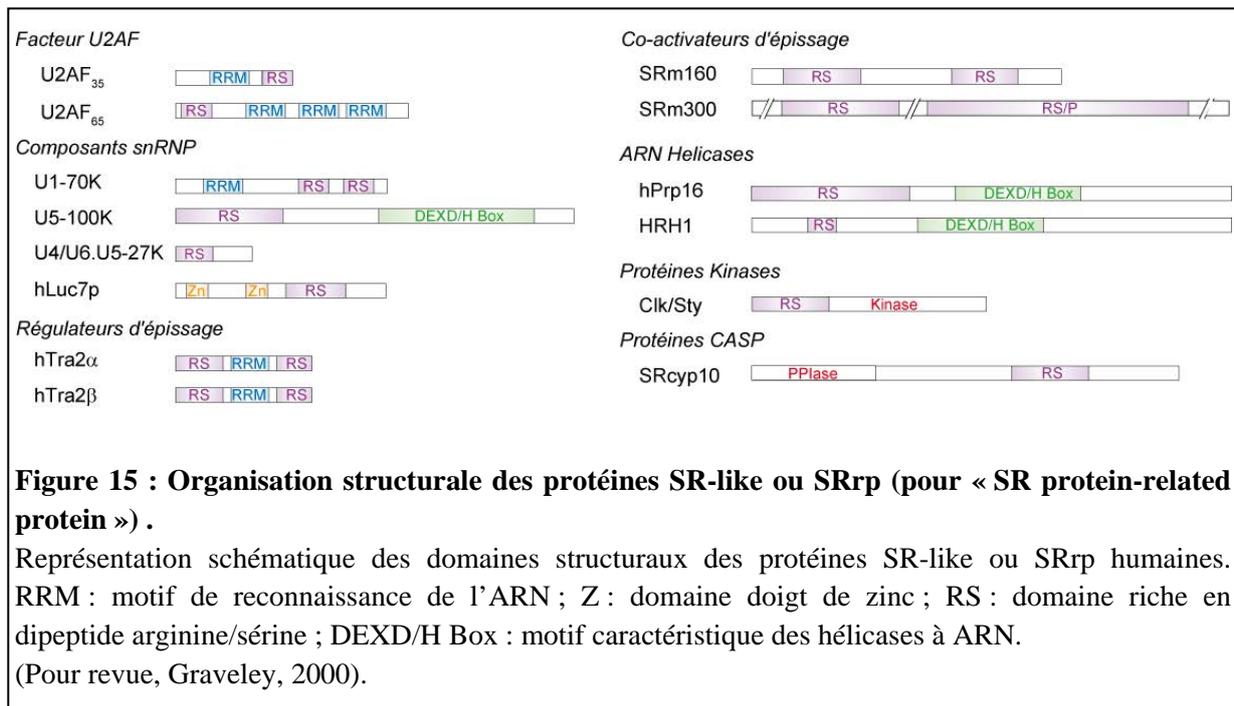


Figure 15 : Organisation structurale des protéines SR-like ou SRrp (pour « SR protein-related protein »).

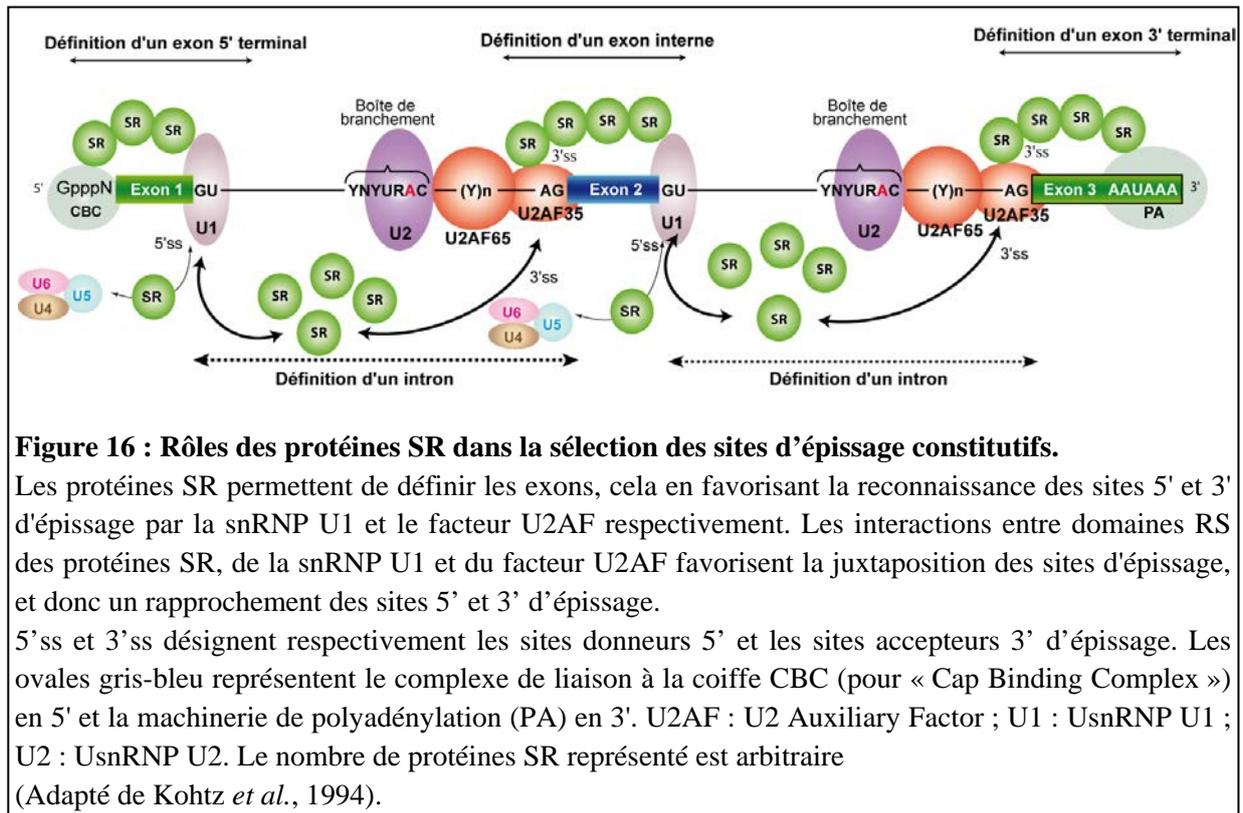
Représentation schématique des domaines structuraux des protéines SR-like ou SRrp humaines. RRM : motif de reconnaissance de l'ARN ; Z : domaine doigt de zinc ; RS : domaine riche en dipeptide arginine/sérine ; DEXD/H Box : motif caractéristique des hélicases à ARN.

(Pour revue, Graveley, 2000).

- Rôles des protéines SR dans l'épissage constitutif

Les données actuelles montrent que ce rôle se situe à différents niveaux (Fig.16) :

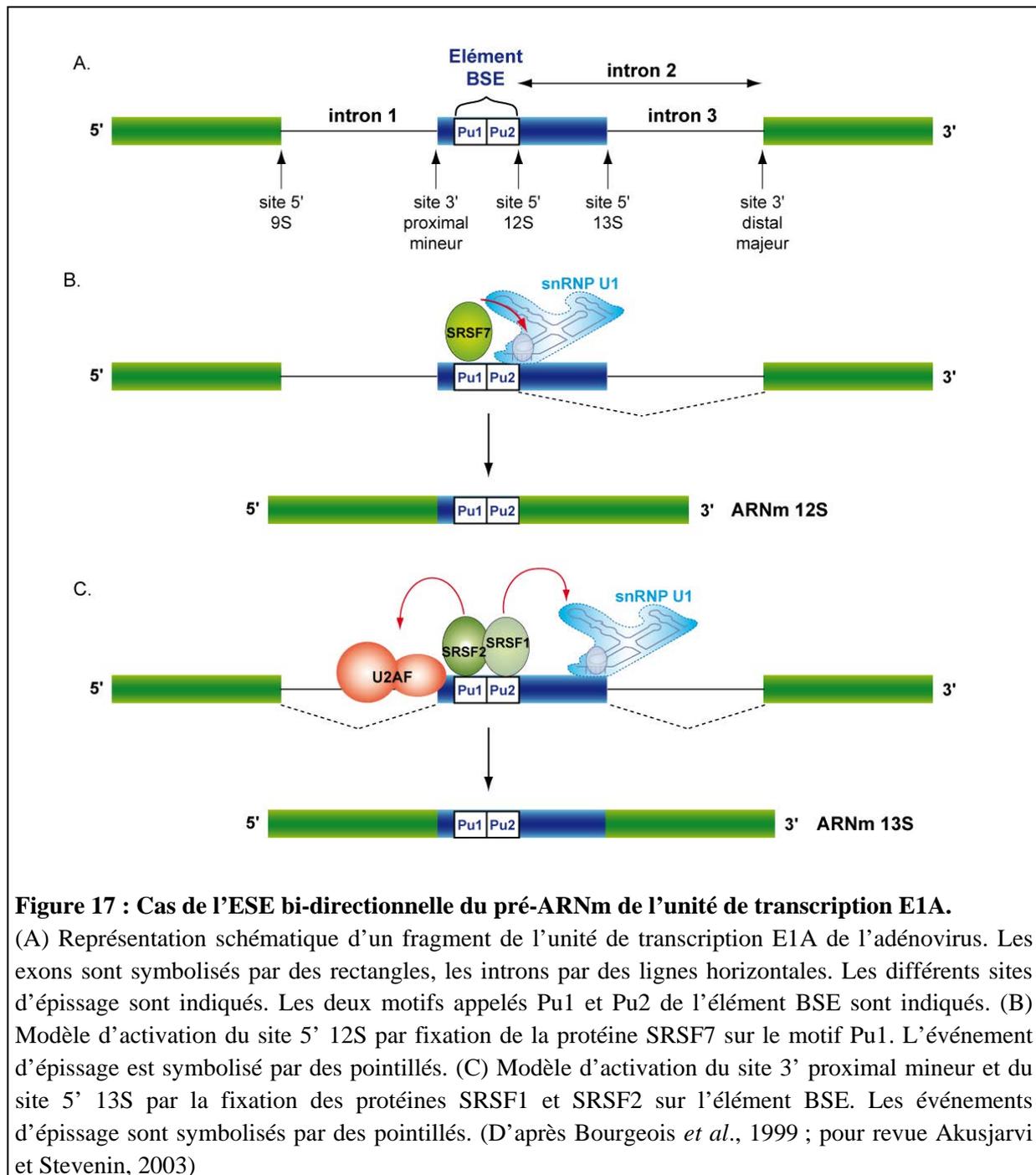
- avec la snRNP U1, elles participent à la définition du site 5'. Cependant, en fonction du degré de complémentarité, la particule peut aussi être stabilisée au niveau du site d'épissage grâce à de multiples interactions avec les protéines SR. Les protéines SR peuvent même, si elles sont présentes en très grande quantité, se substituer à elle (Crispino *et al.*, 1994 ; Tarn et Steitz, 1994 ; Tarn et Steitz, 1995),
- en se fixant au site 3', les protéines SR facilitent également la fixation de la snRNP U2 sur la séquence de branchement (Tarn et Steitz, 1995),
- en se fixant aux parties exoniques des sites 5' et 3', elles pourraient rapprocher ces extrémités pour faciliter la réaction d'épissage (Staknis et Reed, 1994),
- elles jouent également un rôle plus tardif dans l'épissage en escortant la tri-snRNP [U4/U6.U5] vers les précomplexes spliceosomaux (Rosigno et Garcia-Blanco, 1995).



- Rôle dans l'épissage alternatif par sélection du site d'épissage

Cas de l'élément bidirectionnel BSE (pour « Bidirectional Splicing Enhancer ») de l'unité E1A de l'adénovirus (Fig.17) : Certains éléments ESE sont capables de fixer des protéines SR différentes et de favoriser ainsi le choix d'un site par rapport à l'autre. Cette situation a été observée par l'équipe de J. Stevenin pour le pré-ARNm de l'unité E1A de l'adénovirus (Bourgeois *et al.*, 1999 ; pour revue, Akusjarvi et Stevenin, 2003).

L'unité de transcription E1A renferme 3 sites 5' et 2 sites 3' d'épissage. Les 3 sites 5' (9S, 12S et 13S) peuvent chacun être utilisés avec le site 3' distal, et entrent donc en compétition les uns avec les autres, l'événement d'épissage majoritaire étant l'épissage au niveau du site 5' qui conduit à l'ARN 13S. L'élément BSE est localisé immédiatement en amont du site donneur 12S. Il est constitué de 2 séquences riches en purines, Pu1 et Pu2 (Fig.17A). L'élément Pu1 active le site 5' 12S en fixant la protéine SRSF7, ce qui permet de stabiliser la fixation de la snRNP U1 sur le site 5' 12S et conduit à l'élimination de l'intron 2 (Fig.17B). En revanche, l'élément Pu1 associé à Pu2 active l'élimination des introns 1 et 3 en fixant les protéines SRSF1 et SRSF2. Celles-ci participeraient à la définition de l'exon délimité par le site 3' proximal et le site 5' 13S, ceci en favorisant d'une part la fixation du facteur U2AF au site 3' proximal et d'autre part, la stabilisation de la snRNP U1 au site 5' 13S (Fig.17C). Les 2 voies d'activation par l'élément BSE sont mutuellement exclusives.



- Rôle dans l'épissage par antagonisme avec les protéines hnRNP

Un autre modèle proposé pour comprendre l'effet globalement activateur des protéines SR sur l'épissage est celui du phénomène d'antagonisme par rapport aux protéines hnRNP. Le rôle des protéines SR consiste alors à contrecarrer l'effet inhibiteur des protéines hnRNP liées à une séquence inhibitrice d'épissage ESS chevauchante ou juxtaposée à l'élément ESE. Dans ce cas, l'activation de l'épissage est tout à fait indépendante des domaines RS, seuls les

domaines RRM des protéines SR sont nécessaires. Nous reparlerons de ce phénomène dans le paragraphe consacré aux protéines hnRNP.

- Effet inhibiteur sur l'épissage

Bien que les protéines SR aient en majorité un effet activateur sur l'épissage, elles peuvent parfois avoir des effets inhibiteurs du fait de leur fixation sur des éléments de séquences requis pour l'assemblage des éléments du spliceosome ou de leur interaction avec certains de ces composants (pour revue, Pozzoli et Sironi, 2005). La position dans l'intron du site de fixation des protéines SR semble être l'élément déterminant de ce type d'inhibition. On peut citer l'exemple de l'élément intronique 3RE, riche en purines, qui réprime l'épissage du site accepteur IIIa du pré-messager E1A de l'adénovirus en phase précoce d'infection, par fixation de la protéine SRSF1 (Kanopka *et al.*, 1996 ; pour revue, Akusjarvi et Stevenin, 2003). Les auteurs avaient avancé l'hypothèse selon laquelle la fixation de SRSF1 à cette séquence empêcherait la fixation de la snRNP U2 sur le point de branchement. Des données postérieures montrent que l'inhibition est due à l'interaction entre le domaine RRM2 de SRSF1 et d'autres facteurs qui eux sont plus directement impliqués dans le mécanisme de répression (Dauksaite et Akusjarvi, 2002). Ce phénomène d'inhibition est lié à la localisation du site de fixation de SRSF1 dans l'intron car l'inhibition est conservée lorsque cet élément est transporté à une position similaire dans un intron du pré-ARNm de la β -globine de lapin, mais est au contraire perdue s'il est inséré dans un exon (Dauksaite et Akusjarvi, 2002).

Parmi les autres cas connus d'inhibition de l'épissage par une protéine SR, on peut citer :

- ✓ l'exclusion de l'exon 3b du pré-ARNm de la GTPase Rac 1 par la protéine SRSF3 (Goncalves *et al.*, 2009)
- ✓ l'exclusion de l'exon 8 du pré-ARNm de la protéine hnRNP A1 par la protéine SRSF9 fixée à la séquence régulatrice CE9 (Simard et Chabot, 2002)
- ✓ l'exclusion de l'exon 9 de CFTR par SRSF1, 4 et 6 (Buratti *et al.*, 2001)
- ✓ l'exclusion de l'exon 4 de CD45 par SRSF1 et 2 (Lemaire *et al.*, 1999).

III.1.b. Les protéines hnRNP

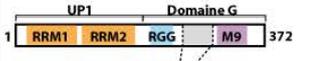
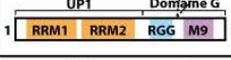
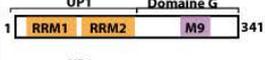
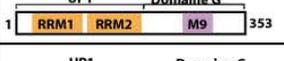
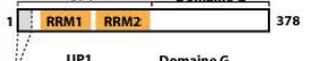
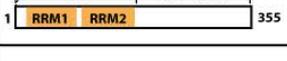
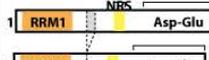
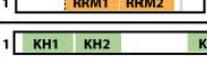
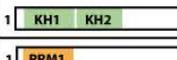
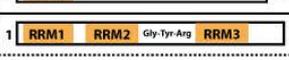
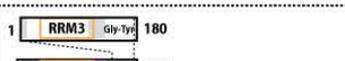
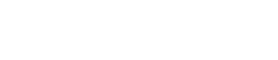
Les protéines hnRNP (pour « heterogeneous nuclear ribonucleoproteins ») forment une famille de protéines très abondantes dans la cellule et structurellement variées. Elles ont initialement été décrites à travers leur capacité à lier les ARN nouvellement transcrits par l'ARN polymérase II, entraînant la formation des particules hétérogènes appelés hnRNP (Dreyfuss, 1986 ; Dreyfuss *et al.*, 1988). Cette famille comprend au moins une vingtaine de

protéines et compte parmi ses membres hnRNP A1, B, C, F, G, H, K, et la hnRNP I, aussi appelée PTB (pour « Polypyrimidine Tract Binding »). Ces membres jouent des rôles variés, qui, au delà de la régulation de l'épissage, peuvent s'étendre à la transcription (Du *et al.*, 1998), à la maturation des extrémités 3' des ARNm (Bagga *et al.*, 1998 ; Kessler *et al.*, 1997) et à la traduction (Ostareck-Lederer et Ostareck, 2004 ; Collier *et al.*, 1998). L'une des protéines hnRNP la plus étudiée, la hnRNP A1, a même été décrite pour jouer un rôle dans la conservation de la longueur des télomères (LaBranche *et al.*, 1998 ; Dallaire *et al.*, 2000).

- Organisation structurale des hnRNP

Les protéines hnRNP (Tableau III) se fixent à l'ARN par leurs domaines de type RRM, RGG ou KH (Dreyfuss *et al.*, 1993). Elles ont toutes une affinité assez large pour l'ARN, avec néanmoins une affinité plus forte pour certains types de séquences : séquences riches en C pour les hnRNP E, I et K, séquences riches en G pour les hnRNP A/B, E, H, H', 2H9, M et U et séquences riches en U pour les protéines hnRNP C et PTB. Certaines protéines hnRNP ont, en plus, la capacité de reconnaître avec une forte affinité des séquences spécifiques (hnRNP A/B, M, 2H9, PTB) (pour revue, Dreyfuss *et al.*, 2002). Par exemple, la protéine hnRNP A1 se lie à l'ARN en général avec une affinité de l'ordre du micromolaire et reconnaît par ailleurs la séquence UAGGG(A/U) avec une affinité de 10^{-8} M (Burd et Dreyfuss, 1994).

Les protéines de la famille hnRNP A/B, D, U et H contiennent également un domaine d'oligomérisation ou domaine *Gly*, riche en résidus glycine, localisé en C-terminal. Des résidus phénylalanine et tyrosine dispersés de manière régulière entre les résidus glycine sont requis pour que les protéines de la famille hnRNP A/B interagissent entre elles. Ces résidus interviendraient par formation d'interactions hydrophobes entre les protéines. Un ensemble d'interactions peut être mis en place de manière coopérative, indépendamment de la fixation à l'ADN ou l'ARN, et conduire à une oligomérisation de ces protéines (Cartegni *et al.*, 1996). Par ailleurs, le domaine *Gly* de ces protéines, en particulier celui de la protéine hnRNP A1, a des propriétés d'hybridation des acides nucléiques, d'où un rôle possible de chaperonne des acides nucléiques pour la protéine hnRNP A1.

hnRNP	organisation structurale	PM (kDa)	L	séquence ARN	rôles cellulaires
A0		31,7	N	G ₄ (AU ₃) ₄ A	Contrôle de la traduction
A1B		34	N,C	UAGGG(A/U) (Py)UAG	Epissage, export des ARN Maintenance des télomères
A1		320			
A2		36	N,C	GCCAAGGAGCC	Epissage, localisation des ARNm, maintenance des télomères
B1		38			
A3(1)		41	N,C	GCCAAGGAGCC	Localisation des ARNm
A3(2)		39			
C2		41	N	riche en U, AGUA(U) ₅ GUGGA	Epissage, contrôle de la tra- duction, contrôle de la loca- lisation des pré-ARNm, maintenance des télomères
C1		43			
D(1)		45	N	riche en A/U, UUAG/G	Maintenance des télomères, stabilisation ou destabilisation des ARNm, recombinaison
D(2)		42			
D(3)		40			
D(4)		37			
E1		37,5	N,C	Poly(C), Poly(G)	Stabilité des ARNm, contrôle traduction
E2		360			
G		42	N	AAGU	Epissage
F		45,5	N	Poly(G)	Epissage
H		56	N	GGGA/C, poly(G)	Epissage, polyadénylation
H'		56	N		
2H9(6)		15,5	N	GGGA	Epissage
2H9(5)		16			
2H9(4)		22			
2H9(1)		37			
2H9(2)		35			
2H9(3)		31,5			
PTB(1)		59	N	riche en C/U, UCUU, UCUCU, CUUCUCUCU	Epissage, localisation des ARNm, polyadénylation
PTB(2)		550			
PTB(4)		556			

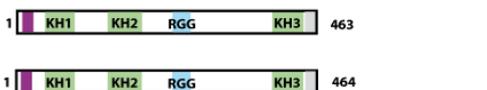
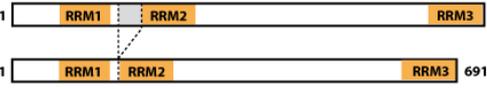
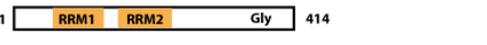
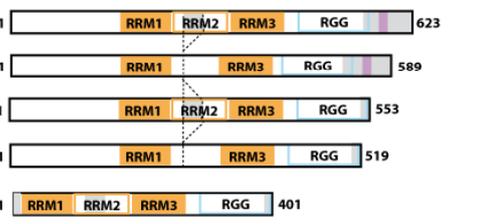
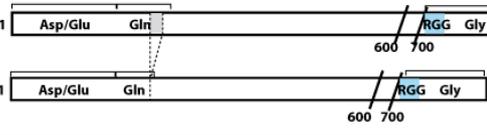
hnRNP	organisation structurale	PM (kDa)	L	séquence ARN	rôles cellulaires																				
K(1) K(2)		62	N,C	poly(C), AU(C) _{3/4} (U/A) ₂	Epissage, régulation de la transcription, contrôle de la traduction																				
L		60	N	répétitions CA	Epissage, export ARNm, stabilité ARNm																				
M(1) M(2)		74 72	N	GAGGAGG	Epissage																				
TDP-43		53	N	répétitions UG	Epissage, représeur transcriptionnel																				
TLS/ FUS		53,5	N	GGUG	Epissage, oncogenèse																				
Q3(1) Q2(2) Q1(3) Q(4) Q(5)		69,6	N,C		Epissage																				
R		70,9	N		Edition, transport des ARNm, polyadénylation																				
U(1) U(2)		120	N	poly(G), riche en A/U	Régulation de la transcription, rétention nucléaire																				
<table border="0"> <tr> <td> RRM</td> <td>domaine RRM</td> <td></td> <td>mutations ou délétions entre les isoformes</td> </tr> <tr> <td> KH</td> <td>domaine KH</td> <td> Gly</td> <td>domaine riche en résidus glycine</td> </tr> <tr> <td> RGG</td> <td>domaine RGG</td> <td> Asp/Glu</td> <td>domaine riche en résidus acide aspartique et glutamique</td> </tr> <tr> <td></td> <td>domaine à doigt de zinc</td> <td></td> <td>domaine NRS ("nuclear retention signal")</td> </tr> <tr> <td> M9</td> <td>domaine M9</td> <td></td> <td>domaine NLS ("nuclear localisation signal")</td> </tr> </table>						 RRM	domaine RRM		mutations ou délétions entre les isoformes	 KH	domaine KH	 Gly	domaine riche en résidus glycine	 RGG	domaine RGG	 Asp/Glu	domaine riche en résidus acide aspartique et glutamique		domaine à doigt de zinc		domaine NRS ("nuclear retention signal")	 M9	domaine M9		domaine NLS ("nuclear localisation signal")
 RRM	domaine RRM		mutations ou délétions entre les isoformes																						
 KH	domaine KH	 Gly	domaine riche en résidus glycine																						
 RGG	domaine RGG	 Asp/Glu	domaine riche en résidus acide aspartique et glutamique																						
	domaine à doigt de zinc		domaine NRS ("nuclear retention signal")																						
 M9	domaine M9		domaine NLS ("nuclear localisation signal")																						

Tableau III : Les protéines hnRNP : structures, rôles et séquences ARN reconnues.

Les différentes protéines hnRNP sont représentées. Chaque isoforme connue pour chaque protéine hnRNP est donnée, ainsi que son organisation structurale à l'échelle, son poids moléculaire prédictif (PM), sa localisation (L) nucléaire (N) et/ou cytoplasmique (C), la séquence ARN reconnue et sa fonction.

(D'après Swanson et Dreyfuss, 1988 ; Dreyfuss *et al.*, 1993 ; Honore *et al.*, 1999 ; Caputi et Zahler, 2001 ; Lerga *et al.*, 2001 ; pour revues, Krecic et Swanson, 1999 ; Dreyfuss *et al.*, 2002).

Bien que les protéines hnRNP soient localisées principalement dans le nucléoplasme des cellules, certaines d'entre elles (les protéines hnRNP A/B, D, E, I, et K) sont capables, comme les protéines SR, de transiter entre le noyau et le cytoplasme. Dans le noyau, elles

sont associées aux pré-ARNm ou à l'ADN pour certaines et dans le cytoplasme aux ARNm (Pinol-Roma et Dreyfuss, 1992 ; Michael *et al.*, 1995). Le motif M9, dans la région C-terminale des protéines hnRNP A1/A1B et A2/B1, constitue un signal d'import nucléaire spécifique (pour revue, Nakielny et Dreyfuss, 1999). Ce motif interagit avec un récepteur d'import nucléaire, la transportine (Pollard *et al.*, 1996). Certaines protéines hnRNP, strictement nucléaires comme la protéine hnRNP C, possèdent un signal de rétention nucléaire NRS31. Comme ces protéines se fixent préférentiellement aux séquences introniques des pré-ARNm (pour revue, Nakielny et Dreyfuss, 1999), elles éviteraient aux ARN non épissés d'être exportés dans le cytoplasme (Pinol-Roma et Dreyfuss, 1992). Ces protéines assureraient donc une rétention nucléaire des pré-ARNm.

- Caractéristiques communes

Pour la plupart des protéines hnRNP, différentes isoformes ont pu être isolées et sont issues de phénomène d'épissage alternatif. Par exemple, les protéines de la famille hnRNP A/B sont issues de l'épissage alternatif des pré-ARNm produits à partir de 2 gènes distincts résultant probablement de la duplication d'un gène initial (Biamonti *et al.*, 1994). Chaque isoforme possède des caractéristiques définies de liaison à l'ARN et il n'y a pas de lien direct entre efficacité de fixation à l'ARN et efficacité d'action dans l'épissage alternatif (Mayeda *et al.*, 1994). Autre exemple, le pré-ARNm de la protéine PTB subit un épissage alternatif conduisant à la production de 3 isoformes. L'isoforme PTB4 est l'isoforme majoritairement produite. Les 2 autres isoformes PTB1 et PTB2 sont également capables de réprimer l'épissage, mais plus faiblement. Les taux relatifs de ces différentes protéines PTB varient suivant les tissus, ce qui participe aux régulations tissu-spécifiques de l'épissage des pré-ARNm (Wollerton *et al.*, 2001 ; pour revue, Wagner et Garcia-Blanco, 2001).

De plus, les protéines hnRNP subissent de nombreuses modifications post-traductionnelles qui sont principalement : des phosphorylations sur les résidus sérines ou tyrosines, des méthylation sur les arginines et des sumoylations sur des résidus lysines. Des différences de taux de modifications post-traductionnelles entraînent des variations dans la capacité des protéines à interagir avec l'ARN. Par exemple, suite à des stress cellulaires (stress oxydatif, réponse à des facteurs de croissance), la protéine hnRNP K est phosphorylée sur des résidus sérines. Ces modifications structurales augmentent l'affinité de la protéine pour d'autres protéines (proto-oncogènes, protéines kinases, répresseurs transcriptionnels) et diminuent sa capacité à interagir avec l'ARN *in vitro* (Schullery *et al.*, 1999). De la même manière, la phosphorylation du peptide F, situé en aval du motif M9, gêne l'interaction avec

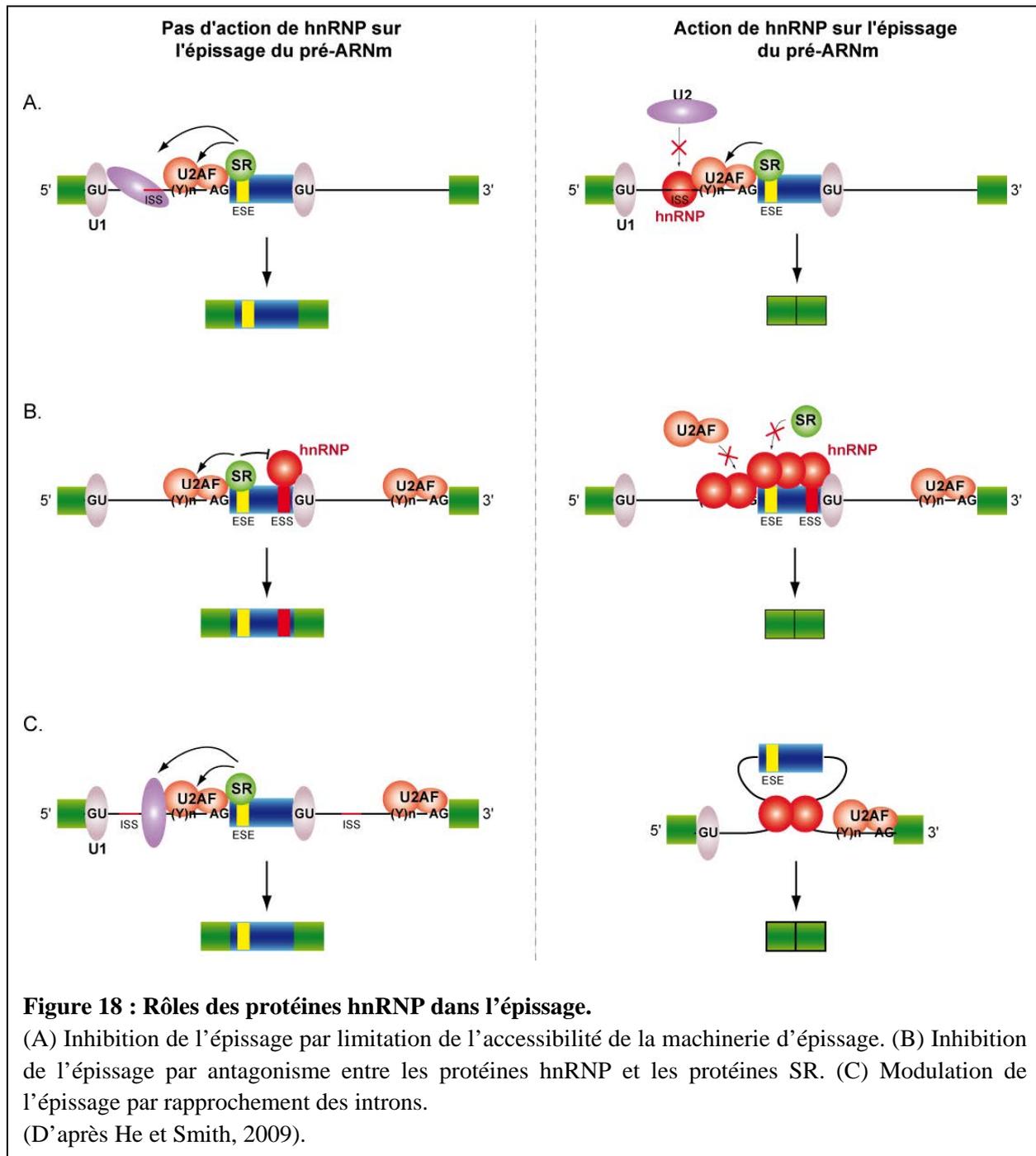
la transportine et diminue l'import nucléaire de hnRNP A1 (Allemand *et al.*, 2005). Autre exemple, les méthylations sur les résidus arginines du domaine RGG de la protéine hnRNP A1 diminuent sa capacité à interagir avec l'ARN et affectent les propriétés d'hybridation des acides nucléiques de hnRNP A1 (Kim *et al.*, 1998). Enfin, les protéines hnRNP C et M, qui sont exclusivement nucléaires et qui ont la capacité d'interagir avec des ARNm et de les retenir dans le noyau, peuvent être sumoylées. Ceci entraîne une diminution de l'affinité des protéines hnRNP pour les ARNm qu'elles retiennent dans le noyau, ce qui faciliterait l'export de ces derniers du noyau vers le cytoplasme (Vassileva et Matunis, 2004).

- Rôles des hnRNP dans l'épissage

Du fait de leur association aux pré-ARNm dès leur synthèse et jusque dans le cytoplasme, les protéines hnRNP jouent des rôles variés dans les différentes étapes de synthèse, maturation, transport, traduction et dégradation des ARNm (Dreyfuss *et al.*, 1993). Cette affinité pour l'ARN fait des protéines hnRNP des acteurs majeurs de l'épissage alternatif.

Inhibition de l'utilisation des sites d'épissage :

De part leur capacité à oligomériser le long de l'ARN, les protéines hnRNP jouent la plupart du temps le rôle d'inhibiteur de l'épissage en empêchant la fixation de la machinerie d'épissage. Cette inhibition prend effet sur les sites accepteurs et donneurs sans distinction. Les éléments sur lesquels se fixent les hnRNP peuvent être de nature exonique (ESS) ou intronique (ISS) (Fig.18). Un exemple de ce type de régulation au niveau du site 3' d'épissage est celui de la régulation de l'épissage du pré-ARNm *FGFR -2* codant le récepteur du facteur de croissance des fibroblastes. La protéine hnRNP A1 se fixe sur des séquences inhibitrices exoniques contenant des triplets UAG et réprime ainsi l'utilisation du site 3' en amont, en empêchant l'assemblage des complexes spliceosomaux autour de ce site d'épissage (Caputi *et al.*, 1999 ; Del Gatto-Konczak *et al.*, 1999). Ce mécanisme d'inhibition par encombrement stérique peut aussi intervenir au niveau du point de branchement et aux sites 5' d'épissage, notamment dans le cas de hnRNP A1 qui est capable d'oligomériser (Wagner et Garcia-Blanco, 2001 ; pour revue He et Smith, 2009).



Par antagonisme entre les protéines hnRNP et les protéines SR

L'inhibition de l'épissage par les protéines hnRNP peut aussi provenir de leur capacité à contrecarrer l'effet activateur des protéines SR liées à une séquence activatrice d'épissage. Cet antagonisme provient du fait que leurs sites de fixation sur l'ARN sont en général juxtaposés voire même chevauchants, ce qui empêche par encombrement stérique la fixation simultanée des 2 types de protéines (pour revue, Smith et Valcarcel, 2000). Ce phénomène a été maintes fois observé et il a conduit à l'hypothèse largement répandue que l'épissage

alternatif peut être régulé de manière tissu-spécifique grâce à des variations des taux relatifs de protéines SR et hnRNP (Fig.18B). L'exemple le plus utilisé est celui de l'antagonisme des protéines SRSF1 et hnRNP A1. Une étude de l'effet de ces facteurs sur la reconnaissance par la snRNP U1 de sites 5' en compétition, a permis de mettre en évidence deux paramètres essentiels de cette notion d'antagonisme qui sont la concentration des protéines et leur affinité pour l'ARN. Le modèle proposé indique qu'à forte concentration, hnRNP A1 multimérise sur l'ARN et empêche la particule snRNP U1 de s'associer au site 5' faible qui par ailleurs possède dans son voisinage un site de haute affinité pour la protéine SRSF1. Lorsque la concentration en protéine SRSF1 est plus élevée, la compétition est plus forte et le site 5' faible est alors utilisé (Eperon *et al.*, 2000). De même, les protéines SRSF7 d'une part et hnRNP H et F d'autre part, possèdent un seul et même site de fixation pour réguler positivement (SRSF7) et négativement (hnRNP H), l'épissage de l'exon 2 de l'ARN de l' α -tropomyosine (Crawford et Patton, 2006).

Modulation de l'épissage par rapprochement d'intron :

Les protéines hnRNP A/B ont également la capacité de rapprocher des séquences introniques très distantes l'une de l'autre (Fig.18C) (pour revue, He et Smith, 2009). Ainsi, si les sites se trouvent dans un même intron, une stimulation de l'épissage par rapprochement des sites 5' donneur et 3' accepteur sera observée. Ce phénomène a déjà été observé dans un contexte hétérologue avec un ARN contenant un large intron, faiblement épissé *in vitro*. L'insertion de séquences hnRNP A1 près des jonctions augmente significativement son épissage (Garneau *et al.*, 2005). De la même façon, un effet inhibiteur indirect peut être observé lorsque les sites de fixation de la protéine encadrent un exon, celui-ci peut se trouver alors exclu de l'épissage. C'est le cas du pré-ARNm de hnRNP A1 qui régule ainsi l'épissage de son propre pré-ARNm. La présence de sites de fixation de haute affinité localisés de part et d'autre de l'exon 7B, induit la formation d'une boucle dans l'ARN, qui conduit à l'exclusion de l'exon 7B sans même empêcher la fixation de la snRNP U1 (Chabot *et al.*, 2003).

Activation de l'épissage :

De la même façon que les protéines SR présentent un effet minoritaire d'inhibiteurs de l'épissage, certaines protéines hnRNP, dans certains contextes, peuvent jouer un rôle positif sur l'épissage. Cependant, les mécanismes sont encore assez mal connus.

En voici quelques exemples :

- ✓ La protéine hnRNP K fait partie d'un complexe activant l'épissage de l'exon 6A de la β -tropomyosine de poulet (Expert-Bezancon *et al.*, 2002).
- ✓ La protéine hnRNP M semble avoir un rôle positif puisque des anticorps dirigés contre elle inhibent l'épissage *in vitro* (Gattoni *et al.*, 1996).
- ✓ La protéine hnRNP G active l'inclusion de l'exon 7 du pré-ARNm *smn17*, en interagissant avec hTra2 β 1, une protéine à domaine RS (Hofmann et Wirth, 2002).
- ✓ Et enfin, la protéine hnRNP H active l'inclusion de l'exon 19 du transcrit du récepteur glutamate de type NMDA (GRIN1) (Han *et al.*, 2005).

III.2. Autres protéines et paramètres intervenant dans la régulation de l'épissage.

En plus des protéines hnRNP et SR, il existe de nombreux autres facteurs de régulation de l'épissage alternatif. Nous décrivons, ci-dessous, 2 exemples de régulation impliquant ces facteurs. De plus, les familles de facteurs de régulation de l'épissage MBNL et CELF seront décrites ultérieurement dans le Chapitre III.

III.2.a. La famille de protéines Fox

Cette famille est constituée des protéines Fox-1 et Fox-2 et de leurs isoformes respectifs. Ce sont des facteurs qui régulent l'épissage de manière tissu-spécifique et qui possèdent un domaine RRM. Les expériences SELEX réalisées avec la protéine Fox-1 humaine montrent qu'elle reconnaît spécifiquement la séquence UGCAUG (Ponthier *et al.*, 2006). Une étude récente montre que ces séquences UGCAUG sont conservées phylogéniquement et retrouvées le plus fréquemment en aval d'exons alternatifs qui sont surtout inclus dans des ARNm produits dans le cerveau ou dans les muscles (Minovitsky *et al.*, 2005). Ces séquences sont retrouvées dans plusieurs gènes comme celui de la fibronectine, c-src, calcitonine/CGRP, NMHC-B, 4.1R et FGFR-2 (Minovitsky *et al.*, 2005).

III.2.b. La famille de protéines Nova

La famille des protéines Nova, constituée de Nova-1 et Nova-2, est la première famille régulant l'épissage décrite avec une distribution tissu-spécifique. Ces protéines sont des facteurs de régulation de l'épissage exprimés exclusivement dans les neurones du système nerveux central (Buckanovich *et al.*, 1993 ; Jensen *et al.*, 2000a). Les protéines Nova contiennent 3 domaines de fixation à l'ARN de type KH. Des expériences de SELEX montrent qu'elles se fixent sur des séquences répétées YCAY où Y est une pyrimidine (Jensen *et al.*, 2000b ; Lewis *et al.*, 2000 ; Dredge *et al.*, 2005). Des études ont montré un lien

entre la localisation des sites de fixation des protéines Nova et leur rôle activateur ou inhibiteur de l'épissage (Ule *et al.*, 2006). Quand le motif Nova est à proximité des sites 5' ou 3' d'épissage de l'exon alternatif, que ça soit au niveau exonique ou intronique, la protéine Nova agit comme répresseur en inhibant la fixation des snRNP U1 ou U2 et des protéines SR. Au contraire, lorsqu'elle est fixée loin des sites donneurs ou accepteurs de l'exon alternatif, la protéine agit comme « activateur » en recrutant les protéines du spliceosome. Des données récentes montrent que les protéines Nova sont impliquées dans la régulation de l'épissage alternatif de pré-ARNm codant des protéines synaptiques, comme par exemple les pré-ARNm de la sous-unité 2 du récepteur GABAA et de la sous-unité $\alpha 2$ du récepteur de la glycine (Ule *et al.*, 2003).

IV. Lien entre transcription et épissage

IV.1. Lien entre l'ARN Pol II et le processus d'épissage

Le couplage entre la transcription et l'épissage a largement été illustré et semble aujourd'hui évident (pour revue, Bentley, 2005). Ce couplage serait dû à la capacité de l'ARN polymérase II et plus particulièrement à son domaine CTD (pour « Carboxy Terminal Domain ») à se lier à certains facteurs de la machinerie de maturation et à recruter les facteurs d'épissage aux abords des sites de transcription (pour revue, Bentley, 2002 ; Misteli et Spector, 1999) Bentley, 2002. En effet, le domaine CTD est nécessaire et probablement suffisant pour la localisation des complexes de transcription à la périphérie des speckles nucléaires qui sont le lieu de stockage de nombreux facteurs requis pour l'épissage (Du et Warren, 1997 ; Misteli et Spector, 1999). Par ailleurs, l'ARN Pol II phosphorylée est capable d'activer l'épissage *in vitro* d'un pré-ARNm contenant un exon interne (Hirose *et al.*, 1999) et cette activation est due au CTD de l'ARN Pol II (Zeng et Berget, 2000). D'où l'hypothèse que l'ARN Pol II serait impliquée à la fois dans la reconnaissance et dans le rapprochement des exons. Récemment, cette hypothèse a été renforcée par l'observation qu'un clivage au niveau de l'intron naissant n'empêche pas son épissage par la suite (Dye *et al.*, 2006). De plus, plusieurs facteurs d'épissage (U2AF₆₅ ou les facteurs PSF (pour « PTB associated Splicing Factor) sont associés à l'ARN Pol II et une analyse protéomique du spliceosome humain a permis de montrer qu'une dizaine de protéines spliceosomales sont communes aux processus d'épissage et de transcription (Tableau IV) (pour revue, Jurica et Moore, 2003 ; Roberts et Smith, 2002 ; Emili *et al.*, 2002). Toutes ces données suggèrent que le complexe transcriptionnel favorise la reconnaissance des exons, leur rapprochement et l'assemblage du spliceosome en regroupant différents facteurs d'épissage à proximité de l'ARN sortant du site

catalytique de l'ARN Pol II. Le CTD de l'ARN Pol II pourrait en particulier avoir un rôle très important en recrutant des facteurs d'épissage aux abords des sites de transcription (Misteli et Spector, 1999).

Protéines	Lien avec la transcription	Lien avec l'épissage
<u>Chez les mammifères</u>		
SCAFs WT-1 PGC-1 CA150	Lie le CTD Facteur de transcription Coactivateur Facteur d'élongation	Protéine SR-like Lie U2AF65 Régule l'épissage alternatif Lie SF1 Présent dans le spliceosome Présent dans le spliceosome Présent dans le spliceosome
XAB2 SKIP TAT-SF1 p54nrb PSF TLS CoAA	Facteur de transcription Facteur de transcription Facteur d'élongation Facteur de transcription Facteur de transcription Facteur de transcription Coactivateur	Lie les UsnRNPs En complexe avec les UsnRNPs En complexe avec les UsnRNPs En complexe avec les UsnRNPs hnRNP-like Régule l'épissage alternatif
<u>Chez la levure</u>		
PrP40 Spt5	Lie le CTD phosphorylé Facteur d'élongation	Facteur d'épissage Active l'épissage

Tableau IV : Protéines candidates au lien entre transcription et épissage.

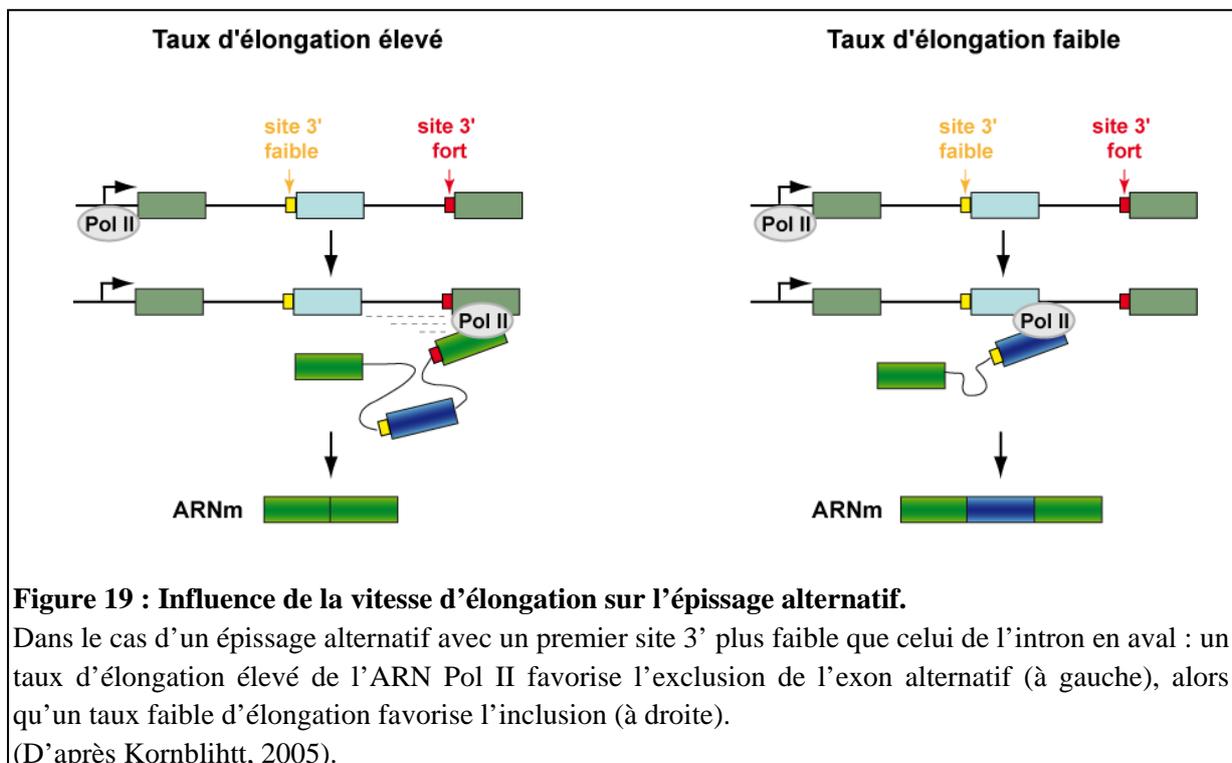
(Yuryev *et al.*, 1996 ; Davies *et al.*, 1998 ; Monsalve *et al.*, 2000 ; Goldstrohm *et al.*, 2001 ; Fong et Zhou, 2001 ; Kameoka *et al.*, 2004 ; Auboeuf *et al.*, 2004 ; Morris et Greenleaf, 2000 ; Lindstrom *et al.*, 2003).

IV.2. Rôle du promoteur et de la vitesse d'élongation de l'ARN Pol II sur le processus d'épissage

Plusieurs études faisant varier le type de promoteur utilisé pour la transcription ont permis de mettre en évidence le lien entre la nature du promoteur et l'efficacité d'épissage. Une telle idée est soutenue par le fait qu'un transcrit Pol II est rapidement associé au spliceosome, alors que l'utilisation du promoteur T7 conduit à la formation d'un complexe H beaucoup moins épissé (Das *et al.*, 2006). Ce phénomène n'est pas dû à des taux différents de transcription mais à des propriétés qualitatives du promoteur (Cramer *et al.*, 1999). De même, le simple changement de promoteur d'un gène peut changer son épissage alternatif (Cramer *et al.*, 1997 ; Pagani *et al.*, 2003). Le cas de l'exon EDI du gène *Fibronectine* illustre bien cet effet. En effet, le contrôle de ce gène par un autre promoteur change son épissage alternatif ainsi que sa réponse aux facteurs d'épissage (Cramer *et al.*, 1997). Le rôle du promoteur dans le contrôle de l'épissage alternatif serait dû à un phénomène d'occupation du promoteur par des facteurs transcriptionnels de nature différente en fonction de la structure du promoteur.

Ainsi, il a été démontré que bon nombre d'activateurs transcriptionnels influencent l'efficacité d'épissage (pour revue, Kornblihtt, 2005). Par conséquent, la régulation transcriptionnelle influence l'épissage alternatif, apportant un niveau de régulation supplémentaire.

Plus récemment, il a été proposé que la vitesse d'élongation de l'ARN polymérase II joue également un rôle (pour revues, Caceres et Kornblihtt, 2002; Caceres et Kornblihtt, 2002, Kornblihtt, 2005). Dans le cas d'une combinaison d'un site accepteur faible et d'un site accepteur fort en aval, une faible vitesse d'élongation ou des moments de pause interne favoriseraient l'inclusion d'exon alternatif, alors qu'une élongation efficace entraînerait plutôt leur exclusion (Fig.19). Cependant, la vitesse d'élongation de la transcription n'a pas d'influence sur des sites forts constitutifs (Nogues *et al.*, 2003).



IV.3. Influence de l'épissage sur la transcription

Comme nous venons de le voir, l'épissage peut être favorisé par la machinerie de transcription, mais différentes données suggèrent qu'en retour la machinerie d'épissage favorise elle aussi la transcription. En effet, la présence d'un intron ou même simplement d'un site 5' d'épissage, à proximité du promoteur augmente significativement le taux de transcription chez les mammifères comme chez la levure (Furger *et al.*, 2002). De même, la présence de la snRNP U1 au niveau du site 5' d'épissage peut activement recruter des facteurs généraux de transcription ainsi que l'ARN polymérase II sur le promoteur de façon épissage

dépendante (Damgaard *et al.*, 2008). De façon similaire, les UsnRNP spliceosomales interagissent avec le facteur humain d'élongation de la transcription p-TEFb et stimulent fortement l'élongation par l'ARN polymérase II probablement par la liaison de TAT-SF1 à p-TEFb. Ce complexe TAT-SF1-UsnRNP stimule aussi l'épissage *in vitro*, ce qui montre que des facteurs d'épissage peuvent stimuler directement l'élongation transcriptionnelle et que ces interactions réciproques existent dans le processus de couplage (Fong et Zhou, 2001 }Fong et Zhou, 2001. Il a aussi été proposé que le snRNA U1 s'associe au facteur TFIIH et que cette interaction stimule *in vitro* les étapes d'initiation et de ré-initiation de la transcription en plus de son rôle sur l'épissage (Kwek *et al.*, 2002). De plus, la présence du facteur U2AF65 permet d'éviter le phénomène de pause de l'ARN Pol II engendré par la synthèse de la séquence polypyrimidine (Ujvari et Luse, 2004). Enfin, les facteurs d'épissage SKIP (c-SKi Interacting Protein) et SRSF2, associés à P-TEFb sont capables de réguler l'élongation de la transcription *in vitro* et *in vivo* (Lin *et al.*, 2008; Bres *et al.*, 2005).

Bien que les mécanismes ne soient pas encore clairement établis, l'idée émerge donc que certains facteurs d'épissage associés au transcrit naissant sont capables de réguler les différentes étapes de la transcription comprenant l'initiation, le passage en phase d'élongation, puis l'élongation (pour revue, Kornbliht *et al.*, 2009).

IV.4. Régulation épigénétique de l'épissage alternatif

Les éléments régulateurs de l'épissage alternatif ainsi que la vitesse d'élongation de l'ARN polymérase II ne sont pas suffisants pour expliquer la complexité des régulations de l'épissage alternatif. Récemment, deux autres paramètres ont été mis en évidence : la structuration de la chromatine et les modifications épigénétiques des histones (Fig.20) (pour revues, Luco *et al.*, 2011 ; Alexander et Beggs, 2010).

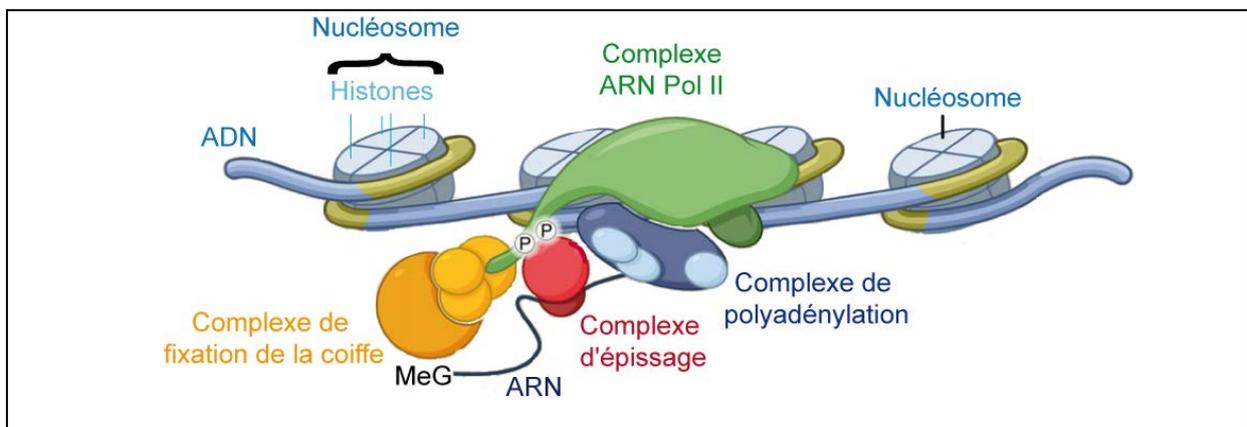


Figure 20 : Couplage entre les différentes machineries et l'épissage.

La régulation de l'épissage alternatif est dépendant de nombreux facteurs : le complexe ARN polymérase II, le complexe d'épissage, la modification des histones et la structure de la chromatine.

IV.4.a. Rôle de la structure de la chromatine

Un lien entre la densité des nucléosomes et l'architecture de la jonction exon/intron au niveau des gènes a été proposé pour la première fois en 1990 par Beckmann et Trifonov, qui ont observé que la distance moyenne entre des sites 5' ou 3' d'épissage présente une périodicité d'environ 205 paires de bases (Beckmann et Trifonov, 1991). Un nucléosome est l'unité structurelle de la chromatine qui détermine sa conformation et son niveau de compaction. Il est constitué d'un octamère de protéines histones enroulé autour de l'ADN avec une périodicité de 147 pb. De façon intéressante, la taille moyenne des exons de mammifère (140 nt) est semblable à la longueur d'ADN enveloppé autour d'un nucléosome, d'où l'idée d'un rôle du nucléosome dans la définition des exons (Schwartz *et al.*, 2009 ; Tilgner *et al.*, 2009).

La cartographie de la localisation des nucléosomes le long du génome par digestion à la nucléase micrococcale a montré que les nucléosomes sont enrichis à la jonction intron-exon, marquant ainsi les exons (Andersson *et al.*, 2009 ; Chodavarapu *et al.*, 2010; Dhami *et al.*, 2010 ; Kolasinska-Zwierz *et al.*, 2009 ; Nahkuri *et al.*, 2009 ; Schwartz *et al.*, 2009 ; Tilgner *et al.*, 2009). De plus, cet enrichissement a été conservé au cours de l'évolution des plantes jusqu'aux mammifères et est trouvé aussi bien dans les cellules somatiques que germinales (Nahkuri *et al.*, 2009). Les exons inclus par épissage alternatif sont plus fortement enrichis dans les nucléosomes que ceux exclus (Schwartz *et al.*, 2009) et la densité de nucléosomes varie selon la force des sites d'épissage avec un positionnement plus fort au niveau des exons définis par des éléments d'épissage plus faibles (Tilgner *et al.*, 2009). Ces données indiquent qu'en plus d'un rôle dans le positionnement des exons, les nucléosomes jouent également un rôle direct dans la régulation de l'épissage.

La position des nucléosomes peut également moduler la densité de l'ARN polymérase II au niveau des exons en induisant des pauses (Hodges *et al.*, 2009). Ainsi, il y a un enrichissement en ARN Pol II au niveau des exons alternatifs épissés par rapport aux exons constitutifs (Brodsky *et al.*, 2005). En outre, la surexpression du complexe SWI/SNF de remodelage de la chromatine dans des cellules humaines induit une accumulation de la forme phosphorylée de l'ARN Pol II dans le bloc central des exons alternatifs du gène CD44 et entraîne une augmentation de l'inclusion de ces exons dans l'ARNm mature (Batsche *et al.*, 2006).

Le positionnement du nucléosome, ainsi que la structure de la chromatine semblent donc intervenir dans la régulation de l'épissage alternatif, cependant, aucune expérience n'a pu à ce jour déterminer si ce rôle était direct ou indirect.

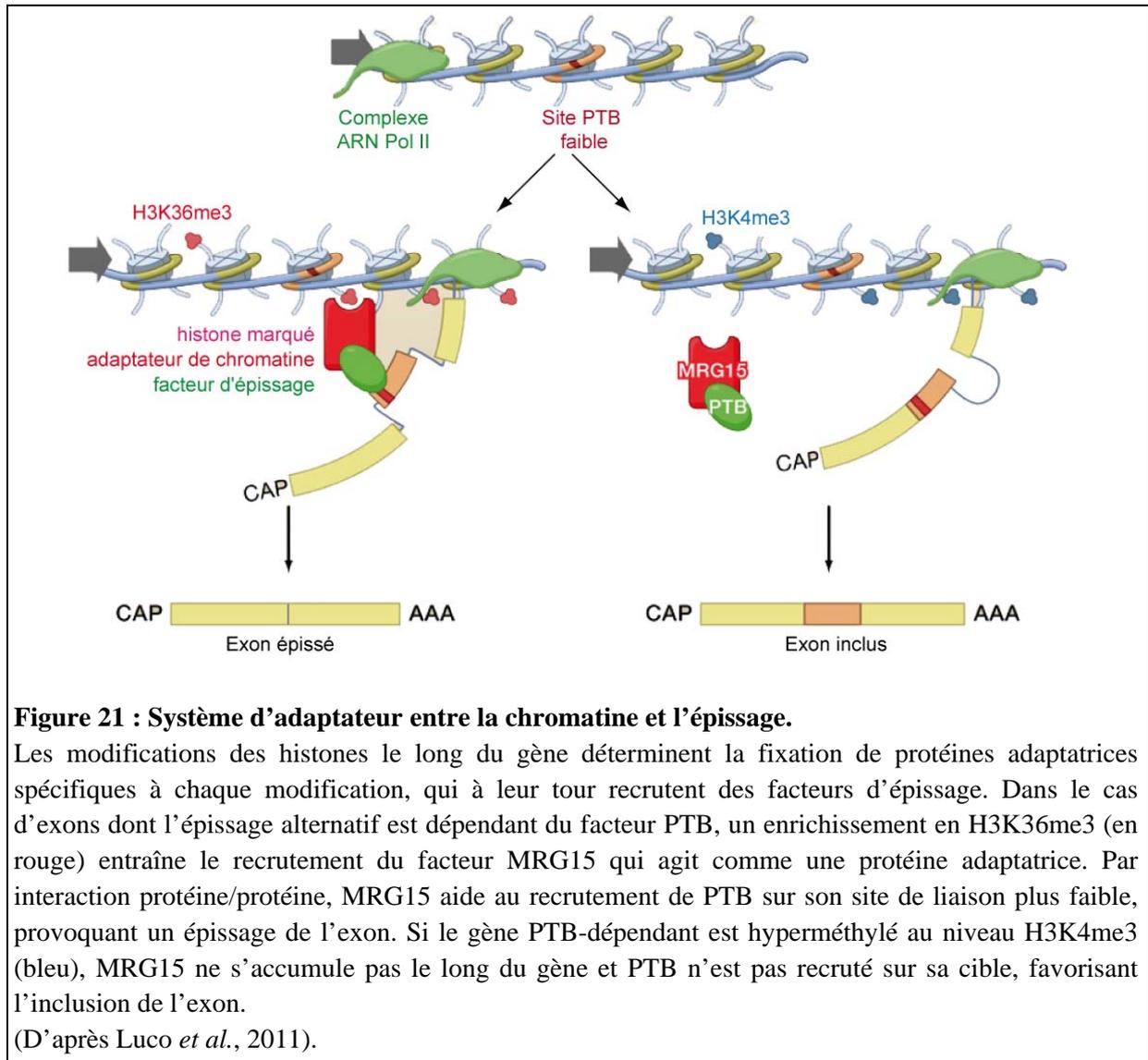
IV.4.b. Implication des modifications épigénétiques des histones

L'analyse de la distribution des 42 modifications des histones au sein du génome révèle que les histones marquées sont distribuées de façon non aléatoire et que plusieurs modifications sont enrichies spécifiquement (Kolasinska-Zwierz *et al.*, 2009 ; Spies *et al.*, 2009 ; Andersson *et al.*, 2009 ; Schwartz *et al.*, 2009). Ainsi, la tri-méthylation de l'histone H3 sur le résidu lysine en position 36 (H3K36me3) ainsi qu'en position 4 (H3K4me3) et la diméthylation de l'histone H3 sur le résidu lysine en position 27 (H3K27me2) sont enrichies au niveau des exons alors que la tri-méthylation de l'histone H3 sur le résidu lysine en position 9 (H3K9me3) y est réduite (Dhami *et al.*, 2010 ; Spies *et al.*, 2009).

Différentes expériences ont pu mettre en évidence un rôle des modifications épigénétiques des histones sur l'épissage alternatif. Ainsi, le traitement de cellules par un inhibiteur d'histone déacétylase induit l'épissage de l'exon alternatif 18 du pré-ARNm NCAM (pour « neural cell adhesion molecule ») (Nogues *et al.*, 2002 ; Allo *et al.*, 2009 ; Schor *et al.*, 2009). Un second exemple est l'étude des deux isoformes mutuellement exclusives et tissus-spécifiques de l'ARNm du récepteur 2 de facteur de croissance des fibroblastes, FGFR2-IIIb et IIIc. Ainsi, dans les cellules mésenchymateuses où l'exon IIIc est inclus, un enrichissement en modifications H3K36me3 et H3K4me1 des histones est observé le long de la région alternativement épissée. A l'inverse, la présence des modifications H3K27me3 et H3K4me3 entraînent l'inclusion de l'exon IIIb dans les cellules épithéliales (Luco *et al.*, 2010). La modulation du niveau de H3K36me3 ou H3K4me3 par surexpression ou répression des méthyltransférases responsables de leur formation provoque des changements des profils d'épissage alternatif tissus-spécifiques (Luco *et al.*, 2010).

Un mécanisme d'action des modifications épigénétiques des histones sur l'épissage a été récemment proposé. Ainsi, en plus d'agir sur la vitesse d'élongation de l'ARN Pol II, les histones influenceraient le recrutement de facteurs régulateurs de l'épissage *via* une protéine associée à la chromatine. Par exemple, pour un certain groupe de gènes dépendants de PTB, un enrichissement en H3K36me3 et un appauvrissement en H3K4me3 sont observés au niveau des régions épissées alternativement. Le mécanisme d'action de H3K36me3 se fait par recrutement du facteur de fixation à la chromatine MRG15, qui agit alors comme une protéine adaptatrice. Elle va à son tour par interaction protéine/protéine recruter le facteur d'épissage PTB sur l'ARN naissant au niveau de son site de liaison, entraînant l'épissage de l'exon. A l'inverse, dans les cellules présentant un faible niveau de H3K36me3, le facteur d'épissage PTB est faiblement recruté sur l'ARN naissant favorisant l'inclusion des exons dépendants de

PTB (Fig.21) (Luco *et al.*, 2010). En plus du système H3K36me3/MRG15/PTB, d'autres systèmes d'adaptateurs entre la chromatine et l'épissage ont été identifiés tels que les systèmes H3K4me3/CHD1 U2 snRNP, H3acétylé/Gcn5/U2 snRNP ou H3K9me3/HP1 α /hnRNP (Luco *et al.*, 2010 ; Sims *et al.*, 2007 ; Gunderson et Johnson, 2009 ; Piacentini *et al.*, 2009 ; Loomis *et al.*, 2009).



Chapitre III : Les dystrophies myotoniques

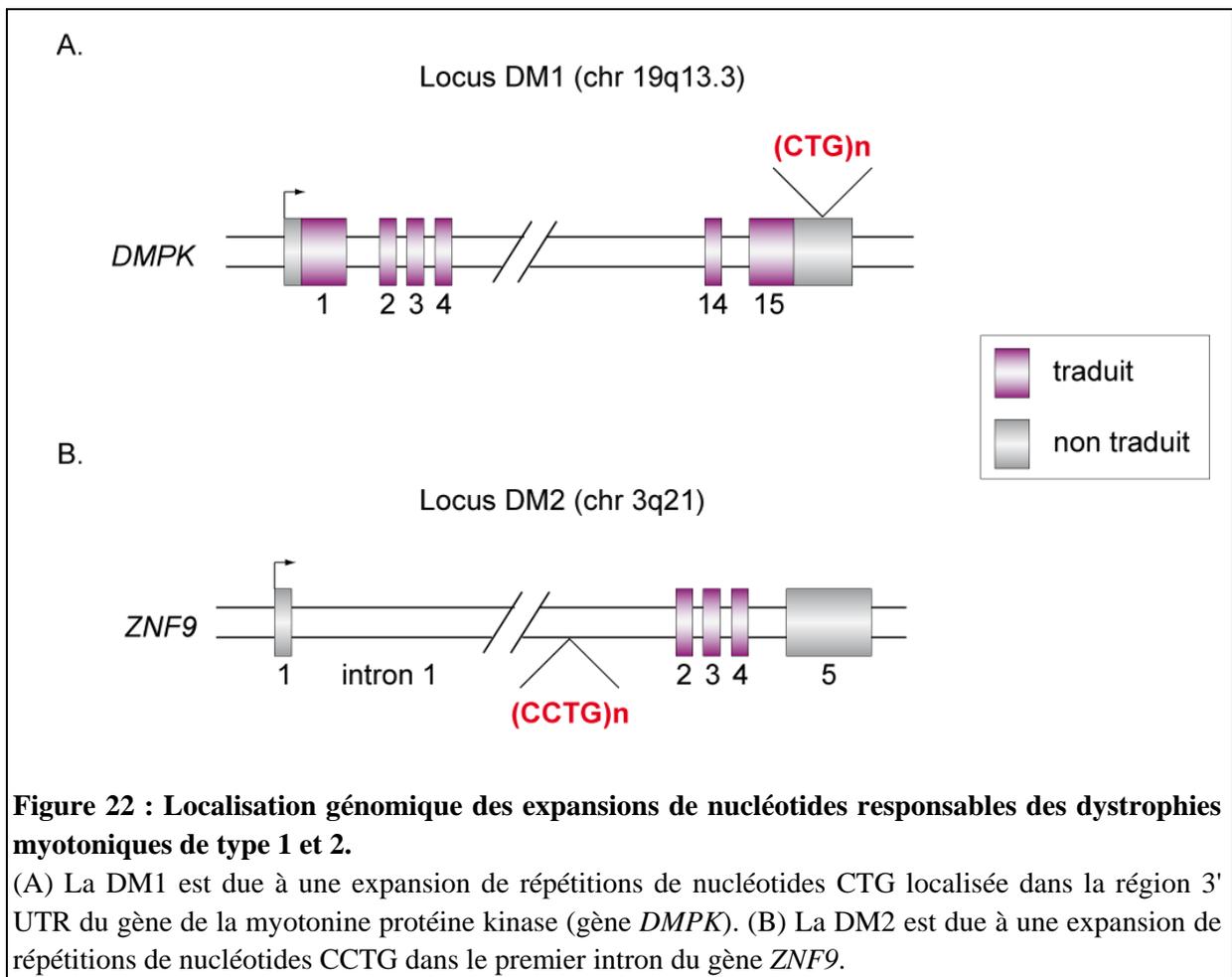
Les dystrophies myotoniques sont les plus fréquentes des dystrophies musculaires chez l'adulte avec une incidence mondiale d'une personne atteinte sur 8000 (Harper, 1989 ; Wieringa, 1994 ; Timchenko *et al.*, 1995). Les dystrophies musculaires sont liées à un groupe de plus de 30 myopathies, ou maladies neuro-musculaires, différentes. Une des caractéristiques communes à toutes ces dystrophies est une faiblesse musculaire qui conduit à une perte de mobilité des patients et une diminution de leur espérance de vie. Seules quelques dystrophies musculaires présentent des atteintes neurologiques. Ces troubles myotoniques correspondent à des difficultés pour détendre des groupes de muscles après leur contraction. Un trait caractéristique de ces maladies est la difficulté des personnes touchées à libérer des objets tels que des outils, des tasses et des poignées de porte par exemple.

I. Historique

En 1909, le docteur Hans Steinert a décrit pour la première fois la maladie de Steinert, appelée par la suite dystrophie myotonique de type 1 (DM1) ou dystrophie de Steinert. Il a noté des symptômes tels que faiblesse et atrophie musculaires, faiblesse faciale, ptosis (chute de la paupière supérieure) et atrophie testiculaire. Après autopsie, il a également observé une fibrose extensive et une dégénérescence des muscles squelettiques (Steinert, 1909). Par la suite, de nombreux travaux ont permis d'affiner les symptômes de la maladie et d'y inclure des troubles neurologiques (Rosman et Rebeiz, 1967). Ce n'est qu'en 1992 que la mutation génétique à l'origine de la pathologie a été identifiée. Il s'agit de la présence d'une expansion de triplets CTG dans la région 3' non traduite (3'-UTR) du gène *DMPK* (pour Dystrophy Myotonic Protein Kinase) situé sur le chromosome 19q13.3 et codant la myotonine, une Protéine Kinase AMPc-dépendante (Fig.22A) (Brook *et al.*, 1992 ; Buxton *et al.*, 1992 ; Fu *et al.*, 1992 ; Mahadevan *et al.*, 1992).

Depuis la découverte de la mutation CTG dans la région 3'-UTR du gène *DMPK*, des tests génétiques ont été utilisés afin de valider le diagnostic clinique de dystrophie myotonique. Ces tests ont révélé que 98% des patients présentant des symptômes cliniques de la dystrophie myotonique étaient porteurs de la mutation DM1 (Mahadevan *et al.*, 1992). Toutefois, entre 1 et 2 % des patients présentant ces symptômes ne portaient pas d'expansions CTG dans la région 3'-UTR du gène *DMPK* (Mahadevan *et al.*, 1992). La première preuve clinique qu'il pourrait exister une deuxième forme de dystrophie myotonique est venue

d'études sur des patients en 1994 (Ricker *et al.*, 1994). Les patients dans ces études présentaient des symptômes cliniques de la DM1 mais sans expansions CTG dans la 3'-UTR de *DMPK*, de plus ils affichaient une faiblesse musculaire proximale par opposition à la faiblesse distale observée classiquement dans la DM1. Les muscles proximaux sont les muscles situés aux racines des membres, c'est à dire les muscles des épaules et des bras pour les membres supérieurs et les muscles de la hanche et des cuisses pour les membres inférieurs. A l'inverse, les muscles distaux, situés aux extrémités des membres, sont les muscles de la main et des avant-bras pour les membres supérieurs et les muscles des pieds et des jambes pour les membres inférieurs. Cette forme particulière de dystrophie myotonique a dès lors été nommée PROMM pour « PROximal Myotonic Myopathy », aujourd'hui appelée dystrophie myotonique de type 2 (DM2) (Ricker *et al.*, 1994). C'est en 2001 que la mutation responsable de la DM2 a été identifiée, elle correspond à une expansion de répétitions CCTG dans l'intron 1 du gène *ZNF9*, également appelé *CNBP* (pour « Cellular Nucleic-acid-Binding Protein »), codant une protéine à doigts de zinc et localisé au niveau du chromosome 3q21 (Fig.22B) (Liquori *et al.*, 2001).



II. Les Dystrophies myotoniques

II.1. DM1

II.1.a. Aspects génétiques

Comme mentionné précédemment, la dystrophie myotonique de type 1 est causée par une expansion instable de répétitions CTG dans la région 3'-UTR du gène *DMPK*. Un individu sain possède naturellement entre 5 et 37 répétitions CTG dans cette région. Lorsque ce nombre de répétitions excède 37, ces répétitions deviennent pathogènes. Les patients présentant entre 38 et 49 répétitions sont à l'état de pré-mutation, ils sont asymptomatiques mais risquent de transmettre un nombre pathologique de répétitions à leur descendance. Entre 50 et 80 répétitions, la pathologie se développe tardivement et n'entraîne que peu de symptômes, c'est ce que l'on appelle la forme tardive non asymptomatique (Machuca-Tzili *et al.*, 2005). La forme adulte, entre 80 et 1000 répétitions, est la forme la plus répandue avec une apparition des symptômes entre 20 et 40 ans. La forme infantile débute avant l'âge de 10 ans et les patients atteints présentent jusqu'à un millier de répétitions. La forme la plus sévère, avec des symptômes présents dès la naissance, est la forme congénitale. Dans ce cas gravissime, l'enfant possède jusqu'à plusieurs milliers de répétitions (4000 CTGs), cette mutation provoque une mortalité infantile élevée (Brook *et al.*, 1992). Ainsi, la longueur des expansions de répétitions CTG est directement corrélée avec l'âge d'apparition de la pathologie et la sévérité des symptômes. Cette corrélation est d'autant plus importante lorsque le nombre de répétitions est inférieur à 400 (Gharehbaghi-Schnell *et al.*, 1998 ; Hamshere *et al.*, 1999).

La mutation DM1 est autosomique dominante et sa transmission dépend du sexe du parent transmetteur (Brunner *et al.*, 1993 ; Rakocevic-Stojanovic *et al.*, 2005). Les expansions de petites tailles très instables sont transmises par le père, tandis que les grandes expansions liées aux formes congénitales sont préférentiellement dues à une transmission maternelle (Zeesman *et al.*, 2002 ; Brunner *et al.*, 1993). Les larges répétitions sont probablement toxiques ou sélectionnées négativement lors de la spermatogenèse et non durant l'ovogenèse (de Die-Smulders *et al.*, 1997).

II.1.b. Les phénomènes de mosaïcisme et d'anticipation du nombre de répétitions

Le gène *DMPK* contenant plus de 37 répétitions CTG est instable et le nombre de répétitions augmente durant la méiose et la mitose au cours de la vie. Cette instabilité génétique des répétitions concerne les cellules somatiques et germinales (Savouret *et al.*, 2004 ; De Temmerman *et al.*, 2004). Le mécanisme majeur pouvant l'expliquer serait un

défaut au moment de la réplication et de la réparation de l'ADN (van den Broek *et al.*, 2007 ; Panigrahi *et al.*, 2005 ; Tome *et al.*, 2009). L'instabilité de l'expansion entraîne une variabilité dans le nombre de répétitions CTG entre les différents tissus d'un même individu ainsi qu'au sein d'un même tissu. Ce phénomène s'appelle le mosaïcisme.

De plus, la transmission de la maladie par des parents atteints de DM1 se manifeste dans les générations suivantes par une aggravation toujours plus prononcée des symptômes qui apparaissent à un âge toujours plus précoce dans la vie de l'individu jusqu'à générer la forme congénitale. Ce phénomène porte le nom d'anticipation (De Temmerman *et al.*, 2004).

II.1.c. Les aspects cliniques

Comme mentionné précédemment, les patients atteints de DM1 peuvent être divisés en quatre sous-types en fonction de l'âge d'apparition de la pathologie mais également en fonction de leur tableau clinique. La particularité de la DM est la présence, en plus des atteintes musculaires, de certaines caractéristiques électrophysiologiques, de symptômes non musculaires et de troubles neurologiques, ce qui renforce le caractère multi-systémique de la pathologie (Tableau V) (Machuca-Tzili *et al.*, 2005 ; Rosman et Rebeiz, 1967).

Système/ Organe	Atteinte
Endocrinien	<ul style="list-style-type: none"> • Résistance à l'insuline • Atrophie testiculaire • Taux élevés de FSH
Yeux	<ul style="list-style-type: none"> • Cataracte • Ptose • Dégénérescence de la rétine
Cœur	<ul style="list-style-type: none"> • Défaut de conduction • Arythmie cardiaque
Respiratoire	<ul style="list-style-type: none"> • Faiblesse et myotonie du diaphragme et des muscles respiratoires • Anomalies de contraction du pharynx et de l'œsophage • Hypoventilation alvéolaire • Aspiration branchiale
Cerveau	<ul style="list-style-type: none"> • Retard mental dans les formes congénitales • Hypersomnie • Déclin cognitif variable • Troubles de la personnalité : apathie élevée, évitement et comportement d'agressivité passive
Peau	<ul style="list-style-type: none"> • Calvitie prématurée
Muscle lisse	<ul style="list-style-type: none"> • Constipation et diarrhée • Douleurs abdominales de type coliques • Syndrome pseudo-occlusif • Contractions non-ordonnées de l'utérus durant l'accouchement

Muscle Squelettique	<ul style="list-style-type: none"> • Myotonie • Faiblesse et perte musculaires • Faiblesse des muscles distaux et tardivement des muscles proximaux • Faiblesse des muscles de la face, de la mâchoire, de la langue et des muscles fléchisseurs du cou • Atrophie du sternocléidomastoïdien
---------------------	---

Tableau V : Principaux symptômes des formes adultes de la dystrophie myotonique de type 1.

- La forme tardive non symptomatique

Une étude sur des patients atteints de DM1 ayant entre 50 et 200 répétitions CTG montre que la majorité des patients ayant entre 50 et 99 répétitions était asymptomatique en dehors de 38% qui présentaient une cataracte (Arsenault *et al.*, 2006). La myotonie et les faiblesses musculaires sont beaucoup plus fréquentes chez les patients ayant entre 100 et 200 répétitions CTG. Certains patients possédant entre 50 et 100 répétitions peuvent néanmoins développer des DM1 graves et à l'inverse des patients ayant entre 300 et 500 répétitions peuvent être asymptomatiques.

- La forme adulte « classique » de DM1

Ces patients DM1 présentent une atteinte des trois types de muscles du corps : squelettique, lisse et cardiaque. Ainsi, une dystrophie musculaire avec une atteinte particulière des muscles de la face, du cou et des muscles distaux des membres (avant-bras, jambe) a été observée, de même que des troubles respiratoires pouvant conduire à la mort de l'individu, une myotonie et des anomalies cardio-respiratoires, responsables de 20% des décès touchant des sujets DM1 (Bassez *et al.*, 2004 ; Harper, 1975 ; Logigian *et al.*, 2005 ; de Die-Smulders *et al.*, 1998 ; Mathieu *et al.*, 1999). Une atteinte de système nerveux central est également détectée chez les patients avec une tendance au retard mental, des troubles du sommeil, des troubles du comportement et une apathie (Woodward *et al.*, 1982 ; Phillips *et al.*, 1999 ; Delaporte, 1998). Les patients DM1 souffrent également d'une cataracte en « arbre de Noël », caractérisée par la présence de « petits points mélangés à des cristaux colorés ». Elle affecte quasiment 100 % des patients et apparaît de façon plus précoce qu'une cataracte ordinaire (Garrott *et al.*, 2004). De plus, des anomalies endocriniennes comprenant des troubles de la thyroïde, du pancréas, de l'hypothalamus, des gonades (hypogonadisme) et une hyperinsulinémie ont été détectées.

- La forme infantile

Cette forme est mal définie et souvent négligée en raison d'un historique familial ne contenant pas de cas répertorié de DM1. Les enfants atteints présentent des faiblesses des muscles faciaux, une myotonie, un faible niveau intellectuel et d'autres problèmes psychosociaux (Steyaert *et al.*, 1997).

- La forme congénitale

La forme congénitale DM1 n'est pas une forme précoce sévère « classique » de DM1 mais possède un phénotype distinct avec ces propres signes cliniques. Des signes sont détectables souvent avant la naissance tels qu'une réduction des mouvements fœtaux. Dès la naissance, les nourrissons sont très hypotoniques et présentent des troubles de succion et de déglutition avec une détresse respiratoire, qui sont à l'origine d'un taux de mortalité infantile élevé. Les enfants qui survivent développent leur capacité de déglutition et de ventilation mais présentent un retard dans le développement moteur et une atrophie cérébrale, et nécessitent un soutien scolaire (Ashizawa, 1998 ; Spranger *et al.*, 1997). Une myopathie progressive et d'autres symptômes vus dans la forme classique de DM1 peuvent se développer bien que normalement ces derniers se déclarent à l'âge adulte et progressent habituellement lentement (Joseph *et al.*, 1997). Les patients développent souvent de graves problèmes cardiorespiratoires vers 30-40 ans.

II.2. DM2

II.2.a. Aspects génétiques

La dystrophie myotonique de type 2 est une maladie génétique dominante présentant des caractéristiques cliniques multi-systémiques. Comme mentionné précédemment, cette pathologie est due à une expansion de répétitions CCTG dans l'intron 1 du gène *ZNF9* (Ranum *et al.*, 1998 ; Liquori *et al.*, 2001). L'intron 1 du gène *ZNF9* contient un motif complexe et polymorphique de répétitions (TG)_n(TCTG)_n(CCTG)_n (Fig.23) Le nombre de répétitions CCTG chez les patients sains est compris entre 11 et 26 et comprend jusque 2 interruptions simples GCTC ou TCTG localisées séparément dans le tract de répétitions CCTG (Liquori *et al.*, 2003). Chez les patients atteints de DM2, le nombre de répétitions CCTG peut aller de 75 à plus de 11000 répétitions CCTG sans interruption, avec en moyenne 5000 répétitions (Liquori *et al.*, 2001). En revanche, il n'a jamais été détecté de taille de répétitions compris entre 26 et 75.



Figure 23 : Complexité des répétitions au niveau de l'intron 1 du gène ZNF9.

Dans le cas des allèles normaux, la région CCTG est composée de 11 à 26 répétitions qui peuvent contenir entre 0 et 2 interruptions simples de type GCTC et TCTG (représentées par les deux lignes verticales). Dans le cas de la pathologie, le mécanisme d'expansion affecte uniquement les répétitions CCTG qui s'étendent alors de 75 à 11000 répétitions CCTG pures.

(D'après Dalton *et al.*, 2007).

Les répétitions CCTG présentent, comme les répétitions CTG, une instabilité somatique avec une augmentation de la taille des expansions avec l'âge (environ 2000bp/3 ans) mais aussi un mosaïcisme somatique (Liquori *et al.*, 2001 ; Day *et al.*, 2003).

II.2.b. Un phénomène d'anticipation ?

Certains symptômes de la DM2 empirent de génération en génération dans certaines familles, suggérant un phénomène d'anticipation similaire à celui observé chez les patients DM1 (Schneider *et al.*, 2000 ; Day *et al.*, 2003). Cependant, à la différence de la DM1, la taille des expansions de répétitions CCTG n'est pas corrélée à l'âge d'apparition de la pathologie ou à la sévérité des symptômes. De plus, aucune forme congénitale de DM2 comparable à la forme congénitale de DM1 n'a jamais été décrite. Enfin, une diminution drastique du nombre de répétitions CCTG a pu être observée dans certaines familles et cela quelque soit le sexe du parent porteur de la mutation. Cependant, le mosaïcisme et l'instabilité somatique compliquent l'interprétation de ces résultats (Day *et al.*, 2003). Ainsi le phénomène d'anticipation, observé chez les patients DM1, ne semble pas exister chez les patients DM2.

II.2.c. Les aspects cliniques

Les patients atteints de DM2 présentent en général des symptômes plus légers que les patients atteints de DM1, de plus la pathologie se déclare plus tardivement.

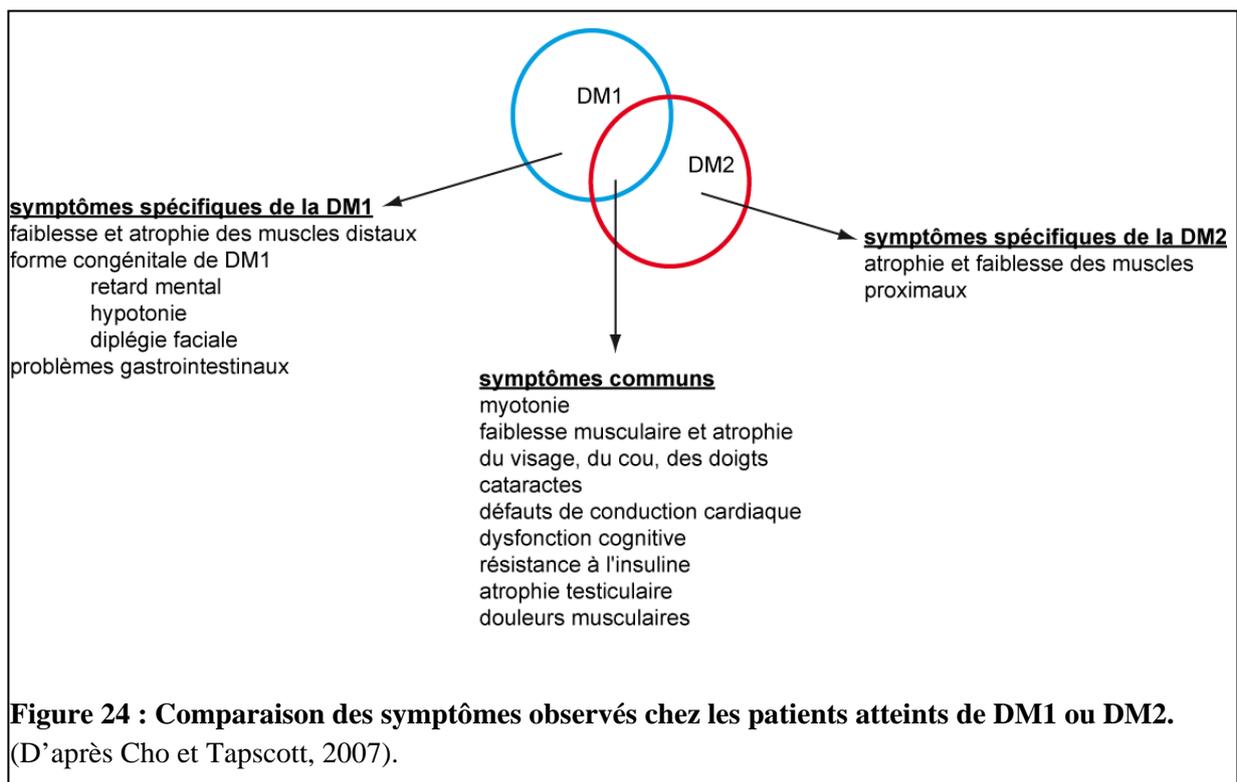
La DM2 est une pathologie multisystémique caractérisée par une myotonie (90 % des cas) et une faiblesse musculaire proximale (82%). Bien que des cas de myotonie aient été détectés chez des enfants atteints de DM2, les premiers symptômes apparaissent généralement vers 30 ans avec une faiblesse musculaire (Day *et al.*, 2003 ; Day *et al.*, 1999). Les patients DM2 présentent également des défauts de conduction cardiaque, des arythmies, des cardiomyopathies, des cataractes, une perte de sensibilité à l'insuline et des anomalies du système nerveux central (Colleran *et al.*, 1997 ; Merino *et al.*, 1998 ; Nguyen *et al.*, 1988 ;

Philips *et al.*, 1998 ; Day *et al.*, 2003 ;, Schoser *et al.*, 2004). L'atteinte du système nerveux central aurait un effet sur la connaissance, le comportement et la personnalité. A la différence de la DM1, la DM2 n'est pas associée à un retard mental (Meola *et al.*, 2002 ; Meola *et al.*, 2003).

III. Bases moléculaires des DM1 et DM2

III.1. Mécanismes de pathogénicité des répétitions CUG et CCUG

Malgré des expansions différentes (CTG/CCTG) localisées dans deux gènes apparemment sans rapport (*DMPK/ZNF9*), les deux maladies partagent beaucoup de manifestations cliniques communes (myotonie, faiblesse de muscle, développement de cataracte, résistance à l'insuline et défauts de conduction cardiaque) (Fig.24). De plus, les ARNm contenant des répétitions CUG ou CCUG sont séquestrés au sein d'amas ou *foci* nucléaires (Mankodi *et al.*, 2001 ; Fardaei *et al.*, 2002). Cependant, des différences cliniques existent entre ces deux pathologies, par conséquent une combinaison de plusieurs mécanismes, communs ou non à la DM1 et à la DM2, a été envisagée.



Comme indiqué précédemment, les expansions de répétitions liées à la DM1 et à la DM2 sont situées en dehors de la région codante. Un gain de fonction des protéines DMPK et ZNF9 est donc improbable. Un certain nombre d'autres mécanismes, que je décrirai plus en détails ci-dessous, peuvent contribuer mutuellement ou exclusivement à l'apparition des

différentes DM ; ceux-ci incluent (a) l'absence d'expression de la protéine DMPK ou ZNF9 due à la séquestration de leur ARNm respectif dans des *foci* nucléaires, (b) des effets locaux des amplifications sur la chromatine et sur l'expression des gènes voisins, et (c) le gain de fonction conféré à l'ARNm par la présence des répétitions (Fig.25) (pour revue, Cummings et Zoghbi, 2000).

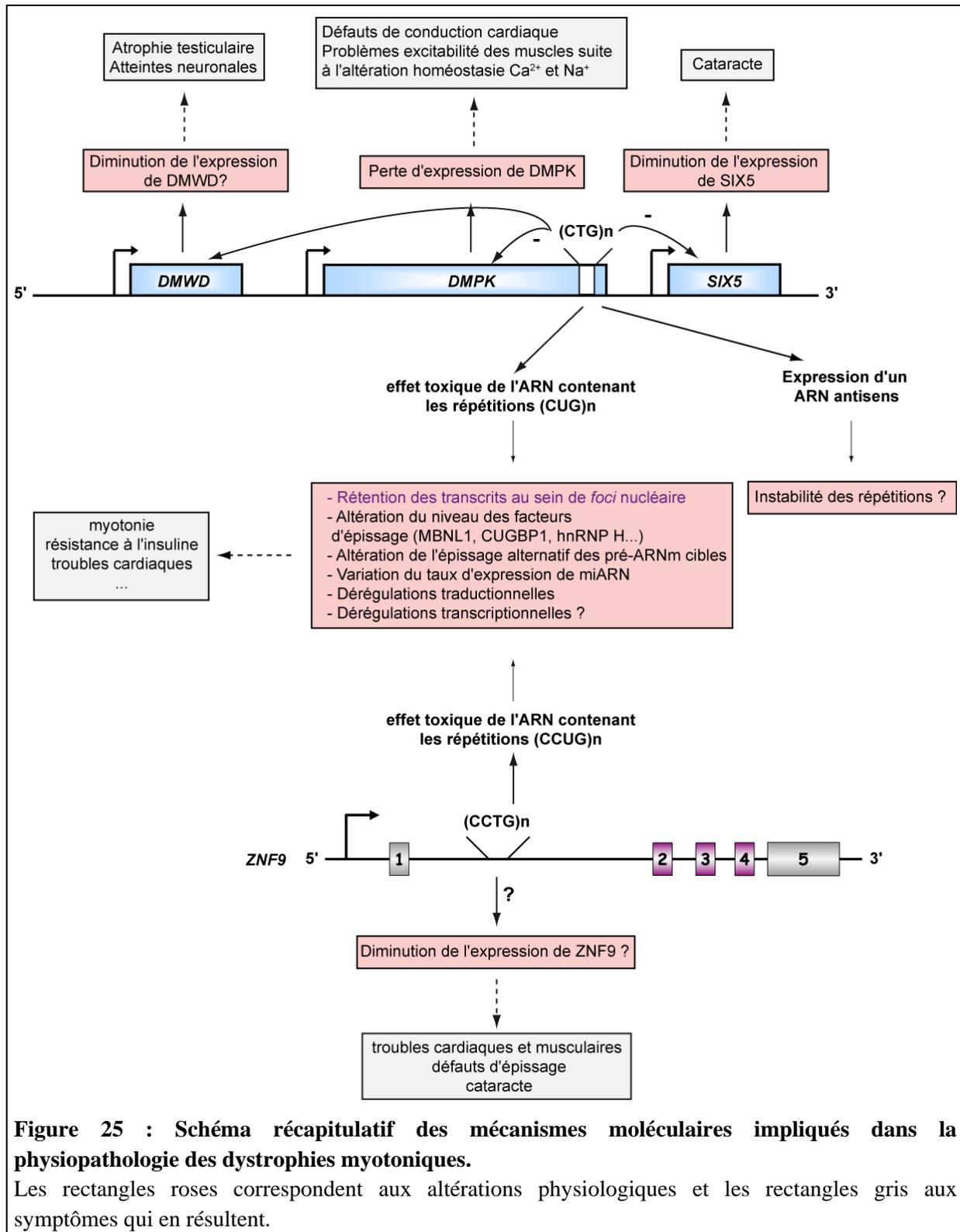


Figure 25 : Schéma récapitulatif des mécanismes moléculaires impliqués dans la physiopathologie des dystrophies myotoniques.

Les rectangles roses correspondent aux altérations physiologiques et les rectangles gris aux symptômes qui en résultent.

III.1.a. Conséquences de la perte d'expression des protéines

- Perte d'expression de la protéine DMPK

Les ARNm *DMPK* mutants contenant des amplifications CUG sont retenus dans le noyau des cellules DM1 et s'agrègent sous forme de *foci* nucléaires (Fig.26) (Taneja *et al.*, 1995).

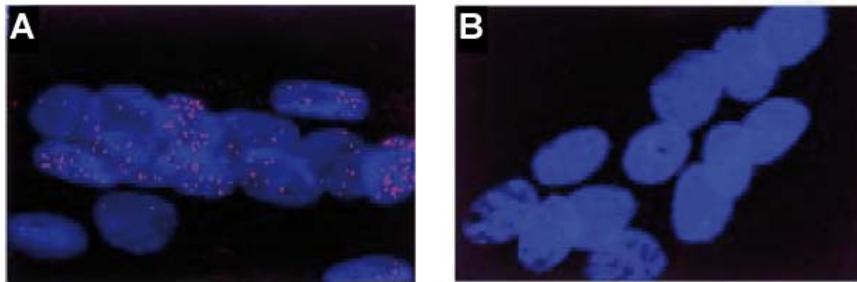


Figure 26 : Visualisation des *foci* d'ARN dans le noyau des cellules DM1 par hybridation *in situ*. Les *foci* composés d'ARN contenant des expansions CUG sont visualisés dans des myoblastes de patients DM1 (A) et de patients contrôles (B). Les *foci* ARN sont détectés par hybridation *in situ* à l'aide d'une sonde (CAG)₃₀ couplée à un fluorochrome rouge. Les noyaux sont identifiés par une coloration DAPI (en bleu).
(D'après Davis *et al.*, 1997).

Cette séquestration de l'ARNm entraîne une baisse d'expression de la protéine correspondante (haplo-insuffisance) qui pourrait être à l'origine de la DM1. En effet, le niveau d'expression de la protéine DMPK est réduit de 50 à 57% dans des myoblastes DM (Furling *et al.*, 2001a ; Furling *et al.*, 2001b). La protéine DMPK est une sérine/thréonine protéine kinase AMP cyclique-dépendante appartenant à la famille Rho. L'ARNm *DMPK* est constitué de 15 exons et 7 isoformes différentes sont produites par épissage alternatif (Groenen *et al.*, 2000 ; Tiscornia et Mahadevan, 2000 ; Wansink *et al.*, 2003). Ces isoformes diffèrent de par leur profil d'expression tissulaire, leur localisation et leur spécificité de substrat (Wansink *et al.*, 2003). Les fonctions de la protéine DMPK sont encore mal connues, mais on sait que cette protéine est activée par la voie des Rho-GTPase (Rac-1 et Raf-1) (Shimizu *et al.*, 2000). Par ailleurs, certains des substrats de DMPK, dont plusieurs sont communs avec la protéine kinase C (PKC) ou la calmoduline kinase II (CaMKII) ont été identifiés (Wansink *et al.*, 2003). Ces substrats sont : i) le phospholemman, une protéine dont la phosphorylation dépend des protéines kinases A et C et qui est impliquée dans la conductance électrique des canaux chlores (Mounsey *et al.*, 2000) ; ii) le phospholambdan, une protéine hyperphosphorylée dans le cœur qui serait impliquée dans des défauts de contractibilité et de conductivité cardiaques (Pall *et al.*, 2003) ; iii) la sous-unité myosine phosphatase (MYPT1) dont l'activité est inhibée par phosphorylation (Muranyi *et al.*, 2001).

En cas d'haploinsuffisance de la DMPK, la phosphorylation de MYPT1 diminue, conduisant à une augmentation de son activité et à une hyperphosphorylation de la chaîne légère de la myosine. Ce mécanisme entraîne une sensibilisation Ca^{2+} -dépendante du muscle lisse (Hartshorne, 1998) et une réorganisation du cytosquelette (Kaibuchi *et al.*, 1999).

Des lignées de souris transgéniques invalidées pour le gène *Dmpk* présentent des anomalies de conduction des canaux sodiques et calciques dans les muscles squelettiques et cardiaques et une myopathie mais celle-ci est d'intensité modérée. Elle apparaît tardivement et il n'y a pas de myotonie (Jansen *et al.*, 1996 ; Berul *et al.*, 1999 ; Matynia *et al.*, 2010). Ainsi, le changement du taux d'expression de la protéine DMPK pourrait contribuer à la maladie mais n'est pas suffisant pour expliquer l'apparition des différents symptômes des patients atteints de DM (Mankodi *et al.*, 2000). D'autres mécanismes sont donc obligatoirement impliqués.

- Perte d'expression de la protéine ZNF9

Dans le cas de la DM2, deux études tendent à prouver que ni l'épissage des pré-ARNm contenant des répétitions CCUG, ni la production de la protéine ZNF9 ne semblent altérés chez les patients atteints de DM2. En effet, les répétitions CCTG présentes au niveau de l'intron 1 du gène *ZNF9* seraient transcrites puis épissées, ainsi seules les répétitions CCUG seraient présentes au niveau des *foci* nucléaires (Fig.27), alors que l'ARNm *ZNF9* serait normalement épissé, exporté dans le cytoplasme puis traduit (Botta *et al.*, 2006 ; Margolis *et al.*, 2006).

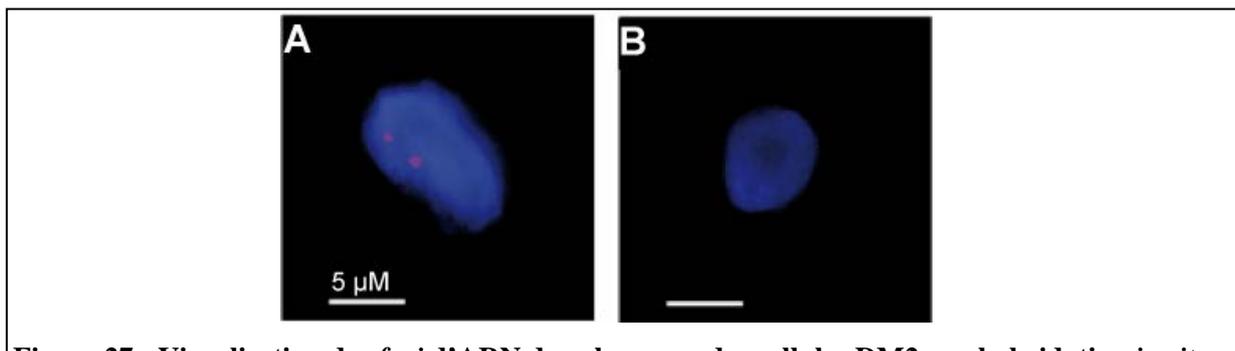


Figure 27 : Visualisation des *foci* d'ARN dans le noyau des cellules DM2 par hybridation *in situ*. Les foci composés d'ARN contenant des répétitions CCUG sont visualisés dans une coupe de muscle squelettique de patient DM2 (A). (B) représente une biopsie du muscle squelettique de patient contrôle. Les *foci* ARN sont détectés par hybridation *in situ* à l'aide d'une sonde $(\text{CAG})_{30}$ couplée à un fluorochrome rouge. Les noyaux sont identifiés par une coloration DAPI (en bleu). (Liquori *et al.*, 2001).

Cependant, des études ont montré qu'une déplétion des protéines orthologues à la protéine ZNF9, chez le zebrafish ou le poussin, entraîne des dommages cérébraux et musculaires (Abe *et al.*, 2006 ; Weiner *et al.*, 2007). De même, un modèle de souris, portant une délétion d'une seule copie du gène *Znf9*, présente des troubles cardiaques et musculaires, une cataracte et des défauts d'épissage des pré-ARNm similaires à ceux observés chez les patients DM2. Ces défauts peuvent être corrigés par la restauration d'un taux normal de protéine ZNF9 (Chen *et al.*, 2007). Ces résultats suggèrent qu'une perte de fonction de ZNF9 pourrait contribuer à la pathologie DM2. De plus, une étude récente démontre l'existence de défauts de traduction et de localisation de la protéine ZNF9 dans les cellules DM2 (Huichalaf *et al.*, 2009).

Cette protéine ZNF9 aurait de nombreuses fonctions, telles qu'une activité chaperonne vis-à-vis des acides nucléiques. Ainsi, elle intervient dans la formation de quartets de G au niveau de la région promotrice de certains gènes, tels que *c-Myc*, agissant ainsi comme un régulateur transcriptionnel (Borgognone *et al.*, 2010 ; Armas *et al.*, 2008 ; Shimizu *et al.*, 2003). La protéine ZNF9 joue également un rôle dans la régulation traductionnelle cap-dépendante et cap-indépendante de nombreuses protéines, en particulier de protéines appartenant à la machinerie traductionnelle telles que les protéines ribosomales, les facteurs d'élongation eEF1A et eEF2, et PABP1 (pour « poly-A binding protein ») (Pellizzoni *et al.*, 1997 ; Meyuhas, 2000 ; Cardinali *et al.*, 2003 ; Gerbasi et Link, 2007 ; Huichalaf *et al.*, 2009 ; Sammons *et al.*, 2010). Or, une diminution du taux global de synthèse des protéines, qui pourrait résulter d'une diminution de l'expression de ZNF9, a été observée dans des myoblastes DM2 (Huichalaf *et al.*, 2009).

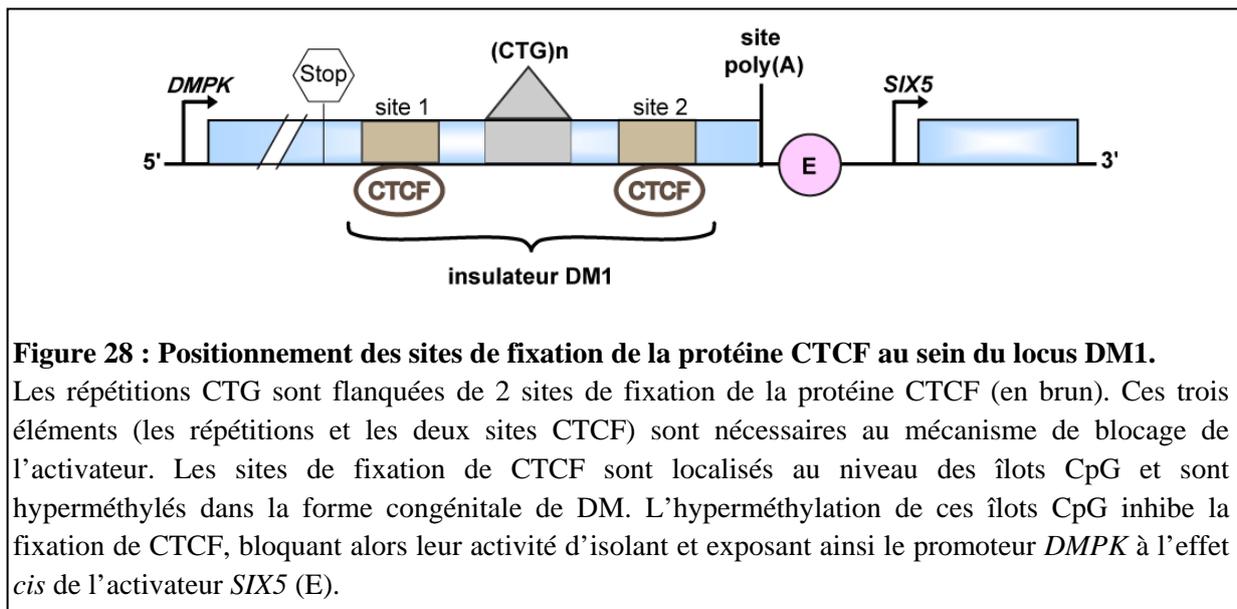
Ainsi, un rôle de ZNF9 dans la DM2 n'est pas à écarter. Cependant, il ne permet pas d'expliquer les similitudes entre DM1 et DM2 et par conséquent d'autres mécanismes doivent être impliqués.

III.1.b. Conséquences de la présence de répétitions sur l'expression des gènes voisins

L'expansion de CTG pourrait changer la structure de la chromatine et avoir des effets locaux sur l'expression des gènes (Day et Ranum, 2005). Comme mentionné dans le Chapitre I de l'Introduction, les expansions de triplets forment une longue structure tige-boucle stable et pourraient fermement enrouler les histones créant de l'hétérochromatine et exercer ainsi un contrôle épigénétique négatif sur la transcription des gènes voisins (Freudenreich *et al.*, 1997 ; Wang *et al.*, 1994).

Le gène *DMPK* est situé en aval du gène *DMWD* (pour « Dystrophia Myotonic Gene with WD repeats ») et en amont du gène *SIX5* (pour « SIne oculis homeoboX homolog 5 »), un facteur de transcription à homéodomaine. Ainsi, des études ont montré que la présence des répétitions CTG dans la région 3'UTR du gène *DMPK* génère une condensation de la chromatine et son hyperméthylation (Otten et Tapscott, 1995 ; Steinbach *et al.*, 1998). Par ailleurs, le niveau d'expression de la protéine *SIX5* est réduit chez les patients DM (Alwazzan *et al.*, 1999 ; Gennarelli *et al.*, 1999 ; Thornton *et al.*, 1997). Des modèles transgéniques murins dans lesquels le gène *Six5* a été délété présentent un symptôme de cataracte, mais pas de pathologies musculaires (Klesert *et al.*, 2000). Des données contradictoires existent en ce qui concerne le niveau d'expression de la protéine *DMWD* dans les cellules de patients atteints de DM. Néanmoins, le profil tissu-spécifique de production de *DMWD* en fait un autre candidat, dont la perte de fonction pourrait contribuer à certains des symptômes de la DM, en particulier, l'atrophie testiculaire (Alwazzan *et al.*, 1999) et une atteinte neuronale (Westerlaken *et al.*, 2003).

Alors que les répétitions de triplet peuvent constituer par elles-mêmes un élément régulateur de l'expression des gènes adjacents au gène *DMPK*, un îlot CpG, localisé à l'extrémité 3' du gène *DMPK*, pourrait également jouer un rôle dans la régulation de l'expression de ces gènes (Boucher et Johnson, 1995 ; Steinbach *et al.*, 1998). Une étude sur le statut de méthylation de l'îlot CpG révèle que bien qu'il n'y ait ni méthylation allèle-spécifique, ni empreinte parentale, une hyperméthylation est présente dans des cellules de patients DM congénitaux (Jansen *et al.*, 1993 ; Steinbach *et al.*, 1998). Ainsi, l'hyperméthylation de l'îlot CpG dans le gène *DMPK* constitue un des seuls changements moléculaires qui permet clairement de distinguer les formes adultes des formes congénitales de DM dans des cultures de cellules primaires (Filippova *et al.*, 2001 ; Steinbach *et al.*, 1998). Cette hyperméthylation perturberait le fonctionnement d'un « insulateur » entre les gènes *DMPK* et *SIX5*, qui protège normalement le promoteur *DMPK* à l'effet *cis* de l'activateur *SIX5* (E) localisé en amont du promoteur *SIX5* (Fig.28) (Filippova *et al.*, 2001). Un insulateur est un site permettant de borner les domaines fonctionnels des gènes, certains constituent une barrière empêchant la propagation de l'hétérochromatine, et d'autres protègent un gène de l'action d'un activateur (Zhan *et al.*, 2001). La protéine CTCF se fixe au niveau de l'insulateur et l'active. Deux sites de fixation de CTCF sont présents de part et d'autre des répétitions CTG (Filippova *et al.*, 2001). La protéine CTCF ne peut plus se fixer lorsque l'îlot CpG est hyperméthylé, ce qui inhibe l'insulateur.



Les patients atteints de DM1 et DM2 présentent un phénotype très similaire alors que les mutations à l'origine de ces 2 maladies ne sont pas situées sur le même chromosome (Liquori *et al.*, 2001). De plus, aucune altération de la quantité d'ARN et de protéine des gènes voisins n'a été observée en présence de répétitions CCTG dans l'intron 1 du gène *ZNF9* (Margolis *et al.*, 2006). L'ensemble de ces éléments invalide donc un mécanisme pathogénique principalement basé sur la perte (ou le gain) de fonction d'une protéine dont le gène est muté.

III.1.c. Gain de fonction *trans*-dominant du transcrit muté

Une autre hypothèse quant au mécanisme de pathogénicité est que les répétitions CTG du gène *DMPK* ou CCTG du gène *ZNF9* génèrent un gain de fonction du transcrit. Bien qu'elles ne soient pas traduites, les expansions de triplets ou quadruplets sont transcrites et recrutent le facteur d'épissage MBNL1, l'ensemble formant des *foci* nucléaires, qui sont localisés à la périphérie des « speckles » nucléaires (Fig.29) (Holt *et al.*, 2007 ; Smith *et al.*, 2007 ; Taneja *et al.*, 1995 ; Davis *et al.*, 1997 ; Jiang *et al.*, 2004 ; Liquori *et al.*, 2001). Ces *foci* nucléaires pourraient avoir un effet toxique *trans*-dominant.

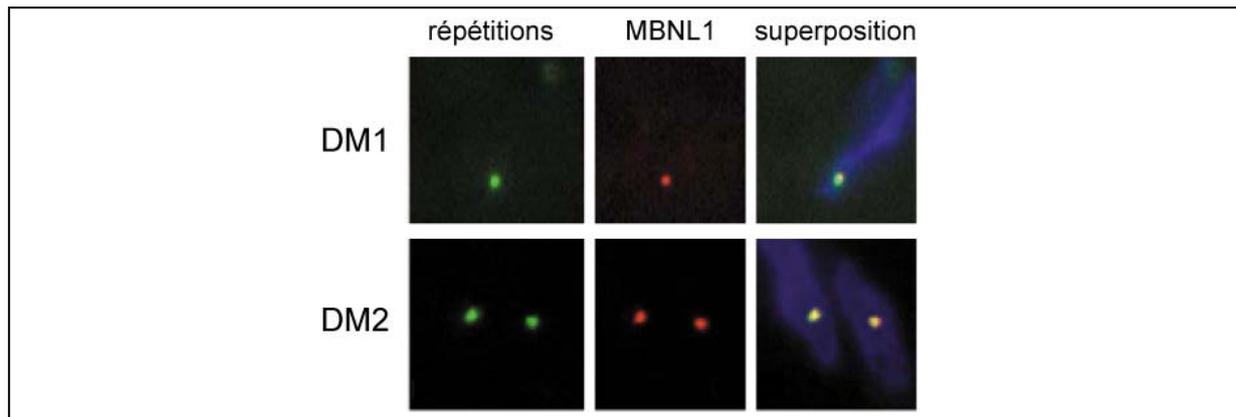


Figure 29 : Co-localisation des agrégats nucléaires formés par des expansions CUG et CCUG avec MBNL1.

Hybridation *in situ* des répétitions (vert) associée à un immunomarquage de la protéine MBNL1 (rouge) dans des fibres musculaires de patients DM1 (en haut) et DM2 (en bas). (D'après Mankodi *et al.*, 2001).

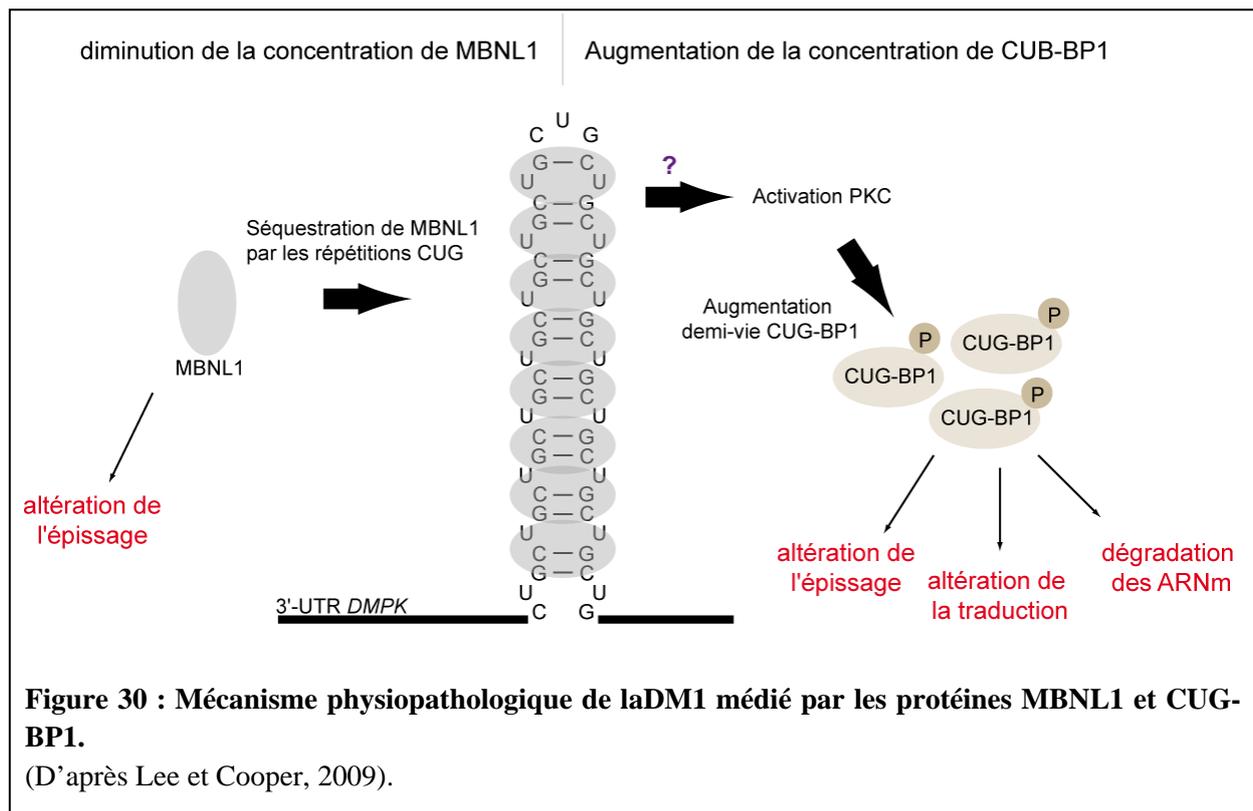
Des travaux sur des cellules myoblastes ont montré que la présence de répétitions CTG induisaient un ralentissement de la différenciation des myoblastes en myotubes (Amack *et al.*, 1999 ; Furling *et al.*, 2001b). De plus, des expériences réalisées chez la souris montrent que l'expression de répétitions CTG, dans la région 3'-UTR d'un gène non impliqué dans la DM1, permet de reproduire certains symptômes de la DM, en particulier la myopathie et la myotonie (Mankodi *et al.*, 2000). De même, l'équipe de Geneviève Gourdon a élaboré un modèle murin qui exprime 300 répétitions CTG dans leur contexte naturel (gène *Dmpk*) sous le contrôle de leur promoteur physiologique. Ces souris présentent un retard de croissance, des atteintes musculaires telles que la myotonie, ainsi que des atteintes cérébrales (Seznec *et al.*, 2001). Depuis, d'autres modèles murins ont permis d'illustrer la contribution des expansions aux atteintes cardiaques et respiratoires (Wang *et al.*, 2007 ; Panaite *et al.*, 2008). De plus, la surexpression de la région 3'UTR normale du gène *Dmpk* (ne contenant que 5 répétitions CTG) suffit à reproduire chez la souris les symptômes majeurs de la DM, et cela en absence d'accumulation des répétitions CUG dans des *foci* nucléaires (Mahadevan *et al.*, 2006). Ce résultat indique que la toxicité des répétitions dépend du nombre total de ces répétitions, qu'elles soient surexprimées en courts motifs ou sous la forme de très longues expansions. De plus, les *foci* ne sont pas indispensables à la toxicité des répétitions CUG.

