



HAL
open science

Une nouvelle fonction pour la DEAD-box ARN hélicase p68/DDX5 dans la Dystrophie Myotonique de type 1

François-Xavier Laurent

► To cite this version:

François-Xavier Laurent. Une nouvelle fonction pour la DEAD-box ARN hélicase p68/DDX5 dans la Dystrophie Myotonique de type 1. Sciences agricoles. Université Paris Sud - Paris XI, 2011. Français. NNT : 2011PA112180 . tel-00662593

HAL Id: tel-00662593

<https://theses.hal.science/tel-00662593>

Submitted on 24 Jan 2012

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.

UNIVERSITE DE PARIS-SUD XI – ORSAY

UFR DES SCIENCES

THESE

Pour l'obtention du Diplôme de

DOCTEUR DE L'UNIVERSITE PARIS-SUD XI

Mention Biologie Moléculaire et Cellulaire

Ecole doctorale : Gènes, Génomes Cellules

Par :

François-Xavier LAURENT

Titre :

**Une nouvelle fonction pour la DEAD-box ARN hélicase
p68/DDX5 dans la Dystrophie Myotonique de type 1.**

Soutenue le 30 Septembre 2011, devant la commission d'examen

Composition du Jury :

| | | |
|---------------------------|---|---------------------------|
| Dr Didier AUBOEUF | Chargé de recherche INSERM | Rapporteur |
| Dr Kyle TANNER | Directeur de recherche CNRS | Rapporteur |
| Dr Nicolas CHARLET | Chargé de recherche INSERM | Examineur |
| Pr Pierre CAPY | Professeur à l'Université Paris-Sud XI | Président |
| Dr Joëlle MARIE | Directeur de recherche INSERM | Directeur de thèse |

LISTE DES ABBREVIATIONS

3'UTR : 3' Untranslated Region
APP : amyloid precursor protein
ARN : Acide Ribonucléique
ARN Pol II : ARN Polymérase II
ARNm : ARN messenger
ATP : Adénosine Triphosphate
BIN1 : Bridging Integrator 1
CELF: CUGBP- and ETR3-like factor
ChIP : Chromatin Immunoprécipitation
CLCN1 : Chloride Channel
CLIP : Cross-Linking Immunoprecipitation
CUGBP1 : CUG Binding Protein 1
DM1 : Dystrophie Myotonique de type 1
DM2 : Dystrophie Myotonique de type 2
DMEM: Dubecco's modification of Eagle's medium
DMPK : Dystrophia Myotonica Protein Kinase
DMWD : Dystrophia myotonica WD repeat-containing protein
DSCAM : Down Syndrome Cell Adhesion Molecule
ER α : Estrogen Receptor alpha
ESE : Exonic Splicing Enhancer
ESS : Exonic Splicing Silencer
FISH : Fluorescent In Situ Hybridization
FRAP : Fluorescence recovery after photobleaching
GFP : Green Fluorescent Protein
GST : Gluthation-S-Transferase
HDAC : Histone Deacetylase
hnRNP : heterogeneous nuclear Ribonucleoprotein
HSALR : Human Skeletal Actin Long Repeats
INSR : Insulin Receptor
ISE : Intronic Splicing Enhancer
ISS : Intronic Splicing Silencer
KO : Knock-Out
MAPT : microtubule-associated protein tau
MBNL1 : Muscleblind-like 1

Mfn2 : Mitofusin 2
miRNA : micro ARN
MTMR1 : Myotubularin-related protein 1
NCAM : Neural Cell Adhesion Molecule
NMD : Non-sense mRNA Decay
nPTB : neuronal Polypyrimidine Tract Binding Protein
NTP : Nucleotide Tri Phosphate
ORF : Open Reading Frame
PDGF : platelet-derived growth factor
PKC : Protein Kinase C
PKR : Protein Kinase R
PTB : Polypyrimidine Tract Binding Protein
RISC: RNA induced silencing complex
RIP : RNA Immunoprecipitation
RRM : RNA Recognition Motif
RT-PCR : Retro-transcriptase Polymerase Chain Reaction
RYR1 : Ryanodine Receptor 1
SELEX : Systematic Evolution of Ligands by EXponential enrichment
SERCA2 : sarco/endoplasmic reticulum Ca²⁺-ATPase
SHAPE : Selective 2'-hydroxylacylation analyzed by primer extension
shRNA : small hairpin RNA
snRNA : small nuclear RNA
snRNP : small nuclear Ribonucleoprotein
SRA : steroid receptor RNA activator
SV40 : Simian Virus 40
TNNT2 : Troponine T cardiaque
tRNA : ARN de Transfert
U2AF : U2 Auxilliary Factor
UV: Ultraviolets
ZAP : Zinc-Finger Antiviral Protein