Les patients atteints de DM présentent également un épissage alternatif aberrant de plusieurs transcrits. Les dérégulations de l'épissage alternatif des transcrits les plus connus ont été reproduites dans des modèles cellulaires et murins exprimant des expansions de

triplets (Charlet *et al.*, 2002b ; Philips *et al.*, 1998 ; Savkur *et al.*, 2001 ; Kanadia *et al.*, 2003a ; Gomes-Pereira *et al.*, 2007 ; Orengo *et al.*, 2008). Ceci a permis de démontrer le lien entre les dérégulations de l'épissage et la pathologie. Ainsi, les répétitions entraîneraient la séquestration de protéines régulatrices de l'épissage au sein des *foci*. Les différents symptômes de la DM ne seraient donc pas directement dus à l'expansion des répétitions, mais à la perte de fonction des protéines séquestrées qui ne pourraient plus jouer leurs rôles habituels (Michalowski *et al.*, 1999 ; Fardaei *et al.*, 2002).

Des expériences *in vitro* ont permis d'identifier la protéine CUG-BP1 et les protéines de la famille MBNL, qui sont toutes deux des facteurs d'épissage, ainsi que la protéine PKR activée, comme pouvant interagir avec les répétitions CUG (Timchenko *et al.*, 1996a; Miller *et al.*, 2000; Tian *et al.*, 2000). Cependant, contrairement à CUG-BP1 et PKR, seules les protéines MBNL sont séquestrées par les expansions de répétitions et sont présentes dans les *foci*. A l'inverse, une hyperphosphorylation de CUG-BP1 a été observée chez les patients DM1, provoquant une augmentation de la stabilité, une réduction de son export vers le cytoplasme et donc de la quantité de CUG-BP1 dans le noyau (Michalowski *et al.*, 1999 ; Mankodi *et al.*, 2001 ; Fardaei *et al.*, 2002 ; Jiang *et al.*, 2004 ; Savkur *et al.*, 2001). Un doublement de la quantité du facteur d'épissage hnRNP H a également été observé dans des myoblastes immortalisés de patients DM1 et une co-localisation entre la protéine hnRNP F et les *foci* a été décrite (Paul *et al.*, 2006 ; Jiang *et al.*, 2004).

Ainsi la grande majorité des symptômes observés chez les patients atteints de DM serait due à un gain de fonction *trans*-dominant du transcrit et plus particulièrement à son effet sur les facteurs régulateurs de l'épissage MBNL1 et CUG-BP1. L'augmentation de la concentration de CUG-BP1, associée à la diminution de protéine MBNL1 libre dans le noyau, provoque un déséquilibre de la balance entre ces deux protéines et induit une forte dérégulation de l'épissage de leurs pré-ARNm cibles. Par ailleurs, comme nous le décrirons plus tard, l'augmentation de protéine CUG-BP1 entraîne une instabilité des ARNm ainsi que des dérégulations de la traduction (Fig 30).



III.2. Lien entre dérégulation de l'épissage et symptômes

Aujourd'hui, l'effet *trans*-dominant des répétitions sur l'épissage alternatif est considéré comme le mécanisme majeur capable d'expliquer le phénotype multi-systémique de la maladie. En effet chaque symptôme a pu être relié à la dérégulation de l'épissage d'un ou de plusieurs pré-ARNm (Tableau VI) (Day et Ranum, 2005 ; Ranum et Cooper, 2006 ; Osborne et Thornton, 2006). La liste des événements d'épissage altérés chez les patients DM1 et DM2 est probablement toujours incomplète à ce jour, mais plusieurs ont été rapportés, parmi lesquels certains ont été caractérisés.

Organe	Pré- ARNm	Dérégulation exon/intron	Inclusion/exclusion	Référence
Coeur	Troponine T cardiaque (TNNT2/cTNT)	Exon 5	inclusion	Philips <i>et al.</i> , 1998
Muscle squelettique	Récepteur à l'insuline (INSR)	Exon 11	exclusion	Savkur <i>et al.</i> , 2001
	Canal chlore (CLCN1)	Intron 2 Exon 7A	inclusion inclusion	Charlet <i>et al.</i> , 2002b Mankodi <i>et al.</i> , 2002
	Troponine T squelettique (TNNT3)	Exon fœtal	inclusion	Kanadia <i>et al.</i> , 2003a
	Amphyphisine 2 (BIN1)	Exon 11	exclusion	Fugier <i>et al.</i> , 2011
	Récepteur à la ryanodine (RyR)	Exon 70	exclusion	Kimura <i>et al.</i> , 2005 Lin <i>et al.</i> , 2006
	Sarcoplasmic/endoplasmic réticulum Ca ²⁺ ATPase 1 (SERCA1)	Exon 22	exclusion	

	Sarcoplasmic/endoplasmic réticulum Ca ²⁺ ATPase 2 (SERCA2)	Intron 19	inclusion	Lin <i>et al.</i> , 2006	
	LIM domain binding 3 (LB3, ZASP)	Exon 11	inclusion		
	Titin (TTN)	Exon Zr4	inclusion		
		Exon Zr5	inclusion		
	Nebulin-related anchoring protein (NRAP)	Exon 12	inclusion		
	Calpain 3 (CAPN3)	Exon 16	exclusion		
	Attractin-like (ATRNL1, ALP)	Exon 5a et 5b	inclusion		
	Forming homology 2 domain containing 1 (FHOD1)	Exon 11a	exclusion		
	Glutamine-fructose_6-phosphate transaminase 1 (GFPT1)	Exon 10	exclusion		
	MBNL1	Exon 7	inclusion		
	MBNL2	Exon 7	inclusion		
	Facteur nucléaire 1	Exon 123	inclusion		Du <i>et al.</i> , 2010
	SET and MYND domain containing 1 (SMYD1)	Exon 39	inclusion		
	Sperm associated antigen 9	Exon 39	inclusion		
Myotubularin-related protein 1 (MTMR1)	Exon 2.1	exclusion	Buj-Bello <i>et al.</i> , 2002 Ho <i>et al.</i> , 2005a		
	Exon 2.3	exclusion			
Alpha-dystrobrevin (DTNA)	Exon 11a	inclusion	Nakamori <i>et al.</i> , 2008		
	Exon 12	inclusion			
Cerveau	Tau (MAPT)	Exon 2	exclusion	Sergeant <i>et al.</i> , 2001	
		Exon 3	exclusion		
		Exon 6	exclusion exon 6c inclusion exon 6d	Leroy <i>et al.</i> , 2006	
		Exon 10	exclusion	Sergeant <i>et al.</i> , 2001 Jiang <i>et al.</i> , 2004	
	Récepteur-N-méthyl-D-aspartate (NMDAR1)	Exon 5	inclusion	Jiang <i>et al.</i> , 2004	
	Amyloid precursor protein (APP)	Exon 7	exclusion		
	MBNL1	Exon 5	inclusion	Dhaenens <i>et al.</i> , 2010	
Exon 7		inclusion			

Tableau VI : Principales anomalies d'épissage dans différents tissus de patients DM1.

L'un des symptômes précoces et prédominants de la DM1 est la myotonie (hyperexcitabilité des fibres musculaires), causée par une perte de fonction des canaux chloriques spécifiques du muscle. Ceci s'explique par le fait que la majorité des ARNm *CLC-1* détectés dans le muscle

squelettique de cellules DM1 contiennent des codons de terminaison précoces, dus à la rétention de l'intron 2 et à l'inclusion de 2 nouveaux exons situés entre les exons 6 et 7, menant à la dégradation de l'ARNm par le mécanisme de NMD (pour « nonsense-mediated mRNA decay ») (Mankodi *et al.*, 2002 ; Charlet *et al.*, 2002b). Les ARNm du récepteur à insuline présentent une exclusion aberrante de l'exon 11, rendant majoritaire l'expression de l'isoforme ayant une moindre affinité pour l'insuline et induisant ainsi le symptôme de résistance à l'insuline (Savkur *et al.*, 2001 ; Savkur *et al.*, 2004). De même, une inclusion anormale de l'exon 5 du gène codant la troponine T cardiaque humaine (hcTNT), forme normalement exprimée lors du développement embryonnaire, entraîne des troubles cardiaques et musculaires (Cooper et Ordahl, 1984 ; Philips *et al.*, 1998). Dans le cerveau, l'épissage du pré-ARNm *MAPT* (pour « Microtubule Associated Protein Tau »), qui code la protéine neuronale Tau est également dérégulé, avec une exclusion majoritaire des exons 2, 3 et 10. Cette altération serait à l'origine des désordres neurodégénératifs présentés par les patients DM (Buee *et al.*, 2000). Des pré-ARNm musculaires voient également leur épissage altéré, tels que ceux codant les pompes à calcium SERCA1 et SERCA2 situées sur la membrane du réticulum endoplasmique, la troponine T squelettique rapide (TNNT3), le récepteur à la ryanodine (RYR1) et la protéine reliée à la myotubularine (MTMR1) (Kimura *et al.*, 2005 ; Dulhunty *et al.*, 2006 ; Kanadia *et al.*, 2003a). L'ensemble des défauts d'épissage génère chez les sujets DM1 un retour à l'expression des isoformes fœtales, ce phénomène pouvant être lié à un profil d'expression des protéines MBNL1 et CUG-BP1 de type fœtal (pour revue, Faustino et Cooper, 2003).

De manière intéressante, cet effet *trans*-dominant sur l'épissage alternatif est également observé chez les patients DM2, où l'on observe aussi une séquestration du facteur d'épissage MBNL1 dans des *foci* (Mankodi *et al.*, 2001 ; Cardani *et al.*, 2006 ; Cardani *et al.*, 2009 ; Lin *et al.*, 2006). Cet effet pourrait expliquer la similitude clinique entre les deux pathologies (Machuca-Tzili *et al.*, 2005). Les événements de dérégulation de l'épissage peuvent être modulés exclusivement par CUG-BP1 ou MBNL1, d'autres sont régulés par les deux protéines qui agissent alors toujours de manière antagoniste (Kalsotra *et al.*, 2008).

III.3. Autres mécanismes de pathogénicité observés

III.3.a. Dérégulation des micro-ARNs

Les micro-ARNs interviennent dans le développement et le fonctionnement du muscle, certains d'entre eux jouant un rôle crucial dans la prolifération et la différenciation cellulaire. La dérégulation de l'expression de micro-ARNs a été reliée fonctionnellement à

l'hypertrophie et aux arythmies cardiaques (Callis *et al.*, 2009), tandis qu'un changement de l'expression de certains micro-ARN a été montré dans des maladies neuromusculaires comme la myopathie de Duchenne (Greco *et al.*, 2009 ; Eisenberg *et al.*, 2009). Ce n'est que très récemment qu'une implication des microARN a été mise en évidence dans les dystrophies myotoniques (Gambardella *et al.*, 2010 ; Perbellini *et al.*, 2011 ; Rau *et al.*, 2011). Cependant, ces études sont contradictoires. Gambardella et ses collaborateurs montrent une surexpression et une localisation anormale de miR-206 dans des biopsies de muscles squelettiques de patients DM1, et ainsi qu'une légère augmentation de miR-1 et miR-181 (Gambardella *et al.*, 2010). Au contraire Perbellini et ses collaborateurs montrent une augmentation de l'expression de miR-335 et miR-1 et une diminution de miR-29b, -29c et miR-33 (Perbellini *et al.*, 2011), sans trouver de différence dans l'expression de miR-206, cela également dans les muscles squelettiques de patients DM1. Tandis que Rau et ses collaborateurs montrent une diminution de l'expression de miR-1 dans le cœur de patients DM1 (Rau *et al.*, 2011). L'augmentation de l'expression de cibles de miR-1 et miR-29 dans les tissus de patients DM1 tend à confirmer leur rôle dans la DM1 (Rau *et al.*, 2011 ; Perbellini *et al.*, 2011).

D'autres travaux préliminaires suggèrent que la protéine CUG-BP1 aurait la capacité de lier directement certains micro-ARNs pour les diriger vers leurs ARNm cibles (Timchenko, congrès IDMC-7, 2009). Une variation du taux d'expression des micro-ARNs (réduction des miR let7a et let7i, augmentation du miR-365) cumulée à l'augmentation de l'activité de la protéine CUG-BP1 pourraient ainsi déréguler l'expression de gènes appartenant à des voies aussi diverses que l'adhésion cellulaire, le cycle cellulaire et la réplication de l'ADN. Par ailleurs, la quantification des micro-ARNs par emploi de sondes TaqMan montre une augmentation du miR-34c dans le muscle de patients DM2 (Perbellini, congrès IDMC-7, 2009). L'expression différentielle de certains micro-ARNs chez les patients DM1 et DM2 représente donc une nouvelle source potentielle de mécanisme pathogénique.

III.3.b. Expression d'un ARN antisens

En 2005, l'existence d'un ARN antisens aux répétitions CUG (ARN CAG) a été montrée pour la première fois (Cho *et al.*, 2005). Cette transcription bidirectionnelle des répétitions aggraverait leur instabilité, que ce soit dans le sens d'une expansion ou d'une contraction, produisant des allèles instables (Nakamori *et al.*, 2011). Cet ARN antisens ne contient aucun cadre de lecture, cependant, il pourrait être traduit sans codon d'initiation et de rares agrégats de polyglutamines ont été observés chez les patients DM (Zu *et al.*, 2011).

Ainsi, le gain de fonction ARN des répétitions CUG pourrait être associé à un gain de fonction protéique comme observé dans les maladies à polyQ.

III.3.c. Dérégulation de la transcription

La comparaison des transcriptomes de muscles squelettiques de souris invalidées pour *MBNLI* et de souris exprimant de longues répétitions CUG (HSALR) a permis de mettre en évidence une dérégulation de l'expression d'un grand nombre de transcrits (Osborne *et al.*, 2009 ; Du *et al.*, 2010). De même, des altérations de l'expression de plusieurs gènes ont été observées dans des échantillons de muscle ou de cellules myoblastes de patients DM (Botta *et al.*, 2007 ; Dansithong *et al.*, 2011). Ces analyses suggèrent une dérégulation transcriptionnelle de nombreux gènes touchant l'homéostasie, la signalisation du calcium ou encore le développement et la fonction des muscles, ainsi que de gènes du réticulum sarcoplasmique et de la matrice extracellulaire.

A l'heure actuelle, des dérégulations de l'expression de deux facteurs de transcription NKX2-5 et SHARP, ainsi que de leurs cibles naturelles ont été détectées chez les patients atteints de DM (Yadava *et al.*, 2008 ; Dansithong *et al.*, 2011). Alors que l'expression du facteur NKX2-5 est induite dans les muscles de patients DM1 par un mécanisme encore inconnu, la concentration du facteur SHARP diminue dans le noyau suite à son export plus important vers le cytoplasme par le système CRM1.

III.3.d. Dérégulation de la traduction

Des études récentes proposent que les longues répétitions CUG inhibent la traduction dans des cellules musculaires de patients DM1 et DM2 (Huichalaf *et al.*, 2010 ; Huichalaf *et al.*, 2009 ; Salisbury *et al.*, 2009). En revanche, les données de la littérature ne sont pas encore claires sur les dérégulations de la traduction des protéines régulées par CUG-BP1. En effet, il a été montré une augmentation de l'expression de deux cibles de CUG-BP1, MEF2A (Myocyte Enhancer Facteur 2A) et de p21, deux facteurs importants pour la myogenèse et la régulation du cycle cellulaire, dans des cellules musculaires de patients DM1. Ces dérégulations contribueraient au défaut de différenciation des cellules musculaires observé chez les patients DM1 (Iakova *et al.*, 2004 ; Timchenko *et al.*, 2004). Cependant, il a également été démontré que les cultures de cellules musculaires DM exhibent une augmentation du taux de CUG-BP1 dans le noyau et à l'inverse une diminution dans le cytoplasme. De plus, CUG-BP1 se localise au niveau des granules de stress présents dans les cellules DM (Fujimura *et al.*, 2008).

III.3.e. Mécanisme d'ARN interférence

En 2005, il a été suggéré que les structures en tige-boucle des répétitions CUG, tout comme l'ensemble des répétitions de type CNG, pouvaient être la cible de la ribonucléase de type III DICER (Malinina, 2005). DICER est impliquée dans la biogenèse des ARN interférents (si- et mi-RNA). Elle clive dans le cytoplasme des ARN double-brin toutes les 21 à 25 paires de bases puis transfère les siRNA et miRNA double-brin au complexe multiprotéique RISC (pour « RNA-Induced Silencing Complex ») (pour revue, Deiters, 2010). Ainsi, les longues répétitions CUG pourraient être clivées par Dicer en de petits ARN CUG se comportant comme de petits ARN interférents et régulant de manière négative l'expression de différents transcrits (Krol *et al.*, 2007). Cependant, les répétitions CUG ne sont pas exportées hors du noyau et DICER est une protéine cytoplasmique. De plus, aucun petit ARN CUG n'a pu être mis en évidence chez les patients DM ou au sein de modèles murins de la DM, invalidant donc ce modèle.

IV. Les protéines de la famille Muscleblind

La famille des protéines Muscleblind (MBNL) a été identifiée comme l'un des principaux acteurs de la DM chez l'homme. En effet, la séquestration de ces protéines au sein de *foci* nucléaires suite à leur interaction avec les répétitions CUG et CCUG, entraîne une perte de leur fonction et ainsi des dérégulations de l'épissage alternatif de nombreux pré-ARNm.

IV.1. Diversité des protéines Muscleblind

Les protéines Muscleblind (Mbl) ont été initialement identifiées chez *Drosophila melanogaster* comme étant requises pour la différenciation terminale chez l'embryon du pharynx, des viscères, des muscles somatiques et des photorécepteurs des ommatidies (Begemann *et al.*, 1997 ; Artero *et al.*, 1998). Cette famille de protéines est fortement conservée entre espèces et n'est absente que chez les bactéries, les champignons et les plantes (pour revue, Pascual *et al.*, 2006). Le génome de toutes les espèces invertébrées analysées, des nématodes (*Caenorhabditis elegans*) aux Chordates primitif (*Ciona intestinalis*), contient un seul gène muscleblind. A l'inverse, la plupart des espèces vertébrées contiennent trois gènes, à l'exception du poisson épineux (*Takifugu rubripes*) qui contient jusqu'à cinq gènes (pour revue, Pascual *et al.*, 2006).

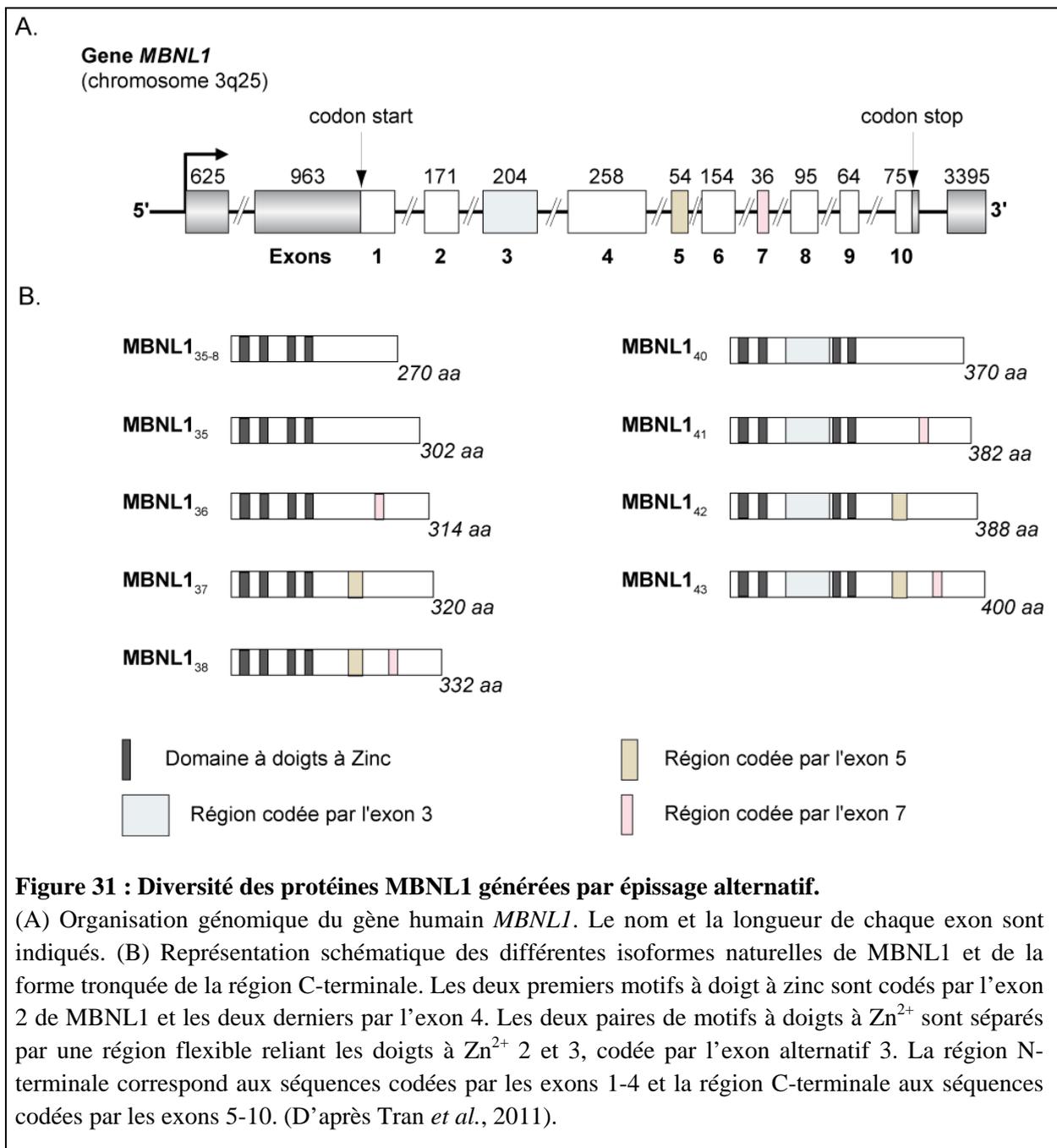
IV.1.a. Les protéines Mbl

Chez *Drosophila melanogaster*, un gène unique code les 7 protéines Mbl (Mbl A à G), qui sont obtenues par épissage alternatif des 11 exons du seul pré-ARNm *mbl* (Begemann *et al.*, 1997 ; Fernandez-Costa et Artero, 2010). Il est important de noter une différence entre la nomenclature des différentes bases de données et celle utilisée dans les différentes publications, cette dernière se basant sur la nomenclature de Begemann (Begemann *et al.*, 1997). Les différences se situent dans l'appellation des exons mais surtout dans l'identification des isoformes Mbl C et Mbl D, qui sont inversées entre les deux nomenclatures. Pour des raisons de facilité de compréhension, j'ai choisi de conserver la nomenclature de Begemann, utilisée dans l'ensemble des articles présentés. De plus, très peu d'informations étant disponibles sur les isoformes E, F et G, seules les isoformes Mbl A à D seront présentées par la suite.

Les protéines Mbl sont exprimées dans les yeux, le système nerveux central et la plupart des types de muscles. La perturbation de l'expression de Mbl entraîne une perturbation de la formation de rhabdomères (structure cellulaire captant la lumière chez les insectes) pendant la différenciation des photorécepteurs et affecte l'attachement de muscles à l'épiderme et la formation de Z-bande des muscles (Artero *et al.*, 1998 ; Begemann *et al.*, 1997). Les différentes isoformes de Mbl possèdent une région 5' commune mais varient au niveau de la région 3'. Après traduction, les protéines codées par les transcrits *mbl* A, B et C partagent les 179 premiers acides aminés, qui incluent deux doigts à Zn²⁺ CCCH, qui sont les domaines de fixation à l'ARN. A l'inverse, l'isoforme Mbl D partage uniquement les deux premiers exons et code pour une petite protéine de 84 aa contenant uniquement 1 seul doigt à Zn²⁺ complet. Les isoformes Mbl présentent des différences d'expression au cours du cycle de vie de la drosophile mais également au niveau des compartiments cellulaires (Vicente *et al.*, 2007). Alors que Mbl A est localisée au niveau de la membrane péri-plasmique, Mbl B est exprimée au niveau du cytoplasme et enfin Mbl C et D sont présentes dans le noyau et le cytoplasme (Fernandez-Costa et Artero, 2010). De plus, ces différentes isoformes ne présentent pas la même capacité à réverser le phénotype embryonique léthal des drosophiles mutées au niveau du gène *mbl*. L'épissage alternatif de l'*a*-actine, altéré chez les drosophiles mutantes, est uniquement régulé par Mbl C. De façon intéressante, seule Mbl C présente un homologue chez l'homme (Vicente *et al.*, 2007).

IV.1.b. Les protéines MBNL humaines

Chez l'homme, 3 gènes codant des protéines homologues à la protéine Mbl de drosophile ont été identifiés : *MBNL1* (ou *MBNL/EXP*), *MBNL2* (ou *MBLL/MPL1*) et *MBNL3* (ou *MBXL/CHCR*). Ils sont localisés respectivement sur les chromosomes 3, 13 et X. L'étude de l'organisation génomique des 3 gènes *MBNL* révèle une haute similarité de composition en exons aussi bien concernant le nombre de ces exons que leur taille (Fardaei *et al.*, 2002). Le gène *MBNL1* comprend 12 exons et la séquence codante est distribuée sur 10 exons numérotés de 1 à 10. Parmi ces 12 exons, 4 sont alternatifs (exons 3, 5, 7 et 9) et permettent la production des 9 isoformes de la protéine MBNL1 (Fig.31).



De même, trois et six isoformes sont respectivement produites pour MBNL2 et MBNL3 (Miller *et al.*, 2000 ; Fardaei *et al.*, 2002). La taille des différentes isoformes de MBNL1 est comprise entre 30 et 43 kDa et l'isoforme majoritaire est l'isoforme 40 kDa (Kino *et al.*, 2004). Il a été observé que l'expression tissulaire des différentes protéines MBNL varie. Alors que MBNL2 est exprimée de façon ubiquitaire dans l'ensemble des organes, MBNL1 est principalement exprimée dans le cœur, le cerveau et les muscles squelettiques, tandis que MBNL3 est surtout exprimée dans le placenta (Fardaei *et al.*, 2002 ; Squillace *et al.*, 2002). Par ailleurs, alors que l'expression de MBNL1 et MBNL2 augmente au cours du développement, celle de MBNL3 diminue (Kanadia *et al.*, 2003b). Il est à noter que l'expression des isoformes des protéines MBNL est également régulée au cours du développement et est altérée dans la DM1. Ainsi, les exons 5 et 7 sont inclus principalement dans les ARNm *MBNL1* présents dans les muscles et le cerveau fœtal plutôt que dans les tissus adultes et sont préférentiellement inclus chez les patients DM1. D'autre part, la localisation subcellulaire de la protéine MBNL change au cours du développement, ainsi MBNL1 est tout d'abord cytoplasmique puis devient majoritairement nucléaire au cours du développement du muscle squelettique (Lin *et al.*, 2006).

Ces changements d'expression des isoformes suggèrent une possible modulation de la fonction de MBNL1 par épissage alternatif de son pré-ARNm. Plusieurs études structure-fonction, dont je vais à présent vous présenter les conclusions, ont été menées afin de vérifier cette hypothèse.

IV.2. La protéine MBNL1 : relation structure-fonction.

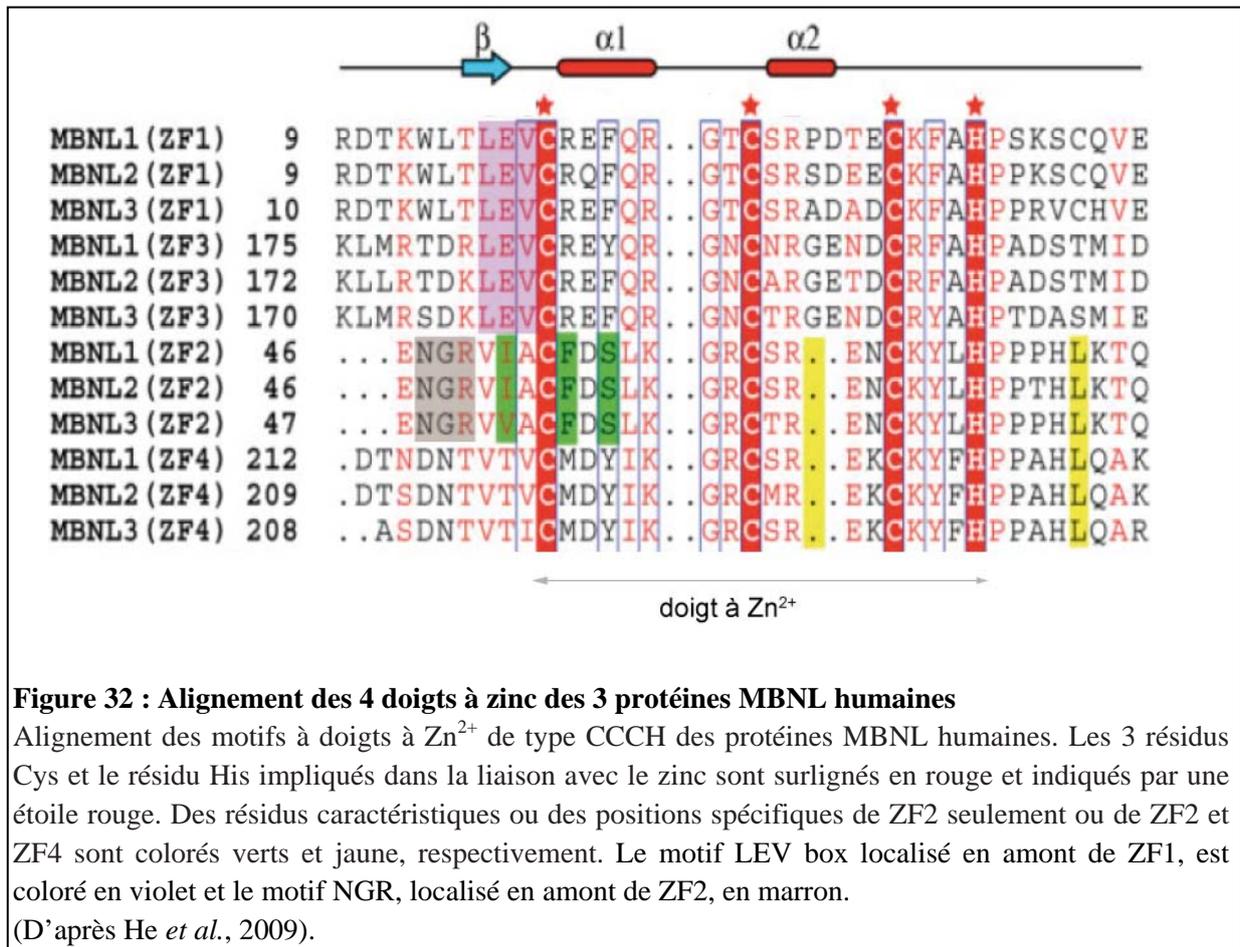
IV.2.a. Architecture générale de la protéine MBNL1

Tous les ARNm produits à partir du pré-ARNm *MBNL1* contiennent les exons 1, 2 et 4. Ces trois exons codent les 4 motifs à doigts à Zn²⁺ (ZF1-4) nécessaires pour la reconnaissance et la fixation de MBNL1 sur ses ARN cibles. La région flexible reliant les motifs ZF2 et ZF3 est codée par l'exon alternatif 3. Les séquences codées par les exons 1 à 4 constituent le domaine N-terminal de MBNL1. Les séquences codées par les exons 5 à 10 correspondent au domaine C-terminal de MBNL1.

IV.2.b. Le domaine N-terminal de MBNL1

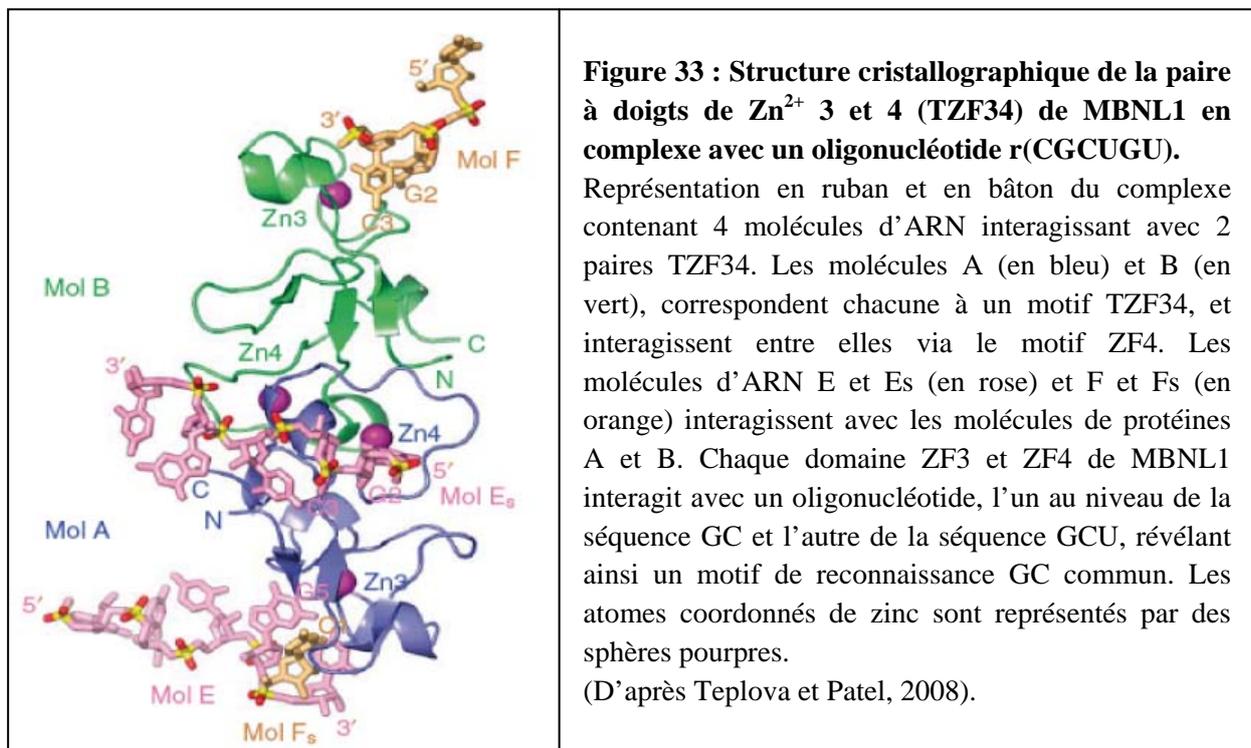
Comme mentionné précédemment, ce domaine correspond aux séquences codées par les exons 1-4 et contient entre autres les motifs ZF1-4. Ces motifs sont très conservés entre les espèces. Ils se caractérisent par la présence de trois résidus Cystéine et d'un résidu Histidine

(CCCH). Les domaines ZF1 et ZF3 sont de type (CX₇CX₆CX₃H), tandis que les motifs ZF2 et ZF4 sont de type (CX₇CX₄CX₃H). Les couples ZF1-ZF2 et ZF3-ZF4 constituent des motifs à doigts à Zn²⁺ en tandem (CX₇CX₆CX₃H-CX₇CX₄CX₃H), appelés TZF12 et TZF34, respectivement (Fardaei *et al.*, 2002). Les motifs TZF12 et TZF34 des trois protéines MBNL humaines présentent plus de 99 % d'identité (Fig.32).



La délétion ou la mutation de l'un des motifs à doigts à Zn²⁺ entraîne une altération de la fonction de la protéine, cependant la présence d'une seule paire à doigts à Zn²⁺ est suffisante pour permettre la fixation à l'ARN (Kino *et al.*, 2004 ; Grammatikakis *et al.*, 2010). La région flexible riche en alanine, située entre ZF2 et ZF3 et codée par l'exon 3, est également nécessaire à la liaison à l'ARN (Kino *et al.*, 2004). Comme mentionné précédemment, l'exon 3 est un exon alternatif, il existe donc naturellement chez l'homme des isoformes de MBNL1 dont la capacité à interagir avec des ARN est altérée (Kino *et al.*, 2004 ; Teplova et Patel, 2008). La seule présence du domaine N-terminal codé par les exons 1-4 (protéine MBNL1_{ACT3}) est suffisante pour reproduire l'affinité de MBNL1 pour ses ARN cibles (Kino *et al.*, 2004) (Fig.31).

La structure des motifs TZF12 et TZF34 des protéines MBNL1 et MBNL2 seuls ou complexés à un ARN a été résolue par cristallographie ou par RMN (Fig.33) (Teplova and Patel, 2008 ; He *et al.*, 2009). Ces structures montrent que les motifs TZF12 et TZF34 au sein des protéines se structurent de manière identique. De plus, la région reliant les doigts à Zn^{2+} au sein d'un motif TZF est fixe et est structurée de façon à ce que les surfaces de fixation à l'ARN de chaque doigt à Zn^{2+} soient éloignées et en orientation antiparallèle. Ces données structurales suggèrent que chaque doigt à Zn^{2+} peut fonctionner de façon indépendante et interagir avec des motifs ARN différents (Teplova et Patel, 2008 ; He *et al.*, 2009). La structure cristallographique d'un complexe TZF34 en interaction avec un oligonucléotide r(CGCU) révèle que les domaines ZF3 et ZF4 de MBNL1 interagissent chacun avec un oligonucléotide, l'un au niveau de la séquence GC et l'autre de la séquence GCU, exposant ainsi un motif de reconnaissance GC commun. Les motifs TZF12 et TZF34 se structurant de manière similaire, on peut supposer que chaque protéine MBNL a la capacité de se fixer sur 4 sites de fixation distincts. De plus, l'observation d'interactions protéine/protéine entre deux motifs ZF4 au sein des cristaux suggère que le motif ZF4 pourrait également intervenir dans des interactions protéine/protéine (Teplova et Patel, 2008).



Une comparaison de séquence des protéines MBNL a permis de mettre en évidence l'existence d'autres motifs conservés au sein du domaine N-terminal. Le premier motif, appelé LEV box, est localisé en amont de ZF1 (Fig.32) (Pascual *et al.*, 2006). Il a une

séquence consensus WLXLEV, où X correspond à n'importe quel acide aminé. Le second motif, appelé NGR box, consiste en une séquence cœur NGR, immédiatement suivie par un résidu arginine ou valine et de deux résidus apolaires localisés en amont et en aval du cœur NGR. Deux NGR box flanquent le second doigt de zinc. La fonction de ces différents motifs est inconnue (Pascual *et al.*, 2006).

IV.2.c. Le domaine C-terminal de MBNL1

Ce domaine correspond aux séquences codées par les exons 5-10. Il ne contient pas de motifs de fixation à l'ARN et sa présence n'est pas requise pour la reconnaissance des ARN puisqu'une protéine MBNL1 dépourvue de ce domaine (MLBNL1 Δ CT3) a la même affinité, voire une meilleure affinité pour ses ARN cibles que la protéine entière (Kino *et al.*, 2004 ; Terenzi et Ladd, 2010). Une étude récente de Terenzi et Ladd suggère même que la présence de l'exon 5 réduirait l'affinité de MBNL1 pour ses ARN cibles, puisqu'ils observent que l'affinité d'une protéine MBNL1 dépourvue de l'exon 5 est 2,5 fois meilleure que celle de la même protéine contenant l'exon 5 (Terenzi et Ladd, 2010). Par ailleurs, il a été observé qu'en l'absence de son domaine C-terminal, MBNL1 perdait sa capacité à réguler l'épissage (Kino *et al.*, 2004), suggérant un rôle essentiel de ce domaine dans la fonction de la protéine MBNL1. Cette importance fonctionnelle du domaine C-terminal et notamment de la séquence codée par l'exon 5 a été confirmée récemment. En effet, il a pu être montré que l'exon alternatif 5 est impliqué dans la localisation subcellulaire de MBNL1 et qu'en son absence MBNL1 perd sa localisation purement nucléaire au profit d'une localisation nucléocytoplasmique (Lin *et al.*, 2006 ; Terenzi et Ladd, 2010 ; Tran *et al.*, 2011). Plus récemment, un nouveau motif d'adressage nucléaire (motif KRAEK) a été mis en évidence chez la Drosophile, ce motif étant présent également au niveau de l'exon 5 de MBNL1. Lorsque ce motif est muté, la protéine n'est plus localisée exclusivement dans le noyau. De plus, lorsqu'il est fusionné à une protéine rapportrice (la β -galactosidase), celle-ci est relocalisée exclusivement dans le noyau (Fernandez-Costa et Artero, 2010). Cependant, le rôle de ce motif au sein de MBNL1 n'a pas été testé. L'épissage étant un processus nucléaire, on peut aisément imaginer les conséquences de ce changement de localisation sur la régulation de l'épissage. D'ailleurs un effet de l'absence de l'exon 5 de MBNL1 sur la régulation de l'épissage du pré-ARNm cTNT de poulet a pu être observé (Terenzi et Ladd, 2010). Par ailleurs, il a été observé que la protéine MBNL1 était capable d'interagir avec elle-même et de former des homodimères. Cette capacité d'oligomérisation de MBNL1 requiert le domaine C-

terminal, cependant les séquences impliquées au sein de ce domaine n'ont pas été identifiées (Yuan *et al.*, 2007).

Plus récemment un rôle des séquences codées par l'exon 6 constitutif et l'exon 7 alternatif, ainsi que la confirmation du rôle de la séquence codée par l'exon 3 dans la fixation des ARN et dans l'épissage a été décrit par Tran et ses collaborateurs. Du fait de ma contribution à cette étude, ces résultats seront présentés dans la partie Résultats de ce manuscrit.

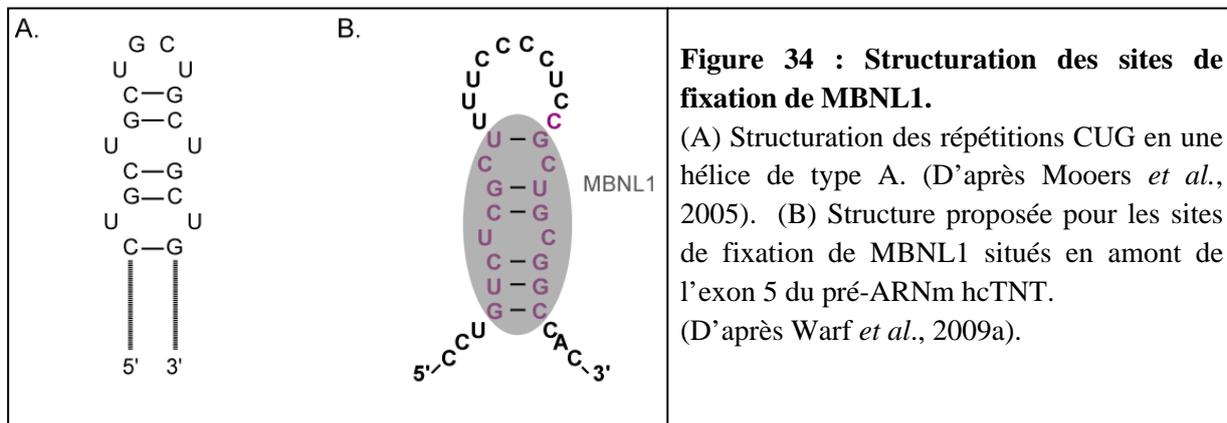
Des régions riches en proline, alanine ou phénylalanine sont également conservées au sein des domaines N- et C-terminaux des différents membres de la famille MBNL. La fonction de ces domaines n'a pas été déterminée, cependant les domaines riches en proline sont généralement impliqués dans les interactions protéine/protéine (Kay *et al.*, 2000). Les protéines MBNL contiennent également différents sites de phosphorylation conservés, qui pourraient moduler leur activité de régulateurs de l'épissage alternatif (Pascual *et al.*, 2006).

IV.3. Les motifs ARN reconnus par MBNL1

Initialement, des expériences de pontage aux U.V ont permis de démontrer la liaison de la protéine MBNL1 à des ARN contenant des répétitions CUG ou CAG (Miller *et al.*, 2000). Puis la fixation de MBNL1 aux répétitions CUG, et plus précisément aux structures ARN en double-brin formées par des répétitions de type CHHG/CHG, (où H=A, U, ou C) a été confirmée par triple-hybride (Kino *et al.*, 2004). Les protéines MBNL2 et MBNL3 sont également capables d'interagir avec ces répétitions et les trois protéines MBNL fusionnées à la GFP sont séquestrées dans les *foci* dans les cellules DM1 et DM2 (Fardaei *et al.*, 2002).

En plus de sa capacité à se fixer sur des répétitions de type CHHG/CHG, la protéine MBNL1 possède des sites de fixation sur ses pré-ARNm cibles. Ainsi, par alignement de séquences des sites de liaison de MBNL1 au sein des pré-ARNm de la troponine T cardiaque humaine et de poulet, Ho et collaborateurs ont mis en évidence la capacité de MBNL1 à se lier à des ARN ayant une séquence répondant au consensus YGCU(U/G)Y (où Y = U ou C) (Ho *et al.*, 2004). Notons que les répétitions CUG s'inscrivent dans cette séquence. Depuis, une étude structurale de l'équipe de Patel a montré que le motif TZF34 se lie à une séquence CGCUGU simple brin (Teplova et Patel, 2008). En 2010, l'équipe de Berglund a affiné le motif minimum de reconnaissance de MBNL1 comme étant YGCY lors d'une expérience de Doped SELEX réalisée à partir d'une région de 50 nucléotides contenant les sites de fixation de MBNL1 de la troponine T cardiaque humaine. Cette même équipe a également proposé

que la fixation de MBNL1 nécessite la présence d'au moins deux sites de fixation, ces sites interagissant entre eux afin de former une structure en tige-boucle reconnue par MBNL1. Cette structure serait assez similaire à celle formée par les répétitions CUG du fait de la présence de mésappariements au niveau de la tige (Fig.34) (Goers *et al.*, 2010 ; Warf *et al.*, 2009a).



Les ARN ne contenant qu'un seul motif YGCY lient MBNL1 avec une très mauvaise affinité comparée aux ARN contenant plusieurs motifs MBNL1. Cette observation a été récemment confirmée lors de la caractérisation du site de fixation de MBNL1 au sein du pré-miR1. En effet, dans ce cas un seul motif YGCU, localisé dans la boucle terminale du pré-miR1, est présent et l'affinité de MBNL1 pour cet ARN est également très mauvaise (Rau *et al.*, 2011).

Une expérience SELEX a également été réalisée sur la protéine Mbl de drosophile. Cette protéine peut être un bon modèle afin d'étudier la protéine MBNL humaine, bien qu'elle ne contienne que 2 doigts à Zn²⁺. En effet, la protéine Mbl a la capacité de se lier sur les cibles naturelles de MBNL1, mais aussi sur des répétitions CUG entraînant la formation de *foci* nucléaires. De plus, Mbl peut réguler l'épissage alternatif de différentes cibles de MBNL1 (Goers *et al.*, 2008 ; Kanadia *et al.*, 2003a ; Vicente-Crespo *et al.*, 2008). De façon intéressante, la protéine MBNL1 humaine a la capacité de réverser la létalité des mutants de la protéine Mbl drosophile (Monferrer et Artero, 2006). Ceci indique que le mécanisme de reconnaissance des cibles semble conservé entre les espèces. Le site consensus de fixation de Mbl identifié par SELEX est : AGUCU. Cette séquence entre dans la formation d'une structure tige-boucle contenant un mésappariement U-C. Cette structuration ressemble à celle proposée pour les motifs MBNL1 de hcTNT (Goers *et al.*, 2008). Cependant, cette séquence n'est pas retrouvée dans les cibles naturelles de MBNL1.

Ainsi, MBNL1 semble se lier sur des ARN double brin formant une structure en tige-boucle, avec la nécessité d'un mésappariement au niveau de la tige, mais également sur des ARN simple brin.

IV.4. Fonctions des protéines MBNL

IV.4.a. Régulateur de l'épissage alternatif

Les protéines MBNL ont été caractérisées comme régulateurs de l'épissage alternatif et régulent l'épissage de plus de 200 pré-ARNm, dont son propre pré-ARNm (Ho *et al.*, 2004 ; Du *et al.*, 2010 ; Gates *et al.*, 2011). En effet, l'inactivation du gène *MBNL1* chez la souris entraîne la dérégulation de l'épissage de plus de 200 pré-ARNm (Du *et al.*, 2010). MBNL peut agir comme activateur ou comme inhibiteur de l'épissage alternatif de différents transcrits. La surexpression de MBNL1, MBNL2 ou MBNL3 entraîne l'exclusion de l'exon 5 de cTNT mais provoque l'inclusion de l'exon 11 de l'IR (Ho *et al.*, 2004). Le mécanisme d'action de MBNL n'est pas encore clairement défini. Cependant, la localisation des sites de fixation par rapport à l'exon régulé pourrait être corrélée au rôle activateur ou inhibiteur de la protéine. Ainsi, la fixation de MBNL en amont d'un exon activerait son inclusion, alors que sa fixation en aval de l'exon entraînerait l'exclusion de l'exon (Goers *et al.*, 2010). Récemment, Warf et Berglund ont montré que MBNL1 régule l'épissage de hcTNT en stabilisant une structure tige-boucle en amont de l'exon 5. Liée à cette structure, MBNL inhiberait la reconnaissance du site 3' accepteur d'épissage par le spliceosome et conduirait ainsi à l'exclusion de l'exon (Warf *et al.*, 2009a). Un mécanisme similaire a également été décrit pour la régulation de l'exon F du pré-ARNm de la troponine T du muscle squelettique (Yuan *et al.*, 2007). L'action de MBNL sur l'épissage pourrait donc être liée à sa capacité à stabiliser des structures secondaires, MBNL séquestrerait alors des éléments activateurs ou inhibiteurs conduisant respectivement à un effet répresseur ou activateur sur l'épissage (Warf et Berglund, 2007 ; Yuan *et al.*, 2007).

En corrélation avec le niveau d'expression des protéines et en particulier avec l'absence d'expression de MBNL3 dans les tissus musculaires matures, MBNL1 et MBNL2 favorisent la différenciation des cellules musculaires alors que MBNL3 l'inhibe (Squillace *et al.*, 2002). Le mécanisme moléculaire par lequel MBNL3 empêche la différenciation musculaire n'est pas encore identifié, cependant il semble lié à des effets opposés à MBNL1 dans la régulation de l'épissage alternatif de transcrits spécifiques. Alors que les 3 isoformes de MBNL régulent de la même manière l'épissage de différents gènes tels que cTNT et IR

(Ho *et al.*, 2004), la surexpression de MBNL3 inhibe des gènes spécifiques des muscles comme ceux de la myogénine et de la myosine (Squillace *et al.*, 2002).

IV.4.b. Les autres fonctions des protéines MBNL

Récemment, il a été proposé que MBNL1 joue également un rôle dans la régulation de la biogenèse des miRNA. En effet, la fixation de MBNL1 au niveau d'un motif YGCY localisé dans la boucle du pré-miR-1 inhibe la fixation et l'action de Lin 28, protéine responsable du blocage de l'action de DICER par poly-uridylation en 3' du pré-miRNA. Ainsi, MBNL1 jouerait un rôle de protecteur de la dégradation de pré-miR-1 mais ne serait pas un activateur direct de sa biogenèse (Rau *et al.*, 2011). De nombreux pré-miRNA contenant un motif YGCY dans leur boucle, un rôle plus général de MBNL1 dans la maturation des pré-miRNA est envisageable. De plus, MBNL1 colocalise en conditions de stress avec les granules de stress cytoplasmiques, d'où l'hypothèse d'un rôle de MBNL1 dans la régulation de la stabilité des ARNm ou dans la traduction (Onishi *et al.*, 2008)

De même, MBNL2 participe dans le cytoplasme à la localisation subcellulaire de l'intégrin- $\alpha 3$, constituant des complexes d'adhésions (Adereth *et al.*, 2005). MBNL2 est alors principalement exprimée dans ou près des larges plaques cytoplasmiques, qui contiennent aussi la pFAK (pour « phospho Focal Adhesion Kinase). Cette localisation est microtubule-dépendente. Afin de contrôler la localisation de l'intégrin- $\alpha 3$, MBNL2 interagit avec le motif ACACCC, également appelé zipcode, au niveau de la région 3'-UTR de l'ARNm de l'intégrin- $\alpha 3$ (Adereth *et al.*, 2005).

IV.5. Implication de MBNL dans la DM

Comme il a été décrit précédemment, le facteur d'épissage MBNL1 se fixe sur les répétitions CUG et CCUG et se retrouve alors séquestré dans des *foci* nucléaires. Ceci conduit à une diminution de concentration de protéine libre dans le noyau et donc à sa perte de fonction (Fardaei *et al.*, 2001 ; Mankodi *et al.*, 2001 ; Mankodi *et al.*, 2005 ; Jiang *et al.*, 2004 ; Kanadia *et al.*, 2003b). Ses paralogues MBNL2 et MBNL3 peuvent également être séquestrés au sein des *foci* (Miller *et al.*, 2000 ; Fardaei *et al.*, 2002 ; Mankodi *et al.*, 2003), cependant seul l'effet de MBNL1 a été clairement étudié.

MBNL1 semble avoir plusieurs rôles au sein de la DM. Tout d'abord, elle participe à la formation et au maintien des *foci*. Un modèle transgénique de Drosophile DM1 a montré la formation de *foci* dans des cellules musculaires, mais pas dans des cellules neuronales et cela

en dépit de la présence de l'ARNm *DMPK* muté. Or la surexpression de MBNL1 dans les cellules neuronales induit la formation de ces structures (Houseley *et al.*, 2005). De plus, l'inactivation de MBNL1 dans des myoblastes de patients DM1 conduit à une diminution drastique du nombre de *foci*, indiquant l'importance de la protéine dans le maintien des *foci* (Dansithong *et al.*, 2005).

La protéine MBNL1 semble également jouer un rôle important dans le phénotype des patients DM. Des souris mutantes KO pour MBNL1 (*Mbnl1* Δ E3/ Δ E3) présentent des caractéristiques communes avec des souris DM1 (HSALR) et les patients DM, telles que des anomalies du muscle squelettique, une myotonie et une cataracte (Kanadia *et al.*, 2003a). La surexpression de MBNL1 dans des souris DM exprimant des expansions de triplets CTG entraîne une réduction considérable de la myotonie (Kanadia *et al.*, 2006). Le rôle de MBNL1 dans le phénotype des patients DM est lié à sa fonction de régulateur de l'épissage alternatif. La comparaison de puces à exons réalisées sur une souris DM1 exprimant des expansions de triplets CUG et une souris déplétée de MBNL1 montre que plus de 80% des dérégulations de l'épissage observées suite à la perte de MBNL1 se retrouvent chez les souris surexprimant des répétitions, mais à l'inverse la perte de MBNL1 n'explique qu'environ 50% des dérégulations de l'épissage observées chez les patients DM (Du *et al.*, 2010). L'implication de MBNL1 a particulièrement été étudiée dans l'épissage alternatif de la troponine T cardiaque (cTNT) dont l'exon 5 devient majoritairement inclus dans la DM1 (Philips *et al.*, 1998). La surexpression de MBNL1 induit bien une augmentation d'exclusion alors que sa perte induit une inclusion de cet exon (Ho *et al.*, 2004). Pour le récepteur à insuline (IR), l'exon 11 est préférentiellement exclu chez les patients DM1 (Savkur *et al.*, 2001 ; Savkur *et al.*, 2004). Dans des cultures primaires de muscles squelettiques de poulet, la surexpression de MBNL1 amène à l'inclusion de l'exon 11, tandis que sa perte de fonction induit son exclusion (Ho *et al.*, 2004). Dans des myoblastes humains, l'inactivation de MBNL1 confirme l'exclusion de l'exon 11, tandis que sa surexpression n'a pas de conséquences (Dansithong *et al.*, 2005). L'effet de MBNL1 est donc dépendant du contexte cellulaire.

V. Les protéines de la famille CELF

La seconde famille de protéines impliquée dans la DM est la famille des protéines CELF (pour CUG-BP and Embryonically lethal abnormal vision-type RNA binding-protein 3-Like Factor), également appelées BRUNOL (pour BRUNO-Like protein) en raison de leur similarité de séquence avec la protéine Bruno présente chez la drosophile (Good *et al.*, 2000 ; Ladd *et al.*, 2001). Cette famille de protéines, conservée chez les eucaryotes supérieurs, joue

un rôle essentiel dans différentes étapes de régulation post-transcriptionnelle de l'expression des gènes telles que l'épissage alternatif, la traduction et la stabilité des ARNm.

V.1. Généralités

En 1995, l'équipe de Timchenko a montré qu'une protéine présente dans un extrait de cellules HeLa avait la capacité de se lier *in vitro* à une séquence ARN (CUG)₈ (Timchenko *et al.*, 1996b). La même équipe a caractérisé cette protéine et l'a nommée CUG-BP pour CUG Binding Protein (Timchenko *et al.*, 1996a). Puis, sur la base de l'homologie de séquences entre les protéines CUG-BP (CUG-BP1/CELF1) et ETR3 (Embryonic Lethal Abnormal Vision Type Ribonuclear Protein/CELF2/CUG-BP2), la famille de protéines appelée CELF (pour « CUG-BP and ETR3-Like Factors ») a été établie (Ladd *et al.*, 2001).

Chez l'Homme, six membres de la famille CELF ont été identifiés : CUG-BP1 (ou CELF1/BRUNOL2) ; ETR-3 (ou CELF2/CUG-BP2/NAPOR/BRUNOL3) ; CELF3 (ou BRUNOL1) ; CELF4 (ou BRUNOL4) ; CELF5 (ou BRUNOL5) et enfin CELF6 (ou BRUNOL6) (Li *et al.*, 2001 ; Good *et al.*, 2000 ; Ladd *et al.*, 2004). Les protéines CELF humaines peuvent être classées en deux sous-familles sur la base de la phylogénie. La première sous-famille contient les protéines CUG-BP1 et ETR-3, dont les séquences présentent 75 % d'identité. La seconde sous-famille correspond aux protéines CELF 3-6, qui présentent 64 % d'identité entre elles et 44 % d'identité avec CUG-BP1 et ETR-3 (Tableau VII) (Ladd et Cooper, 2004 ; Barreau *et al.*, 2006b).

	CELF 1 (Q92879)	CELF 2 (NP_006552)	CELF 3 (AAK07474)	CELF 4 (NP_064565)	CELF 5 (NP_068757)	CELF 6 (NP_443072)
CELF 1	100	76	45	42	44	43
CELF 2		100	43.5	43	45	46
CELF 3			100	61	62	64
CELF 4				100	63.5	65
CELF 5					100	64.5
CELF 6						100

Tableau VII : Pourcentage d'identité de séquence entre les 6 différentes protéines de la famille CELF.

Les protéines des sous-familles 1 (CUG-BP et ETR-3) et 2 (CELF 3-6) sont soulignées en jaune et rouge, respectivement.

(D'après Barreau *et al.*, 2006b).

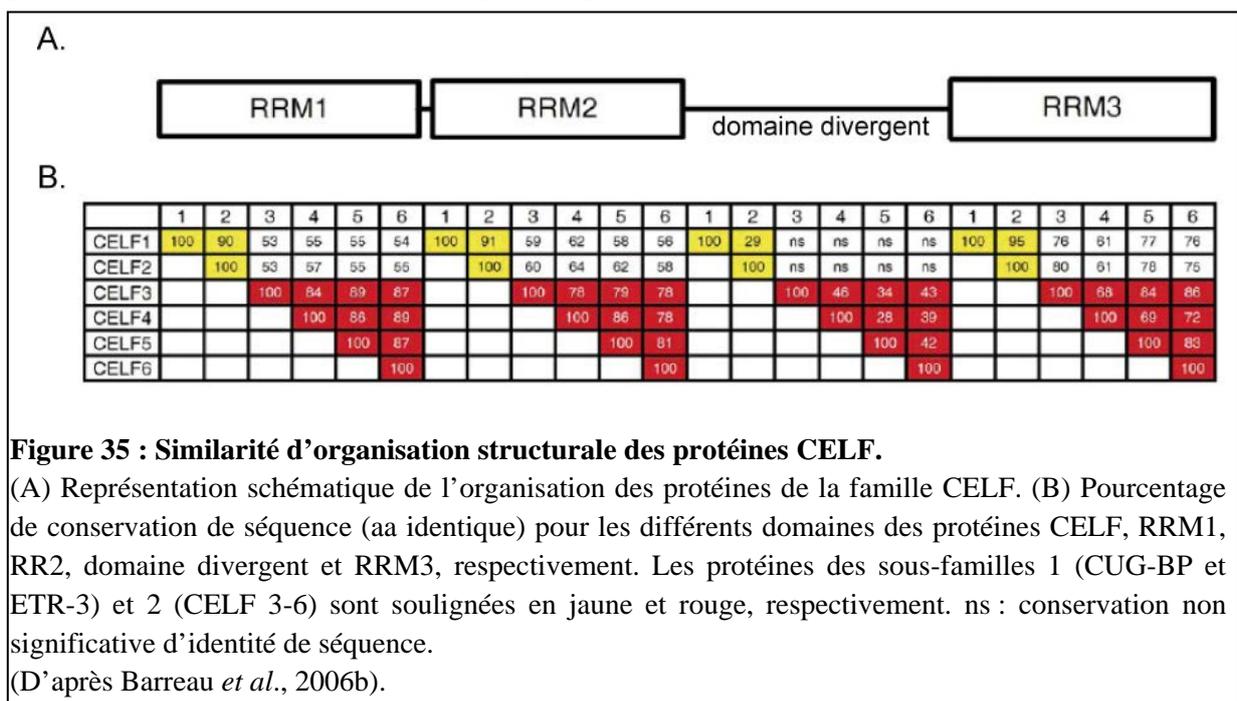
Les protéines CUG-BP1 et ETR-3 sont exprimées de manière ubiquitaire avec un enrichissement dans le cœur, le cerveau et les muscles (Choi *et al.*, 1999 ; Ladd *et al.*, 2001). Les protéines CELF3, 4 et 5 sont uniquement exprimées chez l'adulte au niveau du système nerveux central tandis que la protéine CELF6 est enrichie dans le cerveau et les testicules (Meins *et al.*, 2002 ; Loria *et al.*, 2003 ; Ladd *et al.*, 2004). L'expression des protéines CELF

varie également en fonction du stade de développement. Ainsi, les protéines CUG-BP1 et ETR-3 semblent jouer un rôle vital dans l'embryogenèse avec, par exemple, la protéine ETR-3 qui régule l'épissage alternatif du récepteur au glutamate NMDAR1 dans différentes régions du cerveau au cours du développement (Zhang et Abdel-Rahman, 2002 ; Ladd *et al.*, 2001). De manière intéressante, l'expression de CUG-BP1 diminue au cours du développement dans le muscle squelettique et le cœur (Ladd *et al.*, 2005).

La protéine CUG-BP1 fut découverte par un test de liaison à de courts motifs CUG simple brin (Timchenko *et al.*, 1996a; Timchenko *et al.*, 2001), toutefois, ces résultats furent rapidement discutés, et il apparaît aujourd'hui que CUG-BP1 préfère lier des motifs UGU ou UGUU (Marquis *et al.*, 2006).

V.2. Structure des protéines

Les protéines CELF présentent une forte similarité d'organisation structurale. Elles sont constituées de trois motifs de reconnaissance de l'ARN (RRM). Deux des motifs RRM sont situés dans la région N-terminale (RRM1 et RRM2) et un dans la région C-terminale (RRM3) (Fig.35A). Les motifs RRM2 et 3 sont séparés par un domaine divergent composé de 160 à 230 résidus. Les motifs RRM présentent une forte similarité de séquence, alors que la séquence du domaine divergent est beaucoup moins bien conservée (Ladd *et al.*, 2001). En effet, ce domaine divergent présente 29 % d'identité au sein de la sous-famille 1 et entre 28 et 46 % d'identité au sein de la sous-famille 2, mais aucune similarité n'est observée entre les protéines des deux sous-familles (Fig.35B) (pour revue, Barreau *et al.*, 2006b).



Des expériences de délétion ont permis de montrer que la région N-terminale, contenant les motifs RRM1 et 2, est responsable de la fixation à l'ARN. Suivant le membre considéré de la famille CELF, le 3^{ème} motif RRM a ou n'a pas, de capacité de fixation à l'ARN. Le rôle du domaine divergent n'est pas clairement défini, cependant il pourrait jouer un rôle dans la localisation cellulaire de certaines protéines CELF. Ainsi, la protéine ETR-3 possède deux signaux de localisation nucléaire (NLS), correspondant aux 53 derniers acides aminés du domaine divergent et aux 20 acides aminés C-terminaux. Il a également été proposé que le domaine divergent de CUG-BP1 participe à sa localisation au sein des granules de stress présents dans les cellules DM (Fujimura *et al.*, 2008).

Différentes isoformes de chacune des protéines CELF sont produites par épissage alternatif et co-existent dans les cellules (pour revue, Barreau *et al.*, 2006b). L'expression de ces isoformes semble être régulée en fonction de l'étape de développement et des tissus. Il existe ainsi quatre isoformes de la protéine CUG-BP1, et sept de la protéine ETR-3 (pour revue, Barreau *et al.*, 2006b).

V.3. Les différents rôles des protéines CELF

Les protéines CUG-BP1 et ETR3 sont localisées aussi bien dans le noyau que dans le cytoplasme et l'équilibre entre ces deux compartiments cellulaires dépend du contexte cellulaire environnemental. Cette double localisation suggère que les protéines CELF puissent avoir des rôles différents en fonction de leur localisation cellulaire. Ainsi, puisque les protéines CELF sont impliquées, au niveau du noyau, dans la régulation de l'épissage alternatif d'ARNm cibles, tandis que dans le cytoplasme elles peuvent jouer un rôle dans la régulation de la traduction et de la stabilité des ARNm (pour revue, Barreau *et al.*, 2006a).

V.3.a. Dans l'épissage alternatif

Les protéines CELF ont été caractérisées aussi bien comme régulateurs positifs que négatifs de l'épissage alternatif. Ainsi, dans le cas de l'épissage de l'exon 5 de la troponine T cardiaque ou de l'exon 6B de la β -tropomyosine elles régulent positivement l'épissage, tandis que dans le cas de l'exon 11 du récepteur de l'insuline elles régulent négativement l'épissage. Différents mécanismes d'action des protéines CELF ont été proposés. Dans le cas de l'exon 21 de NMDAR1, ETR-3 joue un rôle activateur en entrant en compétition avec un facteur répresseur. Cette compétition est directe puisque ETR-3 se fixe sur le site du répresseur pour empêcher son action (Zhang et Abdel-Rahman, 2002). Dans d'autres cas, comme celui de la

troponine T cardiaque chez le poulet (cTNT), ETR-3 induit l'inclusion de l'exon 5 par interaction directe avec des composants de la snRNP U2 (17S snRNP) permettant le recrutement et/ou la stabilisation du spliceosome (Goo et Cooper, 2009). Les sites de fixation des protéines CELF peuvent se trouver aussi bien en amont qu'en aval des exons régulés et cette localisation différente pourrait être corrélée aux différents mécanismes d'action proposés pour les protéines CELF (pour revue, Barreau *et al.*, 2006b).

Les protéines CELF sont des antagonistes des protéines MBNL pour la régulation d'un certain nombre de pré-ARNm comme cTNT et IR. Cependant, les sites de fixation des deux protéines sont distincts, éliminant un modèle de compétition de fixation pour expliquer leur antagonisme (Dansithong *et al.*, 2005 ; Ho *et al.*, 2004). De plus, la mutation des sites de fixation des protéines CELF et MBNL a permis de montrer que ces deux familles de protéines agissent de façon indépendante l'une de l'autre (Ho *et al.*, 2004). De même, la protéine PTB agit également comme un antagoniste des protéines CELF dans la régulation de l'épissage de l'exon 6B de la β -tropomyosine (Sureau *et al.*, 2011).

La régulation de l'épissage alternatif par les protéines CELF est associée à leur fixation sur des séquences simple brin riches en U/G (Faustino et Cooper, 2005). Les domaines intervenant dans l'activité de régulation de l'épissage ne sont pas identiques dans tous les membres de la famille CELF. Des délétions effectuées dans ETR-3 ont permis de montrer que la région contenant les motifs RRM1 et 2 et les 71 premiers résidus du domaine divergeant, ainsi que la région contenant le motif RRM3 et les 120 derniers résidus du domaine divergeant, possèdent l'activité d'épissage (Singh *et al.*, 2004 ; Han et Cooper, 2005). De même une étude portant sur CELF4 montre la nécessité de 2 domaines localisés dans le domaine divergeant requis de façon indépendante ou concomitante dans la régulation de l'épissage alternatif des pré-ARNm cibles (Han et Cooper, 2005).

V.3.b. Dans la régulation post-transcriptionnelle

En plus de leur rôle dans l'épissage alternatif, CUG-BP1 et ETR-3 interviennent dans le contrôle de la traduction, la dégradation et la stabilité d'ARNm.

La protéine CUG-BP1 peut fonctionner comme activateur ou répresseur de la traduction par fixation sur des éléments de séquence riches en G/C au sein de la région 5' UTR d'un ARNm (pour revue, Barreau *et al.*, 2006c ; Iakova *et al.*, 2004). Sa fixation sur ces éléments nécessite son hyperphosphorylation (Iakova *et al.*, 2004). De même, les protéines CUG-BP1 et ETR-3 peuvent également contrôler la dégradation et la déadénylation des

ARNm par interaction avec un élément riche en GU (appelé EDEN pour Embryo Deadenylation ElemeNt (EDEN) ou GRE pour GU-Rich Element) localisé dans l'extrémité 3'-UTR des ARNm (Fig.36) (Paillard *et al.*, 2003 ; Moraes *et al.*, 2006 ; pour revue Vlasova et Bohjanen, 2008).

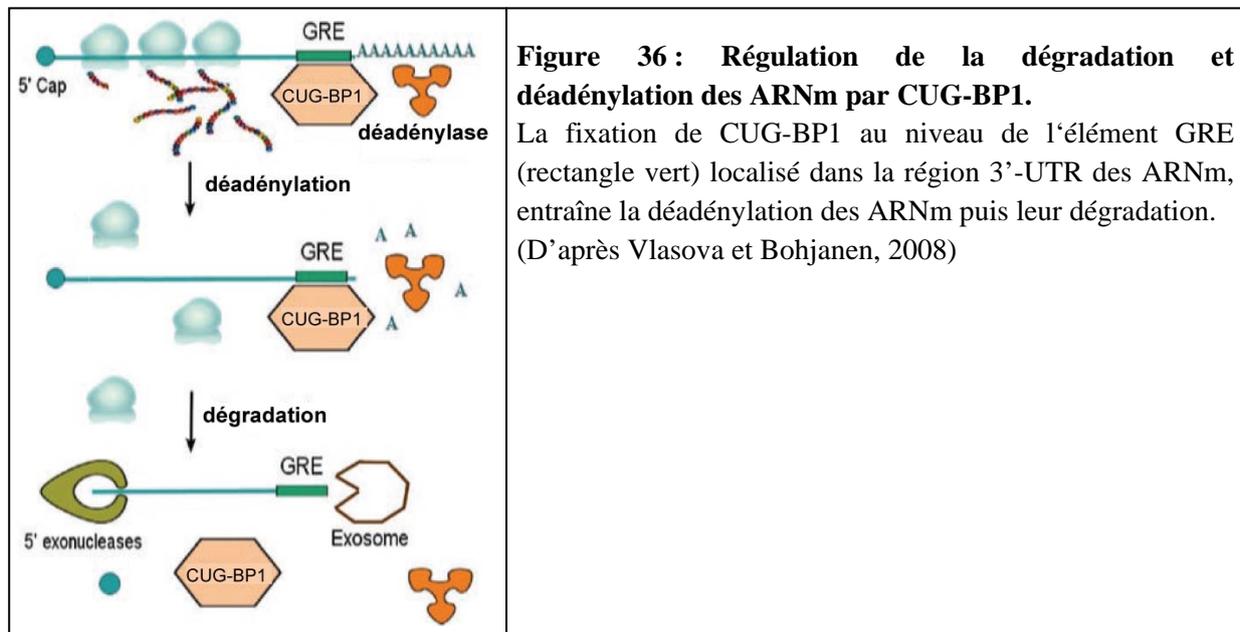


Figure 36 : Régulation de la dégradation et déadénylation des ARNm par CUG-BP1.

La fixation de CUG-BP1 au niveau de l'élément GRE (rectangle vert) localisé dans la région 3'-UTR des ARNm, entraîne la déadénylation des ARNm puis leur dégradation. (D'après Vlasova et Bohjanen, 2008)

A l'inverse de ce qui est observé pour la traduction, seules les protéines CELF déphosphorylées peuvent réguler la dégradation et la déadénylation des ARNm (Detivaud *et al.*, 2003). Ainsi, les différentes fonctions de CUG-BP sont associées aux différents états de phosphorylation de la protéine. La balance entre la fixation avec les régions riches en G/C associée au contrôle de la traduction, et la fixation sur les GRE associée avec la dégradation et déadénylation des ARNm est contrôlée par des voies de signalisation qui activent ou inhibent les kinases appropriées.

Par ailleurs, la capacité des protéines CELF à se fixer *in vivo* sur ces différentes cibles dépend également de la présence ou de l'absence d'autres protéines dans un complexe multifonctionnel, comme eIF2 pour la traduction et PARN (pour Poly A RiboNuclease) pour la déadénylation (Moraes *et al.*, 2006 ; Timchenko *et al.*, 2005). Ainsi, l'association de CUG-BP1 avec ses co-facteurs peut être affectée par l'état de phosphorylation de la protéine ou par le profil d'expression des co-facteurs.

V.4. Implication des protéines CELF dans la DM

La perturbation de la fonction des protéines CELF peut jouer un rôle dans différentes pathologies, telle que la DM1. En effet, l'analyse du niveau d'expression de CUG-BP1 et ETR-3 dans des lignées cellulaires et des souris transgéniques exprimant l'ARNm *DMPK*

muté ainsi que dans des tissus de patients atteints de DM1 révèle une augmentation de la quantité de protéines CELF dans les muscles squelettiques et le cœur (Dansithong *et al.*, 2005 ; Savkur *et al.*, 2001 ; Timchenko *et al.*, 2001 ; Timchenko *et al.*, 2004 ; Wang *et al.*, 2007 ; Orengo *et al.*, 2008). Par ailleurs, la phosphorylation et la distribution cellulaire de CUG-BP1 sont également altérées dans la DM1 (Roberts *et al.*, 1997 ; Kuyumcu-Martinez *et al.*, 2007). Chez les patients atteints de DM1, on observe une accumulation des protéines CUG-BP1 et ETR-3 dans le noyau des cellules. Cependant, contrairement à ce qui est observé pour MBNL1, cette accumulation n'est pas due à une séquestration de la protéine par les répétitions CUG présentes à l'extrémité 3' de l'ARNm *DMPK* (Fig.37).

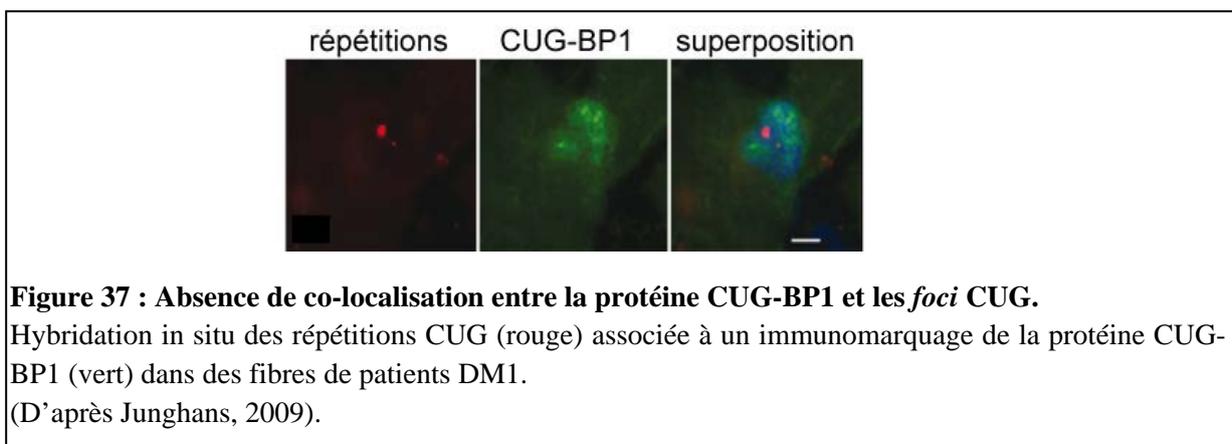


Figure 37 : Absence de co-localisation entre la protéine CUG-BP1 et les foci CUG.

Hybridation in situ des répétitions CUG (rouge) associée à un immunomarquage de la protéine CUG-BP1 (vert) dans des fibres de patients DM1.

(D'après Junghans, 2009).

En effet, cette augmentation résulte d'une augmentation de la stabilité de la protéine CUG-BP1 du fait de son hyperphosphorylation par la protéine PKC (pour Protein Kinase C) (Kuyumcu-Martinez *et al.*, 2007). Le mécanisme par lequel l'ARN mutant *DMPK* active la voie d'activation de PKC est encore inconnu. Il est à noter qu'il existe une controverse quant à la surexpression de CUG-BP1. Ainsi une augmentation de CUG-BP1 est retrouvée dans un modèle murin exprimant de longues répétitions CUG interrompues (Wang *et al.*, 2007), mais pas dans un modèle exprimant des répétitions « pures » plus courtes (Lin *et al.*, 2006). Bien que ces données soient contradictoires, un consensus semble apparaître dans la littérature : une augmentation de la concentration nucléaire de CUG-BP1 chez les patients DM1, corrélée à une diminution de sa concentration cytoplasmique. La situation est toutefois moins claire chez les patients DM2. En effet, deux études analysant des extraits protéiques totaux issus de cellules DM2 ne montrent pas d'augmentation de CUG-BP1 (Pelletier *et al.*, 2009), alors qu'une analyse utilisant des extraits cytoplasmiques issus de myoblastes de DM2 révèle une augmentation (Salisbury *et al.*, 2009). De plus, l'expression de CUG-BP1 serait augmentée dans un modèle murin DM2 exprimant de longues répétitions CCUG (Salisbury *et al.*, 2009).

Plusieurs arguments montrent l'implication de CUG-BP1 dans la DM1 et plus particulièrement dans le phénomène d'épissage aberrant de certains ARNm cibles, tels que cTNT, CIC-1 et IR (Ladd et Cooper, 2004). Dans des souris transgéniques surexprimant la protéine CUG-BP1 humaine, un retour à un profil d'épissage de type fœtal est observé pour un grand nombre de pré-ARNm du cœur et des muscles squelettiques tels que ceux de la troponine T cardiaque (cTNT), du canal à chlore (CIC-1) ou SERCA 1, reproduisant ainsi les altérations d'épissage observées chez les patients atteints de DM (Ho *et al.*, 2005a ; Timchenko *et al.*, 2004 ; Kalsotra *et al.*, 2008 ; Koshelev *et al.*, 2010 ; Ward *et al.*, 2010). De même, des mutations des sites de fixation de CUG-BP1 sur les transcrits hcTNT et CIC-1 les rendent incapables de répondre à l'effet *trans*-dominant des expansions de triplets (Philips *et al.*, 1998 ; Savkur *et al.*, 2001). Si cette incapacité de réponse aux CTG prouve l'implication de CUG-BP1 dans l'effet *trans*-dominant sur ces transcrits, elle montre aussi que MBNL1 n'est pas le seul facteur à y contribuer.

CUG-BP1 pourrait également jouer un rôle dans l'inhibition de la traduction de différents ARNm observée chez les patients atteints de DM1. En effet, les expansions de répétitions CUG entraînent la formation de granules de stress au niveau du cytoplasme où se localise alors CUG-BP1 (Fujimura *et al.*, 2008 ; Huichalaf *et al.*, 2010). Cette localisation de la protéine dans les granules pourrait avoir un impact sur son rôle dans la traduction et la stabilité des ARNm.

VI. Exemple de la dérégulation de l'épissage chez les patients atteints de DM1 : l'exon 5 de la troponine T cardiaque (cTNT)

Comme son nom l'indique, le symptôme caractéristique des DM est une atteinte musculaire, en particulier des muscles squelettiques et cardiaques. L'atteinte cardiaque est à l'origine de 20% des décès des patients atteints de DM1, d'où l'importance de son étude. Elle est caractérisée par des défauts de conduction et de rythme cardiaques (Bushby *et al.*, 2003 ; Vinereanu *et al.*, 2004). Dans le cœur, l'un des transcrits dont l'épissage alternatif est dérégulé est celui de la Troponine T cardiaque (cTNT). Bien que la contribution directe de ce transcrit à l'atteinte cardiaque n'ait pas été démontrée, il reste parmi les plus étudiés.

VI.1. Rôle de la troponine T cardiaque (cTNT) dans la contraction musculaire

La troponine T (TNT), avec la troponine C (TNC) et la troponine I (TNI) entrent dans la formation du complexe de la troponine, qui régule la contraction musculaire des muscles striés auxquels appartient le cœur. Dans ce complexe, la troponine T interagit avec la

tropomyosine, qui inhibe l'interaction entre les filaments d'actine et de myosine et donc la contraction du muscle. La troponine I est la sous-unité inhibitrice qui se lie à l'actine, et la troponine C se lie aux ions calcium. En présence de Ca^{2+} , l'interaction entre TNC et TNI est augmentée, libérant ainsi l'actine. Ainsi, l'affinité de TNT pour la tropomyosine diminue, entraînant un changement de conformation de la tropomyosine. Ce changement de conformation de la tropomyosine permet son déplacement et la libération des sites de fixation spécifique de la myosine présents sur la molécule d'actine. L'actine interagit alors avec la myosine entraînant une contraction musculaire. La troponine T cardiaque (cTNT ou TNNT2) joue un rôle dans la fibrillogenèse du muscle et sa mutation conduit à des cardiomyopathies (White *et al.*, 1987 ; Epstein et Fischman, 1991 ; pour revue Gomes *et al.*, 2004).

VI.2. Généralités

Chez les vertébrés, trois gènes codant la troponine T ont été identifiés. Le gène *TNNT1*, situé au niveau du chromosome 19, est exprimé spécifiquement dans les muscles squelettiques lisses, alors que le gène *TNNT3*, situé au niveau du chromosome 11, est exprimé dans les muscles squelettiques striés. Le gène de la troponine T cardiaque (cTNT), également appelé *TNNT2*, est composé de 17 exons et se situe sur le chromosome 1 en position 1q32. Dans le cœur humain, les exons 4 et 5 sont alternatifs, générant ainsi 4 isoformes différentes, appelées *TnT1* à 4, qui correspondent respectivement aux isoformes 4+5+, 4-5+, 4+5-, et 4-5-. L'expression de ces isoformes évolue au cours du développement. Ainsi, dans le cœur foetal, l'isoforme majeure est l'isoforme *TnT1*, alors que chez l'adulte, l'exon 5 devient exclu et c'est l'isoforme *TnT3* qui devient dominante. Les isoformes *TnT2* et 4 sont également des isoformes fœtales (Gomes *et al.*, 2002). Dans le muscle squelettique, *TNNT2* est exprimé chez le fœtus, mais devient silencieux chez l'adulte (Cooper et Ordahl, 1984). La présence de différentes isoformes de cTNT est importante pour la régulation de la contraction musculaire en fonction de la teneur en calcium. Il a été montré que l'isoforme 4+5+ était plus sensible à la variation du calcium que les isoformes 4+5- et 4-5-. L'exclusion de l'exon 5 entraîne la perte d'une séquence de 10 acides aminés très acides (EEEEWLEEDD) dans la région N-terminale de la protéine (Cooper et Ordahl, 1984 ; Cooper et Ordahl, 1985 ; Long et Ordahl, 1988). Les exons alternatifs de cTNT, et notamment l'exon 5, sont donc importants pour le contrôle de la force musculaire (Gomes *et al.*, 2004 ; Reiser *et al.*, 1992 ; Ogut *et al.*, 1999 ; Nassar *et al.*, 1991).

VI.3. Régulation et dérégulation de l'épissage alternatif de l'exon 5 de cTNT dans la DM

Au cours du développement, le profil d'épissage de l'exon 5 de cTNT change pour passer de l'inclusion vers l'exclusion de cet exon. Ce changement de profil d'épissage est conservé parmi les espèces, on le retrouve ainsi chez l'homme, la souris, le rat ou le poulet, ce qui suggère un mécanisme commun de régulation, bien que la séquence du gène *TNNT2* varie d'une espèce à l'autre (Cooper et Ordahl, 1985 ; Jin et Lin, 1989 ; Townsend *et al.*, 1995). Dans le cas de la pathologie DM1, le défaut d'épissage de cTNT conduit à une inclusion préférentielle de l'exon 5, ce qui correspond au retour à un profil d'épissage de type foetal (Philips *et al.*, 1998).

Les bases moléculaires de la régulation de l'inclusion de l'exon 5 ont été étudiées chez le poulet. Le modèle proposé à l'heure actuelle est présenté dans la Figure 38. Quatre éléments de régulation de l'épissage de l'exon 5, appelés MSE1 à 4 (pour « Muscle Specific Enhancer»), ont été localisés dans les régions introniques bordant l'exon 5, un en amont et 3 en aval de cet exon (Ryan et Cooper, 1996 ; Cooper, 1998). Les protéines MBNL1 et PTB (pour « Polypyrimidine Tract Binding protein »), qui inhibent l'inclusion de l'exon 5, se fixent au niveau des MSE 1 et 4 (Charlet *et al.*, 2002a ; Ho *et al.*, 2004). A l'inverse, toutes les protéines de la famille CELF activent l'inclusion de l'exon, la protéine ETR-3 se fixe au niveau des MSE2 et 3, et CUG-BP1 au niveau du MSE2. Le mode d'action des protéines n'est pas clairement défini, cependant, il a été mis en évidence que la protéine PTB entre en compétition avec CUG-BP1 et empêche sa fixation sur son site de liaison (Ladd *et al.*, 2005). De plus, il a récemment été montré que ETR-3 en se fixant en aval de l'exon recrute la snRNP U2 au niveau du site de branchement présent, lui, en amont de l'exon (Goo et Cooper, 2009). Cette capacité à recruter le spliceosome sur un site présent du côté opposé de l'exon souligne le dynamisme des interactions ARN/protéine et l'importance des remaniements de l'ARN dans ces mêmes interactions. En plus de ces facteurs de régulation introniques, un élément ESE a été localisé au niveau de l'extrémité 5' de l'exon 5 et semble permettre la fixation de différentes protéines SR telles que SRSF5, SRSF6, SRSF4 et SRSF1, qui favorisent l'inclusion de l'exon 5 (Ramchatesingh *et al.*, 1995 ; Nagel *et al.*, 1998 ; Stark *et al.*, 1998).

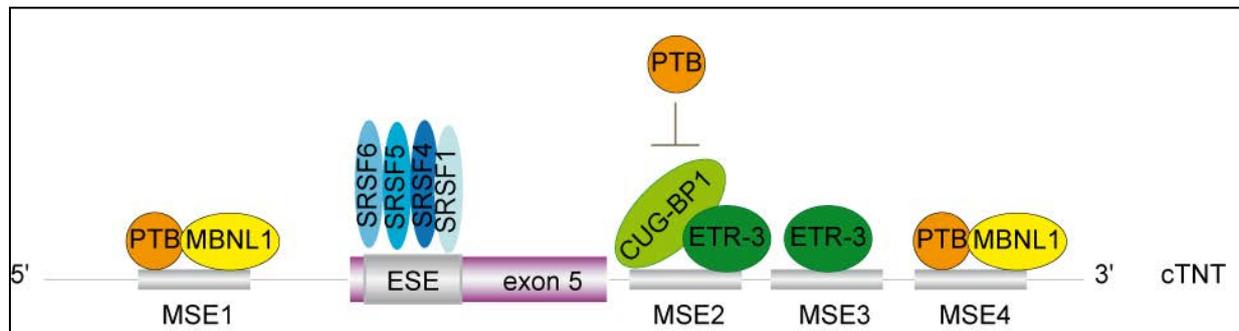


Figure 38 : Régulation de l'épissage alternatif de l'exon 5 de la Troponine T cardiaque (cTNT) du poulet.

L'exon 5 de cTNT est régulé par un «enhancer» exonique (ESE), et quatre autres introniques appelés MSE 1 à 4 (Muscle Specific Enhancers). Ils tirent leur nom de leur implication dans une régulation spécifique du muscle. Plusieurs protéines SR favorisent l'inclusion de l'exon en fixant l'ESE. Au niveau intronique, PTB et MBNL1 qui lient les MSE1 et 4 inhibent l'épissage, tandis que toutes les protéines CELF l'activent. ETR-3 fixe le MSE 2 et 3, tandis que CUGBP1 ne se fixe que sur le MSE2. PTB peut entrer en compétition avec CUGBP1 et empêcher son action sur ce site. En plus de ces facteurs de régulation introniques, un élément ESE a été localisé au niveau de l'extrémité 5' de l'exon 5 et semble permettre la fixation de différentes protéines SR telles que SRSF5, SRSF6, SRSF4 et SRSF1, qui favorisent l'inclusion de l'exon 5.

(D'après Ladd *et al.*, 2005).

Les mécanismes permettant la régulation de l'inclusion de l'exon 5 de hcTNT chez l'Homme ont également été étudiés. Deux sites de fixation de la protéine MBNL1 ont été identifiés entre les positions -40 et -21 en amont de l'exon 5 (Ho *et al.*, 2005b). Ces deux sites encadrent la séquence polypyrimidine essentielle à la réaction d'épissage. Il a été proposé que la fixation de MBNL1 sur ces deux sites inhibe l'inclusion de l'exon 5 en stabilisant une structure tige-boucle formée par appariement des deux sites de fixation de MBNL1 entre eux. Cette structure tige-boucle empêcherait la reconnaissance de la séquence polypyrimidine par le facteur U2AF65 et par conséquent le recrutement de la snRNP U2 (Warf et Berglund, 2007 ; Warf *et al.*, 2009a) (Fig.39). Par ailleurs, des sites de fixation des protéines de la famille CELF ont été identifiés dans les 50 premiers nucléotides de l'intron 5 (Ho *et al.*, 2005b). Cependant, le mécanisme d'activation de l'inclusion de l'exon 5 par les protéines CELF est encore inconnu.

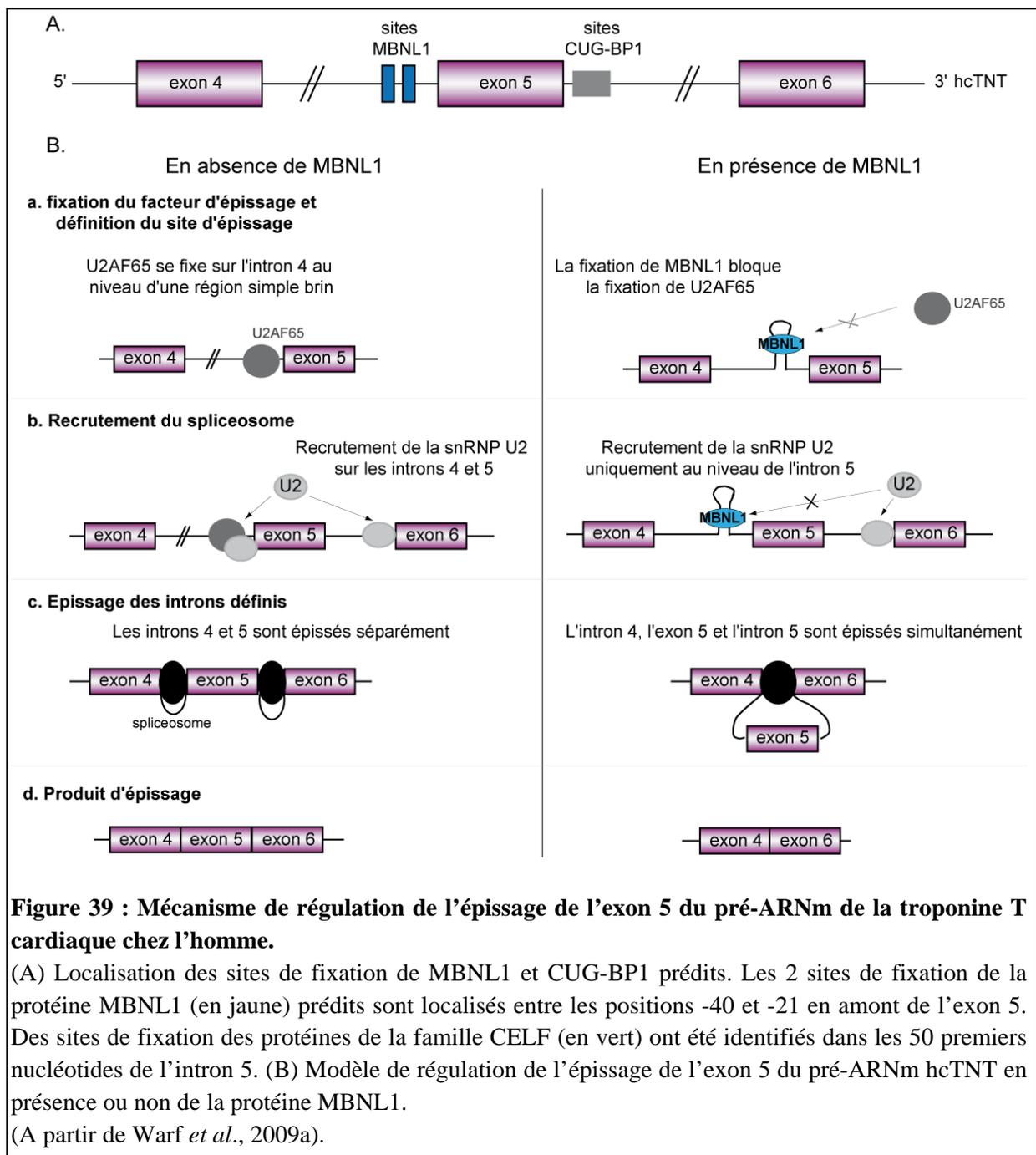


Figure 39 : Mécanisme de régulation de l'épissage de l'exon 5 du pré-ARNm de la troponine T cardiaque chez l'homme.

(A) Localisation des sites de fixation de MBNL1 et CUG-BP1 prédits. Les 2 sites de fixation de la protéine MBNL1 (en jaune) prédits sont localisés entre les positions -40 et -21 en amont de l'exon 5. Des sites de fixation des protéines de la famille CELF (en vert) ont été identifiés dans les 50 premiers nucléotides de l'intron 5. (B) Modèle de régulation de l'épissage de l'exon 5 du pré-ARNm hcTNT en présence ou non de la protéine MBNL1.

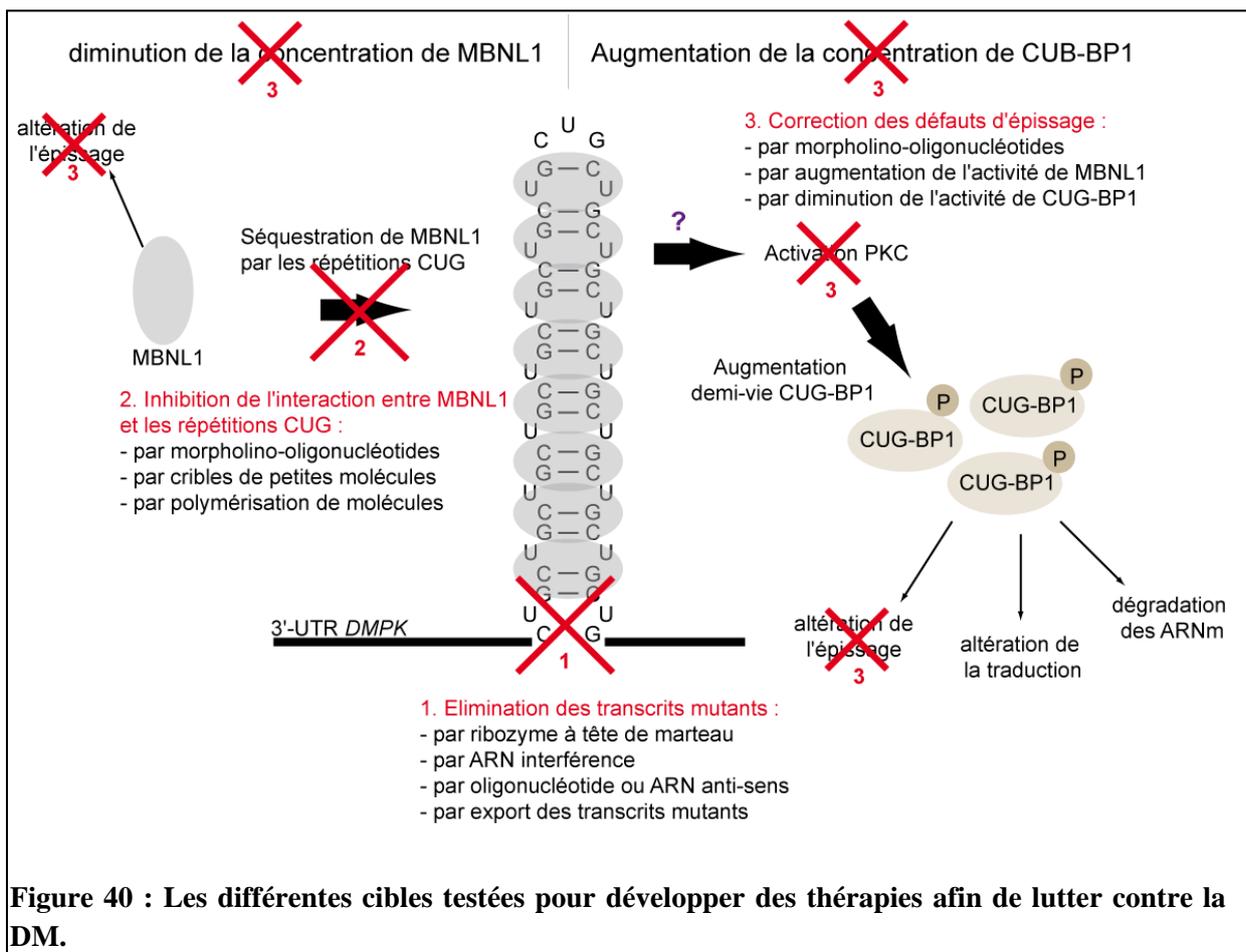
(A partir de Warf *et al.*, 2009a).

Lors du développement du cœur, le changement d'épissage de l'exon 5 est corrélé à des variations d'expression des facteurs régulateurs antagonistes MBNL1 et CUG-BP1. L'expression des activateurs diminue alors que celle des inhibiteurs persiste permettant, de cette manière, l'exclusion de l'exon chez l'adulte (Ladd *et al.*, 2001 ; Ladd *et al.*, 2005 ; Kalsotra *et al.*, 2008 ; Filippova *et al.*, 2001 ; Steinbach *et al.*, 1998 ; Ho *et al.*, 2005b). Dans le cas de la DM, la balance entre ces deux facteurs de régulation de l'épissage est déplacée avec une augmentation de CUG-BP1 (l'activateur dans le cas de cTNT) et surtout une forte diminution du taux de MBNL1 libre (l'inhibiteur). Cette variation de concentration est en

corrélation avec les dérégulations observées, c'est à dire une forte augmentation de l'inclusion de l'exon 5.

VII. Les différentes cibles et approches possibles pour une thérapie globale

A ce jour, aucun remède n'existe pour la DM. Les traitements actuels se limitent à atténuer les signes et les symptômes de la maladie, comme par exemple l'implantation d'un pacemaker afin de maîtriser l'arythmie cardiaque, ou l'absorption de mexiletine ou d'anti-inflammatoires afin d'atténuer les douleurs musculaires. Cependant, ces traitements ne permettent de soigner que certains symptômes et présentent des effets secondaires. De plus il n'existe aucun traitement pour ralentir ou arrêter la progression de la maladie. Comme je vais le décrire par la suite, le but ultime des approches thérapeutiques développées actuellement est l'inversion du phénotype de la maladie. Les différents cibles testées sont l'élimination des transcrits mutants, le blocage de l'interaction entre la protéine MBNL1 et les expansions de répétitions CUG ou CCUG et enfin la correction des défauts d'épissage (Fig.40).



VII.1. Elimination des transcrits mutants

Différents modèles ont permis de montrer qu'une diminution de l'expression des répétitions CTG entraînait une réversion des symptômes et des caractéristiques histopathologiques de la DM1 (Mahadevan *et al.*, 2006 ; Mulders *et al.*, 2009 ; Mulders *et al.*, 2009 ; Francois *et al.*, 2011; Yadava *et al.*, 2008). L'élimination des ARN toxiques chez les patients DM pourrait donc être une stratégie thérapeutique efficace. Différentes approches visant à éliminer les ARN toxiques ont été envisagées :

VII.1.a. Par un ribozyme à tête de marteau

Les ribozymes sont des ARN possédant une activité catalytique permettant de cliver un ARN en 3' d'une séquence spécifique (Rossi, 1999). Les ribozymes à tête de marteau sont de petite taille et peuvent facilement être modifiés pour effectuer un clivage intermoléculaire ciblé sur une molécule d'ARN. Ces propriétés en font un outil thérapeutique potentiel.

L'utilisation d'un ribozyme spécifique du pré-ARNm du gène *DMPK* a montré son efficacité dans le clivage du transcrit muté qu'il réduit jusqu'à 63%. Ses effets s'accompagnent d'une diminution du nombre d'agrégats nucléaires et de la restauration partielle de l'expression de l'isoforme B du récepteur à insuline dans les myoblastes de patients DM1. Cependant, une réduction de 50% l'ARN *DMPK* normal est également observée (Langlois *et al.*, 2003). Une combinaison du ribozyme et du système de promoteur Pol III, permettant un plus fort niveau expression de transcrit ribozyme, pourrait être un outil thérapeutique intéressant pour le traitement de la DM1.

VII.1.b. Par ARN interférence

Une autre stratégie, visant à diminuer la quantité d'ARNm *DMPK* muté, est l'utilisation d'ARNs interférents. La voie de l'ARN interférence est initiée par l'enzyme DICER, une ribonucléase de type III, qui clive les long ARN double brin (ARNdb) en de courts fragments de 21 à 25 nucléotides. L'un des deux brins de chaque fragment, appelé « guide », est ensuite incorporé au complexe RISC (pour RNA-Induced Silencing Complex). Le tout est alors dirigé vers les ARNm possédant une séquence complémentaire au brin guide, le complexe RISC va alors bloquer la traduction ou cliver l'ARNm cible qui est alors dégradé. Un système pour produire artificiellement un ARN interférent dirigé contre une séquence spécifique est l'utilisation de shRNA (pour « short hairpin RNA), également appelé petit ARN à épingle à cheveux. Un shRNA est une molécule d'ARN bicaténaire structurée en épingle à cheveux. Dans la cellule, les shRNA sont pris en charge par Dicer et les siRNA

générés à partir des shRNA sont ensuite incorporés dans le complexe RISC qui va cibler les ARNm possédant une séquence complémentaire aux siRNA.

L'injection d'un shRNA dirigé spécifiquement contre le transcrit *DMPK* à l'aide d'un lentivirus au sein de lignées myoblastes DM1 hétérozygote permet la dégradation du transcrit mutant dans le noyau mais aussi du transcrit sauvage présent dans le cytoplasme (Langlois *et al.*, 2005).

A l'inverse, l'étude de François *et al.* montre que l'injection d'un ARN antisens contenant une séquence poly-CAG insérée dans un snRNA artificiel (U7), permet de dégrader l'ARN muté *DMPK* sans modifier l'expression de l'allèle non muté (François *et al.*, 2011). De plus, cette étude envisage l'utilisation de vecteurs viraux de type AAV (pour « Adeno-Associated Virus) qui permettrait une large distribution et une expression stable.

VII.1.c. Par oligonucléotide ou ARN anti-sens

Une autre approche envisagée basée sur l'utilisation d'ARN antisens est l'emploi d'un oligonucléotide anti-sens 2'-O-méthyl phosphorothioate (2'O MePS) qui ne modifie pas l'épissage ni n'interrompt l'interaction entre MBNL1 et les répétitions CUG, mais qui supprime directement l'expression des expansions d'ARN *DMPK*. Ainsi un oligonucléotide antisens contenant 7 répétitions CAG permet une réduction de 90% de l'expression de l'ARN *DMPK* muté, réduisant l'agrégation de l'ARN au sein des cellules de patients atteints de DM1 et entraînant une réduction partielle des défauts d'épissage (Mulders *et al.*, 2009).

VII.1.d. Par export des transcrits mutants

Une dernière technique pour éliminer les transcrits mutants est de les exporter vers le cytoplasme. En effet, l'export des répétitions dans le cytoplasme ne semble pas avoir d'effet pathogène (Dansithong *et al.*, 2008). Cette stratégie utilise une séquence d'export virale de type WPRE (« Woodchuck Post-transcriptionnal Regulatory Element ») incluse dans un transcrit contenant des répétitions CUG (Mastroiannopoulos *et al.*, 2005). Cependant il reste à déterminer comment insérer cette séquence WPRE dans les répétitions CTG chez les patients DM.

VII.2. Empêcher l'interaction entre la protéine MBNL1 et les expansions de répétitions CUG ou CCUG

L'événement principal de l'apparition des dystrophies myotoniques de types 1 et 2 est la séquestration de MBNL1 par les expansions de répétitions CUG ou CCUG, entraînant la

formation de *foci* nucléaires. Différentes stratégies thérapeutiques visant à inhiber l'interaction entre MBNL1 et l'ARN toxique ont été envisagées. Les *foci* nucléaires n'étant pas toxiques, leur destruction n'était pas nécessaire (Mahadevan *et al.*, 2006). Le domaine de fixation de MBNL1 sur ces différents ARN cibles (pré-ARNm et répétitions CUG ou CCUG) étant identique (les motifs à doigts à Zn²⁺), la réalisation d'un inhibiteur interagissant directement avec MBNL1 n'est pas envisageable. Ce sont donc les répétitions CUG ou CCUG qui ont été choisies comme cibles.

VII.2.a. Morpholino-oligonucléotides

L'utilisation des morpholino-oligonucléotides ou PMO (pour « phosphorodiamidate morpholino oligo ») est une technique d'antisens utilisée pour bloquer l'accès d'autres molécules à une séquence spécifique. Contrairement à de nombreuses techniques d'antisens tels que les petits ARN interférents, les morpholinos ne dégradent pas les molécules d'ARN auxquelles ils se fixent, mais agissent par encombrement stérique. Les morpholino-oligonucléotides sont des molécules synthétiques qui résultent d'un réarrangement de la structure naturelle des acides nucléiques. Ils sont constitués de bases azotées fixées sur des cycles morpholines eux même reliés entre eux par des groupements phosphorodiamidates. Ainsi, l'injection d'un morpholino-oligonucléotide antisens contenant de multiples répétitions de triplet CAG entraîne une inhibition de l'interaction entre MBNL1 et les répétitions CUG, ainsi qu'une redistribution de la protéine MBNL1 provenant des foci. Une dégradation plus importante de l'ARN mutant est également observée, ceci pouvant s'expliquer par une exportation plus importante de l'ARN vers le cytoplasme. Avec la redistribution de MBNL1, les défauts de régulation de l'épissage alternatif sont partiellement corrigés, entraînant une restauration de fonction (Mulders *et al.*, 2009).

VII.2.b. Par cribles de petites molécules

Récemment Warf *et al.* (2009) ont entrepris le criblage d'une banque de petites molécules de molécules ayant une forte affinité pour les acides nucléiques structurés et ont ainsi identifié un composé, la pentamidine, qui bloque l'interaction entre MBNL1 et les répétitions CUG. Ils ont obtenu des résultats encourageants aussi bien *in vitro* qu'*in vivo* dans des modèles de souris DM1 exprimant 250 répétitions dans la région 3'-UTR du gène humain de l' α -actine squelettique. Une diminution du nombre de *foci* CUG ainsi qu'une augmentation de la concentration de MBNL1 libre dans le noyau ont été observées. De plus, l'injection de pentamidine conduit à une réversion partielle du phénotype de dérégulation de l'épissage de

l'exon 7 du pré-ARNm *CIC-1* et de l'exon 22 du pré-ARNm *SERCA1* suggérant une efficacité de la pentamidine (Warf *et al.*, 2009b). Les effets de la pentamidine sont relativement faibles chez les souris, cependant aucun effet secondaire sur des événements d'épissage indépendants des répétitions CUG n'a été observé. Cette molécule pourrait donc servir comme point de départ pour des modifications permettant d'accroître l'efficacité, de réduire la toxicité et d'optimiser la délivrance.

VII.2.c. Par polymérisation de molécules

Une autre approche consiste à polymériser une molécule, connue depuis longtemps pour interagir avec des ARN riches en pyrimidines, et ayant plus particulièrement une forte affinité pour les motifs GC. La polymérisation de ces molécules permet ainsi de limiter la fixation de MBNL1 sur les longues structures tige-boucles formées par les répétitions CUG et CCUG, et non aux motifs GC isolés présents dans les sites de fixations des cibles physiologiques. Trois facteurs importants gouvernent le potentiel inhibiteur de ces molécules pour l'interaction entre les répétitions et MBNL1 : la surface de l'ARN séquestrée par l'assemblage des ligands, l'espace entre les modules de ligands puisqu'une distance plus élevée est nécessaire pour cibler les ARN DM2 (CCUG) par rapport aux ARN DM1 (CUG), et enfin, la flexibilité dans l'architecture utilisée pour présenter les modules de fixation à l'ARN (Disney *et al.*, 2010). C'est précisément cette approche qui a été utilisée par l'équipe de MD. Disney pour mettre au point deux molécules la bisbenzimidazole Hoechst 33258 et la 6'-N-5-hexyolate kanamycin A se liant spécifiquement aux répétitions CUG (Pushechnikov *et al.*, 2009) et CCUG (Lee *et al.*, 2009), respectivement. Ces deux molécules inhibent l'interaction entre MBNL1 et les répétitions et peuvent pénétrer dans les cellules de myoblastes de souris.

Ces approches thérapeutiques sont très prometteuses, cependant il faudra garder à l'esprit que les agents qui ciblent les expansions ARN de type CUG ou CCUG sont également susceptibles d'interagir avec l'ADN car ces expansions se structurent en hélice de type ADN. La taille des expansions augmente au cours de la vie des personnes touchées et cette instabilité somatique contribue à la nature progressive de la maladie. Il est donc important d'envisager la possibilité que les composés visant l'ARN pourraient également avoir des effets involontaires (nocifs ou bénéfiques) sur la propension de l'ADN à augmenter son nombre de répétitions.

VII.3. Correction des défauts d'épissage

La modification de la balance entre les facteurs MBNL1 et CUG-BP1, suite à la diminution de l'activité de MBNL1 et l'augmentation de l'activité de CUG-BP1, entraîne des dérégulations de l'épissage alternatif de plusieurs pré-ARNm. Une correction de la dérégulation de l'épissage alternatif des différentes cibles permettrait de réverser le phénotype de la pathologie. De même, un retour à un taux normal de ces deux protéines permettrait de rétablir le taux normal d'épissage de leurs cibles.

VII.3.a. Correction de l'épissage par utilisation de morpholino-oligonucléotides

Les morpholino-oligonucléotides peuvent également interférer avec les étapes de maturation de pré-ARNm, soit en empêchant l'interaction du complexe snRNP avec leurs cibles au bord des introns sur le brin du pré-ARNm, soit en bloquant la boîte de branchement, soit en interférant avec des protéines régulatrices d'épissage.

Ainsi, il a été montré qu'un morpholino-oligonucléotide dirigé contre le site 3' d'épissage de l'exon 7a du canal chlore permettait de bloquer ce site, favorisant l'exclusion de cet exon alternatif, la restauration du cadre de lecture du pré-ARNm et l'expression du canal à chlore. Une souris transgénique DM1 ainsi infectée ne présente plus de décharges myotoniques. Cette correction de défaut ponctuel de l'épissage permet de guérir l'un des symptômes musculaire majeur (Wheeler et Thornton, 2007). Cependant cette technique ne peut être utilisée pour corriger l'ensemble des symptômes car elle nécessiterait la création d'un trop grand nombre de morpholino-oligonucléotides.

VII.3.b. Augmenter l'activité de MBNL1

Une stratégie permettant d'inverser les effets de la séquestration de MBNL1 au sein de *foci* sur les dérégulations de l'épissage alternatif, consiste à augmenter le niveau de MBNL1 au sein des tissus affectés. L'équipe de M.S. Swanson a réalisé une injection intramusculaire dans un modèle de souris transgénique DM1 sur-exprimant 250 répétitions CTG d'un virus de type AAV permettant une surexpression de MBNL1. Une sur-expression de deux fois de MBNL1 entraîne une saturation des sites de fixation de MBNL1 sur les répétitions CUG et une restauration de la localisation de protéine MBNL1 libre dans le nucléoplasme. Il résulte de ces travaux une normalisation de l'activité de MBNL1 à un niveau suffisant pour réverser la myotonie et rétablir une régulation normale de l'épissage alternatif jusqu'à plusieurs mois après le premier traitement (Kanadia *et al.*, 2006).

Cette stratégie thérapeutique basée sur la surexpression de MBNL1 par l'intermédiaire d'un virus AAV est prometteuse pour le traitement des patients DM1, cependant une application clinique à long-terme est peu envisageable. En effet, pour être une thérapie viable pour les patients DM, la surexpression de MBNL1 par AAV requiert une délivrance systématique (Wheeler, 2008).

VII.3.c. Diminuer l'activité de CUG-BP1

Les approches qui se limitent à la restauration des fonctions de MBNL1 sont peu susceptibles de traiter toutes les conséquences pathologiques des ARN contenant des répétitions CUG ou CCUG. En effet, l'un des effets de la présence de l'ARN mutant est une augmentation de l'activité de CUG-BP1 suite à son hyperphosphorylation, phénomène qui semble être indépendant de l'appauvrissement de MBNL1. Récemment, l'administration d'un inhibiteur de l'activité de PKC au sein d'un modèle de souris DM1 a permis d'atténuer les défauts de conduction cardiaque et de contractions anormales observés chez ces modèles de souris. L'inhibiteur permet également de réduire les dérégulations de l'épissage alternatif régulées par CUG-BP1. Cependant, aucun effet sur la mortalité des souris n'a été détecté (Wang *et al.*, 2009). Ces résultats suggèrent que les effets divergents des ARN toxiques pourraient être réglés par des approches thérapeutiques combinées.

Objectifs

Comme nous venons de le voir dans la partie Introduction, à ce jour l'hypothèse physiopathologique majeure expliquant l'atteinte multi-systémique de la DM1 est celle de l'effet *trans*-dominant des répétitions CUG sur l'épissage alternatif. Il est largement admis que la majorité de ces défauts d'épissage sont dus aux altérations des facteurs régulateurs des familles CELF et Muscleblind. En effet, la protéine MBNL1 piégée au sein de *foci* nucléaires suite à son interaction avec les répétitions CUG, voit sa concentration sous forme libre diminuer et donc sa fonction de régulateur d'épissage réduite. Lorsque j'ai débuté mon travail de thèse en 2006 avec Nathalie Marmier-Gourrier et Isabelle Behm-Ansmant, peu de choses étaient connues sur les propriétés de fixation à l'ARN de MBNL1. Les différentes données disponibles à l'époque étaient la capacité de fixation de MBNL1 aux structures ARN en double-brin formées par des répétitions de type CHHG/CHG, (où H=A, U, ou C), formant une tige-boucle (Kino *et al.*, 2004). De plus, une séquence consensus YGCU(U/G)Y, (où Y = U ou C) de fixation de MBNL1 sur ses pré-ARNm cibles avait été proposée sur la base de l'alignement de séquences de 4 sites de liaison de MBNL1 (Ho *et al.*, 2004). Notre travail s'est donc naturellement orienté vers la compréhension des mécanismes à l'origine des interactions entre MBNL1, les ARNm et les séquences répétées. Parallèlement, mon travail de thèse s'est dirigé vers la compréhension des mécanismes de régulation de l'épissage par MBNL sur un ARN cible.

Après avoir produit différentes protéines MBNL1 recombinantes entières ou tronquées, mon premier objectif a été d'étudier les propriétés de l'ARN nécessaires à la reconnaissance par MBNL1. Est ce que la protéine reconnaît de la même manière les répétitions CUG et ses pré-ARNm cibles, ou présente t'elle au contraire différents modes de fixation? Pour répondre à cette question, nous avons tout d'abord étudié la fixation de MBNL1 sur des structures tige-boucles formés par des répétitions CUG, cela par la technique de protection aux RNases. En parallèle, j'ai cherché à identifier expérimentalement de manière large les différents motifs ARN capables d'interagir avec MBNL1 par la technique SELEX. Ces travaux ont fait l'objet de la rédaction d'un article, dont je suis premier auteur, qui est actuellement en cours de rédaction (Résultats, Chapitre I).