TABLE DES MATIERES

| | |
|--|-----------|
| LISTE DES ABREVIATIONS | 3 |
| TABLE DES MATIERES | 7 |
| LISTE DES FIGURES | 11 |
| LISTE DES TABLEAUX | 13 |
| INTRODUCTION | 17 |
| I. Régulation de l'épissage alternatif : vers le déchiffrement du « code d'épissage »..... | 17 |
| A) Préambule | 17 |
| B) Mécanisme de la réaction d'épissage | 20 |
| 1) Des séquences consensus pour définir les introns | 20 |
| 2) Le spliceosome..... | 21 |
| 3) Variabilités des séquences régulatrices et influence sur l'épissage alternatif | 25 |
| C) Régulation de l'épissage alternatif..... | 27 |
| 1) Séquences activatrices et protéines SR | 27 |
| 2) Séquences inhibitrices et hnRNP | 29 |
| 3) Effets combinatoires entre les activateurs et répresseurs | 31 |
| 4) Les autres facteurs protéiques régulateurs de l'épissage alternatif | 32 |
| 5) La définition d'un « code d'épissage » | 34 |
| D) Nouveaux éléments régulateurs du « code d'épissage » | 35 |
| 1) Influence de la transcription..... | 35 |
| 2) Influence de la structure chromatinienne | 37 |
| 3) Influence des ARN non codants | 40 |
| 4) Influence des structures secondaires ARN | 41 |
| a) Régulation directe | 42 |
| Les structures qui inhibent l'épissage | 43 |
| Les structures qui activent l'épissage | 44 |
| b) Régulation indirecte | 45 |
| E) Analyse globale du « code d'épissage » et prédiction | 47 |
| II. La Dystrophie Myotonique de type I : une maladie caractérisée par un gain de fonction de l'ARN..... | 51 |
| A) Généralités | 51 |
| B) Aspects cliniques | 52 |
| 1) Forme bénigne | 54 |

| | | |
|---|---|-----------|
| 2) | Forme adulte..... | 54 |
| 3) | Forme juvénile | 54 |
| 4) | Forme congénitale | 54 |
| C) | Mécanismes physiopathologiques proposés dans DM1..... | 55 |
| 1) | Haploinsuffisance de la DMPK..... | 56 |
| 2) | Diminution d'expression des gènes adjacents au gène DMPK..... | 58 |
| 3) | Gain de fonction de l'ARN et altération de mécanismes liés au métabolisme de l'ARN..... | 60 |
| a) | Effet trans des répétitions CUG sur la transcription ou "RNA leaching" | 62 |
| b) | Réexpression des isoformes fœtales à l'origine des symptômes cliniques..... | 63 |
| | Gènes musculaires et anomalies musculaires/cardiaques : | 65 |
| | Récepteur à l'insuline et diabète : | 66 |
| | Canal chlore et myotonie : | 67 |
| | Symptômes cérébraux et anomalies neuronales:..... | 68 |
| c) | MBNL1 et sa dérégulation dans la pathologie..... | 68 |
| | Généralités | 68 |
| | Structure/fonction de la protéine MBNL1 | 69 |
| | Séquestration de MBNL1 et conséquence sur l'épissage alternatif | 70 |
| d) | Dérégulation de CUGBP1..... | 72 |
| e) | Dérégulation de la hnRNP H | 75 |
| f) | Dérégulation des micro ARN | 75 |
| g) | Les expansions CUG sont des cibles de Dicer | 77 |
| h) | Effet trans sur la traduction..... | 77 |
| D) | Remise en cause du modèle actuel..... | 78 |
| III. p68 et p72 : DEAD-box ARN hélicases | | 83 |
| A) | Généralités sur les hélicases | 83 |
| B) | Les six superfamilles d'hélicases ADN ou ARN | 83 |
| C) | Motifs conservés et fonctions associées..... | 85 |
| 1) | Motif Q (G-a-x-x-P-o-x-x-Q)..... | 86 |
| 2) | Motif I (A/G-X-X-G-X-G-K-T) | 87 |
| 3) | Motif II (DEAD) | 87 |
| 4) | Motif III (SAT) | 88 |
| 5) | Motif IV | 88 |
| 6) | Motif V | 89 |
| 7) | Motif VI (X-R-X-G-R-X-G-R)..... | 89 |