Après avoir identifié les déterminants ARN nécessaires à la reconnaissance par MBNL1, nous avons cherché à identifier les domaines de MBNL1 importants pour l'interaction avec les ARN. La réponse à cette question a fait partie d'une étude plus globale sur le rôle des différents exons alternatifs de MBNL1, réalisée en collaboration avec l'équipe

de N. Sergeant (Inserm, Lille), et qui a fait l'objet d'un article, paru en 2011 dans *Journal of Biological Chemistry*, dont je suis 5^{ème} auteur (Résultats, Chapitre I).

Parmi les pré-ARNm dont l'épissage alternatif est dérégulé dans la DM1, celui de la troponine T cardiaque (hcTNT) est particulièrement important. En effet, la dérégulation de l'inclusion de l'exon 5 de ce transcrite conduit à une mauvaise contraction musculaire et pourrait être impliquée dans les anomalies cardiaques qui touchent 80 % des patients atteints de DM1 (Philips *et al.*, 1998). Lorsque j'ai débuté ma thèse, la majorité des travaux réalisés sur la régulation de l'épissage du pré-ARNm hcTNT avait été menée sur le poulet et peu d'informations existaient sur la régulation chez l'homme. En collaboration avec l'équipe de N. Sergeant, nous avons tout d'abord déterminé les séquences minimales requises en aval et en amont de l'exon 5 pour récapituler les régulations de son épissage, ainsi que l'effet de la présence d'ARN contenant des répétitions de triplets CUG. En parallèle, j'ai identifié 6 nouveaux sites MBNL1 de fixation à l'ARN par des expériences de retard sur gel et par emploi de la technique d'empreinte enzymatiques. A partir du minigène IR-hcTNT qui permet de reproduire les régulations de l'épissage alternatif, nous avons pu tester la fonctionnalité de ces sites nouvellement identifiés. Un article sur ces travaux est en cours de rédaction et sera prochainement soumis (Résultats, Chapitre II).

L'ensemble des symptômes et des dérégulations de l'épissage alternatif observés chez les patients DM1 ne peut être expliqué uniquement par la perte de fonction de MBNL1. Nous avons donc cherché à mieux comprendre comment les répétitions de triplets CUG interfèrent avec l'épissage alternatif des pré-ARNm et aussi pourquoi les ARN contenant les répétitions de triplets CUG sont localisés dans les *foci* nucléaires. Pour cela, il paraissait important d'identifier les protéines nucléaires capables de s'associer à ces répétitions dans ces *foci*. En effet, des données de la littérature suggéraient la présence, dans les *foci*, de protéines autres que la protéine MBNL1. Certaines de ces protéines pourraient jouer un rôle dans le phénomène de rétention (Kim *et al.*, 2005). Nous avons donc entrepris d'identifier les protéines d'un extrait nucléaire de cellules HeLa capables de se lier à une longue amplification de triplets CUG (51 copies). Ce nombre de répétitions correspond à celui déclenchant la pathologie (pour revue, Usdin et Grabczyk, 2000). Pour cela, j'ai employé une méthode de chromatographie d'affinité basée sur la fixation de la séquence MS2 en 3' de l'ARN étudié afin de le fixer sur billes d'amylose via la protéine de fusion MS2/MBP. Les protéines des complexes purifiés ont été analysées par spectrométrie de masse en collaboration avec l'équipe d'A Van Dorsselaer. Par cette approche, nous avons identifié

plusieurs partenaires potentiels des répétitions CUG, dont la nucléophosmine (NPM) (Résultats, Chapitre III). Cette protéine a retenu notre intérêt car c'est une protéine multifonctionnelle impliquée dans de multiples processus cellulaires, tels que la maintenance de la stabilité du génome, la régulation de la transcription de l'ADN et de l'épissage, processus dérégulés chez les patients DM1. Nous avons émis l'hypothèse que la protéine NPM, en se fixant sur les répétitions CUG, peut déréguler l'épissage alternatif de pré-ARNm. Nous avons dans un premier temps vérifié la présence d'une interaction directe de la protéine NPM avec les répétitions CUG, puis recherché des arguments en faveur d'une signification biologique de l'association observée entre cette protéine et les répétitions CUG. Pour cela, nous avons entre autre entrepris une étude de la localisation de la NPM au sein de myoblastes de patients DM1 par immunofluorescence. Cette étude est présentée dans le Chapitre III de la partie Résultats.

*Matériel et
Méthodes*

Matériel et Méthodes

I Matériel

I. Souches d'*E. coli* utilisées

Souche DH5 α : *deoR*, *supE44*, $\Delta(lacIZYA-argF)U169$, $\Phi80lacZ\Delta M15$, *hsdR17* ($r_K^- m_K^+$), *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*, F^- (Hanahan, 1983) a été utilisée pour la production et la purification des plasmides.

Souche BL21-CodonPlus-RIL : B F^- , *ompT*, *hsdS* ($r_B^- m_B^-$), *dcm+*, Tet^r , *gal*, *endA*, Hte [*argU ileY leuW Cam^r*] (Studier et Moffatt, 1986) a été utilisée pour la production des protéines recombinantes.

Souche XL1-blue : *recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17*, *supE44*, *relA1*, *lacF'*, *proAB*, *lacI^qZAM15*, Tn10 (tet^r) (Bullock *et al.*, 1987) a été utilisée pour les mutagenèses dirigées.

II. Lignées de cellules eucaryotes

Cellules HEK 293T : N°ATCC® : CRL-1573™ (Graham *et al.*, 1977) (Human Embryonic Kidney fibroblast-like) sont des cellules adhérentes issues de tissus rénaux embryonnaires humains. Les lignées utilisées dans ce travail sont des HEK-293T qui expriment l'antigène T de SV40 et permettent donc d'atteindre des niveaux élevés de surexpression de protéines.

Myoblastes immortalisés issus de patients sain (myoblastes WT) : lignée cellulaire de nouveau-né sain immortalisée par surexpression de hTert et cdk4 (Bigot *et al.*, 2009). Lignée fournie par le Dr. D. Furling (Institut de myologie, UMR787, Paris).

Myoblastes immortalisés issus de patients DM (myoblastes DM) : lignée cellulaire de patients atteints de DM1 possédant un nombre de répétitions CUG supérieur à 2000. Les cellules ont été immortalisées par surexpression de hTert et cdk4 (Bigot *et al.*, 2009). Lignée fournie par le Dr. D. Furling.

III. Plasmides utilisés

III.1. Plasmides recombinants utilisés pour la production *in vitro* d'ARN

Le plasmide pCR® 2.1

Ce plasmide est fourni avec le kit « TA cloning kit » (Invitrogen). Il est utilisé pour le clonage de l'ensemble des séquences obtenues par l'expérience de SELEX. Les ARN correspondants à ces séquences sont obtenus à l'aide du couple d'oligonucléotides 2174 et 2176, l'oligonucléotide 2174 contenant le promoteur de l'ARN polymérase T7.

Le plasmide pUC18

Le plasmide pUC18 a servi à réaliser les constructions permettant de purifier les complexes RNP se formant entre des répétitions CUG ou CAG et un extrait nucléaire, cela par la technique de chromatographie d'affinité MBP-MS2. La stratégie complète de production de ce vecteur est détaillée **dans le chapitre III des résultats**. Les vecteurs produits sont appelés pUC18-T7-MS2, pUC18-T7-CTG₅₁, pUC18-T7-CTG₅₁-MS2 et pUC18-T7-CAG₅₉-MS2.

III.2. Vecteurs d'expression eucaryotes

pSVIRB-hcTNT

Ce plasmide nous a été fourni par l'équipe de N. Sergeant. Il provient du Dr A. Andreadis (Université du Massachusetts, Waltham, Massachusetts, USA). Ce minigène a été décrit dans Andreadis *et al.*, (1993 et 1995). Il contient trois exons du gène de l'insuline de rat et une portion de leurs séquences introniques flanquantes. Différentes tailles de séquences de hcTNT ont été insérées au sein d'un site *EcoRI*, situé dans la région intronique entre les exons 2 et 3 de l'insuline. Les séquences hcTNT insérées contiennent l'exon 5 du pré-ARNm hcTNT, ceci avec des régions introniques flanquant l'exon 5 variant de 50 à 250 nt. Des mutants ponctuels ont également été réalisés à partir de ces constructions. L'ensemble des minigènes utilisés dans ce travail est récapitulé dans le Tableau VIII.

Nom de la construction	Description de la région hcTNT insérée
IR-hcTNT 175-E5-420	175 nt de l'intron 4, l'exon 5 et 420 nt de l'intron 5
IR-hcTNT 175-E5-250	175 nt de l'intron 4, l'exon 5 et 250 nt de l'intron 5
IR-hcTNT 175-E5-150	175 nt de l'intron 4, l'exon 5 et 150 nt de l'intron 5
IR-hcTNT 150-E5-150	150 nt de l'intron 4, l'exon 5 et 150 nt de l'intron 5
IR-hcTNT 100-E5-100	100 nt de l'intron 4, l'exon 5 et 100 nt de l'intron 5
IR-hcTNT 75-E5-75	75 nt de l'intron 4, l'exon 5 et 75 nt de l'intron 5
IR-hcTNT 50-E5-50	50 nt de l'intron 4, l'exon 5 et 50 nt de l'intron 5
IR-hcTNT 100-E5-150	100 nt de l'intron 4, l'exon 5 et 150 nt de l'intron 5
IR-hcTNT 50-E5-150	50 nt de l'intron 4, l'exon 5 et 150 nt de l'intron 5
IR-hcTNT 150-E5-100	150 nt de l'intron 4, l'exon 5 et 100 nt de l'intron 5
IR-hcTNT 150-E5-50	150 nt de l'intron 4, l'exon 5 et 50 nt de l'intron 5
IR-hcTNT 150-E5-150 mut1-2-3	Issu de 150-E5-150 + mutation sites 1,2 et 3 de MBNL1 G en A
IR-hcTNT 150-E5-150 mut4-5-6 ΔG	Issu de 150-E5-150 + mutation sites 4,5 et 6 + ΔG
IR-hcTNT 150-E5-150 mut1'-2'	Issu de 150-E5-150 + mutation sites 1' et 2' de MBNL1 G en A
IR-hcTNT 150-E5-150 mutg1-g2-g3ΔG	Issu de 150-E5-150 + mutation de l'ensemble des sites + ΔG

Tableau VIII : Constructions pSVIRB-hcTNT utilisées au cours de ma thèse.

Ces constructions ont également été utilisées pour la production d'ARN hcTNT *in vitro*. Les ARN correspondants à ces séquences sont obtenus à l'aide de différents couples d'oligonucléotides dont l'un contient le promoteur de l'ARN polymérase T7.

pcDNA3.1-(CTG)₉₆₀

La construction contenant les répétitions de CTG provient du laboratoire du Dr. T.A. Cooper (Baylor College of Medicine, Houston, Texas, USA). Ce plasmide est utilisé afin de produire dans des cellules eucaryotes de longues répétitions CUG et de mimer la dystrophie myotonique. Ce plasmide contient les 5 derniers exons du gène codant DMPK, bordés par leurs introns et la totalité de la région 3'UTR. Les 960 répétitions de CTG sont interrompues toutes les 20 répétitions par une séquence CTCGA.

pcDNA3.1-GFP-MBNL1

Le vecteur exprimant le facteur d'épissage MBNL1 provient du Dr. N. Charlet (IGBMC, Strasbourg).

III.3. Vecteurs utilisés pour la production des protéines recombinantes

Le plasmide pGEX-6P1 a été utilisé dans le but de produire les protéines recombinantes MBNL1, CUG-BP1 et NPM, et de leurs formes tronquées, et permet leur purification grâce à une étiquette GST fusionnée à l'extrémité N-terminale. Les fragments correspondant aux séquences des différentes protéines ont été amplifiés par PCR à partir d'ADN génomique avec des oligonucléotides spécifiques portant les sites de restriction *Bam*HI/*Xho*I. Les fragments amplifiés sont clonés dans le vecteur pCR2.1. La validité des séquences clonées est vérifiée par séquençage. Puis, les sites de restrictions permettent de sous-cloner ces fragments dans le vecteur pGEX-6P1. Les oligonucléotides qui ont servi à la production de toutes ces constructions sont présentés des le Tableau IX.

Construction	Oligonucléotide 1 5'-3' Oligonucléotide 2 5'-3'
pGEX-6P1-MBNL1	4427- ATAAAGGATCCGCTGTTAGTGTACACCAA 4428- ATAAACTCGAGCTACATCTGGGTAACATAC
pGEX-6P1-MBNL1 _{ΔCT3}	4427- ATAAAGGATCCGCTGTTAGTGTACACCAA 4845- ATAAACTCGAGTCACATGGCAGCTGCGGTGGCTG
pGEX-6P1-MBNL1 _{ΔCT}	4427- ATAAAGGATCCGCTGTTAGTGTACACCAA 4889- ATAAACTCGAGTCACTTGATCTTGGCTTGCAAATG
pGEX-6P1-CUG-BP1	4402- ATAAAGGATCCAACGGCACCCCTGGACCAC 4403- ATAAACTCGAGTCAGTAGGGCTTGCTGTC
pGEX-6P1-CUG-BP1 _{ΔCT}	4402- ATAAAGGATCCAACGGCACCCCTGGACCAC 4802- ATAAACTCGAGTCACAAATTCTGTAATG
pGEX-6P1-NPM	5499- ATAAAGGATCCATGGAAGATTCGATGGACATG 5500- ATAAACTCGAGTTAAAGAGACTTCTCCACTG

Le plasmide pMBP-MS2 fourni par le Dr Lurhmann a permis de produire et purifier la protéine de fusion MBP/MS2. Il dérive d'un plasmide pMAL (New England Biolabs), qui permet de produire chez *E. coli* des protéines contenant une étiquette MBP en N-ter.

IV. Les anticorps

Tous les anticorps utilisés au cours de ce travail sont répertoriés dans le Tableau X. Les anticorps secondaires proviennent des sociétés Invitrogen et Abcam.

Dénomination de l'anticorps	Clonalité	Utilisation	Origine
anti-MBNL1	monoclonal de souris	Western-blot (7000°), IF (100°)	Sigma
anti-NPM	monoclonal de lapin	Western-blot (1000°), IF (200°)	Cell signaling
anti-CUG-BP1	monoclonal de souris	Western-blot (500°)	GeneTex
anti-PTB	polyclonal de chèvre	Western-blot (1000°)	Abcam
anti-hnRNP H/F	monoclonal de souris	Western-blot (1000°)	Dreyfuss G.
anti-tubuline (T7816)	monoclonal de souris	Western-blot (20000°)	Sigma

V. Milieux de culture, tampons et solutions d'usage courant

Milieu LB (Luria Bertani) : peptone pancréatique animale 5 g/l ; peptone tryptique de caséine 5 g/l ; extrait de levure 5 g/l ; NaCl 10 g/l.

Le pH est ajusté à 7,5 avec NaOH 10 N. Le milieu de culture est ensuite autoclavé 20 min à 121°C sous une pression de 1 bar.

Milieu solide LB : l'agar est ajouté à raison de 17 g par litre de milieu de culture LB avant stérilisation à l'autoclave.

Milieu LBA : l'ampicilline (100 mg/ml) est ajoutée après autoclave dans le milieu de culture LB à une concentration finale de 100 µg/ml.

Milieu LBAC : l'ampicilline (100 mg/ml) et le chloramphénicol (25 mg/ml) sont ajoutés après autoclave dans le milieu de culture LB à une concentration finale de 100 µg/ml et 25 µg/ml, respectivement.

TBE: Tris borate 89 mM, pH 8,3 ; EDTA 2 mM.

Solution de polyacrylamide dénaturante : Acrylamide/Bisacrylamide (23,75/1,25) 25 % (p/v) ; TBE; urée 8 M.

Solution de polyacrylamide native : Acrylamide/Bisacrylamide (38/2) 40 % (p/v) ; TBE 1X.

Solution de bleu de dépôt agarose (6X) : EDTA 125 mM ; SDS 2 % (v/v) ; glycérol 25 % (v/v) ; bleu de bromophénol 0,12 % (p/v).

Solution de bleu de formamide : EDTA 20 mM ; bleu de bromophénol 0,05 % (p/v) ; bleu de xylène cyanol 0,05 % (p/v) dans de la formamide désionisée.

Solution de bleu de dépôt (non dénaturant) pour expériences de retard sur gel (CSB) : HEPES-KOH 20 mM, pH 7,9 ; glycérol 40 % (v/v) ; bleu de bromophénol 0,05 % (p/v) ; bleu de xylène cyanol 0,03 % (p/v).

Tampon d'éluion : Tris-HCl 10 mM, pH 7,5 ; NaCl 300 mM ; EDTA 1 mM ; SDS 1%.

Tampon D: HEPES-KOH 20 mM, pH 7,9 ; KCl 150 mM ; MgCl₂ 1,5 mM ; EDTA 0,2 mM ; glycérol 10 % (v/v).

Tampon D 10X : HEPES-KOH 200 mM, pH 7,9 ; KCl 1,5 M ; MgCl₂ 15 mM ; EDTA 2 mM.

HBS: HEPES-KOH 50 mM, pH 7,05 ; KCl 10 mM ; NaCl 280 mM ; Na₂HPO₄ 1,5 mM.

II. Méthodes

En 2010, C. Branlant a été sollicitée pour rédiger un chapitre d'un ouvrage portant sur l'ensemble des techniques d'étude des ARN. Le chapitre à rédiger concernait les méthodes d'étude des interactions ARN-protéines. Il se trouve que j'avais utilisé la majorité de ces méthodes durant mes travaux de thèse, C. Branlant m'a donc demandé de participer à la rédaction de ce chapitre d'ouvrage, que j'ai inséré à la fin de la partie Matériel et Méthodes de mon manuscrit.

Les techniques suivantes y sont décrites avec des exemples d'applications réalisés au laboratoire :

- cartographie des ARN en solution ou dans des complexes RNP,
- pontages aux UV et immunosélection,
- supershift,

Un des exemples inclus correspond à des travaux que j'ai réalisés lors de l'étude de la structure de répétitions CUG, en particulier des expériences de cartographie de ces répétitions et d'empreintes réalisées sur des complexes formés en présence d'une protéine purifiée recombinante.

L'ensemble des autres méthodes utilisées lors de mes travaux de thèse est présenté en détail ci-dessous.

CHAPITRE DE LIVRE

Analysis of Site-Specific RNA-Protein Interactions

N. Marmier-Gourrier, **A. Vautrin**, C. Branlant and I. Behm-Ansmant

I. Autres techniques d'analyse des interactions ARN/protéines

I.1. Analyses de complexes [ARN/protéine] par gel-retard

La formation d'un complexe entre un ARN et une protéine peut être visualisée par un retard de migration de l'ARN complexé par rapport à l'ARN libre sur un gel natif. Pour cela, un ARN est produit par transcription *in vitro*, déphosphorylé et marqué à son extrémité 5'.

◆ Préparation de la matrice ADN à transcrire

Les matrices ADN utilisées pour la transcription ont été obtenues après une amplification par PCR avec les couples d'oligonucléotides correspondants. L'oligonucléotide 5' comporte à son extrémité la séquence du promoteur du phage T7 qui permet la transcription des matrices ADN.

Le mélange réactionnel est le suivant :

Plasmide	20 ng
Oligonucléotide 5' (20 μ M)	1 μ l
Oligonucléotide 3' (20 μ M)	1 μ l
Mélange de dNTP (1, 25 mM de chaque)	8 μ l
Tampon Pfu (DTT 100 mM ; 0,1 mM EDTA ; KCl 100 mM ; Nonidet P40 0,1% ; Tween 20 0,1% ; glycérol 50% ; Tris-HCl 200 mM, pH 8,8)	5 μ l
Pfu polymérase (5 U/ μ l, MBI Fermentas)	0,5 μ l
H ₂ O	qsp 50 μ l

Après un cycle de dénaturation de 5 minutes à 95°C, l'amplification se déroule en 30 cycles qui sont chacun composés de paliers de 30 secondes à 95°C, 30 secondes à 55°C et 1 minute à 74°C, pour ensuite terminer par une incubation de 10 minutes à 74°C. Le temps d'élongation dépend de la taille du fragment à amplifier, sachant que la processivité de la Pfu polymérase est de 2 min/kb. Les produits PCR sont séparés par électrophorèse en gel d'agarose à 1,8 % (p/v) et purifiés en utilisant le kit "Nucleospin[®] extract" (Macherey-Nagel).

◆ Transcription *in vitro*

Matrice PCR	1,5 pmol
Tampon de transcription 5X : (MgCl ₂ 120 mM ; spermidine 10 mM ; DTT 200 mM ; Hepes-KOH 400 mM, pH 7,5)	6 μ l
rNTP (12,5 mM de chaque)	10 μ l

ARN polymérase T7 (produite au laboratoire)	3 µl
H ₂ O	qsp 30 µl

Le mélange est incubé 4h à 37°C. La matrice ADN est ensuite dégradée pendant 30 minutes à 37°C en présence de 5 U de DNaseI, exempt de RNase (Fermentas). Un volume de 15 µl de bleu de formamide est ajouté et les ARN sont ensuite déposés et purifiés par électrophorèse en gel dénaturant de polyacrylamide 8%-urée 8 M. Les ARN sont alors élués une nuit à 4°C dans du tampon d'éluion à forte concentration saline, extraits au phénol/chloroforme, puis précipités en présence d'éthanol 96% et de 10 µg de glycogène. Les ARN sont ensuite repris dans de l'eau stérile.

♦ Marquage des ARN à leur extrémité 5'

Déphosphorylation de l'extrémité 5' des ARN

ARN à déphosphoryler repris dans de l'eau stérile	30 pmol
Tampon CIP 10X (MgCl ₂ 0.1 M ; Tris/HCl 0.1 M, pH 7,5)	1 µl
Calf Intestine Alkaline Phosphatase (Fermentas, 1 U/µl)	1 µl
H ₂ O	qsp 10 µl

Les ARN sont incubés 1 heure à 37°C, extraits au phénol/chloroforme, puis précipités avec de l'éthanol 96% en présence d'acétate de sodium 0,3 M et repris dans 7 µl d'eau stérile.

Marquage en 5' des ARN

ARN déphosphorylés repris dans de l'eau stérile	7 µl
ATP [γ ³² P] (Amersham, 10 mCi/ml, 3000 Ci/mmol)	1 µl
Tampon PNK 10X	
(MgCl ₂ 0.1 M ; DTT 50 mM ; spermidine 1 mM ; EDTA 1mM ; Tris/HCl 0,5 M, pH 7,6)	1 µl
T4 polynucléotide kinase (MBI Fermentas, 10 U/µl)	1 µl

Le mélange est incubé 45 min à 37°C puis les transcrits sont purifiés sur gel de polyacrylamide dénaturant 8 %, élués, et précipités comme décrit précédemment.

♦ Formation des complexes [ARN/protéine]

Après dénaturation des ARN 5 min à 65°C, puis renaturation par retour lent à 37 °C, les complexes [ARN/protéine] sont formés dans du tampon D 1 en incubant 10 fmol d'ARN marqués en présence de concentrations croissantes en protéines (de 0 à 1 µM), et de MgCl₂ :

ARN radioactif	5 fmol
Tampon D	1,8 µl
MgCl ₂ (62,5 mM)	0,5 µl
H ₂ O	qsp 7 µl

Un volume de 4 µl de protéine purifiée diluée ou non dans le tampon D est ensuite ajouté et le mélange est incubé 30 minutes à température ambiante.

◆ **Fractionnement en gel natif**

Du bleu CSB (4 µl) est ajouté au mélange réactionnel ayant servi à former le complexe. L'échantillon est ensuite déposé sur un gel de polyacrylamide natif 8% (acrylamide/bisacrylamide 38/2), en présence de glycérol 0,5 % et de TBE 0,5X. Le pourcentage de ce gel peut varier de 5 à 8 % en acrylamide/bisacrylamide suivant la taille des ARN utilisés lors de l'expérience. L'électrophorèse se fait à 100 volts et à température ambiante. Le gel est ensuite séché et exposé en autoradiographie durant une nuit. Ainsi, les complexes [ARN/protéine] formés seront retardés dans leur migration par rapport aux ARN non complexés.

◆ **Détermination de la constante de dissociation (Kd) des complexes**

La détermination de la constante de dissociation d'un complexe [ARN/protéine] permet d'évaluer l'affinité qu'a la protéine d'intérêt pour l'ARN. Cette constante correspond à la concentration en protéine pour laquelle 50% de l'ARN est lié à la protéine. Les quantités d'ARN libre et lié sont quantifiées par mesure de la radioactivité dans les bandes correspondantes, en utilisant l'appareil PhosphorImager à l'aide du logiciel « ImageQuant Software Version 3.3 » (Molecular Dynamics). La sigmoïde obtenue par représentation graphique de la concentration en protéine (en abscisse) par rapport au pourcentage d'ARN lié est tracée à l'aide du logiciel « SigmaPlot 2000 ». Le point d'inflexion de la sigmoïde indique la valeur de la constante de dissociation (concentration en protéine pour laquelle 50% des ARN sont liés à la protéine d'intérêt) est donnée par le logiciel.

I.2. Prédiction de structure secondaire par ordinateur

Les prédictions de structure secondaire ont été réalisées par le programme Mfold version 3.2 qui est accessible à l'adresse suivante : <http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi> (Mathews et al., 1999; Zuker, 2003). Pour une séquence d'ARN donnée, ce programme propose les structures secondaires les plus stables d'après le calcul de la stabilité

thermodynamique des appariements des bases de l'ARN. Ce calcul est effectué pour un ARN dans une solution de NaCl 1M à 37°C. Ce programme permet également d'imposer des contraintes de structuration en simple et double brin ce qui permet d'obtenir des modèles de structure secondaire qui sont en accord avec les données expérimentales obtenues lors de l'hydrolyse ménagée de l'ARN par les sondes enzymatiques et chimiques.

II. SELEX (systematic evolution of ligands by exponential enrichment)

Cette approche a été décrite pour la première fois par Tuerk et Gold (1990). Elle consiste à dégénérer la séquence d'un ARN ciblé par une protéine d'intérêt et à isoler après plusieurs cycles de sélection les ARN ayant de l'affinité pour cette protéine. Cette technique repose sur la réalisation de plusieurs cycles de sélection comprenant chacun 4 étapes (Fig. 41) : 1) la synthèse d'ADN double-brin dont la séquence est partiellement dégénérée, 2) la transcription *in vitro* d'ARN dont la séquence est partiellement dégénérée à partir des fragments PCR, 3) la sélection sur billes de Sépharose couplées à du glutathion des transcrits liés à la protéine d'intérêt fusionnée à l'étiquette GST et 4) une amplification par RT-PCR des transcrits sélectionnés.

Un mélange d'oligonucléotides dégénérés est hybridé à un oligonucléotide présentant la séquence du promoteur de l'ARN polymérase du phage T7 pour synthétiser, par PCR, le mélange d'ADN double-brin dégénérés. Les fragments d'ADN double-brin obtenus sont alors transcrits par l'ARN polymérase du phage T7 de manière à obtenir des ARN dégénérés au niveau de leur zone d'interaction avec la protéine. Ces ARN sont ensuite incubés en présence de la protéine d'intérêt fusionnée à une étiquette GST. Ainsi, seuls les ARN liés à cette protéine de fusion seront retenus sur une matrice de Sépharose couplée à du glutathion. Après action de la protéinase K (20 mg/ml), les ARN sont extraits par un traitement au phénol-chloroforme. Des ADN complémentaires aux ARN sélectionnés sont alors synthétisés par transcription inverse puis amplifiés en ADN double-brin par PCR en utilisant un oligonucléotide portant la séquence reconnue par l'ARN polymérase T7. Les produits PCR obtenus servent ensuite de matrices pour transcrire de nouvelles molécules d'ARN qui seront alors soumis à un autre cycle de sélection en présence de la protéine étiquetée.

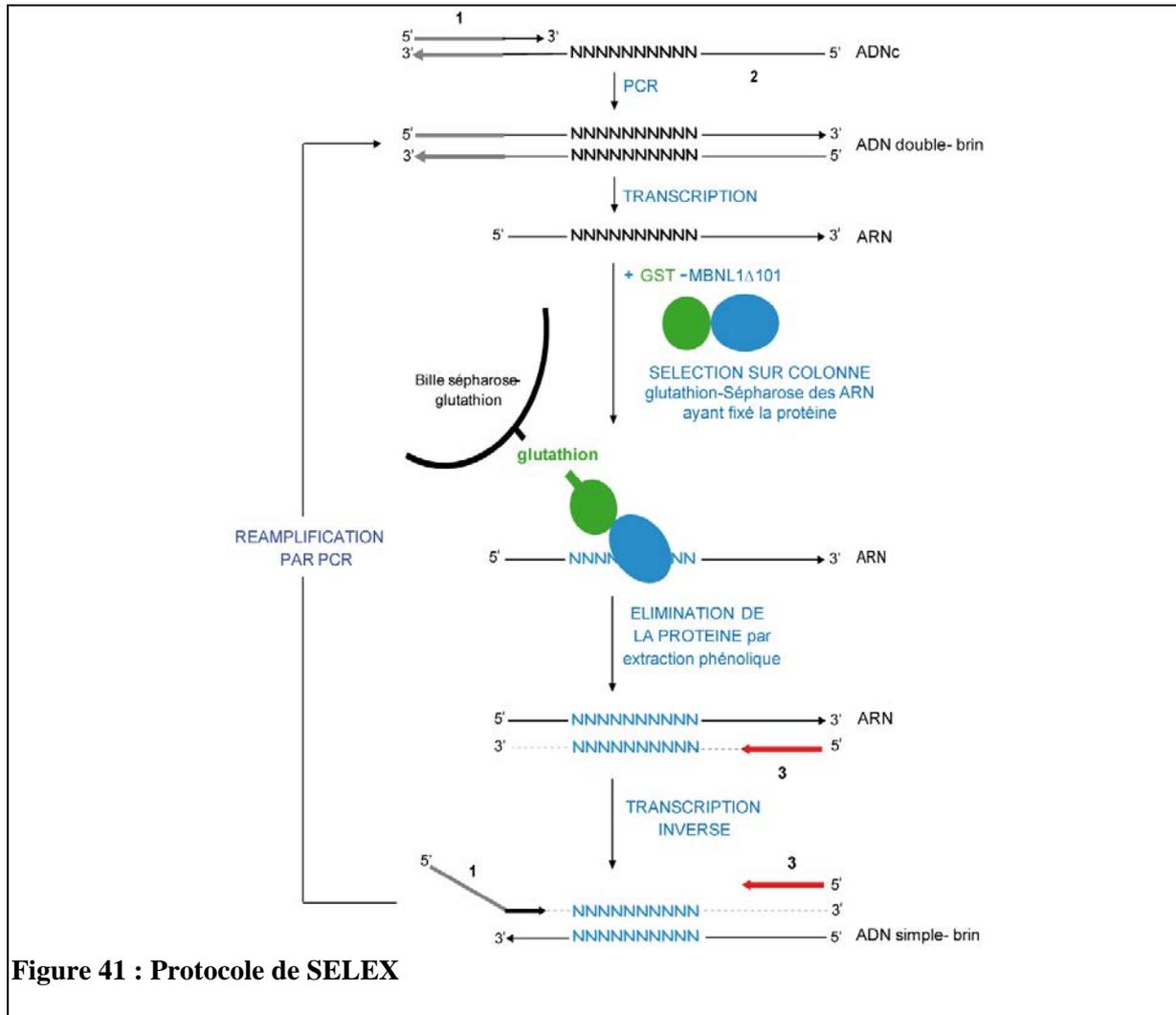


Figure 41 : Protocole de SELEX

◆ **PCR d'amplification de la matrice d'ADN dégénérée**

Oligonucléotide dégénéré	100 μ M
Oligonucléotide 5' portant le promoteur reconnu par la T7 RNA polymérase	100 μ M
Tampon GoTaq 5X (KCl 50 mM ; Triton X-100 0.1% ; Tris/HCl 10 mM, pH 8,8)	10 μ l
MgCl ₂ 25 mM	8 μ l
dNTP (1,25 mM chaque NTP)	8 μ l
GoTaq (Finnzymes, 2U/ μ l)	0,25 μ l
H ₂ O	qsp 50 μ l

Programme PCR (appareil BIORAD) :

5 min à 94°C

30 cycles de : 30 sec à 94°C, 30 sec à 55°C et 30 sec à 74°C

10 min à 74°C

Les produits PCR sont ensuite purifiés sur gel de polyacrylamide natif 6%, élués, puis précipités avec de l'éthanol 96% en présence d'acétate de sodium 0,3 M. Chaque culot obtenu à partir d'une réaction PCR est repris dans 8 µl d'eau stérile.

◆ **Transcription des ARN dégénérés**

Matrice PCR	27 pmol
Tampon de transcription (Tris/HCl 40 mM, pH 8,1 ; MgCl ₂ 20 mM ; DTT 5 mM ; Spermidine 1 mM ; Triton X-100 0,01% ; PEG 8000 80 mg/ml)	10 µl
rNTP (12,5 mM de chaque)	10 µl
RNasin (RNA guard Pharmacia Biotech, 35,7 U/ml)	1 µl
ARN polymérase T7 (produite au laboratoire)	1 µl
H ₂ O	qsp 30 µl

Le mélange est incubé 3h à 37°C. La matrice ADN est ensuite dégradée pendant 30 minutes à 37°C en présence de 7,5 U de DNaseI, exempte de RNase (Fermentas). Un volume de 15 µl de bleu de formamide est ajouté et les ARN sont ensuite déposés et purifiés par électrophorèse en gel dénaturant de polyacrylamide 8%-urée 8 M. Les ARN sont alors élués une nuit à 4°C dans du tampon d'élution à forte concentration saline, extraits au phénol/chloroforme, puis précipités en présence d'éthanol 96% et de 10 µg de glycogène. Les ARN sont ensuite repris dans de l'eau stérile.

◆ **Sélection des ARN ayant de l'affinité pour la protéine d'intérêt**

Afin d'éliminer les ARN présentant de l'affinité pour les billes ou la protéine GST, 0,2 nmol du mélange d'ARN est ensuite incubé en présence de 30 µl de billes de Sépharose couplées à du glutathion durant 2h à 4°C de manière à éliminer les ARN ayant de l'affinité pour la matrice, puis précipité et repris dans de l'eau. Ces ARN sont ensuite renaturés 5 min à 65°C, puis 15 min à température ambiante puis incubés en présence de 0,02 nmoles de protéine GST dans du tampon D durant 30 min à 4°C. Le mélange réactionnel est ensuite incubé pendant 2h en présence de 30 µl de billes de Sépharose couplées à du glutathion afin d'éliminer les ARN présentant de l'affinité pour la protéine GST. Les ARN non retenus par les billes sont ensuite précipités et repris dans de l'eau stérile. Pour chaque cycle de sélection, les complexes [ARN/protéine] sont formés dans du tampon D en incubant 30 min à 4°C, 0,1 nmol d'ARN dégénérés (préalablement renaturés 5 min à 65°C, puis 15 min à température), en présence de 0,01 nmol de protéine fusionnée à la GST et de 2 µg d'ARNt de levure (Roche Molecular

Biochemicals) dans un volume final de 30 μ l. Quinze microlitres de billes de Sépharose couplées à du glutathion sont lavés 3 fois avec du tampon D puis incubés pendant 2 heures à 4°C en présence du mélange réactionnel complété à un volume de 300 μ l avec du tampon D contenant du NP40 0,01% (v/v). Les billes sont ensuite lavées 3 fois avec du tampon D contenant du NP40 0,02% (v/v). Les ARN retenus sur les billes sont incubés 30 min à 37°C en présence de 200 μ g de protéinase K (20 mg/ml) puis extraits au phénol-chloroforme et précipités en présence d'acétate de sodium 0,3 M, d'éthanol absolu et de glycogène (1 μ g). Après lavage du culot d'ARN avec 200 μ l d'éthanol 70%, les ARN sont repris dans 17,2 μ l d'eau.

◆ **Conversion des ARN sélectionnés en ADN complémentaire par RT-PCR**

Hybridation de l'oligonucléotide complémentaire de l'extrémité 3' des ARN sélectionnés

ARN sélectionnés (voir ci-dessus)	17,2 μ l
Oligonucléotide 3' (20 μ M)	2,8 μ l
Tampon AMV 10x (voir 2.1.5.)	5 μ l

Le mélange réactionnel est incubé 10 min à 65°C puis refroidi dans la glace.

Réaction de transcription inverse

Mélange réactionnel précédent	25 μ l
dNTP (5 mM de chaque)	1 μ l
Tampon AMV 10x (voir 2.1.5.)	2,5 μ l
AMV Reverse transcriptase (Quantum Appligene, 20 U/ μ l)	0,25 μ l
H ₂ O	qsp 50 μ l

Le mélange réactionnel est incubé 30 min à 42°C.

Amplification des ADN complémentaires synthétisés

Mélange réactionnel précédent	5 μ l
Oligonucléotide 5' (20 μ M)	1 μ l
Oligonucléotide 3' (20 μ M)	1 μ l
Tampon GoTaq 5X (KCl 50 mM ; Triton X-100 0.1% ; Tris/HCl 10 mM, pH 8,8)	10 μ l
MgCl ₂ 25 mM	8 μ l
dNTP (1,25 mM chaque NTP)	8 μ l
GoTaq (Finnzymes, 2U/ μ l)	0,25 μ l
H ₂ O	qsp 50 μ l

Les 10 produits PCR amplifiés sont ensuite utilisés comme matrice pour la transcription d'ARN qui seront utilisés pour un nouveau cycle de sélection en présence de la protéine fusionnée à l'étiquette GST (GST- MBNL1 Δ CT3).

◆ **Clonage des séquences sélectionnées**

A la fin du 4^{ème} cycle de sélection, les fragments d'ADN sont clonés dans le plasmide pCR2.1 (Invitrogen) selon les indications du fournisseur. Trente-sept clones ont alors été sélectionnés aléatoirement et séquencés. Les séquences ayant une forte affinité pour MBNL1 ont été obtenues par sélection sur gel retard. Une expérience de gel-retard a été réalisée entre les ARN issus du 4^{ème} cycle de sélection non radiomarqué (0.1 nmoles) et la protéine MBNL1 Δ CT3 (0.01 nmoles). La bande correspondant aux complexes [ARN/protéine] est découpée et éluée. Les ARN sélectionnés sont convertis en ADN complémentaire par RT-PCR, puis les séquences ADN sont clonées dans le plasmide pCR2.1. Trente clones sont alors sélectionnés puis séquencés comme décrit précédemment.

III. Production et purification de protéines recombinantes

III.1. Production et purification des différentes isoformes des protéines MBNL, CUG-BP1 et NPM

◆ **Production**

La séquence correspondant à l'extrémité C-terminale de la protéine GST (pour Glutathione S-Transferase) a été fusionnée traductionnellement à l'extrémité N-terminale de séquences correspondant à différentes isoformes des protéines MBNL1, CUG-BP1 et NPM dans le vecteur pGEX-6P-1. Les vecteurs obtenus sont pGEX-6P-1-MBNL1, pGEX-6P-1-MBNL1 Δ Cter, pGEX-6P-1-MBNL1 Δ Cter Δ 3, pGEX-6P-1-CUG-BP1, pGEX-6P-1-CUG-BP1 Δ Cter et pGEX-6P-1-NPM. Leur séquence codante est placée sous le contrôle du promoteur *Ptac* qui est inductible par l'ajout d'IPTG. Cette construction est utilisée pour transformer les bactéries compétentes *E. coli* de la souche BL21-CodonPlus-RIL (Stratagene). Un litre de milieu LB supplémenté en ampicilline (100 μ g/ml) et en ZnSO₄ (100 μ M) estensemencé à 0,15 unité de DO₆₀₀/ml à partir d'une pré-culture, et mise à pousser à 37°C sous agitation. Lorsque l'absorbance de la culture atteint 0,8 U DO₆₀₀/ml, la production de la protéine est induite une nuit par l'ajout d'IPTG à une concentration finale de 0,1 mM. Les cellules sont ensuite récoltées par centrifugation à 4000 rpm à 4°C pendant 20 min et le culot

cellulaire est repris dans du tampon PBS (NaCl 140 mM ; KCl 2,7 mM ; Na₂HPO₄ 10 mM ; KH₂PO₄ 1,8 mM, pH 7,3) à 100 mDO₆₀₀/ml.

◆ Purification

Sonication

Les cellules sont plongées dans la glace et sont traitées aux ultrasons sous une puissance de 40 W pendant 2 min chaque impulsion étant séparée d'une seconde. Après traitement, le lysat bactérien est centrifugé pendant 20 min à 4°C à 26 000 g et seul le surnageant qui contient les protéines solubles est utilisé pour la suite de la purification.

Chromatographie d'affinité sur des billes de Glutathion-Sepharose

Le surnageant de sonication est tout d'abord traité par incubation 20 min dans la glace avec du PEI (0.0125%) afin de supprimer les acides nucléiques, puis centrifugé 20 min à 26 000 g. Le surnageant est alors mis en contact avec 1 ml de billes de Sepharose covalamment liées à du glutathion (Glutathione Sepharose[®] 4B, GE Healthcare) préalablement lavées 3 fois avec du tampon PBS. L'ensemble est incubé durant 3 heures à 4°C sur un agitateur rotatif. Les billes sont ensuite centrifugées 2 min à 2 000 g à 4°C puis lavées 3 fois avec 10 ml de tampon PBS supplémenté à 1 M NaCl et 1 fois avec du PBS.

Dans le cas de la production de protéines fusionnées à la protéine GST, les protéines de fusion sont éluées suite à l'ajout de 500 µl de glutathion réduit à 10 mM et l'éluat est ensuite dialysé contre 1 l de tampon D durant une nuit sous agitation à 4°C. La concentration de la protéine a été mesurée en utilisant le réactif de Bradford et le degré de pureté de la protéine purifiée a été estimé en déposant 0,5 µg de la protéine sur un gel SDS-PAGE.

Dans le cas de la production de protéines sans la protéine GST, les protéines sont éluées par coupure au niveau C-terminale de la protéine GST par incubation avec la protéine Prescission (produite au laboratoire) durant une nuit sous agitation à 4°C. Après traitement, les billes sont centrifugées pendant 20 min à 4°C à 2 000 g et lavées 3X avec du PBS. Les surnageants qui contiennent les protéines sans la protéine GST sont utilisés pour la suite de la purification.

Purification par gel filtration

Cette purification est réalisée dans le but d'éliminer les contaminants en fonction de leur taille. L'éluat de la première chromatographie d'affinité est filtré à 0,22 µm puis déposé par volume de 500 µl sur la colonne gel filtration (Superdex[™]75, GE Healthcare)

préalablement équilibrée avec au moins 5 volumes de colonne de tampon D sans glycérol. Les protéines MBNL1 sont sélectionnées en fonction de leur taille et les fractions contenant les protéines sont rassemblées et dialysées contre 1 l de tampon D durant une nuit sous agitation à 4°C. La concentration de la protéine est mesurée en utilisant le réactif de Bradford et le degré de pureté de la protéine purifiée a été estimé en déposant 0,5 µg de la protéine sur un gel SDS-PAGE.

III.2. Production et purification de la protéine de fusion MBP-MS2

◆ Production

La séquence correspondant à l'extrémité C-terminale de la protéine MBP (pour Maltose Binding Protein) a été fusionnée traductionnellement à l'extrémité N-terminale de la séquence correspondant de la protéine de l'enveloppe MS2 du phage R17 dans le vecteur pMAL pour obtenir le vecteur pMAL-MS2. Cette séquence codante est placée sous le contrôle du promoteur *P_{tac}* qui est inductible par l'ajout d'IPTG. Cette construction est utilisée pour transformer les bactéries compétentes *E. coli* de la souche BL21-CodonPlus-RIL (Stratagene). Après transformation, 800 µl de milieu LB sont ajoutés et les bactéries sont cultivées durant 1 heure à 37°C sous agitation. Un volume de 400 µl de cette culture est ensuite prélevé pour inoculer 25 ml d'un milieu LB supplémenté avec 100 µg/ml d'ampicilline et 25 µg/ml de chloramphénicol. Cette pré-culture est ensuite incubée toute une nuit à 37°C sous agitation. Un volume de 20 ml de cette pré-culture est ensuite prélevé et inoculé dans 1 l de milieu de culture LB de même composition. Lorsque l'absorbance de la culture atteint 0,8 U DO₆₀₀/ml, la production de la protéine est induite jusqu'à obtention d'une densité optique de 1,6 UDO₆₀₀/ml, par l'ajout d'IPTG à une concentration finale de 0,1 mM. Les cellules sont ensuite récoltées par centrifugation à 6 000 g à 4°C pendant 15 minutes et le culot cellulaire est lavé avec 10 ml de tampon de sonication (Tris-HCl 20 mM, pH 7,6 ; NaCl 200 mM ; EDTA 1 mM) puis centrifugé de nouveau à 6 000 g à 4°C pendant 15 minutes. Le poids du culot cellulaire est ensuite mesuré et le culot est repris dans six volumes de tampon de sonication.

◆ Purification

Sonication

Les cellules sont plongées dans de la glace et sont traitées aux ultrasons sous une puissance de 40 W avec des impulsions séparées d'une seconde durant 5x2 min, chaque étape de sonication étant séparée d'une période d'incubation de 2 minutes dans la glace. Après

sonication, le lysat bactérien est centrifugé pendant 15 min à 4°C à 30 000 g et seul le surnageant qui contient la protéine soluble est ensuite utilisé pour la suite de la purification.

Chromatographie d'affinité sur résine d'amylose

Le surnageant de sonication est mis en contact avec 1,8 ml de billes d'amylose (NEB) préalablement lavées 3 fois avec du tampon de sonication. L'ensemble est incubé durant 45 min à 4°C sur un agitateur rotatif. Les billes sont ensuite centrifugées 1 minute à 2 000 g à 4°C puis lavées 2 fois avec 10 ml de tampon de sonication puis une fois avec 10 ml de tampon A (Tris-HCl 20 mM, pH 7,6) afin d'enlever les sels. Les protéines de fusion sont ensuite éluées par deux applications successives de 1,5 ml de tampon d'éluion contenant du maltose (Tris-HCl 20 mM, pH 7,6 ; maltose 15 mM).

Chromatographie d'affinité sur colonne Héparine

Cette chromatographie d'affinité est réalisée dans le but d'éliminer les acides nucléiques liés à la protéine MBP-MS2. L'éluat de la première chromatographie d'affinité est filtré à 0,22 µm puis déposé par volume de 1 ml sur la colonne Héparine-Sepharose (HiTrap Heparin HP, GE Healthcare) préalablement équilibrée avec au moins 5 volumes de colonne de tampon Tris-HCl 20 mM, pH 7,6. Une fois la protéine MBP-MS2 retenue sur le gel, un gradient de concentration en NaCl allant de 0 à 1 M NaCl est appliqué. La protéine MBP-MS2 est éluee à une concentration d'environ 250 mM NaCl et les fractions contenant la protéine sont rassemblées et dialysées dans 1 l de tampon D durant une nuit sous agitation à 4°C. La concentration de la protéine est mesurée en utilisant le réactif de Bradford et le degré de pureté de la protéine purifiée a été estimé en déposant 0,5 µg de la protéine sur un gel SDS-PAGE.

IV. Purification des complexes ribonucléoprotéiques par chromatographie d'affinité avec la technique MS2

Le but est d'identifier les composants se liant à un ARN spécifique. Pour ce faire, trois motifs ARN du bactériophage MS2 d'*E. coli* sont ajoutés cotranscriptionnellement à l'extrémité 3' de l'ARN d'intérêt, à savoir, dans notre cas, des répétitions CUG₅₁. Les motifs ARN du bactériophage MS2 se lient avec une forte affinité ($K_d = 38$ pM) à la protéine de l'enveloppe de ce bactériophage. Cette protéine est produite sous forme de fusion avec la protéine MBP, ce qui permet l'immobilisation de l'ARN « hybride » sur une matrice d'amylose où la protéine est fixée. L'ARN ainsi immobilisé peut être incubé avec un

échantillon d'extrait nucléaire afin de former des complexes, qui sont ensuite élués dans des conditions natives par ajout de maltose.

◆ **Formation des complexes sur l'ARN fixé à la matrice**

L'ARN (500 pmol) fusionné à la séquence MS2 est incubé dans un volume de 50 µl de tampon D en présence de 4,5 mM de MgCl₂ pendant 10 minutes à 65°C puis renaturé 10 minutes à température ambiante et enfin placés dans la glace. Deux mille cinq cents pico moles de protéine MBP-MS2 diluées dans un volume de 100 µl de tampon D contenant 2,8 mM de MgCl₂ sont ensuite incubées avec l'ARN durant 30 minutes à 4°C. Lorsque les complexes ARN-MS2/MBP-MS2 sont formés, ceux-ci sont incubés avec un volume de 6,5 mg d'extrait nucléaire et un excès molaire de 100 fois d'ARNt soit 50 nmol. L'ensemble est incubé durant 45 minutes à 4°C sous agitation rotative. Deux cents microlitres de billes d'amylose (Amylose resin, Biolabs) sont lavés 3 fois avec 1 ml de tampon D et ajoutés aux complexes ribonucléoprotéiques durant 1 heure à 4°C sous agitation rotative. L'excès de protéines de l'extrait nucléaire non retenues sur l'ARN est éliminé par une brève centrifugation des billes à 2 000 g à 4°C (Echantillon NR). Les billes sont ensuite lavées 3 fois avec 2 ml de tampon D, chaque lavage étant suivi d'une brève centrifugation à 4°C à 2 000 g (Echantillons L1, L2, L3). Les complexes ribonucléoprotéiques sont ensuite élués par l'ajout successif de 2x200 µl de tampon d'éluion contenant 10 mM de maltose dans du tampon D, chaque éluion étant suivie d'une brève centrifugation à 4°C à 2 000 g. Afin de suivre la répartition de l'ARN et la formation des complexes au cours de l'expérience, 10 µl de chaque échantillon prélevés au cours de l'expérience sont déposés sur un gel natif de polyacrylamide 5 % qui est ensuite coloré au bromure d'éthidium. Les complexes ribonucléoprotéiques contenus dans le surnageant sont ensuite analysés par les techniques décrites ci-dessous.

◆ **Analyse du contenu en protéine des complexes ribonucléoprotéiques par spectrométrie de masse**

Un volume de 100 µl d'éluat (soit 50 % de l'éluat total) auquel est ajouté 33 µl de bleu de dépôt SBL (Tris 1 M, pH 6,8 ; SDS 32 % ; β-mercaptoéthanol 80 % ; bleu de bromophénol 0,16 %, glycérol 40 %) est chauffé 2 minutes à 96 °C puis déposé sur un gel SDS-PAGE (gel de concentration 5 %, gel de séparation à 10%). L'électrophorèse est réalisée dans du tampon de migration (Tris 25 mM ; glycine 192 mM ; SDS 0,1 %) à 25 mA. Le gel est ensuite coloré

avec du bleu de coomassie colloïdal. Après la coloration, le gel est découpé tous les 2 mm et les bandes sont envoyées au laboratoire de spectrométrie de masse bio-organique ECPM pour analyse. Les protéines contenues dans les bandes de gel subissent une digestion trypsique puis les peptides sont ionisés selon la technique d'électro-spray et analysés par nanoLC-MS/MS selon le protocole détaillé par Miguet *et al.* (Miguet et al., 2006).

V. Méthodes utilisées pour la culture cellulaire

V.1 Culture cellulaire

V.1.a Lignée HEK293

La lignée cellulaire HEK293 est cultivée dans le milieu DMEM (Dulbecco's Modified Eagle Medium) supplémenté avec 10% de sérum de veau foetal (FBS), 10 U/ml de pénicilline/streptomycine et 4 mM de glutamine (Gibco). Les cellules sont cultivées en flasques ou en boîtes de différents diamètres (en fonction de l'utilisation) à 37°C, sous une atmosphère humide et à 5% de CO₂.

Les cellules sont réensemencées à subconfluence (2 à 3 jours).

V.1.b. Lignées cellulaires de myoblastes

Ces lignées sont cultivées dans le milieu HamF10 (Gibco) supplémenté avec 20% de sérum de veau foetal et 50 µg/ml de gentamycine (Gibco) Les cellules sont cultivées en flasques ou en boîtes de différents diamètres (en fonction de l'utilisation) à 37°C, sous une atmosphère humide et à 5% de CO₂.

Les cellules sont réensemencées à subconfluence (5 à 6 jours).

V.1.c Technique de transfection

◆ **Transfection de cellules HEK293 par un plasmide au réactif JetPEI^{PM}**

Les cellules HEK (300 000 cellules par puits dans 3 ml de milieu) sont transfectées dans des plaques 6 puits par 1,5 µg de plasmide pSVIRB-hcTNT ou ses différents mutants et 1,5 µg de plasmide pcDNA3.1-(CTG)₉₆₀ ou 0,1 µg de plasmide pcDNA3.1-MBNL1 ou complété par le plasmide pcDNA3.1 vide en utilisant le réactif JetPEI^{PM} (PolyPlus Transfection). Après 5 h d'incubation avec l'ADN, les cellules sont lavées au PBS et remises en culture pendant 48 h.

◆ **Transfection de cellules humaines par un siRNA au phosphate de calcium**

Les cellules HEK293 (200 000 cellules par puits dans 2 ml de milieu DMEM sans antibiotique) sont transfectées dans des plaques 6 puits. Les siRNA (50-100 pmoles) sont dilués à 25 mM dans 36,8 µl finaux de tampon TE 0,1X, 5,2 µl de CaCl₂ 2 M, 42 µl de HBS pH 7,5. Le mélange est incubé 20 min à température ambiante, avant d'être distribué sur les cellules en culture. Après 16h de culture, le milieu de culture est remplacé par 2 ml de milieu DMEM sans antibiotique frais. Après 48h, les cellules sont co-transfectées par un mélange siRNA (50-100 pmoles)-pSVIRB-hcTNT (0,6 µg) en utilisant le réactif JetPEI™. Après 24h, les cellules sont récoltées dans 1 ml de PBS (70% sert à la préparation des ARN totaux et 30 % sert à la préparation d'extraits protéiques).

V.2. Technique d'immunofluorescence

V.2.a. Culture sur lames, fixation et perméabilisation des cellules

Au préalable, les lames sont stérilisées à la flamme et les cellules myoblastes en cultures sont rincées au PBS et décollées des supports plastiques par ajout de trypsine. Le culot cellulaire obtenu est repris dans 2 ml de milieu de culture puis les cellules sont comptées. Cinq cent mille cellules sontensemencées par lame et mises en culture durant une nuit. Puis, les lames sont lavées au PBS puis recouvertes de paraformaldéhyde (PFA) à 4 % pendant 20 min sur un lit de glace. La PFA 4 % est ensuite aspirée et les lames sont rincées au PBS à 3 reprises pendant 5 min. Les cellules sont fixées par leur mise en contact avec une solution de méthanol froid pendant 5 min. Les lames sont ensuite rincées et lavées à 3 reprises dans du PBS-Tween 0.1% pendant 5 min à température ambiante.

V.2.b. RNA-FISH et immunodétection de protéines

Immunodétection de protéines

Les lames sont mises en contact avec une solution de blocage (3 % BSA ; PBS 1X) pendant 30 min à température ambiante. Les lames sont ensuite rincées et lavées à 3 reprises dans du PBS-Tween 0.1% pendant 5 min à température ambiante. L'anticorps primaire (dilué dans une solution de PBS-Tween 0,1 %) est déposé sur la lame et incubé pendant 1h30 à température ambiante dans une chambre humide. Après 3 lavages successifs, l'anticorps secondaire (dilué dans la solution PBS-Tween 0,1 %) est déposé sur la lame pendant 45 min à température ambiante dans une chambre humide à l'abri de la lumière. Après 3 nouvelles étapes de lavage précédemment décrites, les lames sont prêtes à être montées en absence d'hybridation.

Hybridation de la sonde

Les lames sont ensuite traitées par une solution de formamide déionisée 40 % / SSC 2 X (300 mM CaCl₂ ; 30 mM citrate de sodium pH 7) durant 5 min. Un mélange d'hybridation est ajouté directement sur les lames et incubé dans une chambre humide (SSC 2X/formamide 50%) pendant 16h à 37°C. Ce mélange contient 200 nM de sonde marquée diluée dans du tampon d'hybridation (10 % de sulfate de dextrane, 0,2 % BSA, 0,1 mg/ml de sperme de saumon ; 0,1 mg/ml d'ARNt ; 4 mM VRC). La sonde utilisée est un oligonucléotide Cy3-(CAG)₁₀ (Qiagen operon, Cologne, Allemagne). Les lames sont ensuite rincées et lavées à 3 reprises dans du PBS-Tween 0.1% pendant 5 min à température ambiante.

V.2.c. Montage des lames et acquisition des images

Les lames sont incubées avec une solution DAPI (50 ng/ml diluée dans du PBS) durant 1 min, lavées 3X comme décrit précédemment puis montées. L'acquisition des images est réalisée avec un microscope confocale Leica.

V.3. Préparation d'extrait nucléaire

Toutes les étapes doivent se dérouler à 4°C. Après lavages au PBS 1X, les cellules sont grattées, récoltées puis centrifugées pendant 5 min à 200g. Après suppression du surnageant, le culot cellulaire est repris à une concentration de 10⁷ cellules/ml dans un tampon hypotonique (Tris-HCl 50 mM, pH 7,5 ; NaCl 150 mM ; EDTA 1 mM ; NP-40 0,1 %), contenant un mélange d'inhibiteur de protéase, pendant 30 min à 4°C. Le lysat est régulièrement mélangé, puis centrifugé 15 min à 16 000g durant 15 min. Le surnageant est récupéré et conservé à -80°C. La concentration en protéine est déterminée par la méthode de Bradford.

V.4. Analyse d'extrait nucléaire par Western Blot

Dix microgrammes d'extrait protéiques cellulaires sont séparés sur un gel SDS-PAGE 10 % en parallèle d'un marqueur de poids moléculaire (PageRuler, MBI Fermentas). Les protéines sont alors transférées électriquement sur une membrane de nitrocellulose Trans-Blot[®] Transfert Medium (BIORAD) à 100 V constant 1h dans du tampon de transfert 1X. Les membranes sont incubées dans une solution de PBS -Tween 0,1 % avec 5 % de lait demi-écrémé pendant 1 h sous agitation en présence de l'anticorps primaire utilisé avec les dilutions indiquées dans le Tableau III. Après 3 nouveaux lavages, les membranes sont incubées 40

min sous agitation avec l'anticorps secondaire dilué au 1/20000^{ème}. Les membranes sont ensuite à nouveau lavées trois fois avec du PBS-Tween 0,1 % puis les protéines sont révélées par une solution de luminol et substrats peroxydase mélangés volume à volume provenant du kit ImmobilonWestern (Millipore).

VI. Techniques utilisées pour analyser l'épissage

VI.1. Méthodes spécifiques de l'étude de l'épissage *in vitro*

VI.1.a Préparation de la matrice de transcription

Les matrices ADN utilisées pour la transcription ont été obtenues après une amplification par PCR avec les couples d'oligonucléotides correspondants.

Le mélange réactionnel est le suivant :

Plasmide (pSVIRB-hcTNT)	20 ng
Oligonucléotide 5' (7254) (20 µM)	1 µl
Oligonucléotide 3' (7255) (20 µM)	1 µl
Mélange de dNTP (1, 25 mM de chaque)	8 µl
Tampon Pfu (DTT 100 mM ; 0,1 mM EDTA ; KCl 100 mM ; Nonidet P40 0,1% ; Tween 20 0,1% ; glycérol 50% ; Tris-HCl 200 mM, pH 8,8)	5 µl
Pfu polymérase (5 U/µl, MBI Fermentas)	0,5 µl
H ₂ O	qsp 50 µl

Après un cycle de dénaturation de 5 minutes à 95°C, l'amplification se déroule en 30 cycles qui sont chacun composé de paliers de 30 secondes à 95°C, 30 secondes à 55°C et 1 minute à 74°C pour ensuite terminer par une incubation de 10 minutes à 74°C. Le temps d'élongation dépend de la taille du fragment à amplifier, sachant que la processivité de la Pfu polymérase est de 2 min/kb. Les produits PCR sont séparés par électrophorèse en gel d'agarose à 1,8 % (p/v) et purifiés en utilisant le kit "Nucleospin[®] extract" (Macherey-Nagel).

L'oligonucléotide 7254-TAATACGACTCACTATAGGGCCTTTGTGGTTCTCACTTGG- comporte à son extrémité 5' la séquence du promoteur du phage T7 qui permet la transcription des matrices ADN. Il s'hybride au niveau de l'exon 2 du récepteur à insuline dans le plasmide SVIRB-hcTNT. L'oligonucléotide 7255- GCGCTTCTGCCGGGCCACCT s'hybride sur l'exon 3 du récepteur à insuline présent dans le vecteur SVIRB-hcTNT.

VI.1.b Transcription *in vitro*

Le protocole expérimental utilisé pour la transcription des ARN d'intérêt est le suivant :

Matrice PCR	1,5 pmol
Tp 5X (MgCl ₂ 30 mM ; Spermidine 10 mM ; NaCl 50 mM ; Tris-HCl 200 mM, pH 7,5)	8 µl
rNTP (A, C, U, G ; 5, 5, 5, 2 mM)	4 µl
Cap ⁵ m ⁷ GpppG (10 mM)	2 µl
Inhibiteur de RNases (RNasin, 20 U/µl, Amersham Biosciences)	2 µl
BSA (2 mg/ml)	2 µl
DTT (200 mM)	2 µl
ARN polymérase T7	2 µl
H ₂ O	qsp 40 µl

Incubation pendant 2 h à 37°C

Après une étape de dégradation de la matrice ADN comme décrit précédemment, les ARN sont purifiés sur colonne Sephadex G-50 (mini Quick Spin Columns, Roche).

VI.1.c. Epissage *in vitro*

Les expériences d'épissage *in vitro* sont réalisées avec des extraits nucléaires de cellules HeLa provenant de la société CilBiotech. Le mélange réactionnel est le suivant :

ARN pré-messenger (10 fmol/µl)	2 µl
MgCl ₂ (62,5 mM)	1 µl
ATP (25 mM)	1 µl
Créatine phosphate (625 mM)	1 µl
Inhibiteur de RNases (RNasin, 20 U/µl, Amersham Biosciences)	0,5 µl
Alcool polyvinylique 13 %	6 µl
Extrait nucléaire (10 mg/ml, traité au préalable 10 min à 30°C)	8 µl

Le mélange est complété par 2,5 µl de Tampon D ou de concentration croissante de protéine GST-MBNL1 et de 1 µl d'H₂O ou d'une gamme croissante d'ARN SELEX.

Incubation 5 min à 0°C puis 2h à 30°C

La réaction d'épissage est ensuite arrêtée par ajout de 20 µg de protéinase K (Roche) fonctionnant dans le tampon suivant : EDTA 12,5 mM ; NaCl 150 mM ; SDS 1 % ; Tris-HCl 100 mM, pH 7,4. Après une incubation de 30 minutes à 30°C, les protéines sont extraites au phénol/chloroforme (1/1), puis au chloroforme/alcool isoamylique (24/1) et les ARN sont précipités en présence d'éthanol absolu sans ajout de sels. Le précipité est lavé puis séché et repris dans 10 µl d'eau stérile.

VI.1.d. Conversion des ARN en ADN complémentaires par RT-PCR

Hybridation de l'oligonucléotide complémentaire à l'extrémité 3' des ARN

ARN	2 µl
Amorce (5 µM)	1 µl
H ₂ O	13 µl

Incubation 5 min à 65 °C puis 4°C

Réaction de transcription inverse

dNTP (10 mM de chaque)	3 µl
Tp 5X (Tris/HCl 250 mM pH 8,3 ; KCl 375 mM ; MgCl ₂ 15 mM ; DTT 50 mM)	5 µl
Transcriptase inverse (MMLV, 200 U/µl, Promega)	1 µl

Incuber 1h à 42 °C puis conserver à -20°C

Amplification des ADN complémentaires synthétisés

Matrice	2 µl
dNTP (10 mM)	1 µl
Tp10X (Tris-HCl 10 mM, pH 8,8 ; KCl 50 mM ; Triton X-100 0.1 %)	5 µl
Amorces (25 mM de chaque)	1 µl
GoTaq (Promega, 5 U/µl)	0,5 µl
MgCl ₂ (25 mM)	2 µl
H ₂ O	qsp 50 µl

Après un cycle de dénaturation de 5 minutes à 96°C, l'amplification se déroule en 29 cycles qui sont chacun composés de paliers de 45 s à 96°C, 45 s à 58°C et 1 minute à 72°C, pour ensuite terminer par une incubation de 10 minutes à 72°C. Les produits PCR amplifiés sont ensuite analysés par électrophorèse en gel low melting Nusieve GTG agarose 4 % en présence de Bromure d'éthidium et analysés par le système d'images Genegenius (Syngene). Les séquences des oligonucléotides utilisés pour les expériences de RT-PCR sont répertoriées dans le Tableau XI.

Oligonucléotides	Utilisation
7255- GCGCTTCTGCCGGCCACCT	RT pour épissage <i>in vitro</i>
dT15- TTTTTTTTTTTTTTTT	RT pour épissage <i>in vivo</i>
5991- CCTGGTGTGTGGGGAGCGTGG 8078- TCACCTGCCCGGGCCTCC	PCR

VI.2. Méthodes spécifiques de l'étude de l'épissage *in vivo*

VI.2.a Extraction des ARN totaux

Les cellules HEK293 transfectées sont lavées au PBS 1X et les ARN totaux sont extraits des cellules au Trizol selon les spécifications du fournisseur (Invitrogen).

VI.2.b. RT-PCR

Deux microgrammes d'ARN totaux sont digérés par 2 μ l de DNase RQ1 (1 UE/ μ l, Promega) dans un volume réactionnel de 20 μ l pendant 1h à 42°C puis la réaction est stoppée par ajout de 1 μ l de solution « RQ1 STOP » et incubation de 10 min à 65°C. La réaction de RT-PCR est réalisée selon le protocole décrit précédemment. Seule la quantité de matrice ARN utilisée pour l'étape de transcription inverse varie. Les séquences des oligonucléotides utilisés pour les expériences de RT-PCR sont répertoriées dans le Tableau XI.

Résultats

Résultats et Discussion

Chapitre I : Etude des propriétés de fixation à l'ARN de MBNL1

I. Matériel utilisé

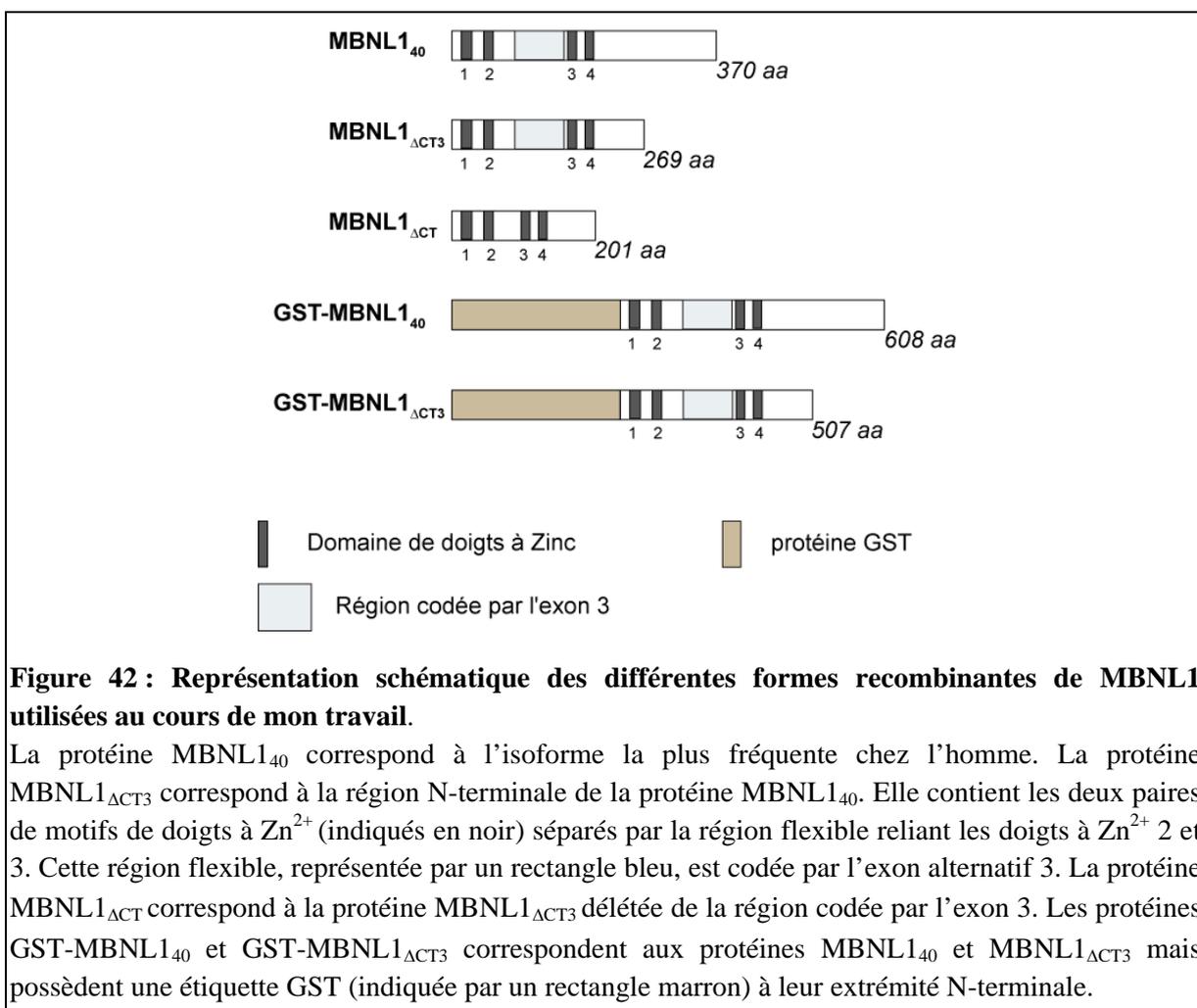
I.1. Production de protéines MBNL1 recombinantes

Un des pré-requis à l'étude des propriétés d'interaction de la protéine MBNL1 avec les ARN était de disposer de quantité suffisante de la protéine MBNL1 recombinante. Avant mon arrivée au laboratoire, N. Marmier-Gourrier (chercheur post-doctoral) avait cloné la séquence ADN correspondant à l'isoforme majoritaire de MBNL1, MBNL1₄₀, dans le vecteur pGEX-6P1. Cette isoforme contient uniquement l'exon alternatif 3 en plus des exons constitutifs. Le clonage avait été réalisé à partir d'un plasmide contenant la séquence codante de cette protéine qui nous avait été fourni par N. Sergeant (INSERM, Lille). Le vecteur pGEX-6P1 permet de fusionner une étiquette GST en N-terminal de la protéine d'intérêt. Cette étiquette peut être clivée par utilisation de la protéase PreScission dont un site de reconnaissance est localisé immédiatement en aval de l'étiquette GST. Différentes conditions de production de la protéine MBNL1₄₀ chez *E. coli* avaient été testées. Cependant, du fait de problèmes de solubilité de la protéine, seules de petites quantités de MBNL1₄₀ avaient pu être purifiées. Ces quantités n'étaient pas suffisantes pour les analyses envisagées. C'est pourquoi nous avons entrepris de réaliser des constructions permettant de produire des formes tronquées de cette protéine afin d'en améliorer la solubilité.

Les données bibliographiques de l'époque indiquaient que la protéine MBNL1 tronquée de 101 aa C-terminaux (MBNL1_{ΔCT3}) est capable d'interagir avec des répétitions CUG et CCUG dans un test triple-hybride, et ceci avec la même affinité que la protéine entière (Kino *et al.*, 2004). Cette forme tronquée semblait donc avoir les mêmes propriétés de fixation à l'ARN que la protéine MBNL1 entière. Les expériences de triple-hybride avaient également mis en évidence l'importance de l'exon alternatif 3 pour la reconnaissance des répétitions CUG ou CCUG par MBNL1 (Kino *et al.*, 2004). Cinq isoformes sur les 9 isoformes naturelles de MBNL1 ne contiennent pas cet exon, d'où l'importance de l'étude de l'effet de l'inclusion de cet exon sur la reconnaissance des ARN par MBNL1. Nous avons donc préparé des constructions pGEX-6P1 permettant d'exprimer des protéines MBNL1

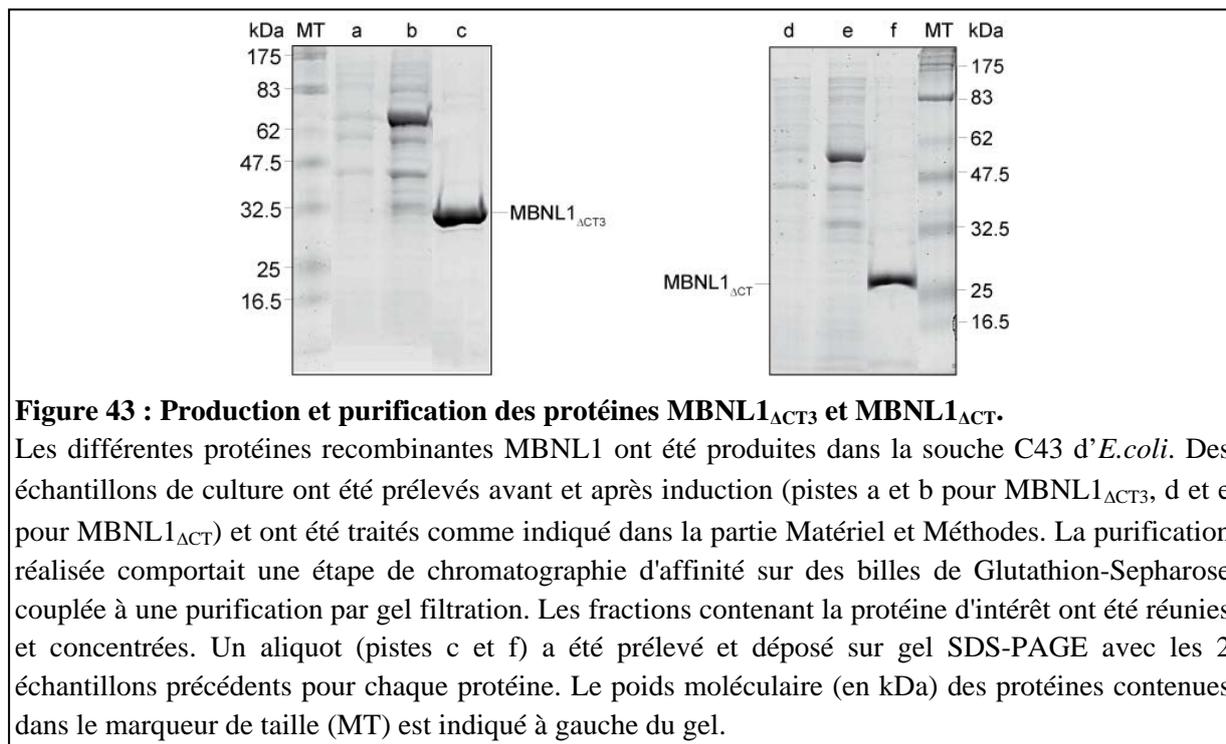
tronquées des 101 aa C-terminaux contenant la séquence codée par l'exon 3 (MBNL1 Δ CT3) ou ne la contenant pas (MBNL1 Δ CT) chez *E. coli*.

Le plasmide pGEX-6P1 MBNL1 Δ CT3 a été obtenu en amplifiant la région codant pour les 269 aa N-terminaux de la protéine MBNL1₄₀ avec des amorces permettant d'insérer les sites de restriction *Bam*HI et *Xho*I. La même approche a été utilisée pour produire le plasmide d'expression MBNL1 Δ CT. Dans ce cas, l'ADN matrice ne contient pas la séquence correspondant à l'exon 3. Ainsi, seule l'absence ou la présence de cet exon 3 différencie les 2 isoformes recombinantes MBNL1 Δ CT3 et MBNL1 Δ CT (Fig. 42).



Sur la base des tests d'expression réalisés par N. Marmier-Gourrier pour la protéine MBNL1₄₀, nous avons pu établir un protocole expérimental de purification des protéines MBNL1 tronquées. Cependant, afin d'améliorer la pureté de la protéine, j'ai ajouté une étape d'élimination des acides nucléiques par précipitation au PEI à ce protocole, ainsi qu'une étape de purification par gel filtration (SuperdexTM75), comme décrit dans la partie Matériel et Méthodes. La concentration et la pureté des protéines recombinantes GST-MBNL1 Δ CT3,

MBNL1 Δ CT3 et MBNL1 Δ CT ont été vérifiées par électrophorèse SDS-PAGE et j'ai mesuré leur concentration par test de BRADFORD (Fig. 43).



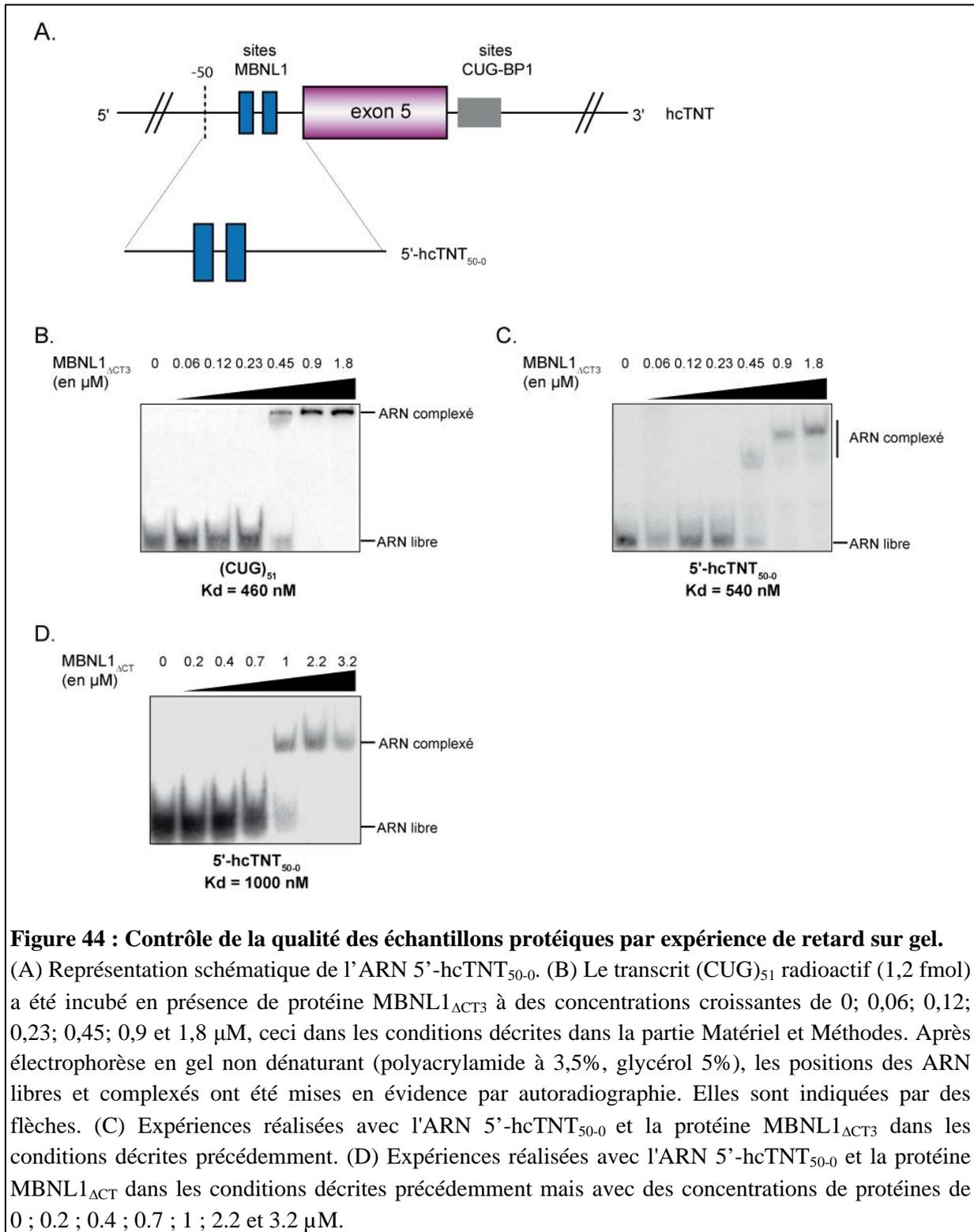
I.2. Contrôle de la qualité des échantillons protéiques par expérience de retard sur gel

Afin d'étudier la capacité de liaison à l'ARN des isoformes tronquées de la protéine MBNL1, j'ai réalisé des expériences de retard sur gel. La capacité de la protéine MBNL1 Δ CT3 à se fixer sur des répétitions CUG (ARN (CUG) $_{51}$), ainsi que sur un fragment du pré-ARNm de la troponine T cardiaque a été testée. Comme décrit dans l'Introduction, deux sites de fixation de MBNL1 avaient été identifiés dans les 50 nucléotides 3' terminaux de l'intron 4 du pré-ARNm hcTNT, nous avons donc produit un ARN contenant ces 50 nt par transcription *in vitro* (5'-hcTNT $_{50-0}$) (Fig. 44A). Nous avons d'abord défini les concentrations en protéines requises pour obtenir une complexation totale de l'ARN. Pour cela, 1,2 fmoles d'ARN radio-marqué en 5' ont été incubées dans un volume final de 11 μ l, en présence de quantités croissantes de protéines allant de 0.06 à 3.2 μ M, ceci dans les conditions décrites dans la partie Matériel et Méthodes. Après électrophorèse en gel de polyacrylamide à 5 ou 8 % non-dénaturant, les positions des ARN libres et des ARN complexés ont été mises en évidence par autoradiographie.

Comme attendu, la protéine MBNL1 Δ CT3 présente une bonne affinité pour les ARN (CUG) $_{51}$, avec une valeur de constante de dissociation (Kd) estimé à 460 nM. Nous observons l'apparition d'agrégats au niveau des poches aux concentrations les plus élevées en protéines,

ceci peut s'expliquer soit par la formation de complexes de haut poids moléculaire par fixation de plusieurs molécules de protéine le long des répétitions CUG, soit par l'agrégation de la protéine lors de l'entrée dans le gel du fait de sa forte concentration (Fig. 44B). De manière intéressante, la protéine MBNL1 Δ CT3 présente une affinité similaire ($K_d = 540$ nM) pour la région intronique du pré-ARNm hcTNT (Fig. 44C). La protéine MBNL1 Δ CT est également capable de se fixer sur l'ARN 5'-hcTNT₅₀₋₀ mais avec une moins bonne affinité ($K_d = 1000$ nM), confirmant l'importance de la séquence codée par l'exon 3 pour la fixation de MBNL1 sur ses ARN cibles (Fig. 44D).

Nous montrions donc, que les deux isoformes tronquées de MBNL1, que nous avons produites étaient capables de se fixer sur des ARN cibles de MBNL1 et pouvaient, par conséquent, être utilisées pour comprendre les modalités de reconnaissance des ARN par MBNL1. Nous pouvions donc utiliser pour notre étude les protéines MBNL1_{CT3} et MBNL1_{CT}, avec ou sans tag GST, ainsi que de petites quantités des protéines GST-MBNL1₄₀ et MBNL1₄₀ préalablement produites par N. Marmier-Gourrier.



I.3. Production des ARN (CUG)₁₆ et (CUG)₁₇

Un ARN contenant de longues répétitions CUG était supposé pouvoir former différentes structures tige-boucles différant par la position de la boucle interne au sein de la séquence répétée (Napierala et Krzyzosiak, 1997). Afin d'obtenir une structuration homogène des molécules d'ARN contenant de telles répétitions, nous avons ajouté aux extrémités de la séquence répétée des séquences supplémentaires pouvant former une hélice très stable constituée de paires de bases G-C canoniques. Ainsi, N. Marmier-Gourrier avait conçu des constructions permettant de produire des ARN renfermant soit 16, soit 17 répétitions de triplets CUG dont la structure 2D était stabilisée par la formation de 6 paires de bases G-C et C-G prolongeant la structure hélicoïdale formée par les répétitions CUG. Les deux constructions génétiques qu'elle avait réalisées, permettaient d'obtenir, par amplification par PCR avec les couples d'oligonucléotides 4508/4506 et 4508/4507, puis transcription *in vitro* par l'ARN polymérase du phage T7, les ARN (CUG)₁₆ et (CUG)₁₇, respectivement. Après déphosphorylation, les transcrits ont été marqués radioactivement en 5' par emploi d'ATP-[γ -³²P] et de T4 polynucléotide kinase dans les conditions décrites dans la partie Matériel et Méthodes.

II. Etude de la fixation de MBNL1 sur des répétitions CUG

Comme nous l'avons vu dans la partie Introduction, la protéine MBNL1 avait été caractérisée initialement comme un facteur nucléaire recruté par les ARN contenant de longues répétitions CUG se structurant en tige-boucle (Miller *et al.*, 2000). Néanmoins, aucune expérience n'avait permis de tester si la fixation de MBNL1 sur la structure tige-boucle particulière formée par ces répétitions était directe. Nous avons donc étudié la fixation de MBNL1 sur des ARN contenant des répétitions CUG en réalisant des empreintes enzymatiques en solution.

En parallèle, nous avons testé la fixation de CUG-BP1 sur les structures tige-boucles formées par les répétitions CUG. En effet, les données concernant cette protéine, qui joue également un rôle majeur dans la DM, n'étaient pas claires. Cette protéine avait été identifiée pour la première fois du fait de sa capacité à se fixer sur un oligonucléotide (CUG)₈, ainsi que sur un ARN contenant des répétitions CUG insérées dans la région 3'-UTR de l'ARNm *DMPK* (Timchenko *et al.*, 1996a et b). Cependant, la capacité de CUG-BP1 à se fixer sur des répétitions CUG avait ensuite été remise en cause par différentes études, qui tendaient à prouver que CUG-BP1 avait une faible affinité pour les répétitions CUG et se fixait sur des répétitions (UG)_n (Michalowski *et al.*, 1999 ; Takahashi *et al.*, 2000 ; Kino *et al.*, 2004 ;

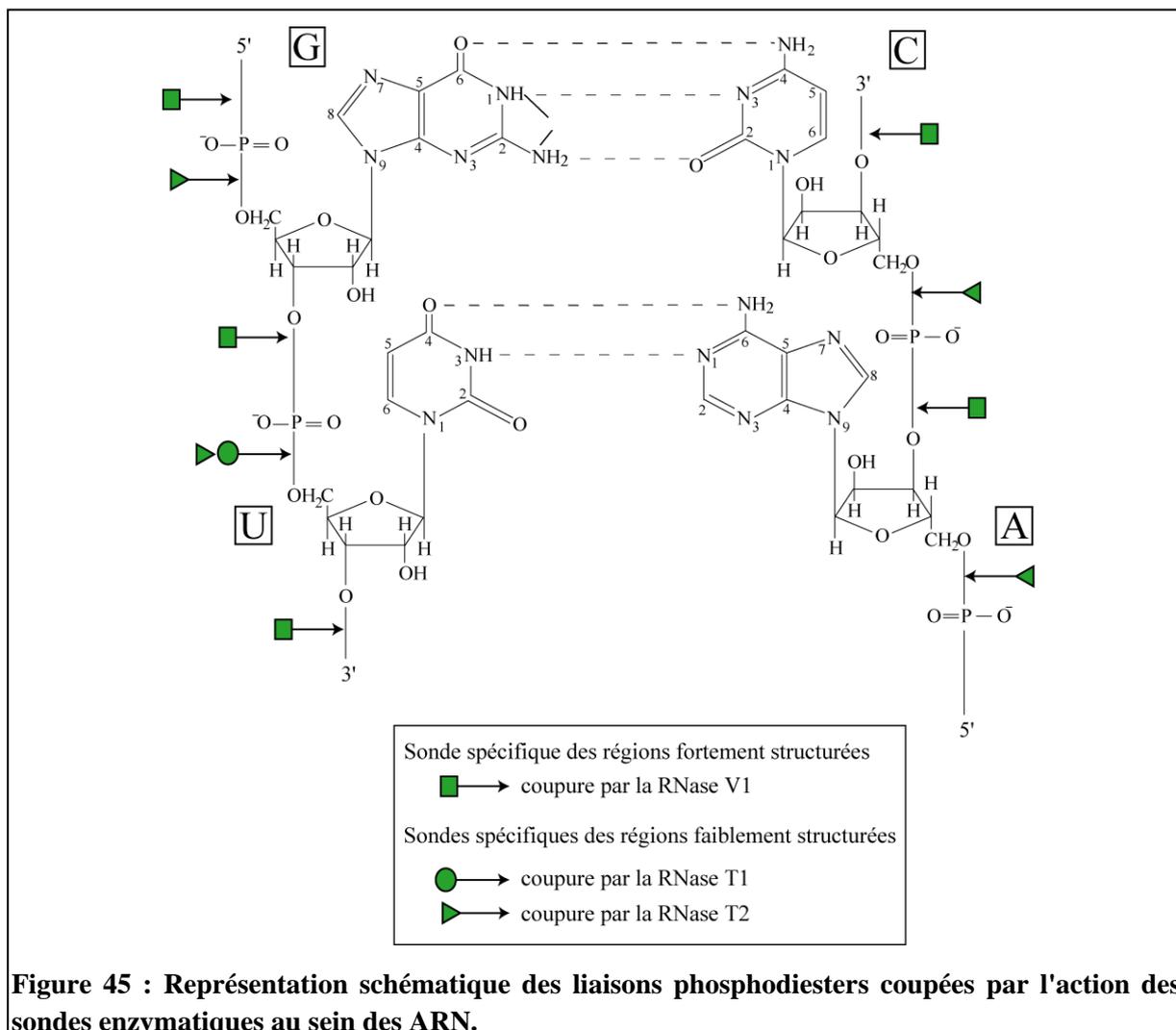
Delaunay *et al.*, 2004). Avant mon arrivée au laboratoire, N. Marmier-Gourrier avait cloné et produit la protéine de fusion GST-CUG-BP1 selon le protocole indiqué dans le Matériel et Méthodes.

II.1. Etude de la structure secondaire des ARN (CUG)₁₆ et (CUG)₁₇

Dans un premier temps, il fallait vérifier que la structure 2D prédite pour les ARN (CUG)₁₆ et (CUG)₁₇ était correcte. Pour cela, la structure 2D des ARN produits par transcription *in vitro* a été analysée par emploi de 3 sondes enzymatiques :

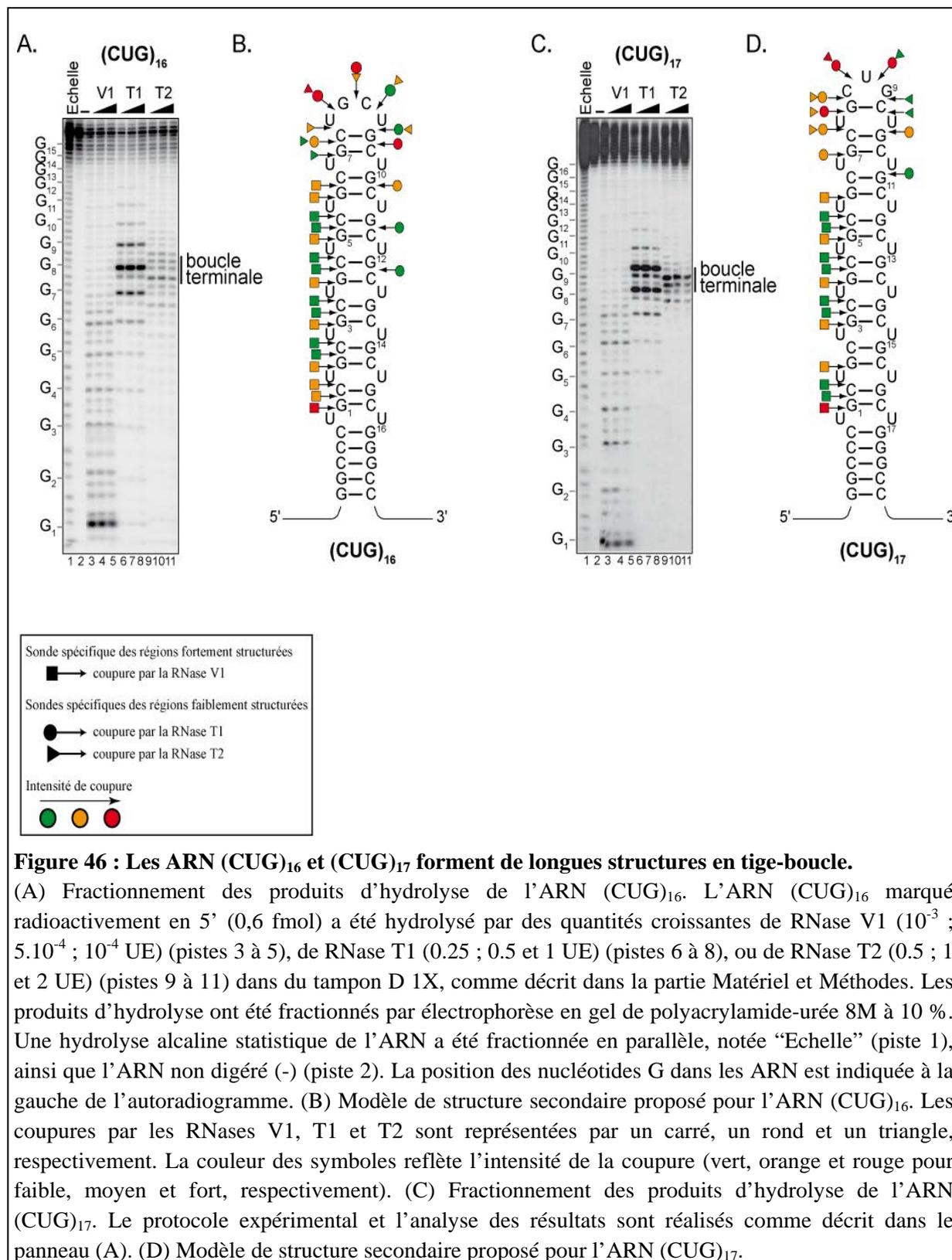
- La RNase T1 qui, en conditions limitantes, hydrolyse spécifiquement les ARN en 3' des résidus guanine et préférentiellement dans les segments non appariés (Fig. 45).
- La RNase T2 qui, en conditions limitantes, hydrolyse aussi préférentiellement les régions en simple brin. Cette enzyme agit sans spécificité marquée vis-à-vis des nucléotides encadrant la liaison phosphodiester clivée et libère, comme la RNase T1, des fragments ARN possédant des extrémités 3' phosphate et 5'-OH (Fig. 45).
- La RNase V1 qui, en conditions limitantes, hydrolyse préférentiellement les régions en double-brin et les régions structurées par empilement de bases, ceci avec peu de spécificité vis-à-vis de l'identité des résidus. Cette enzyme libère des fragments avec des extrémités 3'-OH et 5'-phosphate (Fig. 45).

Des conditions d'utilisation statistiques de ces sondes avaient été mises au point au laboratoire de façon à ce qu'au maximum une coupure ait lieu par molécule. Nous avons dû adapter ces conditions à l'ARN que nous étudions. Les conditions expérimentales utilisées sont décrites dans le chapitre Matériel et Méthodes.



Les données d'hydrolyses ménagées obtenues pour les ARN $(CUG)_{16}$ et $(CUG)_{17}$ sont en accord avec les résultats obtenus précédemment par Sobczak *et al.* (2003), à savoir que les deux ARN formaient une structure tige-boucle régulière fermée à son extrémité par des paires de bases G-C et C-G (Fig. 46). En effet, des coupures par la RNase V1 étaient détectées le long d'un des 2 brins de l'hélice et des coupures par les RNases T1 et T2 étaient localisées dans un segment de 8 nucléotides incluant les résidus supposés former la boucle terminale prédite. Les coupures les plus fortes correspondaient très précisément à cette boucle terminale. L'existence de coupures de plus faible intensité au niveau des résidus encadrant cette boucle suggérait une plus faible stabilité des paires de bases correspondantes ou une légère hétérogénéité de positionnement de la boucle terminale dans la population d'ARN étudiée. Cette instabilité semble plus grande pour la boucle à 4 résidus de l'ARN $(CUG)_{16}$ que pour la boucle à 3 résidus de l'ARN $(CUG)_{17}$. Il est important de noter que les clivages par la

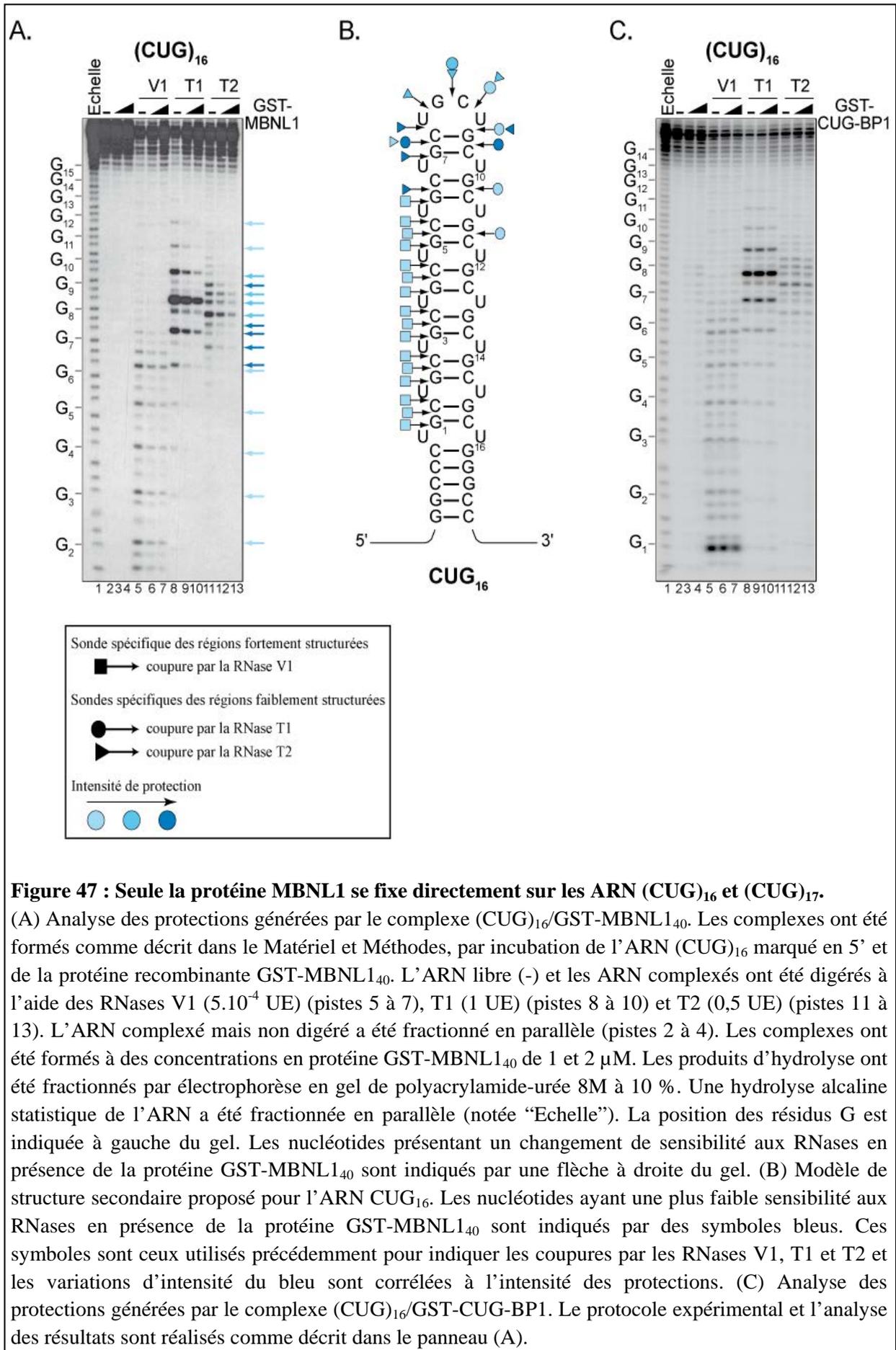
RNase V1 étaient uniquement observés sur le brin 5' et pas sur le brin 3', ce qui peut être lié à la structure particulière de l'hélice formée par les répétitions.



II.2. Etude de la fixation des protéines MBNL1 et CUG-BP1 sur les ARN (CUG)₁₆ et (CUG)₁₇

Afin d'étudier l'interaction des protéines MBNL1 et CUG-BP1 avec les ARN (CUG)₁₆ et (CUG)₁₇, j'ai ensuite réalisé des empreintes enzymatiques en solution. Pour cela, le transcrit a été soumis à une hydrolyse ménagée par les RNases V1, T1 et T2 (5.10^{-4} UE, 0,5 UE et 1 UE/1,2 fmol d'ARN) après incubation en absence ou en présence des protéines GST-MBNL1₄₀ ou GST-CUG-BP1 à des concentrations allant de 1 à 2 μ M. Les expériences réalisées avec la protéine GST-MBNL1₄₀ ont montré qu'elle recouvre entièrement les ARN (CUG)₁₆ et (CUG)₁₇, en induisant une protection à la fois de la partie hélicoïdale et de la boucle terminale (Fig.47).

A l'inverse, aucune protection n'a été détectée avec la protéine GST-CUG-BP1, aussi bien au niveau de la partie hélicoïdale, que de la boucle terminale (Fig.47). Ceci confirmait bien que la protéine CUG-BP1 ne reconnaît pas les grandes structures tige-boucles formées par les ARN à répétitions CUG. Ces résultats apportaient par contre la preuve que MBNL1 s'associe directement aux répétitions CUG. Ces résultats ont été utilisés afin d'illustrer la méthode d'empreintes en solution dans le chapitre « Analysis of Site-Specific RNA-Protein Interactions » du livre « Alternative pre-mRNA splicing », chapitre dont je suis 2^{ème} auteur et qui est inséré dans la partie Matériel et Méthodes de ce manuscrit.



III. Recherche des séquences ARN ayant une forte affinité pour la protéine MBNL1 par la méthode SELEX

En plus des répétitions CUG et CCUG, la protéine MBNL1 a la capacité de se lier à des séquences cibles au sein des pré-ARNm dont elle régule l'épissage et sur certains pré-miRNA dont elle régule la maturation. Une question importante était de savoir si MBNL1 présente un mode de fixation similaire pour chacun de ces différents types d'ARN partenaires ou si différents modes d'interaction sont utilisés selon le type d'ARN ou même pour un même type d'ARN. Le fait de rechercher les séquences et motifs ARN ayant une forte affinité pour MBNL1 pouvait nous apporter des informations permettant de formuler des hypothèses à ce sujet.

Comme indiqué dans l'Introduction et comme nous venons de le voir, MBNL1 reconnaît les motifs CHG/CHHG (H = A, U ou C) qui se structurent en une tige-boucle contenant des mésappariements au niveau de la tige. De façon intéressante, MBNL1 n'interagit pas avec une structure tige-boucle canonique formée par des répétitions CUG/CAG, ce qui suggère une préférence de MBNL1 pour les ARN double-brin contenant des mésappariements (Kino *et al.*, 2004). En ce qui concerne les pré-ARNm endogènes cibles de MBNL1, une séquence consensus YGCU(U/G)Y où Y correspond à U ou C avait été identifiée par alignement des séquences de 4 sites MBNL1 caractérisés (Kino *et al.*, 2004). Les pré-ARNm cibles contiennent généralement plusieurs séquences YGCY relativement proches et une question importante était de savoir si ces motifs répétés s'associent pour former une hélice imparfaite de type A identique à celle formée par les répétitions CUG ou participent à la formation de structures canoniques.

III.1. Présentation de l'article n°1

Afin d'identifier les paramètres de séquences et de structure nécessaires à une fixation efficace de MBNL1 sur les ARN, nous avons réalisé une expérience SELEX. Cette approche consiste à sélectionner des ARN ayant une forte affinité pour une protéine donnée au sein d'une population d'ARN, dont la séquence est dégénérée sur une fenêtre de 10 à 20 résidus. Les ARN dégénérés ainsi produits sont sélectionnés pour leur capacité d'interaction avec la protéine d'intérêt qui est fixée sur une résine. Après 4 à 15 cycles de sélection-amplification, on obtient généralement un fort enrichissement en ARN ayant une forte affinité pour la protéine étudiée (Clery *et al.*, 2007a ; pour revue, Djordjevic, 2007).

L'approche avait déjà été mise en œuvre avec succès au laboratoire pour identifier les motifs ARN reconnus par deux protéines de la famille L7Ae : Snu13p et SBP2 (Clery *et al.*,

2007a,b). L'originalité de l'approche utilisée au laboratoire, par rapport aux approches plus classiques, est d'utiliser des ARN formant une structure hélicoïdale de type Watson-Crick. Ces ARN sont obtenus par dégénérescence d'une séquence de 18 nt au sein d'une structure tige-boucle du snoRNA U3 de *S. cerevisiae*. En raison de la faible solubilité de la forme recombinante de la protéine MBNL1 entière, la protéine de fusion GST-MBNL1 Δ CT3, qui possède les mêmes capacités de fixation à l'ARN que la forme entière, a été utilisée pour réaliser l'expérience SELEX. Cette protéine a ensuite été utilisée pour l'ensemble de nos travaux.

Après 4 cycles d'amplification-sélection une forte augmentation de l'affinité de la protéine MBNL1 Δ CT3 pour les ARN sélectionnés a été détectée par retard sur gel. Les ADNc amplifiés à partir des ARN du cycle 4 ont été clonés dans le plasmide pCR2.1 et 37 clones, appelés clones AC pour Affinity Chromatography ont été arbitrairement sélectionnés parmi les clones obtenus. Par ailleurs, afin d'augmenter la spécificité de la sélection, une étape de sélection par retard sur gel a été ajoutée. Pour cela, une fraction des ADNc amplifiés à partir des ARN du cycle 4 a été utilisée comme matrice pour une réaction de transcription *in vitro* et les transcrits obtenus ont été utilisés pour réaliser un retard sur gel avec la protéine MBNL1 Δ CT3. Les complexes ARN-MBNL1 Δ CT3 ont été excisés du gel et les ADNc amplifiés à partir des complexes purifiés ont été clonés dans le plasmide pCR2.1. Trente clones, appelés clones E pour Electrophoresis ont été arbitrairement sélectionnés parmi les clones obtenus.

A partir des 67 clones (37 clones AC + 30 clones E), 50 ARN différents ayant une forte affinité pour la protéine MBNL1 Δ CT3 ont été identifiés. Trente-quatre de ces ARN (Famille I) contenaient un ou plusieurs motifs YGCY localisé(s) au sein d'une séquence riche en résidus pyrimidine. La fixation d'une ou plusieurs molécules de protéine MBNL1 Δ CT3 sur ces ARN a été confirmée par des expériences de retard sur gel. Aucune autre caractéristique commune n'a pu être détectée pour cette famille d'ARN si ce n'est la localisation d'un motif YGCY à l'extrémité 3' de la séquence dégénérée. Les 16 autres ARN identifiés (Famille II) contenaient tous une séquence conservée de 12 nucléotides dont le consensus est YGUUUGCGCUUG. Cette séquence inclut un motif canonique de fixation à MBNL1 (souligné). De façon intéressante, tous les ARN de la Famille II se replient en une longue structure tige-boucle très stable, préservée par compensation de paires de bases dans les divers ARN sélectionnés. Le motif YGCY est localisé dans la boucle terminale dont la séquence est UUUGCGCU. L'affinité de la protéine MBNL1 Δ CT3 pour les ARN sélectionnés a été estimée par des expériences de retard sur gel et comparée à celle estimée pour les ARN (CUG)₅₁ et 5'-

hcTNT₅₀₋₀. Des valeurs de Kd de la protéine MBNL1_{ΔCT3} allant de 600 à 1200 nM et de 170 à 500 nM ont été estimées pour les ARN des familles I et II, respectivement. Pour comparaison, les valeurs de Kd estimées pour les ARN (CUG)₅₁ et 5'-hcTNT₅₀₋₀ étaient 420 nM et 525 nM. Nous avons ainsi pu montrer que MBNL1 a une meilleure affinité pour les ARN issus de la famille II que pour les ARN issus de la famille I, mais également que pour les ARN (CUG)₅₁ et 5'-hcTNT₅₀₋₀. Nous avons ensuite réalisé une étude détaillée des déterminants moléculaires à l'origine de la très forte affinité de MBNL1 pour les ARN de la famille II. Les données obtenues montrent que la protéine MBNL1 est capable, contrairement à ce qui avait été parfois décrit dans la littérature, de reconnaître avec une très forte affinité un ARN ne contenant qu'un seul motif YGCY, localisé au sein d'une boucle terminale. La nature des résidus encadrant ce motif ainsi que la stabilité de l'hélice apicale modulent l'efficacité de reconnaissance par MBNL1.

Ces résultats font l'objet de la rédaction d'un manuscrit qui sera prochainement soumis et dont je suis 1^{er} auteur.

PUBLICATION 1

**A peculiar stem-loop structure with strong affinity for the
MBNL1 splicing factor**

Audrey Vautrin, Christelle Aigueperse, Isabelle Behm-Ansmant, and Christiane Branlant

Manuscrit en préparation

A peculiar stem-loop structure with strong affinity for the MBNL1 splicing factor

Audrey Vautrin, Christelle Aigueperse, Isabelle Behm-Ansmant[§], and Christiane Branlant[§]

ARN, RNP, Structure-fonction-maturation, Enzymologie Moléculaire et Structurale (AREMS), Nancy Université, UMR7214 CNRS-UHP 1, 9 Avenue de la Forêt de Haye, Batiment Biopôle, Faculté de Médecine, BP 70184, 54506 Vandoeuvre-lès-Nancy, Cedex, France.

§ To whom correspondence should be addressed

isabelle.behm@maem.uhp-nancy.fr

christiane.branlant@maem.uhp-nancy.fr

Summary (247 words)

Myotonic Dystrophies are caused by toxic RNAs transcribed from two genes with CTG or CCTG repeats. The resulting RNAs thus contain UGCY sequences which are bound by the MBNL1 protein and consequently MBNL1 is sequestered in nuclear foci where it cannot perform its usual functions as an alternative pre-mRNA splicing factor. Controlling these diseases will require a deeper understanding of the exact nature of MBNL1's RNA-binding specificity. Previous reports suggest the preferred substrate to be one or preferably several YGCY motifs in a pyrimidine-rich environment while the additional contribution of secondary structure is controversial. We performed an in vitro selection Systematic Evolution of Ligands by EXponential enrichment (SELEX) from a random pool of RNA 18mers using a C-terminal truncated recombinant MBNL1 protein and identified 50 potential target sequences, all of which contained at least one YGCY motif. In addition a subset of 16 of the sequences shared a short stem of at least four base pairs coupled to an invariant loop of 8 nucleotides and an exceptionally high affinity for MBNL1 (<300nM); this new class of RNAs bound at least three times more strongly than other known targets of MBNL1. Collectively our results confirm previous reports of a strict requirement for YGCY repeats but also show that secondary structure and specifically the presentation of the YGCY in a loop can also contribute to MBNL1 binding specificity. The novel high affinity stem loop will likely lead to new insights into the RNA binding specificity of muscleblind proteins and to new therapeutic approaches for Myotonic Dystrophies.

Keywords: Myotonic Dystrophy, alternative splicing, Muscleblind-like, triplet repeat expansion disease, SELEX, RNA secondary structure

Introduction

Splicing of pre-mRNA is a key post-transcriptional step in eukaryotic gene expression. Indeed, 94% of human genes contain introns that are removed from pre-mRNAs to produce mature mRNAs.¹ Alternative splicing, the process by which different 5' and 3' splice sites can be joined together allows individual genes to express multiple mRNAs that encode proteins with diverse and sometimes antagonistic functions. The majority of human genes generate multiple mRNAs by alternative splicing, and 80% of alternative splicing events modify the sequence of the encoded protein; this is thus a major source of human proteome diversity.^{2,3} The regulation of alternative splicing plays a major role in cell differentiation and in development, and depends on the activity of numerous splicing regulatory factors that are differentially expressed during development, and in different types of tissue. Defects in these alternative splicing processes can contribute to pathogenesis, as demonstrated for a growing number of diseases, including neuromuscular diseases such as myotonic dystrophies of type 1 and 2 (DM1 and DM2).^{4,5}

DM1 and DM2 are autosomal dominant disorders caused by an unstable expansion of CTG repeats in the 3'-untranslated region (UTR) of the *DMPK* gene and of CCTG repeats in the first intron of the *ZNF9* gene, respectively.⁶⁻⁸ Pathogenesis results from a toxic RNA gain-of-function mechanism of transcripts bearing long CUG- or CCUG-repeats. The unusual A-form double-stranded RNA structures formed by these repeats sequester RNA-binding proteins, sepecially Muscleblind-like 1 (MBNL1) in nuclear foci.^{9, 10} Consequent loss of MBNL1 activity results in mis-regulation of at least 200 alternative splicing events.^{11, 12} These mis-regulated events include human cardiac Troponin T (hcTNT) exon 5, insulin receptor (IR) exon 11 and exon 7a of the muscle-specific chloride channel (ClC1) and result in specific features of the disease.¹³ Importantly MBNL1 knockout mice recapitulated the DM1 phenotype, consistent with the MBNL1 loss-of-function mechanism proposed to explain DM1

pathogenesis.¹⁴ MBNL1 was shown to participate in post-natal remodeling of skeletal muscles and of the developing heart in both mice and human, by controlling the developmentally regulated switch of a key set of pre-mRNAs by both inhibition or enhancement of different exons.^{16-18,15} Recently MBNL1 was also found to act in the cytoplasm as a regulator of pre-miR biogenesis.³³ All three human paralogs (MBNL1, MBNL2 and MBNL3) contain four CCCH zinc-finger (ZnF1-4) motifs, which are structured in pairs and function as RNA-binding domains.^{19, 20} In agreement with the MBNL1 loss-of-function mechanism, MBNL1 has been shown to bind on both pathogenic CUG- and CCUG-repeats and its intended pre-mRNA targets.^{15, 20, 21} One important question concerning the mode of interaction of MBNL1 with RNAs is to know whether MBNL1 has a similar or a different mode of interaction with its authentic target sites in pre-mRNAs than it does with CUG or CCUG repeats. The first in-depth study of MBNL1's binding to RNA was carried out by Kino et al.²² The affinity for RNAs containing 5'-CHG-3' and 5'-CHHG-3' repeats (where H is A, U or C) was first demonstrated by a yeast 3 hybrid assay.²² Interestingly while MBNL1 does not interact with a genuine double-stranded RNA comprising CAG/CUG repeats, it specifically recognizes the unusual A-form double-stranded RNA structure formed by CUG repeats.²² In this structure pairs or triplets of canonical Watson-Crick base-pairs alternate with mismatched pairs.⁹ Concerning its endogenous pre-mRNA targets, the alignment of the four MBNL1-binding sites described in chicken and human cTNT first revealed the YGCU(U/G)Y consensus sequence (where Y is U or C.¹⁵) In the crystal structure established for MBNL1 ZnF3/4 bound to single-stranded r(CGCUGU) each ZnF motif targets one GC/U sites in the RNA, thereby inducing a chain-reversal trajectory in the bound RNA.²³ However so far no 3D-structure of complexes formed between MBNL1 and CUG repeats has been reported. Recently, the YGCU(U/G)Y consensus established for MBNL1 binding sites has been refined to the YGCU motif with a preference for UGCU.¹³ As many of the MBNL1

pre-mRNA targets contain several YGCY sequences in close vicinity, an important question is to know whether these close repeated sequences can form, like the CHG repeats, imperfect A-type helices. Some data obtained by studying a short piece of the hcTNT pre-mRNA that contains MBNL1 binding sites suggested that this may be the case.^{24, 21} More recently a doped SELEX experiment starting from the same short piece of hcTNT pre-mRNA confirmed that high-affinity MBNL1 sites are pyrimidine-rich and contain multiple copies of the YGCY motif.¹³ It also showed that the RNAs containing zero or one MBNL1 motif do not bind MBNL1 with high affinity and based on these SELEX data it was proposed that no specific RNA structure is required for MBNL1 binding.¹³ Rau and colleagues also observed low-affinity MBNL1 binding to RNAs containing only one YGCY motif. But in this study the MBNL1-binding site was in a single-stranded structure since it was located within the loop of pre-miR-1.³³ Therefore the current view is that MBNL1 binding on RNAs requires the presence of several YGCY motifs in a pyrimidine-rich environment. However, no clear information is available on the contribution of RNA secondary structure to the efficiency of MBNL1 binding to RNA.

In the present study, for a better understanding of MBNL1 RNA-binding properties, we used a SELEX approach which already turned to be highly efficient to define the 2D and sequence parameters which favor binding of a protein to its RNA targets.^{26, 34} This SELEX approach is based on the use of a library of degenerate RNA stem-loop structures with a large terminal loop (18 nt-long) whose sequence is randomized. As the full length recombinant MBNL1 protein is highly insoluble we used a C-terminally truncated version (MBNL1_{ΔCT3}) which is soluble and has been shown to have similar affinity for RNAs as full length MBNL1.²² In agreement with data from the literature all the selected RNAs had pyrimidine-rich sequences and contained at least one YGCY motif. An exceptionally high affinity subset of sixteen of the selected RNAs contained a 16 nt-long stem loop NNYGUUUGCGCUUGNN where the

NNYG formed the head of a stem with the final UGNN. We characterised a key guanine residue directly upstream of the YGCY sequence in the loop and integrity of the stem, by site directed mutagenesis, *in vitro* binding and *in vitro* splicing assays, as being key determinants both for MBNL1 binding and for its action as an alternative splicing factor *in vitro*. The new class of exceptionally high-affinity MBNL1 binding sequences will be an invaluable experimental tool that could lead to novel agents to fight muscular dystrophy.

Results

The SELEX strategy used to study the RNA-binding properties of MBNL1

Because recombinant full length MBNL1 is mainly insoluble we used a truncated recombinant protein (MBNL1 Δ CT3) lacking just the C-terminus but still containing the four Zn finger RNA-binding regions intact. GST -tagged MBNL1 Δ CT3 was produced in *E. coli* as described in Materials and Methods and fixed on glutathione-sepharose beads in order to select RNAs having a high affinity for this protein. A random pool of RNA 18mers presented within a constant RNA stem-loop structure derived from yeast U3 snoRNA was produced as previously described.^{26, 34} Briefly, a DNA oligonucleotide randomized at 18 positions (SELEX N18 primer, see Materials and Methods) was used as a template to produce the double-stranded DNA product by PCR amplification, allowing the production of the library of degenerate RNA stem-loop with a terminal 18 nt-long randomized loop sequence by *in vitro* transcription (Fig. 1A and 1B). Assuming a perfectly statistical synthesis of the degenerated SELEX N18 DNA primer, the expected diversity of the starting RNA sequences was of 6.9×10^{10} molecules such that, each possible RNA sequence was expected to be present 870 times. The evenness of the distribution of sequences in the starting material was verified by sequence analysis (Supplementary Fig. S1). To eliminate RNA molecules having an affinity for the glutathione-sepharose beads or for the GST protein, 4 μ g of the RNA mixture were first incubated with 30 μ l of beads and then with 30 μ l of beads coated with 0.02 nmol of GST protein. The unbound RNAs were recovered and 2 μ g of these RNAs (6×10^{13} molecules) were used to perform the first cycle of selection using a molar [RNA]/[protein] ratio of 10. EMSA assays revealed a strong increase of the affinity of MBNL1 Δ CT3 for the selected RNAs after four rounds of amplification-selection (Fig. 1C). The amplified cDNAs were cloned in plasmid pCR2.1 used to transform *E. coli* cells and 37 clones (named AC clones for Affinity Chromatography clones) were arbitrary selected among the obtained clones. In order to gain

in specificity an additional selection was performed; a fraction of the amplified cDNAs from round four was used as a template for *in vitro* transcription and the obtained transcripts were used to perform an EMSA assay with the MBNL1 Δ CT3 protein. The RNA-MBNL1 Δ CT3 complexes were isolated from the gel and the cDNAs amplified from these complexes were cloned into plasmid pCR2.1. Thirty clones (named E clones for Electrophoresis clones) were arbitrary selected among the obtained clones.

Isolation of MBNL1-binding RNAs

Thirty-three and 24 different RNA sequences were obtained when analysing the 37 AC clones and the 30 E clones, respectively. Among these sequences 5 were common to both batches of clones leading to the final identification of 52 different RNA sequences (Fig. 1D). Fifty of these sequences contained at least one YGCY motif, previously described as being preferentially recognized by MBNL1, which validated our SELEX experiment.^{13, 11} As previously observed by others a significant enrichment in pyrimidines (71%) was detected in the random part of these 50 RNA sequences.¹³ The two sequences that did not contain a YGCY were the only two sequences that we tested that did not bind to MBNL1 therefore these are false positives (FPs).

Among the 50 potential MBNL1-binding selected RNAs, 34 closely resembled the canonical MBNL1-binding site found by other studies, therefore we refer to these henceforth as class I RNAs (Fig. 1D). All but two of these class I clones contained a YGCU motif in the final or penultimate position. Two thirds of the pyrimidines in position one of the motif were uridines and one third cytidines. All but two of these clones contained a further YGCY motif upstream. This coupling of YGCY motifs is entirely consistent with previous literature and validates our experimental approach.¹³ However the positioning in the final position of the randomized region places the final YGCU motif adjacent upstream of a region of base pairing formed in

the construct and this is likely a factor in MBNL1-specificity (see results section below and discussion).

A new class of MBNL1 binding sequences

The remaining 16 out of the 52 RNA sequences that we will further designate as class II RNAs contained a 12 nt-long sequence with the following consensus YGUUUGCGCUUG, (the YGCY MBNL1 motif is underlined). Class II RNAs represented one third of the isolated sequences after four rounds of traditional SELEX (the AC clones) and they were further enriched to constitute two thirds of the further gel purified population (the E clones) (Fig. 1D) suggesting a higher affinity of MBNL1 for class II RNAs.

The affinity of the MBNL1 Δ CT3 protein for the different classes of selected RNAs was tested by EMSA. Class I RNAs had dissociation constants (Kd) ranging from 600 nM to 1200 nM. In contrast Class II RNAs had affinities in the range 170-500 nM with an average Kd value of 311 nM for the 10 tested RNAs (Fig. 1D and supplementary Fig. S2). Representative examples of gels obtained for class I and II RNAs are illustrated in Figure 2 and supplementary Fig. S3 As controls we used two physiological MBNL1 targets, a 50 nt-long fragment of the hcTNT pre-mRNA located upstream from exon 5 (Kd = 525 nM) and an RNA containing 51 CUG repeats (CUG₅₁ Kd = 420 nM) (Fig. 2C). Several RNA/MBNL1 Δ CT3 complexes could be detected for all class I RNAs that contain multiple MBNL1 binding motifs (Fig. 2A). The class I RNAs containing only one MBNL1-binding motif showed lower affinity consistent with previous reports (Fig. 1D).¹³ Concerning the two FP RNAs that did not contain MBNL1 binding motifs, no complex formation was detected by EMSA (Fig. 1D and supplementary Fig. S3).

Class II RNAs fold into a stable stem-loop structure

As our goal was to identify sequence and structural motifs with high affinity for MBNL1, we first checked whether the RNAs from class II can adopt a defined secondary structure by using the M-Fold software (website: <http://mfold.rit.albany.edu/?q=mfold>).²⁶ All the RNAs from this class were predicted to fold into a stable structure consisting of two stacked stems (proximal and distal), a lateral loop and a terminal loop, the degenerate sequence being involved in proximal stem, terminal loop and lateral loop formation. Within class II RNAs the size of the proximal stem and of the lateral loop are variable (Fig. 3B). Remarkably, RNAs AC3, E5, E19 and E25 contained nucleotide insertions allowing the extension of the proximal stem by one to three base-pairs at the expense of the lateral loop. In RNA AC3 the lateral loop was even reduced to a single bulged residue due to the insertion in the proximal stem of three additional base-pairs (Fig. 3B). This observation reinforces the idea of the functional importance of the selected secondary structure for MBNL1 binding (Fig. 1D and 3B).

Most of the structures proposed in Figure 3B were experimentally verified by enzymatic probing using 5'-end labeled transcripts. T1 and T2 RNases identify single-stranded regions, whereas V1 RNase identifies double stranded or stacked RNA regions. An example of cleavage pattern is given for RNA E10 (Fig. 4A). The data obtained for the naked E10 RNA (Fig. 4A, lanes 2, 5, 8 and 11) allowed the building of a secondary structure model (Fig. 4B) that is in good agreement with the M-Fold predicted structure. Indeed, the strong reactivity of residues U22 to G26 and the slight reactivity of residues C27 and U28 towards RNases T1 and/or T2 confirmed the existence of the terminal loop. In the same way the reactivity of residues U10 to A18, U31 and G34 towards RNase V1 confirmed stems 1 and 2 formation. The reactivity of residues U21 towards RNase V1 and U29 towards both RNases V1 and T2 confirmed their localization at a junction between single- and double-stranded RNA regions. No information concerning the existence of the lateral loop was obtained due to extensive

cleavage of the transcript at the level of the lateral loop residues even in the absence of RNases. Evaluation of the free energies of the entire 2D structures of the class II RNAs (at 37°C, in 1M NaCl) by M-Fold predicted that they range from -12.1 to -20.0 Kcal/mole (Fig. 3B). No correlation was observed between the K_d values and the stability of the proposed structures, suggesting that not only the stability of the structure is important for MBNL1 binding.

MBNL1 binds to the terminal loop of class II RNAs

In order to define the localization of the MBNL1-binding site within class II RNAs, we performed enzymatic footprinting experiments using increasing concentrations (1 μM and 2 μM) of the recombinant MBNL1_{ΔCT3} protein. The concentrations of MBNL1_{ΔCT3} used for the footprinting assays were chosen based on EMSA assays: the 1 μM concentration allowed total complex formation for all the tested class II RNAs (Fig. 2). Footprinting experiments were done for most of the class II RNAs and the results obtained for RNA E10 are illustrated in Figure 4A. A strong protection of residues U22 to G26 against T2 or T1 RNase cleavages was observed, indicating that MBNL1 binds to the terminal loop of RNA E10 (Fig. 4A, lanes 8-13 and Fig. 4B). A less pronounced protection of residues A18, C27, U28, U29 against RNase V1 (Fig. 4A, lanes 5-7) and of U31 against RNase T2 (Fig. 4A, lanes 11-13) was also observed suggesting that MBNL1 protects almost the whole terminal loop and the two last base-pairs of the proximal stem (Fig. 4B). Interestingly in the presence of MBNL1_{ΔCT3} residue U21 becomes more reactive to RNase V1 than in the naked RNA, suggesting a partial remodeling of the architecture of the terminal loop when MBNL1 is bound (Fig. 4A, lanes 5-7). Similar results were obtained for the other analyzed class II RNAs (supplementary Fig. S3 and data not shown), confirming the binding of MBNL1 to the terminal loop of class II RNAs.

The size of the variable loop and proximal stem stability do not significantly influence MBNL1 binding efficiency.

In order to identify the primary determinants for MBNL1 binding we first compare the sequence and the structure of all class II RNAs and looked for some variations that could be correlated to variations of MBNL1-binding efficiency. Although sharing similar global 2D-structure, class II RNAs differed by the length of proximal stem and the size of the lateral loop. We looked for a correlation between the size of the lateral loop and the established Kd values, but no direct correlation was observed (Fig. 5A). Indeed, even if the presence of a lateral loop seemed to confer a slight advantage for MBNL1 binding (see RNA AC3), the increase of the size of this loop was not correlated to a better Kd value (Fig. 5A). We also evaluated the free energy of the stem-loop structure containing only the proximal stem and the terminal loop using the M-fold software. Here again no correlation was detected, indicating that neither the size of the variable loop, nor the stability of the proximal stem-terminal loop structure were primary determinants for MBNL1-binding (Fig. 5B)

The 2 G residues in the terminal loop are required for efficient MBNL1 binding

As mentioned in the introduction, the crystal structure of MBNL1 ZnF3/4 bound to r(CGCU) establishes that both ZnF3 and ZnF4 target the GC dinucleotide within the YGCY motif.²³ The terminal loop of class II RNAs contained a single YGCY motif (CGCU) as part of a larger motif containing two GC motifs (GCGCU). To identify the nucleotides of the terminal loop required for MBNL1 binding, we substituted the G residue of each of the GC dinucleotides with an A residue (RNAs AC3m1 and AC3m2) (Fig. 6A). The affinity of MBNL1 $_{\Delta CT3}$ for these two variant RNAs was tested by EMSA and their Kds were estimated. As expected mutation of the G residue of the YGCY motif (G₂₉, RNA AC3m1) completely abolished MBNL1 $_{\Delta CT3}$ binding (Fig. 6A). More surprisingly, the mutation of the G residue

located just upstream from the YGCY motif (G₂₇, RNA AC3m2) also greatly reduced the affinity of MBNL1 Δ CT3. Indeed binding of MBNL1 Δ CT3 on RNA AC3m2 could still be detected but the K_d value was higher than 1800 nM and could not be measured (Fig. 6A and data not shown). Therefore we concluded that in addition to the essential G residue of the YGCY motif, the upstream G₂₇ residue located in the terminal loop is required for MBNL1 efficient binding.

Importance of the U triplet located in the terminal loop

According to our footprinting experiments most of the phosphodiester bonds connecting residues in the terminal loop of RNA AC3 were protected against T1 and T2 RNase cleavage in the presence of MBNL1 Δ CT3. In contrast, cleavage of the phosphodiester bond between residues G20 and U21 was reinforced suggesting a reinforcement of the stability of the G·U pair at the extremity of the structure. To check for the contribution of the U triplet to MBNL1 binding efficiency, we mutated the U triplet of RNA AC 3 to CCC (RNA AC3m3), UCC (RNA AC3m4), CCU (RNA AC3m5) or CUC (RNA AC3m6) triplets (Fig. 6A). The variant RNAs were tested for their capability to form complexes with the MBNL1 Δ CT3 protein. None of the mutations completely abolished MBNL1 binding, however MBNL1 affinity was strongly reduced and the estimated K_ds were expected to be higher than 1800 nM for all the variants (Fig. 6A). The mutations in RNA variants AC3m3, AC3m5 and AC3m6 were expected to change the architecture of the terminal loop. Therefore, we could not distinguish between the sequence and structural effects. On the contrary, for RNA AC3m4 the mutation was not predicted to affect the secondary structure and we can thus conclude from this mutant that residues U25 and U26 (i.e. the two U residues in the sequence UUGC₂₅GC₂₆U, directly upstream of the essential G residue and following YGCY motif) are required for efficient binding of MBNL1.

Importance of the single-stranded context of the MBNL1 motif

As mentioned in the introduction, it was proposed that MBNL1 recognizes several neighboring YGCY sequences that could form imperfect A-type helices.²² The MBNL1 motif in class II RNAs are all located in a single stranded RNA segment. To check whether the structural context of the MBNL1 motif might interfere with MBNL1 binding efficiency, we prepared a variant of the MBNL1 binding motif (AC3M) where the terminal loop was disrupted by a G to A substitution in each of the GC dinucleotides and a MBNL1 binding motif corresponding to the sequence of the terminal loop UUUGCGCU was inserted in the stem (Fig. 6A). According to EMSA data MBNL1 Δ CT3 was not able to form a complex with RNA AC3M. No complex formation was detected, reinforcing the idea of a needed single stranded state of the sequence UUUGCGCU for efficient MBNL1 binding..

Importance of the identity of the base-pair closing the proximal stem

All the class II RNAs except two, RNAs AC12 and E4, had a G·U pair at the apical position in the proximal stem. RNA AC12 and E4 had U·U and C·U mismatches, respectively (Fig. 3). Interestingly, these two RNAs had the lowest affinity for MBNL1 of the class II RNAs tested with estimated Kds of 500 nM and 441 nM, respectively (Fig. 3). To test for the importance of the presence of residues G23 and U32 in RNA AC3, we designed several RNA AC3 variants. RNA AC3m7 contained a G23A substitution which transformed the G·U pair into a more stable A·U base-pair, but preserved the U at position 32 (Fig. 6B). RNA AC3m8 contained a U32C substitution which transforms the G·U base-pair into a more stable G·C pair while preserving the G residue at position 23 (Fig. 6B). RNA AC3m9 contained the two base substitution G23U and U32G substitutions which reversed the G·U base-pair into an equally stable U·G pair (Fig. 6B). Based on EMSA results, all the variant RNAs were able to weakly

form complexes with MBNL1 and only the affinity for RNA AC3m7 was sufficient to be quantified (Kd value 475 nM, Fig. 6B). This Kd value was higher than for RNA AC3 (360 nM), but was in the same range as the Kds obtained for RNA AC12 and E4, that both contained a U residue at the position equivalent to position 32 in RNA AC3. Altogether these data suggest that one determinant for efficient binding of MBNL1 to class I RNA AC3 is the presence of a U residue after the loop at the end of the proximal stem. In addition the stability of the proximal stem last base-pair influences MBNL1 binding efficiency but is not a primary determinant.

As the common 12 nt-long sequence in class II RNAs also included the second last base-pair of the proximal stem, we looked for the importance of the identity of the residues at these positions (positions 19 and 30) in RNAs. Interestingly, RNAs E18, E10 and E22 are almost identical and only differ one with the other by the identity of the second last base-pair of proximal stem: U·G for RNA E18, C·G for RNA E10 and U·A for RNA E22 (Fig. 5A). According to the established Kds these RNAs had very similar affinities for MBNL1_{ΔCT3} (170 nM for RNA E18, 210 nM for RNA E10 and 240 nM for RNA E22), indicating that the stability of the second last base-pair in proximal stem has little if any influence on MBNL1 binding efficiency.

In vitro, class II RNA addition mis-regulates MBNL1-regulated splicing events

We observed that class II RNAs have higher affinity than the natural hcTNT RNA which corresponds to the 50-nucleotide region upstream of hcTNT exon 5. One important question concerning our SELEX RNAs was to know whether they can compete for MBNL1 binding with endogenous pre-mRNAs in splicing conditions. To answer to this question we established an *in vitro* splicing system where we could vary both the SELEX RNA and MBNL1 concentrations. We chose the hcTNT as a splicing model as MBNL1 negatively

regulates the inclusion of exon 5 in cell lines.^{15, 21} We first prepared a capped *in vitro* transcript containing hcTNT exon 5 and bordering intronic sequences. The DNA template used for *in vitro* transcription was amplified from an hcTNT minigene in which the hcTNT sequences were cloned into intron 2 of the rat insulin gene. We first checked that this transcript was efficiently spliced *in vitro* and that exon 5 inclusion rate was correlated to the amount of MBNL1 protein used in the assay (Fig. 7A). The MBNL1 protein used corresponded to a GST-MBNL1 fusion protein of the 40 kDa MBNL1 isoform. GST protein alone was used as a control. Exon 5 inclusion was monitored by RT-PCR analysis. We observed that in the absence of the recombinant GST-MBNL1 protein, exon 5 was included at a rate of 61 %, whereas the addition of increasing amounts of GST-MBNL1 protein led to a decrease of exon 5 inclusion to a rate of 37 % (Fig. 7A). By contrast the addition of increasing amounts of GST protein did not lead to a decrease of exon 5 inclusion (Fig. 7A). We concluded from these data that our *in vitro* hcTNT exon 5 splicing system was functional and responsive to MBNL1 quantity and thus suitable to test for the effect of SELEX RNA addition on splicing efficiency. We next tested this effect by adding increasing amounts (0; 0.1; 0.5 and 1 µg) of RNAs AC3, AC3m1 or E18 in the absence or in the presence of 0.2 µg of the GST-MBNL1 protein (Fig. 7B). We observed that the addition of increasing amounts of RNAs AC3 and E18 increased significantly exon 5 inclusion rate, whereas the addition of RNA AC3m1 which has lost its ability to bind MBNL1 had no effect on exon 5 inclusion (Fig. 7B). This result indicates that the presence of a class II RNA able to bind MBNL1 counteracts the inhibitory effect of MBNL1 on hcTNT exon 5 inclusion. Moreover the addition of 0.2 µg of the GST-MBNL1 protein in the presence of 1 µg of RNAs AC3 and E18 only slightly decreased the exon 5 inclusion rate (Fig. 7B). We even noticed that the extent of the decrease was inversely correlated to the affinity of MBNL1 for the RNAs: 12.9 % decrease for RNA AC3 which has a K_d of 360 nM and 7.3 % of decrease for RNA E18 which has a K_d of 170

nM (Fig. 7B). The initial hcTNT exon 5 inclusion rate of 55.7% observed when adding 0.2 μ g of GST-MBNL1 in the absence of SELEX RNA was only reached when using the RNA AC3m1 (Fig. 7B). Altogether these results indicate that our class II RNAs can efficiently compete with a pre-mRNA target for MBNL1 binding and due to their higher affinity deplete MBNL1 in a splicing assay.

Discussion

Writing in progress

Experimental procedures

Cloning and protein purification- The plasmids pGEX-6P1-MBNL1₄₀ and pGEX-6P1-MBNL1_{ΔCT3} were constructed for the overexpression of recombinant GST-MBNL1₄₀ and GST-MBNL1_{ΔCT3} proteins in *E. coli* as described before.²⁷ They were used to transform *E. coli* C43 cells (Invitrogen). Recombinant GST-MBNL1₄₀ and GST-MBNL1_{ΔCT3} proteins were purified from *E. coli* C43 cells under native conditions, using glutathione-Sepharose 4B as recommended by the manufacturer (GE Healthcare). Briefly, protein expression was induced with 0.1 mM IPTG at an OD₆₀₀ ~0.8. Cells were homogenized in cell lysis buffer (600 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄ at pH 6.8) at 100 DO/ml followed by sonication. Cell extract was centrifuged for 20 min at 20,000 rpm, and the lysate containing the GST-proteins was collected. The supernatants were treated with 0.0125% polyethyleneimine to remove nucleic acids. GST-MBNL proteins were bound to glutathione-Sepharose 4B (GE Healthcare) for 2h at 4°C, washed three times with buffer (1 M NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄ at pH 6.8) and eluted with 10 mM reduced glutathione, diluted to 50 mM Tris-HCl pH 8.0. The protein was subsequently fractionated by Superdex75 (150 mM KCl, 1.5 mM MgCl₂, HEPES-KOH 20 mM at pH 7.9, 20 % glycérol (v/v)) at a flow rate of 1 ml min⁻¹. For purification of untagged proteins, the GST-proteins bound to glutathione-Sepharose 4B beads were cleaved on the beads using 80 units of PreScission protease (GE Healthcare)/ ml of glutathione-Sepharose bead suspension. Cleavage was carried out overnight at 4 °C. The protein was subsequently fractionated by Superdex75 (150 mM KCl, 4.5 mM MgCl₂, 20 mM HEPES at pH 7.9, 20 % glycerol) at a flow rate of 1 ml.min⁻¹. The purified proteins were aliquoted and stored at -80°C.

SELEX experiment- The starting DNA matrix containing a 18-nt long degenerated sequence was produced by PCR amplification using two partially complementary oligonucleotides: SELEX N18 with a 18 nt-long degenerated sequence 5'-CTGCCATTTGTACCCATAAN18 TCTGGGGTACAAAGGTCCTATAGTGAGTCGTATTA and SELEX-5' 5'-TAATACGACTCACTATAGGACCTTTGTACCCAG-3' that generates a T7 RNA polymerase promoter. PCR amplification was as previously described, except that MgCl₂ was added at a 4 mM concentration in the incubation buffer.²⁸ About 500 ng of amplified DNA was used for *in vitro* transcription with T7 RNA polymerase.²⁸ Transcripts were purified by electrophoresis on a 6% denaturing polyacrylamide gel as in Mougín *et al.*²⁹ 0.2 nmoles of transcripts were used for the first round of selection. To eliminate RNA molecules having an affinity for the glutathione-Sepharose beads (Amersham) or GST protein, 0.2 nmol of transcript were first incubated 2 h at 4°C with 30 µl of beads in the absence of the protein followed by an incubation of 30 min at 4°C with 30 µl of beads in the presence of 0.02 nmol of the GST-MBNL1 Δ CT3. For RNP complexes, 0.1 nmol of treated RNAs were incubated with 0.01 nmol of purified GST-MBNL1 Δ CT3 for 30 min at 4°C, in 30 µl of buffer D, in the presence of 2 µg of a yeast tRNA mixture (Roche). The mixture was then incubated with 15 µl of beads equilibrated in buffer D (150 mM KCl, 4.5 mM MgCl₂, 20 mM HEPES at pH 7.9, 10 % glycerol). After extensive washing with buffer D, the selected RNAs were released by a 30 min incubation at 37°C with 20 µg of proteinase K in buffer D-NP40 0.01 %. They were extracted with a phenol–chloroform mixture, ethanol precipitated, dissolved in sterile water, hybridized with 50 pmol of SELEX-3' primer (5'-CTGCCATTTGTACCCA-3') and finally reverse-transcribed with 5 U of AMV Reverse transcriptase (Q.Biogene) for 30 min at 42°C. Next, 30 cycles of PCR amplification were performed in the presence of primers SELEX-5' and SELEX-3' (50 pmol each). The amplified DNA fragments were gel purified and used as the matrix for *in vitro* transcription. At each cycle of the SELEX experiment, an EMSA assay

was performed after incubation of the 5' end-labeled transcripts produced from the DNA pool with the MBNL1 Δ CT3 protein. At the fourth cycle of the amplification-selection experiment, DNA fragments were cloned into plasmid pCR2.1 (Invitrogen). Plasmids were prepared from 37 randomly selected clones and sequenced by the dideoxysequencing method. For sequences S, 0.1 nmol of treated RNAs from cycle 4 were incubated with 0.01 nmol of purified MBNL1 Δ CT3 for 30 min at 4°C, in 30 μ l of buffer D, in the presence of 2 μ g of a yeast tRNA mixture (Roche). The binding reaction was loaded onto an 8 % nondenaturing polyacrylamide gel containing 44.5 mM borate, 1 mM EDTA, 44.5 mM Tris borate, pH 8.3, and 0.5 % glycerol. Gels were run for 45 min at 15 mA. The RNA/protein complexes were extracted and purified with a phenol–chloroform mixture, ethanol precipitated and dissolved in sterile water. The obtained RNAs were hybridized with 50 pmol of SELEX-3' primer, reverse-transcribed with 5 U of AMV Reverse transcriptase (Q.Biogene) for 30 min at 42°C and amplified by PCR in the presence of primers SELEX-5' and SELEX-3' (50 pmol each). The amplified DNA fragments were gel purified and cloned into plasmid pCR2.1. Plasmids were prepared from 30 randomly selected clones and sequenced by the dideoxysequencing method.

Recombinant plasmids- The SELEX DNA variants were obtained by PCR-directed mutagenesis using the oligonucleotides pairs given in Table I and the pCR2.1 SELEX plasmids as templates. Plasmid pUC18-T7-CTG₅₁ used for *in vitro* production of CUG₅₁ RNA was generated by insertion of the *XbaI/BamHI* fragment of T7 and *SmaI/SmaI* inserted of CTG₅₁ repeats into plasmid pUC18 cleaved by the same enzymes. The T7 fragment was obtained by hybridization of complementary oligonucleotides. The CTG repeats were obtained by the method of Takahashi et al.³⁰ To create the pSVIRB-hcTNT plasmid, we inserted an *EcoRI/EcoRI* PCR fragment (corresponding to hcTNT exon 5 and 150 nucleotides

of flanking introns) into a unique *EcoRI* site located within the rabbit insulin receptor intron 2 from the pSVIRB plasmid.³¹

In vitro transcription and labeling- The SELEX matrices used for *in vitro* transcription were obtained by PCR amplification using the SELEX-5' and SELEX-3' primers and plasmids SELEX or oligonucleotides as templates. For CUG₅₁ RNA production, a DNA fragment containing a T7 promoter and the CUG₅₁ coding region was PCR amplified from plasmid pUC18-T7-CUG₅₁ using oligonucleotides O-4633 and O-4634 (Table I). Templates for hcTNT transcripts production by *in vitro* transcription were produced by PCR amplification using the pSVIRB-hcTNT plasmid as template and oligonucleotides O-6133 and O-4390 for the hcTNT50 transcript or oligonucleotides O-7254 and O-7255 for the hcTNT splicing substrate (Table I). For EMSA and RNA secondary structure analysis, RNAs were synthesized by *in vitro* transcription using 500 ng of amplified DNA as the template as previously described (²⁸). Two µg of gel purified RNAs were dephosphorylated and 20 pmol were 5' end-labeled using 10 units of T4 polynucleotide kinase (MBI-Fermentas) and 5 pmol of [γ -³²P] ATP, in a 10 µl reaction mixture containing 10 mM MgCl₂; 5 mM DTT; 0.1 mM spermidine; 0.1 mM EDTA; 50 mM Tris-HCl pH 7.6 at 37°C. The 5' end- labeled RNAs were gel purified on 8 % denaturing polyacrylamide gels.

For *in vitro* splicing assays, the hcTNT splicing substrate was incubated for 2 h in 40 µl of transcription mixture containing 6 mM MgCl₂, 2 mM spermidine, 10 mM NaCl, 40 mM Tris HCl (pH 7.9), 10 mM DTT, 0.1 mg/ml BSA, 40 U of RNasin, 0.5 mM GpppG, 0.5 mM UTP, ATP and CTP, 0.2 mM GTP and 10 units of T7 RNA polymerase. After digestion of the DNA matrix with 7.5 U of RQ1 DNase (Promega), RNAs were purified using the GenElute kit (Sigma-Aldrich) according to manufacturer's recommendations.

Electrophoresis Mobility Shift Assays- For complex formation, 1.2 fmol of 5' end-labeled RNA were preincubated in 7 μ l of buffer D for 10 min at 65 °C followed by slow cooling. Then, 4 μ l of MBNL1 Δ CT3 recombinant protein were added to the renatured RNA for a final protein concentration of 0 μ M to 1.8 μ M. The binding reactions were incubated for 30 min at room temperature and loaded onto an 8 % or 5 % (38/2) nondenaturing polyacrylamide gel containing 44.5 mM borate, 1 mM EDTA, 44.5 mM Tris borate, pH 8.3, and 0.5 % glycerol. Gels were run for 45 min at 15 mA. The amount of radioactivity in the bands, corresponding to free and bound RNAs, was estimated using a Phosphor-Imager and the ImageQuant software. Using these values, apparent K_d values were determined using the SigmaPlot software (SPSS Science Software).

Structure Probing and Footprinting- 5' end-labeled RNAs (25 fmol) were pre-incubated in buffer D for 10 min at 65°C, followed by a slow cooling for renaturation in the presence of 10 μ M of tRNA. MgCl₂ was added when starting the cooling in order to reach a final MgCl₂ concentration of 10 μ M. The renatured RNAs were then incubated for 30 min at room temperature in the absence or the presence of MBNL1 Δ CT3 (10 and 20 pmol) in 10 μ l of buffer D. Digestion was performed for 6 min at 20°C, in the presence of 1 U of T1 RNase (Roche), 0.5 U of T2 RNase (Mobitech) or 0.0005 U of V1 RNase (Kemotex). RNase digestions were stopped by addition of 100mM EDTA and 20 μ g of tRNA followed by phenol extraction and ethanol precipitation. For production of a ladder, an alkaline hydrolysis of the naked RNA was performed for 3 min at 96°C, using 10 fmol of RNA dissolved in 1 ml of 100mM sodium bicarbonate pH 10.4. For production of a T1 ladder, 5' end-labeled RNAs (10 fmol) were pre-incubated in buffer citrate (20 mM citrate NaOH, 1 mM EDTA, 7 M urea, 0.05 % Xylene cyanol, 0.05 % bromophenol blue) for 10 min at 65°C, followed by incubation 10 min at 65°C with 2 U of T1 RNase. The cleavage products were fractionated by electrophoresis on a 10%

polyacrylamide–8M urea gel. The gel was dried and exposed to a phosphoimager screen overnight at 25°C. The free energies of the 2D structures of the selected RNAs were calculated at 378C and in 1M NaCl with the M-fold software.²⁶

In Vitro Splicing- Twenty fmol of purified hcTNT splicing substrate were incubated for 2 h at 30°C in splicing solution containing 53 mM KCl, 2.8 mM MgCl₂, 1 mM ATP, 25 mM creatine phosphate, 3.5% (w/v) polyvinyl alcohol, 10.5% (v/v) glycerol, 79 mM Hepes (pH 7.9), 10 U of RNasin, 8 µl of HeLa cell nuclear extract and 0 to 1 µg of SELEX RNA. To study the influence of the MBNL1 protein on splicing, some splicing reactions were performed in the presence of 100 ng to 2 µg of purified recombinant GST-MBNL1 protein or 2.5 µl of buffer D. Splicing reactions were stopped by addition of 20 µg of Proteinase K, followed by phenol/chloroform extraction. RNAs were ethanol precipitated and dissolved in 10 µl of bi-distilled H₂O. Reverse transcriptions were performed by addition of 2 µl of the splicing products in 25 µl of MMLV buffer (Promega) containing 5 pmol of O-7255 primer, 1 mM of each dNTP and 200 U of MMLV reverse transcriptase (Promega). Incubation was performed for 1 h at 42°C. The splicing products were detected by PCR amplification on 2 µl of the RT products using the oligonucleotides O-5991 and O-8078 and the DreamTaq polymerase (Fermentas) in the conditions recommended by the furnisher. The following protocol was used: initial denaturation for 5 min at 94°C, followed by 30 cycles of amplification. Each cycle included incubation for 30 sec at 94°C, 30 sec at 55°C and 30 sec at 72°C. After agarose gel electrophoresis, gel red fluorescence of RT–PCR products was quantified using a digital ccd camera (GeneGenius, SynGene) and the SynGene GeneTools version 3.08 software.

Acknowledgments

F. Oillo-Blanloeil is acknowledged for her excellent technical assistance. J. Venables is thanked for careful reading of the manuscript. This work was financially supported by the Association Française contre les Myopathies (AFM 14269, 15047), the EURASNET EU contract (EURASNET, FP6 life sciences, genomics and biotechnology for health), the Centre National de la Recherche Scientifique (CNRS), the French Ministry for Youth, National Education and Research, and the Lorraine Region. A. V. was a fellow of the French Ministry for Youth, National Education and Research. C.A., C. B. and I.B-A are Staff Scientists at the Centre National de la Recherche Scientifique (CNRS).

Disclosure: The authors have no conflict of interest to report.

References

1. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 2008; 14:802-13; 18369186; rna.876308 [pii] 10.1261/rna.876308.
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001; 409:860-921; 11237011
3. Modrek B, Lee C. A genomic view of alternative splicing. *Nat Genet* 2002; 30:13-9; 11753382
4. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev* 2003; 17:419-37; 12600935
5. Black DL, Graveley BR. Splicing bioinformatics to biology. *Genome Biol* 2006; 7:317; 16732900; gb-2006-7-5-317 [pii] 10.1186/gb-2006-7-5-317.
6. Pistoni M, Ghigna C, Gabellini D. Alternative splicing and muscular dystrophy. *RNA Biol*; 7:441-52; 20603608; 12258 [pii].
7. Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 1992; 68:799-808; 1310900
8. Fu YH, Pizzuti A, Fenwick RG, Jr., King J, Rajnarayan S, Dunne PW, et al. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 1992; 255:1256-8; 1546326
9. Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, et al. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* 1992; 255:1253-5; 1546325

10. Mooers BH, Logue JS, Berglund JA. The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc Natl Acad Sci U S A* 2005; 102:16626-31; 16269545
11. Fardaei M, Larkin K, Brook JD, Hamshere MG. In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res* 2001; 29:2766-71; 11433021
12. Du H, Cline MS, Osborne RJ, Tuttle DL, Clark TA, Donohue JP, et al. Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat Struct Mol Biol* 2010; 17:187-93; 20098426; nsmb.1720 [pii] 10.1038/nsmb.1720.
13. Goers ES, Purcell J, Voelker RB, Gates DP, Berglund JA. MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res* 2010; 38:2467-84; 20071745; gkp1209 [pii] 10.1093/nar/gkp1209.
14. Wheeler TM, Thornton CA. Myotonic dystrophy: RNA-mediated muscle disease. *Curr Opin Neurol* 2007; 20:572-6; 17885447; 10.1097/WCO.0b013e3282ef606400019052-200710000-00011 [pii].
15. Kanadia RN, Johnstone KA, Mankodi A, Lungu C, Thornton CA, Esson D, et al. A muscleblind knockout model for myotonic dystrophy. *Science* 2003; 302:1978-80; 14671308
16. Ho TH, Charlet BN, Poulos MG, Singh G, Swanson MS, Cooper TA. Muscleblind proteins regulate alternative splicing. *Embo J* 2004; 23:3103-12; 15257297
17. Lin X, Miller JW, Mankodi A, Kanadia RN, Yuan Y, Moxley RT, et al. Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum Mol Genet* 2006; 15:2087-97; 16717059; ddl132 [pii] 10.1093/hmg/ddl132.
18. Kalsotra A, Xiao X, Ward AJ, Castle JC, Johnson JM, Burge CB, et al. A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart.

Proc Natl Acad Sci U S A 2008; 105:20333-8; 19075228; 0809045105 [pii] 10.1073/pnas.0809045105.

19. Terenzi F, Ladd AN. Conserved developmental alternative splicing of muscleblind-like (MBNL) transcripts regulates MBNL localization and activity. RNA Biol 2010; 7:43-55; 20009516; 10401 [pii].

20. Miller JW, Urbinati CR, Teng-Ummuay P, Stenberg MG, Byrne BJ, Thornton CA, et al. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. Embo J 2000; 19:4439-48; 10970838

21. Fardaei M, Rogers MT, Thorpe HM, Larkin K, Hamshere MG, Harper PS, et al. Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. Hum Mol Genet 2002; 11:805-14; 11929853

22. Warf MB, Berglund JA. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. RNA 2007; 13:2238-51; 17942744; rna.610607 [pii] 10.1261/rna.610607.

23. Kino Y, Mori D, Oma Y, Takeshita Y, Sasagawa N, Ishiura S. Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. Hum Mol Genet 2004; 13:495-507; 14722159

24. Teplova M, Patel DJ. Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. Nat Struct Mol Biol 2008; 15:1343-51; 19043415; nsmb.1519 [pii] 10.1038/nsmb.1519.

25. Yuan Y, Compton SA, Sobczak K, Stenberg MG, Thornton CA, Griffith JD, et al. Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. Nucleic Acids Res 2007; 35:5474-86; 17702765; gkm601 [pii] 10.1093/nar/gkm601.

26. Clery A, Bourguignon-Igel V, Allmang C, Krol A, Branlant C. An improved definition of the RNA-binding specificity of SECIS-binding protein 2, an essential component of the selenocysteine incorporation machinery. *Nucleic Acids Res* 2007; 35:1868-84; 17332014
27. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003; 31:3406-15; 12824337
28. Tran H, Gourrier N, Lemercier-Neuillet C, Dhaenens CM, Vautrin A, Fernandez-Gomez FJ, et al. Analysis of exonic regions involved in nuclear localization, splicing activity, and dimerization of Muscleblind-like-1 isoforms. *J Biol Chem* 2011; 286:16435-46; 21454535; M110.194928 [pii] 10.1074/jbc.M110.194928.
29. Marmier-Gourrier N, Clery A, Senty-Segault V, Charpentier B, Schlotter F, Leclerc F, et al. A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA* 2003; 9:821-38; 12810916
30. Mougin A, Gottschalk A, Fabrizio P, Luhrmann R, Branlant C. Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles. *J Mol Biol* 2002; 317:631-49; 11955014; 10.1006/jmbi.2002.5451 S0022283602954513 [pii].
31. Takahashi N, Ishiura S. [DM pathogenesis and expansion of CTG trinucleotide repeat within 3' untranslated region of DMPK gene]. *Nihon Rinsho* 1999; 57:937-42; 10222792
32. Andreadis A, Nisson PE, Kosik KS, Watkins PC. The exon trapping assay partly discriminates against alternatively spliced exons. *Nucleic Acids Res* 1993; 21:2217-21; 7684835
33. Rau F, Freyermuth F, Fugier C, Villemin JP, Fischer MC, Jost B, et al. Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat Struct Mol Biol* 2011; 18:840-45; 21685920

34. Clery A, Senty-Segault V, Leclerc F, Raué HA, Branlant C Analysis of sequence and structural features that identify the B/C motif of U3 small nucleolar RNA as the recognition site for the Snu13p-Rrp9p protein pair. *Mol Cell Biol* 2007; 4:1191-206; 17145781

Figure legends:

Figure 1. Sequences of the RNAs recovered from the SELEX experiment.

(A) The RNA template used to create the initial pool (pool 0) of RNA for SELEX. The grey nucleotides correspond to the constant region, while the central black 18 bases correspond to the variable region.

(B) The secondary structure model proposed for the degenerated RNA.²⁶

(C) Round 4 SELEX RNAs bind significantly tighter than Round 0 RNAs. Illustration of the EMSA data obtained when testing the binding of MBNL1 Δ CT3 on Round 0 RNAs and Round 4 RNAs.

(D) Alignment of the 52 selected RNAs. Selected RNAs were divided into three classes on the basis of their sequence. Class I RNA sequences contain 1 to 4 YGCY motifs but do not display any other obvious common features. Class II RNA sequences contain a 12 nt-long conserved sequence, the YGUUUGCGCUUG sequence, which includes the YGCY MBNL1 motif (grey box). Class FP RNA sequences did not contain MBNL1 motif. The total number of clones obtained for each sequence is indicated on the left with mention in brackets of the number of E clones. The apparent Kd values calculated from EMSA data are indicated on the right of the panel.

Figure 2. The affinity of MBNL1 Δ CT3 for class II RNAs was higher than the one for class I RNAs and natural MBNL1 targets such as the CUG₅₁ and hcTNT RNAs.

Ribonucleoprotein complexes were formed by incubating 1.2 fmol of 5'-labeled RNAs with MBNL1 Δ CT3 recombinant protein at a concentration of 56, 113, 225, 450, 900 or 1800 nM. Incubation conditions were as described in the Materials and Methods. Lanes (-) are control experiments done in the absence of protein. Complexes were fractionated by electrophoresis

in a non-denaturing 5% or 8% polyacrylamide gel. The apparent K_d values calculated for the illustrated experiment are indicated.

Figure 3. Class II RNAs are predicted to adopt a conserved secondary structure corresponding to a long highly stable stem-loop structure with a lateral loop.

Secondary structure prediction for each of the class II RNAs is presented. The apparent K_d values established from EMSA data are indicated. The free energy of the proposed secondary structures, expressed in kcal/mol, was calculated by using the M-Fold software. Nucleotide insertions allowing the extension of the helix length are indicated in green.

Figure 4. MBNL1 binds the terminal loop of the class II RNAs.

(A) Footprinting analysis of the RNA E10-MBNL1_{ΔCT3} complexes. Complexes were formed as described in Materials and Methods, by using 5'-end labeled S10 RNA and recombinant MBNL1_{ΔCT3} protein. Free (-) and complexed RNAs were digested with RNases V1 (lanes 5-7), T1 (lanes 8-10), and T2 (lanes 11-13) under conditions described in Materials and Methods. As a control, undigested RNA was fractionated in parallel (lanes 2-4). The protein concentrations used in the assays are indicated on the top of the gel. The digestion products were fractionated by electrophoresis on a 10 % polyacrylamide gel in the presence of 8 M urea. Lane T1, corresponding to RNase T1 digestion in denaturing conditions, was used for localization of the cleavage sites. Position of G residues is indicated on the left of the autoradiogram. Nucleotides with change sensitivity to RNases in the presence of the MBNL1_{ΔCT3} protein are indicated on the right.

(B) Secondary structure model proposed for SELEX RNA E10. The model was proposed based on thermodynamic considerations and on the results of enzymatic digestions shown in A. V1, T1, and T2 RNase cleavages are represented by arrows surmounted by squares, dots,

and triangles, respectively. Red, orange and green symbols indicate a strong, medium or low cleavage, respectively. Nucleotides with decreased sensitivity to RNases in the presence of the MBNL1 Δ CT3 are represented with blue symbols whereas nucleotides with increased sensitivity to RNases in the presence of MBNL1 Δ CT3 are indicated by pink symbols.

Figure 5. Contribution of the size of the lateral loop and the stability of the proximal-stem/terminal loop structure to MBNL1 binding.

(A) Although favorable the lateral loop is not essential for MBNL1 binding. The residues of the lateral loop are indicated in blue. The name of SELEX RNAs is indicated on the top of each M-fold predicted structure. The apparent Kd values established from EMSA data are indicated.

(B) Although favorable the stability of the proximal-stem/terminal loop structure is not a primary determinant for MBNL1 binding. The Kd in nM of MBNL1 for the class II RNAs is represented as a function of the free energy of the proximal-stem/terminal loop structure.

Figure 6. Importance of the terminal loop residues, the single-stranded structural context for MBNL1 binding and the identity of the two last base-pairs of the proximal stem.

(A) Importance of the terminal loop residues. Different variants of RNA AC3 were obtained by mutating either individually or in combination residues 24 to 29 of the terminal loop. The mutated residues are numbered and indicated in red. The M-Fold predicted secondary structure and the apparent Kd values established from EMSA data are indicated. Y indicates the capability of a variant RNA to form complexes with MBNL1 Δ CT3, whereas N indicates the absence of complex detection. In RNA AC3M the MBNL1 binding site of the terminal loop has been mutated and a MBNL1 binding motif has been inserted in the stem.. The free energy

of the proposed secondary structure, expressed in kcal/mol, was calculated by using the M-Fold software.

(B) The stability of the apical base-pair of the proximal stem influences MBNL1 binding efficiency and the presence of the U32 residue is a limiting parameter. The residues in red indicate mutated residues. The first base-pair of the proximal stem is circled in blue.

(B) The identity of the second last base-pair of the proximal stem has no influence on MBNL1 binding efficiency. The residues involved in the formation of second last base-pair are indicated in blue and the base-pair is circled in blue.

Figure 7. Class II RNA addition mis-regulates MBNL1-regulated splicing events

(A) The GST-MBNL1 recombinant protein inhibits hcTNT exon 5 inclusion in a dose-dependent manner in an *in vitro* splicing system. The hcTNT splicing substrate prepared as indicated in Material and Methods was incubated for 2 h in a mixture of nuclear extracts in the absence or the presence of purified recombinant human GST-MBNL1 or GST proteins (100, 200, 500, 1.000 or 2.000 ng). The splicing products were analyzed by RT-PCR. The amplification products were separated on low melting agarose gel at 4 % and gel red fluorescence of RT-PCR products was quantified. The percentage of exon 5 inclusion has been calculated as follows: $((\text{mRNA} + \text{exon}) / ((\text{mRNA} + \text{exon}) + (\text{mRNA} - \text{exon}))) \times 100$. Results for GST-MBNL1 protein are derived from at least three independent experiments. Data are expressed as means \pm S.E. (error bars). An unpaired t test with two tails was used to compare values. Statistically significant results are indicated by * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).

(B) Addition of a SELEX RNA able to bind MBNL1 counteracts the inhibitory effect of MBNL1 on hcTNT exon 5 inclusion. The hcTNT splicing substrate was incubated for 2 h in a mixture of nuclear extracts in the absence or the presence of 200 ng of purified recombinant

human GST-MBNL1 protein and 0.1, 0.5 or 1 μ g of SELEX RNA AC3, AC3m1 or E18. The experimental procedure and the analysis of the data were performed as described in Material and Methods and Panel A.

Table I : Name, sequence and usage of the oligonucleotides used for this study

Oligonucleotide name	Oligonucleotide sequence	Use
SELEX N18	5'-GGACCUUUGUACCCCAGA(N20)AUGGGUACAAAUGGCAG-3'	PCR
SELEX 5'	5'-TAATACGACTCACTATAGGACCTTTGTACCCCAG-3'	PCR
SELEX 3'	5'-CTGCCATTTGTACCCA-3'	PCR
3 mut 1 G29>A	5'-CCAGACGTTGTTTGCACCTTGATGTTATGGG-3' 5'-CCCATAACATCAAGTGCAAACAACGTCTGG-3'	MUT
3 mut 2 G27>A	5'-CCCCAGACGTTGTTTACGCTTGATGTTATGG-3' 5'-CCATAACATCAAGCGTAAACAACGTCTGGGG-3'	MUT
3 mut 3 UUU>CCC	5'-CTGCCATTTGTACCCATAACATCAAGCGCGGGCAACG-3'	PCR
3 mut 4 UUU>UCC	5'-CTGCCATTTGTACCCATAACATCAAGCGCGGACAACG-3'	PCR
3 mut 5 UUU>CCU	5'-CTGCCATTTGTACCCATAACATCAAGCGCAGGCAACG-3'	PCR
3 mut 6 UUU>CUC	5'-CTGCCATTTGTACCCATAACATCAAGCGCGAGCAACG-3'	PCR
3 mut 7 G23>A	5'-CCTTTGTACCCAGACGTTATTTGCGCTTGATGTTATGG-3' 5'-CCATAACATCAAGCGCAAATAACGTCTGGGGTACAAAGG-3'	MUT
3 mut 8 U32>C	5'-CTGCCATTTGTACCCATAACATCGAGCGC-3'	PCR
3 mut 9 G23>U U32>G	5'-CTGCCATTTGTACCCATAACATCCAGCGCAAAAACG-3'	PCR
3M	5'-CACTGGCGTAAGAGGGTACAAATGGCAG-3' 5'-TAAAAGCGCAAAGGGGTACAAAGGTCC-3'	MUT
O-4633	5'- TAATACGACTCACTATAGGGGCAGGTCGACTCTAGAG-3'	PCR
O-4634	5'- CTATGACATGATTACGAATTC-3'	PCR
O-7254	5'- TAATACGACTCACTATAGGGCCTTTGTGGTTCTCACTTGG-3'	PCR
O-7255	5'- GCGCTTCTGCCGGGCCACCT-3'	RT-PCR
O-6133	5'- TAATACGACTCACTATAGGGACTAACCTGTCTC-3'	PCR
O-4390	5'- GTTCAGGGAGTGGCC-3'	PCR
O-5991	5'- CCTGGTGTGTGGGGAGCGTGG-3'	PCR
O-8078	5'- TCACCTGCCCCGGGCCTCC-3'	PCR

MUT : mutagenesis

PCR : PCR amplification

RT : reverse transcription

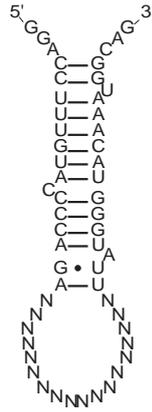
A.

5' constant region random region 3' constant region

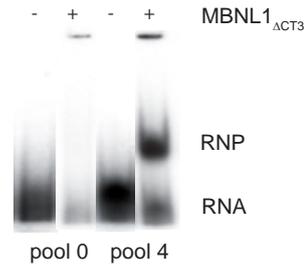
10 20 30 40 50

N18 5' - GGACCUUUGUACCCCAGA NNNNNNNNNNNNNNNNNNNN UUAUGGGUACAAUUGGCAG - 3'

B.



C.



D.

	5' constant region	random region	3' constant region	Kd (nM)
Class I	AC.31. 5' - CCCCAGA	UUCCCCUACACUCGC	UUAUGGG - 3'	> 1800
	E.31. 5' - CCCCAGA	UUUCCCCUACUGCACGCU	UUAUGGG - 3'	
	AC.11. 5' - CCCCAGA	CUGCCCGCGGUCCGCGC	UUAUGGG - 3'	
	AC.23. 5' - CCCCAGA	UCUUGCUGAUCUAUGCGC	UUAUGGG - 3'	
	AC.40. 5' - CCCCAGA	UCAUGCCGCGAUCGACGC	UUAUGGG - 3'	877 ± 29
	E.32. 5' - CCCCAGA	CCUUGCUUGGUCGACGCGC	UUAUGGG - 3'	> 1800
	AC.7. 5' - CCCCAGA	UCAUUGCUGAUCACGCGC	UUAUGGG - 3'	
	AC.21. 5' - CCCCAGA	CACGCCGCUUCCCUUUGC	UUAUGGG - 3'	
	AC.36. 5' - CCCCAGA	CCGCGCGCUUGUCUAUGC	UUAUGGG - 3'	
	AC.27. 5' - CCCCAGA	CUACCGCCGCGCAUUGCU	UUAUGGG - 3'	720 ± 269
	AC.37. 5' - CCCCAGA	CGCCGCGCGUUCUAUGC	UUAUGGG - 3'	
	E.12. 5' - CCCCAGA	ACUGCUUGCUUCCUCGCG	UUAUGGG - 3'	668 ± 263
	AC.17. 5' - CCCCAGA	CCGCGCGCGUUCUAUGC	UUAUGGG - 3'	
	AC.30. 5' - CCCCAGA	CGCCGCGCGUUCUUGCU	UUAUGGG - 3'	
	AC.26. 5' - CCCCAGA	UUGCCGCGCGUUAUAUGC	UUAUGGG - 3'	
	E.17. 5' - CCCCAGA	CCGCGCGCGUACCUUGCU	UUAUGGG - 3'	1050 ± 247
	E.7. 5' - CCCCAGA	UCAUCGCGCGUUCUCCGC	UUAUGGG - 3'	
	E.14. 5' - CCCCAGA	CCGCGCGUUGCUUAUUGC	UUAUGGG - 3'	
	AC.8. 5' - CCCCAGA	ACC CGC ACGCUUAUUGC	UUAUGGG - 3'	
	E.37. 5' - CCCCAGA	UACCGCCACGCUUCACGC	UUAUGGG - 3'	
	E.26. 5' - CCCCAGA	CCUCGCGCGCUUACUGCU	UUAUGGG - 3'	
	E.27. 5' - CCCCAGA	CACGCCGCGUUCUUGC	UUAUGGG - 3'	
	AC.32. 5' - CCCCAGA	CACCGCCACGCUCCCGCU	UUAUGGG - 3'	1110 ± 417
	E.3. 5' - CCCCAGA	UACCGCGCUUGCUUUGC	UUAUGGG - 3'	
	AC.12. 5' - CCCCAGA	ACCCCGCACGCUUUGC	UUAUGGG - 3'	
	AC.14. 5' - CCCCAGA	UUCCCCUACCGCUCCCGC	UUAUGGG - 3'	> 1800
	E.8* 5' - CCCCAGA	CCCGCUCCGUCGUUUGC	UUAUGGG - 3'	687 ± 250
	AC.34. 5' - CCCCAGA	UCACGCGCGUUGCUUUGC	UUAUGGG - 3'	
	E.21. 5' - CCCCAGA	UCGCGCGCUCGCUUUGC	UUAUGGG - 3'	765 ± 308
	AC.25. 5' - CCCCAGA	UUAACGCGCGUUGCUUUGC	UUAUGGG - 3'	
	AC.38. 5' - CCCCAGA	UCCCAACGCUUGCUUUGC	UUAUGGG - 3'	
	AC.10. 5' - CCCCAGA	AUCCCUACCGCUAUGC	UUAUGGG - 3'	
	AC.28. 5' - CCCCAGA	GGUCCAUGCUAAGCUUUA	UUAUGGG - 3'	
	AC.35. 5' - CCCCAGA	AUGUACCUGCGCUACCCA	UUAUGGG - 3'	> 1800
	Class II	number of clones		
2 (1)		AC.3. 5' - CCCCAGA CGU	UGUUUUGCGCUUG AUG	UUAUGGG - 3' 360 ± 76
		E.19. 5' - CCCCAGA A	UGUUUUGCGCUUG U UCUG	UUAUGGG - 3' 292 ± 80
		E.25. 5' - CCCCAGA U	UGUUUUGCGCUUG A UCCU	UUAUGGG - 3' 300 ± 22
		E.5. 5' - CCCCAGA G	UGUUUUGCGCUUG C UCUA	UUAUGGG - 3' 315 ± 73
7 (4)		E.18. 5' - CCCCAGA	UGUUUUGCGCUUG UCUGAA	UUAUGGG - 3' 170 ± 42
		E.10. 5' - CCCCAGA	CGUUUUGCGCUUG UCUGAA	UUAUGGG - 3' 210 ± 35
		E.22. 5' - CCCCAGA	UGUUUUGCGCUUA UCUGAC	UUAUGGG - 3' 240 ± 59
4 (2)		E.36. 5' - CCCCAGA	CGUUUUGCGCUUG UCUCU	UUAUGGG - 3' 282 ± 91
		E.4. 5' - CCCCAGA	UCUUUUGCGCUUG UCUCU	UUAUGGG - 3' 441 ± 120
		AC.12. 5' - CCCCAGA	CUUUUUGCGCUUG UCUCU	UUAUGGG - 3' 500 ± 136
2		AC.4. 5' - CCCCAGA	CGUUUUGCGCUUG UCUCU	UUAUGGG - 3'
2	AC.18. 5' - CCCCAGA	CGUUUUGCGCUUG CCUGAU	UUAUGGG - 3'	
3 (2)	AC.29. 5' - CCCCAGA	UGUUUUGCGCUUG UCUGUA	UUAUGGG - 3'	
2 (2)	E.2. 5' - CCCCAGA	UGUUUUGCGCUUG UCUGAU	UUAUGGG - 3'	
	E.13. 5' - CCCCAGA	UGUUUUGCGCUUG UCUGGU	UUAUGGG - 3'	
	E.35. 5' - CCCCAGA	CGUUUUGCGCUUG UCUUGA	UUAUGGG - 3'	
		YGUUUUGCGCUUG		
Class FP	AC.16. 5' - CCCCAGA	GAACUCGGUAGCGGACCG	UUAUGGG - 3'	No binding
	AC.39. 5' - CCCCAGA	CUCGGUAGCGGACCGUAA	UUAUGGG - 3'	No binding

Figure 1 Vautrin et al.

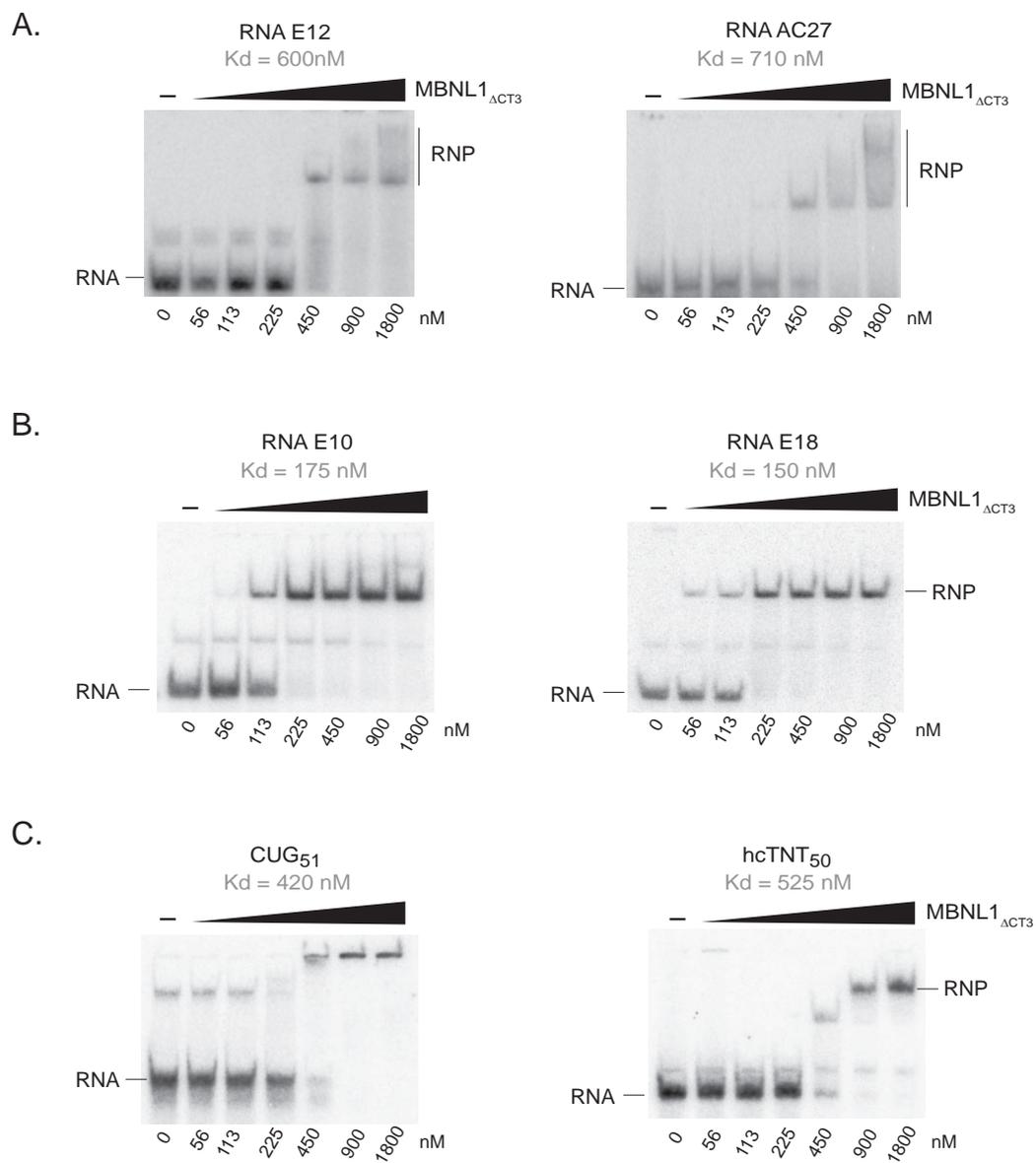


Figure 2 Vautrin et al.

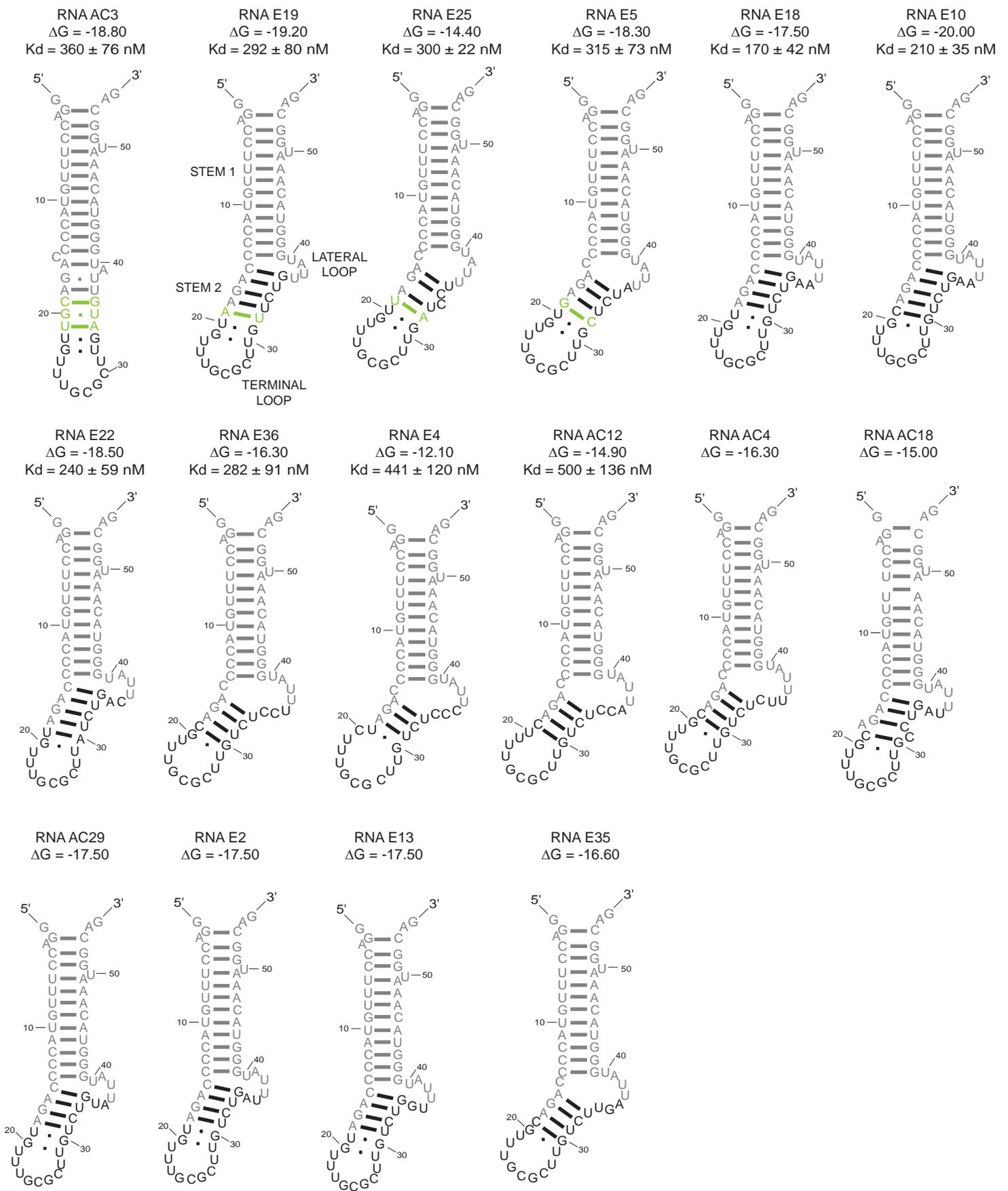


Figure 3 Vautrin et al.

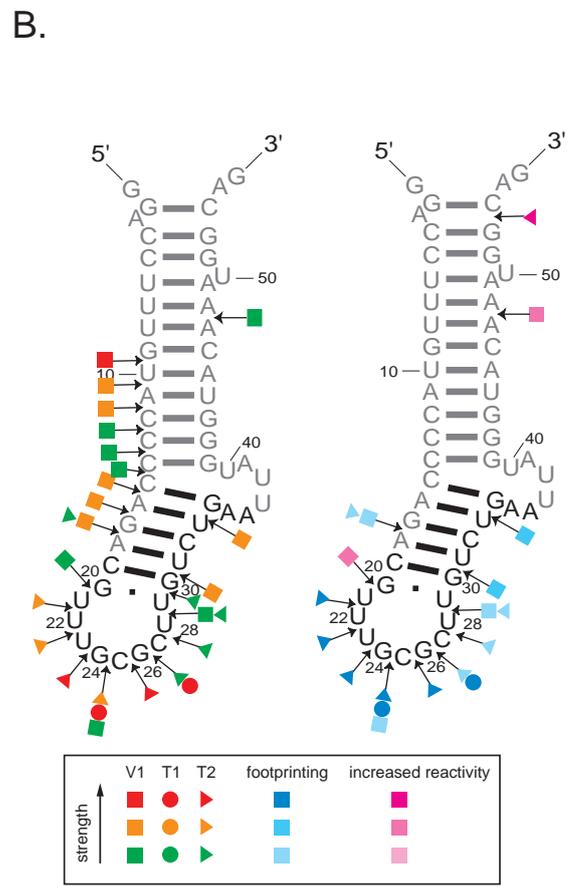
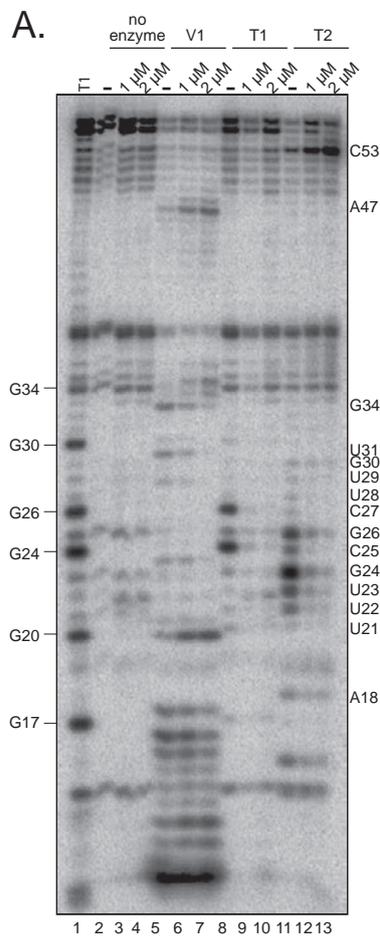
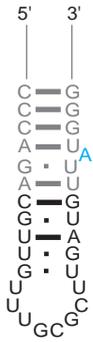


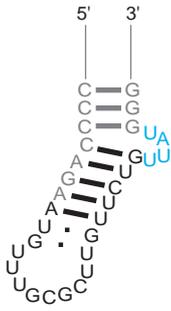
Figure 4 Vautrin et al.

A.

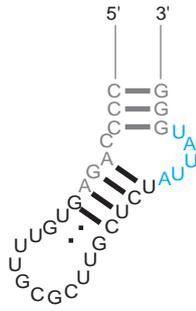
RNA AC3
Kd = 360 ± 76 nM



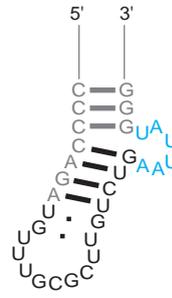
RNA E19
Kd = 292 ± 80 nM



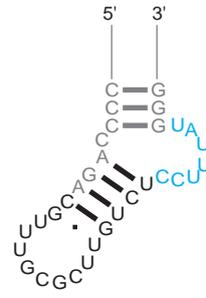
RNA E5
Kd = 315 ± 73 nM



RNA E18
Kd = 170 ± 42 nM



RNA E36
Kd = 282 ± 91 nM



B.

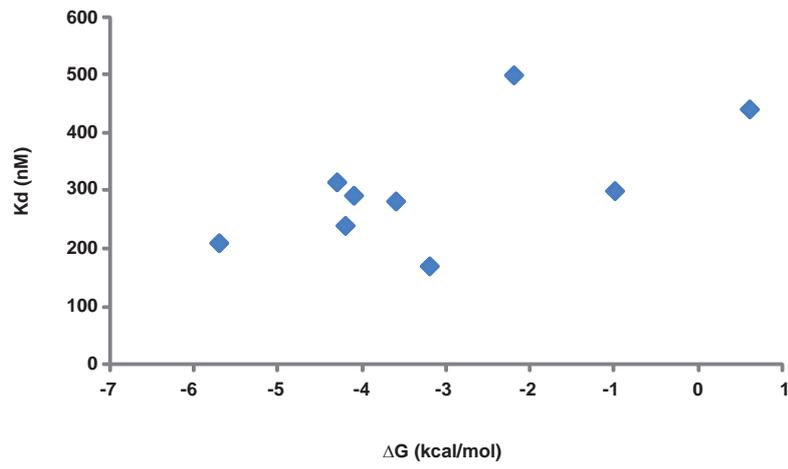
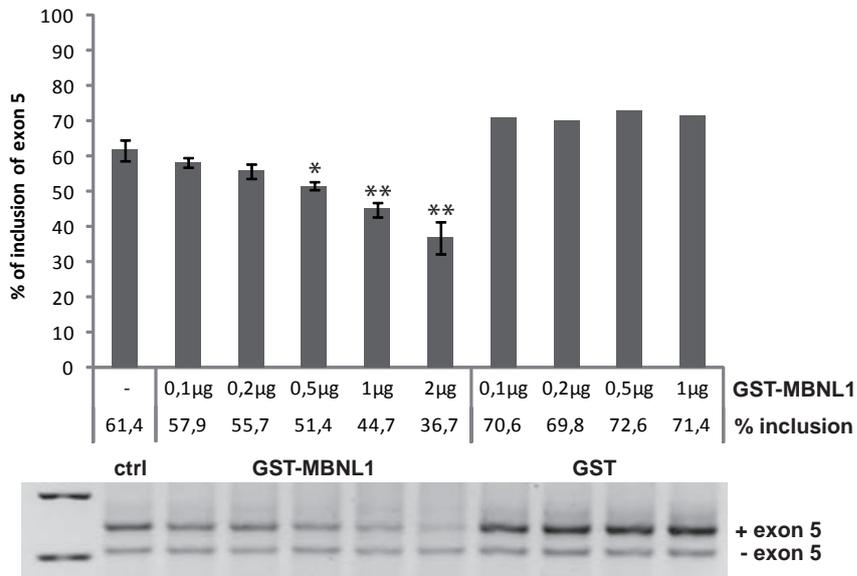


Figure 5 Vautrin et al.

A.



B.

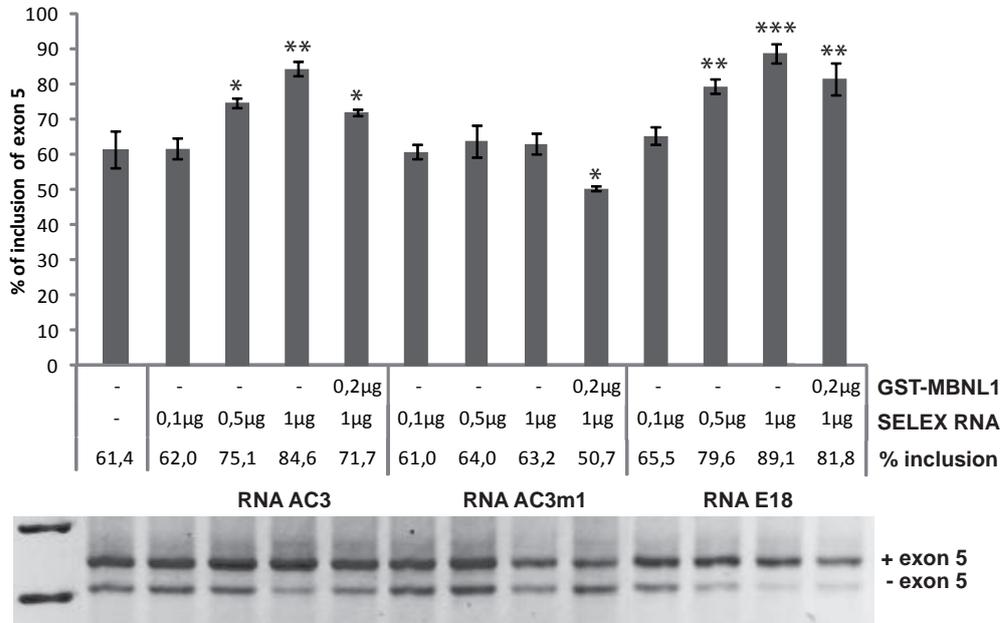


Figure 7 Vautrin et al.

Supplementary Figure legends:

Figure S1. Sequence analysis of the starting material of the SELEX experiment.

Sequence analysis was done by reverse transcription by using oligonucleotide SELEX-3' (Table I, in supplementary data) and a dideoxynTP mixture. RNA yU3A, from which the sequence was degenerated, was sequenced as a control. The nucleotide sequence of yU3A is indicated on the left side of the panel.

Figure S2. Illustration of the EMSA experiments performed on SELEX RNAs.

Ribonucleoprotein complexes were formed by incubating 1.2 fmol of 5'-labeled RNAs with MBNL1 Δ CT3 recombinant protein at a concentration of 56, 113, 225, 450, 900 or 1800 nM. Incubation conditions were as described in the Materials and Methods. Lanes (-) are control experiments done in the absence of protein. Complexes were fractionated by electrophoresis in a non-denaturing 5% or 8% polyacrylamide gel. The name of the RNA used for the EMSA experiment is indicated on the top of the gel. The apparent K_d values calculated for the illustrated experiment are indicated.

Figure S3. The affinity of MBNL1 Δ CT3 for class II RNAs was higher than the one for class I, FP and control RNAs.

The distribution of the estimated average K_d value for class I, class II and FP SELEX RNAs and control (CUG₅₁ and hcTNT) RNAs is represented.

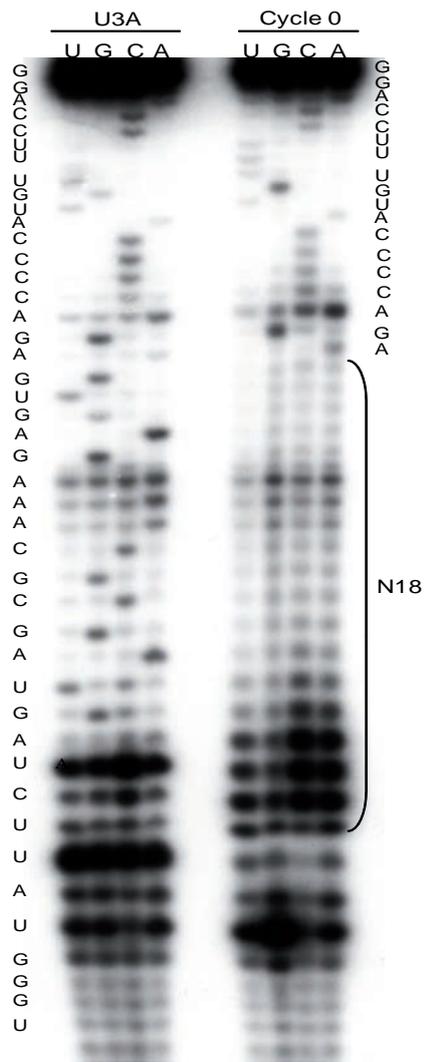


Figure S1 Vautrin et al.

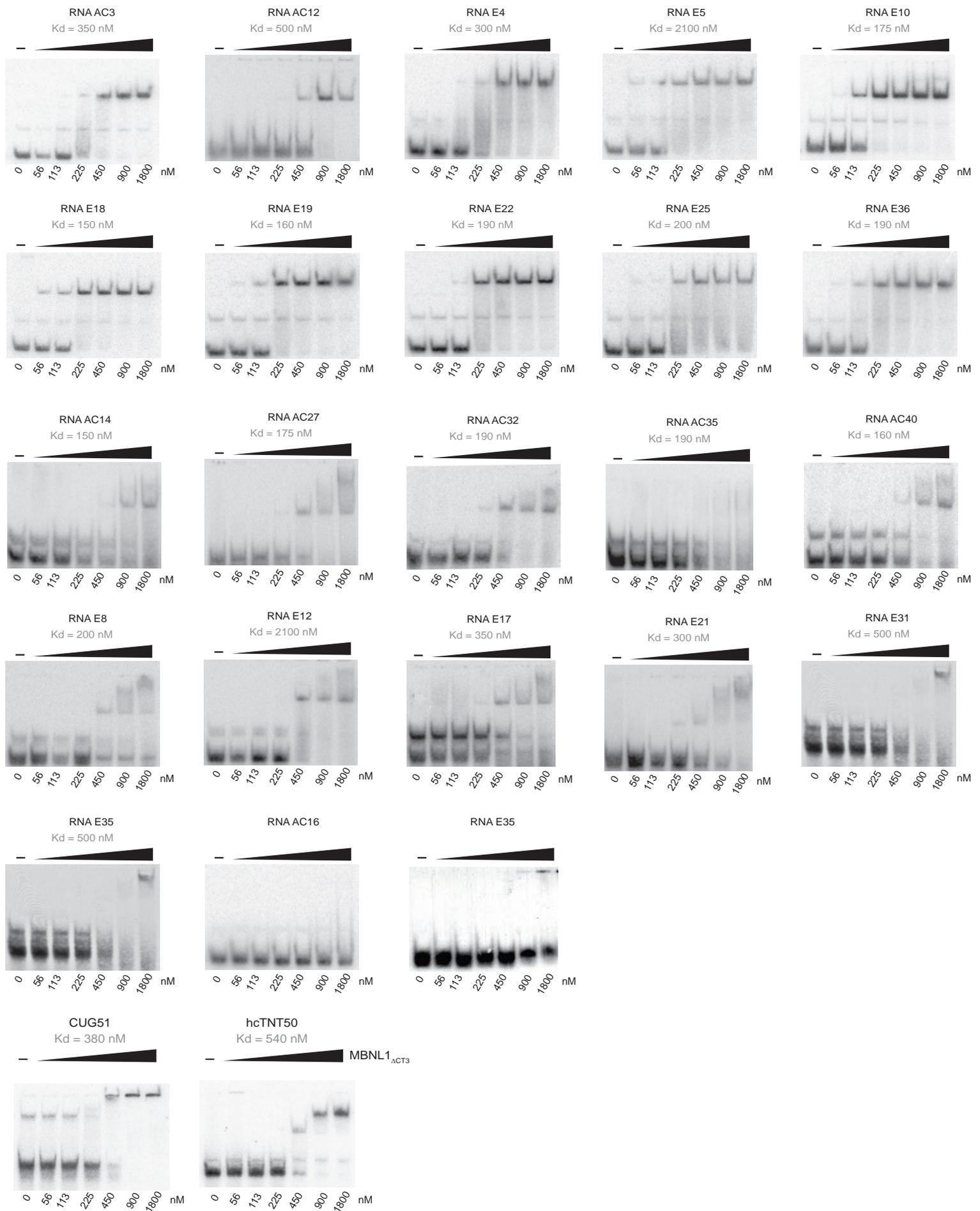


Figure S2 Vautrin et al.

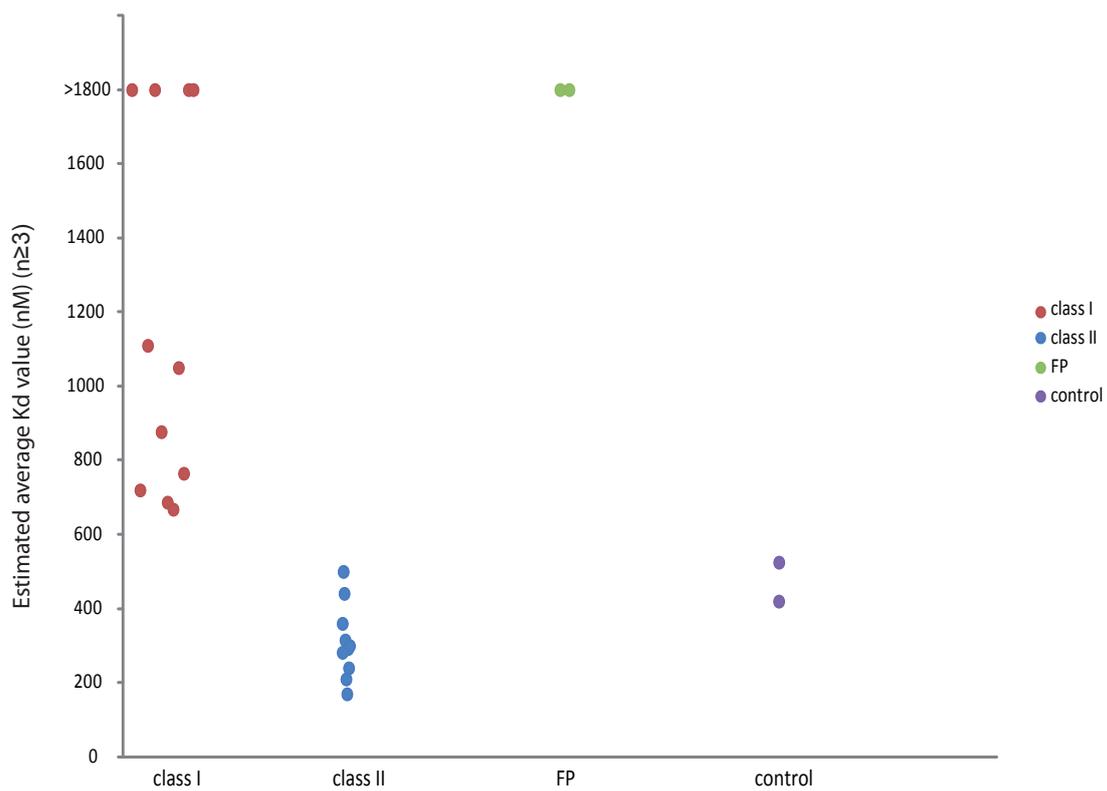


Figure S3 Vautrin et al.

III.2. Résultats complémentaires

III.2.a. Figures supplémentaires concernant les analyses structurales

- ✓ Exemples supplémentaires d'analyse de la structure 2D et de la protection par MBNL1 Δ CT3 des ARN de la classe II

Les Figures 48 et 49 illustrent l'analyse de la structure 2D de l'ARN E18 et l'analyse de la structure 2D et de la protection par MBNL1 Δ CT3 de l'ARN AC3, respectivement.

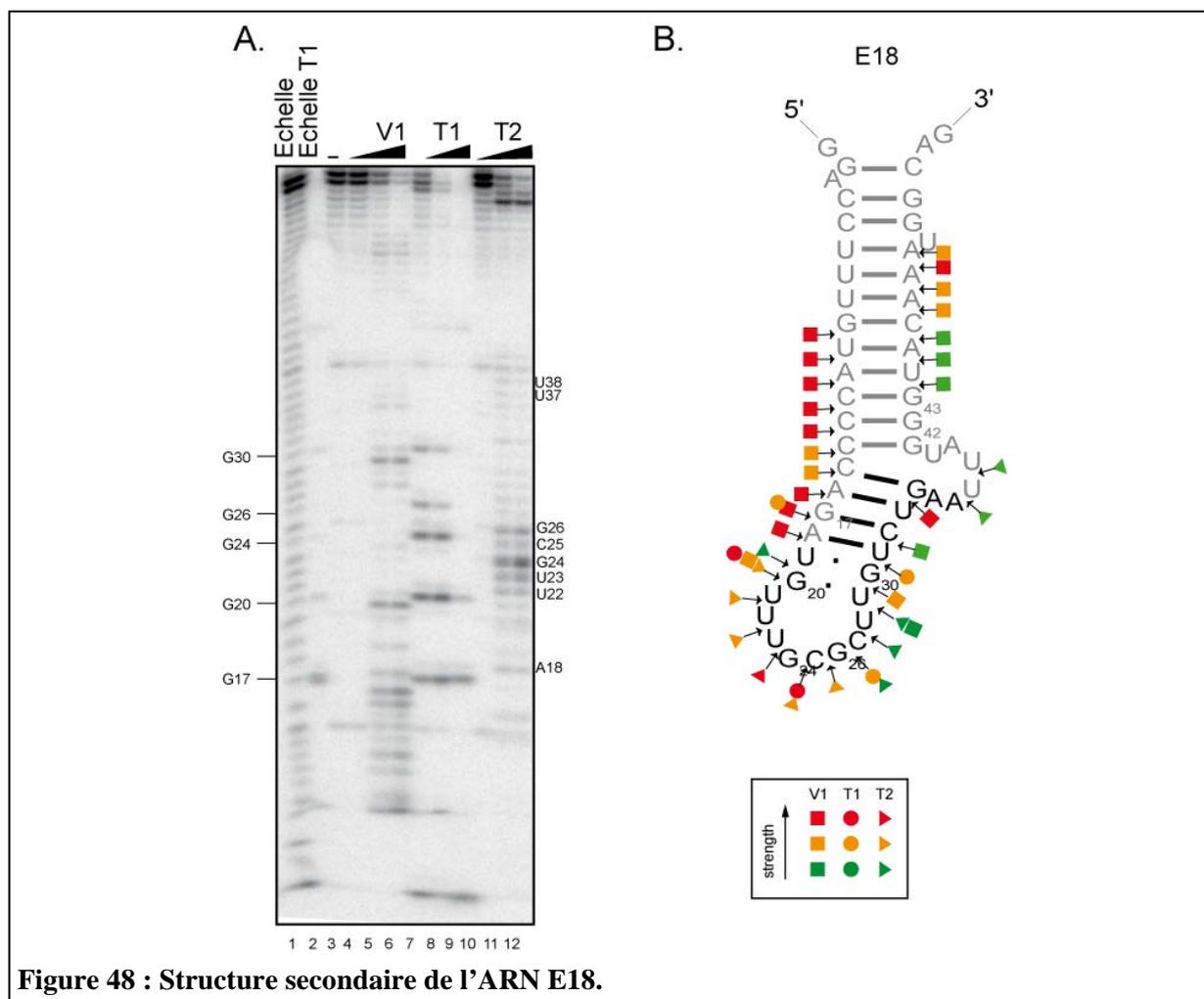
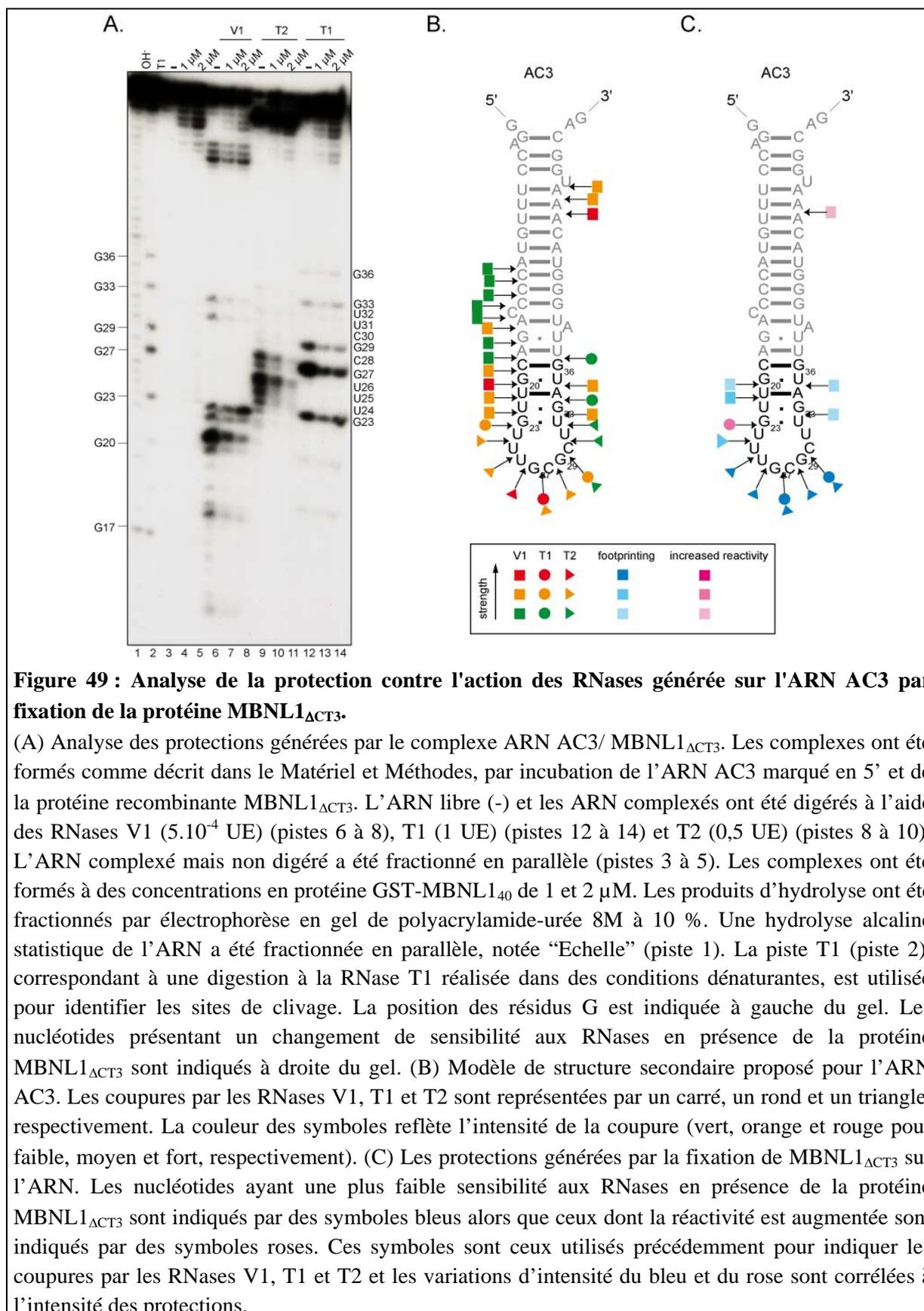


Figure 48 : Structure secondaire de l'ARN E18.

(A) Fractionnement des produits d'hydrolyse de l'ARN E18. L'ARN E18 marqué radioactivement en 5' (0,6 fmol) a été hydrolysé par des quantités croissantes de RNase V1 (10^{-3} ; $5 \cdot 10^{-4}$; 10^{-4} UE) (pistes 4 à 6), de RNase T1 (0.25 ; 0.5 et 1 UE) (pistes 7 à 9), ou de RNase T2 (0.5 ; 1 et 2 UE) (pistes 10 à 12) dans du tampon D 1X, comme décrit dans la partie Matériel et Méthodes. Les produits d'hydrolyse ont été fractionnés par électrophorèse en gel de polyacrylamide-urée 8M à 10 %. Une hydrolyse alcaline statistique de l'ARN notée "Echelle" (piste 1), ainsi que l'ARN non digéré (-) (piste 3) ont été fractionnés en parallèle. La piste T1 (piste 2), correspondant à une digestion à la RNase T1 réalisée dans des conditions dénaturantes, est utilisée pour identifier les sites de clivage. (B) Modèle de structure secondaire proposé pour l'ARN E18. Les coupures par les RNases V1, T1 et T2 sont représentées par un carré, un rond et un triangle, respectivement. La couleur des symboles reflète l'intensité de la coupure (vert, orange et rouge pour faible, moyen et fort, respectivement).



✓ Analyse de la structure 2D des ARN de la classe I

L'analyse de la structure 2D de plusieurs ARN de la classe I a été réalisée dans les mêmes conditions expérimentales que pour les ARN de la classe II. Des exemples des profils de clivage obtenus pour les ARN E7, E14 et E31 sont illustrés dans les Figures 50, 51 et 52. Les données obtenues mettent en évidence de nombreux clivages de la RNase V1 au niveau de la séquence dégénérée, indiquant un degré important de structuration de cette séquence. Cependant, nous n'avons pas réussi à prédire comment des appariements de type Watson-Crick classiques pourraient expliquer les clivages observés, ceci, en envisageant aussi bien la formation de paires de bases au sein de la boucle terminale que la formation de paires de bases au sein d'un homodimère. La structure particulière de ces ARN reste encore énigmatique et seule une analyse de structure 3D par RMN pourra nous apporter une réponse.

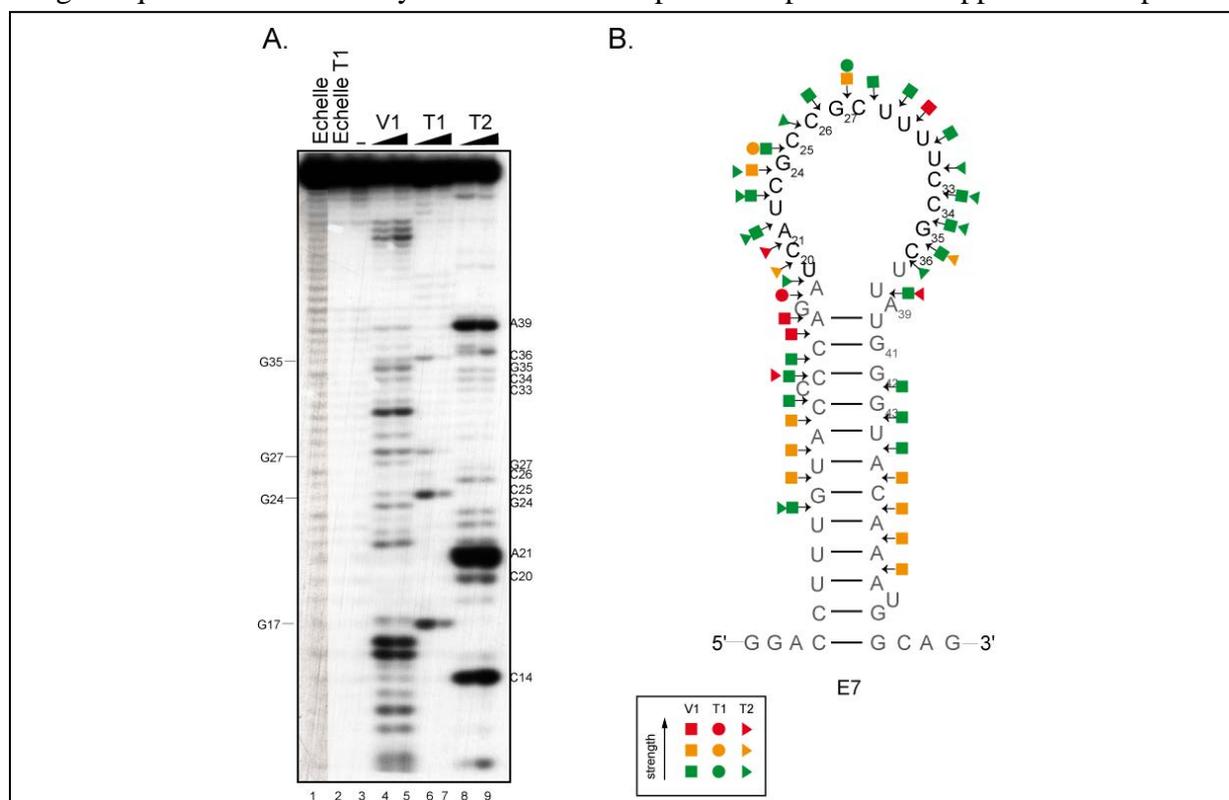


Figure 50 : Structure de l'ARN E7 de la famille I.

(A) Fractionnement des produits d'hydrolyse de l'ARN E7. L'ARN E7 marqué radioactivement en 5' (0,6 fmol) a été hydrolysé par des quantités croissantes de RNase V1 (10^{-3} ; $5 \cdot 10^{-4}$; 10^{-4} UE) (pistes 4 et 5), de RNase T1 (0.25 ; 0.5 et 1 UE) (pistes 6 et 7), ou de RNase T2 (0.5 ; 1 et 2 UE) (pistes 8 et 9) dans du tampon D IX, comme décrit dans la partie Matériel et Méthodes. Les produits d'hydrolyse ont été fractionnés par électrophorèse en gel de polyacrylamide-urée 8M à 10 %. Une hydrolyse alcaline statistique de l'ARN notée "Echelle" (piste 1), ainsi que l'ARN non digéré (-) (piste 3) ont été fractionnés en parallèle. La piste T1 (piste 2), correspondant à une digestion à la RNase T1 réalisée dans des conditions dénaturantes, est utilisée pour identifier les sites de clivage. (B) Modèle de structure secondaire proposé pour l'ARN E7. Les coupures par les RNases V1, T1 et T2 sont représentés par un carré, un rond et un triangle, respectivement. La couleur des symboles reflète l'intensité de la coupure (vert, orange et rouge pour faible, moyen et fort, respectivement).

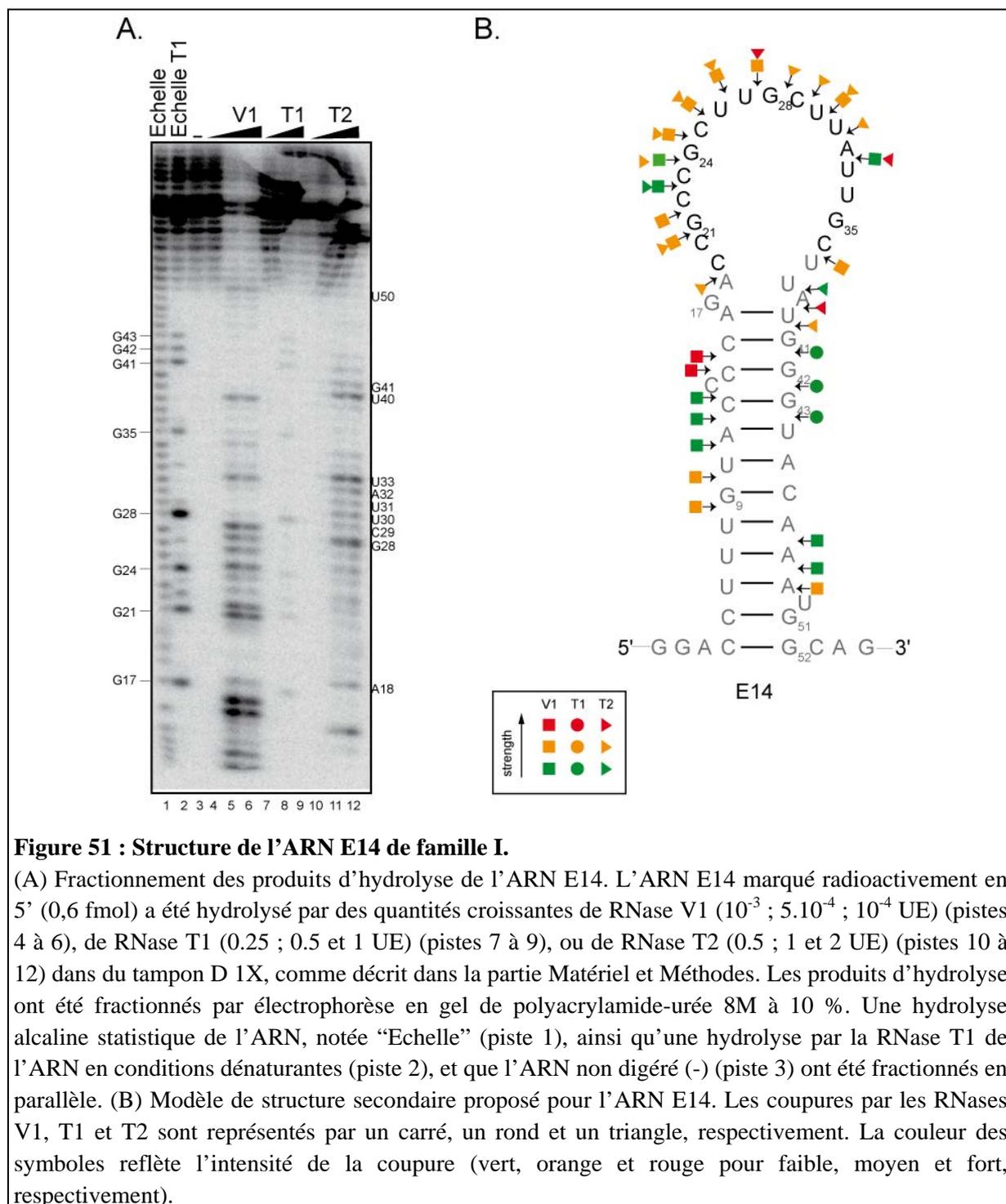


Figure 51 : Structure de l'ARN E14 de famille I.

(A) Fractionnement des produits d'hydrolyse de l'ARN E14. L'ARN E14 marqué radioactivement en 5' (0,6 fmol) a été hydrolysé par des quantités croissantes de RNase V1 (10^{-3} ; $5 \cdot 10^{-4}$; 10^{-4} UE) (pistes 4 à 6), de RNase T1 (0.25 ; 0.5 et 1 UE) (pistes 7 à 9), ou de RNase T2 (0.5 ; 1 et 2 UE) (pistes 10 à 12) dans du tampon D 1X, comme décrit dans la partie Matériel et Méthodes. Les produits d'hydrolyse ont été fractionnés par électrophorèse en gel de polyacrylamide-urée 8M à 10 %. Une hydrolyse alcaline statistique de l'ARN, notée "Echelle" (piste 1), ainsi qu'une hydrolyse par la RNase T1 de l'ARN en conditions dénaturantes (piste 2), et que l'ARN non digéré (-) (piste 3) ont été fractionnés en parallèle. (B) Modèle de structure secondaire proposé pour l'ARN E14. Les coupures par les RNases V1, T1 et T2 sont représentés par un carré, un rond et un triangle, respectivement. La couleur des symboles reflète l'intensité de la coupure (vert, orange et rouge pour faible, moyen et fort, respectivement).

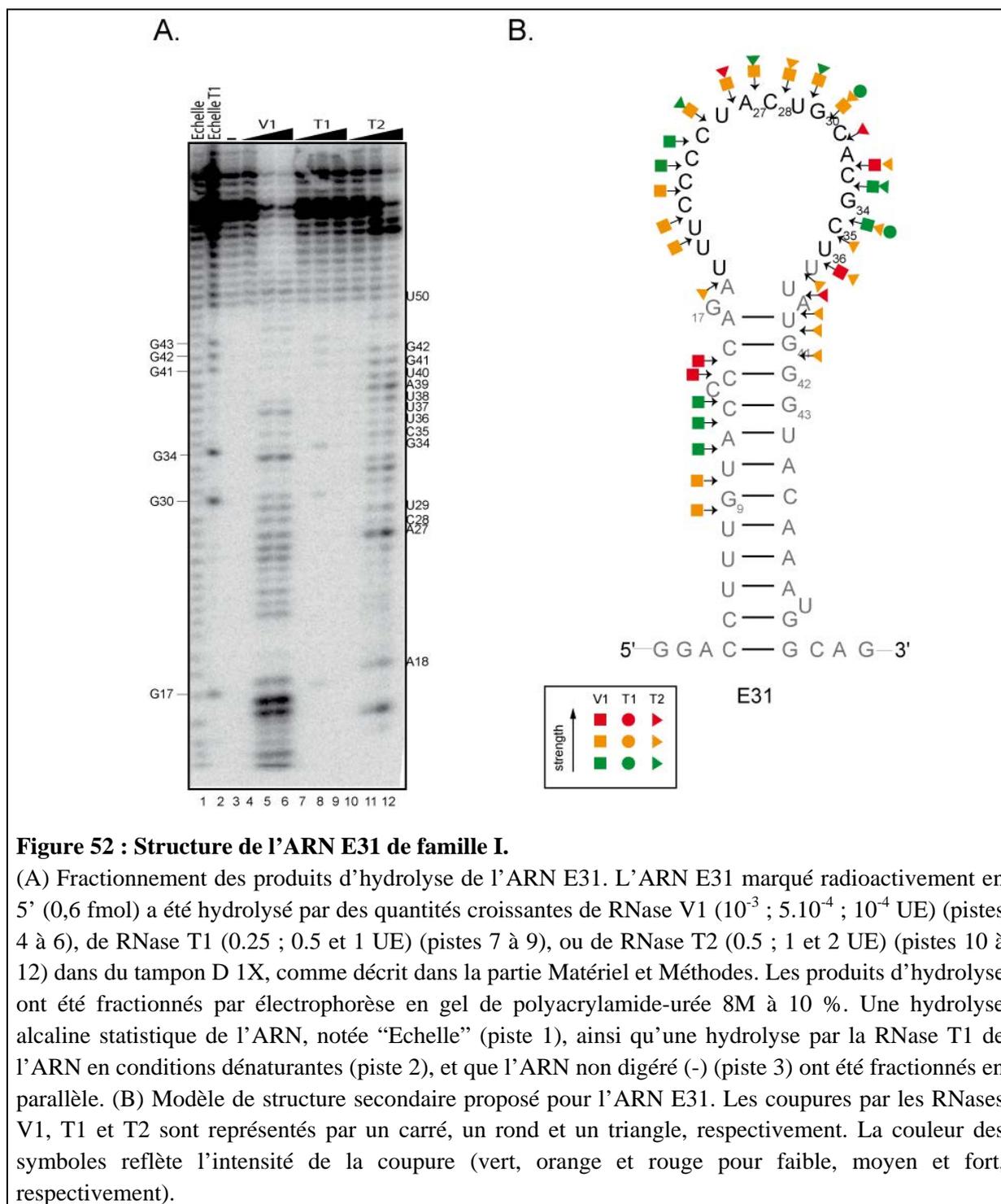


Figure 52 : Structure de l'ARN E31 de famille I.

(A) Fractionnement des produits d'hydrolyse de l'ARN E31. L'ARN E31 marqué radioactivement en 5' (0,6 fmol) a été hydrolysé par des quantités croissantes de RNase V1 (10^{-3} ; $5 \cdot 10^{-4}$; 10^{-4} UE) (pistes 4 à 6), de RNase T1 (0.25 ; 0.5 et 1 UE) (pistes 7 à 9), ou de RNase T2 (0.5 ; 1 et 2 UE) (pistes 10 à 12) dans du tampon D 1X, comme décrit dans la partie Matériel et Méthodes. Les produits d'hydrolyse ont été fractionnés par électrophorèse en gel de polyacrylamide-urée 8M à 10 %. Une hydrolyse alcaline statistique de l'ARN, notée "Echelle" (piste 1), ainsi qu'une hydrolyse par la RNase T1 de l'ARN en conditions dénaturantes (piste 2), et que l'ARN non digéré (-) (piste 3) ont été fractionnés en parallèle. (B) Modèle de structure secondaire proposé pour l'ARN E31. Les coupures par les RNases V1, T1 et T2 sont représentés par un carré, un rond et un triangle, respectivement. La couleur des symboles reflète l'intensité de la coupure (vert, orange et rouge pour faible, moyen et fort, respectivement).

III.2.b. Etude par triple-hybride de l'affinité des différentes isoformes de MBNL1 pour les ARN sélectionnés

Comme mentionné précédemment, l'expérience SELEX a été réalisée avec la protéine tronquée MBNL1 $_{\Delta CT3}$, ceci du fait de la très faible solubilité de la protéine MBNL1 recombinante entière. Afin de confirmer l'affinité des différentes isoformes naturelles de

MBNL1 pour les ARN SELEX, N. Marmier-Gourrier a testé l'efficacité d'interaction entre les ARN que j'avais isolés par l'approche SELEX et les différentes isoformes de MBNL1 par la méthode triple-hybride. Le principe de cette méthode repose sur la reconstitution, dans la levure, d'un facteur de transcription, à partir de deux protéines de fusion et d'un ARN. Une des protéines de fusion contient la protéine d'intérêt fusionnée au domaine activateur du facteur de transcription Gal4 de levure. Si cette protéine est capable d'interagir avec l'ARN, la transcription du gène rapporteur His3 est activée et les cellules ont la capacité de pousser en milieu minimum dépourvu d'histidine. La force de l'interaction est alors évaluée par ajout de 3AT, un compétiteur du produit du gène HIS3 (Fig. 53).

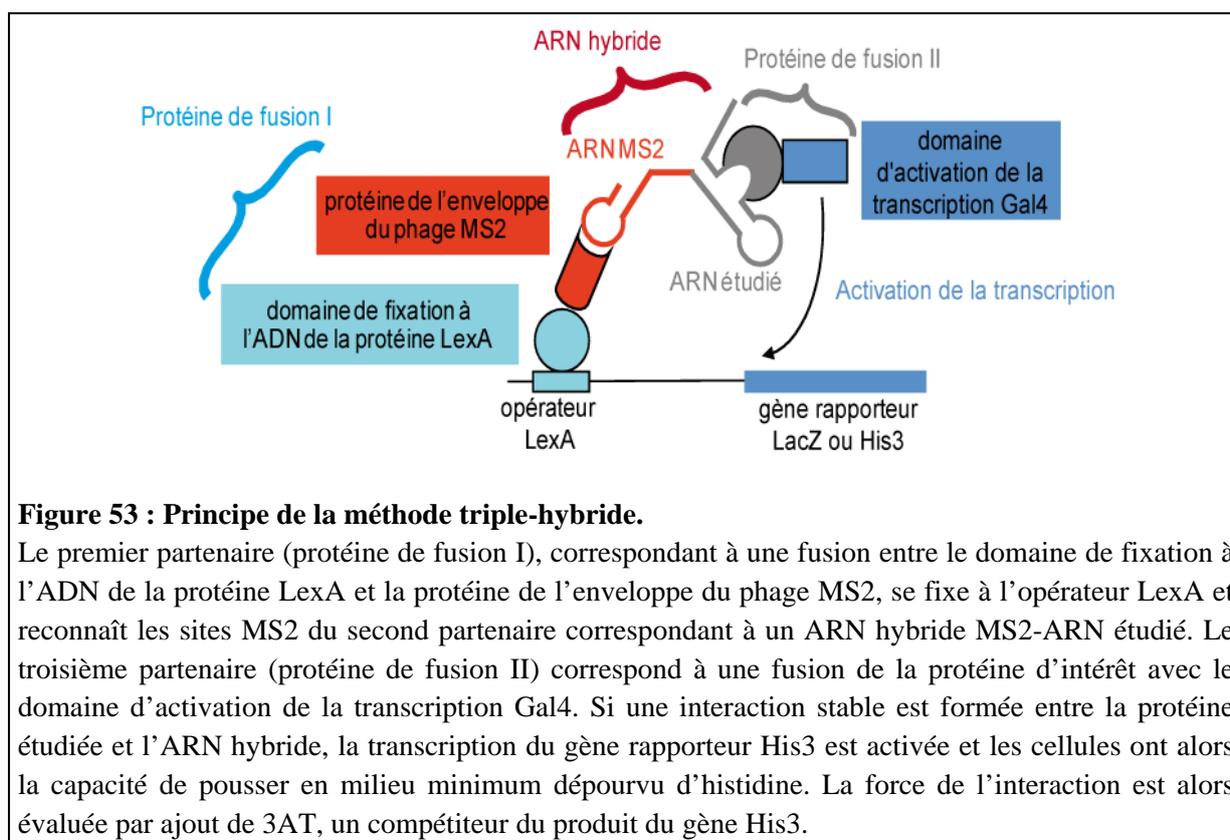


Figure 53 : Principe de la méthode triple-hybride.

Le premier partenaire (protéine de fusion I), correspondant à une fusion entre le domaine de fixation à l'ADN de la protéine LexA et la protéine de l'enveloppe du phage MS2, se fixe à l'opérateur LexA et reconnaît les sites MS2 du second partenaire correspondant à un ARN hybride MS2-ARN étudié. Le troisième partenaire (protéine de fusion II) correspond à une fusion de la protéine d'intérêt avec le domaine d'activation de la transcription Gal4. Si une interaction stable est formée entre la protéine étudiée et l'ARN hybride, la transcription du gène rapporteur His3 est activée et les cellules ont alors la capacité de pousser en milieu minimum dépourvu d'histidine. La force de l'interaction est alors évaluée par ajout de 3AT, un compétiteur du produit du gène His3.

Les résultats obtenus sont présentés dans la Figure 54. Les données obtenues montrent que l'isoforme 40, qui est l'isoforme la plus abondante, et la forme tronquée MBNL1_{ACT3}, que nous avons utilisée pour l'approche SELEX, donnent des résultats très similaires. Ces résultats valident le choix de cette forme tronquée que nous avons utilisée pour effectuer la sélection par la méthode SELEX et que nous utiliserons pour la suite des travaux. De plus, nous constatons que toutes les isoformes de MBNL1 présentent une forte affinité pour les ARN SELEX. Les données de triple-hybride confirment également que l'affinité des différentes isoformes de MBNL1 pour les ARN SELEX est meilleure que celle observée pour les répétitions CUG.

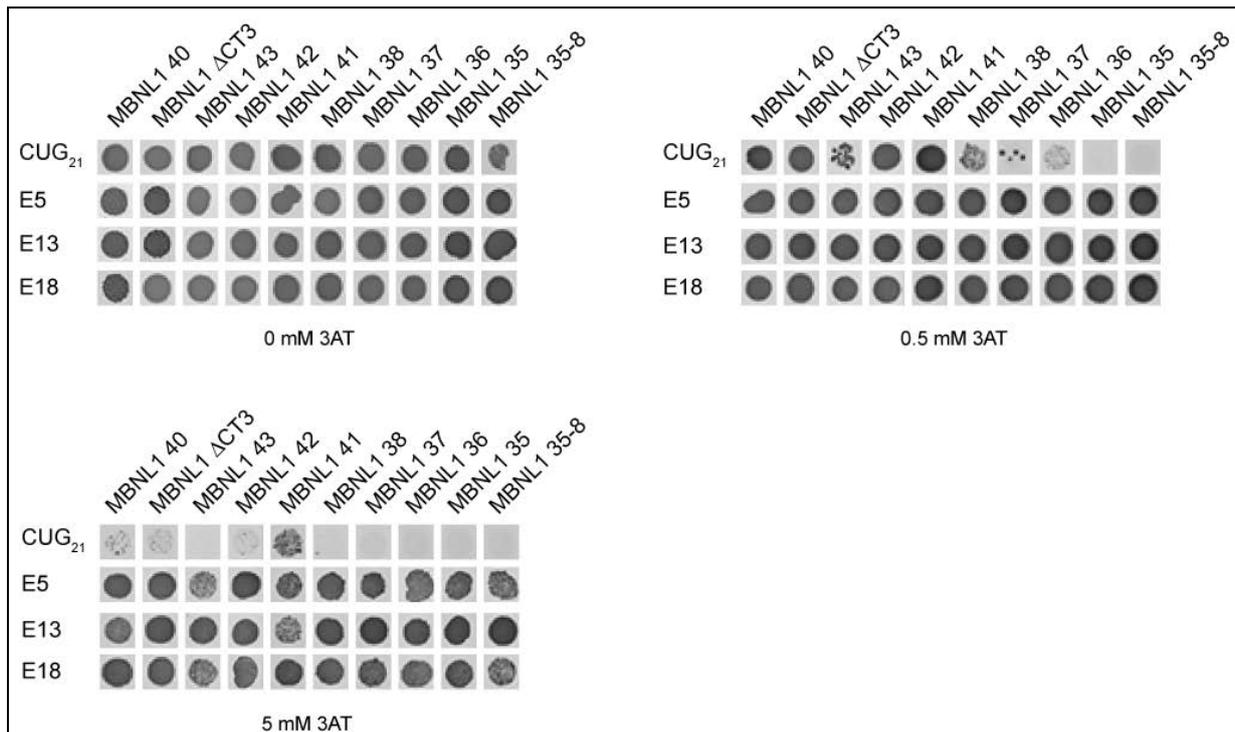


Figure 54 : Les isoformes naturelles de MBNL1 présentent une forte affinité pour les ARN SELEX de la famille II.

Les expériences de triple-hybrides ont été réalisées en utilisant différents ARN SELEX de la famille II et l'ARN (CUG)₂₁ comme ARN d'intérêt et l'ensemble des isoformes naturelles de MBNL1 ainsi que la forme recombinante MBNL1_{ΔCT3} comme protéines d'intérêt. L'ARN MS2.2 nous a servi de contrôle de spécificité d'interaction entre les ARN SELEX et les protéines MBNL1. La force de l'interaction a été évaluée par addition de 3AT à une concentration de 0.5 ou 5 mM.