| | | |
|---------------------------------------|--|------------|
| D) | Les ARN hélicases p68/DDX5 et p72/DDX17..... | 90 |
| 1) | Généralités..... | 90 |
| 2) | Les différents rôles de p68 et p72..... | 93 |
| a) | Activité ATPase et hélicase de p68 et p72..... | 93 |
| b) | Rôle dans la transcription..... | 94 |
| c) | Rôle dans l'épissage..... | 95 |
| d) | Rôle de p68/72 dans les miRNA..... | 96 |
| e) | Rôle dans la différenciation myogénique et ostéoblastique..... | 97 |
| f) | Rôle dans la régulation du cycle cellulaire..... | 98 |
| g) | p68, p72 et leurs modifications post-traductionnelles..... | 99 |
| h) | Rôle dans le développement..... | 100 |
| i) | Implication de p68 et p72 dans les cancers..... | 100 |
| RESULTATS..... | | 107 |
| RESULTATS COMPLEMENTAIRES..... | | 137 |
| A) | Expression de p68 dans la DM1..... | 137 |
| B) | Absence de colocalisation de p68 dans les cellules DM1 et tissus de souris SXL..... | 138 |
| C) | Analyse structure-fonction de p68..... | 141 |
| D) | Interaction de p68 sur la tige-boucle de <i>TNNT2</i> | 143 |
| E) | Purification des partenaires protéiques de MBNL1..... | 145 |
| DISCUSSION..... | | 151 |
| A) | Identification de nouvelles protéines interagissant avec les expansions CUG caractéristiques de la Dystrophie Myotonique de type 1..... | 152 |
| 1) | Affinité spécifique pour les expansions CUG..... | 152 |
| 2) | Affinité suivant le type cellulaire et de l'état de différenciation..... | 152 |
| 3) | Identification et test des candidats putatifs..... | 153 |
| B) | Identification de p68 comme nouveau facteur impliqué dans la DM1 au niveau des expansions CUG..... | 154 |
| 1) | Colocalisation de p68 avec les répétitions CUG dans un modèle cellulaire mimant la Dystrophie Myotonique de type 1..... | 154 |
| 2) | Analyse structure-fonction de la colocalisation de p68..... | 155 |
| 3) | Absence de colocalisation de p68 dans des lignées cellulaires issues de biopsies de patients DM1, ni dans des tissus provenant de modèles de souris DM1..... | 156 |
| C) | p68 est un modificateur de l'interaction de MBNL1 avec les expansions CUG..... | 158 |
| 1) | p68 facilite l'interaction de MBNL1 sur les expansions CUG..... | 158 |
| 2) | L'activité hélicase de p68 est requise pour faciliter la fixation de MBNL1..... | 159 |

| | | |
|----|--|------------|
| 3) | Quel est le mécanisme d'action de p68 sur la fixation de MBNL1 ? | 160 |
| D) | p68 régule l'épissage alternatif de l'exon 5 du transcrit <i>TNNT2</i> , dérégulé dans la DM1 .. | 161 |
| 1) | p68 régule l'épissage alternatif d'un certain type de transcrits | 161 |
| 2) | p68 facilite la fixation de MBNL1 sur des sites intégrés dans une structure secondaire de type tige-boucle | 162 |
| 3) | L'effet de p68 sur l'épissage de <i>TNNT2</i> est dépendant de MBNL1 | 163 |
| E) | Modèle de régulation | 164 |
| F) | Perspectives... Quelques points à élucider ! | 167 |
| 1) | Existe-t-il d'autres cibles régulées par p68/MBNL1 et dérégulées dans la pathologie ? ... | 167 |
| 2) | p68 agit-il dans tous les cas comme un facilitateur de la fixation de MBNL1 ? | 168 |
| 3) | L'altération des complexes p68/MBNL1 pourrait-elle concerner d'autres voies que l'épissage alternatif dans la DM1? | 168 |
| 4) | L'ARN non-codant SRA pourrait-il être le chaînon manquant entre MBNL1 et p68 ? | 169 |
| | CONCLUSION | 173 |
| | REFERENCES BIBLIOGRAPHIQUES | 178 |
| | ANNEXES | 209 |