III.2.c. Etude de la fonctionnalité des séquences SELEX *in cellulo*

Il était important de tester si les ARN SELEX de classe II, qui ont une très forte affinité pour MBNL1 pouvaient être des éléments *in cis* capables de réguler l'épissage alternatif par fixation de MBNL1. Pour cela, j'ai inséré des séquences ADN correspondant à plusieurs de ces ARN au sein d'un minigène dont l'épissage alternatif est régulé par MBNL1. Nous avons utilisé une construction qui permettait de produire un pré-ARNm dans lequel l'exon 5 du pré-ARNm hcTNT et ses régions introniques bordantes étaient insérés dans l'intron 2 du récepteur à l'insuline (IR), flanqué de ses exons bordants (Fig.55). Comme nous l'avons vu dans la partie Introduction, le pré-ARNm hcTNT contient deux sites de fixation de MBNL1 aux positions -40 et -21 en amont de l'exon 5. Ces deux sites encadrent la séquence polypyrimidine essentielle à la réaction d'épissage et il nous était donc impossible de les remplacer par les séquences SELEX. Nous avons alors remplacé les nucléotides -100 à -50 de l'intron 4 du pré-ARNm de hcTNT par les différentes séquences SELEX. Nous avons réalisé les tests sur 4 des ARN SELEX: l'ARN AC16 (minigène hcTNT-AC16) qui ne contient pas

de site de fixation de MBNL1, les ARN E18 et AC3 qui appartiennent à la famille II et contiennent tous deux un motif de haute affinité pour MBNL1 (minigènes hcTNT-E18 et hcTNT-AC3, respectivement) et enfin l'ARN AC3 muté dans le motif YGCY de fixation à MBNL1 (CGCU -> CACU) muté (minigène hcTNT-AC3mut2).

Les minigènes obtenus après insertion des séquences SELEX ont été individuellement co-transfectés dans des cellules HEK293 et l'inclusion de l'exon 5 dans les ARNm produits à partir des minigènes a été analysée par RT-PCR 48 heures après transfection. Les résultats obtenus sont présentés dans la Figure 55.

Lorsque l'exon 5 était environné par ses séquences introniques naturelles, il était inclus dans environ 60 % des produits d'épissage. Par contre, après insertion de la séquence SELEX contrôle (minigène hcTNT-AC16), nous observions une exclusion de plus de 85 % de l'exon 5. Ainsi, le remplacement des 50 nucléotides de la séquence intronique par cette séquence modifiait fortement le profil d'inclusion de l'exon 5. Nous avons donc décidé d'utiliser les données obtenues avec ce minigène contrôle comme référence pour tester l'effet des séquences SELEX ayant un site de fixation pour MBNL1. Dans le cas des minigènes hcTNT-AC3 et hcTNT-E18, nous avons observé une inclusion de l'exon 5 dans environ 60 % et 66 % des produits d'épissage, respectivement. Ainsi la présence d'une séquence SELEX contenant un site fort de fixation de MBNL1 entraînait une forte activation de l'inclusion de l'exon 5. Cette activation était bien liée à MBNL1 puisque la mutation du site de fixation de MBNL1 au sein de la séquence SELEX AC3 (minigène hcTNT-AC3mut2) conduisait à une inclusion de l'exon 5 dans seulement 15 % des produits d'épissage, c'est-à-dire à un taux comparable à celui obtenu avec le minigène contrôle (minigène hcTNT-AC16). L'ensemble de ces résultats montrait une activité MBNL1-dépendante sur l'épissage, et de manière étonnante, dans ce cas, la protéine MBNL1 n'agissait plus comme répresseur mais comme activateur de l'épissage.

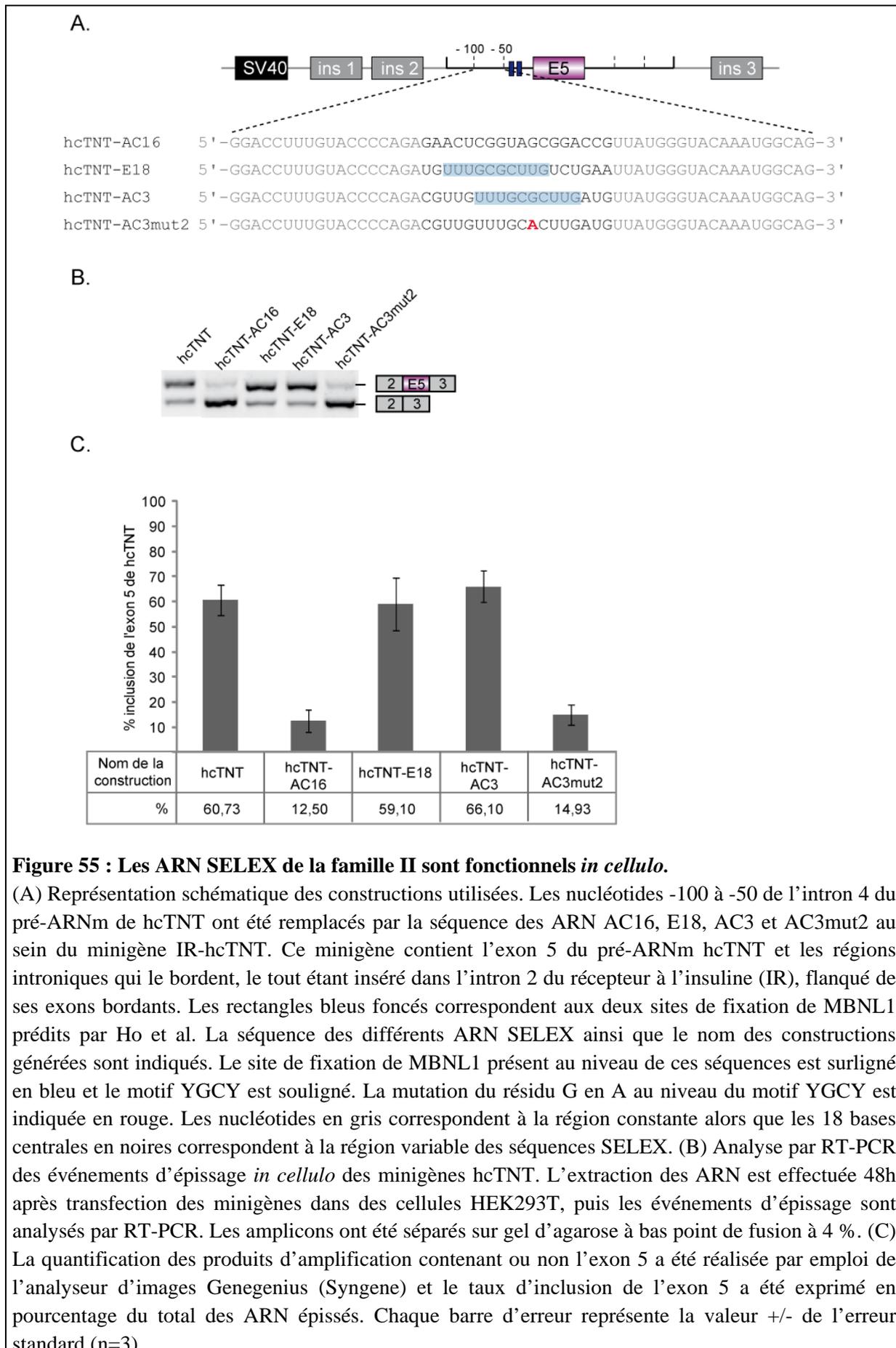


Figure 55 : Les ARN SELEX de la famille II sont fonctionnels *in cellulo*.

(A) Représentation schématique des constructions utilisées. Les nucléotides -100 à -50 de l'intron 4 du pré-ARNm de hcTNT ont été remplacés par la séquence des ARN AC16, E18, AC3 et AC3mut2 au sein du minigène IR-hcTNT. Ce minigène contient l'exon 5 du pré-ARNm hcTNT et les régions introniques qui le bordent, le tout étant inséré dans l'intron 2 du récepteur à l'insuline (IR), flanqué de ses exons bordants. Les rectangles bleus foncés correspondent aux deux sites de fixation de MBNL1 prédits par Ho et al. La séquence des différents ARN SELEX ainsi que le nom des constructions générées sont indiqués. Le site de fixation de MBNL1 présent au niveau de ces séquences est surligné en bleu et le motif YGCY est souligné. La mutation du résidu G en A au niveau du motif YGCY est indiquée en rouge. Les nucléotides en gris correspondent à la région constante alors que les 18 bases centrales en noires correspondent à la région variable des séquences SELEX. (B) Analyse par RT-PCR des événements d'épissage *in cellulo* des minigènes hcTNT. L'extraction des ARN est effectuée 48h après transfection des minigènes dans des cellules HEK293T, puis les événements d'épissage sont analysés par RT-PCR. Les amplicons ont été séparés sur gel d'agarose à bas point de fusion à 4 %. (C) La quantification des produits d'amplification contenant ou non l'exon 5 a été réalisée par emploi de l'analyseur d'images Genegenius (Syngene) et le taux d'inclusion de l'exon 5 a été exprimé en pourcentage du total des ARN épissés. Chaque barre d'erreur représente la valeur +/- de l'erreur standard (n=3).

IV. Rôle des différents domaines de MBNL1 dans la reconnaissance des ARN

Pour compléter l'étude, il nous paraissait important de définir quels segments de MBNL1 étaient impliqués dans la reconnaissance de l'ARN. Comme mentionné précédemment, d'après les données de Kino *et al.*, l'exon 3 semblait jouer un rôle important dans la reconnaissance des répétitions CUG (Kino *et al.*, 2004). Il nous paraissait donc important de déterminer l'importance de cet exon pour la reconnaissance des pré-ARNm cibles. De plus, la protéine Mbl de drosophile ne contenant qu'une seule paire de doigts à Zn^{2+} , et non deux comme la protéine humaine, nous nous sommes demandé si un seul couple de doigts à Zn^{2+} était suffisant pour la fixation de MBNL1 sur ses ARN cibles. Par ailleurs, il avait été observé que la protéine MBNL1 était capable d'interagir avec elle-même et de former des homodimères. Cette capacité d'oligomérisation de MBNL1 nécessitait le domaine C-terminal, cependant les séquences impliquées au sein de ce domaine n'avaient pas été identifiées (Yuan *et al.*, 2007). Nous nous sommes demandé si l'exon 7 alternatif qui code une région appartenant au domaine C-terminal pouvait jouer un rôle dans cette dimérisation et si cette dernière pouvait avoir un impact sur la reconnaissance des ARN.

Les réponses que nous avons pu apporter aux questions concernant le rôle de l'exon 3 et à la dimérisation de MBNL1 ont été publiées dans un article décrivant une étude globale sur la caractérisation des rôles des différents exons alternatifs de MBNL1 (publication n°2, ci-après). Les travaux relatifs aux doigts à zinc seront décrits à la suite de la présentation de cet article.

IV.1. Présentation de l'article n°2

Cet article correspond à étude globale sur la caractérisation des rôles des différents exons alternatifs de MBNL1. Caractériser le rôle de ces différents exons alternatifs était d'autant plus important que l'expression des différentes isoformes de MBNL1 est régulée au cours du développement et est dérégulée chez les patients atteints de DM1, suggérant une modulation des fonctions de MBNL1 par épissage alternatif. Ainsi, les exons alternatifs 5 et 7 sont majoritairement inclus dans les muscles et cerveaux fœtaux et anormalement inclus chez les patients DM1. Alors que l'exon 5 semble jouer un rôle dans la localisation nucléaire de la protéine (Lin *et al.*, 2006a ; Dhaenens *et al.*, 2008), aucun rôle n'avait pu être proposé pour l'exon 7. Par ailleurs, le rôle de ces exons dans la fonction d'épissage alternatif et dans la reconnaissance des ARN n'avait pas été testé.

L'objectif de l'étude globale qui a été menée en collaboration avec les équipes de N. Sergeant (Lille), D. Furling (Paris) et N. Charlet (Strasbourg) était de définir les rôles respectifs des

exons alternatifs 3, 5 et 7, ainsi que de l'exon constitutif 6, dans la localisation nucléaire de MBNL1, la reconnaissance des ARN, l'activité de régulation de l'épissage et les propriétés de dimérisation de MBNL1.

Mon travail sur ce projet a consisté à étudier l'importance de la région C-terminale de MBNL1 et de l'exon 3 dans sa capacité de fixation à l'ARN. Pour cela, j'ai réalisé des expériences de gel retard en incubant l'ARN 5'-hcTNT₅₀₋₀ marqué radioactivement en 5' (1,2 fmoles) avec des quantités croissantes de protéines MBNL1₄₀, MBNL_{ΔCT3} et MBNL1_{ΔCT} (100 à 2500 nM). J'ai ainsi pu mettre en évidence que, comme attendu, la délétion de la région C-terminale de la protéine MBNL1₄₀ affecte peu l'affinité de la protéine pour l'ARN hcTNT. A l'inverse, la délétion de l'exon 3 de la protéine MBNL1 diminue l'affinité de la protéine pour l'ARN hcTNT sans pour autant abolir la fixation (Kd de 1439 nM contre 599 nM). Ainsi, l'exon 3 de MBNL1 n'est pas absolument requis pour que MBNL1 se fixe à l'ARN, néanmoins sa présence augmente l'affinité de MBNL1 pour ses séquences cibles.

Plus globalement, cette étude a permis de tirer les conclusions suivantes :

- L'exon 3 régule l'affinité de MBLN1 pour les répétitions CUG et ses cibles naturelles. Cette propriété est probablement corrélée au rôle de cet exon dans l'activité de régulateur d'épissage de MBNL1. En effet, en l'absence de l'exon 3, MBNL1 modifie à peine l'épissage alternatif de ses pré-ARNm cibles.
- L'exon 5 est impliqué, en association avec l'exon 6 constitutif, dans la localisation subcellulaire de la protéine.
- L'exon 6 intervient également dans la régulation de la fonction de facteur d'épissage.
- L'exon 7 permet de renforcer la propriété de dimérisation de MBNL1.

Ainsi, la région N-terminale de la protéine intervient principalement dans la fixation à l'ARN, et donc dans la régulation de l'épissage, alors que la région C-terminale de MBNL semble nécessaire à la fois pour la régulation de l'épissage, la localisation subcellulaire, la dimérisation et enfin les interactions protéine/protéine.

Nos observations peuvent expliquer les défauts d'épissage de nombreux pré-ARNm observés chez les patients atteints de DM1. En effet, une augmentation de l'inclusion des exons 3, 5 et 7 est observée chez ces patients d'où i) une forte relocalisation nucléaire de MBNL1, ii) sa forte affinité pour les répétitions CUG et iii) son oligomérisation accrue. L'ensemble de ces effets conduit à une séquestration renforcée de MBNL1 par les répétitions CUG localisées dans le noyau.

PUBLICATION 2

Analysis of exonic regions involved in nuclear localization, splicing activity, and dimerization of Muscleblind-like-1 isoforms

Hélène Tran¹, Nathalie Gourrier¹, Camille Lemerrier-Neuillet¹, Claire-Marie Dhaenens, **Audrey Vautrin**, Francisco José Fernandez-Gomez, Ludovic Arandel, Céline Carpentier, Hélène Obriot, Sabiha Eddarkaoui, Lucie Delattre, Edwige Van Brussels, Ian Holt, Glenn E. Morris, Bernard Sablonnière, Luc Buée, Nicolas Charlet-Berguerand, Susanna Schraen-Maschke, Denis Furling, Isabelle Behm-Ansmant, Christiane Branlant, Marie-Laure Caillet-Boudin and Nicolas Sergeant.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 18, pp. 16435–16446,
May 6, 2011

Analysis of Exonic Regions Involved in Nuclear Localization, Splicing Activity, and Dimerization of Muscleblind-like-1 Isoforms^{*S}

Received for publication, October 18, 2010, and in revised form, March 9, 2011. Published, JBC Papers in Press, March 18, 2011, DOI 10.1074/jbc.M110.194928

Hélène Tran^{‡§1,2}, Nathalie Gourrier^{¶1,3}, Camille Lemerrier-Neuillet^{||**††1}, Claire-Marie Dhaenens^{‡§}, Audrey Vautrin^{¶14}, Francisco José Fernandez-Gomez^{‡§}, Ludovic Arandel^{||**††}, Céline Carpentier^{‡§}, Hélène Obriot^{‡§}, Sabiha Eddarkaoui^{‡§}, Lucie Delattre^{‡§}, Edwige Van Brussels^{‡§}, Ian Holt^{§§|||}, Glenn E. Morris^{§§|||}, Bernard Sablonnière^{‡§}, Luc Buée^{‡§§}, Nicolas Charlet-Berguerand^{¶¶}, Susanna Schraen-Maschke^{‡§}, Denis Furling^{||**††§}, Isabelle Behm-Ansmant^{¶15}, Christiane Branlant^{¶15,6}, Marie-Laure Caillet-Boudin^{‡§7}, and Nicolas Sergeant^{‡§7,8}

From the [‡]INSERM, U837, Alzheimer and Tauopathies, place de Verdun, F-59045 Lille, France, the [§]Université Lille Nord de France, Université Droit & Santé de Lille, Faculty of Medicine, Institute of Predictive Medicine and Therapeutic Research, Jean-Pierre Aubert Research Centre, place de Verdun, F-59045 Lille, France, the [¶]Laboratoire Laboratory of ARN-Ribonuclear Particles Maturation-Structure-Function, Molecular and Structural Enzymology, Unité Mixte de Recherche 7214 Nancy University-CNRS, Faculty of Sciences, BP 70239, 54506 Vandoeuvre-les-Nancy Cedex, France, the ^{||}Université Pierre et Marie Curie Université Paris 6, Unité Mixte 76, Institut de Myologie, Paris F-75013, France, the ^{**}CNRS, Unité Mixte de Recherche 7215, Paris F-75013, France, ^{††}INSERM, U974, Paris F-75013, France, ^{¶¶}IGBMC, 1 Rue Laurent FRIES, 67404 Illkirch, France, the ^{§§}Wolfson Centre for Inherited Neuromuscular Disease, Robert Jones and Agnes Hunt Orthopaedic Hospital, Oswestry, United Kingdom, and ^{|||}Institute for Science and Technology in Medicine, Keele University, United Kingdom

Muscleblind-like-1 (MBNL1) is a splicing regulatory factor controlling the fetal-to-adult alternative splicing transitions during vertebrate muscle development. Its capture by nuclear CUG expansions is one major cause for type 1 myotonic dystrophy (DM1). Alternative splicing produces MBNL1 isoforms that differ by the presence or absence of the exonic regions 3, 5, and 7. To understand better their respective roles and the consequences of the deregulation of their expression in DM1, here we studied the respective roles of MBNL1 alternative and constitutive exons. By combining genetics, molecular and cellular approaches, we found that (i) the exon 5 and 6 regions are both needed to control the nuclear localization of MBNL1; (ii) the exon 3 region strongly enhances the affinity of MBNL1 for its pre-mRNA target sites; (iii) the exon 3 and 6 regions are both required for the splicing regulatory activity, and this function is not enhanced by an exclusive nuclear localization of MBNL1;

and finally (iv) the exon 7 region enhances MBNL1-MBNL1 dimerization properties. Consequently, the abnormally high inclusion of the exon 5 and 7 regions in DM1 is expected to enhance the potential of MBNL1 of being sequestered with nuclear CUG expansions, which provides new insight into DM1 pathophysiology.

Splicing of pre-mRNA is a key post-transcriptional step in eukaryotic gene expression. A vast majority of vertebrate pre-mRNAs is alternatively spliced, allowing the production of several protein isoforms from transcripts of a given gene (1). The regulation of alternative splicing plays a major role in cell differentiation and in development and depends on the expression and activity of numerous splicing regulatory factors that are expressed differentially during development, according to the type of tissue. Defects in these alternative-splicing processes can contribute to pathogenesis, as demonstrated for a growing number of diseases, including neuromuscular diseases such as myotonic dystrophy type 1 (DM1)⁹ (2, 3).

DM1 is an autosomal disorder caused by an unstable CTG repeat expansion in the 3'-untranslated region (UTR) of the *DMPK* gene (4–6). One of the main etiological hypotheses of DM1 is based on a toxic RNA gain of function, leading to the dysregulation of alternative splicing. Mutant transcripts bearing long-CUG repeats acquire unusual A-form double-stranded RNA structures (7), accumulate in the nucleus, and lead to small ribonucleoprotein inclusions, named *foci* (8) that sequester RNA-binding proteins such as Muscleblind-like 1

* This work was supported by the Association Française contre les Myopathies Grants 14269 and 15047, the Agence Nationale de Recherche Neuro-Splicing de Tau BLAN 1114 01, the EURASNET EU Contract FP6, life sciences, genomics and biotechnology for health, the Centre National de la Recherche Scientifique, the Institut National pour la Santé et la Recherche Médicale, the French Ministry for Youth, National Education and Research, and the Lorraine Region.

^S The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Experimental Procedures and Figs. 1 and 2.

¹ These authors contributed equally to this work.

² Fellow of the INSERM and the Nord/Pas de Calais Region.

³ Association Française contre le Myopathies postdoctoral fellow.

⁴ Fellow of the French Ministry for Youth, National Education and Research.

⁵ Staff scientists at the CNRS.

⁶ To whom correspondence may be addressed. E-mail: christiane.branlant@maem.uhp-nancy.fr.

⁷ Staff scientists at the INSERM.

⁸ To whom correspondence may be addressed: INSERM U837, Alzheimer and Tauopathies, Université Lille Nord de France, USDL, IMPRT, Batiment Biserte, 1, place de Verdun, rue Polonowski, F-59045 Lille, France. E-mail: nicolas.sergeant@inserm.fr.

⁹ The abbreviations used are: DM1, myotonic dystrophy type 1; 3AT, 3-amino-1,2,4-triazole; BisTris, bis(2-hydroxyethyl)iminotris(hydroxymethyl)methane; hCTNT, human cardiac troponin T; IR, insulin receptor; MBNL1, Muscleblind-like 1.

Structure-Function Analysis of MBNL1

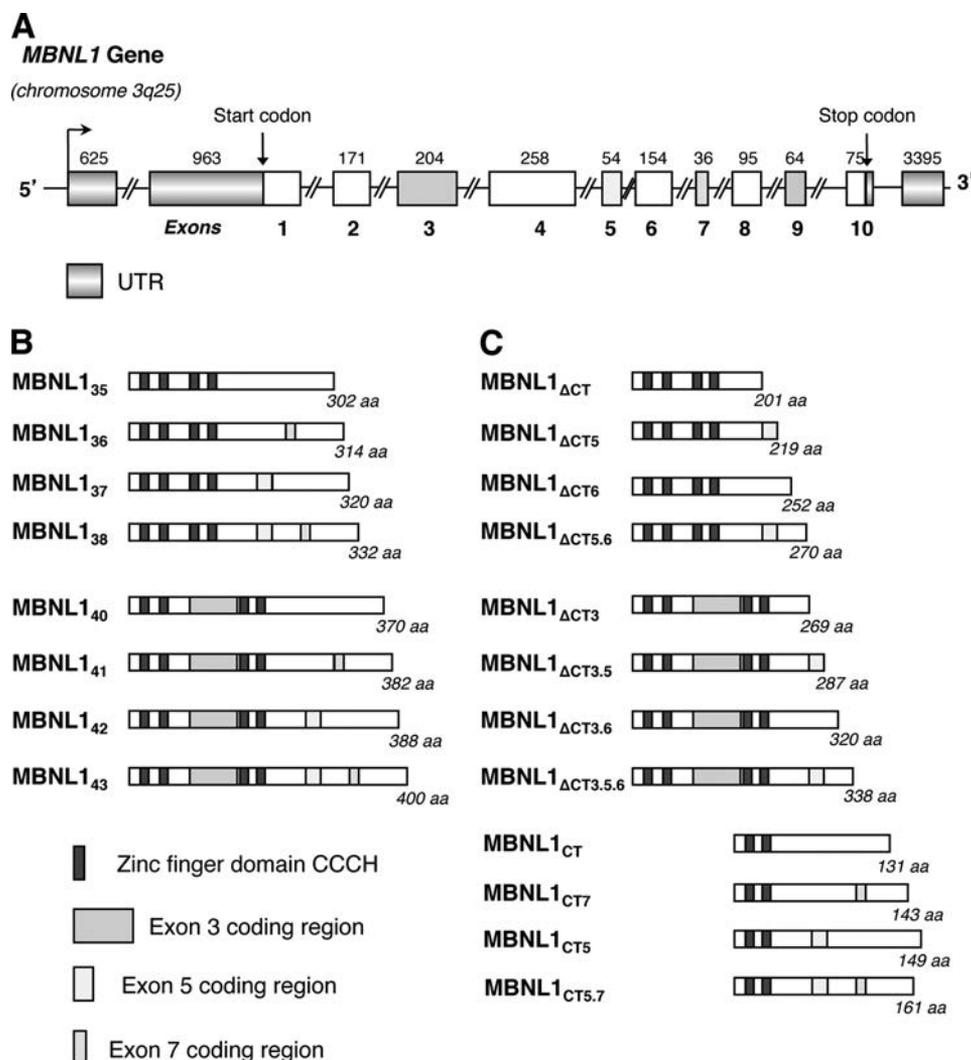


FIGURE 1. Alternative splicing of MBNL1 generates protein diversity. Clear gray boxes represent UTR. Gray boxes represent cassette exons. Empty boxes represent constitutive exons. *A*, genomic DNA organization of human *MBNL1* gene. Order and names of exons used in this work are shown. The length in nucleotides of each exon is shown. *B*, MBNL1 isoforms used in this work. Dark gray boxes represent CCCH zinc finger motifs. Two are located in MBNL1 exon 2, and the two others are located in MBNL1 exon 4. They are separated by cassette exon 3, which acts as a linker between zinc fingers 2 and 3. The name of each isoform is indicated on the left. Gray boxes indicate the presence of alternative exon regions in each isoform scheme. The length in amino acids (aa) is indicated for each isoform used in this work. *C*, truncated MBNL1 mutants used in this work. The name of each construct is indicated on the left. The nomenclature used indicates the protein sequence expressed by exons. The presence of alternative exons is indicated by the number of the exon or by gray boxes in the scheme. The amino acid length is noted for each construct used. When carrying a deletion of the C terminus, the constructs are annotated Δ CT followed where appropriate by the number of the exon. N-terminal truncated MBNL1 constructs are named CT followed by the number of the exon present in the sequence.

(MBNL1). The alternative splicing of several MBNL1 targets is thus abnormally modified in DM1 patients and in a mouse model in which MBNL1 expression is inactivated (9–11). In consequence, a loss-of-function mechanism has been proposed to contribute to DM1 pathogenesis, which is further supported by an *in vivo* model showing that *MBNL1* inactivation leads to many of the symptoms and molecular defects observed in DM1 (10, 11).

MBNL1 is the best example of splicing regulatory factors known to be involved in development and splicing deregulation in disease. Originally identified in *Drosophila melanogaster*, MBNL1 is described as being essential for the terminal differentiation of photoreceptors and muscles (12, 13). In humans and in mice, the splicing factor MBNL1 has been shown to participate in the postnatal remodeling of skeletal muscle and of the developing heart by controlling the developmentally reg-

ulated switch of a key set of pre-mRNAs (9, 14, 15). MBNL1 binds to pre-mRNAs containing YGCY sequence elements (16, 17) and promotes either the inclusion or the exclusion of alternative exons depending on the 5' or 3' localization of *cis*-regulatory elements (17). For instance, the inclusion of exon 5 in human cardiac troponin T (*hcTNT/TNNT2*) mRNA is inhibited by MBNL1, whereas the inclusion of exon 11 in insulin receptor (*IR/INSR*) mRNA is promoted by MBNL1 (16). MBNL1 binds to the polypyrimidine tract of *hcTNT* intron 4 and competes with the splicing factor U2AF for its binding site, explaining the MBNL1 inhibition of exon 5 inclusion (18). Conversely, by binding to elements located 3' to the alternative exon 11 in *IR* pre-mRNA (*IR/INSR*), MBNL1 enhances the inclusion of this exon (17, 19, 20).

The *MBNL1* gene encompasses 12 exons and the coding sequence is distributed over 10 exons numbered 1–10 (Fig. 1A)

(21–23), some of which (exons 3, 5, 7, and 9) are alternatively spliced. Thus, >10 isoforms of MBNL1 have been reported (21). The expression of these isoforms is developmentally regulated and altered in DM1. These observations suggest a modulation of MBNL1 protein function through alternative splicing. Indeed, the cassette exons 5 and 7 are included mainly in fetal brain and muscle rather than in adult tissues and are preferentially included in DM1 patients compared with controls (9, 24). Exons 1, 2, and 4 are always included and encode the four CCCH zinc finger domains involved in RNA binding (25). Exon 3 is often included and may act as a linker joining the second and third zinc finger domains (22, 25). Although the segment encoded by exon 3 is often included in MBNL1 protein, MBNL1 isoforms without this linker have already been found to have a lower affinity for CUG expansions compared with isoforms that include the exon 3-encoded sequence (22). The presence or absence of the sequence encoded by exon 5 has been suggested to modulate the nuclear localization of MBNL1 (9, 15, 24). However, the possible involvement of the bordering amino acid sequences encoded by constitutive exons such as exon 6 in the cellular localization of MBNL1 has not been investigated yet, and up to now, no functional role has been attributed to the highly conserved amino acid sequence encoded by exon 7 cassette, which is also mainly included in fetal brain and muscles and in MBNL1 protein from DM1 patients. It is therefore important to determine the relationship, if any, between the exon composition and function of MBNL1 isoforms.

In the present study, in an effort to define precisely the respective roles of the amino acid sequences encoded by the alternative cassette exons 3, 5, and 7 and constitutive exon 6 of human MBNL1 protein isoforms, we applied a large variety of *in vitro* and *in cellulo* approaches. This allowed us to determine the cellular localization, splicing activity, and RNA binding properties of the eight main normal MBNL1 isoforms as well as of a large number of MBNL1 variant proteins. Experiments were performed using both HeLa cells and human myoblasts that did or did not express CUG expansions. Our results provide a better definition of (i) the respective roles of the amino acid sequences encoded by exons 3 and 5 in the RNA binding property, nuclear localization and splicing regulatory property of MBNL1; (ii) the involvement of the exon 6-encoded sequence in MBNL1 nuclear retention and *in cellulo* splicing regulatory activity; (iii) the identification of a possible role of the exon 7-encoded sequence in MBNL1 self-dimerization.

EXPERIMENTAL PROCEDURES

Plasmids and Cloning—Full-length MBNL1 variant constructs (MBNL1₃₅, MBNL1₃₆, MBNL1₃₇, MBNL1₃₈, MBNL1₄₀, MBNL1₄₁, MBNL1₄₂, and MBNL1₄₃) used in this study have been described previously (24). All truncated forms of MBNL1 were derived from the appropriate full-length MBNL1 cDNAs and cloned using standard techniques. Briefly, MBNL1 cDNA was amplified with DYNAszyme EXT™ Taq polymerase (Finnzymes, Espoo, Finland). The same forward primer sequence was used for all constructs with a truncated C-terminal tail: 5'-atggctgttagtgcacacca-3'; the reverse primer sequences used were 5'-catggcagctgcggg-3' for the MBNL1_{ΔCT} and MBNL1_{ΔCT3} constructs; 5'-caggtcaaaggtgcctc-3' for

MBNL1_{ΔCT5}, and 5'-ctgggggagaaatgctgt-3' for MBNL1_{ΔCT6}, MBNL1_{ΔCT3.6} and MBNL1_{ΔCT3.5.6}, respectively. The cDNA of constructs with a truncated N-terminal tail were amplified using 5'-atgtaccagctcctaagaagaa-3' as the forward primer and 5'-ctacatctgggtaacatactgtg-3' as the reverse primer. PCR products were cloned into the PCR8/GW/TOPO TA cloning vector (Invitrogen) and recombined with LR Clonase in the expression plasmid pDEST53/GW, pcDNA3.1/nV5 plasmid for C-terminal truncated mutants, or in pACTII and pASIIΔΔ previously converted using the Gateway system according to the manufacturer's instructions. All plasmid DNA was double-stranded sequenced and purified using the Nucleobond® AX endotoxin free kit (Macherey Nagel, Düren, Germany). MBNL1 splicing activity was assessed using two minigenes described previously: the RTB300 minigene containing exon 5 *hcTNT* (26) and the *INSR* minigene containing exon 11 of the human IR (27).

Cell Culture and Transfection—HeLa cells (ECACC 93021013) were grown in monolayer cultures in 6-well plates in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen) supplemented with 10% fetal calf serum, 50 units/ml penicillin, 50 μg/ml streptomycin, and 4 mmol/liter glutamine, at 37 °C in a humidified (5% CO₂) incubator. Cells were plated in 6-wells plates, grown to ~70% confluence, and transiently co-transfected with 1 μg of minigene plasmid DNA and 3 μg of MBNL1 plasmid DNA, in triplicate, using the FuGENE HD transfection reagent (Roche Diagnostics) according to the manufacturer's instructions. Human muscle cells were isolated from skeletal muscle biopsies or autopsies as described (28), in accordance with French legislation on ethics. Wild-type (WT) and DM1 myoblasts were grown in Ham's F10 medium supplemented with 20% fetal calf serum and 5 μg/ml gentamycin (Invitrogen), at 37 °C and under 5% CO₂.

Immunofluorescence Microscopy—HeLa cells and myoblasts were transiently transfected with 2 μg of the different MBNL1 constructs, using the transfection reagent FuGENE HD according to the manufacturer's instructions. After 48 h, cells were fixed with 4% paraformaldehyde in 0.1 M phosphate buffer (PBS) for 15 min at room temperature. Cells were washed three times with PBS. The coverslips were mounted onto slides using the mounting medium Vectashield (Vector Laboratories) with DAPI. Images were acquired with a Zeiss apotome microscope or a Leica confocal microscope. All data were analyzed using Photoshop Element 6 software (Adobe).

Fluorescent *In Situ* Hybridization (FISH)—FISH was performed as described (29) using a Cy3-labeled peptide nucleic acid (CAG)₇ probe. Images were captured using a Leica confocal microscope and software (Leica microsystems) and processed with Adobe Photoshop software.

Protein Extraction, Nuclear Fractionation, and Western Blot Analysis—HeLa cells were lysed 48 h after transfection in radio-immune precipitation assay buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Nonidet P-40, 0.5% deoxycholate) supplemented with protease inhibitors (Complete mini; Roche Applied Science). Lysates were sonicated, homogenized at 4 °C, and centrifuged at 12,000 × *g* at 4 °C for 20 min. Myoblasts were washed in 20 mM HEPES at pH 7.9 with 5 mM NaF, 1 mM Na₂MoO₄, and 0.1 mM EDTA and lysed with the

Structure-Function Analysis of MBNL1

addition of 0.5% Nonidet P-40. Nuclear and cytoplasmic fractions were separated by centrifugation (30 s at $16,000 \times g$). Each fraction was lysed in radioimmune precipitation assay buffer supplemented with protease inhibitors (Complete mini) and sonicated. Protein concentration was determined with the BCA protein assay kit (Pierce). HeLa samples were prepared under reducing conditions (NuPage sample buffer with sample reducing agent; Invitrogen) and heated to 100 °C for 10 min. 10 μg of proteins were loaded and separated on a 4–12% BisTris polyacrylamide NuPage gel (Invitrogen) and transferred onto Hybond nitrocellulose membranes (GE Healthcare) using the XCell™ II blot module (Invitrogen). Membranes were blocked with 5% skimmed milk in TBS-T (Tris-buffered saline Tween 20: 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 0.1% Tween 20) or with 2% fetal calf serum (FCS) in PBS-T (PBS + 0.1% Tween 20) and incubated overnight at 4 °C with a primary antibody raised against the GFP tag (anti-GFP B-2; Santa Cruz Biotechnology) or with a primary antibody raised against MBNL1 (MB1a) (30, 31), β -tubulin (Sigma T4026), or H3 histone (Cell Signaling 9715). Horseradish peroxidase-conjugated antibodies (Vector) were used as secondary antibodies, and horseradish peroxidase activity was detected with the ECL™ detection kit on Hyperfilms (GE Healthcare) according to the manufacturer's instructions. Western blots were reproduced at least three times.

RNA Extraction and Semiquantitative Analysis—Total RNA was isolated from cells 48 h after transfection using a total RNA extraction kit (Nucleospin® RNA II kit; Macherey Nagel, Düren, Germany). RNA concentration was determined by measuring the absorbance at 260 nm using Nanodrop ND1000 technology (Labtech). RT-PCR was performed in triplicate with 1 μg of total RNA using random hexamers (5 μM /liter) and the M-MLV reverse transcriptase (Invitrogen) according to standard protocols. No DNA amplification was observed in RT controls. PCR was carried out in a final volume of 25 μl , with a 10 or 15 μM concentration of each primer as described previously (26, 27), 1.5 mM MgCl_2 , and 1 unit of Taq polymerase (Invitrogen), under the following conditions: 5 min at 94 °C, 22–30 cycles of a 1-min denaturation step at 94 °C, annealing for 2 min at 65 °C, 2 min of extension, and 7 min of final extension at 72 °C. 18 S rRNA was used as an internal control. Reaction products were resolved by electrophoresis using a 5% or 8% polyacrylamide gel, and bands were stained with SYBR Gold (Invitrogen). The intensity of SYBR Gold luminescence was measured using a FluoroImager scanner (Clarisvision). PCR experiments were repeated at least three times. For more details of MBNL1 RT-PCR and two-dimensional gel electrophoresis, see [supplemental Experimental Procedures](#).

Yeast Double- and Triple-hybrid Assays—For yeast double-hybrid assays, the appropriate pACT-II and pAS2 $\Delta\Delta$ plasmids were introduced into haploid *Saccharomyces cerevisiae* test strains (CG929 or Y190 and Y187), which were then crossed. Diploid cells were plated on double- or triple-selection media (-Leu, -Trp, or -Leu, -Trp, and -His), and growth was assessed 3 days later. For yeast triple-hybrid assays, *S. cerevisiae* L40-coat strains carrying various pACTII::MBNL1 plasmids were mated with R40-coat strains carrying the plasmid pIIIMS2-2::5'hcTNT or pIIIMS2-2::CUG₂₁ as a control. Diploids were

selected on -Trp, -Leu, and -Ura media and were tested using -Leu -Ura -His media. For both yeast double-hybrid and yeast triple-hybrid assays, the strength of the interactions tested was evaluated by the addition of different concentrations of 3-amino-1,2,4-triazole (3AT), which competes with the *HIS3* gene product.

Recombinant Protein Preparation—The plasmids pGEX-6P1-MBNL1₄₀, pGEX-6P1-MBNL1 Δ CT3', and pGEX-6P1-MBNL1 Δ CT were constructed for the overexpression of recombinant GST-MBNL1₄₀, GST-MBNL1 Δ CT3', and GST-MBNL1 Δ CT proteins in *E. coli*. To this end, MBNL1 ORFs were PCR-amplified from pET28-MBNL1 plasmids using oligonucleotides that generated a BamHI and an XhoI restriction site, respectively. The amplified DNA fragments were digested and inserted into the pGEX-6P1 plasmid (Novagen) cut by the same two nucleases. The resulting constructs were used to transform *E. coli* BL21 CodonPlus (DE3) RIL cells (Stratagene). Recombinant GST-MBNL1₄₀, GST-MBNL1 Δ CT3', and GST-MBNL1 Δ CT proteins were purified from *E. coli* BL21 cells under native conditions, using glutathione-Sepharose 4B as recommended by the manufacturer (Pharmacia). The GST-MBNL1 proteins bound to glutathione-Sepharose 4B beads were cleaved on the beads using 80 units of PreScission protease (GE Healthcare)/ml of glutathione-Sepharose bead suspension. Cleavage was carried out overnight at 4 °C. The purified proteins were dialyzed against buffer D (150 mM KCl, 1.5 mM MgCl_2 , 0.2 mM EDTA, 20 mM HEPES, pH 7.9, 10% glycerol), and aliquots were stored at -80 °C.

Electrophoresis Mobility Shift Assays—Cold hcTNT RNA was synthesized by *in vitro* transcription of a PCR fragment corresponding to the 50 nucleotides upstream of exon 5 in hcTNT pre-mRNA, using T7 RNA polymerase. The hcTNT RNA substrate used for complex formation was 5' end-labeled using [γ -³²P]ATP and T4 polynucleotide kinase and was gel-purified. For complex formation, 1.2 fmol of 5' end-labeled hcTNT RNA was preincubated in 7 μl of buffer D (150 mM KCl, 4.5 mM MgCl_2 , 20 mM HEPES, pH 7.9, 10% glycerol) for 10 min at 65 °C, in the presence of 2 μg of tRNA, followed by slow cooling. Then, 4 μl of MBNL1 recombinant protein was added to the renatured RNA for a final protein concentration of between 0 and 2,150 nM. The binding reactions were incubated for 30 min at room temperature and loaded onto an 8% (38/2) nondenaturing polyacrylamide gel containing 44.5 mM borate, 1 mM EDTA, 44.5 mM Tris borate, pH 8.3, and 5% glycerol. Gels were run for 45 min at 15 mA. The amount of radioactivity in the bands, corresponding to free and bound complexed RNA, was estimated using a PhosphorImager and ImageQuant software. Using these values, apparent K_d values were determined with SigmaPlot software (SPSS Science Software). The overall K_d and the number of binding sites were determined by Hill plotting; $\log [\text{PR}/\text{R}]$ versus $\log [\text{P}]$ (where P = protein concentration; PR = bound RNA, and R = free RNA).

Statistical Analyses—Statistical analyses were performed using unpaired *t* tests with two tails, with the help of Prism software (GraphPad Software, San Diego, CA).

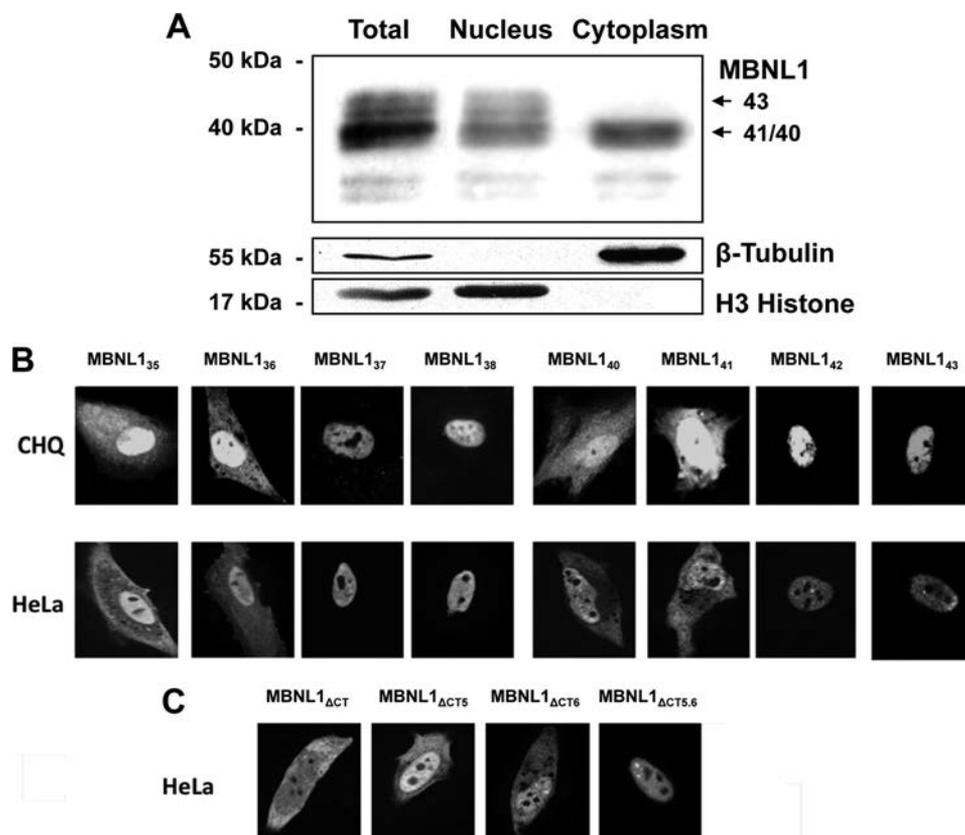


FIGURE 2. Protein sequence encoded by exon 5 modulates MBNL1 nuclear localization. *A*, nucleocytoplasmic localization of endogenous MBNL1 isoforms in human myoblasts. Nuclear fractionation was performed on human myoblasts that endogenously expressed several isoforms of MBNL1. The total fraction (*Total*) is in the *left lane*, the nuclear fraction (*Nucleus*) is in the *middle*, and the cytoplasmic fraction (*Cytoplasm*) is in the *right lane*. Molecular masses are indicated on the *left*, and antibodies used are on the *right*. To visualize MBNL1 isoforms, we used a monoclonal antibody raised against the region encoded by exon 3 (31). β -Tubulin was used as a cytoplasmic marker and H3 histone as a nuclear marker to assess the different fractions. *B*, immunofluorescence of GFP-tagged MBNL1 isoforms in human myoblast and HeLa cells. GFP-tagged MBNL1 isoforms were transiently expressed in human myoblasts and HeLa cells. Fluorescence appears in *gray* and indicates transfected cells and reveals the subcellular localization of MBNL1. The name of the isoform used is indicated at the *top* of each cell, using the same nomenclature as in Fig. 1. *C*, nucleocytoplasmic localization of truncated MBNL1 mutants in HeLa cells. GFP-tagged mutants were transiently expressed in HeLa cells. Fluorescence in *gray* indicates transfected cells and reveals the subcellular localization of truncated MBNL1. The name of the construct used is indicated at the *top* of each cell.

RESULTS

Nuclear Retention of MBNL1 Protein Depends on Segments Encoded by Both Exons 5 and 6—To analyze the subcellular distribution of the various MBNL1 isoforms, nucleocytoplasmic fractionation was performed using human primary muscle cells (Fig. 2*A*). The fractionation protocol was assessed using histone H3 as a nuclear protein marker and β -tubulin as a marker for the cytosolic fraction. MBNL1 isoforms were detected with a monoclonal antibody that specifically recognizes the protein segment encoded by exon 3 (31). After gel electrophoresis, six bands containing proteins ranging from 30 kDa to 45 kDa were observed in the whole cell lysate (Fig. 2*A*, *Total lane*). Based on the epitope recognized by the antibody used, MBNL1 polypeptides detected must contain the exon 3-encoding sequence. According to their apparent molecular masses, the two top bands resolved at 45 kDa were expected to correspond to MBNL1₄₃ and MBNL1₄₂ isoforms, whereas the two middle bands (a major and a minor) at 40 kDa were expected to correspond to the MBNL1₄₁ and MBNL1₄₀ isoforms. The two bottom bands resolved at less than 35 kDa were expected to correspond to proteolytic products because the shortest identified MBNL1 isoform containing exon 3 is

MBNL1₄₀. Based on mRNA analysis, the MBNL1 isoforms 40 S and 41 S were described previously (22); however, they were never identified at the protein level. To support our protein assignment, additional RT-PCR and two-dimensional gel electrophoresis analysis were performed. First, RT-PCR amplification of the MBNL1 mRNAs followed by DNA sequencing identified five distinct mRNAs encoding MBNL1 isoforms 43, 41, 40, 40 S, and 41 S (supplemental Fig. 1*A*). A second Western blot analysis of the two-dimensional gel using MBNL1 antibody revealed three series of four spots aligned in the same range of isoelectric points (supplemental Fig. 1*B*). These spots correspond to MBNL1 isoforms 40, 41, and 43 because they have very close theoretical isoelectric points (pI of 8.9–9.1). The isoforms 40 S and 41 S have theoretical pI values of 8.92–8.96 and a molecular mass of 35–37 kDa, indicating that the isoforms with a molecular mass below 30 kDa and pI values lower than 8.0 were therefore most probably catabolic products of MBNL1 rather than 40 S or 41 S MBNL1 isoforms. Altogether, our results indicate that MBNL1 isoforms 40, 41 and 43 are expressed in human myoblasts.

The MBNL1 isoforms detected in the whole cell lysate were also found in the nuclear fraction (Fig. 2*A*, *Nucleus lane*). How-

Structure-Function Analysis of MBNL1

ever, the upper bands assigned as isoform 43 and containing the exon 5 region were not detected in the cytoplasm. In contrast, the proteins that we assigned to isoforms 41 and 40, which do not contain the exon 5 region, were present in both the cytoplasmic and nuclear fraction. In agreement with this subcellular localization, the exon 5 region was previously proposed to be involved in nuclear localization of MBNL1. Therefore, these data bring additional support to our isoform assignment and extend to human muscle cells the observations previously made in mice (9) and chickens (15).

To complete the analysis of the sequence dependence of MBNL1 nuclear retention, we investigated the nucleocytoplasmic distribution of all WT MBNL1 isoforms, using GFP-tagged constructs (Fig. 1B). First, the eight-tagged isoforms were ectopically expressed in myoblasts and HeLa cells (Fig. 2B). As expected, all of the MBNL1 isoforms containing exon 5 (MBNL1_{37, 38, 42, and 43}) were detected in the nucleus but not in the cytoplasm. Conversely, isoforms lacking this sequence (MBNL1_{35, 36, 40, and 41}) showed both nuclear and cytoplasmic distributions. The localization of MBNL1 isoforms was not modified by the inclusion of exon 3 and/or 7. These results further support the idea that the region encoded by exon 5 regulates the nuclear retention of MBNL1. However, it did not exclude the possibility that other regions of the protein might have contributed to MBNL1 subcellular localization. To test this hypothesis, we generated four GFP-tagged truncated MBNL1 constructs: MBNL1_{ΔCT}, MBNL1_{ΔCT5}, MBNL1_{ΔCT6}, and MBNL1_{ΔCT5,6}, respectively (Fig. 1C). The proteins expressed by these constructs all contained the four zinc finger domains (exons 1, 2, and 4 were present in all of them), with or without exon 5. The truncated protein MBNL1_{ΔCT6} contained, in addition, the entire exon 6 region, but not the exon 5 region, whereas MBNL1_{ΔCT5,6} contained both the exon 5 and 6 regions. These truncated MBNL1 proteins were expressed in HeLa cells, and their subcellular localizations were visualized using GFP fluorescence. Surprisingly, MBNL1_{ΔCT} and MBNL1_{ΔCT5} showed both nuclear and cytoplasmic distribution, even though MBNL1_{ΔCT5} contained the exon 5 region (Fig. 2C). In contrast, the MBNL1_{ΔCT5,6} protein containing both the exon 5 and 6 regions had an exclusive nuclear localization. Furthermore, its variant lacking the exon 5 region had both a nuclear and cytoplasmic distribution. Taken together, our results show that the exon 5 region is necessary but not sufficient for the nuclear retention of MBNL1. The downstream region encoded by exon 6 is necessary to mediate the nuclear retention of MBNL1 isoforms. Therefore, the nuclear retention of MBNL1 depends on both the exon 5 and exon 6 regions.

MBNL1 Exon 3-encoded Region Is Necessary to Modulate Alternative Splicing of Both hcTNT and Human IR Pre-mRNAs—We next investigated whether the splicing function of MBNL1 was isoform-dependent. To assess the splicing activity of the various MBNL1 isoforms, we took advantage of two well characterized MBNL1 targets, hcTNT exon 5 and IR exon 11, and used their respective minigenes (26, 27). The eight MBNL1 isoforms (including or not the exon 3, 5, and 7 regions) were co-expressed with hcTNT or IR minigenes in HeLa cells. The transient expression of MBNL1 constructs yielded very similar

expression levels of MBNL1 proteins (Fig. 3A): only the two largest isoforms (MBNL1₄₂ and MBNL1₄₃) were expressed at slightly lower levels. The splicing products of the hcTNT exon 5 and IR exon 11 minigenes were analyzed by RT-PCR, and the data obtained demonstrated that the presence or absence of the MBNL1 exon 5 and 7 regions had no detectable effect on the MBNL1 splicing regulatory activity (Fig. 3B). In contrast, the presence of the exon 3 region in MBNL1 isoforms was found to be essential for the repression of hcTNT exon 5 inclusion and the activation of IR exon 11 inclusion. Therefore, we concluded that the exon 5 and 7 regions have no obvious splicing regulatory function, in sharp contrast to the exon 3 region (Fig. 3B). To rule out potential bias introduced by the GFP tag, we generated similar constructs in which the GFP tag was replaced by five amino acids (V5 tag). Similar results were obtained with these constructs (data not shown).

Our data suggest that the sequence encoded by exon 3 is essential to the splicing regulatory activity of protein MBNL1. This amino acid sequence links the second and third zinc finger domains (Fig. 1B), and this linker sequence has previously been shown to be required for the efficient binding of MBNL1 to RNAs containing CUG repeats (22). Therefore, we hypothesized that this linker sequence also increases MBNL1 affinity for its target sites in pre-mRNAs. To address this question, a yeast triple-hybrid assay was developed (Fig. 3D). In this assay, the interaction between RNA and protein was detected by the activity of the *HIS3* reporter gene. The strength of the interaction was assessed by the addition of increasing concentrations of 3AT to the medium. 3AT is a competitive inhibitor of the activity of the *HIS3* gene product. The MBNL1-Gal4 transcriptional activation domain fusion proteins used in these assays corresponded to: (i) the MBNL1₄₀ isoform, (ii) its homolog lacking exon 3, MBNL1₃₅, (iii) the MBNL1₄₁ isoform, (iv) its homolog lacking exon 3, MBNL1₃₆, (v) MBNL1_{ΔCT3} including the exon 3 region, and (vi) MBNL1_{ΔCT} lacking the exon 3 region. The RNA used in the tests corresponded to the 3' end of the fourth intron of hcTNT pre-mRNA (hcTNT RNA). This RNA has been previously shown to interact with the MBNL1₄₀ isoform (16, 19). The truncation of the C-terminal region in MBNL₄₀ did not change the affinity for the hcTNT RNA (Fig. 3D). Conversely, this affinity was strongly reduced in the absence of the exon 3 region (proteins MBNL1₃₅, MBNL1₃₆, and MBNL1_{ΔCT}). Indeed, almost no interaction of MBNL1₃₅ and MBNL1_{ΔCT} was detected at a 3AT concentration of 0.5 mM (Fig. 3D). Interestingly, at 1 mM 3AT concentration, a stronger interaction was observed for MBNL1₄₁ including the exon 7 region compared with MBNL1₄₀ missing this region. In addition, at this 3AT concentration, some remaining interaction was also detected for MBNL1₃₆ compared with MBNL1_{ΔCT}. Therefore, the exon 7 region may partially compensate for the absence of exon 3 (Fig. 3D). One possible explanation will be given in a following paragraph describing properties of the exon 7 region.

We then confirmed the role of exon 3 region in MBNL1 binding to its target sites in hcTNT pre-mRNAs by gel shift experiments using a 5' end-labeled hcTNT RNA (1.2 fmol) and the recombinant proteins MBNL1₄₀, MBNL1_{ΔCT3}, and MBNL1_{ΔCT} at concentrations ranging from 100 to 2,150 nM

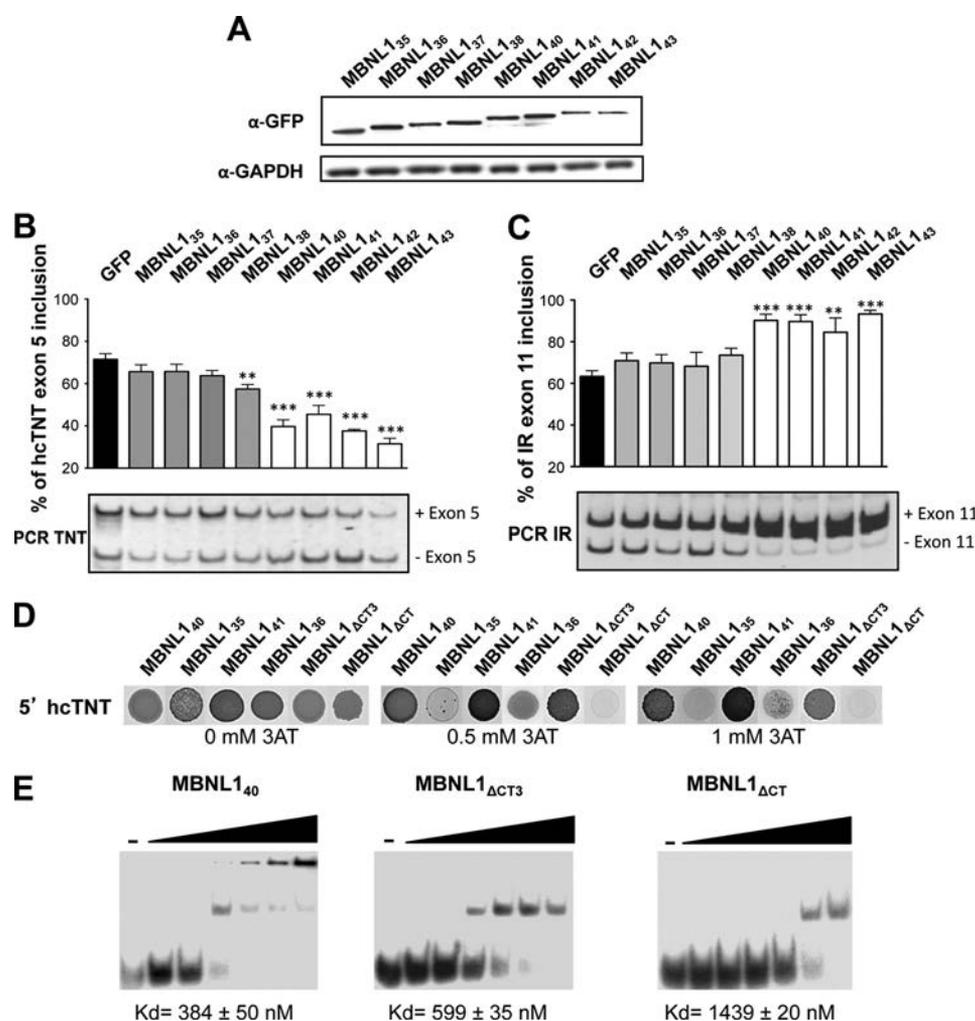


FIGURE 3. Protein sequence encoded by exon 3 modulates MBNL1 splicing and RNA binding properties. *A*, eight GFP-tagged MBNL1 isoforms were transiently expressed in HeLa cells. The level of expression among the different isoforms was verified to be the same by Western blotting, using an antibody raised against the GFP tag. GAPDH was used as a loading control. *B* and *C*, MBNL1 isoforms were co-expressed in HeLa cells in the presence of hcTNT exon 5 (*B*) or IR exon 11 (*C*) minigenes. The name of each isoform is indicated at the top. Cells were lysed 48 h after transfection, and splicing products were assayed by RT-PCR. Data are expressed as means \pm S.E. (error bars). The percentage of inclusion is calculated as follows: $((\text{mRNA} + \text{exon})/(\text{mRNA} + \text{exon}) + (\text{mRNA} - \text{exon})) \times 100$. Results are derived from at least three independent experiments. An unpaired *t* test with two tails was used to compare values. Statistically significant results are indicated by ** ($p < 0.001$) and *** ($p < 0.0001$). *D*, triple-hybrid assays were performed using MBNL1₄₀, MBNL1₃₅, MBNL1₄₁, MBNL1₃₆, MBNL1 Δ CT, and MBNL1 Δ CT₃ fusion proteins and 5' hcTNT as the bait RNA. The strength of the interaction was evaluated by the addition of 3AT at a concentration of 0.5 or 1 mM. *E*, ribonucleoprotein complexes were formed by the incubation of 1.2 fmol of 5'-labeled hcTNT RNA in the presence of MBNL1₄₀, MBNL1 Δ CT, or MBNL1 Δ CT₃ recombinant proteins at a concentration of 100, 215, 430, 750, 1,000, or 2,150 nM. All complexes were formed at room temperature in the presence of 2 μ g of yeast tRNAs under the conditions described under "Experimental Procedures." Lanes (–) are control experiments in the absence of protein. Complexes were fractionated by electrophoresis in a nondenaturing 8% polyacrylamide gel.

(for incubation conditions, see "Experimental Procedures"). No gel shift experiments were performed with the MBNL₄₁, MBNL₃₆, and MBNL₃₅ recombinant proteins because of their very low solubility during the purification steps. In agreement with the above triple-hybrid data, protein MBNL1 Δ CT, lacking the exon 3 region, bound to the hcTNT RNA with a lower affinity (apparent K_d 1,439 \pm 20 nM) compared with protein MBNL1 Δ CT₃, containing the exon 3 region (apparent K_d 599 \pm 35 nM) (Fig. 3E). The MBNL1₄₀ protein isoform, containing both the exon 3 region and additional C-terminal sequences, had a slightly lower apparent K_d (384 \pm 50 nM). In addition, due to its aggregation at high concentration, RNP complexes formed with this protein were retained in the loading well at high protein concentrations.

For the three proteins studied, the curves obtained by plotting the ratio of bound versus bound plus unbound RNA (PR/

PR+R) against the protein concentration [P] (where PR = bound RNA and R = free RNA) indicated the presence of multiple binding sites that might be filled either individually, or in a co-operative manner because no binding intermediates were visible (supplemental Fig. 2A). By determining the overall K_d and studying the binding co-operativity by Hill plots (log (PR/R) versus log [P]) (supplemental supplementary Fig. 2), we concluded that each of the studied proteins (MBNL1₄₀, MBNL1 Δ CT₃, and MBNL1 Δ CT) binds co-operatively on two MBNL1 binding sites which are present in the hcTNT RNA.

Taken together, these results indicate that the exon 3 region is not absolutely required for MBNL1 RNA binding activity. However, in its absence, the affinity of MBNL1 for its target site in hcTNT pre-mRNA is reduced by a factor of about 2. Although we cannot exclude a possible interaction of the exon 3 region with splicing factors, the decreased RNA affinity of

Structure-Function Analysis of MBNL1

MBNL1 isoforms lacking this region likely strongly participates in their reduced capabilities to regulate splicing.

Exon 6 Region Is Also Needed for the Splicing Regulatory Property of MBNL1—Having shown the importance of the alternative MBNL1 exon 3 region for MBNL1 splicing regulatory activity and the absence of the involvement of the amino acid regions coded by the two other alternative exons (5 and 7) in this process, we decided to investigate whether the amino acid regions encoded by some of the constitutive exons are also involved in splicing regulation. The contribution of the regions encoded by exons 1, 2, and 4 is obvious because they correspond to the two pairs of zinc finger domains needed for RNA binding (25). Therefore, we focused our study on the possible role of the exon 6 region, which was present in all of the isoforms tested for their capabilities to regulate splicing in the above experiments (Figs. 1B and 3A). To test for a possible effect of this region in both the presence or and the absence of the exon 3 or 5 region, we prepared a series of constructs expressing proteins lacking exon 3 and/or 5 region(s) (MBNL1 Δ CT, MBNL1 Δ CT3, MBNL1 Δ CT6, MBNL1 Δ CT5,6, MBNL1 Δ CT3,6, and MBNL1 Δ CT3,5,6 (Fig. 1C). These constructs were co-transfected in HeLa cells together with the hcTNT or IR minigenes, and the expression of the truncated proteins was verified by Western blot analysis (Fig. 4A). Minigene splicing products were assessed by RT-PCR 48 h after transfection. As expected based on the above results, MBNL1 variants lacking the exon 3 region had no detectable effect on the alternative splicing of the hcTNT exon 5 and IR exon 11 minigenes (Fig. 4, B and C). Interestingly, the presence of the exon 3 region conferred some splicing regulatory activity on the MBNL1 protein even in the absence of the exon 5 and 6 regions, revealing some splicing regulatory properties of the Δ CT3 N-terminal part of MBNL1. However, the simultaneous presence of both exons 3 and 6 regions was needed for recovering a full splicing activity (Fig. 4, B and C). In agreement with the above results, the deletion or addition of the exon 5 region had no significant effect on MBNL1 splicing regulatory properties. Similar results were obtained with V5-tagged MBNL1 isoforms (data not shown). Combined together, our results reveal the essential role of both the alternative exon 3 region and the constitutive exon 6 region for full MBNL1 splicing regulation activity.

All MBNL1 Isoforms Are Trapped in DM1 Foci—One of the primary causes of DM1 pathophysiology is the abnormal sequestration of MBNL1 by CUG repeats. To determine whether the trapping of MBNL1 in nuclear aggregates that is triggered by expanded CUG repeats was dependent on the presence of the exon 3-encoded region, we examined the ability of each MBNL1 isoform to co-localize with *foci* in primary muscle cells isolated from a muscle biopsy of a DM1 patient. CUG-RNA *foci* were visualized by FISH using a (CAG)₇-Cy3 probe. Various MBNL1 isoforms were ectopically expressed in their wild-type form and visualized using their GFP tag.

The whole GFP-tagged MBNL1 isoforms co-localized with discrete *foci* containing CUG-RNA repeats, indicating that MBNL1 isoforms with or without the exon 3 region were sequestered by CUG repeats (Fig. 5). This was surprising, considering the previously described low affinity of MBNL1

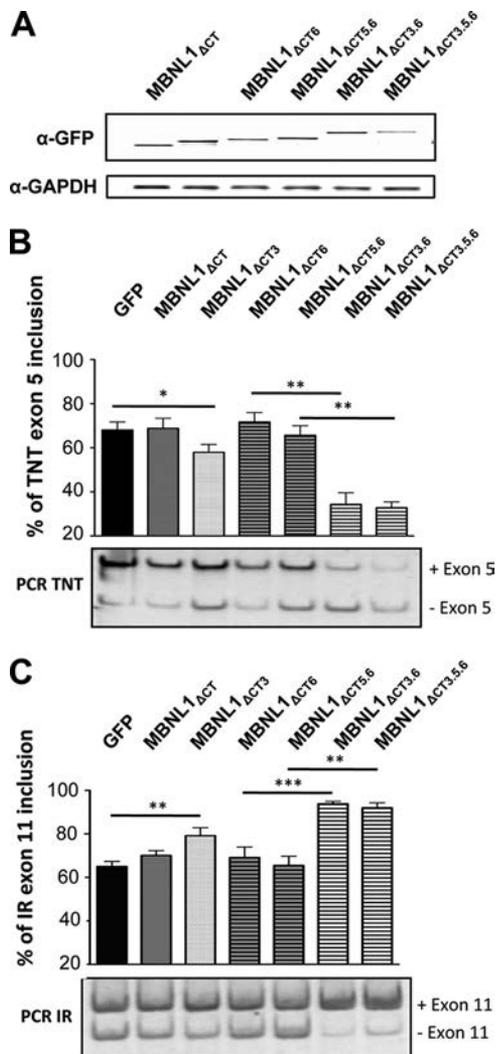


FIGURE 4. MBNL1 splicing function requires the region encoded by MBNL1 exon 6. A, truncated MBNL1 constructs were transiently expressed in HeLa cells. The name of each construct is indicated at the top. The regions expressed are indicated, and the presence of alternative exons is shown in *subscripts*. The level of expression of the different isoforms was verified to be the same by Western blotting using an antibody raised against the GFP tag. GAPDH was used as a loading control. B and C, truncated MBNL1 constructs were co-expressed in HeLa cells in the presence of hcTNT exon 5 (B) or IR exon 11 (C) minigenes. The name of each construct is indicated at the top and the regions expressed with the presence of alternative exons in *subscripts*. Cells were lysed 48 h after transfection, and splicing products were assayed by RT-PCR. Data are expressed as indicated for Fig. 3. An unpaired *t* test with two tails was used to compare values. Statistically significant results are indicated by * ($p < 0.05$), ** ($p < 0.001$), and *** ($p < 0.0001$). Error bars, S.E.

isoforms lacking the exon 3 region for CUG repeats (22). Noteworthy, the dual nucleocytoplasmic localization of the MBNL1_{35,36,40,41} isoforms was maintained even in presence of *foci* (Fig. 5).

Our detection of all isoforms of MBNL1 in *foci* containing CUG expansions could be explained by the capability of all isoforms to interact directly with the CUG RNA expansions or by their capability to establish protein-protein interactions. To get insight into this question, we first compared the affinity for CUG repeats of all MBNL1 isoforms, using triple-hybrid assays in the same conditions as in Fig. 3D, but by replacing the hcTNT RNA by an RNA containing 21 successive CUG triplets (CUG₂₁ RNA; Fig. 6A). All MBNL1 isoforms interacted with

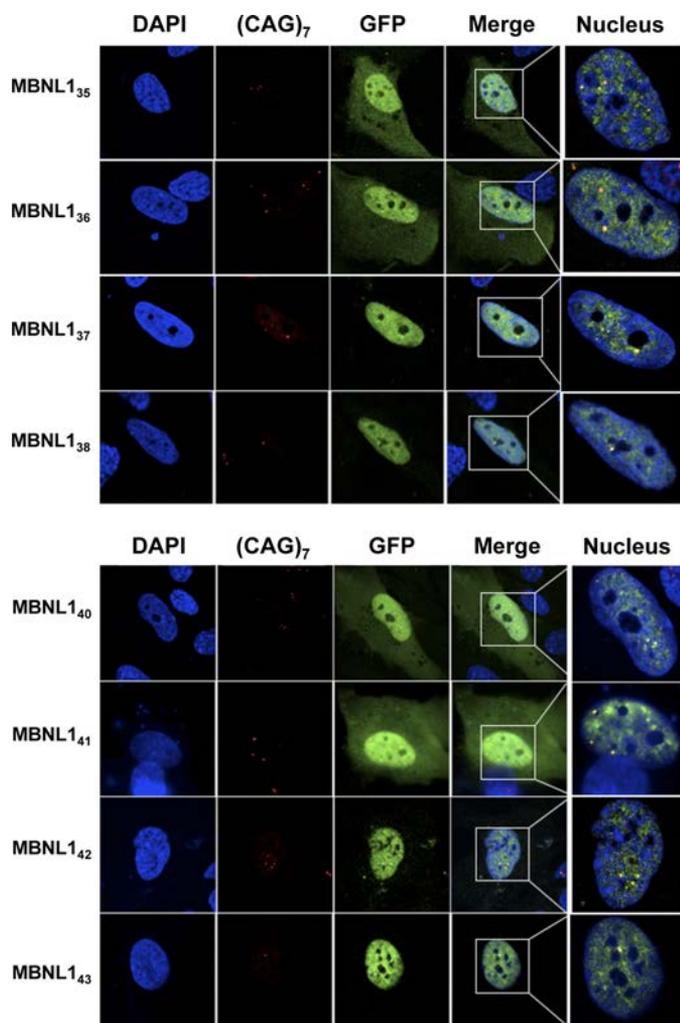


FIGURE 5. All MBNL1 isoforms are trapped by the CUG repeats. Eight GFP-tagged MBNL1 isoforms (MBNL1_{35,36,37,38,40,41,42,43}) were transiently expressed in DM1 myoblasts. Green fluorescence indicates transfected cells. CUG repeats appear in red after FISH using a (CAG)₇-Cy3 probe. Merged images show that all MBNL1 isoforms co-localize with the foci of expanded DMPK transcripts. A higher magnification of nuclei is also showed in the fifth column.

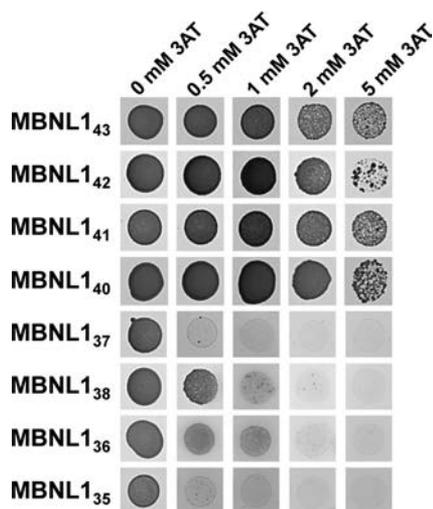


FIGURE 6. MBNL1 isoforms that include the exon 3 region bind to CUG₂₁ RNA with higher affinity than those lacking this region. Triple-hybrid assays were performed using CUG₂₁ RNA as the bait RNA and all MBNL1 isoforms as fusion proteins. The strength of the interaction was evaluated by the addition of 3AT at a concentration of 0.5, 1, 2, or 5 mM.

this RNA in the absence of 3AT. In agreement with data in Fig. 3D, all isoforms MBNL1_{40,41,42,43} containing the exon 3 region had a stronger affinity for the CUG₂₁ RNA than the ones missing this region (MBNL1 isoforms 35–37). Indeed, an interaction was detected up to 5 mM 3AT for MBNL1 isoforms 40–43 whereas this interaction was abolished at 2 mM 3AT for MBNL1 isoforms 35–37 (Fig. 6A). Similar to data obtained for the hcTNT RNA (Fig. 3D), the presence of the exon 7 region was found to increase the capability of MBNL1 isoforms missing the exon 3 region (MBNL1₃₈ and MBNL1₃₆) to interact with CUG₂₁ RNA (Fig. 6A). One possible explanation for this latter observation was that by enhancing MBNL1-MBNL1 interactions, the exon 7 region reinforces the co-operative binding of MBNL1 to the RNA. We therefore undertook a study of homo- and heterodimerization of the WT MBNL1 isoforms.

Exon 7-Encoded Sequence Contributes to MBNL1 Dimerization—Previous studies performed in yeast using two-hybrid assays or in mammalian cells have shown that MBNL1 self-interacts via its C-terminal region (amino acids 239–382) (32). The C-terminal domains of MBNL1 isoforms mainly differ one from the other by the presence or absence of the regions encoded by exons 5 and 7. To test whether these two amino acid sequences or constitutive amino acid sequences in the MBNL1 C-terminal region are implicated in protein dimerization, we performed two-hybrid assay experiments in yeast.

Each of the eight full-length MBNL1 isoforms studied above and each of the four different MBNL1 C-terminal regions were used in the assays. We also used the truncated MBNL1_{ΔCT} and MBNL1_{ΔCT3} versions of MBNL1, as controls of MBNL1 isoforms lacking the C-terminal domain. The four C-terminal domains all contained part of the exon 4 region located downstream to the zinc finger 4 and the exon 6, 8, and 10 regions with or without the exon 5 and/or 7 regions (MBNL1_{CT}, MBNL1_{CT5}, MBNL1_{CT7}, and MBNL1_{CT5,7}) (Figs. 1 and 7). To test the interaction capabilities of each protein pair in the two possible orientations, recombinant two-hybrid pAS2 and pACT2 plasmids, allowing the expression of the Gal4 DNA binding domain or Gal4 transcriptional activation domain in fusion with each of the tested proteins, were produced (Fig. 7). As for the three-hybrid experiments, protein-protein interactions were detected by the activity of the *HIS3* reporter gene, and their interaction strengths were challenged by the addition of increasing concentrations of 3AT. We tested the capability of homo- and heterodimerization of (i) the C-terminal domains and (ii) the C-terminal domain with full-length MBNL1 isoforms and MBNL1_{ΔCT} and MBNL1_{ΔCT3} proteins.

No interaction of the MBNL1_{ΔCT} and MBNL1_{ΔCT3} proteins was detected with any of the C-terminal domains or full-length MBNL1 proteins, even in the absence of 3AT. This result demonstrated the absence of interaction of the MBNL1 Zinc finger domains with the C-terminal domain (Fig. 7). In contrast, in the absence of 3AT, several of the C-terminal regions formed homo- or heterodimers with C-terminal domains or full-length proteins. The weakest interactions were detected for the all MBNL1 isoforms, such as MBNL1_{CT} and MBNL1_{CT5}, which lack the exon 7 region (Fig. 7, blue squares compared with orange squares). These interactions almost disappeared at a 2

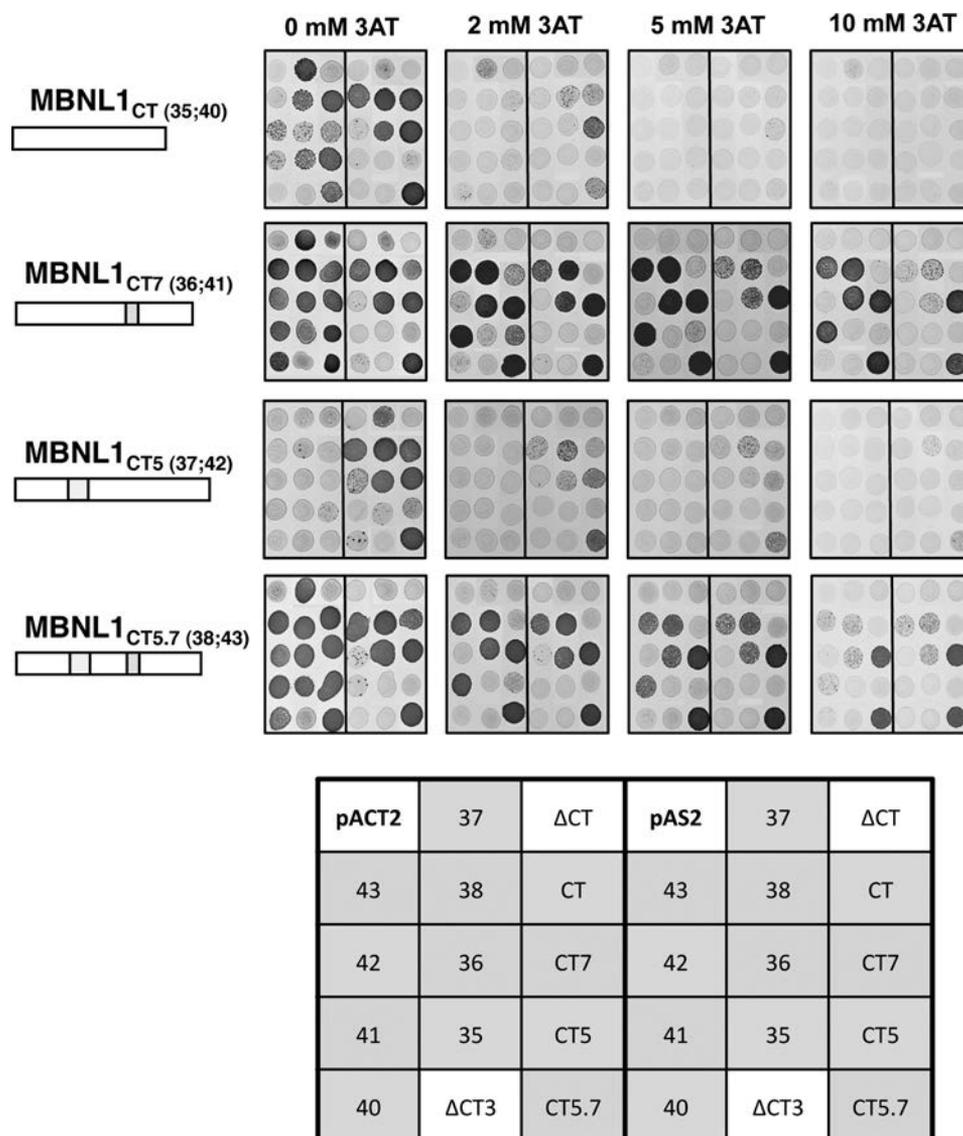


FIGURE 7. Protein sequence encoded by exon 7 modulates MBNL1 dimerization properties. The interaction of the four MBNL1 C-terminal tails (A–D) with all MBNL1 isoforms and with themselves was assessed by yeast two-hybrid assay. cDNAs encoding the proteins of interest were PCR-amplified and cloned into both pAS2 and pACT2 vectors, containing the GAL4 DNA binding domain and the GAL4 activation domain, respectively. Protein-protein interactions were detected by the activity of the *HIS3* reporter gene and the strength of the interaction assessed by the addition of incremental concentrations of 3AT. In each panel, the three left columns correspond to the MBNL1 C-terminal tail cloned into the pAS2 vector, whereas the three right columns correspond to the pACT2 clone. The presence of exon 5- and exon 7-encoded sequences in the C-terminal tails is schematically indicated by gray boxes, and the name of the full-length isoforms containing a given C-terminal tail is indicated in brackets. The identity of bait proteins is indicated in the scheme on the bottom, and the presence of the exon 7-encoded sequence corresponds to squares CT5.7, CT7, 36, 37, 38, 41, and 43 whereas the isoforms lacking the exon 7 correspond to squares CT, CT5, 35, 37, 40, and 42. The controls are indicated in the white squares.

mM 3AT concentration. At 2 mM 3AT, these two MBNL1_{CT} and MBNL1_{CT5} C-terminal regions only showed a low level of interaction with the C-terminal regions containing the exon 7 or both the exon 5 and 7 regions. In contrast, some of the interactions formed by the MBNL1_{CT7} and MBNL1_{CT5.7} C-terminal domains (CT7 or CT5.7) containing the exon 7 region and the full-length proteins containing one of these two C-terminal domains (36, 38, 41, and 43) were highly stable. They were detected at 3AT concentrations up to 5 or 10 mM (Fig. 7). Altogether, these data confirmed the fact that only the C-terminal domain of MBNL1 is involved in MBNL1 dimerization and revealed the previously unknown importance of the exon 7 region in this process.

DISCUSSION

Alternative splicing produces several MBNL1 isoforms, differing mainly in the inclusion or exclusion of the in-frame sequences encoded by exons 3, 5, and 7. The exon 3 region has already been proposed to modulate RNA binding affinity, and here we provide new information as to how the presence or absence of the exon 3 region modifies the binding of MBNL1 to its target sites in pre-mRNA and to CUG expansions. We show for the first time the crucial role of this region in the splicing regulatory property of MBNL1. The exon 5 region has previously been proposed to play a critical role in the nuclear retention of the MBNL1 protein. Here, we show that this nuclear retention also depends on amino acid sequences encoded by

exon 6 and that the exon 6 region is required for splicing regulation. Finally, we have identified the role of the alternative exon 7 region in MBNL1 dimerization and showed also that exon 8 is alternative and not constitutive. We will now discuss the implications of our findings for the normal and pathological activities of the MBNL1 splicing factor.

Nuclear Localization of MBNL1 and Splicing Activity—MBNL1 isoforms displayed different nucleocytoplasmic localization patterns in HeLa cells and in human myoblasts. Isoforms lacking the exon 5 region were distributed in both the nucleus and the cytoplasm whereas the subcellular localization of those including the exon 5 region was restricted to the nucleus. Cassette exon 5 therefore probably encodes a region involved in the nuclear localization of MBNL1. This corroborates previously published data (9, 15, 24) and extends them to human myoblasts.

Because all MBNL1 isoforms were located in the nucleus, with or without the inclusion of the alternative exon 5 region, we can rule out the idea that this region is solely driving the nuclear localization of MBNL1. This observation was further confirmed using truncated proteins with or without the exon 5 region and lacking the C-terminal region. Indeed, diffuse staining for the truncated proteins MBNL1 $_{\Delta CT}$ and MBNL1 $_{\Delta CT5}$ was present in both the nucleus and the cytoplasm. The apparent molecular size of these truncated proteins fused to the GFP tag was around 60 kDa, excluding passive diffusion through the nuclear pore complex. Thus, the RNA binding domain composed of four CCCH zinc finger motifs was sufficient to target MBNL1 to the nucleus. Furthermore, this result also suggests that the MBNL1 exon 5 region co-operates with another region to maintain MBNL1 in the nucleus. In line with our hypothesis, we observed that both exons 5 and 6 were required to induce a nuclear localization of MBNL1 exclusively, as seen in Fig. 2C. Interestingly, a new type of highly conserved nuclear localization signal motif (the KRAEK motif) has recently been identified in *D. melanogaster* Muscleblind-like C and E isoforms (33). This motif has been shown to be sufficient to target Muscleblind-like to the nucleus of *D. melanogaster* Schneider 2 cells. Although this motif is conserved in the exon 6 region of human MBNL1, but not in exon 5, our data strongly suggest that additional regions might be required for MBNL1 nuclear retention. Thus, the nuclear retention of MBNL1, according to our results, necessitates a bipartite signal with one motif in the exon 5 and the second in the exon 6, but separately, these motifs are nonfunctional. We suggest that the MBNL1 exon 6 region contains a nuclear localization signal that is either functional in the presence of the exon 5 region or is regulated by an adjacent upstream sequence. Importantly, we cannot exclude the possibility that post-translational modification around the exon 5 sequence also regulates the subcellular localization of MBNL1.

Because splicing occurs in the nucleus, we asked whether there was a relationship between strict nuclear localization and the splicing activity of MBNL1. Surprisingly, with a similar level of expression, we observed no difference in the splicing activity of the different MBNL1 isoforms on hCNT exon 5 or IR exon 11 splicing, as assayed using minigenes. This result raises the question of the interest of a restrictive nuclear localization. One possibility is that nuclear retention is a cellular mechanism that impedes MBNL1 cytoplasmic activity while maintaining its

splicing activity when located in the nucleus. Few studies have reported the potential cytoplasmic functions of MBNL1; judging by its similarity to MBNL2, another member of the MBNL protein family, MBNL1 could be involved in mRNA transport and/or stability (34). In line with this observation, some recent studies have demonstrated a global change in gene expression in MBNL1 knock-out mice (11, 35). Stability and translatability may also be regulated by MBNL1 under conditions of stress, in which MBNL1 co-localizes with stress granules (36). In this latter study, the authors used the isoform MBNL1 $_{40}$, which displayed both nuclear and cytoplasmic localization at base line. However, this property cannot be extrapolated to other isoforms of MBNL1, especially to MBNL1 isoforms having the exon 5 and 6 sequences. Additional work is needed to determine whether MBNL1 isoforms located only in the nucleus also co-localize with stress granules under conditions of stress. Accordingly, the function of MBNL1 isoforms lacking exon 3 also remains to be established.

MBNL1 Splicing Activity Depends on Both Exon 3- and Exon 6-Encoded Regions—MBNL1 splicing factor can enhance or repress the inclusion of alternative exons (16). Based on our data, when an isoform is active in splicing, it can have a silencing or an enhancing activity on splicing, depending on the targeted splicing site. Our results exclude the possibility that some isoforms may be dedicated to splicing inhibition, whereas others would have splicing activation properties. As suggested previously, MBNL1 splicing activity might rather depend on the localization of its RNA-binding elements (17). Indeed, the MBNL1 exon 3 region is located between the zinc finger 2 and zinc finger 3 motifs and acts as a linker between the zinc finger 1/2 and zinc finger 3/4 domains that structure the MBNL1 RNA binding domain (22, 25) and probably modulate its RNA binding affinity. Triple-hybrid assays and gel shift experiments confirmed that the presence of the exon 3 region was essential for MBNL1 efficient binding to its pre-mRNA target. Moreover, we observed that all MBNL1 isoforms that included the exon 3 region (MBNL1 $_{40}$, MBNL1 $_{41}$, MBNL1 $_{42}$, and MBNL1 $_{43}$) strongly modified MBNL1 splicing activity with respect to hCNT exon 5 and IR exon 11, whereas those that lacked exon 3 (MBNL1 $_{35}$, MBNL1 $_{36}$, MBNL1 $_{37}$, and MBNL1 $_{38}$) barely did. Also in accordance with the need for an RNA binding domain for MBNL1 activity, experiments performed with a truncated protein revealed that the splicing activity of the exon 3 region was weak but accurate (Fig. 4, MBNL1 $_{\Delta CT3}$). The recovery of full splicing activity was observed with proteins that expressed both the exon 3 and exon 6 regions (Fig. 4, MBNL1 $_{\Delta CT3}$). This result suggests that MBNL1 splicing activity mainly resides in the exon 6 region. Furthermore, it also suggests that MBNL1 splicing activity likely depends on its affinity for its RNA targets and the presence of both exon 3 and 6 regions.

Relevance to DM1 Pathology—Myotonic dystrophy type 1 belongs to the triplet repeat disorder family. The unstable CTG expansions are transcribed into RNA and retained in the nucleus. These toxic RNAs have been shown to sequester MBNL1 proteins (8, 37). There are various MBNL1 protein isoforms, which in our study differed in the presence or absence of the exon 3, 5, and 7 regions. All of them co-localized with the CUG repeats expressed in DM1 myoblasts. The MBNL1 RNA

Structure-Function Analysis of MBNL1

binding domain alone (MBNL1_{ΔCT}, MBNL1_{ΔCT3}) was sufficient to mediate MBNL1 binding to the CUG repeats, as previously suggested (19), with or without the exon 3 region. Furthermore, using the yeast triple-hybrid system, we observed that the presence of the exon 3 region strongly enhanced the binding affinity of MBNL1 to the CUG repeats. Therefore, MBNL1 isoforms that lack the exon 3 region likely interact with CUG repeats very weakly. We thus suggest that their sequestration is not direct but mediated by the dimerization of MBNL1 with itself. The self-dimerization of MBNL1 has been previously reported and shown to occur via the C-terminal tail of MBNL1 (32). In our study, double-hybrid assays performed with the MBNL1 C-terminal tail revealed that these C-terminal fragments only dimerized with the full-length MBNL1. The presence of exon 7 also improved MBNL1 binding on both CUG repeats and hcTNT mRNA, most probably by promoting MBNL1 dimerization. Interestingly, the expression of MBNL1 protein isoforms, including the exon 7, region is greatly increased with the development of DM1 (9, 24). These isoforms could therefore contribute to the focal localization of all MBNL1 isoforms, with or without the exon 3 region. Altogether, our results suggest that the binding of MBNL1 proteins to the CUG repeats could either be direct or indirect, through the dimerization of MBNL1 with itself. Thus, the MBNL1 missplicing observed in DM1 could enhance its own sequestration in the *foci*.

CONCLUSION

Altogether, our study reveals the importance of the alternative splicing of MBNL1 in the post-transcriptional regulation of the subcellular localization and function of MBNL1. We attempted to understand better the function of the various MBNL1 isoforms. Our results show that (i) the presence of the exon 3 region is essential to ensure a high binding affinity of MBNL1 for pre-mRNA, and thus its splicing activity; (ii) the presence of the exon 5 region ensures a restrictive nuclear localization; (iii) exons 3 and 5 are fully functional only in the presence of the exon 6 region; and (iv) the exon 7 region modulates the dimerization properties of MBNL1. Our data therefore indicate that association of MBNL1 with CUG repeats depends both on RNA-protein interactions and on self-dimerization. During fetal development and in DM1 pathophysiology, both MBNL1 exons 5 and 7 are included. This could contribute to the exclusively nuclear localization of MBNL1 and to its self-dimerization, two key features for the sequestration within *foci*. However, further studies will be needed for a better understanding of all the molecular consequences of MBNL1 sequestration in DM1. This knowledge is important to define precisely the mechanism for a potential therapeutic development.

Acknowledgments—We thank Dr. Tom Cooper for the hcTNT minigene and Dr. S Reddy for the IR minigene. We acknowledge the Muscular Dystrophy Association and the Monoclonal Antibody Resource.

REFERENCES

- Black, D. L. (2003) *Annu. Rev. Biochem.* **72**, 291–336
- Ranum, L. P., and Cooper, T. A. (2006) *Annu. Rev. Neurosci.* **29**, 259–277
- Lee, J. E., and Cooper, T. A. (2009) *Biochem. Soc. Trans.* **37**, 1281–1286
- Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., et al. (1992) *Cell* **69**, 385
- Fu, Y. H., Pizzuti, A., Fenwick, R. G., Jr., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., de Jong, P., et al. (1992) *Science* **255**, 1256–1258
- Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barceló, J., and O'Hoy, K. (1992) *Science* **255**, 1253–1255
- Mooers, B. H., Logue, J. S., and Berglund, J. A. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16626–16631
- Fardaei, M., Larkin, K., Brook, J. D., and Hamshere, M. G. (2001) *Nucleic Acids Res.* **29**, 2766–2771
- Lin, X., Miller, J. W., Mankodi, A., Kanadia, R. N., Yuan, Y., Moxley, R. T., Swanson, M. S., and Thornton, C. A. (2006) *Hum. Mol. Genet.* **15**, 2087–2097
- Kanadia, R. N., Johnstone, K. A., Mankodi, A., Lungu, C., Thornton, C. A., Esson, D., Timmers, A. M., Hauswirth, W. W., and Swanson, M. S. (2003) *Science* **302**, 1978–1980
- Du, H., Cline, M. S., Osborne, R. J., Tuttle, D. L., Clark, T. A., Donohue, J. P., Hall, M. P., Shiue, L., Swanson, M. S., Thornton, C. A., and Ares, M., Jr. (2010) *Nat. Struct. Mol. Biol.* **17**, 187–193
- Begemann, G., Paricio, N., Artero, R., Kiss, I., Pérez-Alonso, M., and Mlodzik, M. (1997) *Development* **124**, 4321–4331
- Artero, R., Prokop, A., Paricio, N., Begemann, G., Pueyo, I., Mlodzik, M., Perez-Alonso, M., and Baylies, M. K. (1998) *Dev. Biol.* **195**, 131–143
- Kalsotra, A., Xiao, X., Ward, A. J., Castle, J. C., Johnson, J. M., Burge, C. B., and Cooper, T. A. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20333–20338
- Terenzi, F., and Ladd, A. N. (2010) *RNA Biol.* **7**, 43–55
- Ho, T. H., Charlet, B. N., Poulos, M. G., Singh, G., Swanson, M. S., and Cooper, T. A. (2004) *EMBO J.* **23**, 3103–3112
- Goers, E. S., Purcell, J., Voelker, R. B., Gates, D. P., and Berglund, J. A. (2010) *Nucleic Acids Res.* **38**, 2467–2484
- Warf, M. B., Diegel, J. V., von Hippel, P. H., and Berglund, J. A. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9203–9208
- Warf, M. B., and Berglund, J. A. (2007) *RNA* **13**, 2238–2251
- Sen, S., Talukdar, I., Liu, Y., Tam, J., Reddy, S., and Webster, N. J. (2010) *J. Biol. Chem.* **285**, 25426–25437
- Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshere, M. G., Harper, P. S., and Brook, J. D. (2002) *Hum. Mol. Genet.* **11**, 805–814
- Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N., and Ishiura, S. (2004) *Hum. Mol. Genet.* **13**, 495–507
- Pascual, M., Vicente, M., Monferrer, L., and Artero, R. (2006) *Differentiation* **74**, 65–80
- Dhaenens, C. M., Schraen-Maschke, S., Tran, H., Vingdeux, V., Ghanem, D., Leroy, O., Delplanque, J., Vanbrussel, E., Delacourte, A., Vermersch, P., Maurage, C. A., Gruffat, H., Sergeant, A., Mahadevan, M. S., Ishiura, S., Buée, L., Cooper, T. A., Caillet-Boudin, M. L., Charlet-Berguerand, N., Sablonnière, B., and Sergeant, N. (2008) *Exp. Neurol.* **210**, 467–478
- Teplava, M., and Patel, D. J. (2008) *Nat. Struct. Mol. Biol.* **15**, 1343–1351
- Cooper, T. A. (1998) *Mol. Cell. Biol.* **18**, 4519–4525
- Kosaki, A., Nelson, J., and Webster, N. J. (1998) *J. Biol. Chem.* **273**, 10331–10337
- Edom, F., Mouly, V., Barbet, J. P., Fiszman, M. Y., and Butler-Browne, G. S. (1994) *Dev. Biol.* **164**, 219–229
- Taneja, K. L. (1998) *BioTechniques* **24**, 472–476
- Holt, I., Mittal, S., Furling, D., Butler-Browne, G. S., Brook, J. D., and Morris, G. E. (2007) *Genes Cells* **12**, 1035–1048
- Holt, I., Jacquemin, V., Fardaei, M., Sewry, C. A., Butler-Browne, G. S., Furling, D., Brook, J. D., and Morris, G. E. (2009) *Am. J. Pathol.* **174**, 216–227
- Yuan, Y., Compton, S. A., Sobczak, K., Stenberg, M. G., Thornton, C. A., Griffith, J. D., and Swanson, M. S. (2007) *Nucleic Acids Res.* **35**, 5474–5486
- Fernandez-Costa, J. M., and Artero, R. (2010) *Mol. Cells* **30**, 65–70
- Adereth, Y., Dammai, V., Kose, N., Li, R., and Hsu, T. (2005) *Nat. Cell Biol.* **7**, 1240–1247
- Osborne, R. J., Lin, X., Welle, S., Sobczak, K., O'Rourke, J. R., Swanson, M. S., and Thornton, C. A. (2009) *Hum. Mol. Genet.* **18**, 1471–1481
- Onishi, H., Kino, Y., Morita, T., Futai, E., Sasagawa, N., and Ishiura, S. (2008) *J. Neurosci. Res.* **86**, 1994–2002
- Miller, J. W., Urbinati, C. R., Teng-Umuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., and Swanson, M. S. (2000) *EMBO J.* **19**, 4439–4448

IV.2. Rôle des paires de doigts à zinc dans la reconnaissance des ARN

La protéine Mbl de drosophile ne contient qu'une seule paire de doigts à Zn^{2+} contrairement à la protéine MBNL1 humaine qui en contient deux. Or, Mbl a la capacité de se lier sur les pré-ARNm cibles de MBNL1 et de réguler leur épissage alternatif. Elle se fixe aussi sur des répétitions CUG, ce qui entraîne la formation de *foci* nucléaires (Goers *et al.*, 2008 ; Kanadia *et al.*, 2003b ; Vicente-Crespo *et al.*, 2008). Le motif TZF de la protéine de drosophile étant homologue à ceux de la protéine humaine, un seul motif TZF de la protéine humaine pourrait être suffisant à sa fixation à l'ARN.

A l'époque où nous avons réalisé ces études, aucune donnée structurale n'était disponible sur les motifs TZF de MBNL1. I. Behm-Ansmant et X. Manival ont fait des prédictions informatiques de la structure des motifs TZF12 et TZF34 afin de délimiter quels aminoacides situés de part et d'autre de ces motifs TZF étaient requis pour maintenir une structuration de ces motifs. A partir de ces prédictions, I. Behm-Ansmant a cloné, produit et purifié les protéines MBNL1_{Zn12} et MBNL1_{Zn34}. La protéine MBNL1_{Zn12} correspond aux 81 acides aminés N-terminaux de la protéine MBNL1 et contient le motif TZF12 (aa 14 à 72). La protéine MBNL1_{Zn34} (aa 164 à 248) contient le motif TZF34 (aa 179 à 240). Nous avons aussi utilisé la protéine recombinante MBNL1_{Δ122} préalablement décrite par Kino comme pouvant interagir avec l'ARN. Cette protéine correspond aux aa 1 à 248 de MBNL1 (Fig.56).

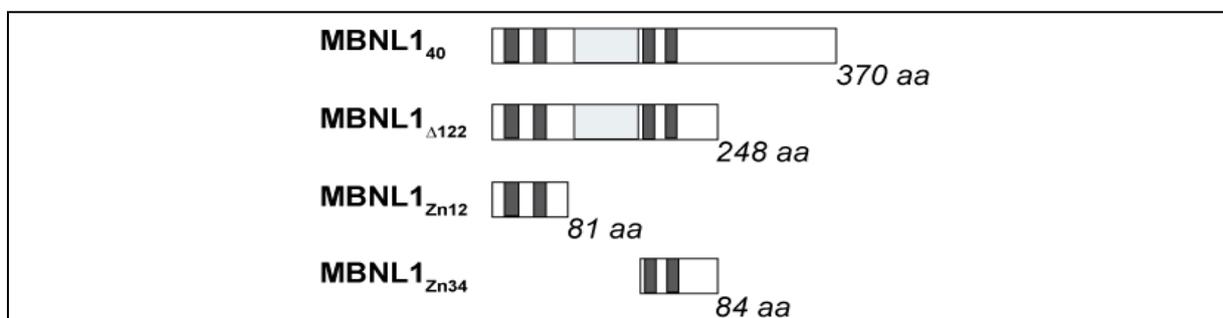


Figure 56 : Représentation schématique des formes recombinantes de MBNL1₄₀, MBNL1_{Zn12}, MBNL1_{Zn34} et MBNL1_{Δ122}.

La protéine MBNL1_{Δ122} correspond à la protéine MBNL1₄₀ déléetée des 122 acides aminés C-terminaux. Elle contient les deux paires de motifs à doigts à Zn^{2+} , séparés par la région flexible reliant les doigts à Zn^{2+} 2 et 3. La protéine MBNL1_{Zn12} correspond aux 81 premiers acides aminés de la protéine MBNL1 et contient le motif TZF12 (aa 14 à 72). La protéine MBNL1_{Zn34} (aa 164-248) contient le motif TZF34 (aa 179 à 240).

J'ai testé la capacité de ces différentes protéines à se fixer sur des répétitions CUG (CUG)₁₇, et sur l'ARN 5'-hcTNT₅₀₋₀. Pour cela, un mélange contenant 100 fmol d'ARN non marqué et une très faible quantité de ce même ARN marqué en 5' a été incubé en présence de quantités croissantes de protéines allant de 225 à 9000 nM, ceci dans les conditions décrites dans la partie Matériel et Méthodes. Après électrophorèse en gel de polyacrylamide à 5 ou 8 % non-dénaturant, les positions des ARN libres et des ARN complexés ont été mises en évidence par autoradiographie (Fig.57).

Comme attendu, la protéine MBNL1_{Δ122} présente une affinité similaire pour les répétitions CUG et l'ARN 5'-hcTNT₅₀₋₀ avec un Kd apparent d'environ 1000 nM. L'affinité de cette protéine est inférieure à celle de MBNL1_{ΔCT3} (environ 600 nM), mais reste comparable. Par contre, aucun complexe n'était formé ni avec les répétitions CUG, ni avec l'ARN 5'-hcTNT₅₀₋₀, en présence des protéines MBNL1_{Zn12} et MBNL1_{Zn34}. Nous avons alors testé si l'ajout simultané des deux protéines MBNL1_{Zn12} et MBNL1_{Zn34} dans un rapport 1:1, pouvait permettre la formation d'un complexe, mais aucun complexe n'était visible.

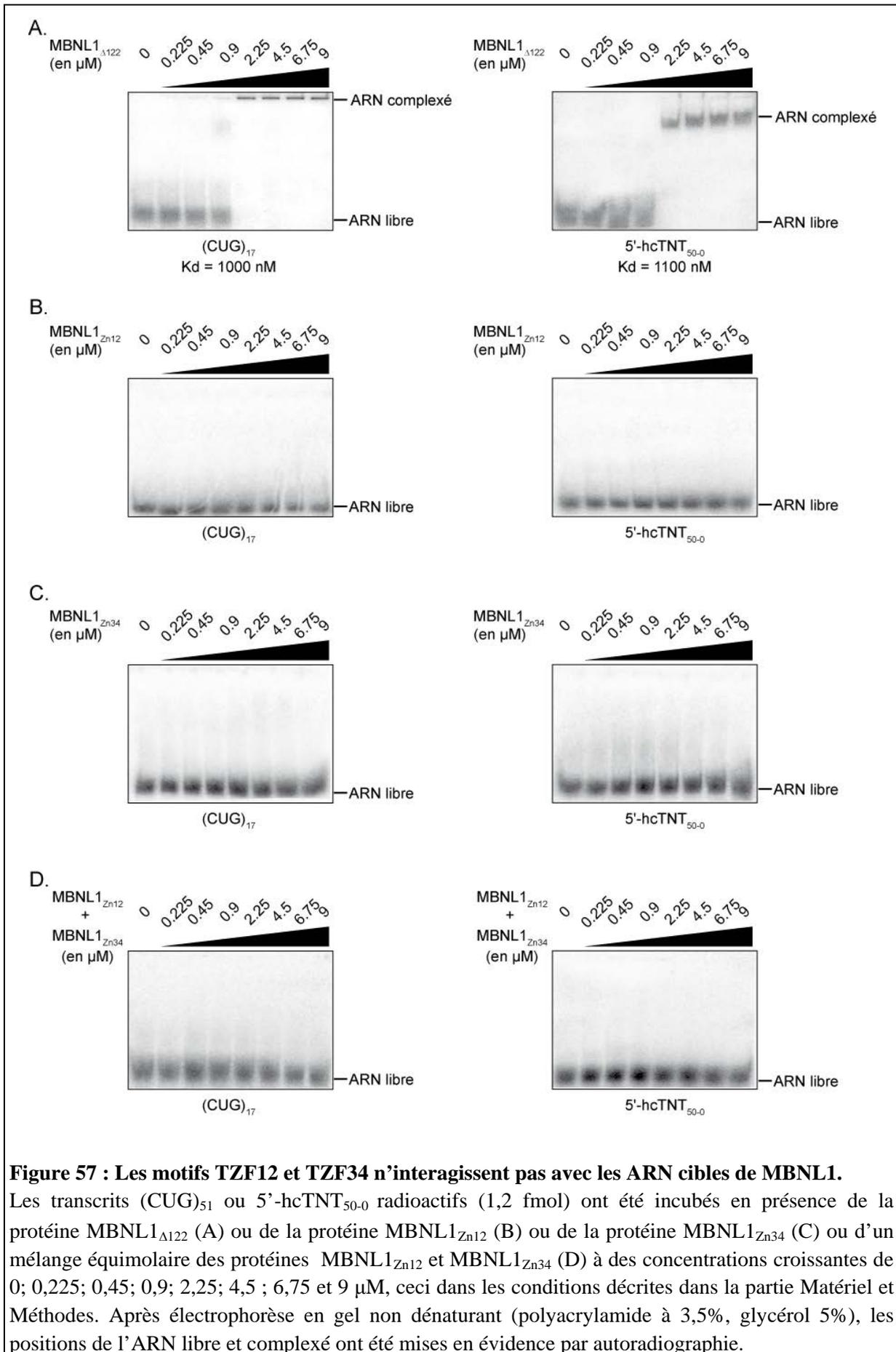


Figure 57 : Les motifs TZF12 et TZF34 n'interagissent pas avec les ARN cibles de MBNL1.

Les transcrits (CUG)₅₁ ou 5'-hcTNT₅₀₋₀ radioactifs (1,2 fmol) ont été incubés en présence de la protéine MBNL1^{Δ122} (A) ou de la protéine MBNL1^{Zn12} (B) ou de la protéine MBNL1^{Zn34} (C) ou d'un mélange équimolaire des protéines MBNL1^{Zn12} et MBNL1^{Zn34} (D) à des concentrations croissantes de 0; 0,225; 0,45; 0,9; 2,25; 4,5; 6,75 et 9 μM, ceci dans les conditions décrites dans la partie Matériel et Méthodes. Après électrophorèse en gel non dénaturant (polyacrylamide à 3,5%, glycérol 5%), les positions de l'ARN libre et complexé ont été mises en évidence par autoradiographie.

V. Conclusions et perspectives

Cette partie de mon travail de thèse avait pour but la compréhension des mécanismes de reconnaissance des ARN par la protéine MBNL1. La reconnaissance des ARN par une protéine donnée repose généralement sur l'implication de domaines spécifiques au sein de la protéine et d'autres part de séquences ARN présentant des caractéristiques de séquences et/ou structurales particulières. C'est pourquoi une partie de notre étude a porté sur la caractérisation des domaines de la protéine MBNL1 impliqués dans la reconnaissance des ARN. Au cours de ce travail, nous avons démontré que i) la délétion de la région C-terminale de la protéine MBNL1 (codée par les exons 5 à 10) n'affecte pas ou très peu l'affinité de la protéine pour l'ARN et ii) bien que l'exon 3 de MBNL1 ne soit pas absolument requis pour la fixation de MBNL1 à l'ARN, sa présence augmente fortement les interactions de MBNL1 avec ses ARN cibles. Ainsi, nous montrons que l'ensemble de la région N-terminale (codée par les exons 1 à 4) et non uniquement les deux couples de doigts à Zn²⁺ de la protéine MBNL1 intervient dans la fixation à l'ARN.

Par ailleurs nous avons cherché par l'approche SELEX à identifier des séquences ARN de haute affinité pour MBNL1. Les données obtenues montrent la capacité de MBNL1 à se fixer sur deux catégories différentes d'ARN :

1) les ARN de la famille I qui contiennent de 1 à 4 motifs YGCY localisés dans une séquence enrichie en pyrimidine et qui ne semblent pas présenter de caractéristiques structurales communes. Ces ARN sont similaires aux ARN trouvés récemment par l'équipe de Berglund par doped SELEX (Goers *et al.*, 2010), mais également aux segments de pré-ARNm contenant des sites MBNL1. Nos données confirment également que les ARN ne contenant aucun ou un seul motif YGCY lient MBNL1 avec une mauvaise affinité comparée à celle observée pour des ARN contenant 2, 3 ou 4 motifs YGCY.

2) les ARN de la famille II qui contiennent un seul motif YGCY et se replient en une longue structure tige-boucle très stable dans laquelle le motif YGCY est localisé dans la boucle terminale, dont la séquence est UUUGCGCU. Nous montrons donc que, contrairement à ce qui été parfois décrit dans la littérature, MBNL1 peut se fixer avec une très forte affinité sur un ARN ne contenant qu'un seul motif YGCY présent dans une région en simple brin. De façon intéressante un changement de contexte structural du motif YGCY présent au sein des ARN SELEX de la famille II par placement de la séquence de la boucle terminale au sein d'une hélice inhibe totalement la fixation de MBNL1, confirmant l'importance de la présence de ce motif unique dans un contexte simple brin. Le motif de haute affinité présent au sein des

ARN de la famille II n'a encore jamais été décrit naturellement, cependant on note une similitude des ARN de cette famille avec les pré-miRNA dont la maturation est dépendante de MBNL1 (Rau *et al.*, 2011). En effet ces pré-miRNA contiennent également un seul motif YGCY placé dans une boucle. Cependant dans le cas des pré-miRNA l'affinité de MBNL1 pour le motif YGCY est beaucoup plus faible que celle observée pour les ARN de la famille II. Cela s'explique sans doute par les caractéristiques de séquence et de structure (taille de la boucle et de l'appariement qui la ferme) des ARN de la famille II et l'architecture différente de la boucle des pré-miRNA. Par ailleurs, il est normal que l'affinité de MBNL1 pour les pré-miRNA ne soit pas trop forte puisqu'il s'agit d'un mécanisme de régulation de la maturation du pré-miR qui dépend en particulier des conditions physiologiques.

La découverte d'un motif de très haute affinité pour MBNL1 (affinité supérieure à celle de MBNL1 pour les répétitions CUG) ouvre par ailleurs de nouvelles perspectives thérapeutiques. En effet, il est possible d'envisager de « délivrer » spécifiquement la protéine MBNL1 au voisinage d'un site d'épissage donné afin de corriger son épissage par utilisation d'un oligonucléotide antisens bifonctionnel dont une partie s'hybride à l'ARN cible, l'autre partie contenant le motif de haute affinité pour MBNL1 et permettant son ancrage. Ce type de stratégie a déjà été utilisé avec succès pour délivrer un facteur d'épissage, en l'occurrence SRSF1, au niveau de son site d'action et ainsi corriger des défauts d'épissage à l'origine de l'amyotrophie spinale (Skordis *et al.*, 2003).

Chapitre II : Etude des mécanismes régulant le taux d'inclusion de l'exon 5 du pré-ARNm de la troponine T cardiaque humaine.

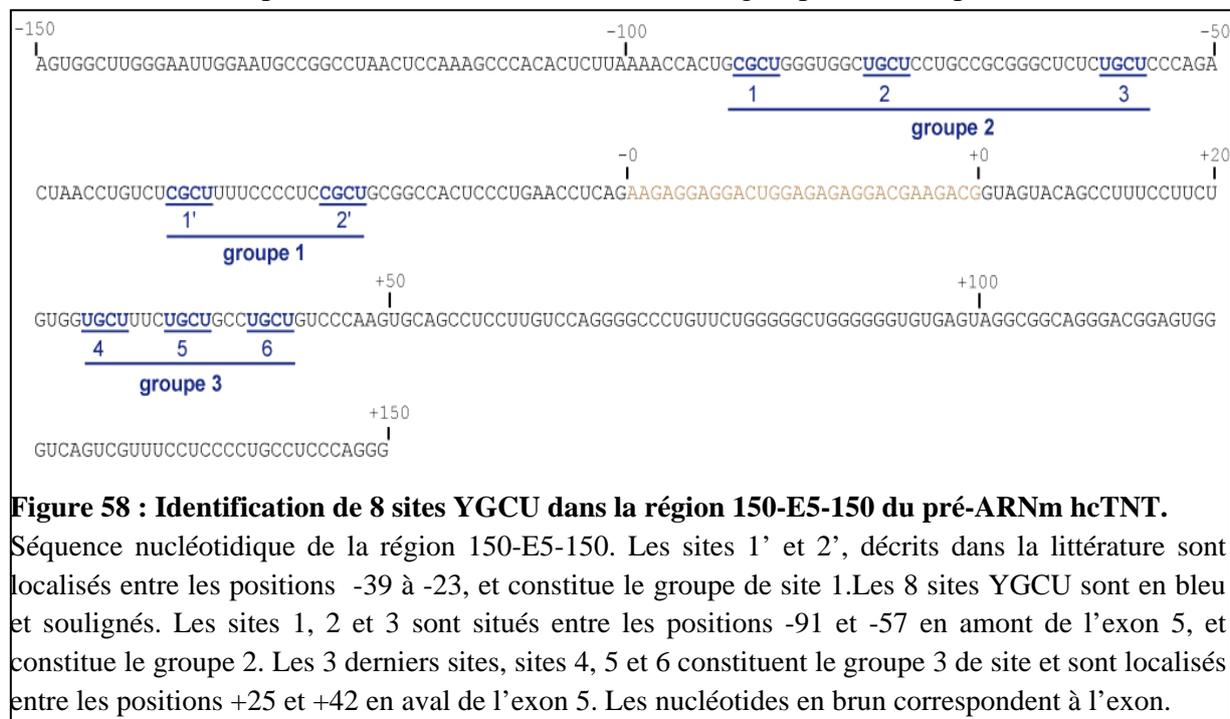
Comme mentionné précédemment, l'un de nos objectifs initiaux était de comprendre comment MBNL1 interagit avec les pré-ARNm pour réguler leur épissage. Parmi les pré-ARNm dont l'épissage alternatif est dérégulé chez les patients atteints de DM1, celui de la troponine T cardiaque humaine (hcTNT) est particulièrement important. En effet, la dérégulation de l'inclusion de l'exon 5 de ce transcrit conduit à une mauvaise contraction musculaire et pourrait être impliquée dans les anomalies cardiaques qui touchent 80 % des patients atteints de DM1 (Philips *et al.*, 1998). Lorsque j'ai débuté ma thèse, la majorité des travaux réalisés sur la régulation de l'épissage du pré-ARNm hcTNT avait été effectuée chez le poulet et peu d'informations existaient sur la régulation de l'épissage du pré-ARNm humain. Deux sites de fixation de la protéine MBNL1 avaient été identifiés entre les positions -40 et -21 en amont de l'exon 5 (Ho *et al.*, 2005). Ces deux sites encadrent la séquence polypyrimidine. Par ailleurs, des sites de fixation des protéines de la famille CELF avaient été identifiés dans les 50 premiers nucléotides de l'intron 5 (Ho *et al.*, 2005).

I. Mise en évidence par examen des séquences de nombreux sites potentiels de fixation de MBNL1 situés en amont et en aval de l'exon 5 du pré-ARNm de la Troponine T cardiaque humaine.

Sur la base des données que nous avons obtenues par l'expérience SELEX et du consensus établi pour les sites de fixation connus de MBNL1 (Ho *et al.*, 2004), nous avons recherché l'existence de sites potentiels de fixation de MBNL1 en amont et en aval de l'exon 5 de hcTNT. Dans un premier temps, N. Marmier-Gourrier a mis en évidence la présence de 3 sites potentiels en aval de l'exon 5 en plus des deux sites situés en amont et déjà décrits dans la littérature (Ho *et al.*, 2005). Ces 3 sites en aval sont localisés entre les positions +25 et +42 par rapport à l'extrémité 3' de l'exon 5 et ils chevauchent les sites proposés de fixation de CUG-BP1. Par la suite, j'ai moi aussi identifié 3 sites potentiels supplémentaires non décrits dans la littérature, et localisés entre les positions -91 et -57 par rapport à l'extrémité 5'. Pour faciliter la compréhension, nous désignerons les deux sites préalablement décrits par Ho *et al.* comme les sites 1' et 2' et nous considérerons qu'ils forment un groupe de site (groupe 1). Les trois nouveaux sites potentiels que j'ai identifié entre les positions -91 et -57 seront appelés sites 1, 2 et 3, et nous considérerons qu'ils forment un autre groupe de sites (groupe

2). Enfin, les trois derniers sites potentiels supplémentaires situés en aval de l'exon 5 seront appelés sites 4, 5 et 6, et nous considérerons qu'ils appartiennent à un 3^{ème} groupe de sites (groupe 3) (Fig.58).