LISTE DES FIGURES

| | | |
|---------------------------|---|----|
| <u>Figure 1 :</u> | Séquences régulatrices consensus au sein des introns de type GU-AG | 20 |
| <u>Figure 2 :</u> | Les deux étapes de trans-estérification de la réaction d'épissage alternatif | 21 |
| <u>Figure 3 :</u> | Etapes précoces d'assemblage et désassemblage du spliceosome majeur | 22 |
| <u>Figure 4 :</u> | Eléments régulateurs et différents types d'épissage alternatif | 26 |
| <u>Figure 5 :</u> | Mécanismes de régulation de l'épissage par les protéines SR | 28 |
| <u>Figure 6 :</u> | Différents mécanismes d'inhibition de l'épissage par les protéines hnRNP | 30 |
| <u>Figure 7 :</u> | Mécanisme de régulations des facteurs antagonistes PTB et nPTB | 33 |
| <u>Figure 8 :</u> | Influence de la vitesse d'élongation par l'ARN polymérase II sur l'épissage alternatif | 36 |
| <u>Figure 9 :</u> | Marques épigénétiques principales correspondant à la méthylation des lysines sur l'extrémité N-terminale des histones H3 et H4 | 38 |
| <u>Figure 10 :</u> | Rôle de MALAT1 dans la régulation de l'épissage alternatif | 41 |
| <u>Figure 11 :</u> | Les structures qui inhibent l'épissage | 43 |
| <u>Figure 12 :</u> | Les structures qui activent l'épissage | 44 |
| <u>Figure 13 :</u> | Mécanisme de la régulation de l'épissage alternatif de l'exon 5 du transcrit codant la troponine T cardiaque (<i>TNNT2</i>) | 46 |
| <u>Figure 14 :</u> | Mécanisme putatif de la régulation de l'épissage alternatif de l'exon 4 du transcrit CD44 | 47 |
| <u>Figure 15 :</u> | Représentation schématique des résultats obtenus par Barash et al. concernant la redondance des sites consensus de facteurs d'épissage suivant les tissus | 49 |
| <u>Figure 16 :</u> | Représentation schématique du gène DMPK contenant la mutation responsable de la pathologie | 51 |
| <u>Figure 17 :</u> | Patients atteints de Dystrophie Myotonique de type 1 | 53 |
| <u>Figure 18 :</u> | Les expansions CTG interfèrent avec la régulation de l'expression génique à plusieurs niveaux | 55 |
| <u>Figure 19 :</u> | Possibles implications pathologiques de l'activité kinase de DMPK dans la Dystrophie Myotonique de type 1 | 57 |
| <u>Figure 20 :</u> | Organisation du locus DMPK humain | 59 |
| <u>Figure 21 :</u> | Visualisation des foci nucléaires dans des myoblastes humains | 61 |
| <u>Figure 22 :</u> | Dérégulation de l'épissage alternatif de <i>TNNT2</i> dans la DM1 | 65 |
| <u>Figure 23 :</u> | Dérégulation de l'épissage alternatif de <i>BIN1</i> dans la DM1 | 66 |
| <u>Figure 24 :</u> | Dérégulation de l'épissage alternatif du gène codant le récepteur à l'insuline dans la DM1 | 66 |
| <u>Figure 25 :</u> | Dérégulation de l'épissage alternatif du canal chlore dans la DM1 | 67 |