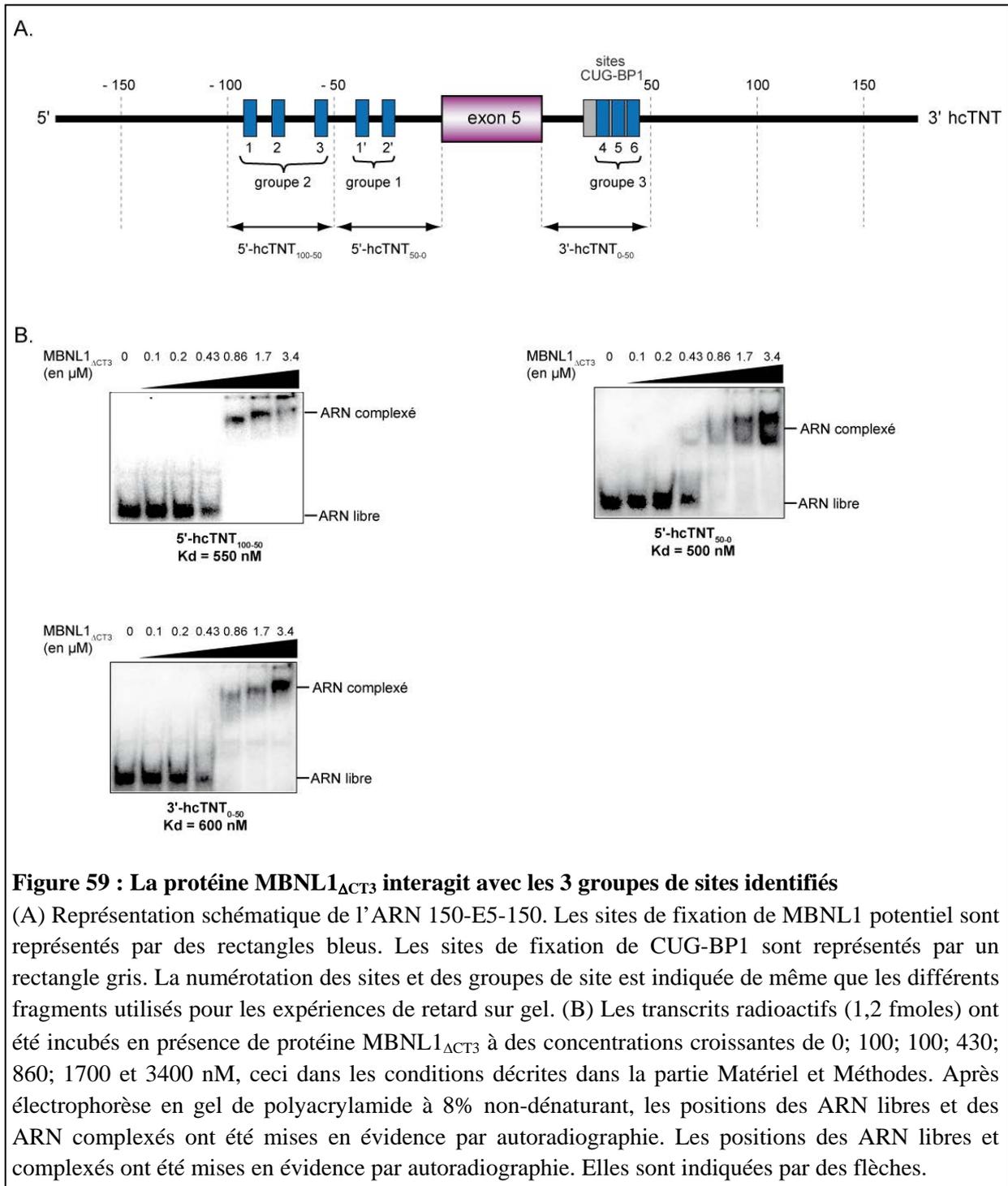
J'ai alors testé la capacité de fixation de MBNL1 sur les groupes de sites potentiels 2 et 3.



1.1. Test de fixation de MBNL1 sur ses différents sites potentiels de fixation

Afin de vérifier que les nouveaux sites potentiels identifiés correspondaient bien à des sites de fixation de MBNL1, j'ai réalisé des expériences de retard sur gel en utilisant la protéine recombinante MBNL1 Δ CT3. J'ai utilisé comme contrôle le transcrit de 50 nt 5'-hcTNT₅₀₋₀ qui permettait de vérifier la fixation de MBNL1 sur les sites 1' et 2', comme décrit dans le Chapitre I de la partie Résultats. Par ailleurs, j'ai utilisé deux autres transcrits. Un transcrit correspondait aux résidus compris entre les positions -100 à -50 (5'-hcTNT₁₀₀₋₅₀) en amont de l'exon 5. Il permettait de tester la reconnaissance des sites du groupe 2. Le second transcrit correspondait aux résidus compris entre les positions 0 à 50 (3'-hcTNT₀₋₅₀) en aval de l'exon 5. Il permettait de tester la capacité de fixation des sites du groupe 3.

Les expériences de retard sur gel ont été réalisées en incubant 1,2 fmoles d'ARN radio-marqué en 5' en présence de concentrations croissantes de protéines MBNL1 Δ CT3 allant de 100 à 3400 nM, ceci dans les conditions décrites dans la partie Matériel et Méthodes. Après électrophorèse en gel de polyacrylamide à 8 % non-dénaturant, la position des ARN libres et complexés a été mise en évidence par autoradiographie. Les résultats obtenus sont illustrés dans la Figure 59.



Comme décrit dans le Chapitre I de la partie Résultats, la protéine MBNL1_{ΔCT3} présentait une bonne affinité pour l'ARN 5'-hcTNT₅₀₋₀ (Kd apparent d'environ 500 nM). MBNL1 possède une affinité du même ordre de grandeur pour les deux autres ARN testés, 5'-hcTNT₁₀₀₋₅₀ (contenant les sites du groupe 2) et 3'-hcTNT₀₋₅₀ (contenant les sites du groupe 3), les Kd apparents étant d'environ 550 et 600 nM, respectivement. De plus, nous observions la formation de plusieurs complexes ayant des mobilités électrophorétiques différentes pour chacun des trois fragments suggérant la fixation de plusieurs molécules de protéine

MBNL1 Δ CT3 sur ces fragments. Alors que deux formes de complexes coexistaient pour l'ARN 5'-hcTNT₅₀₋₀ avec les trois plus fortes concentrations en protéine testées, nous observions pour les ARN 5'-hcTNT₁₀₀₋₅₀ et 3'-hcTNT₀₋₅₀ une diminution progressive de la mobilité électrophorétique des complexes formés lorsque la concentration en protéine MBNL1 Δ CT3 augmentait de 0,83 à 3,5 μ M.

Ces résultats confirmaient la fixation de MBNL1 sur les différents groupes de sites, en particulier les 2 nouveaux groupes de sites que nous avons identifiés. La persistance de deux types de complexes dans le cas de l'ARN 5'-hcTNT₅₀₋₀ suggère l'existence d'une hétérogénéité de conformation de l'ARN, l'une des conformations ne conduisant à la fixation que d'une seule molécule de protéine, l'autre à la fixation de 2 molécules. Les 3 sites du segment 3'-hcTNT₀₋₅₀ sont quasiment juxtaposés. On peut donc penser que la fixation de 3 molécules de MBNL1 Δ CT3 sur ces 3 sites soit limitée par des problèmes d'encombrement stérique. Néanmoins, les mobilités électrophorétiques des différents complexes observées pour cet ARN, comme pour l'ARN 5'-hcTNT₁₀₀₋₅₀ étaient en accord avec la fixation possible de 3 molécules de MBNL1 Δ CT3 sur chacun de ces ARN. Nous en concluons que soit MBNL1 Δ CT3 est capable de se lier sur des sites proches, soit ce que nous observons correspond à une multimérisation de la protéine fixée à un des sites.

Dans le Chapitre I de la partie Résultats, nous avons montré que la région de MBNL1 codée par l'exon 7 joue un rôle important dans la multimérisation de cette protéine. Cette région n'est pas présente dans la protéine MBNL1 Δ CT3, ce qui pourrait être en faveur de la fixation de 3 molécules dans nos expériences. Néanmoins, il faut souligner que la région codée par l'exon 7 n'est apparemment pas la seule région impliquée dans des interactions protéine/protéine.

I.2. Abolition de la fixation de MBNL1 suite à la mutation de ces sites de fixation

Ayant observé précédemment (Chapitre I) que la mutation du résidu G du motif YGCY en résidu A est suffisante pour abolir la capacité de fixation de MBNL1, j'ai ensuite réalisé ce type de mutation dans chacun des sites présents dans les ARN 5'-hcTNT₁₀₀₋₅₀, 5'-hcTNT₅₀₋₀ et 3'-hcTNT₀₋₅₀, et testés la capacité de fixation de MBNL1 Δ CT3 par expérience de retard sur gel. Les ARN mutés sont désignés : 5'-hcTNT₁₀₀₋₅₀mut1-2-3, 5'-hcTNT₅₀₋₀mut1'-2' et 3'-hcTNT₀₋₅₀mut4-5-6.

Comme précédemment, ces expériences ont été réalisées avec 1,2 fmoles d'ARN radio-marqués en 5' et incubées en présence de concentrations croissantes de protéines

MBNL1 $_{\Delta CT3}$, allant de 100 à 3400 nM, ceci dans les conditions décrites dans la partie Matériel et Méthodes. Après électrophorèse en gel de polyacrylamide à 8 % non-dénaturant, la position des ARN libres et ARN complexés a été mise en évidence par autoradiographie (Fig.60).

Les résultats obtenus montraient que la mutation G->A des motifs YGCU des groupes de sites 1 et 3 (ARN 5'-hcTNT $_{50-0}$ mut1'-2' et 3'-hcTNT $_{0-50}$ mut4-5-6) abolissait totalement la fixation de MBNL1. Par contre, un complexe était encore visible après mutation des sites du groupe 2 (5'-hcTNT $_{100-50}$ mut1-2-3), cependant l'affinité de MBNL1 pour l'ARN muté était faible. La fixation de MBNL1 sur ce mutant pourrait s'expliquer par la présence d'un motif YGCC connu pour fixer MBNL1 avec une affinité plus faible que les motifs YGCU (Goers *et al.*, 2010).

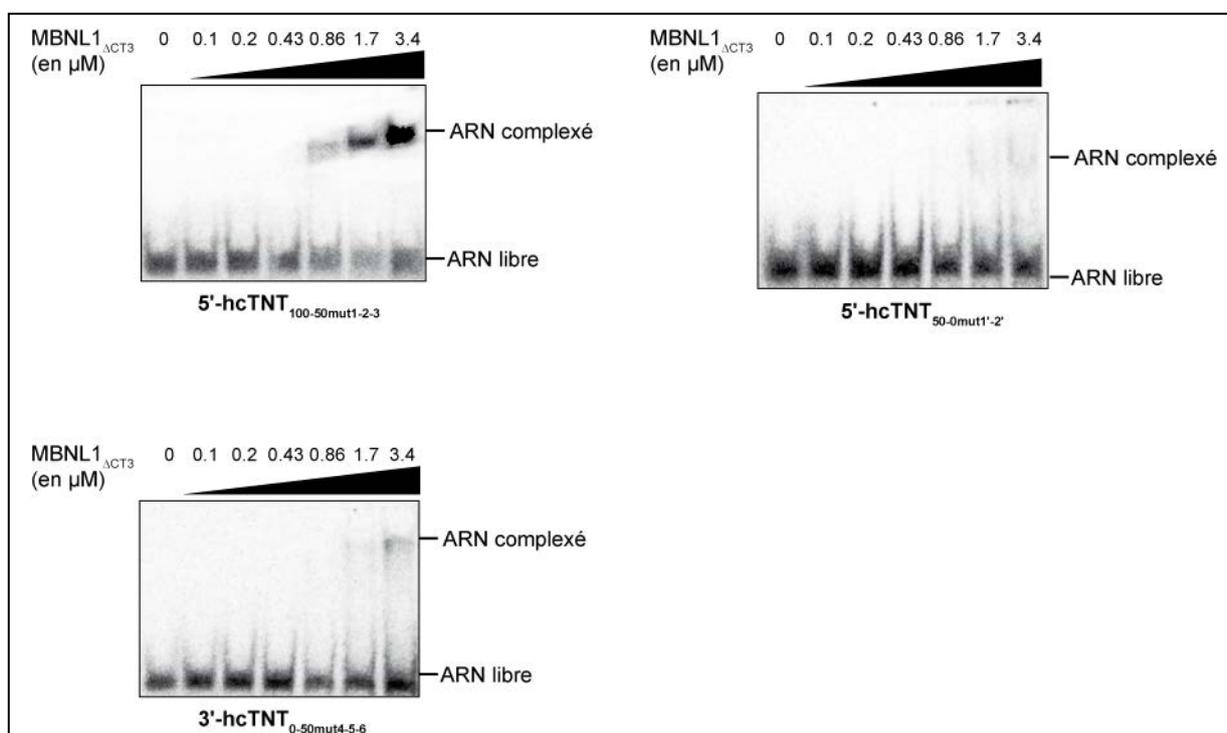


Figure 60 : La mutation des résidus G en résidus A des motifs YGCU inhibe la fixation de MBNL1 sur les groupes de sites

Les transcrits radioactifs (1,2 fmoles) ont été incubés en présence de protéine MBNL1 $_{\Delta CT3}$ à des concentrations croissantes de 0; 100; 100; 430; 860; 1700 et 3400 nM, ceci dans les conditions décrites dans la partie Matériel et Méthodes. Après électrophorèse en gel de polyacrylamide à 8% non-dénaturant, les positions des ARN libres et des ARN complexés ont été mises en évidence par autoradiographie. Les positions des ARN libres et complexés ont été mises en évidence par autoradiographie. Elles sont indiquées par des flèches.

II. Test d'une compétition éventuelle entre CUG-BP1 et MBNL1 pour leur fixation aux séquences introniques bordant l'exon 5

Les données de Philips et al, montraient que la protéine CUG-BP1 est capable de se lier à la séquence intronique située en aval de l'exon 5, ceci par des expériences de pontage aux UV réalisées avec le fragment s'étendant des positions +17 à +46 par rapport à l'extrémité 3' de l'exon 5. La mutation de résidus U en résidus A avaient permis d'identifier 5 sites potentiels de fixation de CUG-BP1 au sein de cette région. Cependant, l'importance de chacun de ces sites n'a jamais été testée. Parmi ces sites, présents dans l'ARN 3'-hcTNT₀₋₅₀, 4 chevauchaient les sites MBNL1 (Fig.61). Afin de savoir si les deux protéines entraînent en compétition dans leur fixation à l'ARN 3'-hcTNT₀₋₅₀, nous avons d'abord du produire une forme soluble de la protéine CUG-BP1 afin de réaliser des tests de fixation par retard sur gel.

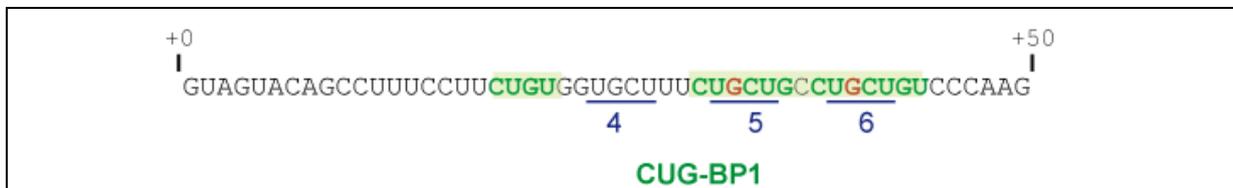


Figure 61 : Chevauchement des sites de fixation de MBNL1 et CUG-BP1 au niveau de l'ARN 3'-hcTNT₀₋₅₀

Séquence nucléotidique de l'ARN 3'-hcTNT₀₋₅₀. Les sites de fixation de CUG-BP1 décrits dans la littérature sont indiqués en vert. Les sites 4, 5 et 6 de fixation de MBNL1 sont soulignés en bleu et la mutation des résidus G en résidus A du motif YGCU de ces sites est indiquée en rouge.

II.1. Construction génétique et production d'une forme soluble de la protéine CUG-BP1

En raison de la faible solubilité de la protéine CUG-BP1 recombinante produite par N. Marmier-Gourrier (voir Chapitre I de la partie Résultat), nous avons décidé d'utiliser une forme tronquée de cette protéine. Les données de la littérature indiquaient que les 2 motifs RRM N-terminaux de la protéine CUG-BP1 ainsi qu'une partie du domaine divergent séparant les motifs RRM1 et 2 du motif RRM3, étaient suffisants pour assurer la fixation à l'ARN (Fig.62). Par ailleurs, une protéine contenant ces éléments de séquences était connue pour conserver ses capacités de régulation de l'épissage (Singh *et al.*, 2004 ; Han et Cooper., 2005). Nous avons donc décidé d'utiliser cette forme tronquée de sa région C-terminale, nous l'appellerons CUG-BP1_{ΔCT}, et j'ai réalisé une construction permettant de l'exprimer. A partir de cette construction réalisée dans le plasmide pGEX6P1, N. Marmier-Gourrier a produit la protéine en grande quantité chez *E. coli*, et l'a purifiée par fixation sur billes de glutathion suivie d'une coupure par la protéase PréScission afin d'éliminer spécifiquement la partie

GST. Après concentration, la protéine a été dialysée contre du tampon D 1X. J'ai utilisé cette protéine pour réaliser des expériences de retard sur gel.

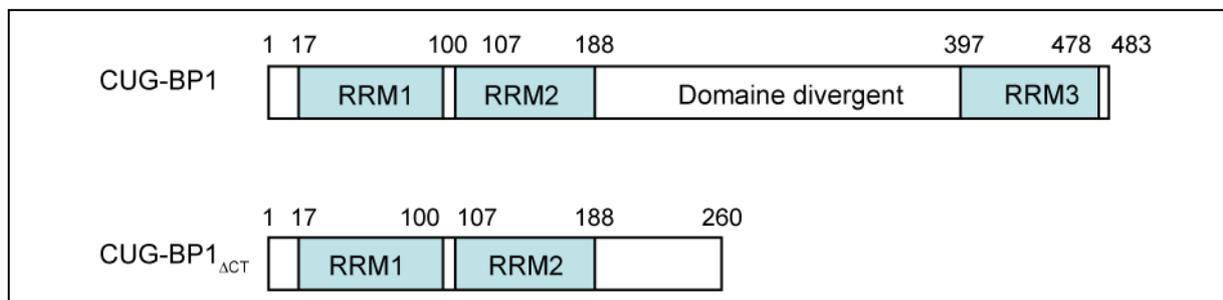


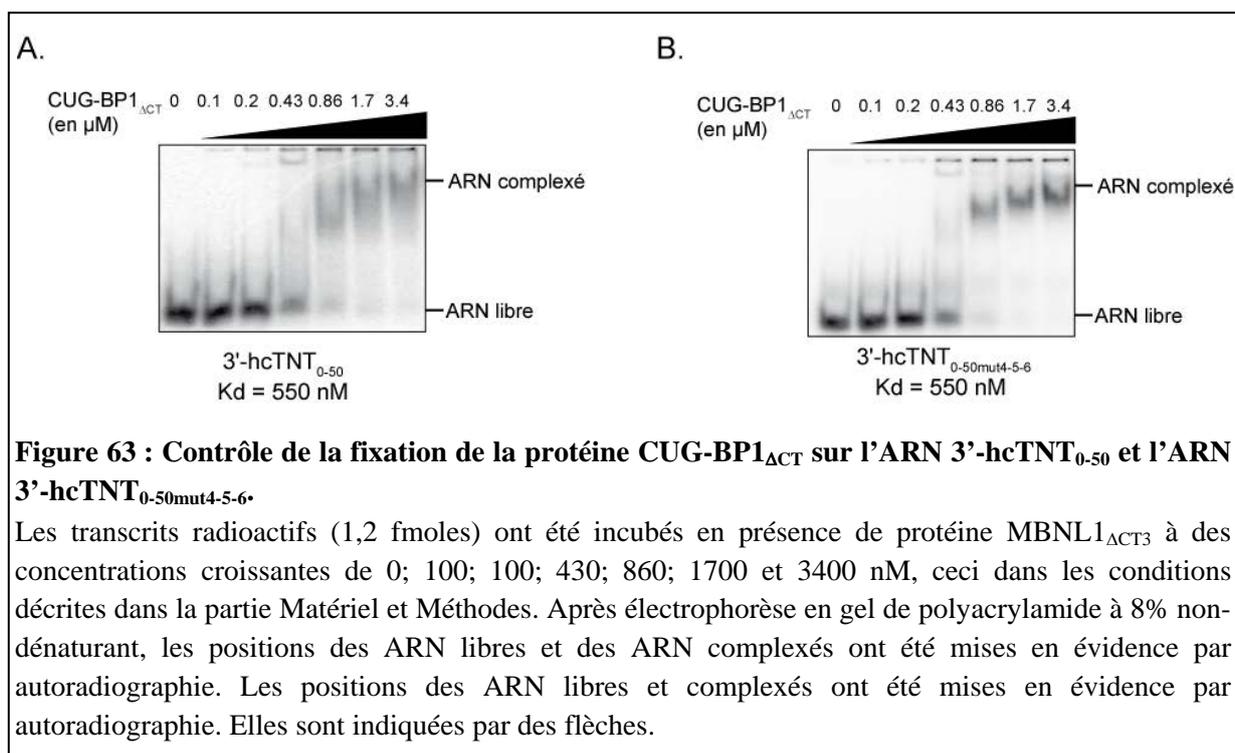
Figure 62 : Représentation schématique de la forme recombinante de CUG-BP1 utilisée au cours de mon travail : CUG-BP1_{ΔCT}.

La protéine CUG-BP1_{ΔCT} correspond à la région N-terminale de la protéine CUG-BP1. Elle contient les motifs RRM1 et RRM2 ainsi que 72 acides aminés du domaine divergent, région qui sépare les RRM2 et RRM3.

II.2. Test de fixation des protéines par expériences de retard sur gel

Les expériences de retard sur gel ont été effectuées avec l'ARN 3'-hcTNT₀₋₅₀ dans les conditions utilisées pour MBNL1, de façon à pouvoir comparer les résultats, à savoir par incubation de 1,2 fmoles d'ARN radio-marqué en 5' en présence de quantités croissantes de protéines CUG-BP1_{ΔCT} allant de 100 à 3400 nM, ceci dans les conditions décrites dans la partie Matériel et Méthodes. Après électrophorèse en gel de polyacrylamide à 8 % non-dénaturant, la position des ARN libres et complexés a été mise en évidence par autoradiographie. Les résultats obtenus sont illustrés dans la Figure 63.

Comme pour MBNL1_{ΔCT3}, l'ARN était presque totalement complexé à partir d'une concentration en protéine CUG-BP1_{ΔCT} de 0,86 μM (Fig.63). J'ai ensuite testé par la même approche la fixation de CUG-BP1_{ΔCT} sur l'ARN 3'-hcTNT₀₋₅₀mut4-5-6 où les 3 sites de fixation de MBNL1 avaient été mutés. Nous n'avons pas noté de perte d'affinité significative de CUG-BP1_{ΔCT} pour l'ARN muté. Ceci pourrait s'expliquer par la fixation prédominante de CUG-BP1 sur les séquences UGU entre les positions 22 et 24 et les positions 42 et 44, qui n'ont pas été mutées dans l'ARN 3'-hcTNT₀₋₅₀mut4-5-6. Néanmoins, nous observons une diminution de la mobilité électrophorétique du complexe formé en augmentant la concentration en protéine, ce qui soulève la question de l'existence d'autres sites de fixation ou de la multimérisation de la protéine.



II.3. Expérience de compétition par pontage covalent aux UV entre la protéine MBNL1 et CUG-BP1

Les données obtenues suggéraient qu'au niveau de l'ARN 3'-hcTNT $_{0-50}$, les sites MBNL1 et CUG-BP1 n'étaient en fait pas chevauchants. Néanmoins par encombrement stérique la fixation d'une des protéines pouvait empêcher la fixation de l'autre. Afin de le tester, j'ai entrepris des études de compétition entre les 2 protéines en employant une méthode basée sur le pontage à l'ARN des protéines qui lui sont directement liées par irradiation aux UV. Ce type d'expérience a été réalisé avec un ARN 3'-hcTNT $_{0-50}$ uniformément marqué au cours de la transcription par incorporation de [α - 32 P]-UTP. Après formation d'un complexe entre l'ARN marqué et la protéine, le complexe est soumis à une irradiation aux UV à 256 nm selon les conditions décrites dans la partie Matériel et Méthode. L'ARN est ensuite hydrolysé par un mélange de RNase T1 et de RNase A, puis les protéines qui, si elles ont été pontées à l'ARN, comportent des nucléotides radioactifs, sont fractionnées par électrophorèse SDS-PAGE.

Dans un premier temps, j'ai vérifié que chacune des deux protéines MBNL1 $_{\Delta CT3}$ et CUG-BP1 $_{\Delta CT}$ était bien pontée à l'ARN 3'-hcTNT $_{0-50}$. Ensuite j'ai utilisé la stratégie suivante pour tester la compétition entre les 2 protéines. J'ai formé un complexe entre l'ARN 3'-

hcTNT₀₋₅₀ et soit MBNL1_{ΔCT3}, soit CUG-BP1_{ΔCT}, et j'ai ensuite ajouté la seconde protéine à différentes concentrations croissantes.

Pour former le complexe initial, comme les deux protéines avaient des Kd apparent similaires, j'ai choisi la même concentration pour chacune de ces protéines, à savoir 1 μM qui permet une complexation totale de l'ARN. Les compétitions ont ensuite été réalisées avec des concentrations en protéine allant de 1 à 4 μM. L'incubation avec la première protéine a été réalisée pendant 15 min à température ambiante. La compétition avec la seconde protéine a également été réalisée pendant 15 min à température ambiante. L'irradiation aux UV était alors réalisée pendant 2 min. Après hydrolyse de l'ARN les protéines ont été fractionnées par électrophorèse SDS-PAGE avec en témoin une expérience réalisée sans compétition. Les résultats obtenus sont illustrés dans la Figure 64. La fixation simultanée des deux protéines sur l'ARN 3'-hcTNT₀₋₅₀ a été observée et aucune compétition entre les 2 protéines n'a été détectée

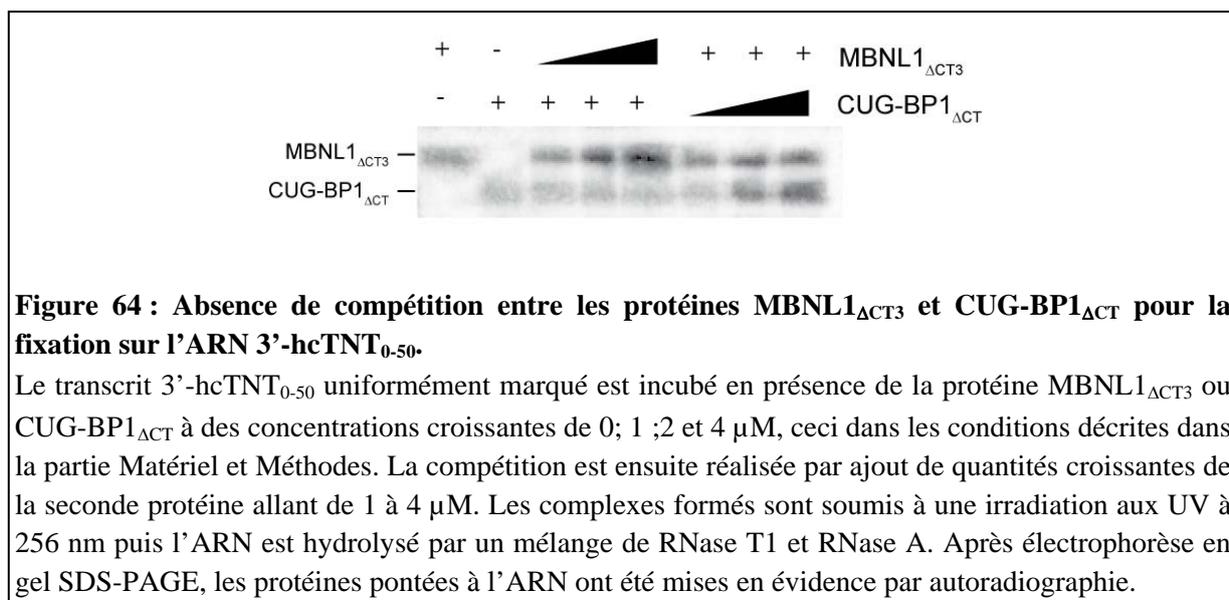


Figure 64 : Absence de compétition entre les protéines MBNL1_{ΔCT3} et CUG-BP1_{ΔCT} pour la fixation sur l'ARN 3'-hcTNT₀₋₅₀.

Le transcrit 3'-hcTNT₀₋₅₀ uniformément marqué est incubé en présence de la protéine MBNL1_{ΔCT3} ou CUG-BP1_{ΔCT} à des concentrations croissantes de 0; 1 ;2 et 4 μM, ceci dans les conditions décrites dans la partie Matériel et Méthodes. La compétition est ensuite réalisée par ajout de quantités croissantes de la seconde protéine allant de 1 à 4 μM. Les complexes formés sont soumis à une irradiation aux UV à 256 nm puis l'ARN est hydrolysé par un mélange de RNase T1 et RNase A. Après électrophorèse en gel SDS-PAGE, les protéines pontées à l'ARN ont été mises en évidence par autoradiographie.

III. Caractérisation des séquences nucléotidiques requises pour la régulation de l'inclusion de l'exon 5 dans l'ARNm hcTNT

II.1. Identification des séquences nucléotidiques requises pour la régulation de l'inclusion de l'exon 5 dans l'ARNm hcTNT par des études *in cellulo*

Ayant mis en évidence des sites potentiels de fixation de MBNL1 en amont des sites déjà décrits dans l'intron 4 de hcTNT, ainsi que des sites supplémentaires en aval de l'exon 5, il nous a paru indispensable d'avoir une définition précise des séquences des introns 4 et 5 indispensables à la régulation du taux d'inclusion de cet exon et aussi à sa dérégulation en

présence d'un ARN contenant des répétitions CUG. Nous avons donc cherché à identifier les séquences minimales requises en amont et en aval de l'exon 5 pour pouvoir récapituler *in cellulo* avec une construction modèle, les régulations de l'inclusion de l'exon 5 et l'effet *trans*-dominant positif des répétitions de triplets CUG.

Pour cela, nous avons développé une collaboration avec l'équipe de N. Sergeant (Lille, Inserm). En effet, cette équipe avait mis au point un système d'étude *in cellulo* reposant sur la transfection de cellules T98 (cellules de glioblastomes) par deux constructions : l'une appelée pcDNA-CTG₉₆₀ exprimait un ARN correspondant à la région 3'-UTR de l'ARNm *DMPK* dans lequel un segment contenant 960 répétitions CUG était inséré, l'autre IR-hcTNT₁₇₅₋₄₂₀ dérivait du minigène pSVIRB (Andreadis *et al.*, 1993), et avait été utilisé préalablement par l'équipe de N. Sergeant pour étudier l'épissage alternatif du pré-ARNm Tau. Dans ce minigène IR-hcTNT₁₇₅₋₄₂₀, l'exon 5 et ses séquences introniques flanquantes avaient été insérés dans l'intron 2 du récepteur à l'insuline (IR) et lui-même était flanqué de ses exons bordants. Ce minigène exprimait un ARN contenant les séquences introniques entières des introns 4 et 5, 175 et 420 nt respectivement. L'équipe de N. Sergeant avait montré qu'il permettait bien de récapituler à la fois la régulation naturelle de l'exon 5 au sein du pré-ARNm hcTNT et sa dérégulation en présence d'un ARN contenant des répétitions CUG. En effet, lorsque ce minigène était transfecté dans des cellules T98, le taux d'inclusion de l'exon 5 était d'environ 50 % et comme attendu, lors de sa co-transfection avec le plasmide pcDNA-CTG₉₆₀, le taux d'inclusion de l'exon 5 était d'environ 95%.

Afin de définir la longueur minimale des séquences des introns 4 et 5 requises pour la régulation de l'inclusion de l'exon 5, N. Marmier-Gourrier et I. Behm-Ansmant ont préparé des minigènes IR-hcTNT qui comportaient 250, 150, 100, 75 ou 50 nt de part et d'autre de l'exon 5. D. Ghanem et M-L. Caillet-Boudin de l'équipe de N. Sergeant ont ensuite analysé le taux d'inclusion de l'exon 5 au sein des ARNm produits à partir de ces différents minigènes dans des cellules T98. Pour cela les différents minigènes IR-hcTNT ont été cotransfectés soit avec le plasmide contrôle pcDNA, pour analyser leur épissage dit « normal », soit avec le plasmide pcDNA-CTG₉₆₀ pour tester l'effet des répétitions CUG.

Les résultats obtenus, qui sont illustrés dans la Figure 65, nous ont permis de conclure que la régulation naturelle du taux d'inclusion de l'exon 5 et l'effet positif des répétitions CUG sur son inclusion nécessitent la présence d'au moins 150 nt en amont et en aval de cet exon. En effet, lorsque l'exon 5 est bordé par 50 nt issus des introns 4 et 5 (IR-hcTNT₅₀₋₅₀), il est totalement exclu de l'ARNm, et ceci, aussi bien en présence, qu'en absence des répétitions CUG (Fig.65). Lorsque des régions introniques de 75 nt sont présentes (IR-hcTNT₇₅₋₇₅), nous

observations une forte inclusion de l'exon 5 (environ 75 %), et ce taux n'augmentait que très légèrement en présence des répétitions CUG, il restait inférieur à celui observé avec la construction contrôle où l'ensemble des séquences introniques étaient présentes. Enfin, lorsque des régions introniques de 100 nt étaient présentes (IR-hcTNT₁₀₀₋₁₀₀), nous observions au contraire une inclusion presque totale de l'exon 5 et ce taux n'était pas modifié en présence des répétitions CUG (Fig.65.). Ces résultats suggéraient l'existence d'un enchaînement d'éléments régulateurs positifs et négatifs dans les séquences en amont et en aval de l'exon 5 s'étendant des positions -150 à 0 d'une part et 0 à +150 d'autre part, et donc une grande complexité de la régulation de l'inclusion de l'exon 5. Nous avons donc décidé de poursuivre les travaux en utilisant le minigène IR-hcTNT₁₅₀₋₁₅₀.

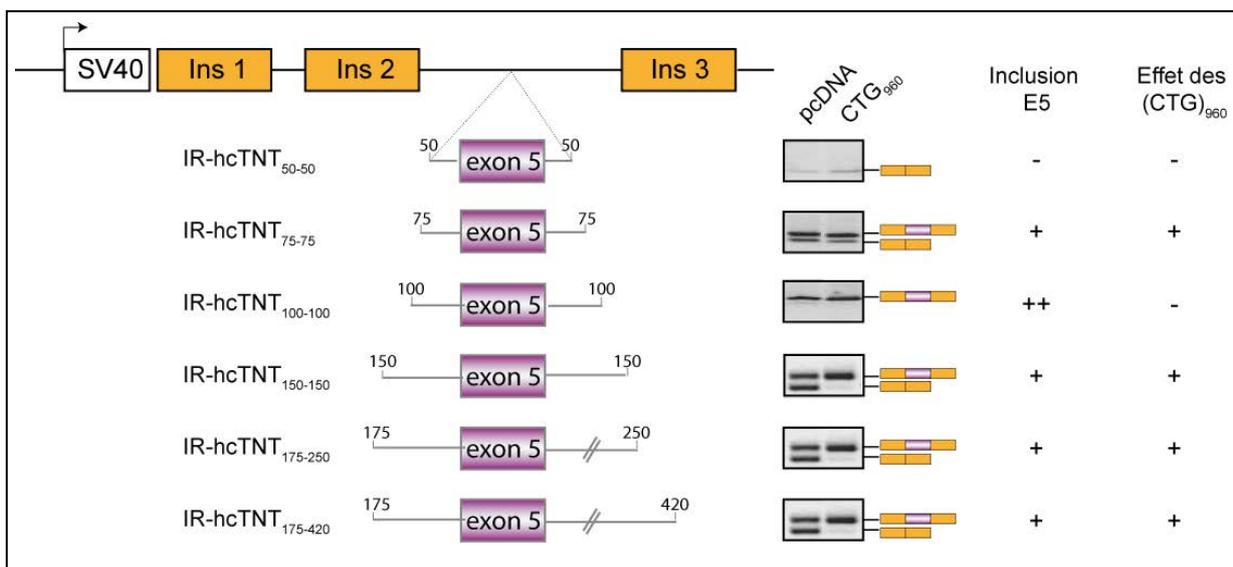


Figure 65 : Détermination de la taille des régions introniques en amont et en aval de l'exon 5 du pré-ARNm hcTNT requises pour récapituler l'ensemble des régulations de l'inclusion de l'exon 5.

L'exon 5 du pré-ARNm hcTT et ses séquences introniques flanquantes ont été insérés dans l'intron 2 du récepteur à l'insuline de rat, lui-même flanqué de ses exons bordants et placé sous le contrôle du promoteur SV40. L'épissage de l'exon 5 *in cellulo* des transcrits des différentes minigènes est analysé par RT-PCR. La taille des séquences introniques flanquantes est indiquée. Chaque minigène a été cotransfecté dans des cellules T98 avec une construction exprimant ou non des répétitions CTG dans le contexte du gène *DMPK*.

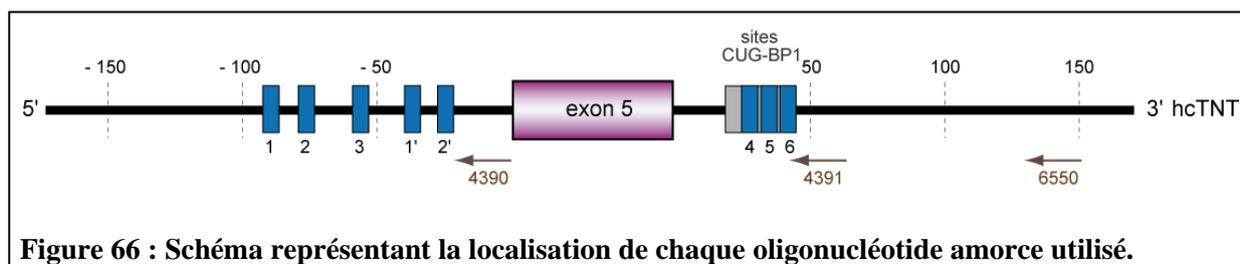
+ correspond à une inclusion d'environ 50 % de l'exon 5 ; ++ correspond à une inclusion de plus de 90 % de l'inclusion de l'exon 5 ; - correspond à une forte exclusion de l'exon 5

II.2. Analyse de la structure 2D de la région du pré-ARNm hcTNT requise pour la régulation de l'inclusion de l'exon 5 et test de la fixation de MBNL1 par expériences

II.2.a. Analyse de la structure 2D

La structure secondaire des ARN est connue pour pouvoir jouer un rôle important dans les régulations de l'épissage alternatif, en particulier, du fait qu'une forte structuration peut diminuer l'accessibilité du pré-ARNm aux composants de la machinerie d'épissage ou aux facteurs régulant l'épissage alternatif (Sirand-Pugnet *et al.*, 1995). Il nous a donc paru important d'essayer d'établir la structure 2D de la région 150-E5-150. Dans la mesure où cette région était nécessaire et suffisante pour que toutes les régulations aient lieu, nous avons formulé l'hypothèse qu'en absence des séquences qui la borde dans le pré-ARNm, elle doit adopter une structure 2D compatible avec la mise en place de toutes les régulations.

Avec I. Behm-Ansmnt, j'ai donc analysé la structure 2D de l'ARN 150-E5-150 en employant 3 sondes enzymatiques : les RNases V1, T1 et T2 décrites dans le Chapitre I de la partie Résultats. L'ARN 150-E5-150 a été produit par transcription avec l'ARN polymérase T7 en utilisant un produit PCR obtenu par l'utilisation d'un oligonucléotide amont contenant un site de fixation à la polymérase T7. Les hydrolyses enzymatiques ont été effectuées sur un ARN non marqué dans le tampon utilisé pour réaliser les expériences d'épissage *in vitro*, à 20°C dans les conditions décrites dans la partie Matériel et Méthodes. Les positions des coupures ont été identifiées par extension d'amorce par la transcriptase inverse suivie d'un fractionnement en gel dénaturant en parallèle avec les produits des 4 réactions de séquençage réalisée avec le même oligonucléotide amorce. Une extension d'amorce contrôle a été systématique réalisée en absence de clivage de l'ARN par les RNases, afin d'identifier les pauses de la transcriptase inverse. Afin de couvrir la totalité de l'ARN 150-E5-150, 3 oligonucléotides ont été utilisés pour amorcer la transcription inverse : 4390, 4391 et 6550. L'oligonucléotide 4390 s'hybride sur les positions -6 à -20 en amont de l'exon 5, l'oligonucléotide 4391 sur les positions +6 à +20 en aval de l'exon 5 et l'oligonucléotide 6550 sur les positions +139 à +150 en aval de l'exon (Fig.66).



La transcription inverse permet d'étudier les ARN de 3' en 5'. Il faut noter que dans le cas d'une hydrolyse par la RNase T1, qui libère des produits de clivage ayant une extrémité 5'-OH, l'incorporation des dNTP s'arrête au niveau du résidu précédant la liaison phosphodiester clivée, tandis que dans le cas des RNases V1 et T2, qui libèrent des produits de clivage ayant une extrémité 5'-P, l'incorporation des dNTP s'arrête au niveau du résidu dont la liaison phosphodiester a été clivée. Les analyses de structure 2D des ARN étant très délicates, elles ont toutes été répétées au moins trois fois et ceci sur des lots distincts d'ARN. Un modèle de structure 2D a pu être proposé en introduisant nos données expérimentales comme contraintes dans le programme M-Fold, permettant de prédire la structure 2D des ARN sur une base thermodynamique.

Selon le modèle ainsi établi, l'ARN 150-E5-150 possède un taux élevé de structuration (Fig. 67). Il comporte une longue hélice centrale irrégulière formée par les 31 nt en amont de l'exon 5, l'exon 5 et les 93 nt en aval de l'exon 5. Cette hélice irrégulière comporte plusieurs petites structures tige-boucle latérales, ainsi que des boucles latérales et des boucles internes. L'estimation de son énergie libre par le logiciel M-fold est de -74,58, elle n'est donc pas très stable. Le modèle proposé comporte aussi deux grandes structures tige-boucles formées l'une par les 45 nt 5' terminaux, l'autre par les nucléotides compris entre les positions +96 à +139 dans l'intron 5.

Dans ce modèle (Fig.68), les sites de fixation de MBNL1 du groupe 1 sont localisés dans des boucles et sont donc simple brin, comme en atteste leur clivage par les RNases T1 et T2. Les sites de fixation du groupe 2 sont situés dans une région qui semble se replier indépendamment du reste de l'ARN sous la forme de deux petites structures tige-boucles. Au sein de ce groupe de sites, le site 1 est clairement très accessible et en simple-brin vue la très forte intensité des clivages par les RNases T1 et T2, le site 3 est aussi clivé par les RNases T2 mais à un moindre degré. Par contre, nous n'avons observé aucun clivage pour le site 2, ni par les RNases clivant les simple-brins ni par la RNase V1 coupant les double-brins, nous n'avons donc pas d'information expérimentales sur l'état de structuration de ce site. Selon le modèle, il pourrait être impliqué dans la formation d'une hélice. Enfin, en ce qui concerne les sites du groupe 3, nous disposons de peu d'informations expérimentales, et selon le modèle les 3 sites seraient impliqués dans la formation d'hélices, ce qui expliquerait l'absence de clivage par les RNases T1 et T2.

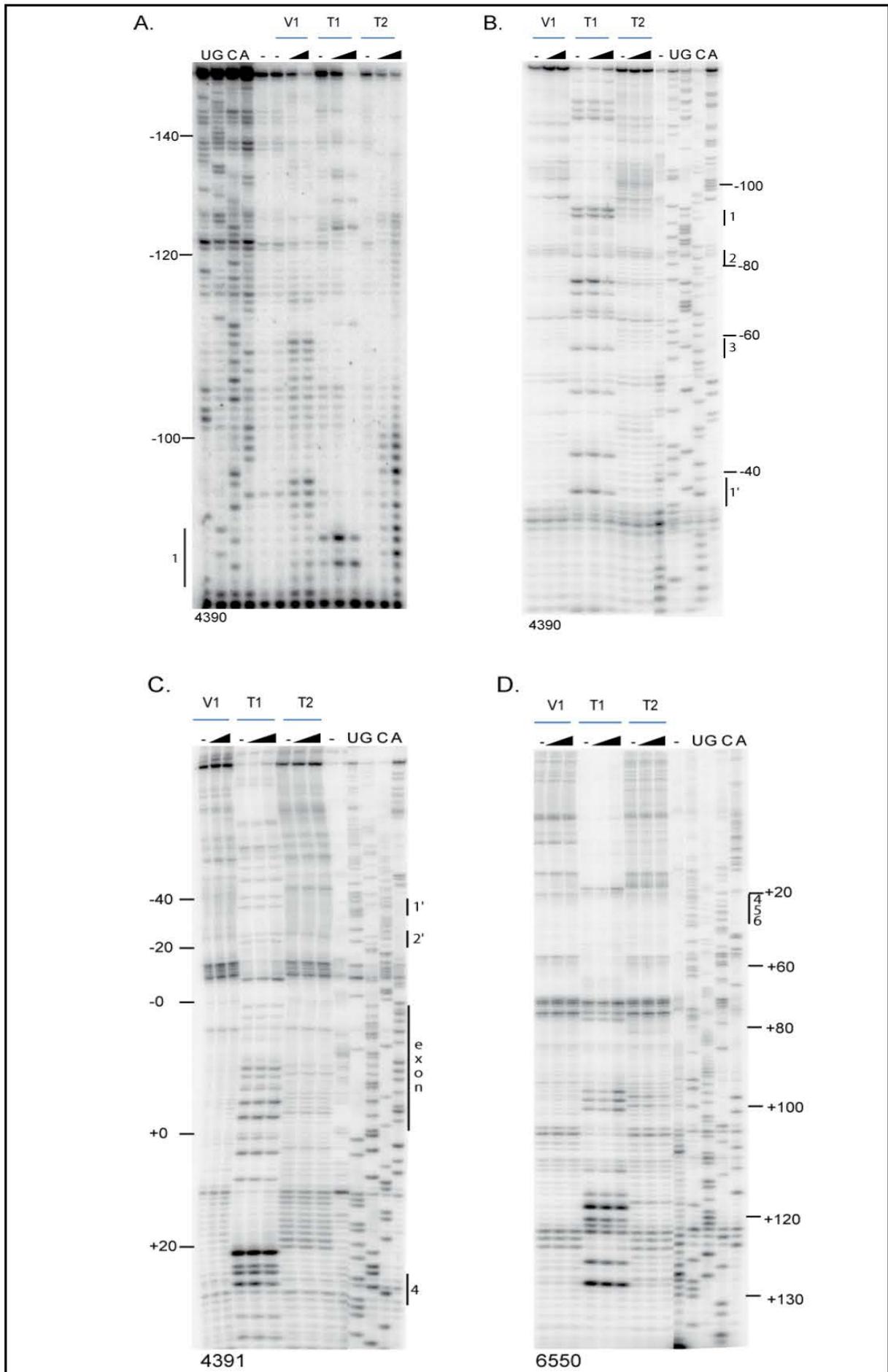


Figure 67 : Analyse des coupures enzymatiques de la région du pré-ARNm hcTNT requise pour récapituler l'ensemble des régulations et dérégulations de l'inclusion de l'exon 5.

Le transcrit a été soumis à une hydrolyse ménagée par les RNase V1 (10^{-3} ; 5.10^{-4} ; 10^{-4} UE), de RNase T1 (0.25 ; 0.5 et 1 UE) ou de RNase T2 (0.5 ; 1 et 2 UE) dans du tampon D 1X, comme décrit dans la partie Matériel et Méthodes. Les positions des coupures ont été identifiées par extension des oligonucléotides 4390 (A et B), 4391 (C) et 6550 (D) marqués en 5'. Les arrêts d'extension ont été positionnés dans la séquence grâce à des séquençages du transcrit réalisés avec les mêmes oligonucléotides (pistes U, G, C, A). Les ADNc ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 7%. Les positions des nucléotides sont indiquées à côté de l'autoradiographie. Les sites de fixation de MBNL1 et l'exon sont également indiqués.

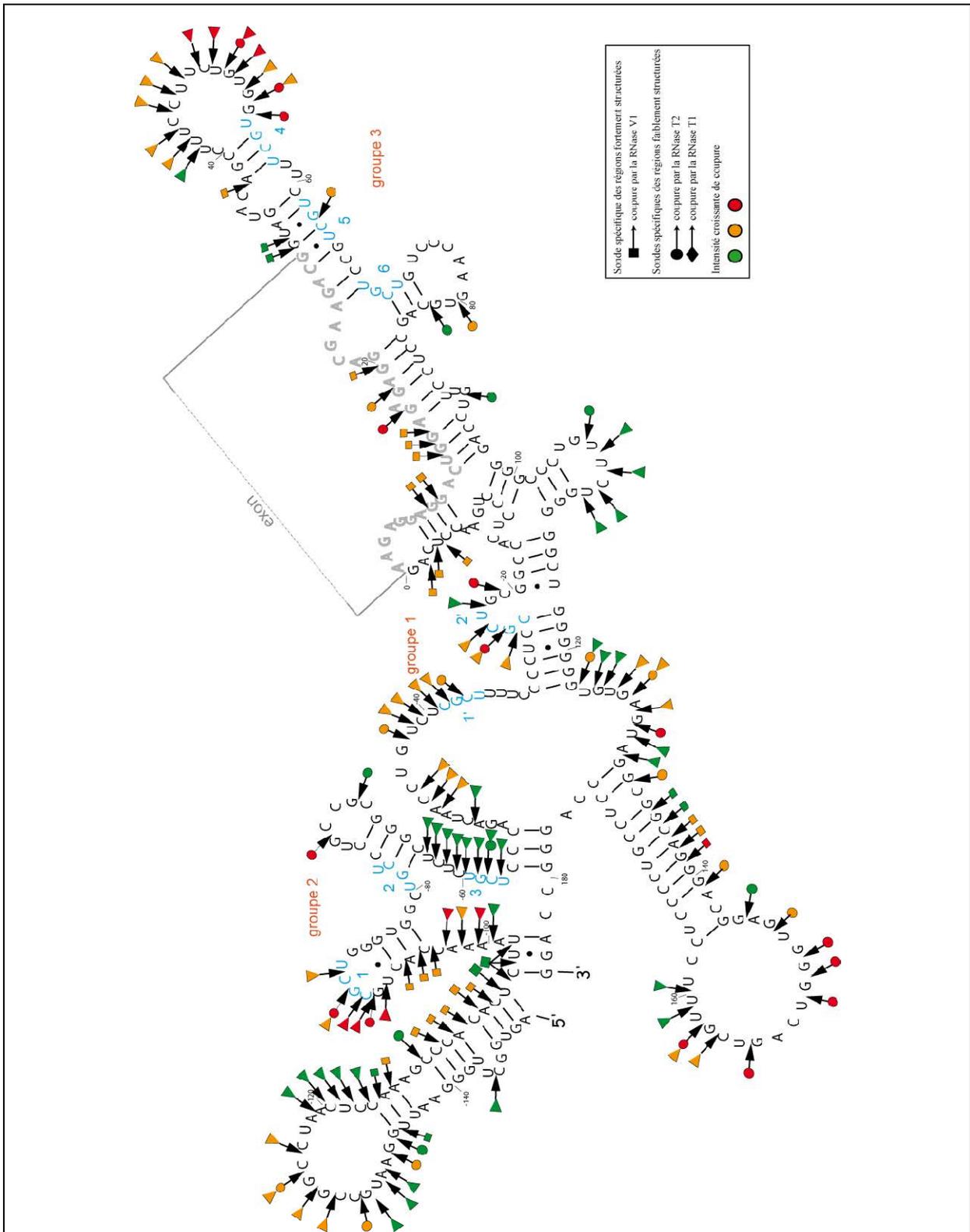


Figure 68 : Analyse de la structure 2D de la région du pré-ARNm hcTNT qui est requise pour récapituler l'ensemble des régulations et dérégulations de l'inclusion de l'exon 5.

Modèle de structure secondaire proposé pour l'ARN 150-E5-150. Les clivages par les RNases V1, T1 et T2 sont représentés par des carrés, des losanges et des ronds, respectivement. La couleur des symboles représente l'intensité du clivage (faible en vert, moyenne en orange et forte en rouge). La séquence de l'exon 5 est indiquée en gris. Les huit sites MBNL1 sont indiqués en bleu et les groupes de sites (groupes 1, 2 et 3) sont indiqués.

II.2.b. Protection de l'ARN 150-E5-150 par la protéine MBNL1_{CT3}

Disposant des données de structure, j'ai pu ensuite m'en servir pour réaliser des expériences d'empreinte sur les complexes formés par MBNL1_{ΔCT3} et l'ARN 150-E5-150. Pour cela, j'ai soumis le transcrit 150-E5-150 à une hydrolyse ménagée par les RNases V1, T1 et T2 (5.10⁻⁴ UE, 1 UE et 0,5 UE/1,2 fmol d'ARN) dans les mêmes conditions que ci-dessus mais après incubation en présence de deux concentrations différentes en protéine MBNL1_{ΔCT3}. Nous avons choisi ces 2 concentrations sur la base des données de retard sur gel réalisées avec des sous fragments de l'ARN 150-E5-150. Nous savions d'après ces données qu'à une concentration en protéine MBNL1_{ΔCT3} de 1 μM les sous fragments et donc vraisemblablement l'ARN 150-E5-150 étaient totalement complexés à la protéine. A la concentration supérieure 2 μM, nous devions donc avoir fixation de MBNL1_{ΔCT3} sur tous les sites. Nos données montraient qu'à la plus faible des 2 concentrations testées MBNL1_{ΔCT3} était déjà associée à tous les sites que nous avons détectés, puisque nous observions une protection spécifique de tous les résidus G des motifs YGCY identifiés comme sites potentiels. Nous confirmons ainsi les données obtenues par retard sur gel. De manière intéressante (Fig.69), nous n'observions que très peu de protections autres que celles des résidus G des motifs YGCY. En particulier, très peu de protections supplémentaires étaient détectées au niveau des coupures par la RNase T2, ce qui suggère que seul le résidu G est vraiment en forte interaction avec la protéine. Par ailleurs, l'association de MBNL1_{ΔCT3} semble très peu affecter la structure 2D de l'ARN puisque nous observions des clivages par la RNase V1 très similaires en présence et en absence de la protéine (Fig.70).

Nos données d'empreintes réalisées en solution sont étaient donc en accord avec l'existence de 3 segments distincts de fixation de MBNL1 sur les régions régulatrices de l'exon 5.

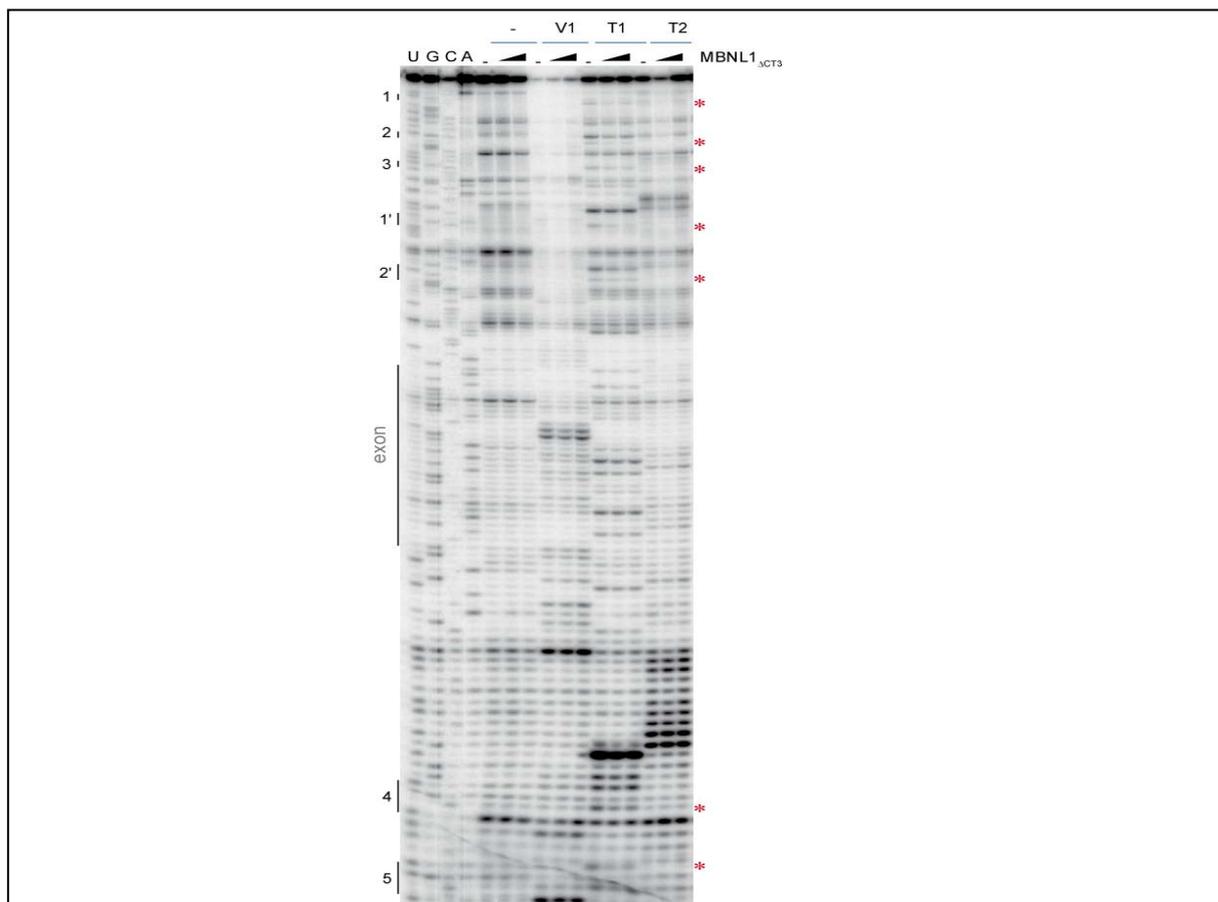


Figure 69 : Analyse des protections générées par la fixation de MBNL1 Δ CT3 sur l'ARNm hcTNT 150-E5-150.

Le transcrit a été soumis à une hydrolyse ménagée par les RNases V1 (5.10^{-4} UE) (pistes 5 à 7), T1 (1 UE) (pistes 8 à 10) et T2 (0,5 UE) après incubation en présence ou en absence (-) de deux concentrations de MBNL1 Δ CT3 (1 μ M et 2 μ M). Les conditions sont décrites dans la partie Matériel et Méthodes. Les positions des coupures ont été identifiées par extension des oligonucléotides 4391 marqués en 5'. Les arrêts d'extension ont été positionnés dans la séquence grâce à des séquençages du transcrit réalisés avec les mêmes oligonucléotides (pistes U, G, C, A). Les ADNc ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 7%. Les positions des nucléotides sont indiquées à gauche de l'autoradiographie. Les zones protégées sont indiquées à droite des autoradiographies par une étoile rouge.

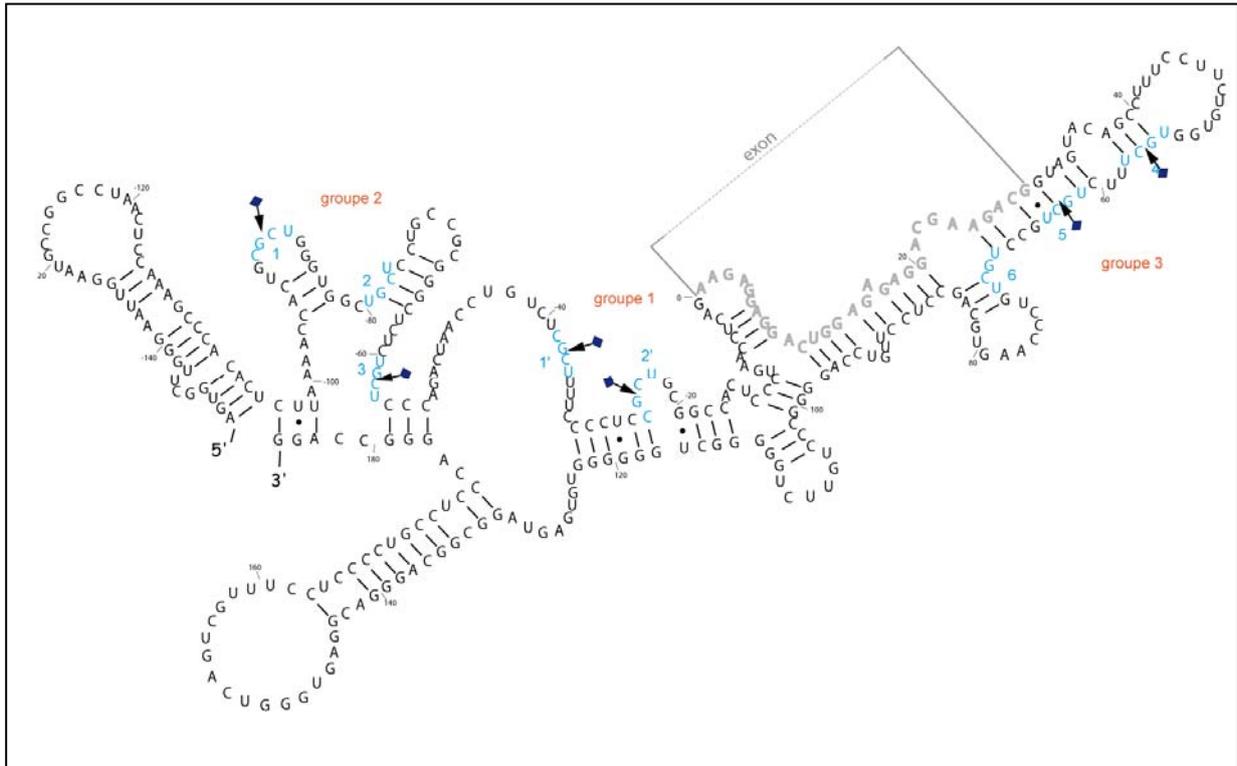


Figure 70 : Protection contre les RNases des résidus G des motifs YGCU de l'ARN 150-E5-150 par la protéine MBNL1 Δ CT3.

Les nucléotides ayant une plus faible sensibilité aux RNases en présence de la protéine MBNL1 Δ CT3 sont indiqués par des symboles bleus alors que ceux dont la réactivité est augmentée sont indiqués par des symboles roses. Ces symboles sont ceux utilisés précédemment pour indiquer les coupures par les RNases V1, T1 et T2 et les variations d'intensité du bleu et du rose sont corrélées à l'intensité des protections.

IV. Test de l'importance des différents groupes de sites de fixation de MBNL1 mis en évidence dans la régulation de l'inclusion de l'exon 5 *in cellulo*

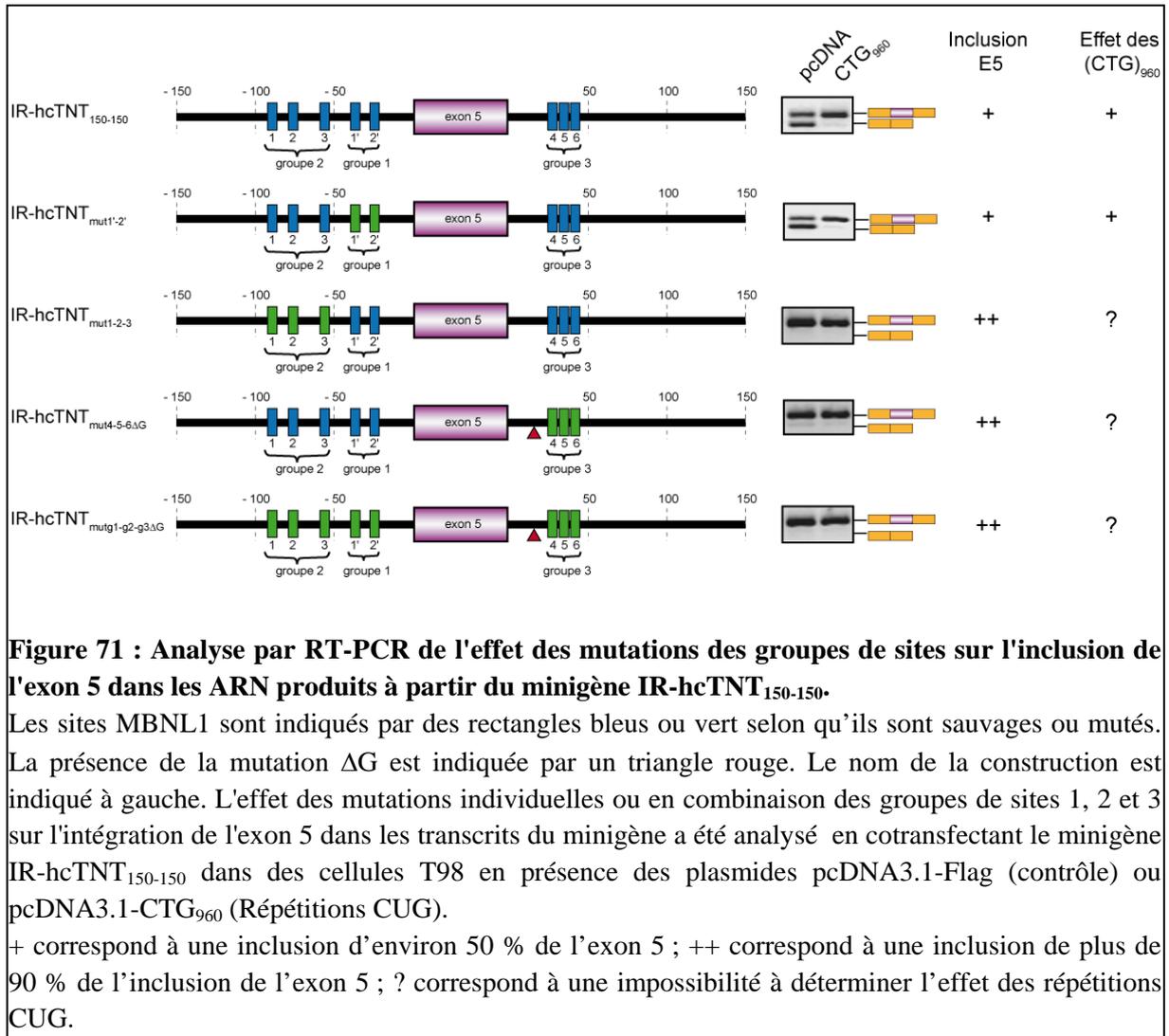
Afin de déterminer l'importance fonctionnelle des différents sites de fixation de MBNL1 que j'avais identifié par des expériences réalisées *in vitro*, nous avons décidé de tester l'effet de leur mutation sur l'incorporation de l'exon 5 *in cellulo*. Comme la mutation du résidu G des motifs YGCY en résidu A était suffisante pour abolir la fixation de MBNL1 sur les sites, nous avons choisi dans un premier temps de muter ensemble tous les sites d'un même groupe en introduisant les mutations YGCY \rightarrow YACY au sein du minigène IR-hcTNT₁₅₀₋₁₅₀. L'effet de ces mutations sur la régulation *in cellulo* de l'inclusion de l'exon 5 du pré-ARNm hcTNT a ensuite été testé par l'équipe de N. Sergeant. Au cours de ces expériences, il est apparu que la mutation du résidu G en A du site 4 entraînait l'apparition d'un site 5' cryptique d'épissage. La délétion du résidu G24 (Δ G) permettait d'abolir l'utilisation de ce site cryptique d'épissage et ne semblait pas interférer avec la régulation de l'épissage de

l'exon 5 aussi cette mutation a été systématiquement introduite en plus des 3 mutations G->A pour l'étude de l'importance des sites du groupe 3.

Les différents variants du minigène IR-hcTNT générés et testés in cellulo sont décrits dans la Figure 71. Le minigène muté au niveau des sites du groupe 1 a été appelé mut1'-2', le minigène muté au niveau des sites du groupe 2 a été appelé mut1-2-3 et le minigène muté au niveau des trois sites du groupe 3 et du résidu G24 a été nommé mut4-5-6ΔG. Enfin le minigène contenant l'ensemble des mutations, sites 1', 2', 1, 2, 3, 4, 5, 6 et G24 a été appelé mutg1-g2-g3ΔG.

Comme précédemment, les différents variants du minigène hcTNT ont été cotransfectés dans des cellules T98 avec une construction exprimant ou non des répétitions CTG insérées dans le contexte de séquence de la région 3' UTR du gène DMPK.

Les résultats obtenus, sont illustrés dans la Figure 71. Ils montrent que la présence de chacun des groupes de sites 2 et 3 est nécessaire à la régulation de l'inclusion de l'exon 5 du pré-ARNm hcTNT. En effet, lorsque les sites du groupe 2 (mut1-2-3) ou du groupe 3 (mut4-5-6ΔG) étaient mutés, une forte augmentation du taux d'inclusion de l'exon 5 était observée. Aucun effet sur l'inclusion de l'exon 5 n'a été observé en présence des répétitions CUG, cependant ce résultat était attendu du fait de l'inclusion quasi-totale de l'exon 5 en présence des mutations des sites du groupe 2 ou 3. De façon intéressante, le profil des produits d'épissage obtenu pour les mutants mut1-2-3 (groupe 2) et mut4-5-6ΔG (groupe 3) étaient identiques à celui obtenu lors de la mutation de l'ensemble des 8 sites de fixation de MBNL1 (mutg1g2g3ΔG). D'ailleurs dans nos expériences, la mutation des sites de fixation du groupe 1 (mut1'-2'), n'entraînait pas de modification du taux d'inclusion de l'exon 5 aussi bien en absence qu'en présence des répétitions CUG (Fig.71), suggérant que ces sites n'ont qu'une faible influence sur le taux d'inclusion de l'exon 5, contrairement à ce qui était proposé jusque là (Ho *et al.*, 2004).



V. Recherche des autres facteurs d'épissage qui pourraient être impliqués dans la régulation de l'épissage alternatif de l'exon 5 du pré-ARNm hcTNT

Etant donné que la région requise pour la régulation de l'inclusion de l'exon 5 s'étendait bien au delà des sites MBLN1 que nous venions d'identifier, nous avons formulé l'hypothèse que d'autres facteurs régulant l'épissage devaient intervenir dans les régulations mis en jeu.

Dans l'expérience qui avait été décrite ci-dessus pour définir la région requise pour les régulations, nous avons réalisé des réductions symétriques au niveau des introns 4 et 5. Afin d'avoir une définition plus précise des séquences régulatrices et de leur fonction, N. Marmier-Gourrier a réalisé de nouvelles délétions dans la construction IR-hcTNT₁₅₀₋₁₅₀, cette fois-ci asymétriques. L'effet de ces délétions sur la régulation *in cellulo* de l'inclusion de l'exon 5 du pré-ARNm hcTNT a ensuite été testé par l'équipe de N. Sergeant. Les données qu'ils ont obtenues en utilisant ces constructions pour transfecter des cellules T98 avec ou sans

expression de l'ARN (CUG)₉₆₀ sont illustrées dans la Figure 72. Ces données montrent que la séquence de 50 nt située en amont du groupe de site 2 est nécessaire à la régulation et qu'elle a des propriétés inhibitrices de l'inclusion de l'exon 5. Le segment situé entre les positions +50 et +150 est requis en aval de l'exon pour avoir toutes les régulations et s'est avéré avoir des propriétés activatrices. Par ailleurs, la région -150 à +50 est requise pour avoir une réponse aux répétitions CUG.

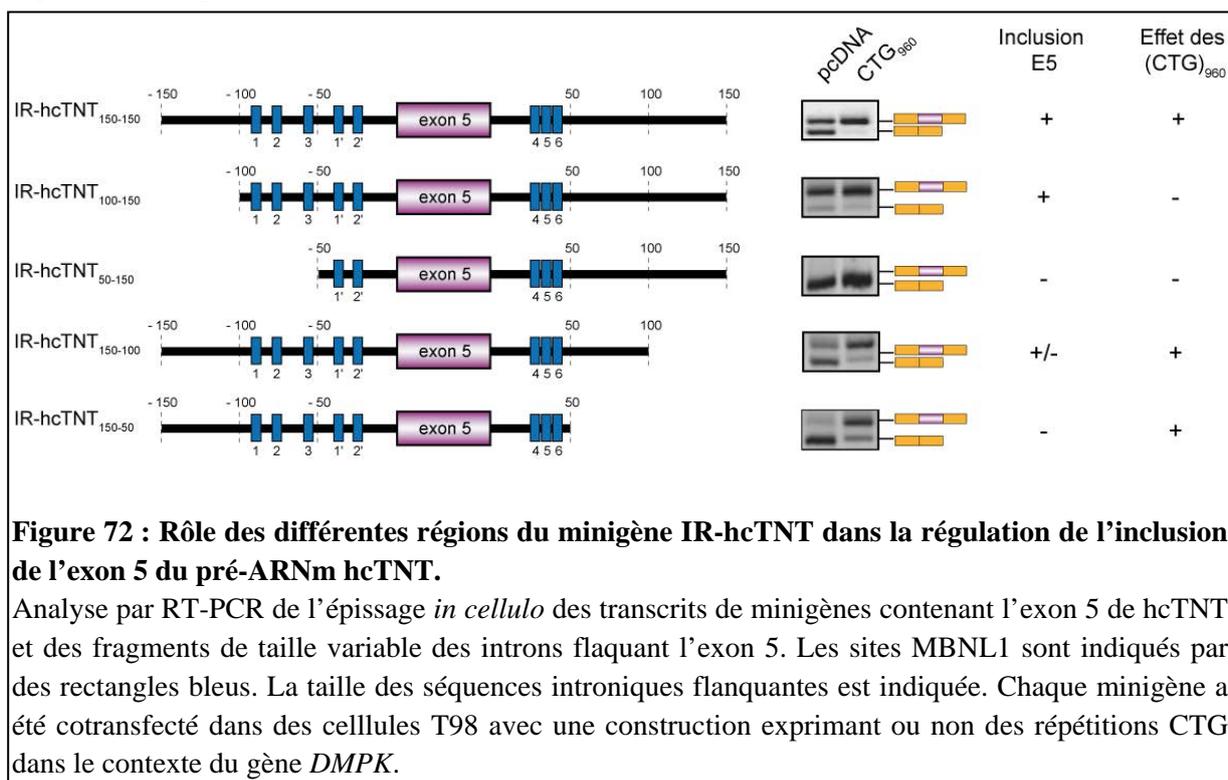
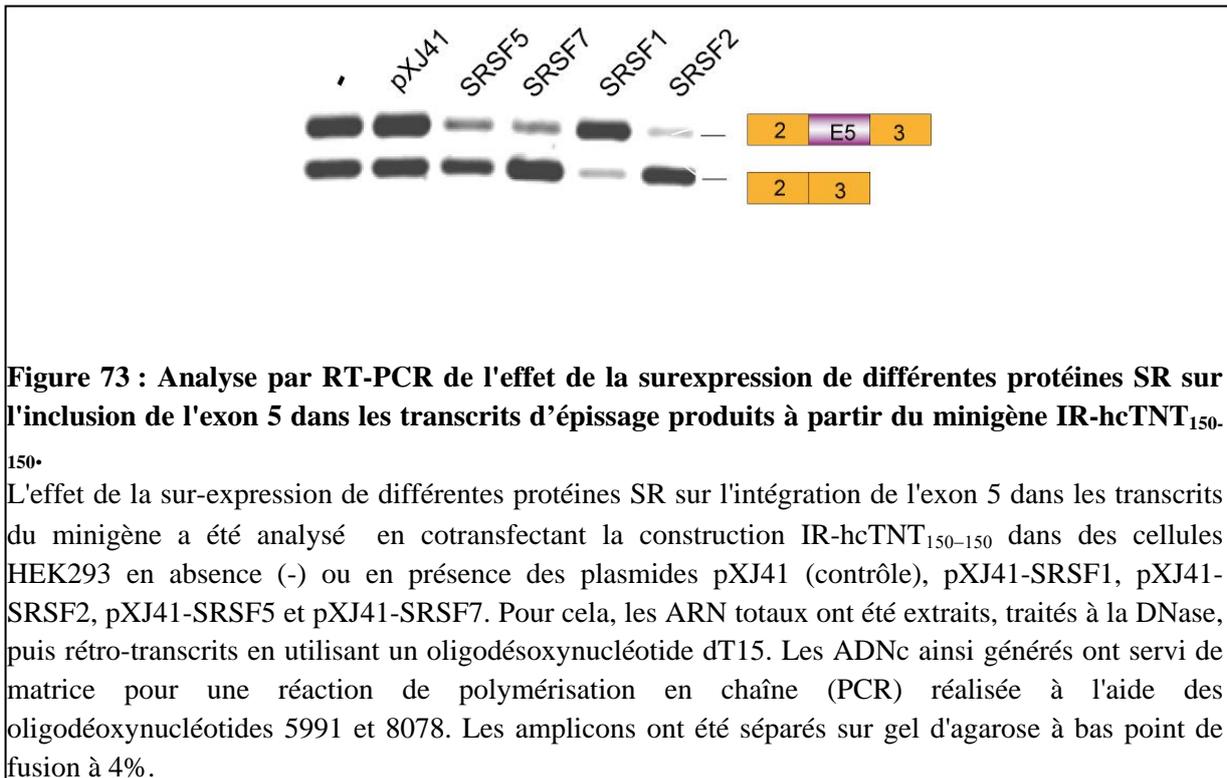


Figure 72 : Rôle des différentes régions du minigène IR-hcTNT dans la régulation de l'inclusion de l'exon 5 du pré-ARNm hcTNT.

Analyse par RT-PCR de l'épissage *in cellulo* des transcrits de minigènes contenant l'exon 5 de hcTNT et des fragments de taille variable des introns flaquant l'exon 5. Les sites MBNL1 sont indiqués par des rectangles bleus. La taille des séquences introniques flanquantes est indiquée. Chaque minigène a été cotransfecté dans des cellules T98 avec une construction exprimant ou non des répétitions CTG dans le contexte du gène *DMPK*.

Suite à ces résultats, I Behm-Ansmant a entrepris d'identifier les facteurs d'épissage impliqués dans la régulation de l'inclusion de l'exon 5, autres que PTB déjà mise en évidence (Charlet *et al.*, 2002), CUG-BP1 et MBNL1. Pour cela, le minigène IR-hcTNT₁₅₀₋₁₅₀ a été cotransfecté avec différentes constructions permettant de sur-exprimer différentes protéines SR : SRSF1 (ASF/SF2), SRSF2 (SC35), SRSF5 (SRp40) et SRSF7 (9G8). Les résultats qu'I. Behm-Ansmant a obtenus sont présentés dans la Figure 73. Les protéines SR sont connues pour être majoritairement des facteurs activateurs de l'épissage, alors que les protéines hnRNP sont au contraire plus généralement des inhibiteurs. Étonnamment, dans la construction IR-hcTNT₁₅₀₋₁₅₀ utilisée, la surexpression des protéines SR : SRSF2, 5 et 7 entraînait une diminution du taux d'inclusion de l'exon 5 (environ 20 % d'inclusion), elle pourrait donc être un répresseur de l'inclusion de l'exon 5. Seule la surexpression de la protéine SRSF1 entraînait une augmentation du taux d'inclusion de l'exon 5, suggérant que SRSF1 est un activateur de l'inclusion de l'exon 5.



Afin d'étudier le rôle de ces différentes protéines dans la régulation de l'épissage de l'exon 5, nous devons confirmer leur fixation sur l'ARN hcTNT 150-E5-150 par retard sur gel. J'ai réalisé ces expériences en utilisant un mélange de 100 fmoles d'ARN hcTNT 150-E5-150 non marqués et d'environ 0,6 fmoles de cet ARN radioactifs (10 cps) que j'ai incubé avec des quantités croissantes de ces protéines recombinantes présentes au laboratoire (concentrations allant de 125 nM à 4,2 µM), ceci dans les conditions décrites dans la partie Matériel et Méthodes. Ces protéines, fournies par J. Stevenin, avaient été produites dans le système baculovirus.

Les résultats obtenus, sont illustrés dans la Figure 74. Ils montrent que la protéine SRSF2 présente une faible affinité pour l'ARN (Kd apparent d'environ 3000 nM), de même aucun complexe n'est visible avec la protéine SRSF1. L'absence ou la faible fixation de ces protéines sur l'ARN suggère que soit l'effet de ces protéines sur l'inclusion de l'exon 5 est lié au contexte du minigène utilisé qui comportait les exons du récepteur de l'insuline de rat, soit que leur fixation à l'ARN est stabilisé par d'autres protéines. Pour les protéines SRSF5 et SRSF7, nous observons la formation de complexes au niveau des poches à des concentrations en protéines allant de 1 à 4 µM. Nous ne pouvons donc conclure sur une fixation des protéines à l'ARN ou à une agrégation des protéines entraînant une rétention de l'ARN dans les poches.

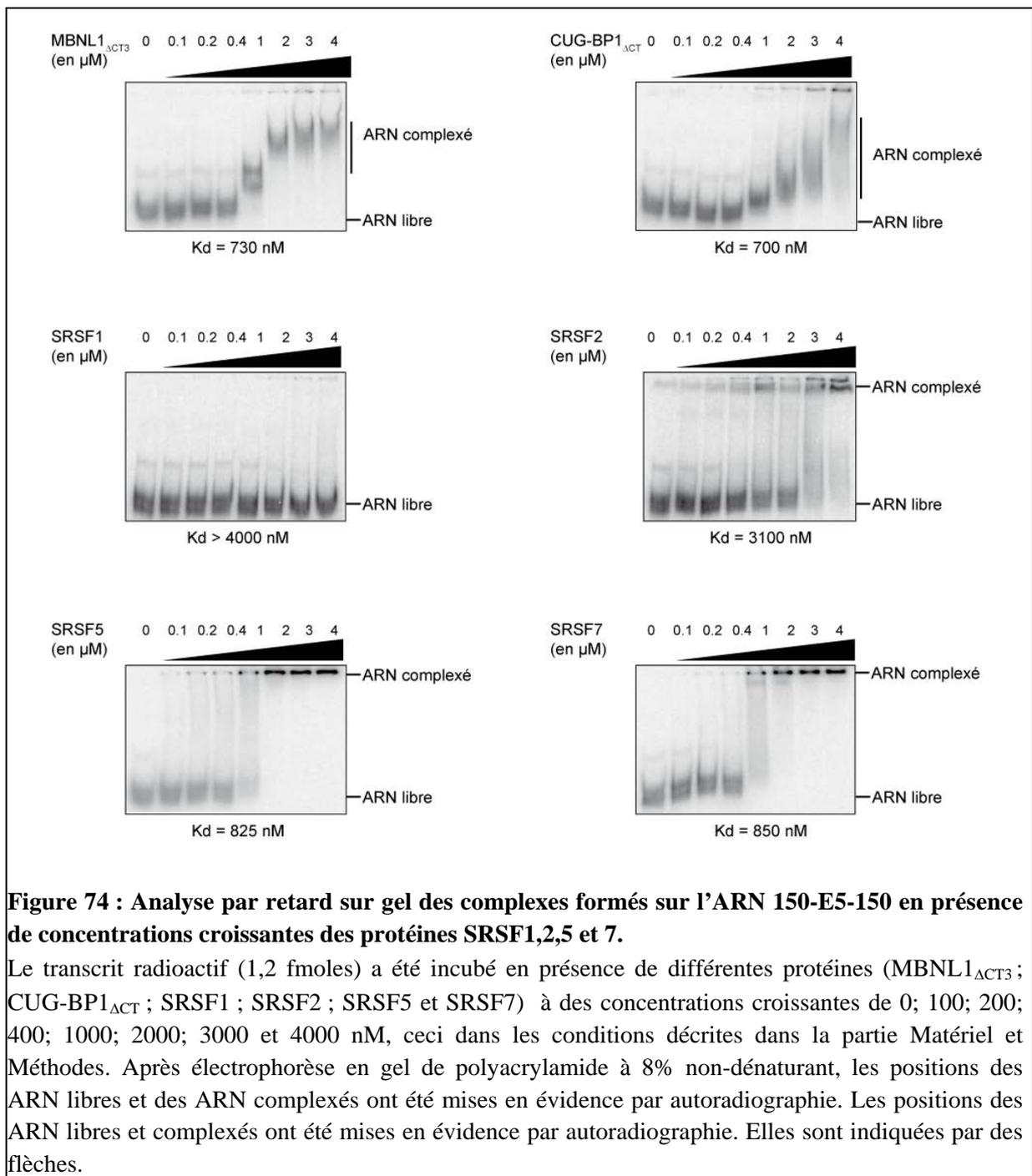


Figure 74 : Analyse par retard sur gel des complexes formés sur l'ARN 150-E5-150 en présence de concentrations croissantes des protéines SRSF1,2,5 et 7.

Le transcrit radioactif (1,2 fmoles) a été incubé en présence de différentes protéines (MBNL1_{ΔCT3} ; CUG-BP1_{ΔCT} ; SRSF1 ; SRSF2 ; SRSF5 et SRSF7) à des concentrations croissantes de 0; 100; 200; 400; 1000; 2000; 3000 et 4000 nM, ceci dans les conditions décrites dans la partie Matériel et Méthodes. Après électrophorèse en gel de polyacrylamide à 8% non-dénaturant, les positions des ARN libres et des ARN complexés ont été mises en évidence par autoradiographie. Les positions des ARN libres et complexés ont été mises en évidence par autoradiographie. Elles sont indiquées par des flèches.

VI. Conclusions et perspectives

Les données que nous avons obtenues apportent des informations qui vont être essentielles pour décortiquer les mécanismes en jeu dans la régulation de l'inclusion de l'exon 5 du pré-ARNm hcTNT. En effet, nous avons identifié 6 nouveaux sites de fixation de MBNL1 et montré qu'ils sont nécessaires au mécanisme de régulation. De manière surprenante la mutation des deux sites de fixation de MBNL1 (sites 1' et 2') identifiés par Ho et al n'abolit pas l'action négative de MBNL1 sur l'inclusion de l'exon 5. De plus, l'épissage

de cet exon est toujours dérégulé lors de la sur-expression de MBNL1 et répond à la présence des répétitions CUG. Ces résultats remettent en question le modèle proposé par l'équipe de Berglund, basé sur la formation d'une structure tige-boucle entre les sites 1' et 2', stabilisée par la fixation de MBNL1, qui empêcherait la reconnaissance de la séquence polypyrimidine par le facteur U2AF65 et par conséquent le recrutement de la snRNP U2 (Warf et Berglund, 2009). Dans cet article la confirmation de la fonctionnalité des sites 1' et 2' et de l'hypothèse de la formation d'une structure tige -boucle par appariement des sites 1' et 2' avait été obtenue par mutation de 4 résidus G, potentiellement impliqués dans l'appariement dont 2 appartenaient aux sites 1' et 2', en résidus U ou C (Fig.75). Ces mutations certes diminuent la fixation de MBNL1, cependant elles entraînent également un renforcement de la boîte de polypyrimidine, qui passe d'une longueur de 12 à 25 résidus. Nous ne pouvons donc pas exclure que l'effet qu'ils observaient sur l'inclusion de l'exon 5 soit en réalité non pas du à une perte de fixation de MBNL1 mais à la création d'un site 5' d'épissage plus fort. De plus, la structure de l'ARN 150-E5-150 que j'ai réalisée renforce notre hypothèse, en effet les sites 1' et 2' dans un contexte plus large que celui testé pour cette équipe ne forme pas une structure tige-boucle mais sont structurés en simple. Afin de confirmer notre hypothèse, il serait intéressant d'augmenter la taille de la séquence polypyrimidine sans mutation des sites de fixation de MBNL1.

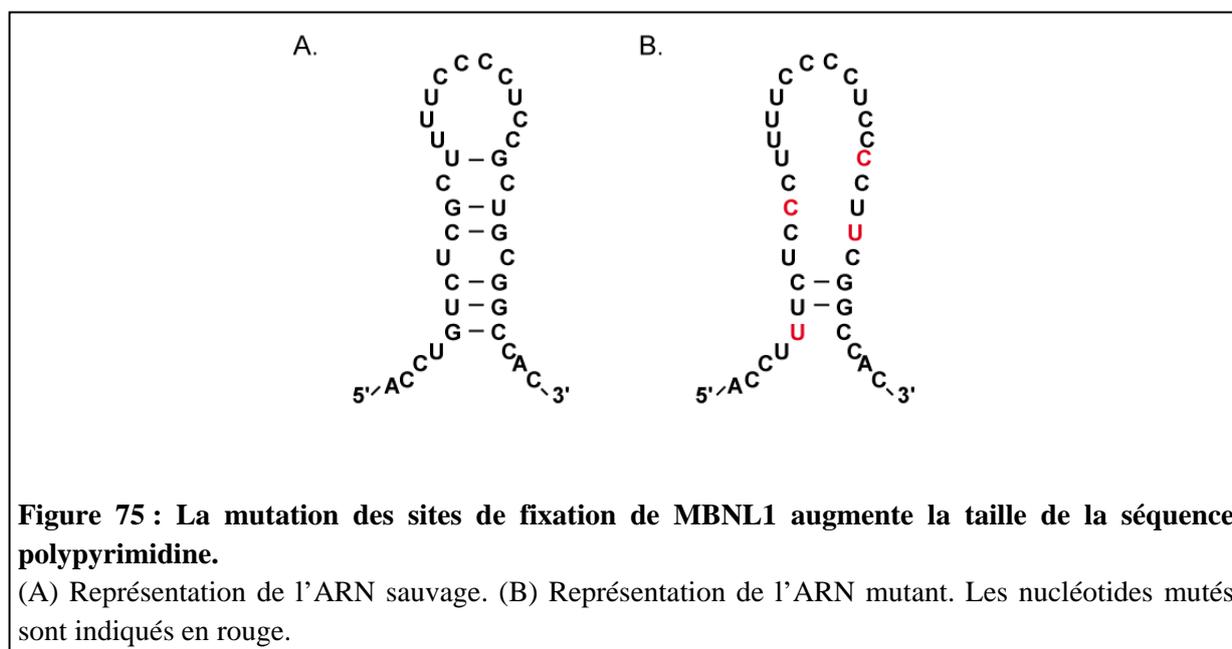


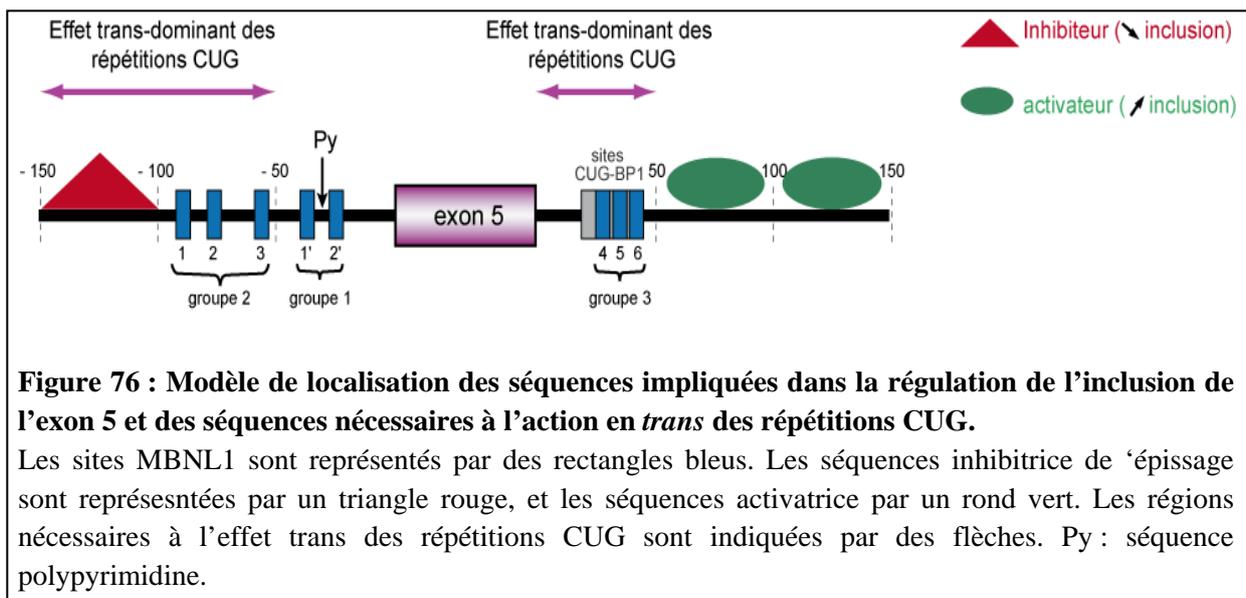
Figure 75 : La mutation des sites de fixation de MBNL1 augmente la taille de la séquence polypyrimidine.

(A) Représentation de l'ARN sauvage. (B) Représentation de l'ARN mutant. Les nucléotides mutés sont indiqués en rouge.

En revanche, nous observons que la mutation de l'un des deux groupes de sites (groupes 2 et 3) que nous avons identifiés limite fortement l'effet négatif de MBNL1 aux concentrations normales de MBNL1. Ainsi la présence simultanée de ces deux groupes de sites est nécessaire pour la régulation de l'inclusion de l'exon 5 du pré-ARNm hcTNT. Cette observation nous a conduit à proposer un nouveau modèle de régulation de l'inclusion de l'exon 5 du pré-ARNm hcTNT par MBNL1. Dans ce modèle, la régulation de cet événement d'épissage nécessiterait la formation d'une structure à longue distance au sein du pré-ARNm par interaction des protéines fixées sur les deux groupes de sites, ce qui isolerait l'exon 5 dans une boucle et limiterait la fixation des composants du spliceosome comme cela a déjà été proposé pour d'autres facteurs d'épissage comme hnRNP A1 et PTB (Wagner et Garcia-Blanco, 2001 ; Nasim *et al.*, 2002). Les données d'empreintes obtenues avec la protéine MBNL1_{ACT3} pourraient sembler contradictoires avec ce modèle puisque nous n'avons pas observé de changement de la structure 2D de l'ARN en présence de cette protéine. Cependant, cette absence de détection de changements structuraux pourrait s'expliquer par l'utilisation d'une protéine MBNL1 tronquée lors de nos expériences. En effet la protéine MBNL1_{ACT3} ne possède pas de propriétés de dimérisation et par conséquent ne pourrait pas former la structure à longue distance que nous proposons. C'est pourquoi il sera important de répéter les expériences d'empreintes en solution en utilisant la protéine GST-MBNL1₄₀ que nous avons utilisée lors des expériences d'empreintes sur les ARN (CUG)₁₆ et (CUG)₁₇. Par ailleurs l'existence de cette structure à longue distance, ainsi que son implication dans la régulation de l'inclusion de l'exon 5 devront être prouvées expérimentalement. Pour cela, 3 types d'expériences pourront être réalisées. Dans un premier temps, afin de vérifier la formation de la structure à longue distance nous proposons d'utiliser la méthode de FRET. Le FRET est une méthode de choix pour suivre les interactions moléculaires. Elle est sensible à la distance qui sépare deux fluorophores d'intérêt. En effet, le phénomène de FRET résulte d'un transfert d'énergie entre un fluorophore dit « donneur » et un fluorophore dit « accepteur ». En pratique, des oligodésoxynucléotides portant à leur extrémité un fluorophore d'intérêt (donneur ou accepteur) et s'hybridant à proximité des groupes de sites 2 et 3 seront hybridés sur l'ARN 150-E5-150 en présence ou en absence de la protéine MBNL1. La proximité dans l'espace de ces 2 oligodésoxynucléotides sera testée par leur capacité de transfert d'énergie. De façon complémentaire, la capacité des segments d'ARN contenant les groupes de sites 2 et 3 à interagir entre eux via la protéine MBNL1 pourra être testée en utilisant une variante de l'approche MS2/MBP (décrite dans le chapitre III de la partie Résultats). Un ARN hybride MS2-hcTNT₁₀₀₋₅₀ sera fixé sur des billes d'amylose via la protéine MS2/MBP et la rétention

de l'ARN hcTNT₀₋₅₀ radiomarqué sur les billes d'amylose en présence ou en absence de la protéine MBNL1 sera testée. Afin de tester la contribution éventuelle des propriétés d'oligomérisation de MBNL1 dans cette interaction, les expériences seront réalisées soit avec la protéine MBNL1₄₀ soit avec la protéine MBNL1_{ΔCT3}. Enfin, afin de montrer l'importance de cette structure à distance dans la régulation de l'inclusion de l'exon 5, nous remplacerons les sites des groupes 2 et 3 par des séquences complémentaires permettant de former une structure très stable en absence de la protéine MBNL1 et nous testerons l'effet de la formation de cette structure sur la régulation de l'inclusion de l'exon 5 de hcTNT.

Par ailleurs les données obtenues lors de l'utilisation des minigènes hcTNT « asymétriques » mettent en évidence la complexité de la régulation de l'inclusion de l'exon 5 hcTNT avec la présence d'un enchainement d'éléments régulateurs positifs et négatifs dans les séquences en amont et en aval de l'exon 5 s'étendant des positions -150 à 0 d'une part et 0 à +150 d'autre part (Fig. 76).



L'implication de différentes protéines SR et hnRNP a été testée par épissage *in cellulo* et j'ai pu confirmer la fixation de 3 d'entre elles sur l'ARN *in vitro*. Cependant les résultats obtenus ne sont pas très concluants et par cette approche nous n'avons ciblé que quelques-uns des facteurs d'épissage potentiellement impliqués dans cette régulation. C'est pourquoi, dans le but d'identifier de façon plus exhaustive les facteurs de régulation qui se fixent au niveau des éléments régulateurs présents dans les 150 nt en amont et en aval de l'exon 5, JV. Philippe (étudiant en M2, qui débute actuellement sa 1^{ère} année de thèse) a utilisé l'approche MS2/MBP. Cette approche, lorsqu'elle est couplée à une analyse par spectrométrie de masse

(cf chapitre III de la partie Résultats), permet la purification des complexes ribonucléoprotéiques formés sur un ARN d'intérêt et l'identification des protéines appartenant à ces complexes. J.V. Philippe et C. Aigueperse (IE, CNRS) ont généré des constructions permettant la production d'ARN hybrides hcTNT-MS2 contenant chacun un fragment de 50 nucléotides de l'ARN 150-E5-150. Les complexes formés sur ces ARN après incubation avec un extrait nucléaire ont été purifiés et l'analyse par spectrométrie de masse des protéines présentes au sein des complexes est en cours. Cependant l'analyse par western blot des complexes purifiés a d'ores et déjà permis de mettre en évidence l'association de la protéine hnRNP H sur les séquences régulatrices situées en aval de l'exon 5 et un rôle activateur de hnRNP H sur l'inclusion de l'exon 5 a été observé *in vitro* et *in cellulo*. L'étude du rôle de hnRNP H dans cette régulation sera poursuivie.

Chapitre III : Recherche de nouvelles protéines associées aux répétitions CUG : identification de la nucléophosmine

I. Recherche de nouvelles protéines associées aux répétitions CUG

Comme nous l'avons vu dans le chapitre Introduction, des données de la littérature suggéraient que d'autres protéines que MBNL étaient présentes dans les *foci* nucléaires renfermant les ARNm *DMPK* contenant des amplifications de triplets CUG. Certaines de ces protéines pourraient avoir un rôle dans le phénomène de rétention mais également dans les phénomènes de dérégulation de l'épissage (Kim *et al.*, 2005). Pour apporter des informations dans ce domaine, nous avons décidé d'identifier les protéines d'un extrait nucléaire de cellules HeLa capables de se lier à une longue amplification de triplets CUG (51 copies). Ce nombre de répétitions avait été choisi car il correspondait à celui connu pour déclencher la pathologie (pour revue, Usdin et Grabczyk, 2000). Pour cette recherche de partenaires, nous avons utilisé une méthode de chromatographie d'affinité dite « MS2/MBP », qui avait été appliquée par l'équipe de R. Lührmann du Max Planck Institut de Goettingen pour purifier des complexes pré-spliceosomaux. Elle avait été adaptée au laboratoire par I. Motorin, V. Marchand et C. Aigueperse dans l'objectif de purifier les complexes ribonucléoprotéiques (RNP) formés avec un segment d'ARN du VIH placé dans un extrait nucléaire de cellules HeLa.

I.1. Principe de la méthode de chromatographie d'affinité basée sur le système MS2/MBP

Cette méthode repose sur l'insertion à l'extrémité 3' de l'ARN d'intérêt, de trois motifs de l'ARN du bactériophage MS2 d'*E. coli* (appelés par la suite motifs MS2) qui sont des sites de fixation de la protéine de l'enveloppe de ce phage. L'affinité entre les partenaires ARN MS2/protéine de l'enveloppe du phage MS2 est très forte (Kd de l'ordre du nM) (Jurica *et al.*, 2002). Le domaine de fixation à l'ARN de la protéine de l'enveloppe de ce phage MS2 est produit sous forme fusionnée avec la protéine MBP (Maltose Binding Protein), ce qui permet de fixer ce domaine sur des billes d'amylose et ensuite d'immobiliser l'ARN « hybride » sur les billes auxquelles la protéine de fusion MS2-MBP est fixée (Jurica *et al.*, 2002). L'ARN « hybride » ainsi immobilisé peut être incubé avec un extrait nucléaire afin de former des complexes ribonucléoprotéiques. Après plusieurs étapes de lavages destinées à éliminer les protéines liées de manière non spécifique, le mélange de complexes formés sur l'ARN est élué en conditions non-dénaturantes par ajout de maltose (Fig.77). Les protéines du

mélange de complexes purifiés sont ensuite séparées par électrophorèse en gel SDS-PAGE. Après coloration au bleu de coomassie colloïdal, les bandes de gel sont découpées et les protéines qu'elles contiennent sont soumises à des analyses par spectrométrie de masse, afin d'identifier les partenaires protéiques de l'ARN.

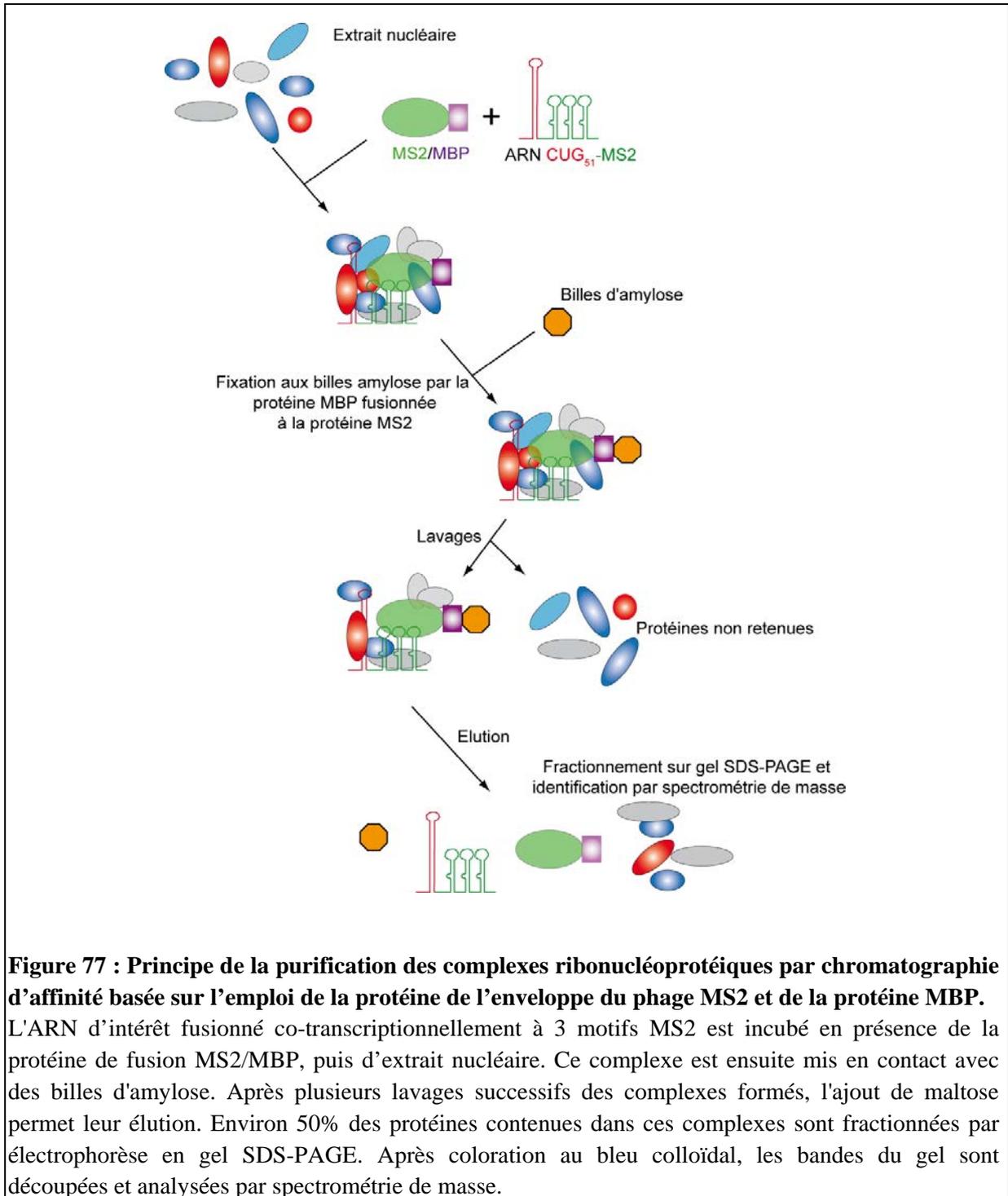


Figure 77 : Principe de la purification des complexes ribonucléoprotéiques par chromatographie d'affinité basée sur l'emploi de la protéine de l'enveloppe du phage MS2 et de la protéine MBP. L'ARN d'intérêt fusionné co-transcriptionnellement à 3 motifs MS2 est incubé en présence de la protéine de fusion MS2/MBP, puis d'extrait nucléaire. Ce complexe est ensuite mis en contact avec des billes d'amylose. Après plusieurs lavages successifs des complexes formés, l'ajout de maltose permet leur élution. Environ 50% des protéines contenues dans ces complexes sont fractionnées par électrophorèse en gel SDS-PAGE. Après coloration au bleu colloïdal, les bandes du gel sont découpées et analysées par spectrométrie de masse.

I.2. Constructions utilisées

La première étape a consisté à réaliser une construction permettant de produire par transcription *in vitro* un ARN « hybride » appelé ARN (CUG)₅₁-MS2 comportant 51 répétitions CUG (CUG)₅₁ fusionnées à 3 motifs MS2 (Fig.78).

Nom de la construction	Schéma de la construction	Enzyme de restriction utilisée pour linéariser le plasmide	ARN transcrits
pUC18-CUG ₅₁			
pUC18-T7-CUG ₅₁		KpnI	
pUC18-T7-CUG ₅₁ -MS2		SacI	
pUC18-T7-MS2		SacI	
pUC18-T7-CAG ₅₉ -MS2		SacI	

- ADN codant la séquence correspondant au promoteur de l'ARN polymérase du phage T7
- ADN codant les répétitions CUG₅₁
- ADN codant les trois motifs MS2
- ADN codant les répétitions CAG₅₉

Figure 78 : Constructions réalisées pour les expériences de purification par chromatographie d'affinité MS2/MBP des RNP formées sur les répétitions CUG et CAG après incubation en présence d'extrait nucléaire de cellules HeLa.

Pour cela, j'ai utilisé le vecteur pUC18 contenant la séquence codant l'ARN (CUG)₅₁, qui avait été construit par N. Marmier-Gourrier (pUC18-CTG₅₁). J'ai utilisé les sites de restrictions uniques *XbaI* et *BamHI* afin d'insérer, d'abord en amont des répétitions (CTG)₅₁, une séquence correspondant à un promoteur reconnu par l'ARN polymérase du phage T7. Pour réaliser cela, j'ai utilisé deux oligodésoxyribonucléotides complémentaires (4915/4916) comportant la séquence ADN promotrice et, à l'extrémité 5' de l'un et l'extrémité 3' de l'autre respectivement, une séquence générant un site *XbaI* et une séquence générant un site

*Bam*HI. Le plasmide recombinant obtenu (pUC18-T7-CTG₅₁) permettait, après linéarisation par l'enzyme *Kpn*I, de produire l'ARN (CUG)₅₁ par transcription *in vitro*.

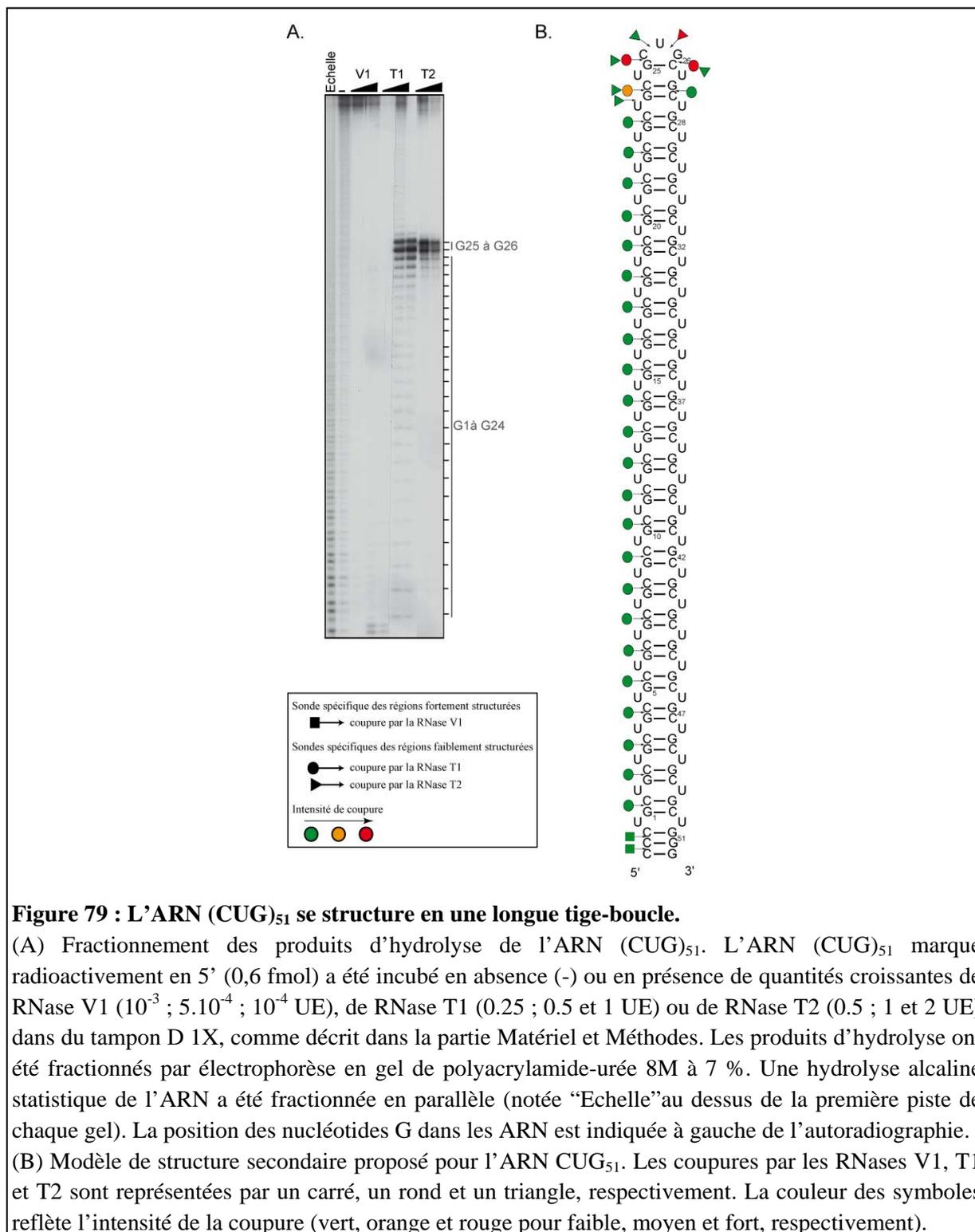
J'ai ensuite inséré dans ce plasmide, en aval des répétitions (CTG)₅₁ un segment codant 3 motifs MS2, ceci en utilisant les sites de restriction *Kpn*I et *Sac*I. Ce segment avait été préalablement amplifié par PCR à partir du plasmide pAdML3 fourni par R. Lührmann ceci à l'aide du couple d'oligodésoxyribonucléotides 4913/4914. Le plasmide recombinant obtenu (pUC18-T7-CTG₅₁-MS2) permettait, après linéarisation par l'enzyme *Sac*I, de produire par transcription *in vitro* un ARN hybride comportant les répétitions (CUG)₅₁, suivies des 3 motifs MS2.

La séquence codant les 3 motifs MS2 précédés du promoteur du phage T7 a également été insérée dans le plasmide pUC18 par la même approche. Après linéarisation par l'enzyme de restriction *Sac*I, le plasmide recombinant obtenu (pUC18-T7-MS2) permettait de produire par transcription *in vitro* un ARN comportant uniquement les 3 motifs MS2, ARN qui nous a servi de contrôle.

I.3. Etude de la structuration de l'ARN CUG₅₁

La construction du plasmide pUC18-CTG₅₁ contenant la séquence codant pour l'ARN (CUG)₅₁ avait été réalisée avant mon arrivée par N. Marmier-Gourrier. Mon premier travail a consisté à vérifier la structuration et l'homogénéité de l'ARN produit à partir de cette construction. Pour cela, l'ARN a été marqué en 5' par utilisation d'ATP[γ -³²P], et nous avons réalisé des hydrolyses enzymatiques par les RNases V1, T1 et T2. Les résultats obtenus sont illustrés dans la Figure 79.

L'hydrolyse par la RNase V1 n'a pas été satisfaisante, en revanche, les hydrolyses par les RNases T1 et T2 mettent clairement en évidence la présence d'une boucle terminale correspondant à la 26^{ème} répétition CUG, comme nous l'avons montré pour l'ARN (CUG)₁₇ au Chapitre I de la partie Résultats. L'ARN (CUG)₅₁ adopte bien une structure tige-boucle avec, vraisemblablement, une boucle terminale à 3 résidus. On note également que les paires de bases adjacentes à la boucle terminale ont une plus faible stabilité que celles du reste de l'hélice.



I.4. Mise au point des conditions de purification des RNP

Dans un premier temps, nous avons dû :

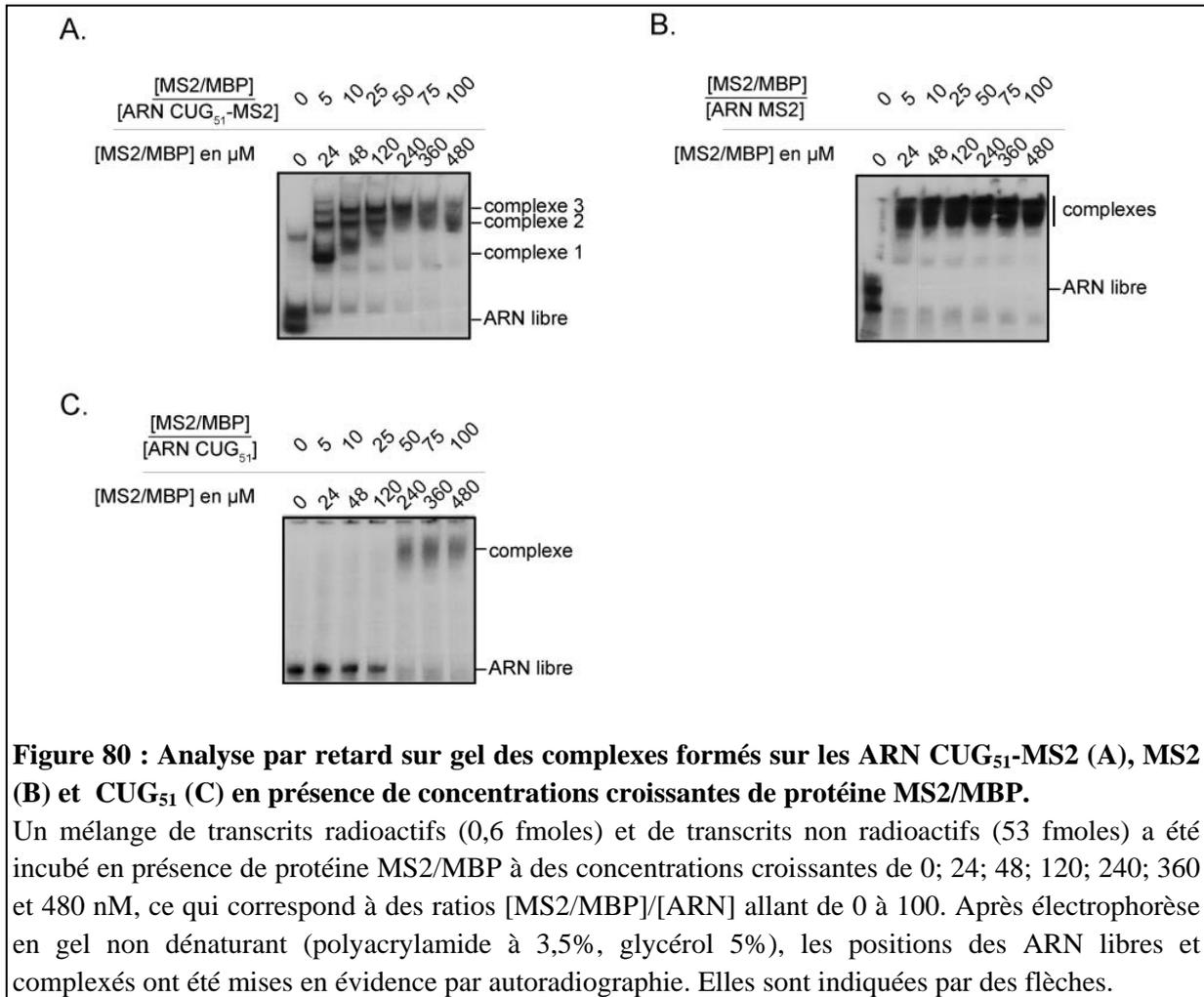
1) vérifier la capacité et la spécificité de fixation de la protéine de fusion MS2/MBP produite sous forme recombinante chez *E. coli* sur l'ARN (CUG)₅₁-MS2

2) définir les concentrations respectives d'ARN et de protéines à utiliser pour avoir une fixation efficace et spécifique de l'ARN.

Pour cela, nous avons réalisé des expériences de retard sur gel, en utilisant un mélange d'ARN uniformément marqué par incorporation d'UTP[α - 32 P] (environ 0,6 fmoles) et d'ARN non marqués (environ 53 fmoles). Ces ARN ont été produits par transcription *in vitro* par l'ARN polymérase du phage T7 à partir des matrices pUC18-T7-CTG₅₁-MS2 et pUC18-T7-MS2 linéarisées avec *Sac*I et de la matrice pUC18-T7-CTG₅₁ linéarisée avec *Kpn*I. Le transcrit (CUG)₅₁ servait aussi de contrôle. Les mélanges d'ARN froid et radioactif ont été incubés avec des quantités croissantes de la protéine de fusion MS2/MBP (concentrations allant de 24 nM à 480 nM) dans les conditions décrites dans la partie Matériel et Méthodes.

Comme le montre la Figure 80, la majorité des molécules d'ARN CUG₅₁-MS2 présentait un retard de migration pour un rapport molaire [protéine]/[ARN] de 5. Plusieurs bandes de complexes étaient observées, ce qui s'explique sans doute par la présence des trois sites de fixation de la protéine MS2, qui devaient être occupés séquentiellement lorsque les concentrations croissantes étaient utilisées. Dans le cas de l'ARN MS2, la totalité de l'ARN avait également une migration retardée pour un rapport molaire [protéine]/[ARN] de 5 (Fig. 80), cependant une seule bande était observée, ce qui suggérait une fixation coopérative de la protéine MBP-MS2 sur les trois sites. Cette coopérativité semblait être perdue dans le cas de l'ARN hybride. Un retard sur gel était également détecté dans le cas de l'ARN (CUG)₅₁, mais ceci uniquement en présence d'un très fort excès de protéine (rapport molaire [protéine]/[ARN] de 50) (Fig. 80). Pour un rapport [protéine]/[ARN] compris entre 5 et 50, nous avons donc une bonne spécificité de reconnaissance des motifs MS2 par la protéine de fusion sans avoir de fixation aspécifique sur la séquence (CUG)₅₁.

Nous avons donc décidé de poursuivre les études en utilisant un rapport molaire [protéine]/[ARN] de 5 pour purifier les complexes. Ce rapport permettait une fixation efficace tout en préservant la spécificité et qui nous permettait aussi d'utiliser moins de protéine MS2/MBP, qui après électrophorèse donne une très grosse bande susceptible de gêner la détection des partenaires de l'ARN ayant une masse moléculaire similaire à celle de MS2/MBP.



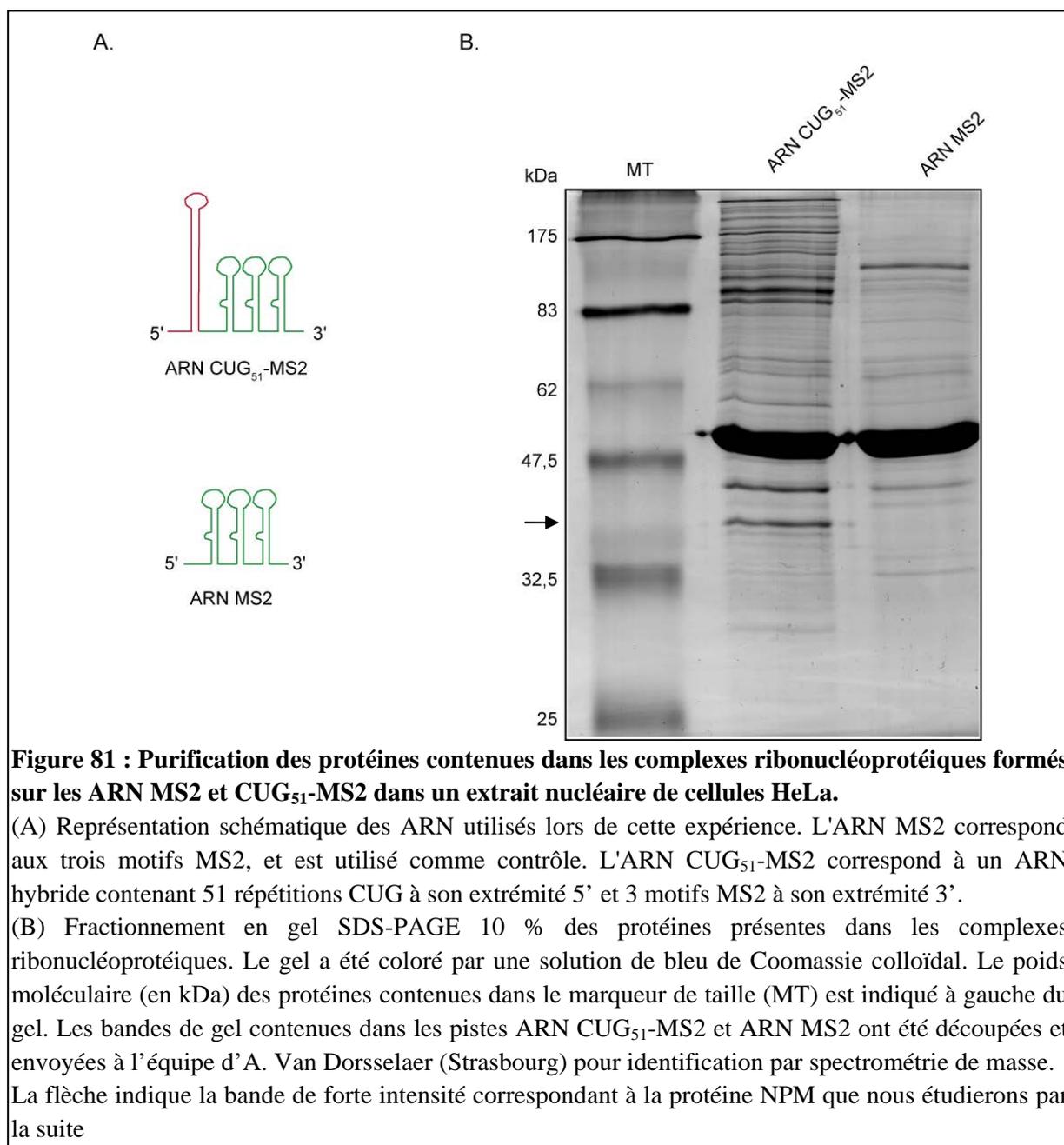
I.5. Purification des RNP formées avec l'ARN (CUG)₅₁-MS2 dans un extrait nucléaire de cellules HeLa

Les complexes RNP ont été formés par incubation de 500 pmoles d'ARN (CUG)₅₁-MS2 froid avec 2500 fmoles de la protéine MS2/MBP, puis avec 500 μl de protéines d'un extrait nucléaire de cellules HeLa. Les complexes RNP ainsi formés ont alors été fixés sur billes d'amylose, purifiés par lavages successifs puis élués en présence de maltose. Une expérience contrôle a été réalisée dans les mêmes conditions avec l'ARN MS2.

Pour que l'approche soit pertinente, il fallait que le contenu en protéines des RNP formées avec l'ARN-MS2 contrôle et l'ARN hybride soient différents. Le contenu en protéines des fractions d'éluion a été analysé par électrophorèse en conditions dénaturantes (SDS-PAGE), suivie d'une coloration au bleu de Coomassie colloïdal. Les profils de fractionnement des protéines obtenus après éluion des complexes formés avec chacun des 2 ARN étaient bien différents (Fig.81). De plus, étant donné qu'une coloration au bleu de Coomassie colloïdal

permettait de détecter efficacement les protéines éluées, nous avons conclu que les quantités protéiques purifiées étaient suffisantes pour une identification par spectrométrie de masse.

L'analyse par spectrométrie de masse a été réalisée par S. Sanglier et L. Fouillen de l'équipe d'A. Van Dorsselaar (LSMBO, Strasbourg). Les bandes de gel découpées ont été traitées à la trypsine une nuit à température ambiante. Les peptides générés à partir de chaque bande ont ensuite été analysés par spectrométrie de masse en tandem (MS/MS). L. Fouillen a ensuite interprété les spectres obtenus en utilisant différentes banques de données (NCBI, Swissprot).



I.6. Bilan de l'analyse par spectrométrie de masse des complexes formés par l'ARN (CUG)₅₁-

Au total l'analyse par spectrométrie de masse a permis d'identifier 148 protéines. Contrairement à ce qui était attendu, la protéine MBNL1 n'a pas été retrouvée parmi les protéines identifiées. A l'inverse, la protéine hnRNP F, identifiée dans les *foci* DM (Jiang et al., 2004), a été détectée au sein des RNP purifiées. Parmi les 148 protéines identifiées, 101 étaient uniquement associées aux répétitions CUG (Tableau XII), 17 sont plus fortement associées aux répétitions CUG (Tableau XIII) et 30 étaient majoritairement ou uniquement associées à l'ARN MS2. Les analyses par spectrométrie de masse réalisées ne permettant pas d'avoir des données quantitatives, l'association majoritaire à un ARN donné a été déterminée par comparaison du nombre de peptides identifiés et du pourcentage de recouvrement de séquence de la protéine. Parmi les 101 protéines uniquement trouvées associées aux répétitions CUG, on note un enrichissement en protéines ribosomales, protéines hnRNP, facteurs d'épissage, et hélicases.

Etant donné que beaucoup d'expériences de chromatographie de RNP basées sur le système MS2 avaient été réalisées au laboratoire et qu'il en ressortait que certaines protéines nucléaires étaient systématiquement retrouvées quel que soit l'ARN d'intérêt, nous avons décidé du moins dans un premier temps, d'éliminer ces protéines de notre liste et de nous intéresser à celles retrouvées spécifiquement avec l'ARN (CUG)₅₁-MS2. Nous avons ainsi pu réduire la liste à 55 protéines pouvant représenter des candidats intéressants. Cependant ce nombre de protéines était toujours trop important pour réaliser une étude exhaustive de chacune d'entre elles. L'interaction entre les répétitions CUG et leurs protéines partenaires étant supposée forte, nous avons alors envisagé de modifier notre protocole expérimental afin d'être plus stringent au niveau de la sélection des protéines partenaires.

Numéro d'accèsion	Noms des protéines	Taxonomie	Masse
hnRNPs			
P07910 HNRPC_HUMAN	Heterogeneous nuclear ribonucleoproteins C1/C2 (hnRNP C1 / hnRNP C2)	Homo sapiens	33650
P52597 HNRPF_HUMAN	Heterogeneous nuclear ribonucleoprotein F (hnRNP F) (Nucleolin-like protein mcs9)	Homo sapiens	45512
O75643 U520_HUMAN	U5 small nuclear ribonucleoprotein 200 kDa helicase (EC 3.6.1.-) (U5 snRNP-specific)	Homo sapiens	244353
P14678 RSMB_HUMAN ou P63162 RSMN_HUMAN	Small nuclear ribonucleoprotein-associated proteins B and B' (snRNP-B) ou Small nuclear ribonucleoprotein-associated protein N (snRNP-N)	Homo sapiens	24594
Facteurs d'épissage			
O75533 SF3B1_HUMAN	Splicing factor 3B subunit 1 (Spliceosome-associated protein 155) (SAP 155) (SF3B1)	Homo sapiens	145723
Q07955 SFRS1_HUMAN	Splicing factor, arginine/serine-rich 1 (pre-mRNA-splicing factor SF2, P33 subunit)	Homo sapiens	27597
P84103 SFRS3_HUMAN	Splicing factor, arginine/serine-rich 3 (Pre-mRNA-splicing factor SRP20)	Homo sapiens	19318
Histones			
P16401 H15_HUMAN	Histone H1.5 (Histone H1a) - Homo sapiens	Homo sapiens	22435
P16403 H12_HUMAN	Histone H1.2 (Histone H1d) - Homo sapiens	Homo sapiens	21221
Q09028 RBBP4_HUMAN	Histone-binding protein RBBP4 (Retinoblastoma-binding protein 4) (RBBP-4)	Homo sapiens	47495
Facteurs de réplication			
P35251 RFC1_HUMAN	Replication factor C subunit 1 (Replication factor C large subunit) (RF-C 140 kDa)	Homo sapiens	128175
P35250 RFC2_HUMAN	Replication factor C subunit 2 (Replication factor C 40 kDa subunit) (RF-C 40 kDa)	Homo sapiens	39132
P40938 RFC3_HUMAN	Replication factor C subunit 3 (Replication factor C 38 kDa subunit) (RFC38)	Homo sapiens	40530
P35249 RFC4_HUMA	Replication factor C subunit 4 (Replication factor C 37 kDa subunit) (RF-C 37 kDa)	Homo sapiens	39657
P40937 RFC5_HUMAN	Replication factor C subunit 5 (Replication factor C 36 kDa subunit) (RF-C 36 kDa)	Homo sapiens	38472
Protéines ribosomales			
P15880 RS2_HUMAN	40S ribosomal protein S2 (S4) (LLRep3 protein)	Homo sapiens	31305
P23396 RS3_HUMAN	40S ribosomal protein S3	Homo sapiens	26671
P08865 RSSA_HUMAN	40S ribosomal protein SA (p40) (34/67 kDa laminin receptor)	Homo sapiens	32702
P62701 RS4X_HUMAN	40S ribosomal protein S4, X isoform (Single copy abundant mRNA protein) (SCR10)	Homo sapiens	29448
P62753 RS6_HUMAN	40S ribosomal protein S6 (Phosphoprotein NP33) - Homo sapiens	Homo sapiens	28663

vP62241 RS8_HUMAN	40S ribosomal protein S8	Homo sapiens	24059
P46781 RS9_HUMAN	40S ribosomal protein S9	Homo sapiens	22447
P39023 RL3_HUMAN	60S ribosomal protein L3 (HIV-1 TAR RNA-binding protein B) (TARBP-B)	Homo sapiens	45949
P62906 RL10A_HUMAN	60S ribosomal protein L10a (CSA-19)	Homo sapiens	24684
P40429	60S ribosomal protein L13a	Homo sapiens	23460
P50914 RL14_HUMAN	60S ribosomal protein L14 (CAG-ISL 7)	Homo sapiens	23144
Q07020 RL18_HUMAN	60S ribosomal protein L18	Homo sapiens	21490
P05388 RLA0_HUMAN	60S acidic ribosomal protein P0 (L10E)	Homo sapiens	34252
Q32969 RL9_HUMAN	60S ribosomal protein L9	Homo sapiens	21779
Septines			
Q16181 SEPT7_HUMAN	Septin-7 (CDC10 protein homolog)	Homo sapiens	50648
Q9UHD8 SEPT9_HUMAN	Septin-9 (MLL septin-like fusion protein) (MLL septin-like fusion protein MSF-A)	Homo sapiens	65361
Q9NVA2 SEP11_HUMAN	Septin-11	Homo sapiens	49236
Hélicases ARN ATP-dépendante			
Q9GZR7 DDX24_HUMAN	ATP-dependent RNA helicase DDX24 (EC 3.6.1.-) (DEAD box protein 24)	Homo sapiens	96271
Q8NHQ9 DDX55_HUMAN	ATP-dependent RNA helicase DDX55 (EC 3.6.1.-) (DEAD box protein 55)	Homo sapiens	68445
Autres			
P62258 1433E_HUMAN	14-3-3 protein epsilon (14-3-3E)	Homo sapiens	29155
P11021 GRP78_HUMAN	78 kDa glucose-regulated protein precursor (GRP 78) (Heat shock 70 kDa protein 5)	Homo sapiens	72288
P60709 ACTB_HUMAN ou P63261 ACTG_HUMAN	Actin, cytoplasmic 1 (Beta-actin) ou Actin, cytoplasmic 2 (Gamma-actin)	Homo sapiens	41710
P07355 ANXA2_HUMAN	Annexin A2 (Annexin II) (Lipocortin II) (Calpactin I heavy chain)	Homo sapiens	38449
Q8NE71 ABCF1_HUMAN	ATP-binding cassette sub-family F member 1 (ATP-binding cassette 50)	Homo sapiens	95866
P12956 KU70_HUMAN	ATP-dependent DNA helicase 2 subunit 1 (ATP-dependent DNA helicase II 70 kDa)	Homo sapiens	69668
Q96T60 PNKP_HUMAN	Bifunctional polynucleotide phosphatase/kinase (Polynucleotide kinase-3'-phosphate)	Homo sapiens	57040
Q9NRL2 BAZ1A_HUMAN	Bromodomain adjacent to zinc finger domain protein 1A (ATP-utilizing chromatin a)	Homo sapiens	178592
Q13111 CAF1A_HUMAN	Chromatin assembly factor 1 subunit A (CAF-1 subunit A) (Chromatin assembly fact)	Homo sapiens	105158
O43809 CPSF5_HUMAN	Cleavage and polyadenylation specificity factor 5 (Cleavage and polyadenylation)	Homo sapiens	26211

P11388 TOP2A_HUMAN	DNA topoisomerase 2-alpha (EC 5.99.1.3) (DNA topoisomerase II, alpha isozyme)	Homo sapiens	174276
Q02880 TOP2B_HUMAN	DNA topoisomerase 2-beta (EC 5.99.1.3) (DNA topoisomerase II, beta isozyme)	Homo sapiens	183152
P06746 DPOLB_HUMAN	DNA polymerase beta (EC 2.7.7.7) (EC 4.2.99.-)	Homo sapiens	38023
P18887 XRCC1_HUMAN	DNA-repair protein XRCC1 (X-ray repair cross-complementing protein 1)	Homo sapiens	69483
P49916 DNL3_HUMAN	DNA ligase 3 (EC 6.5.1.1) (DNA ligase III) (Polydeoxyribonucleotide synthase)	Homo sapiens	102625
Q9H9Y6 RPA2_HUMAN	DNA-directed RNA polymerase I 135 kDa polypeptide (EC 2.7.7.6) (RNA polymerase)	Homo sapiens	128146
Q95602 RPA1_HUMAN	DNA-directed RNA polymerase I largest subunit (EC 2.7.7.6) (RNA polymerase I 194)	Homo sapiens	194068
Q9GZS1 RPF53_HUMAN	DNA-directed RNA polymerase I-associated factor 53 kDa subunit (EC 2.7.7.6)	Homo sapiens	53928
P55265 DSRAD_HUMAN	Double-stranded RNA-specific adenosine deaminase (EC 3.5.4.-) (DRADA)	Homo sapiens	135910
P68104 EF1A1_HUMAN	Elongation factor 1-alpha 1 (EF-1-alpha-1) (Elongation factor 1 A-1) (eEF1A-1)	Homo sapiens	50109
P05198 IF2A_HUMAN	Eukaryotic translation initiation factor 2 subunit 1 (Eukaryotic translation init)	Homo sapiens	35958
Q13868 EXOS2_HUMAN	Exosome complex exonuclease RRP4 (EC 3.1.13.-) (Ribosomal RNA-processing)	Homo sapiens	32768
Q9NPD3 EXOS4_HUMAN	Exosome complex exonuclease RRP41 (EC 3.1.13.-) (Ribosomal RNA-processing)	Homo sapiens	26235
Q9NQT4 EXOS5_HUMAN	Exosome complex exonuclease RRP46 (EC 3.1.13.-) (Ribosomal RNA-processing)	Homo sapiens	25233
P21333 FLNA_HUMAN	Filamin-A (Alpha-filamin) (Filamin-1) (Endothelial actin-binding protein)	Homo sapiens	280433
O75369 FLNB_HUMAN	Filamin-B (FLN-B) (Beta-filamin) (Actin-binding-like protein) (Thyroid autoantige)	Homo sapiens	278021
P62826 RAN_HUMAN	GTP-binding nuclear protein Ran (GTPase Ran) (Ras-like protein TC4)	Homo sapiens	24277
P63244 GBLP_HUMAN	Guanine nucleotide-binding protein subunit beta 2-like 1 (Guanine nucleotide-bind)	Homo sapiens	34924
O60832 DKC1_HUMAN	H/ACA ribonucleoprotein complex subunit 4 (EC 5.4.99.-) (Dyskerin)	Homo sapiens	57507
P08107 HSP71_HUMAN	Heat shock 70 kDa protein 1 (HSP70.1) (HSP70-1/HSP70-2)	Homo sapiens	69995
P11142 HSP7C_HUMAN	Heat shock cognate 71 kDa protein (Heat shock 70 kDa protein 8)	Homo sapiens	70854
Q9H583 HEAT1_HUMAN	HEAT repeat-containing protein 1 (Protein BAP28)	Homo sapiens	242215
P52292 IMA2_HUMAN	Importin alpha-2 subunit (Karyopherin alpha-2 subunit) (SRP1-alpha)	Homo sapiens	57826
P42166 LAP2A_HUMAN	Lamina-associated polypeptide 2 isoform alpha (Thymopoietin isoform alpha)	Homo sapiens	75315

P05455 LA_HUMAN	Lupus La protein (Sjogren syndrome type B antigen) (SS-B) (La ribonucleoprotein)	Homo sapiens	46808
Q14676 MDC1_HUMAN	Mediator of DNA damage checkpoint protein 1 (Nuclear factor with BRCT domains 1)	Homo sapiens	226529
Q9NPJ6 MED4_HUMAN	Mediator complex subunit 4 (Mediator of RNA polymerase II transcription subunit 4)	Homo sapiens	29727
O00159 MYO1C_HUMAN	Myosin Ic (Myosin I beta) (MMI-beta) (MMIb)	Homo sapiens	117965
Q9H0A0 NAT10_HUMAN	N-acetyltransferase 10 (EC 2.3.1.-)	Homo sapiens	115631
O00567 NOP56_HUMAN	Nucleolar protein Nop56 (Nucleolar protein 5A)	Homo sapiens	66009
Q14978 NOLC1_HUMAN	Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa nucleolar pho)	Homo sapiens	73677
O75607 NPM3_HUMAN	Nucleoplasmin-3	Homo sapiens	19200
P06748 NPM_HUMAN	Nucleophosmin (NPM) (Nucleolar phosphoprotein B23) (Numatrin)	Homo sapiens	32555
P09874 PARP1_HUMAN	Parp 1	Homo sapiens	112881
Q15269 PWP2_HUMAN	Periodic tryptophan protein 2 homolog	Homo sapiens	102387
Q6P2Q9 PRP8_HUMAN	Pre-mRNA-processing-splicing factor 8 (Splicing factor Prp8) (PRP8 homolog) (220)	Homo sapiens	273427
Q9UMS4 PRP19_HUMAN	Pre-mRNA-processing factor 19 (EC 6.3.2.-) (PRP19/PSO4 homolog) (hPso4) (Nuclear)	Homo sapiens	55146
Q15648 PPRB_HUMAN	Peroxisome proliferator-activated receptor-binding protein (PBP) (PPAR-binding)	Homo sapiens	168333
Q92979 NEP1_HUMAN	Probable ribosome biogenesis protein NEP1 (Nucleolar protein EMG1 homolog)	Homo sapiens	26572
P46087 NOL1_HUMAN	Proliferating-cell nucleolar antigen p120	Homo sapiens	94020
Q9UKS6 PACN3_HUMAN	Protein kinase C and casein kinase substrate in neurons protein 3	Homo sapiens	48457
Q96P11 NSUN5_HUMAN	Putative methyltransferase NSUN5 (EC 2.1.1.-) (NOL1/NOP2/Sun domain family member)	Homo sapiens	46662
Q13283 G3BP_HUMAN	Ras-GTPase-activating protein-binding protein 1 (EC 3.6.1.-) (Homo sapiens	52132
Q99575 POP1_HUMAN	Ribonucleases P/MRP protein subunit POP1 (EC 3.1.26.5) (hPOP1)	Homo sapiens	114636
P78346 RPP30_HUMAN	Ribonuclease P protein subunit p30 (EC 3.1.26.5) (RNaseP protein p30)	Homo sapiens	29303
P35637 FUS_HUMAN	RNA-binding protein FUS (Oncogene FUS) (Oncogene TLS) (Translocated in liposarcoma)	Homo sapiens	53394
P22087 FBRL_HUMAN	rRNA 2'-O-methyltransferase fibrillarlin (EC 2.1.1.-) (34 kDa nucleolar scleroderm)	Homo sapiens	33763
Q01082 SPTB2_HUMAN	Spectrin beta chain, brain 1 (Spectrin, non-erythroid beta chain 1)	Homo sapiens	274439
Q13813 SPTA2_HUMAN	Spectrin alpha chain, brain (Spectrin, non-erythroid alpha chain)	Homo sapiens	284364

P57772 SELB_HUMAN	Selenocysteine-specific elongation factor (Elongation factor sec)	Homo sapiens	65297
P62136 PP1A_HUMAN ou P62140 PP1B_HUMAN	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit (EC 3.1.3.16) ou Serine/threonine-protein phosphatase PP1-beta catalytic subunit (EC 3.1.3.16)	Homo sapiens	37488
Q5UIP0 RIF1_HUMAN	Telomere-associated protein RIF1 (Rap1-interacting factor 1 homolog)	Homo sapiens	274294
P30048 PRDX3_HUMAN	Thioredoxin-dependent peroxide reductase, mitochondrial precursor (EC 1.11.1.15)	Homo sapiens	27675
Q13428 TCOF_HUMAN	Treacle protein (Treacher Collins syndrome protein)	Homo sapiens	144226
Q9NUW8 TYDP1_HUMAN	Tyrosyl-DNA phosphodiesterase 1 (EC 3.1.4.-) (Tyr-DNA phosphodiesterase 1)	Homo sapiens	68377
P08670 VIME_HUMAN	Vimentin	Homo sapiens	53488
Q9UNX4 WDR3_HUMAN	WD repeat protein 3	Homo sapiens	106032

Tableau XII : Résultats d'identification par spectrométrie de masse des protéines contenues spécifiquement dans les RNP formées sur les répétitions (CUG)₅₁.

Le numéro d'accension, le nom ainsi que la masse en Da des protéines identifiées sont indiqués. Les protéines sont regroupées selon leur fonction lorsque cela est possible. Les lignes surlignées en orange correspondent aux protéines dont l'identification est à confirmer du fait du faible nombre de peptide détecté. Les protéines surlignées en violet sont celles dont il est fait mention dans le texte. Les protéines surlignées en bleu ont été spécifiquement détectées dans toutes les expériences de purification de complexes que j'ai réalisées dans le cadre de cette étude et constituent les meilleurs candidats.

Numéro d'accension	Noms des protéines	Taxonomie	Masse	MS2		(CUG) ₅₁ -MS2	
				Pourcentage de recouvrement	Nombre de peptides	Pourcentage de recouvrement	Nombre de peptides
Histones							
Q92522 H1X_HUMAN	Histone H1x - Homo sapiens (Human)	Homo sapiens	22474	10%	2	17%	3
Protéines ribosomales							
Q02878 RL6_HUMAN	60S ribosomal protein L6 (TAX-responsive enhancer element-binding protein 107)	Homo sapiens	32577	8%	2	32%	9
P18124 RL7_HUMAN	60S ribosomal protein L7	Homo sapiens	29207	4%	1	21%	5
P62424 RL7A_HUMAN	60S ribosomal protein L7a (Surfeit locus protein 3) (PLA-X polypeptide)	Homo sapiens	29846	9%	2	30%	11
O76021 RL1D1_HUMAN	Ribosomal L1 domain-containing protein 1	Homo sapiens	54939	8%	5	28%	16
Septines							
Q15019 SEPT2_HUMAN	Septin-2 (Protein NEDD5) - Homo sapiens	Homo sapiens	41461	6%	2	11 %/ 26 %	3 / 6

Hélicases ARN ATP-dépendante							
O00571 DDX3X_HUMAN	ATP-dependent RNA helicase DDX3X (EC 3.6.1.-) (DEAD box protein 3, X-chromosom)	Homo sapiens	73067	19%	10	35%	12
P17844 DDX5_HUMAN	Probable ATP-dependent RNA helicase DDX5 (EC 3.6.1.-) (DEAD box protein 5)	Homo sapiens	69105	14 % / 6 %	10 / 5	11 %/ 21 %	7 / 12
Q9NR30 DDX21_HUMAN	Nucleolar RNA helicase 2 (EC 3.6.1.-) (Nucleolar RNA helicase II)	Homo sapiens	87290	28%	19	32 %/ 50 %	22 / 49
Q08211 DHX9_HUMAN	ATP-dependent RNA helicase A (EC 3.6.1.-) (Nuclear DNA helicase II) (NDH II)	Homo sapiens	140869	16 % / 13 %	19 / 17	41%	58
Autres							
P63261 ACTG_HUMAN	Actin, cytoplasmic 2 (Gamma-actin)	Homo sapiens	41766	67%	35	82 %/ 40 %	42 / 12
O43707 ACTN4_HUMAN	Alpha-actinin-4 (Non-muscle alpha-actinin 4) (F-actin cross-linking protein)	Homo sapiens	104788	12%	9	32%	27
P11387 TOP1_HUMAN	DNA topoisomerase 1 (EC 5.99.1.2) (DNA topoisomerase I)	Homo sapiens	90669	11%	9	24 %/ 15 %	27 / 15
P19525 E2AK2_HUMAN	Interferon-induced, double-stranded RNA-activated protein kinase (EC 2.7.11.1)	Homo sapiens	62056	6%	3	20%	12
Q9BQG0 MBB1A_HUMAN	Myb-binding protein 1A	Homo sapiens	148762	6%	6	25%	32
Q86V81 THOC4_HUMAN	THO complex subunit 4 (Tho4) (Ally of AML-1 and LEF-1)	Homo sapiens	26741	9%	3	29%	5
P68363 TBAK_HUMAN ou Q9BQE3 TBA6_HUMAN	Tubulin alpha-ubiquitous chain ou Tubulin alpha-6 chain	Homo sapiens	50120	15%	5	30%	13

Tableau XIII : Protéines identifiées majoritairement présentes dans les complexes formés sur les répétitions (CUG)₅₁.

Le numéro d'accèsion, le nom, la masse en Da, le pourcentage de recouvrement ainsi que le nombre de peptides identifiés pour une protéine donnée au sein des échantillons contrôle (MS2) et CUG51-MS2 sont indiqués. Les protéines surlignées en violet sont celles dont il est fait mention dans le texte. Les protéines surlignées en bleu ont été détectées dans toutes les expériences de purification de complexes que j'ai réalisées dans le cadre de cette étude et constituent les meilleurs candidats.

I.7. Amélioration de la technique de purification de complexes

La purification des complexes RNP formés sur l'ARN (CUG)₅₁ était initialement réalisée en conditions faiblement stringentes (150 mM KCl) et ceci pouvait être à l'origine de la rétention aspécifique de nombreuses protéines. La rétention de ces protéines pouvait résulter d'interactions protéine/protéine ou d'interactions non-spécifiques ARN/protéine. Afin

d'améliorer les conditions de spécificité d'interaction, nous avons rajouté une étape de pontage covalent aux UV à 254 nm après incubation de l'ARN avec l'extrait nucléaire et la protéine MBP-MS2. L'ajout de cette étape couplé à l'augmentation de la stringence par augmentation de la quantité de sel (300 mM KCl) et à l'ajout d'un détergent non-ionique le NP40, devait permettre de limiter le nombre de protéines présentes dans les RNP purifiées.

Par ailleurs, comme décrit dans l'Introduction de ce manuscrit, il avait été observé que certaines protéines, telles que MBNL1, pouvaient se fixer sur l'ensemble des répétitions de type CNG, alors que d'autres, telles que Sam68 ou Puro α étaient spécifiques, d'un type donné de répétitions. Afin de pouvoir distinguer les protéines reconnaissant spécifiquement les répétitions CUG des protéines reconnaissant différents types de répétitions, nous avons réalisé une construction permettant d'exprimer un ARN contenant 59 répétitions CAG fusionnées à l'ARN MS2. Les répétitions CAG forment, tout comme les répétitions CUG, des hélices de type A et la surexpression de telles répétitions entraîne la formation de *foci* nucléaires contenant MBNL1, ainsi que des dérégulations de l'épissage alternatif de pré-ARNm connus comme dérégulé chez les patients DM (Kiliszek *et al.*, 2010 ; Ho *et al.*, 2005; Mykowska *et al.*, 2011). L'analyse des complexes protéiques formés sur l'ARN CAG₅₉-MS2 avait donc le double-avantage d'identifier les protéines reconnaissant les structures en tige-boucle formées par les répétitions CUG ou CAG, mais également de différencier les protéines reconnaissant spécifiquement les répétitions CUG.

J'ai donc réalisé deux expériences de purification de complexes RNP associés aux répétitions CUG et CAG, l'une selon le protocole « classique » et une seconde selon le protocole « stringent ». Les profils de fractionnement des protéines obtenus après élution des complexes formés sont présentés dans la Figure 82

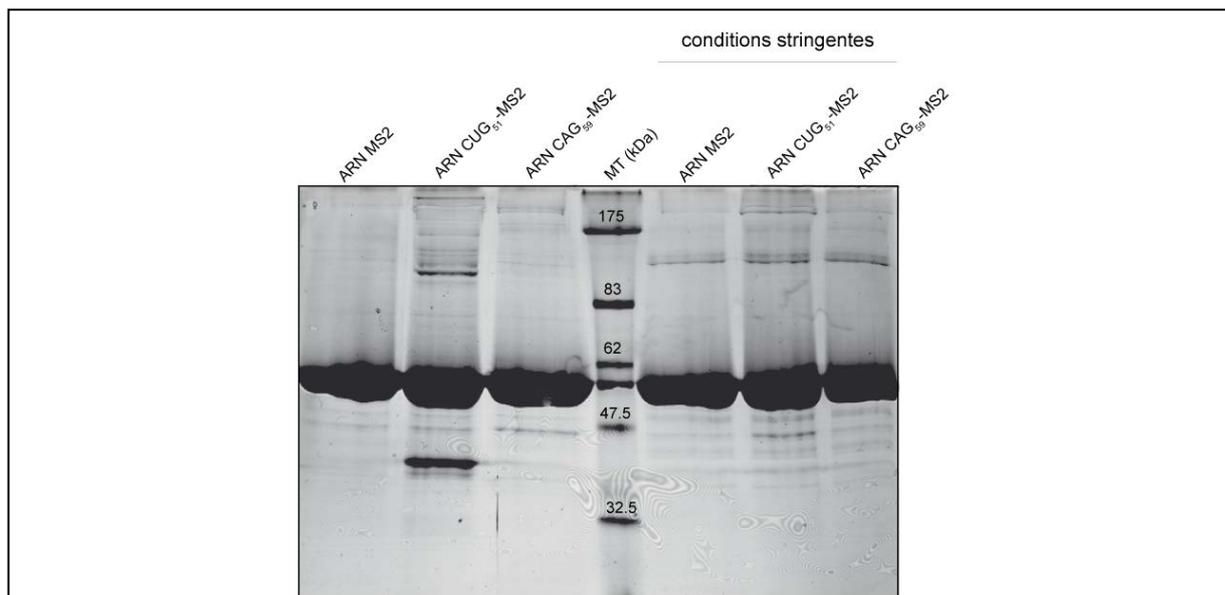


Figure 82 : Analyse par électrophorèse en gel SDS-PAGE 10 % des protéines contenues dans les RNP formées sur les ARN (CUG)₅₁-MS2, (CAG)₅₉-MS2 et MS2, purifiées par chromatographie d'affinité MS2/MBP réalisée en conditions « normales » ou « stringentes »

Les RNP formées sur les ARN (CUG)₅₁-MS2, (CAG)₅₉-MS2 et MS2 ont été purifiées par l'approche MS2/MBP selon le protocole « classique » ou « stringent ». Après lavages, les RNP ont été éluées par ajout de maltose 10 mM et 50 % des produits élués ont été fractionnés en gel SDS-PAGE 10 %. Le gel a ensuite été coloré par une solution de bleu de Coomassie colloïdal. Un marqueur de poids moléculaire a été déposé à gauche du gel (MT).

Etonnamment, une différence de profil est observée entre les deux expériences de purification de complexe RNP selon le protocole classique pour l'ARN (CUG)₅₁-MS2 et l'ARN control MS2. Ce résultat pose la question de la reproductibilité de cette expérience. Une différence de profil est nettement visible entre la purification classique et la purification « stringente », de même qu'entre les profils de purification des complexes RNP CUG et CAG. Afin de limiter le nombre d'échantillons à analyser par spectrométrie de masse, nous avons sélectionné les 11 bandes les plus intenses, en dehors de la bande correspondant à MBP-MS2, pour chaque profil issu de la purification de complexes réalisée en conditions stringentes. En revanche, nous avons procédé à un découpage systématique des profils issus de la purification « classique », afin de pouvoir comparer les résultats obtenus avec ceux de l'expérience précédente. L'analyse par spectrométrie de masse a aussi été réalisée par S. Sanglier et L. Fouillen de l'équipe d'A. Van Dorsselaar comme décrit précédemment. Malheureusement, un problème de quantité de matériel est apparu lors de l'analyse par spectrométrie de masse pour l'ensemble des expériences, et en particulier pour les bandes issues du profil associé au pontage covalent. Par conséquent, seules quelques protéines ont pu être détectées (Tableaux XIV et XV).

Numéro d'accèsion	Nom des Protéines	Taxonomie	Masse	MS2	(CUG) ₅₁	(CAG) ₅₉
P60709 ACTB_HUMAN	Actin, cytoplasmic 1 (Beta-actin)	Homo sapiens	41710	X	X	X
P69171 COAT_BPZR	Coat protein - Bacteriophage ZR	Enterobacteria phage ZR	13719	X	X	X
SRP14_HUMAN	Signal recognition particle 14 kDa protein (SRP14) (18 kDa Alu RNA-binding protein)	Homo sapiens		X	X	
Q9NR30 DDX21_HUMAN	Nucleolar RNA helicase 2 (EC 3.6.1.-)	Homo sapiens	87290	X	X	
Q08211 DHX9_HUMAN	ATP-dependent RNA helicase A (EC 3.6.1.-) (Nuclear DNA helicase II) (NDH II) (DEAH box protein 9)	Homo sapiens	140869		X	X
Q15019 SEPT2_HUMAN	Septin-2 (Protein NEDD5)	Homo sapiens	41461		X	X
P06748 NPM_HUMAN	Nucleophosmin (NPM) (Nucleolar phosphoprotein B23) (Numatrin) (Nucleolar protein NO38)	Homo sapiens	32555		X	
O75607 NPM3_HUMAN	Nucleoplasmin-3	Homo sapiens	19200		X	
P09874 PARP1_HUMAN	Parp 1	Homo sapiens	112881		X	
O76021 RL1D1_HUMAN	Ribosomal L1 domain-containing protein 1	Homo sapiens	54939		X	
P11387 TOP1_HUMAN	DNA topoisomerase 1 (EC 5.99.1.2) (DNA topoisomerase I)	Homo sapiens	90669		X	
P11388 TOP2A_HUMAN	DNA topoisomerase 2-alpha (EC 5.99.1.3) (DNA topoisomerase II, alpha isozyme)	Homo sapiens	174276		X	

Tableau XIV : Protéines contenues dans les RNP formées sur les ARN (CUG)₅₁-MS2, (CAG)₅₉-MS2 ou MS2.

Le numéro d'accèsion, le nom ainsi que la masse en Da des protéines identifiées sont indiqués. La présence d'une protéine donnée dans les complexes purifiés à partir des différents ARN est indiquée par une croix. Les protéines surlignées en violet sont celles dont il est fait mention dans le texte. Les protéines surlignées en bleu ont été détectées dans toutes les expériences de purification de complexes que j'ai réalisées dans le cadre de cette étude et constituent les meilleurs candidats.

Numéro d'accension	Nom des protéines	Taxonomie	Masse	MS2	(CUG) ₅₁	(CAG) ₅₉
P60709 ACTB_HUMAN ou P63261 ACTG_HUMAN	Actin	Homo sapiens	41710	X	X	X
A004003 MYH10_HUMAN	Myosin-10 (Myosin heavy chain 10)	Homo sapiens	228798			X
A004003 MYH9_HUMAN	Myosin-9 (Myosin heavy chain 9)	Homo sapiens	226392		X	
Q9NR30 DDX21_HUMAN	Nucleolar RNA helicase 2	Homo sapiens	87290		X	
P06748 NPM_HUMAN	Nucleophosmin (NPM)	Homo sapiens	32555		X	
Q08211 DHX9_HUMAN	ATP-dependent RNA helicase A	Homo sapiens	140869		X	X

Tableau XV : Protéines contenues dans les RNP formées en conditions stringentes sur les ARN (CUG)₅₁-MS2, (CAG)₅₉-MS2 ou MS2.

Le numéro d'accension, le nom ainsi que la masse en Da des protéines identifiées sont indiqués. La présence d'une protéine donnée dans les complexes purifiés à partir des différents ARN est indiquée par une croix. Les protéines surlignées en violet sont celles dont il est fait mention dans le texte. Les protéines surlignées en bleu ont été détectées dans toutes les expériences de purification de complexes que j'ai réalisées dans le cadre de cette étude et constituent les meilleurs candidats.

Parmi les 12 protéines identifiées lors de la purification de complexes réalisée selon le protocole « classique » : 2 protéines (Actin et Coat protein) étaient présentes dans les complexes formés sur les 3 ARN et correspondent soit à des contaminations soit à des protéines aspécifiques ; 2 protéines (SRP14 et DDX21) étaient présentes dans les complexes formés sur l'ARN contrôle MS2 et sur l'ARN (CUG)₅₁-MS2 ; 2 protéines (DHX9 et SEPT2) étaient uniquement retrouvées dans les complexes formés sur les ARN (CUG)₅₁-MS2 et CAG₅₉-MS2 et enfin 6 protéines (NPM, NPM3, PARP1, RL1D1, TOP1 et TOP2A) étaient spécifiquement associées à l'ARN (CUG)₅₁-MS2 (Tableau XIV). Les 8 protéines co-purifiées avec l'ARN (CUG)₅₁-MS2 avaient déjà été détectées lors de la 1^{ère} purification de complexes, mais dans cette 1^{ère} expérience seules 4 d'entre elles (NPM, NPM3, PARP1 et TOP2A) étaient spécifiques des répétitions CUG.

En ce qui concerne la purification de complexes réalisée selon le protocole « stringent », 6 protéines ont été détectées : 1 protéine (Actin) était probablement aspécifique ; 1 protéine (MYH10) était seulement trouvée avec l'ARN (CAG)₅₉-MS2 ; 1 protéine (DHX9) était seulement détectée dans les complexes formés sur les ARN CUG₅₁-MS2 et CAG₅₉-MS2 et enfin 3 protéines (MYH9, DDX21 et NPM) étaient seulement associées à l'ARN (CUG)₅₁-MS2 (Tableau XV). Deux de ces protéines (MYH10 et MYH9) n'avaient encore jamais été détectées. Ces protéines sont impliquées dans la contraction musculaire.

Au cours de ces différentes études une seule protéine, la nucléophosmine ou NPM, a été retrouvée uniquement associée à l'ARN (CUG)₅₁-MS2 et ceci quelque soit le protocole expérimental utilisé. Par ailleurs, les données de spectrométrie de masse mettaient en évidence un très grand nombre de peptides NPM avec un fort pourcentage de recouvrement de la séquence protéique au niveau de la bande de plus forte intensité retrouvée au sein du profil électrophorétique aux environs de 35 kDa (Figure 81).

Nous avons donc décidé de focaliser notre attention sur cette protéine et les expériences réalisées dans le but de caractériser l'interaction de NPM avec les répétitions CUG font l'objet du paragraphe suivant. D'autres protéines, telles que les hélicases DHX9 et DDX21, ainsi que les protéines capables de se lier à l'ADN tel que les histones et les protéines du facteur de réplication C ont cependant retenu notre attention et leur implication potentielle dans la DM1 sera discutée dans la partie conclusion-perspectives de ce chapitre.

II. Identification d'une nouvelle protéine capable de s'associer à un ARN contenant des répétitions CUG : la nucléophosmine.

Sur la base de l'observation reproductible d'une association entre l'ARN (CUG)₅₁-MS2 et la nucléophosmine, nous avons entrepris une analyse bibliographique sur cette protéine et nous résumerons ici l'état des connaissances sur cette protéine.

II.1. Présentation de la nucléophosmine

La protéine nucléophosmine NPM, aussi appelée NPM1, B23, NO38 chez les amphibiens ou numatrine chez les animaux appartient à la famille des protéines chaperonne de type nucléophosmine/nucléoplasmine (NPM) (Kang *et al.*, 1974 ; Prestaylo *et al.*, 1974 ; Feuerstein et Mond, 1987 ; Schmidt-Zachmann *et al.*, 1987). Cette famille comporte également la nucléoplasmine (NPM2) et la nucléoplasmine 3 (NPM3). La NPM est une phosphoprotéine très abondante et exprimée de manière ubiquitaire, tout comme NPM3, alors que NPM2 est exprimée uniquement dans les oocytes et les œufs (pour revue, Okuwaki, 2008).

Le gène codant la protéine NPM1 est localisé au niveau de la position 5q35. Il comporte 12 exons chez l'Homme et le rat, mais seulement 11 chez la souris. Le gène humain NPM1 peut produire par épissage alternatif trois isoformes :

- ✓ NPM1.1 : isoforme majoritaire et exprimée dans l'ensemble des tissus. Il s'agit de l'isoforme la plus longue, elle fait 294 acides aminés et a un poids moléculaire de 38

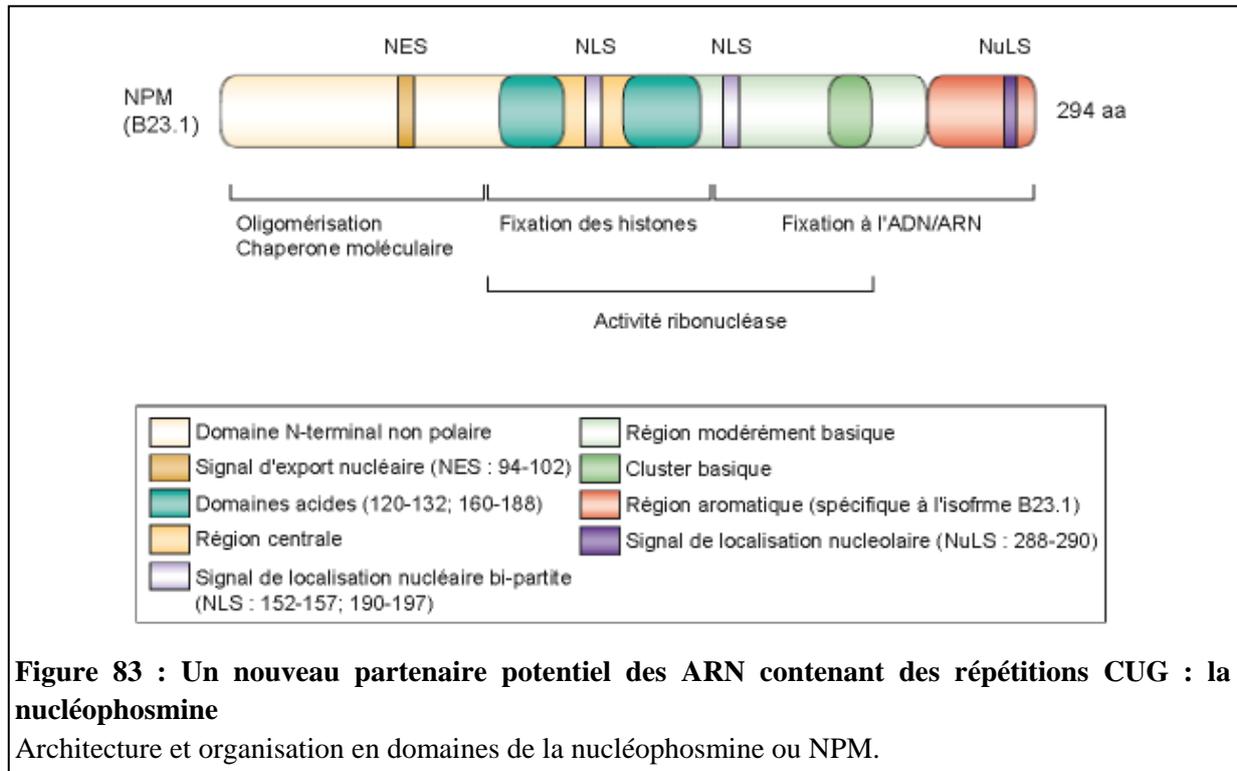
kDa. La protéine résultant de cette isoforme est localisée principalement au niveau du nucléole, mais a également la capacité de naviguer entre le noyau et le cytoplasme (Fig.83)

- ✓ NPM1.2 : isoforme tronquée des 35 derniers acides aminés C-terminaux suite à l'utilisation d'un site 3'-accepteur alternatif. Cette isoforme est faiblement exprimée et est localisée aussi bien au niveau du noyau que du cytoplasme
- ✓ NPM1.3 : isoforme tronquée de la région centrale par exclusion de l'exon 8. Très peu d'informations sont disponibles aujourd'hui sur cette isoforme (Lim et Wang, 2006).

La protéine NPM1 est caractérisée par un domaine d'oligomérisation N-terminal grâce auquel elle peut former un homohexamère, elle est présente dans les cellules sous forme monomérique ou hexamérique. De plus, toujours par son domaine N-terminal hydrophobe, NPM1 exerce également une activité de chaperonne, favorisant l'assemblage des histones pour former les nucléosomes, et augmentant l'activité de la machinerie transcriptionnelle acétylation-dépendante. La région centrale de NPM1 contient un domaine de fixation aux histones, et un domaine à activité ribonucléase chevauche la région centrale et la région C-terminale. Cette région C-terminale contient également un domaine de fixation aux acides nucléiques. Les capacités de transfert de NPM1 du cytoplasme vers le noyau et vice versa dépendent respectivement de l'élément NLS bi-partite situé dans la région centrale et à son signal d'export nucléaire (NES), riche en leucine, localisé dans la région N-terminale. La localisation de l'isoforme 1 de NPM1 dans le nucléole est due à un signal de localisation nucleolaire (NuLS) présent dans les 35 derniers acides aminés et spécifique de cette isoforme (Fig. 59)(pour revue, Okuwaki, 2008).

La protéine NPM1 joue un rôle dans un très grand nombre de processus (pour revue, Lindström, 2010). Elle a tout d'abord été identifiée pour son rôle dans la biogénèse et le transport des particules pré-ribosomales. De plus, NPM1 est également impliquée dans le transport nucléocytoplasmique, la duplication du centrosome, la régulation de la transcription via la modulation des événements de condensation et de décondensation de la chromatine, dans l'interaction et l'altération de la structure secondaire des acides nucléiques et dans la réponse à certains stress (irradiation aux UV, hypoxie) (pour revue, Frehlick *et al.*,2006). NPM1 est principalement connue pour son rôle dans la régulation de l'activité et de la stabilité de suppresseurs de tumeur cruciaux tels que p53 et ARF (pour revue, Grisendi *et al.* 2006). Plus récemment, il a été proposé que NPM soit un facteur d'épissage, en effet un rôle de NPM dans la régulation de l'épissage alternatif du pré-ARNm de la protéine ribosomale L3

a été observé (Russo *et al.*, 2011). Dans ce système, NPM réprimerait l'épissage alternatif en contrecarrant, probablement par encombrement stérique, l'effet activateur des protéines KSRP et hnRNP H1 sur l'utilisation d'un site 3' cryptique qui permet à la protéine rpL3 d'autoréguler négativement son expression via le NMD (Russo *et al.*, 2011).



Comme mentionné précédemment NPM1 est une phosphoprotéine qui contient de multiples sites de phosphorylation impliqués dans la régulation de ses différentes fonctions. Différentes kinases interviennent dans la phosphorylation de NPM1, telles que la caséine kinase II (CKII), la kinase nucléaire NII, la polo-like kinase (Plk), la kinase de type cdc2 et les kinases cycline-dépendantes (CDK1/cyclin B, CDK2/cyclin E, et CDK2/cyclin A) (Peter *et al.*, 1990 ; Chan *et al.*, 1990 ; Jones *et al.*, 1981 ; Feuerstein et Randazzo, 1991). Parmi les nombreux sites de phosphorylation, la thréonine T199 joue un rôle crucial dans plusieurs processus. Sa phosphorylation a tout d'abord été montrée intervenir dans la duplication du centrosome. En effet, la phosphorylation du résidu T199 affecte l'affinité de NPM pour les centrosomes. Après phosphorylation, la protéine se détache de ces derniers, ce qui constitue la première étape requise pour la duplication des centrosomes (Tokuyama *et al.*, 2001). De manière intéressante, la forme phosphorylée NPM-T199 est localisée non pas dans le nucléole mais au niveau des speckles nucléaires, zone de « stockage » et/ou d'assemblage des protéines régulant l'épissage des pré-ARNm (Tarapore *et al.*, 2006). Or, les *foci* nucléaires contenant les répétitions CUG sont localisés à la périphérie des speckles (Smith *et al.*, 2007 ; Holt *et al.*,

2007). De plus, la forme NPM-T199 peut jouer un rôle dans l'épissage. En effet, il a été montré que cette forme, et uniquement cette forme, peut réprimer *in vitro* l'épissage du pré-ARNm issu d'un minigène codant pour la α -globine (Tarapore *et al.*, 2006).

De cet ensemble de données de la littérature, nous retiendrons les propriétés suivantes au travers desquelles nous pourrions imaginer un lien entre NPM et DM1 :

- elle interagit avec l'ADN et l'ARN (pour revue, Grisendi *et al.* 2006),
- elle est impliquée *in vivo* dans la régulation de l'épissage alternatif du pré-ARNm de la protéine ribosomale L3 (Russo *et al.*, 2011),
- sous sa forme phosphorylée au niveau du résidu Thr199 elle est localisée dans les speckles, à la périphérie desquels sont localisés les *foci* DM1 (Tarapore *et al.* 2006; Smith *et al.* 2007 ; Holt *et al.*, 2007),
- toujours sous sa forme phosphorylée au niveau du résidu Thr199, elle a été décrite comme ayant une forte affinité pour les ARN et comme pouvant réguler *in vitro* l'épissage d'un ARN pré-messenger reporter β -globine (Tarapore *et al.*, 2006).

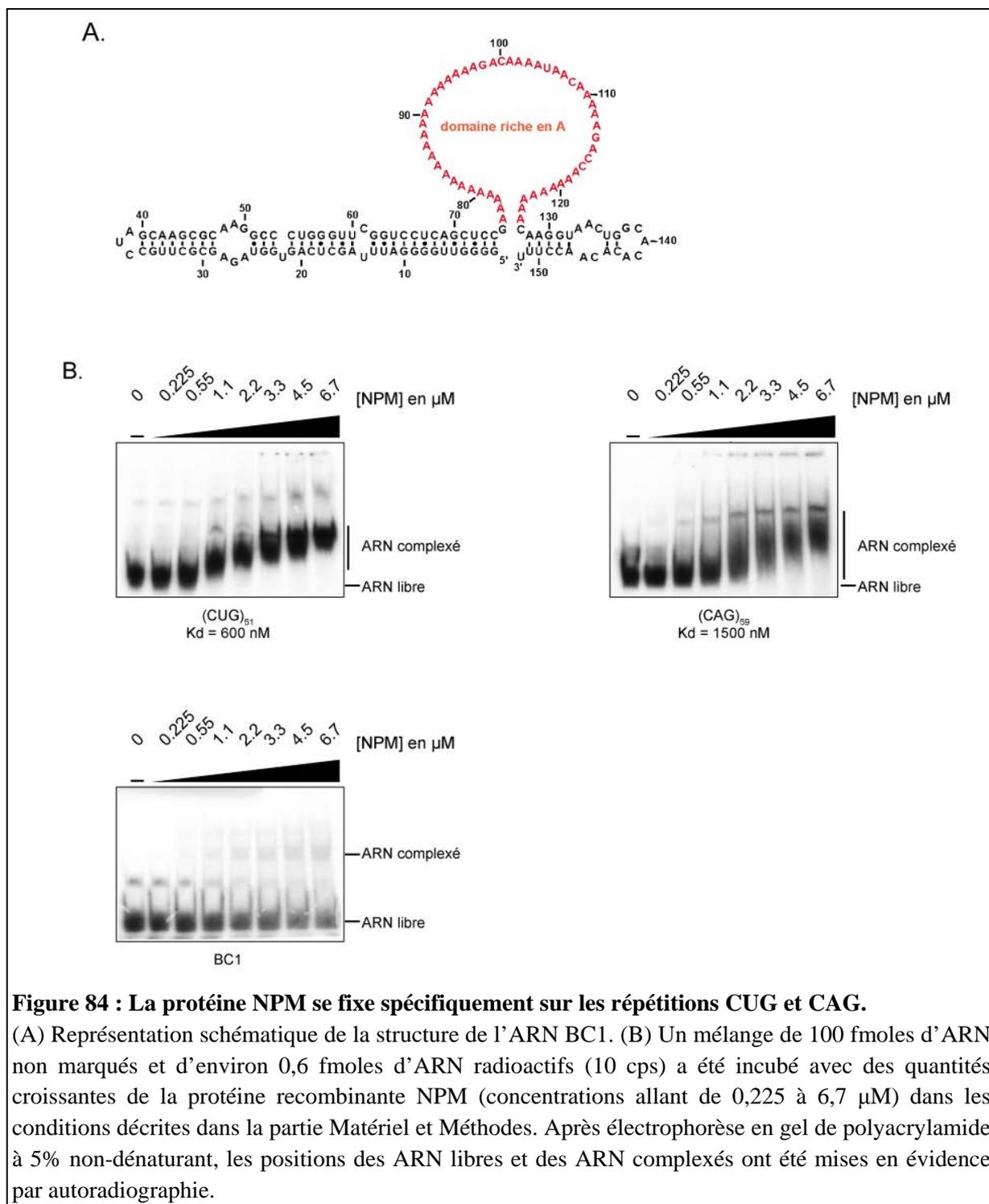
La détection de NPM lors de la purification des complexes formés sur des ARN contenant des répétitions CUG réalisée par l'équipe de J. Marie (Gif-sur-Yvette) en utilisant une autre approche nous a également conforté dans notre choix (com. pers.).

II.2. La nucléophosmine interagit directement avec l'ARN (CUG)₅₁

La présence de NPM dans les complexes formés *in vitro* avec l'ARN (CUG)₅₁ ne signifiait pas forcément que cette protéine soit liée directement à cet ARN. Elle pouvait être associée aux complexes par des interactions protéine-protéine. Afin de tester si une interaction directe était possible, I. Behm-Ansmant a amplifié et cloné la séquence codant l'isoforme NPM1.1 humaine dans le plasmide d'expression bactérien pGEX-6P1 décrit précédemment. La souche d'*E. coli* BL21 codon+ a été choisie comme souche hôte. I. Behm-Ansmant a ensuite produit et purifié la protéine NPM recombinante.

J'ai utilisé cette protéine pour réaliser des expériences de retard sur gel. Les ARN (CUG)₅₁ et (CAG)₅₉ ont été produits par transcription *in vitro*, puis déphosphorylés et marqués radioactivement en 5' en présence d'ATP-[γ ³²P]. L'ARN BC1 qui a une taille similaire à celle des ARN (CUG)₅₁ et (CAG)₅₉, mais qui ne contient pas de répétitions de type CNG a servi de témoin négatif (Fig. 84). Le test de fixation de NPM sur chacun de ces ARN a été réalisé avec un mélange de 100 fmoles d'ARN non marqué et d'environ 0,6 fmoles d'ARN radioactifs (10 cps). Les mélanges d'ARN froid et marqué ont été incubés avec des quantités croissantes de protéine recombinante NPM (concentrations allant de 0,225 à 6,7 μ M) dans les conditions

décrites dans la partie Matériel et Méthodes. Après électrophorèse en gel de polyacrylamide à 5% non-dénaturant, la position des ARN libres et complexés a été mise en évidence par autoradiographie.



Comme illustré dans la Figure 84, la protéine NPM présente une bonne affinité pour les répétitions $(\text{CUG})_{51}$ avec un K_d apparent d'environ 600 nM. Cette affinité est du même

ordre de grandeur que celle de MBNL1 pour les répétitions CUG ($K_d \sim 500$ nM). La protéine NPM est également capable d'interagir avec des répétitions $(CAG)_{59}$ mais avec une plus faible affinité (K_d de l'ordre de 1500 nM). Pour les deux ARN $(CUG)_{51}$ et $(CAG)_{59}$, les complexes formés étaient de plus en plus retardés lors de l'augmentation de la quantité de NPM, ce qui suggère, comme pour MBNL1, la fixation de plusieurs molécules de protéines sur une même molécule d'ARN.

Ainsi, NPM interagit directement avec les répétitions CUG, mais également avec une moindre affinité avec les répétitions CAG.

II.3. La nucléophosmine pourrait entrer en compétition avec la fixation de MBNL1 sur les répétitions $(CUG)_{51}$

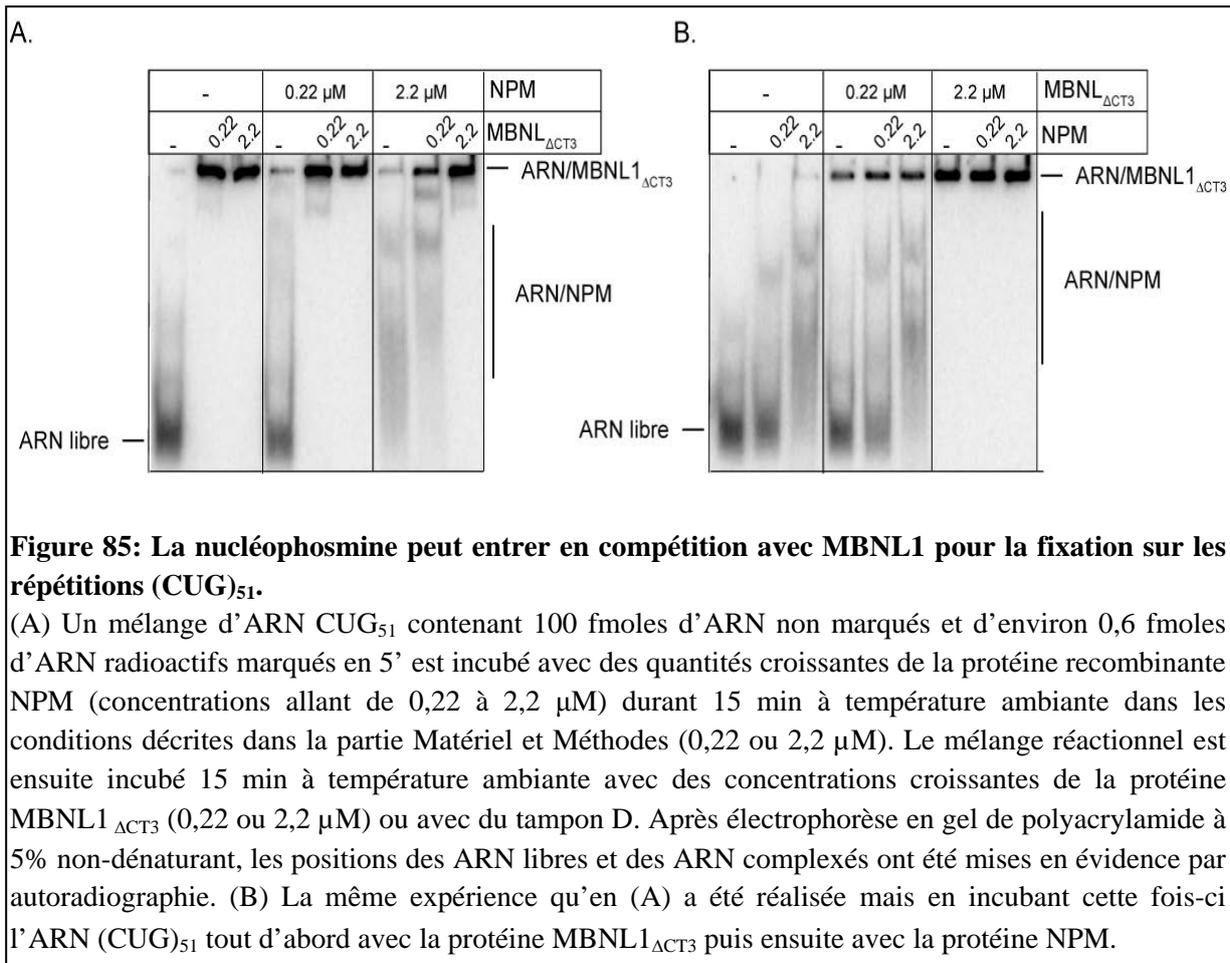
Une question importante était de savoir si la fixation de NPM1 et MBNL1 sur un ARN contenant des répétitions CUG pouvait se faire simultanément ou si elles entraient en compétition pour se fixer à un tel ARN. Afin de tester l'existence d'une compétition, j'ai réalisé des expériences de retard sur gel en présence de quantités variables de chacune des deux protéines. La protéine NPM1 et la protéine tronquée $MBNL1_{\Delta CT3}$ ont été utilisées pour ces expériences.

Ces expériences ont été réalisées en incubant un mélange d'ARN contenant 100 fmoles d'ARN CUG_{51} non marqués et environ 0,6 fmoles d'ARN marqué en 5', avec différentes concentrations de protéine $MBNL1_{\Delta CT3}$ ou NPM1 (0,22 ou 2,2 μ M) durant 15 min à température ambiante. Le mélange réactionnel a ensuite été incubé 15 min à température ambiante avec des concentrations croissantes de la seconde protéine (0,22 ou 2,2 μ M), ou avec du tampon D. Après électrophorèse en gel de polyacrylamide à 5% non-dénaturant, la position des ARN libres et des ARN complexés a été mise en évidence par autoradiographie (Fig.85). Les deux concentrations en protéines choisies (0,22 et 2,2 μ M) permettaient de se placer à une concentration de protéine où peu de complexes [ARN/protéine] sont formés (0,22 μ M) et où la totalité de l'ARN était supposé être complexé (2,2 μ M).

Malgré la présence d'agrégats dans les poches en présence de la protéine $MBNL1_{\Delta CT3}$, les données obtenues montrent qu'à une faible concentration en protéine MBNL1, la fixation à la fois des protéines NPM1 et $MBNL1_{\Delta CT3}$ est détectée et que NPM1 pourrait entrer en compétition avec MBNL1 pour la fixation sur des ARN $(CUG)_{51}$. Par contre à de fortes concentrations en MBNL1, MBNL1 déplacerait la nucléophosmine de l'ARN, quelle que soit la concentration initiale en nucléophosmine. Cependant à fortes concentrations en MBNL1

tout l'ARN est dans les poches et nous ne pouvons pas être sûrs de l'absence de NPM dans les complexes formés. Afin de répondre à cette question la même expérience devra être réalisée mais sur des ARN (CUG)₅₁ uniformément marqués au niveau des résidus G et en ajoutant une étape de crosslink aux UV.

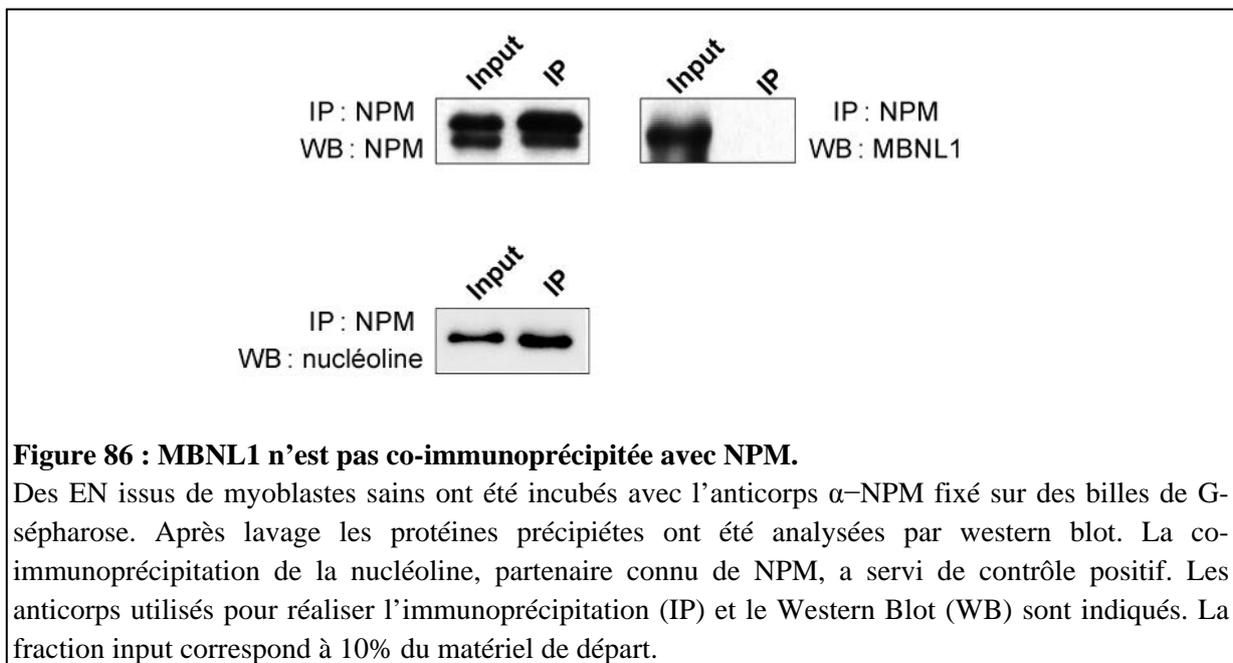
Nous en concluons qu'à faible concentration en protéine MBNL1, ce qui est le cas dans les noyaux, NPM1, qui elle est très abondante, pourrait empêcher MBNL1 de se fixer aux répétitions CUG.



II.4. La protéine NPM n'interagit pas avec la protéine MBNL1

Sur la base des résultats ci-dessus, dans les cellules de patients atteints de DM1, le recrutement de NPM1 au sein des *foci* nucléaires, pourrait résulter d'une part de l'interaction de NPM1 avec les ARNm *DMPK* contenant les répétitions CUG et, d'autre part, de l'interaction de NPM1 avec MBNL1 ou toute autre protéine ayant une forte affinité pour les répétitions CUG. Afin de vérifier la possibilité d'une interaction de type protéine-protéine entre NPM1 et MBNL1, j'ai réalisé des expériences de co-immunoprécipitation (Fig.86).

Dans le but de détecter une association possible de NPM1 et MBLN1 dans des cellules saines, I. Behm-Ansmant a préparé un extrait à partir de myoblastes sains, ceci afin d'immunoselectionner par un anticorps spécifique de NPM les complexes présents dans l'extrait et contenant NPM. L'extrait a été incubé avec l'anticorps fixé sur des billes de G-sépharose, selon le protocole décrit dans la partie Matériels et Méthodes. Cet anticorps monoclonal reconnaît les 3 isoformes de NPM. Dans ces conditions, la protéine NPM et ses partenaires protéiques et/ou ARN devaient être retenus sur les billes de G-sépharose. L'efficacité de l'immunoprécipitation a été vérifiée par western blot en utilisant l'anticorps dirigé contre NPM (Fig.86). Afin de tester l'efficacité de co-immunoprécipitation, nous avons utilisé un anticorps polyclonal dirigé contre la nucléoline, car cette protéine nucléolaire est connue pour interagir avec NPM1 (Li *et al.*, 1996) (Fig.86). Une réponse positive pour la nucléoline a bien été obtenue dans l'expérience de western-blot, par contre aucun signal avec l'anticorps anti-MBNL1 n'a été détecté, alors que MBNL1 était bien détectée dans l'extrait non traité avec cet anticorps (Fig.86). Ce résultat suggère une absence d'association entre MBNL1 et NPM dans le nucléoplasme de cellules saines. Il faut tout de même noter que l'anticorps anti-MBNL1 utilisé semble moins efficace que l'anticorps anti-NPM et aussi que la concentration de MBNL1 dans les cellules est plus faible que celle de NPM, de ce fait nous n'avons pas réussi à immunoprécipiter les complexes formés par MBNL1 et à tester s'ils contenaient NPM1.



II.5. La quantité de NPM libre dans le noyau diminue dans les cellules DM1

Si NPM est capable de se lier aux répétitions CUG, on pouvait s'attendre à une diminution de son taux dans les cellules de patients atteints de DM1 et/ou à des variations de son taux de phosphorylation. Pour tester cette hypothèse, nous avons analysé le niveau d'expression et de phosphorylation de NPM dans des lignées immortalisées de myoblastes de patients sains ou atteints de DM1 qui nous avaient été fournies par l'équipe de D. Furling (Institut de Myologie, Paris). Partant du constat que lors de la préparation d'un lysat total de cellules de patients atteints de DM1 dans les conditions expérimentales que nous utilisons, les *foci* nucléaires sont éliminés. Si NPM était piégée dans les *foci*, nous nous attendions à observer une diminution de son taux dans le lysat. C'est bien ce que nous avons observé pour MBNL1 en comparant les lysats de myoblastes normaux et ceux issus de patients atteints de DM1. En contrôle, nous n'observons pas de variations du taux de la protéine PTB, absente des *foci* (Fig.87). Deux anticorps différents ont été utilisés pour détecter la nucléophosmine : 1) l'anticorps anti-P-NPM qui reconnaît spécifiquement la forme phosphorylée au niveau du résidu Thr199, forme décrite comme pouvant réguler l'épissage *in vitro*, 2) l'anticorps anti-NPM qui reconnaît toutes les isoformes de NPM indépendamment de leur degré de phosphorylation (Tarapore *et al*, 2006).

De manière très intéressante, nous avons constaté que la quantité globale de nucléophosmine, ainsi que celle de la protéine phosphorylée au niveau du résidu Thr199 (NPM-T199) diminuaient fortement dans le lysat de cellules de patients atteints de DM1 (Fig. 87). Ce résultat confortait l'idée que NPM pourrait se fixer sur les répétitions CUG et être présente dans les *foci*. De ce fait, j'ai entrepris une étude de la localisation *in cellulo* de NPM en présence ou en absence de répétitions CUG.

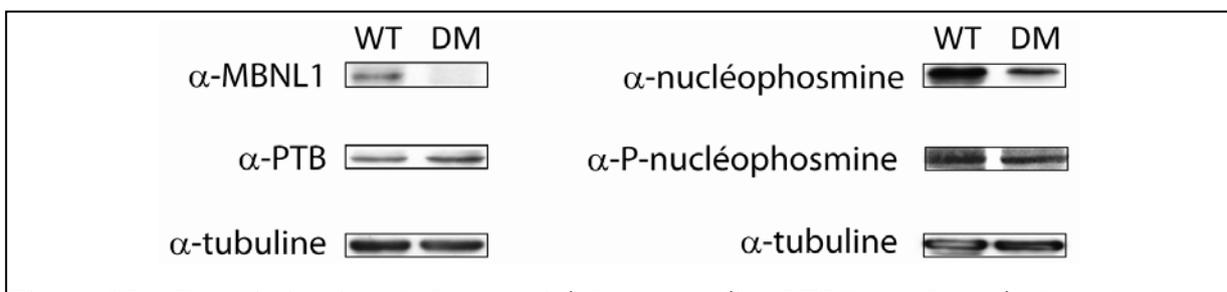


Figure 87 : Une diminution de la quantité de la protéine NPM est observée dans les lysats cellulaires préparés à partir de myoblastes de patients atteints de DM1.

Analyse par Western Blot de la quantité des différentes protéines au sein de lysats cellulaires préparés à partir de lignées immortalisées de myoblastes de patients sains (lysat WT) ou atteints de DM1 (lysat DM). L' α -tubuline est utilisée comme contrôle de dépôt. Le nom de l'anticorps utilisé pour réaliser le Western Blot est indiqué à gauche du gel.

II.6. Etude par immunofluorescence de la localisation cellulaire de la protéine NPM dans des myoblastes de patients atteints de DM1

J'ai donc entrepris des études d'imagerie cellulaires en vue d'observer les variations de localisation de NPM dans des cellules de patients atteints de DM1. On savait que dans des cellules normales, la forme phosphorylée NPM-T199 de NPM, est localisée au niveau des speckles nucléaires à la périphérie desquels sont localisés les *foci* CUG dans les cellules de patients atteints de DM1 (Smith *et al.*, 2007 ; Holt *et al.*, 2007). Ce travail a été réalisé en collaboration avec D. Furling (Institut de Myologie) et j'ai effectué plusieurs séjours dans son équipe afin de réaliser ces expériences. Dans un premier temps, j'ai mis au point des conditions expérimentales permettant d'utiliser les anticorps pour des analyses d'imagerie confocale. J'ai ainsi pu définir les dilutions optimales pour chacun des 2 anticorps (1/200^e pour l'anticorps anti-NPM et 1/100^e pour l'anticorps anti-P-NPM). Le protocole retenu est décrit dans la partie Matériel et Méthode.

J'ai ensuite utilisé les 2 anticorps dans ces conditions pour étudier la localisation par immunofluorescence de NPM et NPM-T199 à la fois dans des myoblastes de patients sains et des myoblastes de patients atteints de DM1. Les cellules ont été examinées sur un microscope confocal Leica TCS SP5 (Leica Microsystems, Milton Keynes, UK) et l'analyse des images a été réalisée avec le programme Imagej 1.36b (Rasband, <<http://rsb.info.nih.gov/ij/>>, 1997–2006). Aucune différence significative de localisation de NPM n'a pu être observée entre les 2 types cellulaires. En effet, NPM a toujours une localisation majoritairement nucléolaire et NPM-T199 a toujours un profil de distribution nucléaire de type « speckles » (Fig.88) (Tarapore *et al.*, 2006).

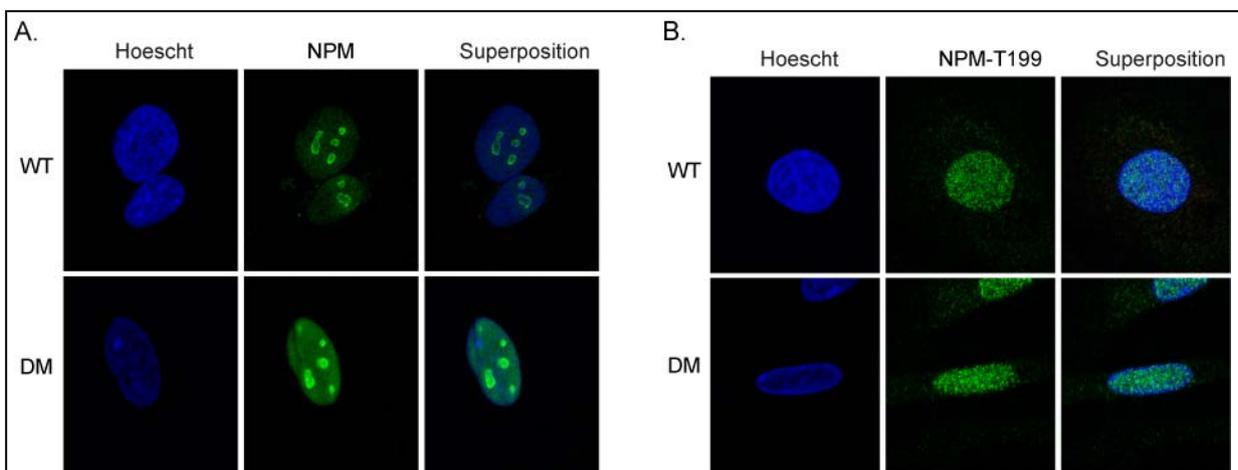


Figure 88 : Etude de localisation des protéines NPM et NPM-T199 au sein de myoblastes de patients sains ou de patients atteints de DM1.

Localisation des protéines NPM (A) et NPM-T199 (B) par immunofluorescence au sein de lignées immortalisées de myoblastes de patients sains (WT) ou atteints de DM1 (DM). Le noyau est coloré en

bleu (Hoescht) et la protéine NPM en vert. Superposition correspond à la superposition des images issues de la coloration du noyau et de NPM ou NPM-T199.

J'ai ensuite testé par immunofluorescence si la protéine NPM-T199 colocalisait avec MBNL1 dans des myoblastes de patients atteints de DM1 avec en comparaison des myoblastes normaux. Dans les deux cas, j'ai observé très peu de co-localisation des deux protéines. Cependant, peu de *foci* MBNL1 étaient détectés par cellules dans ces expériences (Fig.89).

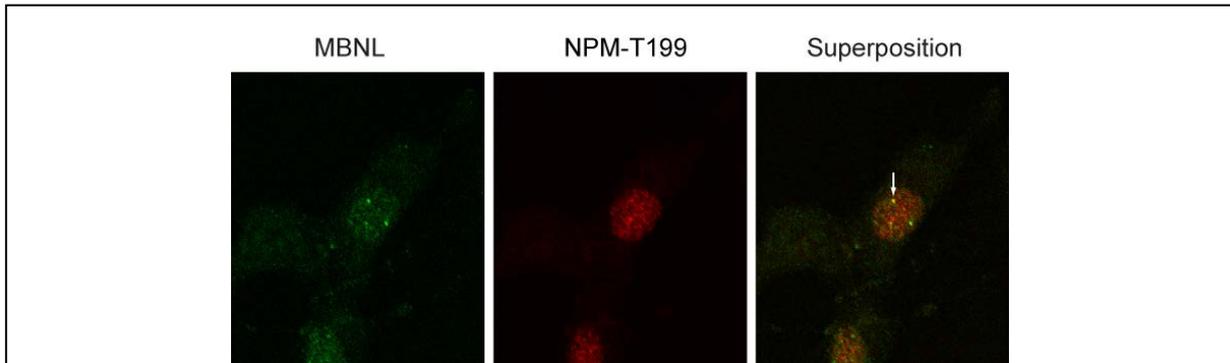


Figure 89 : Co-localisation de MBNL1 et NPM-T199 au sein de myoblastes de patients atteints de DM1.

Immunolocalisation des protéines NPM-T199 et MBNL1 dans des myoblastes de patients DM1.

La protéine MBNL1 est colorée en rouge et la protéine NPM en vert. Superposition correspond à la superposition des images issues de la coloration de MBNL1 et de NPM-T199.

La flèche blanche indique une superposition entre le signal des répétitions CUG et celui de la protéine NPM-T199.

Afin d'améliorer l'identification des foci nucléaires, nous avons alors utilisé à la place de l'anticorps anti-MBNL1, une sonde oligonucléotide Cy3-(CAG)₁₀ qui reconnaît spécifiquement les répétitions CUG. Le protocole expérimental utilisé est décrit dans la partie Matériel et Méthode. J'ai ainsi réalisé des études de co-localisation de NPM-T199 par immunofluorescence couplé au RNA-FISH, dans des cellules de patients atteints de DM1 (Fig.90). Nous avons détecté une faible co-localisation de NPM-T199 avec les *foci* nucléaires. Nous ne pouvons donc pas exclure la présence de NPM sous sa forme NPM-T199 dans les *foci* nucléaires.

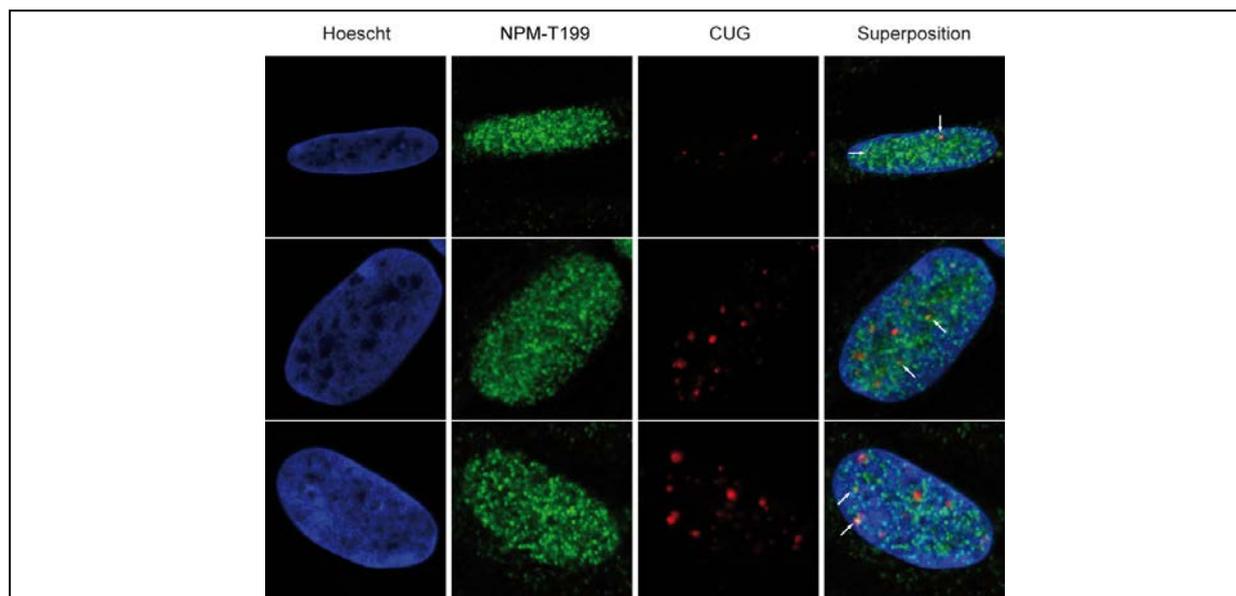


Figure 90: Co-localisation de NPM-T199 et des *foci* nucléaires au sein de myoblastes de patients atteints de DM1.

Localisation des *foci* nucléaires par détection de l'ARN CUG (RNA-FISH) par une sonde oligonucléotide Cy3-(CAG)₁₀ et par immunolocalisation de la protéine NPM-T199 dans des myoblastes de patients DM1.

Le noyau est coloré en bleu (Hoescht), la protéine NPM-T199 en vert et les répétitions CUG en rouge. Superposition correspond à la superposition des trois images issues de la coloration du noyau, de NPM-T199 et des répétitions CUG.

Les flèches blanches indiquent une superposition entre le signal des répétitions CUG et celui de la protéine NPM-T199.

III. Conclusion et perspectives

La recherche par l'approche MS2/MBP des protéines nucléaires capables d'interagir avec l'ARN (CUG)₅₁-MS2 nous a permis d'identifier 101 protéines partenaires potentielles. En répétant l'expérience dans des conditions plus stringentes et en incluant l'ARN (CAG)₅₉-MS2, nous avons trouvé un nombre beaucoup plus limité de protéines qui pourraient se lier de manière stable aux répétitions CUG. De façon surprenante la protéine MBNL1 n'a jamais été détectée au sein des complexes purifiés. Cette absence de détection de MBNL1 pourrait être liée à : i) la faible quantité de protéine MBNL1 présente au sein des extraits nucléaires utilisés ; ii) la faible capacité à détecter la protéine MBNL1 par spectrométrie de masse du fait d'une mauvaise ionisation des fragments tryptiques générés. Il est intéressant de noter que MBNL1 n'a également pas pu être détectée lors de la purification des complexes formés sur des ARN contenant des répétitions CUG réalisée par l'équipe de J. Marie (Gif-sur-Yvette) (com. pers.)

Parmi les protéines identifiées par spectrométrie de masse la protéine NPM a tout particulièrement retenu notre attention du fait de : i) son abondance dans les complexes purifiés et la reproductibilité de sa détection au sein des différentes expériences réalisées, ii) sa capacité à réguler l'épissage alternatif *in vitro* et *in vivo*, iii) la localisation dans les speckles de la forme phosphorylée NPM-T199. J'ai pu montrer *in vitro* la fixation directe de la protéine NPM1 recombinante sur des ARN contenant 51 répétitions CUG ou 59 répétitions CAG. L'affinité de NPM pour les ARN CUG₅₁ (Kd ~ 600 nM) est similaire à celle de la protéine MBNL1_{ΔCT3} pour ces mêmes ARN (Kd ~ 500 nM), ce qui fait de NPM un très bon candidat pour interagir avec les répétitions CUG. Les données de la littérature et nos données d'immunofluorescence montrent que NPM est majoritairement localisée dans le nucléole, ce qui pourrait sembler incompatible avec une interaction de NPM avec l'ARNm *DMPK* muté. Cependant NPM est une protéine navette et par conséquent on peut supposer qu'une faible fraction de NPM libre est présente dans le nucléoplasme. Un autre argument en la faveur de l'existence de la protéine NPM en dehors du nucléole est la localisation spécifique de la forme phosphorylée de NPM (NPM-T199) dans les « speckles ». Comme cette forme phosphorylée a une meilleure affinité pour les ARN que la protéine NPM non phosphorylée, il serait intéressant de tester son affinité pour les ARN (CUG)₅₁. Cela pourra être réalisé en utilisant un mutant phosphomimétique de la protéine NPM-T199 : le mutant NPM T199D (Tarapore *et al.*, 2006). Une construction permettant la production de ce variant phosphomimétique sous forme recombinante a été produite par I. Behm-Ansmant et la protéine purifiée à partir de cellules bactériennes exprimant cette construction sera utilisée pour des expériences de retard sur gel.

Les expériences d'immunofluorescence et de FISH que j'ai réalisées montrent une faible co-localisation entre la protéine NPM-T199 et les *foci* nucléaires. Il semblerait donc que la protéine NPM sous sa forme NPM-T199 puisse interagir *in cellulo* avec les répétitions CUG, cependant ces données sont préliminaires et nécessitent d'être confirmées. L'observation d'une diminution de la quantité de NPM dans les lysats protéiques dépourvus de *foci* préparés à partir de myoblastes DM1 est également un argument en faveur d'une localisation partielle de NPM dans les *foci* nucléaires.

NPM peut réguler l'épissage de certains pré-ARNm (Tarapore *et al.*, 2006 ; Russo *et al.*, 2011). Afin de confirmer un rôle de NPM dans la DM1 il sera important de tester si des variations de la quantité de NPM intracellulaire ont un effet sur l'épissage alternatif des pré-ARNm dont l'épissage est dérégulé chez les patients DM1. Ceci pourra être testé dans des cellules Hek et des myoblastes sains et DM1 par surexpression de NPM ou diminution de son

expression par emploi de siRNA. Nous avons observé que NPM a une plus faible affinité pour l'ARN (CAG)₅₉ que pour l'ARN (CUG)₅₁. Cette affinité doit être insuffisante pour que NPM soit détectée dans les complexes formés par les ARN (CAG)₅₉-MS2 *in vitro* puisque nous ne l'avons pas détecté par spectrométrie de masse. Néanmoins, ces expériences de purification de complexe formés avec l'ARN CAG n'ont été réalisées qu'une fois et devront être répétées.

Lors des différentes expériences de purification de complexes réalisées d'autres protéines que NPM ont été détectées comme partenaires potentiels des répétitions CUG (Tableaux XII, XIII, XIV et XV). Ces protéines peuvent être classées dans différentes catégories fonctionnelles : i) des facteurs régulateurs de l'épissage, ii) des protéines impliquées dans le métabolisme de l'ARN, et iii) des protéines impliquées dans le métabolisme de l'ADN.

. Les facteurs régulateurs de l'épissage

Contrairement aux autres expériences de purification de RNP réalisées au laboratoire sur d'autres ARN, peu de protéines régulatrices de l'épissage ont été détectées dans nos expériences. Ainsi seules 2 protéines hnRNP (hnRNP C1/C2 et hnRNP F), 2 protéines SR (SRSF1 ou ASF/SF2 et SRSF3 ou SRp20) et 5 protéines appartenant ou associées au spliceosome (U5-200kD, PRP8, PRP19, snRNP-B et SF3b) ont été identifiées.

La colocalisation de la protéine hnRNP F avec les répétitions CUG chez les patients atteints de DM1 avait déjà été décrite (Jiang *et al.*, 2004). Par contre, les protéines hnRNP C1/C2 et SRSF1 que nous avons détectés dans les complexes n'avaient pas été détectés dans les *foci* de cellules de patients atteints de DM (Fardaei *et al.*, 2002 ; Paul *et al.*, 2006 ; Perdoni *et al.*, 2009). La présence de la protéine SRSF3 dans les *foci* nucléaires dans les cellules de patients DM n'avait jamais été étudiée. On notera que SRSF3 est impliquée dans la régulation de l'épissage alternatif de différents pré-ARNm tels que l'exon 11 de l'insuline récepteur (IR) ou l'exon E III A de la fibronectine (FN1), et que l'épissage de ces exons est dérégulé chez les patients atteints de DM1. Il serait donc intéressant de tester la localisation et le taux de ce facteur d'épissage dans les cellules de patients DM.

. Protéines impliquées dans le métabolisme de l'ARN

Nous avons également identifié de nombreuses ARN hélicases telles que les protéines DDX21, DHX9, DDX24, DDX55, DDX3X et DDX5. Ces hélicases jouent un rôle important dans la modulation de la structuration de l'ARN dans des mécanismes cellulaires tels que

l'export des ARN, l'assemblage des ribosomes, la traduction, la transcription ou encore la régulation de l'épissage. Ainsi, la protéine DDX5 régule l'intégration de l'exon 10 du pré-ARN tau par déstabilisation d'une structure tige-boucle, ce qui facilite la fixation de la U1snRNP sur le site 5' d'épissage et active l'inclusion de cet exon (Kar *et al.*, 2011).

L'ARN hélicase nucléolaire RH-II ou DDX21, présente dans l'ensemble des expériences de purification de complexes, est un régulateur de la transcription et la maturation des ARN ribosomiques. Cette protéine est normalement exprimée dans le nucléole et est redirigée vers le nucléoplasme, où elle interagit avec le facteur de transcription c-Jun dans des conditions de stress (pour revue, Fuller-Pace, 2006).

La seconde protéine détectée également dans les 3 expériences de purification de complexes est le régulateur de transcription DHX9 ou ARN hélicase A. De manière intéressante, cette protéine intervient également dans la maturation des ARNm et semble avoir la capacité de déstructurer les ARN double-brin (pour revue, Linder et Jankowsky, 2011). Récemment, il a été montré que DHX9 interagit directement avec la protéine MBNL1, de même que la protéine DDX5, également détectée dans la purification de complexe (Paul *et al.*, 2011). Une augmentation de la demi-vie de ces deux protéines a été observée dans les cellules de myoblastes de patients atteints de DM1. Une augmentation du taux de DDX5 entraîne une dérégulation de l'épissage de l'exon 11 du pré-ARNm IR, mimant le profil observé chez les patients DM1. Cependant, malgré leur interaction avec la protéine MBNL1, ces deux protéines ne semblent pas séquestrées au sein des *foci* nucléaires de patients atteints de DM1 (Paul *et al.*, 2011). On peut envisager comme cela a été proposé que ces hélicases facilitent la fixation de MBNL1 sur certains ARN structurés. On peut aussi envisager que la fixation de MBNL1 sur ces protéines puisse limiter leur activité et de ce fait que la séquestration de MBNL1 dans les *foci* permette d'augmenter l'action de ces hélicases connues pour réguler beaucoup de mécanismes cellulaires dont l'épissage alternatif

. Protéines impliquées dans le métabolisme de l'ADN

Un grand nombre des protéines identifiées dans les complexes que nous avons purifiés sont des protéines impliquées dans le métabolisme de l'ADN : des ADN hélicases, des protéines de réparation de l'ADN, des histones et des topo-isomérases. Nous avons notamment identifié l'ensemble des protéines du facteur de réplication C (RFC). RFC est un complexe constitué de cinq sous-unités (RFC1-5), qui se lie aux amorces d'ARN et favorise la fixation du facteur PCNA sur l'ADN, qui augmente la processivité des ADN polymérases (Stillman, 2008). Ce complexe en association avec PARP-1, XRCC1 et l'ADN ligase III

intervient aussi dans les mécanismes de réparation de l'ADN. Comme nous l'avons vu dans l'introduction, les expansions de répétitions CTG se produisent entre autre au cours des processus de réplication et de réparation. On pourrait se demander sur la base de nos données si la présence de longues répétitions pourraient séquestrer certains facteurs de réparation de l'ADN et donc favoriser l'expansion des amplifications. La détection de nombreuses protéines spécifiques de l'ADN pourrait s'expliquer par le fait que les répétitions CUG se structurent en une hélice de type ADN.

Par ailleurs, de nombreuses protéines identifiées dans les complexes que nous avons purifiés sont impliquées dans la structuration de la chromatine et les modifications épigénétiques de l'ADN. Nous avons détecté 3 des protéines histones H1, une protéine de fixation aux histones (RBBP4), CAF1 qui intervient dans l'enroulement de l'ADN lors de la formation des nucléosomes et BAZ1A qui est un composant d'un complexe de remodelage de la chromatine. La nucléophosmine agit également comme une protéine chaperonne des histones. Certaines protéines impliquées dans le processus de régulation de l'épissage jouent également un rôle dans la méthylation de l'ADN ou des histones tels que la protéine U5-200kD ou les protéines hnRNP C1/C2, citées précédemment. En effet, les protéine hnRNP C1/C2 participent à la formation d'un complexe de remodelage de l'ADN qui intervient au niveau du locus de la région contrôle de la β -globine humaine (Mahajan *et al.*, 2004). De même la protéine U5-200kD intervient dans la triméthylation de la lysine 36 de l'histone 3 (H3K36me3). Il n'y a cependant pas à l'heure actuelle de lien établi entre la présence de répétitions CUG et des dérégulations des mécanismes épigénétiques. Seule une hyperméthylation de l'îlot CpG dans le gène DMPK chez les patients atteints de la forme congénitale a été caractérisée mais peut être que cette absence de lien connu est liée à une absence de recherche dans ce domaine.

*Discussion et
Perspectives*

L'effet *trans*-dominant des répétitions CUG sur l'épissage alternatif est considéré comme l'hypothèse physiopathologique majeure permettant d'expliquer le phénotype multi-systémique de la DM1. La majorité des défauts d'épissage a été jusque là attribuée à la séquestration de MBNL1 au sein de *foci* nucléaires suite à son interaction avec les longues structures tige-boucle imparfaites formées par les répétitions CUG. En plus de sa capacité à fixer cette structure particulière, la protéine MBNL1 se lie également sur ses séquences cibles au sein des pré-ARNm et des pré-miRNA. Comme mentionné précédemment beaucoup de stratégies thérapeutiques développées actuellement visent à inhiber la liaison de MBNL1 aux répétitions CUG sans altérer son interaction avec ses pré-ARNm cibles. Il est donc primordial de comprendre le mode de liaison de la protéine MBNL1 à ces différents types d'ARN, afin de savoir s'il est possible de découpler sa fixation sur les répétitions CUG, de la fixation sur les autres ARN cellulaires. Lorsque j'ai débuté ce travail de thèse, peu de choses étaient connues sur les propriétés de fixation à l'ARN de MBNL1. Notamment, aucune étude expérimentale n'avait permis de déterminer si MBNL1 reconnaissait uniquement les motifs structuraux similaires aux hélices de type A formées par les répétitions CUG, ou si elle utilise d'autres modes de fixation sur ses cibles potentielles. Par ailleurs les domaines de MBNL1 intervenant dans la reconnaissance des ARN n'avaient pas été identifiés et peu de données sur le mécanisme de régulation de l'épissage alternatif des pré-ARNm impliquant MBNL1 étaient disponibles.

Mon travail de thèse a permis de mettre en évidence l'existence d'au moins trois modes distincts de reconnaissance des ARN par MBNL1 :

- une reconnaissance de type « répétitions CUG » avec une fixation de MBNL1 le long de la structure tige-boucle particulière formée par les répétitions CUG,
- une reconnaissance de type « pré-ARNm » avec généralement la présence de plusieurs sites MBNL1 adjacents, les régions des pré-ARNm correspondant ne semblent pas présenter de caractéristiques structurales communes. Les ARN SELEX de la famille I que nous avons obtenus et caractérisés appartiennent à cette catégorie,
- une reconnaissance de type « pré-miRNA » impliquant la reconnaissance d'un seul motif YGCY localisé au sein d'une boucle. Les ARN SELEX de la famille II appartiennent à cette catégorie, mais présentent des caractéristiques tant en séquence qu'en structure (taille de la boucle et de l'appariement qui la ferme) conduisant à une très forte affinité de MBNL1 pour ces ARN. La plus faible affinité de MBNL1 pour le motif YGCY présent au sein de certains pré-miRNA s'explique sans doute par l'architecture différente de la

boucle et le fait que l'affinité de MBNL1 pour ces ARN ne doit pas être trop forte puisqu'il s'agit d'un mécanisme de régulation de la maturation du pré-miR qui dépend en particulier des conditions physiologiques.

Il apparaît donc que la reconnaissance des ARN par MBNL1 est beaucoup plus complexe qu'initialement proposé mais il est possible d'envisager des stratégies thérapeutiques basées sur nos résultats de SELEX. On peut en effet imaginer l'utilisation d'un oligonucléotide bifonctionnel comportant d'une part le motif de forte affinité pour MBNL1 que nous avons caractérisé et d'autre part une séquence antisens, permettant de délivrer MBNL1 spécifiquement à proximité d'un site d'épissage où elle est nécessaire, afin de corriger le défaut d'épissage généré par un déficit en MBNL1 en particulier ceux liés à l'expansion de répétitions CUG.

Toujours dans le but de comprendre le mode de reconnaissance des différents ARN par MBNL1, il était nécessaire d'identifier les éléments de la protéine nécessaires à sa fixation sur l'ARN. Au cours de notre travail, nous avons démontré que l'ensemble de la région N-terminale (codée par les exons 1 à 4) et notamment la séquence codée par l'exon 3 de la protéine MBNL1 intervient dans la fixation à l'ARN. Cette région contient les deux paires de doigts à Zn^{2+} permettant la fixation sur l'ARN. Une explication possible de l'existence de différents modes d'interaction de MBNL1 avec les ARN serait une implication différentielle des différents motifs à doigts à Zn^{2+} (ZnF1 à 4) dans la reconnaissance des différents ARN partenaires. Afin de tester cette hypothèse et d'avoir une idée plus précise des modalités de reconnaissance des ARN par MBNL1, nous avons débuté une collaboration avec H. Urlaub (MPI Goettingen). Le but de cette collaboration était d'identifier les acides aminés de la protéine MBNL1 directement impliqués dans l'interaction avec les ARN par une approche basée sur l'utilisation combinée de pontage aux UV et d'identification par spectrométrie de masse des résidus pontés, réalisée après hydrolyse des complexes pontés par un mélange de ribonucléases et d'endoprotéases. Une première analyse des complexes formés entre la protéine MBNL1 $_{\Delta CT3}$ et l'ARN AC3 issu de la famille II du SELEX (contenant le long motif YGUUUGCGCUUG) a été réalisée. Les résultats préliminaires obtenus ont permis de vérifier que la stœchiométrie de ce complexe était de 1 :1 et d'identifier deux peptides crosslinkés correspondants aux motifs ZnF2 et ZnF4. Soit nous avons à faire à un mélange de complexes dans lesquels soit ZnF2 soit ZnF4 interagissait avec l'ARN, soit les 2 paires de doigts à Zn^{2+} interagissaient avec l'ARN. Il faudra tester ces 2 hypothèses en étudiant la formation de complexes avec chacun des motifs doigts à Zn^{2+} pris individuellement, ce qui nous amènera

à retravailler les problèmes de production de ces motifs pris individuellement. Ensuite, il sera important de poursuivre cette étude afin d'identifier plus précisément les nucléotides pontés. Par la même approche, nous pourrions essayer d'identifier les acides aminés impliqués dans l'interaction de MBNL1 avec des ARN contenant des répétitions CUG et des segments d'ARN présents dans les pré-ARNm cibles et des pré-miRNA. Nous pourrions ainsi déterminer si les peptides impliqués dans la reconnaissance de ces différents ARN sont identiques ou différents. L'utilisation dans ce type d'expériences de la protéine MBNL1 Δ CT, nous permettra d'étudier le rôle de la séquence codée par l'exon 3 dans la reconnaissance des ARN. Par ailleurs une collaboration a été débutée avec l'équipe de F. Allain (ETH, Zurich) afin de résoudre par RMN la structure de complexes formés entre la protéine MBNL1 et ses différents ARN partenaires dont les ARN SELEX de la famille II. Cette approche devrait nous permettre de décrypter les mécanismes de reconnaissance à l'échelle atomique. Une telle connaissance sera essentielle au développement de molécules à potentialité thérapeutiques.

Une meilleure compréhension du mécanisme de dérégulation de l'épissage alternatif dans la DM1 et plus particulièrement du rôle de MBNL1 dans la physiopathologie représente aussi une étape nécessaire dans le développement de nouvelles thérapies pour la DM ou d'autres maladies humaines à gain de fonction présentant également des défauts d'épissage liés à la protéine MBNL1. Une partie de mon travail de thèse a consisté à approfondir les connaissances sur les mécanismes de régulation de l'épissage alternatif par MBNL1. En effet, lorsque j'ai débuté ma thèse peu de données étaient disponibles dans ce domaine. Lors de notre étude de la régulation de l'inclusion de l'exon 5 du pré-ARNm hcTNT, nous avons identifié 6 nouveaux sites de fixation de MBNL1, localisés en amont (3 sites, groupe 2) et en aval (3 sites, groupe 3) de l'exon 5. L'étude de l'importance de ces sites dans la régulation de l'inclusion de l'exon 5 du pré-ARNm hcTNT, nous a conduit à proposer un nouveau modèle de régulation de cet événement d'épissage par MBNL1. Dans ce modèle, la régulation de l'inclusion de l'exon 5 par MBNL1 nécessiterait la formation d'une structure à longue distance au sein du pré-ARNm par interaction des protéines fixées sur les deux groupes de sites, ce qui isolerait l'exon 5 dans une boucle et limiterait la fixation des composants du spliceosome comme cela a déjà été proposé pour d'autres facteurs d'épissage comme hnRNP A1 et PTB (Wagner et Garcia-Blanco, 2001 ; Nasim *et al.*, 2002). Ce type de modèle de régulation par MBNL1 n'avait encore jamais été évoqué.

Nous devons d'abord vérifier expérimentalement les capacités de MBNL1 à former cette structure à longue distance. Il sera ensuite intéressant de rechercher si ce mode de régulation

négative de l'épissage par MBNL1 est utilisé pour la régulation de l'épissage d'autres pré-ARNm. L'équipe de Berglund a recherché la présence de motifs YGCY dans les 150 nt introniques situés en amont et en aval des 24 exons dont l'inclusion est dérégulée chez les patients atteints de DM (Goers *et al.*, 2010). Certains de ces exons, régulés négativement par MBNL1, possèdent des sites de fixation de MBNL1 à la fois dans les séquences introniques présentes en amont et en aval de l'exon et pourraient donc être régulés selon le modèle de régulation que nous proposons. En revanche, d'autres exons régulés positivement ou négativement par MBNL1 ne présentent des sites de fixation qu'en amont ou en aval, indiquant que d'autres modes de régulation par MBNL1 existent. Nous pourrions dans un premier temps vérifier si les exons susceptibles d'être régulés selon notre modèle le sont effectivement. Par ailleurs, MBNL1 étant impliquée dans la régulation d'événements d'épissage alternatif n'ayant pas de lien avec la DM, il sera intéressant de tester si notre modèle de régulation peut être utilisé pour la régulation négative de ces événements. J. Venables, qui a rejoint récemment notre équipe, a réalisé une analyse des dérégulations de l'épissage se produisant lors de la déplétion de la protéine MBNL1 par emploi de siRNA. La recherche des motifs MBNL1 au sein des régions introniques flanquant les exons dérégulés est actuellement en cours et l'existence de notre modèle de régulation pourra être testée au sein de ces pré-ARNm. L'établissement d'une liste plus exhaustive des pré-ARNm dont l'épissage est régulé par MBNL1 pourra également être réalisé par l'identification exhaustive des ARN partenaires de MBNL1 au sein d'une cellule par emploi de la méthode CLIP (Cross-Linking and ImmunoPrecipitation). Nous pourrions là encore voir si le mécanisme de régulation que nous proposons peut être généralisé. Une collaboration avec l'équipe de J. Ule (Cambridge) qui a mis au point la méthode CLIP a été développée. Deux expériences CLIP MBNL1 ont déjà été réalisées en collaboration avec J. Ule. Cependant, aucun résultat concluant n'a été obtenu en raison de la faible efficacité des anticorps utilisés à immunoprécipiter MBNL1. Cette expérience devra être répétée en utilisant d'autres anticorps dirigés contre la protéine MBNL1 ou un mélange de ces anticorps afin d'augmenter le rendement de l'immunosélection.

Comme mentionné précédemment la majorité des défauts d'épissage observés chez les patients DM est généralement attribuée à la séquestration de MBNL1 au sein de *foci* nucléaires. Cependant les données de la littérature suggéraient la présence dans les *foci* nucléaires de protéines autres que MBNL1 qui pourraient avoir un rôle dans le phénomène de rétention et/ou dans la dérégulation de l'épissage. Notre recherche par l'approche MBP-MS2

des protéines nucléaires capables d'interagir avec un ARN contenant 51 répétitions CUG conforte cette idée, puisqu'elle a permis l'identification d'un grand nombre de protéines potentiellement partenaires de ces répétitions. L'un de ces candidats, la protéine NPM, a retenu notre attention et la caractérisation de l'interaction de cette protéine avec les répétitions CUG a été débutée. Cette étude sera poursuivie. Par ailleurs, d'autres protéines candidates, telles que des ARN hélicases, des protéines histones ou les composants du facteur de réplication et de réparation de l'ADN sont très intéressants et leur présence dans les *foci* mérite d'être étudiée. Ainsi les ARN hélicases DDX5 et DHX9, retrouvées au sein des complexes que nous avons purifiés, ont été décrites comme formant un complexe avec la protéine MBNL1 dans les cellules (Paul et *al.*, 2011). On peut envisager comme cela a été proposé que ces hélicases facilitent la fixation de MBNL1 sur certains ARN structurés. On peut aussi envisager que la fixation de MBNL1 sur ces protéines puisse limiter leur activité et de ce fait que la séquestration de MBNL1 dans les *foci* permette d'augmenter l'action de ces hélicases connues pour réguler beaucoup de mécanismes cellulaires dont l'épissage alternatif.

Bibliographie

- Abe, Y., Chen W., Huang W., Nishino M. and Li Y. P.** (2006). CNBP regulates forebrain formation at organogenesis stage in chick embryos. *Dev Biol*, **295**(1): 116-27.
- Achsel, T., Brahms H., Kastner B., Bachi A., Wilm M. and Luhrmann R.** (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J*, **18**(20): 5789-802.
- Adereth, Y., Dammai V., Kose N., Li R. and Hsu T.** (2005). RNA-dependent integrin alpha3 protein localization regulated by the Muscleblind-like protein MLP1. *Nat Cell Biol*, **7**(12): 1240-7.
- Akusjarvi, G. and Stevenin J.** (2003). Remodelling of the host cell RNA splicing machinery during an adenovirus infection. *Curr Top Microbiol Immunol*, **272**(253-86).
- Alexander, R. and Beggs J. D.** (2010). Cross-talk in transcription, splicing and chromatin: who makes the first call? *Biochem Soc Trans*, **38**(5): 1251-6.
- Allemand, E., Guil S., Myers M., Moscat J., Caceres J. F. and Krainer A. R.** (2005). Regulation of heterogenous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock. *Proc Natl Acad Sci U S A*, **102**(10): 3605-10.
- Allo, M., Buggiano V., Fededa J. P., Petrillo E., Schor I., De La Mata M., Agirre E., Plass M., Eyras E., Elela S. A., Klinck R., Chabot B. and Kornblihtt A. R.** (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol*, **16**(7): 717-24.
- Alwazzan, M., Newman E., Hamshere M. G. and Brook J. D.** (1999). Myotonic dystrophy is associated with a reduced level of RNA from the DMWD allele adjacent to the expanded repeat. *Hum Mol Genet*, **8**(8): 1491-7.
- Amack, J. D., Paguio A. P. and Mahadevan M. S.** (1999). Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model. *Hum Mol Genet*, **8**(11): 1975-84.
- Andersson, R., Enroth S., Rada-Iglesias A., Wadelius C. and Komorowski J.** (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res*, **19**(10): 1732-41.
- Armas, P., Aguero T. H., Borgognone M., Aybar M. J. and Calcaterra N. B.** (2008). Dissecting CNBP, a zinc-finger protein required for neural crest development, in its structural and functional domains. *J Mol Biol*, **382**(4): 1043-56.
- Artero, R., Prokop A., Paricio N., Begemann G., Pueyo I., Mlodzik M., Perez-Alonso M. and Baylies M. K.** (1998). The muscleblind gene participates in the organization of Z-bands and epidermal attachments of Drosophila muscles and is regulated by Dmef2. *Dev Biol*, **195**(2): 131-43.
- Ashizawa, T.** (1998). Myotonic dystrophy as a brain disorder. *Arch Neurol*, **55**(3): 291-3.
- Auboeuf, D., Dowhan D. H., Li X., Larkin K., Ko L., Berget S. M. and O'malley B. W.** (2004). CoAA, a nuclear receptor coactivator protein at the interface of transcriptional coactivation and RNA splicing. *Mol Cell Biol*, **24**(1): 442-53.
- Bagga, R., Armstrong J. A. and Emerson B. M.** (1998). Role of chromatin structure and distal enhancers in tissue-specific transcriptional regulation in vitro. *Cold Spring Harb Symp Quant Biol*, **63**(569-76).
- Barreau, C., Paillard L., Mereau A. and Osborne H. B.** (2006a). Mammalian CELF/Bruno-like RNA-binding proteins: molecular characteristics and biological functions. *Biochimie*.
- Barreau, C., Paillard L., Mereau A. and Osborne H. B.** (2006b). Mammalian CELF/Bruno-like RNA-binding proteins: molecular characteristics and biological functions. *Biochimie*, **88**(5): 515-25.
- Barreau, C., Watrin T., Beverley Osborne H. and Paillard L.** (2006c). Protein expression is increased by a class III AU-rich element and tethered CUG-BP1. *Biochem Biophys Res Commun*, **347**(3): 723-30.

- Bassez, G., Lazarus A., Desguerre I., Varin J., Laforet P., Becane H. M., Meune C., Arne-Bes M. C., Ounnoughene Z., Radvanyi H., Eymard B. and Duboc D.** (2004). Severe cardiac arrhythmias in young patients with myotonic dystrophy type 1. *Neurology*, **63**(10): 1939-41.
- Batsche, E., Yaniv M. and Muchardt C.** (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol*, **13**(1): 22-9.
- Beckmann, J. S. and Trifonov E. N.** (1991). Splice junctions follow a 205-base ladder. *Proc Natl Acad Sci U S A*, **88**(6): 2380-3.
- Bedard, K. M., Daijogo S. and Semler B. L.** (2007). A nucleo-cytoplasmic SR protein functions in viral IRES-mediated translation initiation. *EMBO J*, **26**(2): 459-67.
- Begemann, G., Paricio N., Artero R., Kiss I., Perez-Alonso M. and Mlodzik M.** (1997). muscleblind, a gene required for photoreceptor differentiation in Drosophila, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. *Development*, **124**(21): 4321-31.
- Bensaid, M., Melko M., Bechara E. G., Davidovic L., Berretta A., Catania M. V., Gecz J., Lalli E. and Bardoni B.** (2009). FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic Acids Res*, **37**(4): 1269-79.
- Bentley, D.** (2002). The mRNA assembly line: transcription and processing machines in the same factory. *Curr Opin Cell Biol*, **14**(3): 336-42.
- Bentley, D. L.** (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol*, **17**(3): 251-6.
- Berul, C. I., Maguire C. T., Aronovitz M. J., Greenwood J., Miller C., Gehrman J., Housman D., Mendelsohn M. E. and Reddy S.** (1999). DMPK dosage alterations result in atrioventricular conduction abnormalities in a mouse myotonic dystrophy model. *J Clin Invest*, **103**(4): R1-7.
- Beyersmann, D.** (2000). Regulation of mammalian gene expression. *Exs*, **89**(11-28).
- Biamonti, G., Ruggiu M., Saccone S., Della Valle G. and Riva S.** (1994). Two homologous genes, originated by duplication, encode the human hnRNP proteins A2 and A1. *Nucleic Acids Res*, **22**(11): 1996-2002.
- Black, D. L.** (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, **72**(291-336).
- Blencowe, B. J.** (2006). Alternative splicing: new insights from global analyses. *Cell*, **126**(1): 37-47.
- Blencowe, B. J., Issner R., Nickerson J. A. and Sharp P. A.** (1998). A coactivator of pre-mRNA splicing. *Genes Dev*, **12**(7): 996-1009.
- Borgognone, M., Armas P. and Calcaterra N. B.** (2010). Cellular nucleic-acid-binding protein, a transcriptional enhancer of c-Myc, promotes the formation of parallel G-quadruplexes. *Biochem J*, **428**(3): 491-8.
- Botta, A., Caldarola S., Vallo L., Bonifazi E., Fruci D., Gullotta F., Massa R., Novelli G. and Loreni F.** (2006). Effect of the [CCTG]_n repeat expansion on ZNF9 expression in myotonic dystrophy type II (DM2). *Biochim Biophys Acta*, **1762**(3): 329-34.
- Botta, A., Vallo L., Rinaldi F., Bonifazi E., Amati F., Biancolella M., Gambardella S., Mancinelli E., Angelini C., Meola G. and Novelli G.** (2007). Gene expression analysis in myotonic dystrophy: indications for a common molecular pathogenic pathway in DM1 and DM2. *Gene Expr*, **13**(6): 339-51.
- Boucher, C. and Johnson K.** (1995). Triplet repeats in neuromuscular disorders. *Ann Med*, **27**(1): 3-5.
- Boucher, L., Ouzounis C. A., Enright A. J. and Blencowe B. J.** (2001). A genome-wide survey of RS domain proteins. *RNA*, **7**(12): 1693-701.