| | | |
|---------------------------|--|-----|
| <u>Figure 26 :</u> | Effet antagoniste des protéines CUGBP1 et MBNL1 dans la régulation de l'épissage alternatif | 68 |
| <u>Figure 27 :</u> | Dérégulation du facteur d'épissage MBNL1 dans la pathologie DM1 | 71 |
| <u>Figure 28 :</u> | Absence de colocalisation de CUGBP1 dans les tissus musculaires de patients DM1 | 73 |
| <u>Figure 29 :</u> | Augmentation de la quantité relative de CUGBP1 chez les patients DM1 | 74 |
| <u>Figure 30 :</u> | Modèle de la dérégulation du processing du pre-miR-1 dans la DM1 | 76 |
| <u>Figure 31 :</u> | Représentation schématique du modèle proposé par R.P. Junghans | 79 |
| <u>Figure 32 :</u> | Représentation schématique des différents motifs conservés des DEAD-box ARN hélicases et de leur interaction avec l'ATP et leur substrat ARN | 86 |
| <u>Figure 33 :</u> | Alignement des séquences protéiques DDX5 et DDX17 humaines et caractérisation des motifs conservés | 91 |
| <u>Figure 34 :</u> | L'inactivation conjointe de p68/p72 altère l'expression de gènes musculaires et la différenciation myogénique | 98 |
| <u>Figure 35 :</u> | Expression de la protéine p68 dans des extraits protéiques totaux de myoblastes et myotubes contrôles et DM1 | 137 |
| <u>Figure 36 :</u> | Absence de colocalisation de p68 avec les foci CUG dans des myoblastes DM1 | 139 |
| <u>Figure 37 :</u> | Absence de colocalisation de p68 avec les foci CUG dans tissus provenant de souris DM1 SXL | 139 |
| <u>Figure 38 :</u> | Liste des différents mutants principaux de p68 utilisés au cours de ma thèse | 140 |
| <u>Figure 39 :</u> | La mutation d'acides aminés dans les motifs Q et I abolit la colocalisation de p68 avec les foci | 141 |
| <u>Figure 40 :</u> | La délétion des extrémités N-terminale et C-terminale de p68 altèrent la colocalisation avec les foci CUG | 142 |
| <u>Figure 41 :</u> | La délétion de la région 52-76 suffit à reproduire la localisation complète de l'extrémité N-terminale | 143 |
| <u>Figure 42 :</u> | p68 Δ Ct2-CBP-His et GST-MBNL interagissent avec la tige-boucle de <i>TNNT2</i> en gel retard | 144 |
| <u>Figure 43 :</u> | p68 Δ Ct2-CBP-His semble interagir avec la tige-boucle de <i>TNNT2</i> sauvage ou mutante avec une affinité comparable | 144 |
| <u>Figure 44 :</u> | Mise au point de la purification des partenaires protéiques de MBNL1 | 145 |
| <u>Figure 45 :</u> | Modèle proposé par Berglund revisité avec l'intervention de p68 | 164 |
| <u>Figure 46 :</u> | Modèle proposé sur le rôle de p68 dans la Dystrophie Myotonique de type 1 | 166 |
| <u>Figure 47 :</u> | Prédiction d'une partie de la structure secondaire du pre-miR-1 humain par Sfold | 169 |
| <u>Figure 48 :</u> | Prédiction de la sous-structure STR7 de l'ARN non-codant SRA par Sfold | 170 |

LISTE DES TABLEAUX

| | | |
|---------------------------|--|----|
| <u>Tableau 1 :</u> | Liste des protéines SR | 28 |
| <u>Tableau 2 :</u> | Liste des protéines hnRNP | 30 |
| <u>Tableau 3 :</u> | Liste des facteurs d'épissage alternatif tissu-spécifiques | 33 |
| <u>Tableau 4 :</u> | Liste des symptômes présents chez les patients DM1 adulte suivant les tissus | 53 |
| <u>Tableau 5 :</u> | Liste des gènes dont l'épissage alternatif est dérégulé dans la DM1 suivant les tissus | 64 |
| <u>Tableau 6 :</u> | Liste des superfamilles d'hélicases | 84 |

INTRODUCTION

INTRODUCTION

I. Régulation de l'épissage alternatif : vers le déchiffrement du « code d'épissage »

A) Préambule

L'expression génique est le processus par lequel l'information contenue dans un gène est utilisée pour la synthèse d'un produit fonctionnel. Ce cheminement est composé d'étapes complexes : la transcription, la maturation, l'export de l'ARN mature au cytoplasme, la traduction en protéine et les modifications post-traductionnelles. Chaque étape nécessite l'intervention d'une machinerie ribonucléoprotéique complexe, cependant des études récentes suggèrent que chaque étape de l'expression génique est physiquement et fonctionnellement connectée avec la suivante, formant un processus continu (Kornbliht *et al.*, 2004).

Au cours de la transcription, l'ARN pré-messager subit plusieurs étapes de maturation avant son export dans le cytoplasme en tant qu'ARN messenger mature fonctionnel où il sera traduit en protéine.

L'une de ces étapes essentielles est l'épissage des ARN pré-messagers. Du fait de la structure fragmentée des gènes eucaryotes, la plupart des séquences codantes (ou exons) sont interrompues par de longues séquences non-codantes (ou introns) qui doivent être excisées afin de réunir les parties codantes dans un cadre de lecture correct. L'épissage est dit constitutif lorsqu'un exon est systématiquement inséré dans le transcrit mature. Il est à l'inverse alternatif lorsqu'un exon peut être soit inclus, soit exclu de l'ARN mature. Ainsi, l'épissage alternatif est un incroyable mécanisme contribuant de deux manières différentes à la vie cellulaire.

Tout d'abord de façon qualitative, l'épissage alternatif étant la source majeure à l'origine de la diversité du protéome. A l'opposé du dogme originel qui supposait qu'un gène codait une protéine unique, les données actuelles montrent que le nombre de protéines différentes est bien plus grand que le nombre de gènes chez les métazoaires. Plusieurs observations chez de nombreux eucaryotes ont montré qu'un gène peut avoir une douzaine, voire des milliers de produits fonctionnels. Le cas le plus extrême concerne le gène DSCAM chez la drosophile, qui peut potentiellement générer plus de 38 000 isoformes protéiques

différentes (Schmucker *et al.*, 2000). Récemment, des études protéomiques ont montré qu'une grande majorité de ces isoformes était effectivement exprimée en protéine (Tress *et al.*, 2008).

L'épissage alternatif peut également agir de façon quantitative, en réduisant l'expression génique par la voie du nonsense-mediated mRNA decay (NMD). En effet, au cours de l'épissage alternatif, la rétention de séquences introniques ou exoniques entraînant un décalage du cadre de lecture permettent l'insertion de codon de terminaison de la traduction prématuré. D'une manière générale, le codon stop authentique réside dans le dernier exon et n'est pas suivi par une jonction exon-exon. Les transcrits possédant des codons stop prématurés situés à plus de 50pb en amont de la dernière jonction exon-exon sont dans leur majorité dégradés par le NMD afin d'empêcher la production de protéines délétères (Nagy & Maquat, 1998). Un nombre important de facteurs d'épissage et de composants du spliceosome sont autorégulés par ce mécanisme. Les protéines PTB et SC35 régulent l'épissage alternatif de leur propre transcrit en jouant sur l'inclusion d'un exon alternatif contenant codon stop prématuré. Cet ARNm est dégradé par le mécanisme du NMD aboutissant en une baisse de production protéique (Sureau *et al.*, 2001; Wollerton *et al.*, 2004; Lareau *et al.*, 2007). Ceci crée une boucle d'autorégulation négative qui stabilise la concentration relative d'un facteur d'épissage au cours du temps.

L'épissage est un mécanisme ubiquitaire chez les eucaryotes et il existe très peu d'exemple d'épissage chez les bactéries (Edgell *et al.*, 2000) ou les archées (Watanabe *et al.*, 2002). L'importance de l'épissage alternatif peut être constatée par l'analyse de sa prévalence parmi l'arbre évolutionnaire eucaryote. En effet, l'épissage alternatif est plus abondant chez les eucaryotes multicellulaires que chez les eucaryotes unicellulaires et le pourcentage de gènes et d'exons subissant l'épissage alternatif est plus important chez les vertébrés que chez les invertébrés (Alekseyenko *et al.*, 2007).

Quels sont les éléments à l'origine de l'apparition de l'épissage alternatif au cours du temps ? L'une des réponses les plus probables pourrait résider dans l'apparition successive des séquences introniques, puis de l'épissage constitutif chez les eucaryotes.