- Bourgeois, C. F., Popielarz M., Hildwein G. and Stevenin J.** (1999). Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation. *Mol Cell Biol*, **19**(11): 7347-56.
- Branlant, C., Krol A., Ebel J. P., Lazar E., Gallinaro H., Jacob M., Sri-Widada J. and Jeanteur P.** (1980). Nucleotide sequences of nuclear U1A RNAs from chicken, rat and man. *Nucleic Acids Res*, **8**(18): 4143-54.
- Branlant, C., Krol A., Ebel J. P., Lazar E., Haendler B. and Jacob M.** (1982). U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *EMBO J*, **1**(10): 1259-65.
- Branlant, C., Krol A., Lazar E., Haendler B., Jacob M., Galego-Dias L. and Pousada C.** (1983). High evolutionary conservation of the secondary structure and of certain nucleotide sequences of U5 RNA. *Nucleic Acids Res*, **11**(23): 8359-67.
- Bres, V., Gomes N., Pickle L. and Jones K. A.** (2005). A human splicing factor, SKIP, associates with P-TEFb and enhances transcription elongation by HIV-1 Tat. *Genes Dev*, **19**(10): 1211-26.
- Brodsky, A. S., Meyer C. A., Swinburne I. A., Hall G., Keenan B. J., Liu X. S., Fox E. A. and Silver P. A.** (2005). Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol*, **6**(8): R64.
- Brook, J. D., Mccurrach M. E., Harley H. G., Buckler A. J., Church D., Aburatani H., Hunter K., Stanton V. P., Thirion J. P., Hudson T. and Et Al.** (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, **68**(4): 799-808.
- Brow, D. A. and Guthrie C.** (1988). Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature*, **334**(6179): 213-8.
- Brunner, H. G., Bruggenwirth H. T., Nillesen W., Jansen G., Hamel B. C., Hoppe R. L., De Die C. E., Howeler C. J., Van Oost B. A., Wieringa B. and Et Al.** (1993). Influence of sex of the transmitting parent as well as of parental allele size on the CTG expansion in myotonic dystrophy (DM). *Am J Hum Genet*, **53**(5): 1016-23.
- Buckanovich, R. J., Posner J. B. and Darnell R. B.** (1993). Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron*, **11**(4): 657-72.
- Buee, L., Bussiere T., Buee-Scherrer V., Delacourte A. and Hof P. R.** (2000). Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Res Brain Res Rev*, **33**(1): 95-130.
- Buj-Bello, A., Furling D., Tronchere H., Laporte J., Lerouge T., Butler-Browne G. S. and Mandel J. L.** (2002). Muscle-specific alternative splicing of myotubularin-related 1 gene is impaired in DM1 muscle cells. *Hum Mol Genet*, **11**(19): 2297-307.
- Buratti, E., Dork T., Zuccato E., Pagani F., Romano M. and Baralle F. E.** (2001). Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J*, **20**(7): 1774-84.
- Burd, C. G. and Dreyfuss G.** (1994). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J*, **13**(5): 1197-204.
- Bushby, K., Muntoni F. and Bourke J. P.** (2003). 107th ENMC international workshop: the management of cardiac involvement in muscular dystrophy and myotonic dystrophy. 7th-9th June 2002, Naarden, the Netherlands. *Neuromuscul Disord*, **13**(2): 166-72.
- Buxton, J., Shelbourne P., Davies J., Jones C., Van Tongeren T., Aslanidis C., De Jong P., Jansen G., Anvret M., Riley B. and Et Al.** (1992). Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature*, **355**(6360): 547-8.
- Caceres, J. F. and Kornblihtt A. R.** (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet*, **18**(4): 186-93.

- Callis, T. E., Pandya K., Seok H. Y., Tang R. H., Tatsuguchi M., Huang Z. P., Chen J. F., Deng Z., Gunn B., Shumate J., Willis M. S., Selzman C. H. and Wang D. Z.** (2009). MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. *J Clin Invest*, **119**(9): 2772-86.
- Campuzano, V., Montermini L., Molto M. D., Pianese L., Cossee M., Cavalcanti F., Monros E., Rodius F., Duclos F., Monticelli A., Zara F., Canizares J., Koutnikova H., Bidichandani S. I., Gellera C., Brice A., Trouillas P., De Michele G., Filla A., De Frutos R., Palau F., Patel P. I., Di Donato S., Mandel J. L., Coccozza S., Koenig M. and Pandolfo M.** (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*, **271**(5254): 1423-7.
- Caputi, M., Mayeda A., Krainer A. R. and Zahler A. M.** (1999). hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *EMBO J*, **18**(14): 4060-7.
- Caputi, M. and Zahler A. M.** (2001). Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J Biol Chem*, **276**(47): 43850-9.
- Cardani, R., Baldassa S., Botta A., Rinaldi F., Novelli G., Mancinelli E. and Meola G.** (2009). Ribonuclear inclusions and MBNL1 nuclear sequestration do not affect myoblast differentiation but alter gene splicing in myotonic dystrophy type 2. *Neuromuscul Disord*, **19**(5): 335-43.
- Cardani, R., Mancinelli E., Rotondo G., Sansone V. and Meola G.** (2006). Muscleblind-like protein 1 nuclear sequestration is a molecular pathology marker of DM1 and DM2. *Eur J Histochem*, **50**(3): 177-82.
- Cardinali, B., Carissimi C., Gravina P. and Pierandrei-Amaldi P.** (2003). La protein is associated with terminal oligopyrimidine mRNAs in actively translating polysomes. *J Biol Chem*, **278**(37): 35145-51.
- Carmel, I., Tal S., Vig I. and Ast G.** (2004). Comparative analysis detects dependencies among the 5' splice-site positions. *RNA*, **10**(5): 828-40.
- Cartegni, L., Maconi M., Morandi E., Cobianchi F., Riva S. and Biamonti G.** (1996). hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J Mol Biol*, **259**(3): 337-48.
- Chabot, B., Lebel C., Hutchison S., Nasim F. H. and Simard M. J.** (2003). Heterogeneous nuclear ribonucleoprotein particle A/B proteins and the control of alternative splicing of the mammalian heterogeneous nuclear ribonucleoprotein particle A1 pre-mRNA. *Prog Mol Subcell Biol*, **31**(59-88).
- Charlet, B. N., Logan P., Singh G. and Cooper T. A.** (2002a). Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. *Mol Cell*, **9**(3): 649-58.
- Charlet, B. N., Savkur R. S., Singh G., Philips A. V., Grice E. A. and Cooper T. A.** (2002b). Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol Cell*, **10**(1): 45-53.
- Chen, W., Wang Y., Abe Y., Cheney L., Udd B. and Li Y. P.** (2007). Haploinsufficiency for Znf9 in Znf9^{+/-} mice is associated with multiorgan abnormalities resembling myotonic dystrophy. *J Mol Biol*, **368**(1): 8-17.
- Cho, D. H. and Tapscott S. J.** (2007). Myotonic dystrophy: emerging mechanisms for DM1 and DM2. *Biochim Biophys Acta*, **1772**(2): 195-204.
- Cho, D. H., Thienes C. P., Mahoney S. E., Analau E., Filippova G. N. and Tapscott S. J.** (2005). Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol Cell*, **20**(3): 483-9.
- Chodavarapu, R. K., Feng S., Bernatavichute Y. V., Chen P. Y., Stroud H., Yu Y., Hetzel J. A., Kuo F., Kim J., Cokus S. J., Casero D., Bernal M., Huijser P., Clark A. T.,**

- Kramer U., Merchant S. S., Zhang X., Jacobsen S. E. and Pellegrini M.** (2010). Relationship between nucleosome positioning and DNA methylation. *Nature*, **466**(7304): 388-92.
- Choi, D. K., Ito T., Tsukahara F., Hirai M. and Sakaki Y.** (1999). Developmentally-regulated expression of mNapor encoding an apoptosis-induced ELAV-type RNA binding protein. *Gene*, **237**(1): 135-42.
- Cleary, J. D. and Pearson C. E.** (2005). Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. *Trends Genet*, **21**(5): 272-80.
- Cleary, J. D., Tome S., Lopez Castel A., Panigrahi G. B., Foiry L., Hagerman K. A., Sroka H., Chitayat D., Gourdon G. and Pearson C. E.** (2010). Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus. *Nat Struct Mol Biol*, **17**(9): 1079-87.
- Cocquempot, O., Brault V., Babinet C. and Herault Y.** (2009). Fork stalling and template switching as a mechanism for polyalanine tract expansion affecting the DYC mutant of HOXD13, a new murine model of synpolydactyly. *Genetics*, **183**(1): 23-30.
- Coffee, B., Zhang F., Warren S. T. and Reines D.** (1999). Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. *Nat Genet*, **22**(1): 98-101.
- Colleran, J. A., Hawley R. J., Pinnow E. E., Kokkinos P. F. and Fletcher R. D.** (1997). Value of the electrocardiogram in determining cardiac events and mortality in myotonic dystrophy. *Am J Cardiol*, **80**(11): 1494-7.
- Collier, B., Goobar-Larsson L., Sokolowski M. and Schwartz S.** (1998). Translational inhibition in vitro of human papillomavirus type 16 L2 mRNA mediated through interaction with heterogenous ribonucleoprotein K and poly(rC)-binding proteins 1 and 2. *J Biol Chem*, **273**(35): 22648-56.
- Colwill, K., Feng L. L., Yeakley J. M., Gish G. D., Caceres J. F., Pawson T. and Fu X. D.** (1996). SRPK1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors. *J Biol Chem*, **271**(40): 24569-75.
- Cooper, T. A.** (1998). Muscle-specific splicing of a heterologous exon mediated by a single muscle-specific splicing enhancer from the cardiac troponin T gene. *Mol Cell Biol*, **18**(8): 4519-25.
- Cooper, T. A. and Ordahl C. P.** (1985). A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternate splicing. *J Biol Chem*, **260**(20): 11140-8.
- Cooper, T. A. and Ordahl C. P.** (1984). A single troponin T gene regulated by different programs in cardiac and skeletal muscle development. *Science*, **226**(4677): 979-82.
- Cramer, P., Caceres J. F., Cazalla D., Kadener S., Muro A. F., Baralle F. E. and Kornblihtt A. R.** (1999). Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell*, **4**(2): 251-8.
- Cramer, P., Pesce C. G., Baralle F. E. and Kornblihtt A. R.** (1997). Functional association between promoter structure and transcript alternative splicing. *Proc Natl Acad Sci U S A*, **94**(21): 11456-60.
- Crawford, J. B. and Patton J. G.** (2006). Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol Cell Biol*, **26**(23): 8791-802.
- Crispino, J. D., Blencowe B. J. and Sharp P. A.** (1994). Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science*, **265**(5180): 1866-9.
- Cummings, C. J. and Zoghbi H. Y.** (2000). Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum Mol Genet*, **9**(6): 909-16.

- Dallaire, F., Dupuis S., Fiset S. and Chabot B.** (2000). Heterogeneous nuclear ribonucleoprotein A1 and UP1 protect mammalian telomeric repeats and modulate telomere replication in vitro. *J Biol Chem*, **275**(19): 14509-16.
- Damgaard, C. K., Kahns S., Lykke-Andersen S., Nielsen A. L., Jensen T. H. and Kjems J.** (2008). A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell*, **29**(2): 271-8.
- Dansithong, W., Jog S. P., Paul S., Mohammadzadeh R., Tring S., Kwok Y., Fry R. C., Marjoram P., Comai L. and Reddy S.** (2011). RNA steady-state defects in myotonic dystrophy are linked to nuclear exclusion of SHARP. *EMBO Rep*, **12**(7): 735-42.
- Dansithong, W., Paul S., Comai L. and Reddy S.** (2005). MBNL1 is the primary determinant of focus formation and aberrant insulin receptor splicing in DM1. *J Biol Chem*, **280**(7): 5773-80.
- Dansithong, W., Wolf C. M., Sarkar P., Paul S., Chiang A., Holt I., Morris G. E., Branco D., Sherwood M. C., Comai L., Berul C. I. and Reddy S.** (2008). Cytoplasmic CUG RNA foci are insufficient to elicit key DM1 features. *PLoS One*, **3**(12): e3968.
- Das, R., Dufu K., Romney B., Feldt M., Elenko M. and Reed R.** (2006). Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev*, **20**(9): 1100-9.
- Daughters, R. S., Tuttle D. L., Gao W., Ikeda Y., Moseley M. L., Ebner T. J., Swanson M. S. and Ranum L. P.** (2009). RNA gain-of-function in spinocerebellar ataxia type 8. *PLoS Genet*, **5**(8): e1000600.
- Dauksaite, V. and Akusjarvi G.** (2002). Human splicing factor ASF/SF2 encodes for a repressor domain required for its inhibitory activity on pre-mRNA splicing. *J Biol Chem*, **277**(15): 12579-86.
- Davies, R. C., Calvio C., Bratt E., Larsson S. H., Lamond A. I. and Hastie N. D.** (1998). WT1 interacts with the splicing factor U2AF65 in an isoform-dependent manner and can be incorporated into spliceosomes. *Genes Dev*, **12**(20): 3217-25.
- Davis, B. M., Mccurrach M. E., Taneja K. L., Singer R. H. and Housman D. E.** (1997). Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc Natl Acad Sci U S A*, **94**(14): 7388-93.
- Day, J. W. and Ranum L. P.** (2005). RNA pathogenesis of the myotonic dystrophies. *Neuromuscul Disord*, **15**(1): 5-16.
- Day, J. W., Ricker K., Jacobsen J. F., Rasmussen L. J., Dick K. A., Kress W., Schneider C., Koch M. C., Beilman G. J., Harrison A. R., Dalton J. C. and Ranum L. P.** (2003). Myotonic dystrophy type 2: molecular, diagnostic and clinical spectrum. *Neurology*, **60**(4): 657-64.
- Day, J. W., Roelofs R., Leroy B., Pech I., Benzow K. and Ranum L. P.** (1999). Clinical and genetic characteristics of a five-generation family with a novel form of myotonic dystrophy (DM2). *Neuromuscul Disord*, **9**(1): 19-27.
- De Die-Smulders, C. E., Howeler C. J., Thijs C., Mirandolle J. F., Anten H. B., Smeets H. J., Chandler K. E. and Geraedts J. P.** (1998). Age and causes of death in adult-onset myotonic dystrophy. *Brain*, **121** (Pt 8)(1557-63.
- De Die-Smulders, C. E., Smeets H. J., Loots W., Anten H. B., Mirandolle J. F., Geraedts J. P. and Howeler C. J.** (1997). Paternal transmission of congenital myotonic dystrophy. *J Med Genet*, **34**(11): 930-3.
- De Mezer, M., Wojciechowska M., Napierala M., Sobczak K. and Krzyzosiak W. J.** (2011). Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. *Nucleic Acids Res*, **39**(9): 3852-63.
- De Temmerman, N., Sermon K., Seneca S., De Rycke M., Hilven P., Lissens W., Van Steirteghem A. and Liebaers I.** (2004). Intergenerational instability of the expanded CTG

repeat in the DMPK gene: studies in human gametes and preimplantation embryos. *Am J Hum Genet*, **75**(2): 325-9.

Deiters, A. (2010). Small molecule modifiers of the microRNA and RNA interference pathway. *AAPS J*, **12**(1): 51-60.

Del Gatto-Konczak, F., Olive M., Gesnel M. C. and Breathnach R. (1999). hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. *Mol Cell Biol*, **19**(1): 251-60.

Delagoutte, E., Goellner G. M., Guo J., Baldacci G. and McMurray C. T. (2008). Single-stranded DNA-binding protein in vitro eliminates the orientation-dependent impediment to polymerase passage on CAG/CTG repeats. *J Biol Chem*, **283**(19): 13341-56.

Delaporte, C. (1998). Personality patterns in patients with myotonic dystrophy. *Arch Neurol*, **55**(5): 635-40.

Dere, R., Napierala M., Ranum L. P. and Wells R. D. (2004). Hairpin structure-forming propensity of the (CCTG.CAGG) tetranucleotide repeats contributes to the genetic instability associated with myotonic dystrophy type 2. *J Biol Chem*, **279**(40): 41715-26.

Dere, R. and Wells R. D. (2006). DM2 CCTG*CAGG repeats are crossover hotspots that are more prone to expansions than the DM1 CTG*CAG repeats in *Escherichia coli*. *J Mol Biol*, **360**(1): 21-36.

Detivaud, L., Pascreau G., Karaiskou A., Osborne H. B. and Kubiak J. Z. (2003). Regulation of EDEN-dependent deadenylation of Aurora A/Eg2-derived mRNA via phosphorylation and dephosphorylation in *Xenopus laevis* egg extracts. *J Cell Sci*, **116**(Pt 13): 2697-705.

Deutsch, M. and Long M. (1999). Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res*, **27**(15): 3219-28.

Dhaenens, C. M., Tran H., Frandemiche M. L., Carpentier C., Schraen-Maschke S., Sistiaga A., Goicoechea M., Eddarkaoui S., Van Brussels E., Obriot H., Labudeck A., Gevaert M. H., Fernandez-Gomez F., Charlet-Berguerand N., Deramecourt V., Maurage C. A., Buee L., De Munain A. L., Sablonniere B., Caillet-Boudin M. L. and Sergeant N. (2010). Mis-splicing of Tau exon 10 in myotonic dystrophy type 1 is reproduced by overexpression of CELF2 but not by MBNL1 silencing. *Biochim Biophys Acta*, **1812**(7): 732-42.

Dhami, P., Saffrey P., Bruce A. W., Dillon S. C., Chiang K., Bonhoure N., Koch C. M., Bye J., James K., Foad N. S., Ellis P., Watkins N. A., Ouwehand W. H., Langford C., Andrews R. M., Dunham I. and Vetrie D. (2010). Complex exon-intron marking by histone modifications is not determined solely by nucleosome distribution. *PLoS One*, **5**(8): e12339.

Dictenberg, J. B., Swanger S. A., Antar L. N., Singer R. H. and Bassell G. J. (2008). A direct role for FMRP in activity-dependent dendritic mRNA transport links filopodial-spine morphogenesis to fragile X syndrome. *Dev Cell*, **14**(6): 926-39.

Disney, M. D., Lee M. M., Pushechnikov A. and Childs-Disney J. L. (2010). The role of flexibility in the rational design of modularly assembled ligands targeting the RNAs that cause the myotonic dystrophies. *Chembiochem*, **11**(3): 375-82.

Dredge, B. K., Stefani G., Engelhard C. C. and Darnell R. B. (2005). Nova autoregulation reveals dual functions in neuronal splicing. *EMBO J*, **24**(8): 1608-20.

Dreyfuss, G. (1986). Structure and function of nuclear and cytoplasmic ribonucleoprotein particles. *Annu Rev Cell Biol*, **2**(459-98).

Dreyfuss, G., Kim V. N. and Kataoka N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol*, **3**(3): 195-205.

Dreyfuss, G., Matunis M. J., Pinol-Roma S. and Burd C. G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem*, **62**(289-321).

- Dreyfuss, G., Swanson M. S. and Pinol-Roma S.** (1988). Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci*, **13**(3): 86-91.
- Du, C., McGuffin M. E., Dauwalder B., Rabinow L. and Mattox W.** (1998). Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in *Drosophila*. *Mol Cell*, **2**(6): 741-50.
- Du, H., Cline M. S., Osborne R. J., Tuttle D. L., Clark T. A., Donohue J. P., Hall M. P., Shiue L., Swanson M. S., Thornton C. A. and Ares M., Jr.** (2010). Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat Struct Mol Biol*, **17**(2): 187-93.
- Du, L. and Warren S. L.** (1997). A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. *J Cell Biol*, **136**(1): 5-18.
- Dulhunty, A. F., Beard N. A., Pouliquin P. and Kimura T.** (2006). Novel regulators of RyR Ca²⁺ release channels: insight into molecular changes in genetically-linked myopathies. *J Muscle Res Cell Motil*, **27**(5-7): 351-65.
- Dye, M. J., Gromak N. and Proudfoot N. J.** (2006). Exon tethering in transcription by RNA polymerase II. *Mol Cell*, **21**(6): 849-59.
- Eisenberg, I., Alexander M. S. and Kunkel L. M.** (2009). miRNAs in normal and diseased skeletal muscle. *J Cell Mol Med*, **13**(1): 2-11.
- Emili, A., Shales M., Mccracken S., Xie W., Tucker P. W., Kobayashi R., Blencowe B. J. and Ingles C. J.** (2002). Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. *RNA*, **8**(9): 1102-11.
- Eperon, I. C., Makarova O. V., Mayeda A., Munroe S. H., Caceres J. F., Hayward D. G. and Krainer A. R.** (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): 8303-18.
- Epstein, H. F. and Fischman D. A.** (1991). Molecular analysis of protein assembly in muscle development. *Science*, **251**(4997): 1039-44.
- Estes, P. S., O'shea M., Clasen S. and Zarnescu D. C.** (2008). Fragile X protein controls the efficacy of mRNA transport in *Drosophila* neurons. *Mol Cell Neurosci*, **39**(2): 170-9.
- Expert-Bezancon, A., Le Caer J. P. and Marie J.** (2002). Heterogeneous nuclear ribonucleoprotein (hnRNP) K is a component of an intronic splicing enhancer complex that activates the splicing of the alternative exon 6A from chicken beta-tropomyosin pre-mRNA. *J Biol Chem*, **277**(19): 16614-23.
- Fardaei, M., Larkin K., Brook J. D. and Hamshere M. G.** (2001). In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res*, **29**(13): 2766-71.
- Fardaei, M., Rogers M. T., Thorpe H. M., Larkin K., Hamshere M. G., Harper P. S. and Brook J. D.** (2002). Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum Mol Genet*, **11**(7): 805-14.
- Farrell, B. T. and Lahue R. S.** (2006). CAG*CTG repeat instability in cultured human astrocytes. *Nucleic Acids Res*, **34**(16): 4495-505.
- Faustino, N. A. and Cooper T. A.** (2005). Identification of putative new splicing targets for ETR-3 using sequences identified by systematic evolution of ligands by exponential enrichment. *Mol Cell Biol*, **25**(3): 879-87.
- Faustino, N. A. and Cooper T. A.** (2003). Pre-mRNA splicing and human disease. *Genes Dev*, **17**(4): 419-37.
- Fededa, J. P. and Kornblihtt A. R.** (2008). A splicing regulator promotes transcriptional elongation. *Nat Struct Mol Biol*, **15**(8): 779-81.

- Fernandez-Costa, J. M. and Artero R.** (2010). A conserved motif controls nuclear localization of Drosophila Muscleblind. *Mol Cells*, **30**(1): 65-70.
- Filippova, G. N., Thienes C. P., Penn B. H., Cho D. H., Hu Y. J., Moore J. M., Klesert T. R., Lobanenko V. V. and Tapscott S. J.** (2001). CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat Genet*, **28**(4): 335-43.
- Fong, Y. W. and Zhou Q.** (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, **414**(6866): 929-33.
- Fouche, N., Ozgur S., Roy D. and Griffith J. D.** (2006). Replication fork regression in repetitive DNAs. *Nucleic Acids Res*, **34**(20): 6044-50.
- Francois, V., Klein A. F., Beley C., Jollet A., Lemercier C., Garcia L. and Furling D.** (2011). Selective silencing of mutated mRNAs in DM1 by using modified hU7-snrRNAs. *Nat Struct Mol Biol*, **18**(1): 85-7.
- Freudenreich, C. H., Stavenhagen J. B. and Zakian V. A.** (1997). Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol Cell Biol*, **17**(4): 2090-8.
- Fu, X. D.** (1993). Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature*, **365**(6441): 82-5.
- Fu, Y. H., Pizzuti A., Fenwick R. G., Jr., King J., Rajnarayan S., Dunne P. W., Dubel J., Nasser G. A., Ashizawa T., De Jong P. and Et Al.** (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science*, **255**(5049): 1256-8.
- Fugier, C., Klein A. F., Hammer C., Vassilopoulos S., Ivarsson Y., Toussaint A., Tosch V., Vignaud A., Ferry A., Messaddeq N., Kokunai Y., Tsuburaya R., De La Grange P., Dembele D., Francois V., Precigout G., Boulade-Ladame C., Hummel M. C., De Munain A. L., Sergeant N., Laquerriere A., Thibault C., Deryckere F., Auboeuf D., Garcia L., Zimmermann P., Udd B., Schoser B., Takahashi M. P., Nishino I., Bassez G., Laporte J., Furling D. and Charlet-Berguerand N.** (2011). Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. *Nat Med*, **17**(6): 720-5.
- Fujimura, K., Kano F. and Murata M.** (2008). Dual localization of the RNA binding protein CUGBP-1 to stress granule and perinucleolar compartment. *Exp Cell Res*, **314**(3): 543-53.
- Furger, A., O'sullivan J. M., Binnie A., Lee B. A. and Proudfoot N. J.** (2002). Promoter proximal splice sites enhance transcription. *Genes Dev*, **16**(21): 2792-9.
- Furling, D., Coiffier L., Mouly V., Barbet J. P., St Guily J. L., Taneja K., Gourdon G., Junien C. and Butler-Browne G. S.** (2001a). Defective satellite cells in congenital myotonic dystrophy. *Hum Mol Genet*, **10**(19): 2079-87.
- Furling, D., Lemieux D., Taneja K. and Puymirat J.** (2001b). Decreased levels of myotonic dystrophy protein kinase (DMPK) and delayed differentiation in human myotonic dystrophy myoblasts. *Neuromuscul Disord*, **11**(8): 728-35.
- Gacy, A. M., Goellner G., Juranic N., Macura S. and McMurray C. T.** (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell*, **81**(4): 533-40.
- Gacy, A. M. and McMurray C. T.** (1998). Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. *Biochemistry*, **37**(26): 9426-34.
- Gallego, M. E., Sirand-Pugnet P., Durosay P., Clouet D'orval B., D'aubenton-Carafa Y., Brody E., Expert-Bezancon A. and Marie J.** (1996). Tissue-specific splicing of two mutually exclusive exons of the chicken beta-tropomyosin pre-mRNA: positive and negative regulations. *Biochimie*, **78**(6): 457-65.

- Gambardella, S., Rinaldi F., Lepore S. M., Viola A., Loro E., Angelini C., Vergani L., Novelli G. and Botta A.** (2010). Overexpression of microRNA-206 in the skeletal muscle from myotonic dystrophy type 1 patients. *J Transl Med*, **8**(48).
- Garcia-Arocena, D. and Hagerman P. J.** (2011). Advances in understanding the molecular basis of FXTAS. *Hum Mol Genet*, **19**(R1): R83-9.
- Garneau, D., Revil T., Fiset J. F. and Chabot B.** (2005). Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J Biol Chem*, **280**(24): 22641-50.
- Garrott, H. M., Walland M. J. and O'day J.** (2004). Recurrent posterior capsular opacification and capsulorhexis contracture after cataract surgery in myotonic dystrophy. *Clin Experiment Ophthalmol*, **32**(6): 653-5.
- Gatchel, J. R. and Zoghbi H. Y.** (2005). Diseases of unstable repeat expansion: mechanisms and common principles. *Nat Rev Genet*, **6**(10): 743-55.
- Gates, D. P., Coonrod L. A. and Berglund J. A.** (2011). Auto-regulated splicing of the muscleblind-like 1 (MBNL1) pre-mRNA. *J Biol Chem*.
- Gattoni, R., Mahe D., Mahl P., Fischer N., Mattei M. G., Stevenin J. and Fuchs J. P.** (1996). The human hnRNP-M proteins: structure and relation with early heat shock-induced splicing arrest and chromosome mapping. *Nucleic Acids Res*, **24**(13): 2535-42.
- Gecz, J., Oostra B. A., Hockey A., Carbonell P., Turner G., Haan E. A., Sutherland G. R. and Mulley J. C.** (1997). FMR2 expression in families with FRAXE mental retardation. *Hum Mol Genet*, **6**(3): 435-41.
- Gennarelli, M., Pavoni M., Amicucci P., Angelini C., Menegazzo E., Zelano G., Novelli G. and Dallapiccola B.** (1999). Reduction of the DM-associated homeo domain protein (DMAHP) mRNA in different brain areas of myotonic dystrophy patients. *Neuromuscul Disord*, **9**(4): 215-9.
- Gerbasi, V. R. and Link A. J.** (2007). The myotonic dystrophy type 2 protein ZNF9 is part of an ITAF complex that promotes cap-independent translation. *Mol Cell Proteomics*, **6**(6): 1049-58.
- Gharehbaghi-Schnell, E. B., Finsterer J., Korschineck I., Mamoli B. and Binder B. R.** (1998). Genotype-phenotype correlation in myotonic dystrophy. *Clin Genet*, **53**(1): 20-6.
- Goedert, M., Spillantini M. G., Jakes R., Rutherford D. and Crowther R. A.** (1989). Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron*, **3**(4): 519-26.
- Goers, E. S., Purcell J., Voelker R. B., Gates D. P. and Berglund J. A.** (2010). MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res*, **38**(7): 2467-84.
- Goers, E. S., Voelker R. B., Gates D. P. and Berglund J. A.** (2008). RNA binding specificity of Drosophila muscleblind. *Biochemistry*, **47**(27): 7284-94.
- Goldstrohm, A. C., Greenleaf A. L. and Garcia-Blanco M. A.** (2001). Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene*, **277**(1-2): 31-47.
- Gomes-Pereira, M., Foiry L., Nicole A., Huguet A., Junien C., Munnich A. and Gourdon G.** (2007). CTG trinucleotide repeat "big jumps": large expansions, small mice. *PLoS Genet*, **3**(4): e52.
- Gomes, A. V., Barnes J. A., Harada K. and Potter J. D.** (2004). Role of troponin T in disease. *Mol Cell Biochem*, **263**(1-2): 115-29.
- Gomes, A. V., Guzman G., Zhao J. and Potter J. D.** (2002). Cardiac troponin T isoforms affect the Ca²⁺ sensitivity and inhibition of force development. Insights into the role of troponin T isoforms in the heart. *J Biol Chem*, **277**(38): 35341-9.

- Goncalves, V., Matos P. and Jordan P.** (2009). Antagonistic SR proteins regulate alternative splicing of tumor-related Rac1b downstream of the PI3-kinase and Wnt pathways. *Hum Mol Genet*, **18**(19): 3696-707.
- Goo, Y. H. and Cooper T. A.** (2009). CUGBP2 directly interacts with U2 17S snRNP components and promotes U2 snRNA binding to cardiac troponin T pre-mRNA. *Nucleic Acids Res*, **37**(13): 4275-86.
- Good, P. J., Chen Q., Warner S. J. and Herring D. C.** (2000). A family of human RNA-binding proteins related to the Drosophila Bruno translational regulator. *J Biol Chem*, **275**(37): 28583-92.
- Goula, A. V., Berquist B. R., Wilson D. M., 3rd, Wheeler V. C., Trottier Y. and Merienne K.** (2009). Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. *PLoS Genet*, **5**(12): e1000749.
- Grabczyk, E. and Usdin K.** (2000). The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res*, **28**(14): 2815-22.
- Grammatikakis, I., Goo Y. H., Echeverria G. V. and Cooper T. A.** (2010). Identification of MBNL1 and MBNL3 domains required for splicing activation and repression. *Nucleic Acids Res*, **39**(7): 2769-80.
- Graveley, B. R.** (2000). Sorting out the complexity of SR protein functions. *Rna*, **6**(9): 1197-211.
- Greco, S., De Simone M., Colussi C., Zaccagnini G., Fasanaro P., Pescatori M., Cardani R., Perbellini R., Isaia E., Sale P., Meola G., Capogrossi M. C., Gaetano C. and Martelli F.** (2009). Common micro-RNA signature in skeletal muscle damage and regeneration induced by Duchenne muscular dystrophy and acute ischemia. *FASEB J*, **23**(10): 3335-46.
- Groenen, P. J., Wansink D. G., Coerwinkel M., Van Den Broek W., Jansen G. and Wieringa B.** (2000). Constitutive and regulated modes of splicing produce six major myotonic dystrophy protein kinase (DMPK) isoforms with distinct properties. *Hum Mol Genet*, **9**(4): 605-16.
- Gu, Y., Shen Y., Gibbs R. A. and Nelson D. L.** (1996). Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island. *Nat Genet*, **13**(1): 109-13.
- Gui, J. F., Tronchere H., Chandler S. D. and Fu X. D.** (1994). Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. *Proc Natl Acad Sci U S A*, **91**(23): 10824-8.
- Gunderson, F. Q. and Johnson T. L.** (2009). Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly. *PLoS Genet*, **5**(10): e1000682.
- Hamshire, M. G., Harley H., Harper P., Brook J. D. and Brookfield J. F.** (1999). Myotonic dystrophy: the correlation of (CTG) repeat length in leucocytes with age at onset is significant only for patients with small expansions. *J Med Genet*, **36**(1): 59-61.
- Han, J. and Cooper T. A.** (2005). Identification of CELF splicing activation and repression domains in vivo. *Nucleic Acids Res*, **33**(9): 2769-80.
- Han, K., Yeo G., An P., Burge C. B. and Grabowski P. J.** (2005). A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol*, **3**(5): e158.
- Harper, P. S.** (1975). Congenital myotonic dystrophy in Britain. I. Clinical aspects. *Arch Dis Child*, **50**(7): 505-13.
- Harper, P. S.** (1989). Postoperative complications in myotonic dystrophy. *Lancet*, **2**(8674): 1269.
- Hartshorne, D. J.** (1998). Myosin phosphatase: subunits and interactions. *Acta Physiol Scand*, **164**(4): 483-93.

- Hashimoto, C. and Steitz J. A.** (1984). U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res*, **12**(7): 3283-93.
- Hastings, M. L. and Krainer A. R.** (2001). Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol*, **13**(3): 302-9.
- He, F., Dang W., Abe C., Tsuda K., Inoue M., Watanabe S., Kobayashi N., Kigawa T., Matsuda T., Yabuki T., Aoki M., Seki E., Harada T., Tomabechi Y., Terada T., Shirouzu M., Tanaka A., Guntert P., Muto Y. and Yokoyama S.** (2009). Solution structure of the RNA binding domain in the human muscleblind-like protein 2. *Protein Sci*, **18**(1): 80-91.
- He, Y. and Smith R.** (2009). Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell Mol Life Sci*, **66**(7): 1239-56.
- Hiller, M., Zhang Z., Backofen R. and Stamm S.** (2007). Pre-mRNA secondary structures influence exon recognition. *PLoS Genet*, **3**(11): e204.
- Hirose, Y., Tacke R. and Manley J. L.** (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev*, **13**(10): 1234-9.
- Ho, T. H., Bundman D., Armstrong D. L. and Cooper T. A.** (2005a). Transgenic mice expressing CUG-BP1 reproduce splicing mis-regulation observed in myotonic dystrophy. *Hum Mol Genet*, **14**(11): 1539-47.
- Ho, T. H., Charlet B. N., Poulos M. G., Singh G., Swanson M. S. and Cooper T. A.** (2004). Muscleblind proteins regulate alternative splicing. *Embo J*, **23**(15): 3103-12.
- Ho, T. H., Savkur R. S., Poulos M. G., Mancini M. A., Swanson M. S. and Cooper T. A.** (2005b). Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy. *J Cell Sci*, **118**(Pt 13): 2923-33.
- Hodges, C., Bintu L., Lubkowska L., Kashlev M. and Bustamante C.** (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science*, **325**(5940): 626-8.
- Hofmann, Y. and Wirth B.** (2002). hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1. *Hum Mol Genet*, **11**(17): 2037-49.
- Holmes, S. E., O'hearn E., Rosenblatt A., Callahan C., Hwang H. S., Ingersoll-Ashworth R. G., Fleisher A., Stevanin G., Brice A., Potter N. T., Ross C. A. and Margolis R. L.** (2001). A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. *Nat Genet*, **29**(4): 377-8.
- Holt, I., Mittal S., Furling D., Butler-Browne G. S., Brook J. D. and Morris G. E.** (2007). Defective mRNA in myotonic dystrophy accumulates at the periphery of nuclear splicing speckles. *Genes Cells*, **12**(9): 1035-48.
- Honore, B., Vorum H. and Baandrup U.** (1999). hnRNPs H, H' and F behave differently with respect to posttranslational cleavage and subcellular localization. *FEBS Lett*, **456**(2): 274-80.
- Houseley, J. M., Wang Z., Brock G. J., Soloway J., Artero R., Perez-Alonso M., O'dell K. M. and Monckton D. G.** (2005). Myotonic dystrophy associated expanded CUG repeat muscleblind positive ribonuclear foci are not toxic to Drosophila. *Hum Mol Genet*, **14**(6): 873-83.
- Hsu, R. J., Hsiao K. M., Lin M. J., Li C. Y., Wang L. C., Chen L. K. and Pan H.** (2011). Long tract of untranslated CAG repeats is deleterious in transgenic mice. *PLoS One*, **6**(1): e16417.
- Huang, Y., Gattoni R., Stevenin J. and Steitz J. A.** (2003). SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell*, **11**(3): 837-43.
- Huang, Y. and Steitz J. A.** (2001). Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol Cell*, **7**(4): 899-905.

- Huichalaf, C., Sakai K., Jin B., Jones K., Wang G. L., Schoser B., Schneider-Gold C., Sarkar P., Pereira-Smith O. M., Timchenko N. and Timchenko L.** (2010). Expansion of CUG RNA repeats causes stress and inhibition of translation in myotonic dystrophy 1 (DM1) cells. *FASEB J*, **24**(10): 3706-19.
- Huichalaf, C., Schoser B., Schneider-Gold C., Jin B., Sarkar P. and Timchenko L.** (2009). Reduction of the rate of protein translation in patients with myotonic dystrophy 2. *J Neurosci*, **29**(28): 9042-9.
- Iakova, P., Wang G. L., Timchenko L., Michalak M., Pereira-Smith O. M., Smith J. R. and Timchenko N. A.** (2004). Competition of CUGBP1 and calreticulin for the regulation of p21 translation determines cell fate. *EMBO J*, **23**(2): 406-17.
- Iwahashi, C. K., Yasui D. H., An H. J., Greco C. M., Tassone F., Nannen K., Babineau B., Lebrilla C. B., Hagerman R. J. and Hagerman P. J.** (2006). Protein composition of the intranuclear inclusions of FXTAS. *Brain*, **129**(Pt 1): 256-71.
- Jady, B. E. and Kiss T.** (2001). A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J*, **20**(3): 541-51.
- Jakupciak, J. P. and Wells R. D.** (1999). Genetic instabilities in (CTG.CAG) repeats occur by recombination. *J Biol Chem*, **274**(33): 23468-79.
- Jansen, G., Bartolomei M., Kalscheuer V., Merckx G., Wormskamp N., Mariman E., Smeets D., Ropers H. H. and Wieringa B.** (1993). No imprinting involved in the expression of DM-kinase mRNAs in mouse and human tissues. *Hum Mol Genet*, **2**(8): 1221-7.
- Jansen, G., Groenen P. J., Bachner D., Jap P. H., Coerwinkel M., Oerlemans F., Van Den Broek W., Gohlsch B., Pette D., Plomp J. J., Molenaar P. C., Nederhoff M. G., Van Echteld C. J., Dekker M., Berns A., Hameister H. and Wieringa B.** (1996). Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. *Nat Genet*, **13**(3): 316-24.
- Jasinska, A., Michlewski G., De Mezer M., Sobczak K., Kozlowski P., Napierala M. and Krzyzosiak W. J.** (2003). Structures of trinucleotide repeats in human transcripts and their functional implications. *Nucleic Acids Res*, **31**(19): 5463-8.
- Jensen, K. B., Dredge B. K., Stefani G., Zhong R., Buckanovich R. J., Okano H. J., Yang Y. Y. and Darnell R. B.** (2000a). Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron*, **25**(2): 359-71.
- Jensen, K. B., Musunuru K., Lewis H. A., Burley S. K. and Darnell R. B.** (2000b). The tetranucleotide UCAAY directs the specific recognition of RNA by the Nova K-homology 3 domain. *Proc Natl Acad Sci U S A*, **97**(11): 5740-5.
- Jiang, H., Mankodi A., Swanson M. S., Moxley R. T. and Thornton C. A.** (2004). Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum Mol Genet*, **13**(24): 3079-88.
- Jiang, H., Nucifora F. C., Jr., Ross C. A. and Defranco D. B.** (2003). Cell death triggered by polyglutamine-expanded huntingtin in a neuronal cell line is associated with degradation of CREB-binding protein. *Hum Mol Genet*, **12**(1): 1-12.
- Jin, J. P. and Lin J. J.** (1989). Isolation and characterization of cDNA clones encoding embryonic and adult isoforms of rat cardiac troponin T. *J Biol Chem*, **264**(24): 14471-7.
- Joseph, J. T., Richards C. S., Anthony D. C., Upton M., Perez-Atayde A. R. and Greenstein P.** (1997). Congenital myotonic dystrophy pathology and somatic mosaicism. *Neurology*, **49**(5): 1457-60.
- Jung, J. and Bonini N.** (2007). CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. *Science*, **315**(5820): 1857-9.
- Junghans, R. P.** (2009). Dystrophia myotonica: why focus on foci? *Eur J Hum Genet*, **17**(5): 543-53.

- Jurica, M. S., Licklider L. J., Gygi S. R., Grigorieff N. and Moore M. J.** (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*, **8**(4): 426-39.
- Jurica, M. S. and Moore M. J.** (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell*, **12**(1): 5-14.
- Kaibuchi, K., Kuroda S. and Amano M.** (1999). Regulation of the cytoskeleton and cell adhesion by the Rho family GTPases in mammalian cells. *Annu Rev Biochem*, **68**(459-86).
- Kalsotra, A., Xiao X., Ward A. J., Castle J. C., Johnson J. M., Burge C. B. and Cooper T. A.** (2008). A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc Natl Acad Sci U S A*, **105**(51): 20333-8.
- Kameoka, S., Duque P. and Konarska M. M.** (2004). p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. *EMBO J*, **23**(8): 1782-91.
- Kanadia, R. N., Johnstone K. A., Mankodi A., Lungu C., Thornton C. A., Esson D., Timmers A. M., Hauswirth W. W. and Swanson M. S.** (2003a). A muscleblind knockout model for myotonic dystrophy. *Science*, **302**(5652): 1978-80.
- Kanadia, R. N., Shin J., Yuan Y., Beattie S. G., Wheeler T. M., Thornton C. A. and Swanson M. S.** (2006). Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. *Proc Natl Acad Sci U S A*, **103**(31): 11748-53.
- Kanadia, R. N., Urbinati C. R., Crusselle V. J., Luo D., Lee Y. J., Harrison J. K., Oh S. P. and Swanson M. S.** (2003b). Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3. *Gene Expr Patterns*, **3**(4): 459-62.
- Kanopka, A., Muhlemann O. and Akusjarvi G.** (1996). Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature*, **381**(6582): 535-8.
- Karthikeyan, G., Santos J. H., Graziewicz M. A., Copeland W. C., Isaya G., Van Houten B. and Resnick M. A.** (2003). Reduction in frataxin causes progressive accumulation of mitochondrial damage. *Hum Mol Genet*, **12**(24): 3331-42.
- Kay, B. K., Williamson M. P. and Sudol M.** (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J*, **14**(2): 231-41.
- Kessler, M. M., Henry M. F., Shen E., Zhao J., Gross S., Silver P. A. and Moore C. L.** (1997). Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. *Genes Dev*, **11**(19): 2545-56.
- Kilizek, A., Kierzek R., Krzyzosiak W. J. and Rypniewski W.** (2010). Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucleic Acids Res*, **38**(22): 8370-6.
- Kilizek, A., Kierzek R., Krzyzosiak W. J. and Rypniewski W.** (2011). Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic Acids Res*.
- Kim, M. H.** (2008). Protein phosphatase 1 activation and alternative splicing of Bcl-X and Mcl-1 by EGCG + ibuprofen. *J Cell Biochem*, **104**(4): 1491-9.
- Kim, S., Park G. H. and Paik W. K.** (1998). Recent advances in protein methylation: enzymatic methylation of nucleic acid binding proteins. *Amino Acids*, **15**(4): 291-306.
- Kimura, T., Nakamori M., Lueck J. D., Pouliquin P., Aoike F., Fujimura H., Dirksen R. T., Takahashi M. P., Dulhunty A. F. and Sakoda S.** (2005). Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca²⁺-ATPase in myotonic dystrophy type 1. *Hum Mol Genet*, **14**(15): 2189-200.
- Kino, Y., Mori D., Oma Y., Takeshita Y., Sasagawa N. and Ishiura S.** (2004). Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum Mol Genet*, **13**(5): 495-507.

- Kiss, T.** (2004). Biogenesis of small nuclear RNPs. *J Cell Sci*, **117**(Pt 25): 5949-51.
- Klesert, T. R., Cho D. H., Clark J. I., Maylie J., Adelman J., Snider L., Yuen E. C., Soriano P. and Tapscott S. J.** (2000). Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy. *Nat Genet*, **25**(1): 105-9.
- Kohtz, J. D., Jamison S. F., Will C. L., Zuo P., Luhrmann R., Garcia-Blanco M. A. and Manley J. L.** (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, **368**(6467): 119-24.
- Kolasinska-Zwierz, P., Down T., Latorre I., Liu T., Liu X. S. and Ahringer J.** (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet*, **41**(3): 376-81.
- Koob, M. D., Moseley M. L., Schut L. J., Benzow K. A., Bird T. D., Day J. W. and Ranum L. P.** (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat Genet*, **21**(4): 379-84.
- Kornblihtt, A. R.** (2005). Promoter usage and alternative splicing. *Curr Opin Cell Biol*, **17**(3): 262-8.
- Kornblihtt, A. R., Schor I. E., Allo M. and Blencowe B. J.** (2009). When chromatin meets splicing. *Nat Struct Mol Biol*, **16**(9): 902-3.
- Koshelev, M., Sarma S., Price R. E., Wehrens X. H. and Cooper T. A.** (2010). Heart-specific overexpression of CUGBP1 reproduces functional and molecular abnormalities of myotonic dystrophy type 1. *Hum Mol Genet*, **19**(6): 1066-75.
- Kosik, K. S., Orecchio L. D., Bakalis S. and Neve R. L.** (1989). Developmentally regulated expression of specific tau sequences. *Neuron*, **2**(4): 1389-97.
- Kovtun, I. V., Liu Y., Bjoras M., Klungland A., Wilson S. H. and McMurray C. T.** (2007). OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature*, **447**(7143): 447-52.
- Kovtun, I. V. and McMurray C. T.** (2008). Features of trinucleotide repeat instability in vivo. *Cell Res*, **18**(1): 198-213.
- Krecic, A. M. and Swanson M. S.** (1999). hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol*, **11**(3): 363-71.
- Krol, A., Branlant C., Lazar E., Gallinaro H. and Jacob M.** (1981). Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs. *Nucleic Acids Res*, **9**(12): 2699-716.
- Krol, J., Fiszer A., Mykowska A., Sobczak K., De Mezer M. and Krzyzosiak W. J.** (2007). Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Mol Cell*, **25**(4): 575-86.
- Kunkel, G. R., Maser R. L., Calvet J. P. and Pederson T.** (1986). U6 small nuclear RNA is transcribed by RNA polymerase III. *Proc Natl Acad Sci U S A*, **83**(22): 8575-9.
- Kuyumcu-Martinez, N. M., Wang G. S. and Cooper T. A.** (2007). Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. *Mol Cell*, **28**(1): 68-78.
- Kwek, K. Y., Murphy S., Furger A., Thomas B., O'gorman W., Kimura H., Proudfoot N. J. and Akoulitchev A.** (2002). U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nat Struct Biol*, **9**(11): 800-5.
- La Spada, A. R., Wilson E. M., Lubahn D. B., Harding A. E. and Fischbeck K. H.** (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*, **352**(6330): 77-9.
- Labranche, H., Dupuis S., Ben-David Y., Bani M. R., Wellinger R. J. and Chabot B.** (1998). Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat Genet*, **19**(2): 199-202.

- Ladd, A. N., Charlet N. and Cooper T. A. (2001). The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol Cell Biol*, **21**(4): 1285-96.
- Ladd, A. N. and Cooper T. A. (2004). Multiple domains control the subcellular localization and activity of ETR-3, a regulator of nuclear and cytoplasmic RNA processing events. *J Cell Sci*, **117**(Pt 16): 3519-29.
- Ladd, A. N., Nguyen N. H., Malhotra K. and Cooper T. A. (2004). CELF6, a member of the CELF family of RNA-binding proteins, regulates muscle-specific splicing enhancer-dependent alternative splicing. *J Biol Chem*, **279**(17): 17756-64.
- Ladd, A. N., Taffet G., Hartley C., Kearney D. L. and Cooper T. A. (2005). Cardiac tissue-specific repression of CELF activity disrupts alternative splicing and causes cardiomyopathy. *Mol Cell Biol*, **25**(14): 6267-78.
- Lam, S. L., Wu F., Yang H. and Chi L. M. (2011). The origin of genetic instability in CCTG repeats. *Nucleic Acids Res*, **39**(14): 6260-8.
- Lamond, A. I. and Spector D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol*, **4**(8): 605-12.
- Lander, E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J., Devon K., Dewar K., Doyle M., Fitzhugh W., Funke R., Gage D., Harris K., Heaford A., Howland J., Kann L., Lehoczky J., Levine R., Mcewan P., Mckernan K., Meldrim J., Mesirov J. P., Miranda C., Morris W., Naylor J., Raymond C., Rosetti M., Santos R., Sheridan A., Sougnez C., Stange-Thomann N., Stojanovic N., Subramanian A., Wyman D., Rogers J., Sulston J., Ainscough R., Beck S., Bentley D., Burton J., Clee C., Carter N., Coulson A., Deadman R., Deloukas P., Dunham A., Dunham I., Durbin R., French L., Grafham D., Gregory S., Hubbard T., Humphray S., Hunt A., Jones M., Lloyd C., McMurray A., Matthews L., Mercer S., Milne S., Mullikin J. C., Mungall A., Plumb R., Ross M., Shownkeen R., Sims S., Waterston R. H., Wilson R. K., Hillier L. W., Mcpherson J. D., Marra M. A., Mardis E. R., Fulton L. A., Chinwalla A. T., Pepin K. H., Gish W. R., Chisoe S. L., Wendl M. C., Delehaunty K. D., Miner T. L., Delehaunty A., Kramer J. B., Cook L. L., Fulton R. S., Johnson D. L., Minx P. J., Clifton S. W., Hawkins T., Branscomb E., Predki P., Richardson P., Wenning S., Slezak T., Doggett N., Cheng J. F., Olsen A., Lucas S., Elkin C., Uberbacher E., Frazier M., Gibbs R. A., Muzny D. M., Scherer S. E., Bouck J. B., Sodergren E. J., Worley K. C., Rives C. M., Gorrell J. H., Metzker M. L., Naylor S. L., Kucherlapati R. S., Nelson D. L., Weinstock G. M., Sakaki Y., Fujiyama A., Hattori M., Yada T., Toyoda A., Itoh T., Kawagoe C., Watanabe H., Totoki Y., Taylor T., Weissenbach J., Heilig R., Saurin W., Artiguenave F., Brottier P., Bruls T., Pelletier E., Robert C., Wincker P., Smith D. R., Doucette-Stamm L., Rubenfield M., Weinstock K., Lee H. M., Dubois J., Rosenthal A., Platzer M., Nyakatura G., Taudien S., Rump A., Yang H., Yu J., Wang J., Huang G., Gu J., Hood L., Rowen L., Madan A., Qin S., Davis R. W., Federspiel N. A., Abola A. P., Proctor M. J., Myers R. M., Schmutz J., Dickson M., Grimwood J., Cox D. R., Olson M. V., Kaul R., Shimizu N., Kawasaki K., Minoshima S., Evans G. A., Athanasiou M., Schultz R., Roe B. A., Chen F., Pan H., Ramser J., Lehrach H., Reinhardt R., McCombie W. R., De La Bastide M., Dedhia N., Blocker H., Hornischer K., Nordsiek G., Agarwala R., Aravind L., Bailey J. A., Bateman A., Batzoglou S., Birney E., Bork P., Brown D. G., Burge C. B., Cerutti L., Chen H. C., Church D., Clamp M., Copley R. R., Doerks T., Eddy S. R., Eichler E. E., Furey T. S., Galagan J., Gilbert J. G., Harmon C., Hayashizaki Y., Haussler D., Hermjakob H., Hokamp K., Jang W., Johnson L. S., Jones T. A., Kasif S., Kasprzyk A., Kennedy S., Kent W. J., Kitts P., Koonin E. V., Korf I., Kulp D., Lancet D., Lowe T. M., Mclysaght A., Mikkelsen T., Moran J. V., Mulder N., Pollara V. J., Ponting C. P., Schuler G., Schultz J., Slater G., Smit A. F., Stupka E., Szustakowski J.,

- Thierry-Mieg D., Thierry-Mieg J., Wagner L., Wallis J., Wheeler R., Williams A., Wolf Y. I., Wolfe K. H., Yang S. P., Yeh R. F., Collins F., Guyer M. S., Peterson J., Felsenfeld A., Wetterstrand K. A., Patrinos A., Morgan M. J., De Jong P., Catanese J. J., Osoegawa K., Shizuya H., Choi S. and Chen Y. J.** (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**(6822): 860-921.
- Langlois, M. A., Boniface C., Wang G., Alluin J., Salvaterra P. M., Puymirat J., Rossi J. J. and Lee N. S.** (2005). Cytoplasmic and nuclear retained DMPK mRNAs are targets for RNA interference in myotonic dystrophy cells. *J Biol Chem*, **280**(17): 16949-54.
- Langlois, M. A., Lee N. S., Rossi J. J. and Puymirat J.** (2003). Hammerhead ribozyme-mediated destruction of nuclear foci in myotonic dystrophy myoblasts. *Mol Ther*, **7**(5 Pt 1): 670-80.
- Lear, A. L., Eperon L. P., Wheatley I. M. and Eperon I. C.** (1990). Hierarchy for 5' splice site preference determined in vivo. *J Mol Biol*, **211**(1): 103-15.
- Lee, J. E. and Cooper T. A.** (2009). Pathogenic mechanisms of myotonic dystrophy. *Biochem Soc Trans*, **37**(Pt 6): 1281-6.
- Lee, M. M., Pushechnikov A. and Disney M. D.** (2009). Rational and modular design of potent ligands targeting the RNA that causes myotonic dystrophy 2. *ACS Chem Biol*, **4**(5): 345-55.
- Lemaire, R., Winne A., Sarkissian M. and Lafyatis R.** (1999). SF2 and SRp55 regulation of CD45 exon 4 skipping during T cell activation. *Eur J Immunol*, **29**(3): 823-37.
- Leppert, J., Urbinati C. R., Hafner S., Ohlenschlager O., Swanson M. S., Gorlach M. and Ramachandran R.** (2004). Identification of NH...N hydrogen bonds by magic angle spinning solid state NMR in a double-stranded RNA associated with myotonic dystrophy. *Nucleic Acids Res*, **32**(3): 1177-83.
- Lerga, A., Hallier M., Delva L., Orvain C., Gallais I., Marie J. and Moreau-Gachelin F.** (2001). Identification of an RNA binding specificity for the potential splicing factor TLS. *J Biol Chem*, **276**(9): 6807-16.
- Leroy, O., Dhaenens C. M., Schraen-Maschke S., Belarbi K., Delacourte A., Andreadis A., Sablonniere B., Buee L., Sergeant N. and Caillet-Boudin M. L.** (2006). ETR-3 represses Tau exons 2/3 inclusion, a splicing event abnormally enhanced in myotonic dystrophy type I. *J Neurosci Res*, **84**(4): 852-9.
- Lewis, H. A., Musunuru K., Jensen K. B., Edo C., Chen H., Darnell R. B. and Burley S. K.** (2000). Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell*, **100**(3): 323-32.
- Li, D., Bachinski L. L. and Roberts R.** (2001). Genomic organization and isoform-specific tissue expression of human NAPOR (CUGBP2) as a candidate gene for familial arrhythmogenic right ventricular dysplasia. *Genomics*, **74**(3): 396-401.
- Li, J. W., Yuan D., Li H., Liang X. W., Lu W. H. and Gu Y. Q.** (2008). [Effect of (CAG) n polymorphism of androgen receptor gene on hormonal male contraception]. *Zhonghua Nan Ke Xue*, **14**(2): 126-30.
- Li, X. and Manley J. L.** (2006). Cotranscriptional processes and their influence on genome stability. *Genes Dev*, **20**(14): 1838-47.
- Lin, S., Coutinho-Mansfield G., Wang D., Pandit S. and Fu X. D.** (2008). The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol*, **15**(8): 819-26.
- Lin, X., Miller J. W., Mankodi A., Kanadia R. N., Yuan Y., Moxley R. T., Swanson M. S. and Thornton C. A.** (2006a). Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum Mol Genet*, **15**(13): 2087-97.
- Lin, Y., Dent S. Y., Wilson J. H., Wells R. D. and Napierala M.** (2010). R loops stimulate genetic instability of CTG.CAG repeats. *Proc Natl Acad Sci U S A*, **107**(2): 692-7.

- Lin, Y., Dion V. and Wilson J. H.** (2006b). Transcription promotes contraction of CAG repeat tracts in human cells. *Nat Struct Mol Biol*, **13**(2): 179-80.
- Lin, Y., Hubert L., Jr. and Wilson J. H.** (2009). Transcription destabilizes triplet repeats. *Mol Carcinog*, **48**(4): 350-61.
- Lindstrom, D. L., Squazzo S. L., Muster N., Burckin T. A., Wachter K. C., Emigh C. A., Mcclery J. A., Yates J. R., 3rd and Hartzog G. A.** (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol Cell Biol*, **23**(4): 1368-78.
- Liquori, C. L., Ikeda Y., Weatherspoon M., Ricker K., Schoser B. G., Dalton J. C., Day J. W. and Ranum L. P.** (2003). Myotonic dystrophy type 2: human founder haplotype and evolutionary conservation of the repeat tract. *Am J Hum Genet*, **73**(4): 849-62.
- Liquori, C. L., Ricker K., Moseley M. L., Jacobsen J. F., Kress W., Naylor S. L., Day J. W. and Ranum L. P.** (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science*, **293**(5531): 864-7.
- Liu, H. X., Zhang M. and Krainer A. R.** (1998). Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev*, **12**(13): 1998-2012.
- Logigian, E. L., Blood C. L., Dilek N., Martens W. B., Moxley R. T. T., Wiegner A. W., Thornton C. A. and Moxley R. T., 3rd.** (2005). Quantitative analysis of the "warm-up" phenomenon in myotonic dystrophy type 1. *Muscle Nerve*, **32**(1): 35-42.
- Long, C. S. and Ordahl C. P.** (1988). Transcriptional repression of an embryo-specific muscle gene. *Dev Biol*, **127**(1): 228-34.
- Long, J. C. and Caceres J. F.** (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem J*, **417**(1): 15-27.
- Loomis, R. J., Naoe Y., Parker J. B., Savic V., Bozovsky M. R., Macfarlan T., Manley J. L. and Chakravarti D.** (2009). Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. *Mol Cell*, **33**(4): 450-61.
- Lopez Castel, A., Nakamori M., Tome S., Chitayat D., Gourdon G., Thornton C. A. and Pearson C. E.** (2010). Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Hum Mol Genet*, **20**(1): 1-15.
- Loria, P. M., Duke A., Rand J. B. and Hobert O.** (2003). Two neuronal, nuclear-localized RNA binding proteins involved in synaptic transmission. *Curr Biol*, **13**(15): 1317-23.
- Lou, H. and Gagel R. F.** (1998). Alternative RNA processing--its role in regulating expression of calcitonin/calcitonin gene-related peptide. *J Endocrinol*, **156**(3): 401-5.
- Luco, R. F., Allo M., Schor I. E., Kornblihtt A. R. and Misteli T.** (2011). Epigenetics in alternative pre-mRNA splicing. *Cell*, **144**(1): 16-26.
- Luco, R. F., Pan Q., Tominaga K., Blencowe B. J., Pereira-Smith O. M. and Misteli T.** (2010). Regulation of alternative splicing by histone modifications. *Science*, **327**(5968): 996-1000.
- Lund, E. and Dahlberg J. E.** (1992). Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. *Science*, **255**(5042): 327-30.
- Machuca-Tzili, L., Brook D. and Hilton-Jones D.** (2005). Clinical and molecular aspects of the myotonic dystrophies: a review. *Muscle Nerve*, **32**(1): 1-18.
- Mahadevan, M., Tsilfidis C., Sabourin L., Shutler G., Amemiya C., Jansen G., Neville C., Narang M., Barcelo J., O'hoy K. and Et Al.** (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science*, **255**(5049): 1253-5.
- Mahadevan, M. S., Yadava R. S., Yu Q., Balijepalli S., Frenzel-Mccardell C. D., Bourne T. D. and Phillips L. H.** (2006). Reversible model of RNA toxicity and cardiac conduction defects in myotonic dystrophy. *Nat Genet*, **38**(9): 1066-70.

- Malinina, L.** (2005). Possible involvement of the RNAi pathway in trinucleotide repeat expansion diseases. *J Biomol Struct Dyn*, **23**(3): 233-5.
- Mangiarini, L., Sathasivam K., Mahal A., Mott R., Seller M. and Bates G. P.** (1997). Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat Genet*, **15**(2): 197-200.
- Mankodi, A., Lin X., Blaxall B. C., Swanson M. S. and Thornton C. A.** (2005). Nuclear RNA foci in the heart in myotonic dystrophy. *Circ Res*, **97**(11): 1152-5.
- Mankodi, A., Logigian E., Callahan L., McClain C., White R., Henderson D., Krym M. and Thornton C. A.** (2000). Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science*, **289**(5485): 1769-73.
- Mankodi, A., Takahashi M. P., Jiang H., Beck C. L., Bowers W. J., Moxley R. T., Cannon S. C. and Thornton C. A.** (2002). Expanded CUG repeats trigger aberrant splicing of *Clc-1* chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol Cell*, **10**(1): 35-44.
- Mankodi, A., Teng-Umuay P., Krym M., Henderson D., Swanson M. and Thornton C. A.** (2003). Ribonuclear inclusions in skeletal muscle in myotonic dystrophy types 1 and 2. *Ann Neurol*, **54**(6): 760-8.
- Mankodi, A., Urbinati C. R., Yuan Q. P., Moxley R. T., Sansone V., Krym M., Henderson D., Schalling M., Swanson M. S. and Thornton C. A.** (2001). Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum Mol Genet*, **10**(19): 2165-70.
- Margolis, J. M., Schoser B. G., Moseley M. L., Day J. W. and Ranum L. P.** (2006). DM2 intronic expansions: evidence for CCUG accumulation without flanking sequence or effects on ZNF9 mRNA processing or protein expression. *Hum Mol Genet*, **15**(11): 1808-15.
- Marquis, J., Paillard L., Audic Y., Cosson B., Danos O., Le Bec C. and Osborne H. B.** (2006). CUG-BP1/CELF1 requires UGU-rich sequences for high-affinity binding. *Biochem J*, **400**(2): 291-301.
- Martorell, L., Gamez J., Cayuela M. L., Gould F. K., Mcabney J. P., Ashizawa T., Monckton D. G. and Baiget M.** (2004). Germline mutational dynamics in myotonic dystrophy type 1 males: allele length and age effects. *Neurology*, **62**(2): 269-74.
- Massenet, S. and Branlant C.** (1999). A limited number of pseudouridine residues in the human ataxin-1 pseudouridine-rich region as compared to human major spliceosomal UsnRNAs. *RNA*, **5**(11): 1495-503.
- Mastroiannopoulos, N. P., Feldman M. L., Uney J. B., Mahadevan M. S. and Phylactou L. A.** (2005). Woodchuck post-transcriptional element induces nuclear export of myotonic dystrophy 3' untranslated region transcripts. *EMBO Rep*, **6**(5): 458-63.
- Mathieu, J., Allard P., Potvin L., Prevost C. and Begin P.** (1999). A 10-year study of mortality in a cohort of patients with myotonic dystrophy. *Neurology*, **52**(8): 1658-62.
- Matilla, A., Roberson E. D., Banfi S., Morales J., Armstrong D. L., Burright E. N., Orr H. T., Sweatt J. D., Zoghbi H. Y. and Matzuk M. M.** (1998). Mice lacking ataxin-1 display learning deficits and decreased hippocampal paired-pulse facilitation. *J Neurosci*, **18**(14): 5508-16.
- Matsuura, T., Yamagata T., Burgess D. L., Rasmussen A., Grewal R. P., Watase K., Khajavi M., McCall A. E., Davis C. F., Zu L., Achari M., Pulst S. M., Alonso E., Noebels J. L., Nelson D. L., Zoghbi H. Y. and Ashizawa T.** (2000). Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat Genet*, **26**(2): 191-4.
- Matynia, A., Ng C. H., Dansithong W., Chiang A., Silva A. J. and Reddy S.** (2010). Muscleblind1, but not Dmpk or Six5, contributes to a complex phenotype of muscular and motivational deficits in mouse models of myotonic dystrophy. *PLoS One*, **5**(3): e9857.

- Mayeda, A., Munroe S. H., Caceres J. F. and Krainer A. R.** (1994). Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J*, **13**(22): 5483-95.
- Mayeda, A., Sreaton G. R., Chandler S. D., Fu X. D. and Krainer A. R.** (1999). Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Mol Cell Biol*, **19**(3): 1853-63.
- Mayes, A. E., Verdone L., Legrain P. and Beggs J. D.** (1999). Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J*, **18**(15): 4321-31.
- Mcmurray, C. T.** (1999). DNA secondary structure: a common and causative factor for expansion in human disease. *Proc Natl Acad Sci U S A*, **96**(5): 1823-5.
- Mcmurray, C. T.** (2010). Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet*, **11**(11): 786-99.
- Meins, M., Schlickum S., Wilhelm C., Missbach J., Yadav S., Glaser B., Grzmil M., Burfeind P. and Laccone F.** (2002). Identification and characterization of murine Brunol4, a new member of the elav/bruno family. *Cytogenet Genome Res*, **97**(3-4): 254-60.
- Meola, G., Sansone V., Marinou K., Cotelli M., Moxley R. T., 3rd, Thornton C. A. and De Ambroggi L.** (2002). Proximal myotonic myopathy: a syndrome with a favourable prognosis? *J Neurol Sci*, **193**(2): 89-96.
- Meola, G., Sansone V., Perani D., Scarone S., Cappa S., Dragoni C., Cattaneo E., Cotelli M., Gobbo C., Fazio F., Siciliano G., Mancuso M., Vitelli E., Zhang S., Krahe R. and Moxley R. T.** (2003). Executive dysfunction and avoidant personality trait in myotonic dystrophy type 1 (DM-1) and in proximal myotonic myopathy (PROMM/DM-2). *Neuromuscul Disord*, **13**(10): 813-21.
- Merino, J. L., Carmona J. R., Fernandez-Lozano I., Peinado R., Basterra N. and Sobrino J. A.** (1998). Mechanisms of sustained ventricular tachycardia in myotonic dystrophy: implications for catheter ablation. *Circulation*, **98**(6): 541-6.
- Mermoud, J. E., Cohen P. T. and Lamond A. I.** (1994). Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J*, **13**(23): 5679-88.
- Messaed, C. and Rouleau G. A.** (2009). Molecular mechanisms underlying polyalanine diseases. *Neurobiol Dis*, **34**(3): 397-405.
- Meyuhas, O.** (2000). Synthesis of the translational apparatus is regulated at the translational level. *Eur J Biochem*, **267**(21): 6321-30.
- Michael, W. M., Choi M. and Dreyfuss G.** (1995). A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell*, **83**(3): 415-22.
- Michalowski, S., Miller J. W., Urbinati C. R., Paliouras M., Swanson M. S. and Griffith J.** (1999). Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. *Nucleic Acids Res*, **27**(17): 3534-42.
- Michlewski, G., Sanford J. R. and Caceres J. F.** (2008). The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Mol Cell*, **30**(2): 179-89.
- Miller, J. W., Urbinati C. R., Teng-Umuay P., Stenberg M. G., Byrne B. J., Thornton C. A. and Swanson M. S.** (2000). Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *Embo J*, **19**(17): 4439-48.
- Minovitsky, S., Gee S. L., Schokrpur S., Dubchak I. and Conboy J. G.** (2005). The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. *Nucleic Acids Res*, **33**(2): 714-24.
- Mirkin, S. M.** (2007). Expandable DNA repeats and human disease. *Nature*, **447**(7147): 932-40.
- Mirkin, S. M. and Smirnova E. V.** (2002). Positioned to expand. *Nat Genet*, **31**(1): 5-6.
- Misteli, T. and Spector D. L.** (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol Cell*, **3**(6): 697-705.

- Modrek, B. and Lee C.** (2002). A genomic view of alternative splicing. *Nat Genet*, **30**(1): 13-9.
- Monferrer, L. and Artero R.** (2006). An interspecific functional complementation test in *Drosophila* for introductory genetics laboratory courses. *J Hered*, **97**(1): 67-73.
- Monsalve, M., Wu Z., Adelmant G., Puigserver P., Fan M. and Spiegelman B. M.** (2000). Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1. *Mol Cell*, **6**(2): 307-16.
- Mooers, B. H., Logue J. S. and Berglund J. A.** (2005). The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc Natl Acad Sci U S A*, **102**(46): 16626-31.
- Moore, M. J. and Sharp P. A.** (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature*, **365**(6444): 364-8.
- Moraes, K. C., Wilusz C. J. and Wilusz J.** (2006). CUG-BP binds to RNA substrates and recruits PARN deadenylase. *Rna*, **12**(6): 1084-91.
- Morris, D. P. and Greenleaf A. L.** (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem*, **275**(51): 39935-43.
- Moseley, M. L., Zu T., Ikeda Y., Gao W., Mosemiller A. K., Daughters R. S., Chen G., Weatherspoon M. R., Clark H. B., Ebner T. J., Day J. W. and Ranum L. P.** (2006). Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat Genet*, **38**(7): 758-69.
- Mougin, A., Gottschalk A., Fabrizio P., Luhrmann R. and Branlant C.** (2002). Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles. *J Mol Biol*, **317**(5): 631-49.
- Mounsey, J. P., John J. E., 3rd, Helmke S. M., Bush E. W., Gilbert J., Roses A. D., Perryman M. B., Jones L. R. and Moorman J. R.** (2000). Phospholemman is a substrate for myotonic dystrophy protein kinase. *J Biol Chem*, **275**(30): 23362-7.
- Mount, S. M. and Steitz J. A.** (1981). Sequence of U1 RNA from *Drosophila melanogaster*: implications for U1 secondary structure and possible involvement in splicing. *Nucleic Acids Res*, **9**(23): 6351-68.
- Mulders, S. A., Van Den Broek W. J., Wheeler T. M., Croes H. J., Van Kuik-Romeijn P., De Kimpe S. J., Furling D., Platenburg G. J., Gourdon G., Thornton C. A., Wieringa B. and Wansink D. G.** (2009). Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy. *Proc Natl Acad Sci U S A*, **106**(33): 13915-20.
- Muranyi, A., Zhang R., Liu F., Hirano K., Ito M., Epstein H. F. and Hartshorne D. J.** (2001). Myotonic dystrophy protein kinase phosphorylates the myosin phosphatase targeting subunit and inhibits myosin phosphatase activity. *FEBS Lett*, **493**(2-3): 80-4.
- Nagel, R. J., Lancaster A. M. and Zahler A. M.** (1998). Specific binding of an exonic splicing enhancer by the pre-mRNA splicing factor SRp55. *RNA*, **4**(1): 11-23.
- Nahkuri, S., Taft R. J. and Mattick J. S.** (2009). Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle*, **8**(20): 3420-4.
- Nakamori, M., Kimura T., Kubota T., Matsumura T., Sumi H., Fujimura H., Takahashi M. P. and Sakoda S.** (2008). Aberrantly spliced alpha-dystrobrevin alters alpha-syntrophin binding in myotonic dystrophy type 1. *Neurology*, **70**(9): 677-85.
- Nakamori, M., Pearson C. E. and Thornton C. A.** (2011). Bidirectional transcription stimulates expansion and contraction of expanded (CTG)ⁿ(CAG)^m repeats. *Hum Mol Genet*, **20**(3): 580-8.
- Nakamura, K., Jeong S. Y., Uchihara T., Anno M., Nagashima K., Nagashima T., Ikeda S., Tsuji S. and Kanazawa I.** (2001). SCA17, a novel autosomal dominant cerebellar ataxia

- caused by an expanded polyglutamine in TATA-binding protein. *Hum Mol Genet*, **10**(14): 1441-8.
- Nakielny, S. and Dreyfuss G.** (1999). Transport of proteins and RNAs in and out of the nucleus. *Cell*, **99**(7): 677-90.
- Napierala, M. and Krzyzosiak W. J.** (1997). CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J Biol Chem*, **272**(49): 31079-85.
- Nassar, R., Malouf N. N., Kelly M. B., Oakeley A. E. and Anderson P. A.** (1991). Force-pCa relation and troponin T isoforms of rabbit myocardium. *Circ Res*, **69**(6): 1470-5.
- Nemes, J. P., Benzow K. A., Moseley M. L., Ranum L. P. and Koob M. D.** (2000). The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1). *Hum Mol Genet*, **9**(10): 1543-51.
- Nguyen, H. H., Wolfe J. T., 3rd, Holmes D. R., Jr. and Edwards W. D.** (1988). Pathology of the cardiac conduction system in myotonic dystrophy: a study of 12 cases. *J Am Coll Cardiol*, **11**(3): 662-71.
- Nilson, I., Reichel M., Ennas M. G., Greim R., Knorr C., Siegler G., Greil J., Fey G. H. and Marschalek R.** (1997). Exon/intron structure of the human AF-4 gene, a member of the AF-4/LAF-4/FMR-2 gene family coding for a nuclear protein with structural alterations in acute leukaemia. *Br J Haematol*, **98**(1): 157-69.
- Nogues, G., Kadener S., Cramer P., Bentley D. and Kornblihtt A. R.** (2002). Transcriptional activators differ in their abilities to control alternative splicing. *J Biol Chem*, **277**(45): 43110-4.
- Nogues, G., Munoz M. J. and Kornblihtt A. R.** (2003). Influence of polymerase II processivity on alternative splicing depends on splice site strength. *J Biol Chem*, **278**(52): 52166-71.
- Ogut, O., Granzier H. and Jin J. P.** (1999). Acidic and basic troponin T isoforms in mature fast-twitch skeletal muscle and effect on contractility. *Am J Physiol*, **276**(5 Pt 1): C1162-70.
- Onishi, H., Kino Y., Morita T., Futai E., Sasagawa N. and Ishiura S.** (2008). MBNL1 associates with YB-1 in cytoplasmic stress granules. *J Neurosci Res*, **86**(9): 1994-2002.
- Orengo, J. P., Chambon P., Metzger D., Mosier D. R., Snipes G. J. and Cooper T. A.** (2008). Expanded CTG repeats within the DMPK 3' UTR causes severe skeletal muscle wasting in an inducible mouse model for myotonic dystrophy. *Proc Natl Acad Sci U S A*, **105**(7): 2646-51.
- Osborne, R. J., Lin X., Welle S., Sobczak K., O'rourke J. R., Swanson M. S. and Thornton C. A.** (2009). Transcriptional and post-transcriptional impact of toxic RNA in myotonic dystrophy. *Hum Mol Genet*, **18**(8): 1471-81.
- Osborne, R. J. and Thornton C. A.** (2006). RNA-dominant diseases. *Hum Mol Genet*, **15** Spec No 2(R162-9).
- Ostareck-Lederer, A. and Ostareck D. H.** (2004). Control of mRNA translation and stability in haematopoietic cells: the function of hnRNPs K and E1/E2. *Biol Cell*, **96**(6): 407-11.
- Otten, A. D. and Tapscott S. J.** (1995). Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc Natl Acad Sci U S A*, **92**(12): 5465-9.
- Pagani, F., Stuani C., Zuccato E., Kornblihtt A. R. and Baralle F. E.** (2003). Promoter architecture modulates CFTR exon 9 skipping. *J Biol Chem*, **278**(3): 1511-7.
- Paillard, L., Legagneux V. and Beverley Osborne H.** (2003). A functional deadenylation assay identifies human CUG-BP as a deadenylation factor. *Biol Cell*, **95**(2): 107-13.
- Pall, G. S., Johnson K. J. and Smith G. L.** (2003). Abnormal contractile activity and calcium cycling in cardiac myocytes isolated from DMPK knockout mice. *Physiol Genomics*, **13**(2): 139-46.

- Panaite, P. A., Gantelet E., Kraftsik R., Gourdon G., Kuntzer T. and Barakat-Walter I.** (2008). Myotonic dystrophy transgenic mice exhibit pathologic abnormalities in diaphragm neuromuscular junctions and phrenic nerves. *J Neuropathol Exp Neurol*, **67**(8): 763-72.
- Pandolfo, M.** (2002). The molecular basis of Friedreich ataxia. *Adv Exp Med Biol*, **516**(99-118).
- Panigrahi, G. B., Lau R., Montgomery S. E., Leonard M. R. and Pearson C. E.** (2005). Slipped (CTG)^{*}(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. *Nat Struct Mol Biol*, **12**(8): 654-62.
- Pascual, M., Vicente M., Monferrer L. and Artero R.** (2006). The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation*, **74**(2-3): 65-80.
- Patel, A. A. and Steitz J. A.** (2003). Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol*, **4**(12): 960-70.
- Patel, N. A., Kaneko S., Apostolatos H. S., Bae S. S., Watson J. E., Davidowitz K., Chappell D. S., Birnbaum M. J., Cheng J. Q. and Cooper D. R.** (2005). Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase CbetaII alternative splicing via phosphorylation of serine/arginine-rich splicing factor SRp40. *J Biol Chem*, **280**(14): 14302-9.
- Paul, S., Dansithong W., Kim D., Rossi J., Webster N. J., Comai L. and Reddy S.** (2006). Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. *EMBO J*, **25**(18): 4271-83.
- Paushkin, S., Gubitz A. K., Massenet S. and Dreyfuss G.** (2002). The SMN complex, an assemblysome of ribonucleoproteins. *Curr Opin Cell Biol*, **14**(3): 305-12.
- Pearson, C. E., Nichol Edamura K. and Cleary J. D.** (2005). Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet*, **6**(10): 729-42.
- Pearson, C. E. and Sinden R. R.** (1996). Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry*, **35**(15): 5041-53.
- Pearson, C. E., Tam M., Wang Y. H., Montgomery S. E., Dar A. C., Cleary J. D. and Nichol K.** (2002). Slipped-strand DNAs formed by long (CAG)^{*}(CTG) repeats: slipped-out repeats and slip-out junctions. *Nucleic Acids Res*, **30**(20): 4534-47.
- Pelletier, R., Farrell B. T., Miret J. J. and Lahue R. S.** (2005). Mechanistic features of CAG^{*}CTG repeat contractions in cultured cells revealed by a novel genetic assay. *Nucleic Acids Res*, **33**(17): 5667-76.
- Pelletier, R., Hamel F., Beaulieu D., Patry L., Haineault C., Tarnopolsky M., Schoser B. and Puymirat J.** (2009). Absence of a differentiation defect in muscle satellite cells from DM2 patients. *Neurobiol Dis*, **36**(1): 181-90.
- Pellizzoni, L., Lotti F., Maras B. and Pierandrei-Amaldi P.** (1997). Cellular nucleic acid binding protein binds a conserved region of the 5' UTR of *Xenopus laevis* ribosomal protein mRNAs. *J Mol Biol*, **267**(2): 264-75.
- Perbellini, R., Greco S., Sarra-Ferraris G., Cardani R., Capogrossi M. C., Meola G. and Martelli F.** (2011). Dysregulation and cellular mislocalization of specific miRNAs in myotonic dystrophy type 1. *Neuromuscul Disord*, **21**(2): 81-8.
- Perie, S., Mamchaoui K., Mouly V., Blot S., Bouazza B., Thornell L. E., St Guily J. L. and Butler-Browne G.** (2006). Premature proliferative arrest of cricopharyngeal myoblasts in oculo-pharyngeal muscular dystrophy: Therapeutic perspectives of autologous myoblast transplantation. *Neuromuscul Disord*, **16**(11): 770-81.
- Philips, A. V., Timchenko L. T. and Cooper T. A.** (1998). Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science*, **280**(5364): 737-41.

- Phillips, M. F., Smith P. E., Carroll N., Edwards R. H. and Calverley P. M.** (1999). Nocturnal oxygenation and prognosis in Duchenne muscular dystrophy. *Am J Respir Crit Care Med*, **160**(1): 198-202.
- Piacentini, L., Fanti L., Negri R., Del Vescovo V., Fatica A., Altieri F. and Pimpinelli S.** (2009). Heterochromatin protein 1 (HP1a) positively regulates euchromatic gene expression through RNA transcript association and interaction with hnRNPs in *Drosophila*. *PLoS Genet*, **5**(10): e1000670.
- Pinol-Roma, S. and Dreyfuss G.** (1992). Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature*, **355**(6362): 730-2.
- Pollard, V. W., Michael W. M., Nakielny S., Siomi M. C., Wang F. and Dreyfuss G.** (1996). A novel receptor-mediated nuclear protein import pathway. *Cell*, **86**(6): 985-94.
- Ponthier, J. L., Schlupe C., Chen W., Lersch R. A., Gee S. L., Hou V. C., Lo A. J., Short S. A., Chasis J. A., Winkelmann J. C. and Conboy J. G.** (2006). Fox-2 splicing factor binds to a conserved intron motif to promote inclusion of protein 4.1R alternative exon 16. *J Biol Chem*, **281**(18): 12468-74.
- Pozzoli, U. and Sironi M.** (2005). Silencers regulate both constitutive and alternative splicing events in mammals. *Cell Mol Life Sci*, **62**(14): 1579-604.
- Prasad, J., Colwill K., Pawson T. and Manley J. L.** (1999). The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. *Mol Cell Biol*, **19**(10): 6991-7000.
- Pushechnikov, A., Lee M. M., Childs-Disney J. L., Sobczak K., French J. M., Thornton C. A. and Disney M. D.** (2009). Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3. *J Am Chem Soc*, **131**(28): 9767-79.
- Rakocevic-Stojanovic, V., Savic D., Pavlovic S., Lavrnic D., Stevic Z., Basta I., Romac S. and Apostolski S.** (2005). Intergenerational changes of CTG repeat depending on the sex of the transmitting parent in myotonic dystrophy type 1. *Eur J Neurol*, **12**(3): 236-7.
- Ramchatesingh, J., Zahler A. M., Neugebauer K. M., Roth M. B. and Cooper T. A.** (1995). A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Mol Cell Biol*, **15**(9): 4898-907.
- Ranum, L. P. and Cooper T. A.** (2006). RNA-mediated neuromuscular disorders. *Annu Rev Neurosci*, **29**(259-77).
- Ranum, L. P., Rasmussen P. F., Benzow K. A., Koob M. D. and Day J. W.** (1998). Genetic mapping of a second myotonic dystrophy locus. *Nat Genet*, **19**(2): 196-8.
- Rau, F., Freyermuth F., Fugier C., Villemin J. P., Fischer M. C., Jost B., Dembele D., Gourdon G., Nicole A., Duboc D., Wahbi K., Day J. W., Fujimura H., Takahashi M. P., Auboeuf D., Dreumont N., Furling D. and Charlet-Berguerand N.** (2011). Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat Struct Mol Biol*, **18**(7): 840-5.
- Ravache, M., Abou-Sleymane G. and Trottier Y.** (2010). [Neurodegenerative polyglutamine expansion diseases: physiopathology and therapeutic strategies]. *Pathol Biol (Paris)*, **58**(5): 357-66.
- Reddy, K., Tam M., Bowater R. P., Barber M., Tomlinson M., Nichol Edamura K., Wang Y. H. and Pearson C. E.** (2011). Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res*, **39**(5): 1749-62.
- Reiser, P. J., Greaser M. L. and Moss R. L.** (1992). Developmental changes in troponin T isoform expression and tension production in chicken single skeletal muscle fibres. *J Physiol*, **449**(573-88).

- Resch, A., Xing Y., Alekseyenko A., Modrek B. and Lee C.** (2004). Evidence for a subpopulation of conserved alternative splicing events under selection pressure for protein reading frame preservation. *Nucleic Acids Res*, **32**(4): 1261-9.
- Richard, G. F., Kerrest A. and Dujon B.** (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev*, **72**(4): 686-727.
- Richard, G. F. and Paques F.** (2000). Mini- and microsatellite expansions: the recombination connection. *EMBO Rep*, **1**(2): 122-6.
- Richards, R. I. and Sutherland G. R.** (1992). Heritable unstable DNA sequences. *Nat Genet*, **1**(1): 7-9.
- Ricker, K., Koch M. C., Lehmann-Horn F., Pongratz D., Otto M., Heine R. and Moxley R. T., 3rd.** (1994). Proximal myotonic myopathy: a new dominant disorder with myotonia, muscle weakness, and cataracts. *Neurology*, **44**(8): 1448-52.
- Rinke, J. and Steitz J. A.** (1985). Association of the lupus antigen La with a subset of U6 snRNA molecules. *Nucleic Acids Res*, **13**(7): 2617-29.
- Roberts, G. C. and Smith C. W.** (2002). Alternative splicing: combinatorial output from the genome. *Curr Opin Chem Biol*, **6**(3): 375-83.
- Roberts, R., Timchenko N. A., Miller J. W., Reddy S., Caskey C. T., Swanson M. S. and Timchenko L. T.** (1997). Altered phosphorylation and intracellular distribution of a (CUG)_n triplet repeat RNA-binding protein in patients with myotonic dystrophy and in myotonin protein kinase knockout mice. *Proc Natl Acad Sci U S A*, **94**(24): 13221-6.
- Rolseth, V., Runden-Pran E., Neurauter C. G., Yndestad A., Luna L., Aukrust P., Ottersen O. P. and Bjoras M.** (2008). Base excision repair activities in organotypic hippocampal slice cultures exposed to oxygen and glucose deprivation. *DNA Repair (Amst)*, **7**(6): 869-78.
- Roscigno, R. F. and Garcia-Blanco M. A.** (1995). SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA*, **1**(7): 692-706.
- Rosman, N. P. and Rebeiz J. J.** (1967). The cerebral defect and myopathy in myotonic dystrophy. A comparative clinicopathological study. *Neurology*, **17**(11): 1106-12.
- Ross, C. A. and Poirier M. A.** (2004). Protein aggregation and neurodegenerative disease. *Nat Med*, **10** Suppl(S10-7).
- Rossi, F., Labourier E., Forne T., Divita G., Derancourt J., Riou J. F., Antoine E., Cathala G., Brunel C. and Tazi J.** (1996). Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature*, **381**(6577): 80-2.
- Rossi, J. J.** (1999). Ribozymes, genomics and therapeutics. *Chem Biol*, **6**(2): R33-7.
- Rowen, L., Young J., Birditt B., Kaur A., Madan A., Philipps D. L., Qin S., Minx P., Wilson R. K., Hood L. and Graveley B. R.** (2002). Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. *Genomics*, **79**(4): 587-97.
- Rudnicki, D. D., Holmes S. E., Lin M. W., Thornton C. A., Ross C. A. and Margolis R. L.** (2007). Huntington's disease--like 2 is associated with CUG repeat-containing RNA foci. *Ann Neurol*, **61**(3): 272-82.
- Ryan, K. J. and Cooper T. A.** (1996). Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle. *Mol Cell Biol*, **16**(8): 4014-23.
- Salgado-Garrido, J., Bragado-Nilsson E., Kandels-Lewis S. and Seraphin B.** (1999). Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J*, **18**(12): 3451-62.
- Salinas-Rios, V., Belotserkovskii B. P. and Hanawalt P. C.** (2011). DNA slip-outs cause RNA polymerase II arrest in vitro: potential implications for genetic instability. *Nucleic Acids Res*.

- Salisbury, E., Schoser B., Schneider-Gold C., Wang G. L., Huichalaf C., Jin B., Sirito M., Sarkar P., Krahe R., Timchenko N. A. and Timchenko L. T.** (2009). Expression of RNA CCUG repeats dysregulates translation and degradation of proteins in myotonic dystrophy 2 patients. *Am J Pathol*, **175**(2): 748-62.
- Sammons, M. A., Antons A. K., Bendjennat M., Udd B., Krahe R. and Link A. J.** (2010). ZNF9 activation of IRES-mediated translation of the human ODC mRNA is decreased in myotonic dystrophy type 2. *PLoS One*, **5**(2): e9301.
- Sanford, J. R. and Bruzik J. P.** (2001). Regulation of SR protein localization during development. *Proc Natl Acad Sci U S A*, **98**(18): 10184-9.
- Sanford, J. R., Gray N. K., Beckmann K. and Caceres J. F.** (2004). A novel role for shuttling SR proteins in mRNA translation. *Genes Dev*, **18**(7): 755-68.
- Savkur, R. S., Philips A. V. and Cooper T. A.** (2001). Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat Genet*, **29**(1): 40-7.
- Savkur, R. S., Philips A. V., Cooper T. A., Dalton J. C., Moseley M. L., Ranum L. P. and Day J. W.** (2004). Insulin receptor splicing alteration in myotonic dystrophy type 2. *Am J Hum Genet*, **74**(6): 1309-13.
- Savouret, C., Brisson E., Essers J., Kanaar R., Pastink A., Te Riele H., Junien C. and Gourdon G.** (2003). CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J*, **22**(9): 2264-73.
- Savouret, C., Garcia-Cordier C., Megret J., Te Riele H., Junien C. and Gourdon G.** (2004). MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. *Mol Cell Biol*, **24**(2): 629-37.
- Schneider, C., Ziegler A., Ricker K., Grimm T., Kress W., Reimers C. D., Meinck H., Reiners K. and Toyka K. V.** (2000). Proximal myotonic myopathy: evidence for anticipation in families with linkage to chromosome 3q. *Neurology*, **55**(3): 383-8.
- Schor, I. E., Rascovan N., Pelisch F., Allo M. and Kornblihtt A. R.** (2009). Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc Natl Acad Sci U S A*, **106**(11): 4325-30.
- Schoser, B. G., Ricker K., Schneider-Gold C., Hengstenberg C., Durre J., Bultmann B., Kress W., Day J. W. and Ranum L. P.** (2004). Sudden cardiac death in myotonic dystrophy type 2. *Neurology*, **63**(12): 2402-4.
- Schullery, D. S., Ostrowski J., Denisenko O. N., Stempka L., Shnyreva M., Suzuki H., Gschwendt M. and Bomsztyk K.** (1999). Regulated interaction of protein kinase Cdelta with the heterogeneous nuclear ribonucleoprotein K protein. *J Biol Chem*, **274**(21): 15101-9.
- Schwartz, S., Meshorer E. and Ast G.** (2009). Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol*, **16**(9): 990-5.
- Sellier, C., Rau F., Liu Y., Tassone F., Hukema R. K., Gattoni R., Schneider A., Richard S., Willemsen R., Elliott D. J., Hagerman P. J. and Charlet-Berguerand N.** (2010). Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J*, **29**(7): 1248-61.
- Seraphin, B.** (1995). Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J*, **14**(9): 2089-98.
- Sergeant, N., Delacourte A. and Buee L.** (2005). Tau protein as a differential biomarker of tauopathies. *Biochim Biophys Acta*, **1739**(2-3): 179-97.
- Sergeant, N., Sablonniere B., Schraen-Maschke S., Ghestem A., Muraige C. A., Watzel A., Vermersch P. and Delacourte A.** (2001). Dysregulation of human brain microtubule-associated tau mRNA maturation in myotonic dystrophy type 1. *Hum Mol Genet*, **10**(19): 2143-55.

- Seznec, H., Agbulut O., Sergeant N., Savouret C., Ghestem A., Tabti N., Willer J. C., Ourth L., Duros C., Brisson E., Fouquet C., Butler-Browne G., Delacourte A., Junien C. and Gourdon G.** (2001). Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities. *Hum Mol Genet*, **10**(23): 2717-26.
- Shen, H. and Green M. R.** (2004). A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol Cell*, **16**(3): 363-73.
- Shimizu, K., Chen W., Ashique A. M., Moroi R. and Li Y. P.** (2003). Molecular cloning, developmental expression, promoter analysis and functional characterization of the mouse CNBP gene. *Gene*, **307**(51-62).
- Shimizu, M., Wang W., Walch E. T., Dunne P. W. and Epstein H. F.** (2000). Rac-1 and Raf-1 kinases, components of distinct signaling pathways, activate myotonic dystrophy protein kinase. *FEBS Lett*, **475**(3): 273-7.
- Shin, J., Charizanis K. and Swanson M. S.** (2009). Pathogenic RNAs in microsatellite expansion disease. *Neurosci Lett*, **466**(2): 99-102.
- Shishkin, A. A., Voineagu I., Matera R., Cherg N., Chernet B. T., Krasilnikova M. M., Narayanan V., Lobachev K. S. and Mirkin S. M.** (2009). Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Mol Cell*, **35**(1): 82-92.
- Simard, M. J. and Chabot B.** (2002). SRp30c is a repressor of 3' splice site utilization. *Mol Cell Biol*, **22**(12): 4001-10.
- Sims, R. J., 3rd, Millhouse S., Chen C. F., Lewis B. A., Erdjument-Bromage H., Tempst P., Manley J. L. and 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): 665-76.
- Sirand-Pugnet, P., Durosay P., Clouet D'orval B. C., Brody E. and Marie J.** (1995). beta-Tropomyosin pre-mRNA folding around a muscle-specific exon interferes with several steps of spliceosome assembly. *J Mol Biol*, **251**(5): 591-602.
- Slean, M. M., Panigrahi G. B., Ranum L. P. and Pearson C. E.** (2008). Mutagenic roles of DNA "repair" proteins in antibody diversity and disease-associated trinucleotide repeat instability. *DNA Repair (Amst)*, **7**(7): 1135-54.
- Smith, C. W. and Valcarcel J.** (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci*, **25**(8): 381-8.
- Smith, K. P., Byron M., Johnson C., Xing Y. and Lawrence J. B.** (2007). Defining early steps in mRNA transport: mutant mRNA in myotonic dystrophy type I is blocked at entry into SC-35 domains. *J Cell Biol*, **178**(6): 951-64.
- Sobczak, K., De Mezer M., Michlewski G., Krol J. and Krzyzosiak W. J.** (2003). RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res*, **31**(19): 5469-82.
- Sobczak, K., Michlewski G., De Mezer M., Kierzek E., Krol J., Olejniczak M., Kierzek R. and Krzyzosiak W. J.** (2010). Structural diversity of triplet repeat RNAs. *J Biol Chem*, **285**(17): 12755-64.
- Sorek, R., Lev-Maor G., Reznik M., Dagan T., Belinky F., Graur D. and Ast G.** (2004). Minimal conditions for exonization of intronic sequences: 5' splice site formation in alu exons. *Mol Cell*, **14**(2): 221-31.
- Spies, N., Nielsen C. B., Padgett R. A. and Burge C. B.** (2009). Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell*, **36**(2): 245-54.
- Spranger, M., Spranger S., Tischendorf M., Meinck H. M. and Cremer M.** (1997). Myotonic dystrophy. The role of large triplet repeat length in the development of mental retardation. *Arch Neurol*, **54**(3): 251-4.

- Squillace, R. M., Chenault D. M. and Wang E. H.** (2002). Inhibition of muscle differentiation by the novel muscleblind-related protein CHCR. *Dev Biol*, **250**(1): 218-30.
- Staknis, D. and Reed R.** (1994). Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. *Mol Cell Biol*, **14**(5): 2994-3005.
- Stamm, S., Zhang M. Q., Marr T. G. and Helfman D. M.** (1994). A sequence compilation and comparison of exons that are alternatively spliced in neurons. *Nucleic Acids Res*, **22**(9): 1515-26.
- Stark, J. M., Bazett-Jones D. P., Herfort M. and Roth M. B.** (1998). SR proteins are sufficient for exon bridging across an intron. *Proc Natl Acad Sci U S A*, **95**(5): 2163-8.
- Steinbach, P., Glaser D., Vogel W., Wolf M. and Schwemmle S.** (1998). The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. *Am J Hum Genet*, **62**(2): 278-85.
- Steyaert, J., Umans S., Willekens D., Legius E., Pijckels E., De Die-Smulders C., Van Den Berghe H. and Fryns J. P.** (1997). A study of the cognitive and psychological profile in 16 children with congenital or juvenile myotonic dystrophy. *Clin Genet*, **52**(3): 135-41.
- Sureau, A., Sauliere J., Expert-Bezancon A. and Marie J.** (2011). CELF and PTB proteins modulate the inclusion of the beta-tropomyosin exon 6B during myogenic differentiation. *Exp Cell Res*, **317**(1): 94-106.
- Swanson, M. S. and Dreyfuss G.** (1988). Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol Cell Biol*, **8**(5): 2237-41.
- Swartz, J. E., Bor Y. C., Misawa Y., Rekosh D. and Hammarskjold M. L.** (2007). The shuttling SR protein 9G8 plays a role in translation of unspliced mRNA containing a constitutive transport element. *J Biol Chem*, **282**(27): 19844-53.
- Szymczyna, B. R., Bowman J., Mccracken S., Pineda-Lucena A., Lu Y., Cox B., Lambermon M., Graveley B. R., Arrowsmith C. H. and Blencowe B. J.** (2003). Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. *Genes Dev*, **17**(4): 461-75.
- Tacke, R. and Manley J. L.** (1995). The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J*, **14**(14): 3540-51.
- Taneja, K. L., Mccurrach M., Schalling M., Housman D. and Singer R. H.** (1995). Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J Cell Biol*, **128**(6): 995-1002.
- Tarn, W. Y. and Steitz J. A.** (1995). Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps. *Proc Natl Acad Sci U S A*, **92**(7): 2504-8.
- Tarn, W. Y. and Steitz J. A.** (1997). Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci*, **22**(4): 132-7.
- Tarn, W. Y. and Steitz J. A.** (1994). SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes Dev*, **8**(22): 2704-17.
- Tassone, F., Hagerman R. J., Garcia-Arocena D., Khandjian E. W., Greco C. M. and Hagerman P. J.** (2004). Intranuclear inclusions in neural cells with premutation alleles in fragile X associated tremor/ataxia syndrome. *J Med Genet*, **41**(4): e43.
- Teplova, M. and Patel D. J.** (2008). Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat Struct Mol Biol*, **15**(12): 1343-51.
- Terenzi, F. and Ladd A. N.** (2010). Conserved developmental alternative splicing of muscleblind-like (MBNL) transcripts regulates MBNL localization and activity. *RNA Biol*, **7**(1): 43-55.

- Thornton, C. A., Wymer J. P., Simmons Z., McClain C. and Moxley R. T., 3rd.** (1997). Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nat Genet*, **16**(4): 407-9.
- Tian, B., White R. J., Xia T., Welle S., Turner D. H., Mathews M. B. and Thornton C. A.** (2000). Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA*, **6**(1): 79-87.
- Tilgner, H., Nikolaou C., Althammer S., Sammeth M., Beato M., Valcarcel J. and Guigo R.** (2009). Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol*, **16**(9): 996-1001.
- Timchenko, L., Monckton D. G. and Caskey C. T.** (1995). Myotonic dystrophy: an unstable CTG repeat in a protein kinase gene. *Semin Cell Biol*, **6**(1): 13-9.
- Timchenko, L. T., Miller J. W., Timchenko N. A., Devore D. R., Datar K. V., Lin L., Roberts R., Caskey C. T. and Swanson M. S.** (1996a). Identification of a (CUG)_n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res*, **24**(22): 4407-14.
- Timchenko, L. T., Timchenko N. A., Caskey C. T. and Roberts R.** (1996b). Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy. *Hum Mol Genet*, **5**(1): 115-21.
- Timchenko, N. A., Cai Z. J., Welm A. L., Reddy S., Ashizawa T. and Timchenko L. T.** (2001). RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. *J Biol Chem*, **276**(11): 7820-6.
- Timchenko, N. A., Patel R., Iakova P., Cai Z. J., Quan L. and Timchenko L. T.** (2004). Overexpression of CUG triplet repeat-binding protein, CUGBP1, in mice inhibits myogenesis. *J Biol Chem*, **279**(13): 13129-39.
- Timchenko, N. A., Wang G. L. and Timchenko L. T.** (2005). RNA CUG-binding protein 1 increases translation of 20-kDa isoform of CCAAT/enhancer-binding protein beta by interacting with the alpha and beta subunits of eukaryotic initiation translation factor 2. *J Biol Chem*, **280**(21): 20549-57.
- Tiscornia, G. and Mahadevan M. S.** (2000). Myotonic dystrophy: the role of the CUG triplet repeats in splicing of a novel DMPK exon and altered cytoplasmic DMPK mRNA isoform ratios. *Mol Cell*, **5**(6): 959-67.
- Tome, S., Holt I., Edelmann W., Morris G. E., Munnich A., Pearson C. E. and Gourdon G.** (2009). MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice. *PLoS Genet*, **5**(5): e1000482.
- Townsend, P. J., Barton P. J., Yacoub M. H. and Farza H.** (1995). Molecular cloning of human cardiac troponin T isoforms: expression in developing and failing heart. *J Mol Cell Cardiol*, **27**(10): 2223-36.
- Tran, H., Gourrier N., Lemercier-Neuillet C., Dhaenens C. M., Vautrin A., Fernandez-Gomez F. J., Arandel L., Carpentier C., Obriot H., Eddarkaoui S., Delattre L., Van Brussels E., Holt I., Morris G. E., Sablonniere B., Buee L., Charlet-Berguerand N., Schraen-Maschke S., Furling D., Behm-Ansmant I., Branlant C., Caillet-Boudin M. L. and Sergeant N.** (2011). Analysis of exonic regions involved in nuclear localization, splicing activity, and dimerization of Muscleblind-like-1 isoforms. *J Biol Chem*, **286**(18): 16435-46.
- Ujvari, A. and Luse D. S.** (2004). Newly Initiated RNA encounters a factor involved in splicing immediately upon emerging from within RNA polymerase II. *J Biol Chem*, **279**(48): 49773-9.
- Ule, J., Jensen K. B., Ruggiu M., Mele A., Ule A. and Darnell R. B.** (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science*, **302**(5648): 1212-5.

- Ule, J., Stefani G., Mele A., Ruggiu M., Wang X., Taneri B., Gaasterland T., Blencowe B. J. and Darnell R. B.** (2006). An RNA map predicting Nova-dependent splicing regulation. *Nature*, **444**(7119): 580-6.
- Usdin, K.** (1998). NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases. *Nucleic Acids Res*, **26**(17): 4078-85.
- Usdin, K. and Grabczyk E.** (2000). DNA repeat expansions and human disease. *Cell Mol Life Sci*, **57**(6): 914-31.
- Van Den Broek, W. J., Wansink D. G. and Wieringa B.** (2007). Somatic CTG*CAG repeat instability in a mouse model for myotonic dystrophy type 1 is associated with changes in cell nuclearity and DNA ploidy. *BMC Mol Biol*, **8**(61).
- Vassileva, M. T. and Matunis M. J.** (2004). SUMO modification of heterogeneous nuclear ribonucleoproteins. *Mol Cell Biol*, **24**(9): 3623-32.
- Verkerk, A. J., Pieretti M., Sutcliffe J. S., Fu Y. H., Kuhl D. P., Pizzuti A., Reiner O., Richards S., Victoria M. F., Zhang F. P. and Et Al.** (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, **65**(5): 905-14.
- Vicente-Crespo, M., Pascual M., Fernandez-Costa J. M., Garcia-Lopez A., Monferrer L., Miranda M. E., Zhou L. and Artero R. D.** (2008). Drosophila muscleblind is involved in troponin T alternative splicing and apoptosis. *PLoS One*, **3**(2): e1613.
- Vicente, M., Monferrer L., Poulos M. G., Houseley J., Monckton D. G., O'dell K M., Swanson M. S. and Artero R. D.** (2007). Muscleblind isoforms are functionally distinct and regulate alpha-actinin splicing. *Differentiation*, **75**(5): 427-40.
- Vidal, V. P., Verdone L., Mayes A. E. and Beggs J. D.** (1999). Characterization of U6 snRNA-protein interactions. *RNA*, **5**(11): 1470-81.
- Vinereanu, D., Bajaj B. P., Fenton-May J., Rogers M. T., Madler C. F. and Fraser A. G.** (2004). Subclinical cardiac involvement in myotonic dystrophy manifesting as decreased myocardial Doppler velocities. *Neuromuscul Disord*, **14**(3): 188-94.
- Vlasova, I. A. and Bohjanen P. R.** (2008). Posttranscriptional regulation of gene networks by GU-rich elements and CELF proteins. *RNA Biol*, **5**(4): 201-7.
- Wagner, E. J. and Garcia-Blanco M. A.** (2001). Polypyrimidine tract binding protein antagonizes exon definition. *Mol Cell Biol*, **21**(10): 3281-8.
- Wahl, M. C., Will C. L. and Luhrmann R.** (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**(4): 701-18.
- Wang, G. S., Kearney D. L., De Biasi M., Taffet G. and Cooper T. A.** (2007). Elevation of RNA-binding protein CUGBP1 is an early event in an inducible heart-specific mouse model of myotonic dystrophy. *J Clin Invest*, **117**(10): 2802-11.
- Wang, G. S., Kuyumcu-Martinez M. N., Sarma S., Mathur N., Wehrens X. H. and Cooper T. A.** (2009). PKC inhibition ameliorates the cardiac phenotype in a mouse model of myotonic dystrophy type 1. *J Clin Invest*, **119**(12): 3797-806.
- Wang, H. Y., Lin W., Dyck J. A., Yeakley J. M., Songyang Z., Cantley L. C. and Fu X. D.** (1998). SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J Cell Biol*, **140**(4): 737-50.
- Wang, Q. and Bag J.** (2006). Ectopic expression of a polyalanine expansion mutant of poly(A)-binding protein N1 in muscle cells in culture inhibits myogenesis. *Biochem Biophys Res Commun*, **340**(3): 815-22.
- Wang, Y. H., Amirhaeri S., Kang S., Wells R. D. and Griffith J. D.** (1994). Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science*, **265**(5172): 669-71.

- Wang, Z. and Burge C. B.** (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, **14**(5): 802-13.
- Wansink, D. G., Van Herpen R. E., Coerwinkel-Driessen M. M., Groenen P. J., Hemmings B. A. and Wieringa B.** (2003). Alternative splicing controls myotonic dystrophy protein kinase structure, enzymatic activity, and subcellular localization. *Mol Cell Biol*, **23**(16): 5489-501.
- Ward, A. J., Rimer M., Killian J. M., Dowling J. J. and Cooper T. A.** (2010). CUGBP1 overexpression in mouse skeletal muscle reproduces features of myotonic dystrophy type 1. *Hum Mol Genet*, **19**(18): 3614-22.
- Warf, M. B. and Berglund J. A.** (2007). MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA*, **13**(12): 2238-51.
- Warf, M. B., Diegel J. V., Von Hippel P. H. and Berglund J. A.** (2009a). The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc Natl Acad Sci U S A*, **106**(23): 9203-8.
- Warf, M. B., Nakamori M., Matthys C. M., Thornton C. A. and Berglund J. A.** (2009b). Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc Natl Acad Sci U S A*, **106**(44): 18551-6.
- Warren, S. T.** (1997). Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13. *Science*, **275**(5298): 408-9.
- Watase, K., Weeber E. J., Xu B., Antalffy B., Yuva-Paylor L., Hashimoto K., Kano M., Atkinson R., Sun Y., Armstrong D. L., Sweatt J. D., Orr H. T., Paylor R. and Zoghbi H. Y.** (2002). A long CAG repeat in the mouse Sca1 locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration. *Neuron*, **34**(6): 905-19.
- Weiner, A. M., Allende M. L., Becker T. S. and Calcaterra N. B.** (2007). CNBP mediates neural crest cell expansion by controlling cell proliferation and cell survival during rostral head development. *J Cell Biochem*, **102**(6): 1553-70.
- Wells, R. D., Dere R., Hebert M. L., Napierala M. and Son L. S.** (2005). Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res*, **33**(12): 3785-98.
- Westerlaken, J. H., Van Der Zee C. E., Peters W. and Wieringa B.** (2003). The DMWD protein from the myotonic dystrophy (DM1) gene region is developmentally regulated and is present most prominently in synapse-dense brain areas. *Brain Res*, **971**(1): 116-27.
- Wheeler, T. M.** (2008). Myotonic dystrophy: therapeutic strategies for the future. *Neurotherapeutics*, **5**(4): 592-600.
- Wheeler, T. M. and Thornton C. A.** (2007). Myotonic dystrophy: RNA-mediated muscle disease. *Curr Opin Neurol*, **20**(5): 572-6.
- White, M. C., Gao R., Xu W., Mandal S. M., Lim J. G., Hazra T. K., Wakamiya M., Edwards S. F., Raskin S., Teive H. A., Zoghbi H. Y., Sarkar P. S. and Ashizawa T.** (2010). Inactivation of hnRNP K by expanded intronic AUUCU repeat induces apoptosis via translocation of PKCdelta to mitochondria in spinocerebellar ataxia 10. *PLoS Genet*, **6**(6): e1000984.
- White, S. P., Cohen C. and Phillips G. N., Jr.** (1987). Structure of co-crystals of tropomyosin and troponin. *Nature*, **325**(6107): 826-8.
- Wieringa, B.** (1994). Myotonic dystrophy reviewed: back to the future? *Hum Mol Genet*, **3**(1): 1-7.
- Will, C. L. and Luhrmann R.** (1997). Protein functions in pre-mRNA splicing. *Curr Opin Cell Biol*, **9**(3): 320-8.
- Will, C. L. and Luhrmann R.** (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol*, **13**(3): 290-301.

- Willemsen, R., Oostra B. A., Bassell G. J. and Dichtenberg J.** (2004). The fragile X syndrome: from molecular genetics to neurobiology. *Ment Retard Dev Disabil Res Rev*, **10**(1): 60-7.
- Wojciechowska, M. and Krzyzosiak W. J.** (2011). Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum Mol Genet*.
- Wolff, S.** (1989). Are radiation-induced effects hormetic? *Science*, **245**(4918): 575, 621.
- Wolin, S. L. and Cedervall T.** (2002). The La protein. *Annu Rev Biochem*, **71**(375-403).
- Wollerton, M. C., Gooding C., Robinson F., Brown E. C., Jackson R. J. and Smith C. W.** (2001). Differential alternative splicing activity of isoforms of polypyrimidine tract binding protein (PTB). *RNA*, **7**(6): 819-32.
- Woodward, J. B., 3rd, Heaton R. K., Simon D. B. and Ringel S. P.** (1982). Neuropsychological findings in myotonic dystrophy. *J Clin Neuropsychol*, **4**(4): 335-42.
- Wu, J. Y. and Maniatis T.** (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, **75**(6): 1061-70.
- Xiao, S. H. and Manley J. L.** (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev*, **11**(3): 334-44.
- Yadava, R. S., Frenzel-Mccardell C. D., Yu Q., Srinivasan V., Tucker A. L., Puymirat J., Thornton C. A., Prall O. W., Harvey R. P. and Mahadevan M. S.** (2008). RNA toxicity in myotonic muscular dystrophy induces NKX2-5 expression. *Nat Genet*, **40**(1): 61-8.
- Yang, J. and Freudenreich C. H.** (2007). Haploinsufficiency of yeast FEN1 causes instability of expanded CAG/CTG tracts in a length-dependent manner. *Gene*, **393**(1-2): 110-5.
- Yang, Z., Lau R., Marcadier J. L., Chitayat D. and Pearson C. E.** (2003). Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type 1 disease locus in human cells. *Am J Hum Genet*, **73**(5): 1092-105.
- Yoo, S. Y., Pennesi M. E., Weeber E. J., Xu B., Atkinson R., Chen S., Armstrong D. L., Wu S. M., Sweatt J. D. and Zoghbi H. Y.** (2003). SCA7 knockin mice model human SCA7 and reveal gradual accumulation of mutant ataxin-7 in neurons and abnormalities in short-term plasticity. *Neuron*, **37**(3): 383-401.
- Yoon, T. and Cowan J. A.** (2003). Iron-sulfur cluster biosynthesis. Characterization of frataxin as an iron donor for assembly of [2Fe-2S] clusters in ISU-type proteins. *J Am Chem Soc*, **125**(20): 6078-84.
- Yu, Y. T., Shu M. D. and Steitz J. A.** (1998). Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J*, **17**(19): 5783-95.
- Yuan, Y., Compton S. A., Sobczak K., Stenberg M. G., Thornton C. A., Griffith J. D. and Swanson M. S.** (2007). Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res*, **35**(16): 5474-86.
- Yuryev, A., Patturajan M., Litingtung Y., Joshi R. V., Gentile C., Gebara M. and Corden J. L.** (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A*, **93**(14): 6975-80.
- Zahler, A. M., Lane W. S., Stolk J. A. and Roth M. B.** (1992). SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev*, **6**(5): 837-47.
- Zavolan, M., Kondo S., Schonbach C., Adachi J., Hume D. A., Hayashizaki Y. and Gaasterland T.** (2003). Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res*, **13**(6B): 1290-300.

- Zeesman, S., Carson N. and Whelan D. T.** (2002). Paternal transmission of the congenital form of myotonic dystrophy type 1: a new case and review of the literature. *Am J Med Genet*, **107**(3): 222-6.
- Zeng, C. and Berget S. M.** (2000). Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. *Mol Cell Biol*, **20**(21): 8290-301.
- Zhan, H. C., Liu D. P. and Liang C. C.** (2001). Insulator: from chromatin domain boundary to gene regulation. *Hum Genet*, **109**(5): 471-8.
- Zhang, J. and Abdel-Rahman A. A.** (2002). The hypotensive action of rilmenidine is dependent on functional N-methyl-D-aspartate receptor in the rostral ventrolateral medulla of conscious spontaneously hypertensive rats. *J Pharmacol Exp Ther*, **303**(1): 204-10.
- Zhang, Z. and Krainer A. R.** (2004). Involvement of SR proteins in mRNA surveillance. *Mol Cell*, **16**(4): 597-607.
- Zoghbi, H. Y. and Orr H. T.** (2000). Glutamine repeats and neurodegeneration. *Annu Rev Neurosci*, **23**(217-47).
- Zu, T., Gibbens B., Doty N. S., Gomes-Pereira M., Huguet A., Stone M. D., Margolis J., Peterson M., Markowski T. W., Ingram M. A., Nan Z., Forster C., Low W. C., Schoser B., Somia N. V., Clark H. B., Schmechel S., Bitterman P. B., Gourdon G., Swanson M. S., Moseley M. and Ranum L. P.** (2011). Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A*, **108**(1): 260-5.
- Zumwalt, M., Ludwig A., Hagerman P. J. and Dieckmann T.** (2007). Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the fragile X mental retardation 1 (FMR1) gene. *RNA Biol*, **4**(2): 93-100.

Summary:

Amplifications of CTG motifs in the human *DMPK* gene are responsible for Myotonic Dystrophy of type 1. The resulting CUG repeats in pre-mRNAs capture the MBNL1 splicing factor, leading to mis-regulation of MBNL1 pre-mRNA targets. Due to the recent discovery of MBNL1 and its numerous isoforms (9) resulting from alternative splicing, little is known on how MBNL1 regulates splicing and how a decreased level of available MBNL1 generates splicing miss-regulations. First, we defined which of the MBNL1 alternative and constitutive exons are required for: i) RNA binding, ii) splicing activity and, iii) MBNL1 sub-cellular localization. Second, for a more precise definition of the MBNL1 RNA binding properties, we performed SELEX experiments using a library of RNA stem-loop structures containing a 18-nt long randomized sequence. Its leads to the identification of 12-nt long sequence adopting a peculiar stem-loop structure, whose importance for MBNL1 binding was revealed by its preservation by compensatory base-pair mutations. Finally, based on the above data, we studied the mechanisms involved in regulation of hcTNT exon 5 splicing. By *in cellulo* assays, we defined the hcTNT pre-mRNA region required for both normal inclusion and for the trans-dominant effect of CUG repeats. Within this region, we identified six new potential MBNL1 sites and demonstrated their functional role by *in vitro* and *in cellulo* assays. We also identified several additional splicing regulatory elements involved in normal and CUG-deregulated exon 5 inclusion and already showed a role of hnRNP H in splicing regulation. Altogether, our data bring new information important for understanding the pathology.

Key words:

DM1, MBNL1, alternative splicing, SELEX, 2D RNA structure, triplet repeat expansion

Résumé:

Les amplifications de répétitions de triplets CTG dans le gène *DMPK* humain sont à l'origine de la dystrophie myotonique de type 1 ou DM1. Les répétitions CUG présentes dans les ARNm *DMPK* séquestrent le facteur d'épissage MBNL1 au sein de foci nucléaires et dérégulent l'épissage des pré-ARNm cibles de MBNL1. Par ailleurs, 9 isoformes de MBNL1, produites par épissage alternative, coexistent dans les cellules. Dans un premier temps nous avons recherché quels exons constitutifs ou alternatifs étaient requis pour la reconnaissance des ARN, la régulation de l'épissage et la localisation cellulaire de MBNL1. Nous avons ensuite entrepris de rechercher par l'approche SELEX les séquences de haute affinité pour MBNL1. Nous avons ainsi identifié une séquence conservée de 12 nucléotides de long, contenant un seul motif de fixation pour MBNL1 et adoptant une structuration tige-boucle particulière. L'importance de cette structuration a été confirmée par l'existence de mutants compensatoires au sein des ARN sélectionnés. Finalement nous avons étudié les mécanismes de régulation de l'inclusion de l'exon 5 du pré-ARNm de la troponine T cardiaque humaine (hcTNT). Une approche *in cellulo* nous a permis d'identifier les séquences minimales requises pour la régulation de l'épissage en conditions normales et en présence des répétitions CUG. Au sein de ces séquences nous avons identifié 6 nouveaux sites MBNL1 dont nous avons montré l'importance fonctionnelle *in cellulo* et *in vitro*. Nous avons également mis en évidence l'implication d'autres séquences régulatrices dans la régulation de l'inclusion de l'exon 5 du pré-ARNm hcTNT et un rôle de la protéine hnRNP H dans ces régulations. L'ensemble de ces données apportent de nouveaux éléments d'information importants pour la compréhension de la DM1.

Mots clés:

DM1, MBNL1, épissage alternatif, SELEX, structure secondaire, expansion de triplets