Bien qu'il existe des introns dans les génomes de la plupart des eucaryotes, l'épissage alternatif est comme nous l'avons vu principalement présent chez les eucaryotes multicellulaires. La levure *Saccharomyces cerevisiae* possède des introns dans 3% de ses gènes (environ 250 introns) et seulement 6 de ses gènes contiennent deux introns (Barrass & Beggs, 2003). A l'opposé, dans une autre levure, *Schizosaccharomyces pombe*, 43% des gènes contiennent des introns, et une grande partie de ces gènes en possède plus d'un (Wood

et al., 2002). Cependant, aucune forme d'épissage alternatif n'a été décrite chez ces organismes. A l'inverse des introns chez les mammifères qui sont relativement longs (500-30000 nucléotides), les introns chez la levure sont courts, de l'ordre de 40-75 nucléotides chez *S. pombe* et 270 nucléotides chez *S. cerevisiae* (Wood *et al.*, 2002). Ceci pourrait suggérer que l'épissage alternatif puisse être préférentiellement observable dans les organismes possédant de grandes séquences introniques.

Le récent séquençage des génomes et ARN messagers de plusieurs organismes a facilité l'analyse à grande échelle des structures exon-intron et du type d'épissage pour de nombreux gènes. La conservation au cours de l'évolution de certaines séquences régulatrices parmi différents organismes indique que ces séquences subissent une pression de sélection et pourraient donc avoir des fonctions importantes.

Comment l'épissage constitutif a pu se diversifier et aboutir à de l'épissage alternatif ? Deux modèles non-contradictaires sont envisagés pour expliquer cette apparition au cours de l'évolution.

Le premier modèle suggère l'apparition de mutations conduisant à l'affaiblissement des sites consensus d'épissage ou au renforcement de sites cryptiques, ce qui confère à la machinerie d'épissage la possibilité d'inclure ou d'exclure un exon en particulier. Cette possibilité serait à l'origine de la production d'un nouveau transcrit possédant potentiellement une nouvelle fonction, sans compromettre le répertoire original de transcrits produits par ce gène (Sorek *et al.*, 2004a). En faveur de ce modèle, il a été montré que les exons alternatifs possèdent des sites d'épissage plus faibles que les exons constitutifs, ce qui pourrait réduire la détection de ces exons par le spliceosome et entraîner à l'épissage alternatif (Stamm *et al.*, 1994; Carmel *et al.*, 2004; Sorek *et al.*, 2004a).

Le second modèle quant à lui propose que l'évolution des protéines régulatrices de l'épissage, telles que les hnRNP ou les protéines SR (qui seront abordées par la suite), appliqueraient une pression sélective sur les exons constitutifs pour qu'ils deviennent alternatifs. Par exemple, la fixation de protéines inhibitrices à proximité d'un exon constitutif réduit la sélection de cet exon, ce qui diminuerait la pression de sélection au niveau de ses sites d'épissage et résulterait en l'apparition de mutations qui fragiliseraient ces sites (Ast, 2004). Ce second modèle s'appuie sur une adaptation de la théorie de l'évolution des facteurs de transcription (Moss, 2002; Ast, 2004) et n'a pas fait l'objet d'études scientifiques plus poussées et reste donc à l'état d'hypothèse.

A sa découverte, l'épissage alternatif était considéré comme un événement limité, touchant dans le meilleur des cas quelques transcrits impliqués dans la différenciation sexuelle chez la drosophile. En 1994, Philip Sharp, lauréat du Prix Nobel de médecine pour la découverte de ce mécanisme, estimait à moins de 5% le nombre de gènes humains régulés par épissage alternatif (Sharp, 1994). Depuis ce pourcentage n'a cessé d'augmenter chaque année. L'équipe de Blencowe a récemment estimé, en combinant des techniques de mRNA-Seq et des bases de données EST-cDNA, que 95% des gènes seraient régulés par épissage alternatif, ce qui représenterait plus de 100 000 événements d'épissage alternatif distincts dans les tissus humains (Pan *et al.*, 2008).

B) Mécanisme de la réaction d'épissage

1) Des séquences consensus pour définir les introns

Les gènes eucaryotes étant constitués d'exons interrompus par de grandes séquences introniques, il est nécessaire pour la machinerie d'épissage d'identifier de façon précise tous ces exons. Cette reconnaissance nécessite au moins quatre types de séquences, représentées dans la Figure 1 : les sites 5' et 3' d'épissage, respectivement appelés « site donneur » et « site accepteur », situés en amont et en aval de l'intron au niveau de la jonction intron/exon, un « point de branchement », localisé en amont du site 3' d'épissage (18 à 40 nucléotides en moyenne) et une séquence riche en pyrimidine située entre le point de branchement et le site 3' accepteur. Les sites 5' et 3' d'épissage sont définis par des séquences consensus courtes (9 bases pour le site 5' et une quinzaine de bases pour le site 3'). Par contre, le point de branchement possède une séquence peu conservée (chez l'humain : YNYCRAY). La distance variable entre le point de branchement et le site 3' d'épissage et sa faible conservation en font un élément peu fiable pour la prédiction bioinformatique des sites 3' d'épissage réels car ils sont générateurs de nombreux faux-positifs.

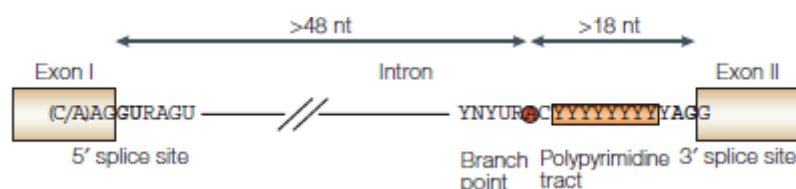


Figure 1 : Séquences régulatrices consensus au sein des introns de type GU-AG. L'intron situé entre les deux exons est caractérisé par un site 5' et 3' d'épissage, un point de branchement et une séquence riche en pyrimidines. Ces séquences permettent la détection des introns et leur épissage par le spliceosome (Matlin *et al.*, 2005).

La réaction d'épissage se déroule en deux étapes successives, impliquant chacune une réaction de transestérification (Figure 2). Lors de la première étape, le groupement 2' hydroxyle de l'adénosine du site de branchement attaque la liaison phosphodiester qui sépare le dernier nucléotide de l'exon du premier nucléotide de l'intron. Cette réaction de transestérification aboutit à la formation d'un exon libre au niveau de l'extrémité 5' d'épissage de l'intron et une structure intermédiaire en lasso contenant le deuxième exon. La seconde étape consiste en l'attaque de la liaison phosphodiester qui sépare le dernier nucléotide de l'intron et le premier nucléotide de l'exon aval par l'exon amont. Ceci aboutit à la production d'un ARN messager et d'un intron sous forme de lasso.

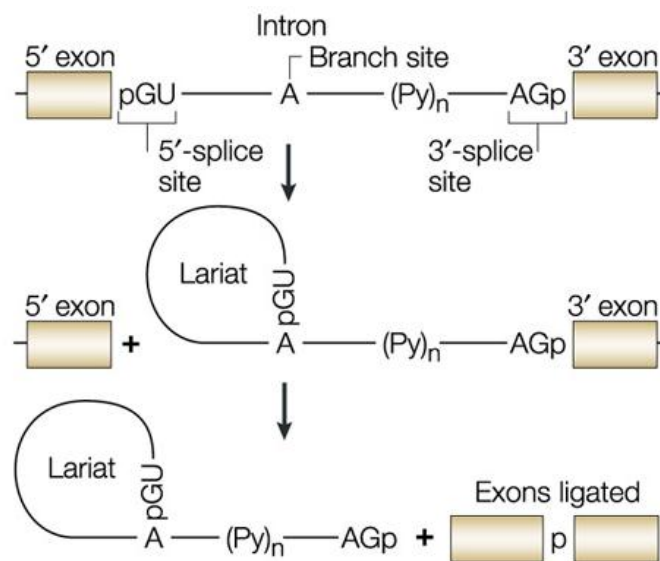


Figure 2 : Les deux étapes de transestérification de la réaction d'épissage alternatif. La première étape produit deux intermédiaires d'épissage : l'exon 5' libre et l'exon 3' lié à l'intron structuré en lasso. La seconde étape lie les deux exons et libère le lasso (Pagani & Baralle, 2004).

2) Le spliceosome

L'épissage est assuré par un complexe ribonucléoprotéique appelé spliceosome (de l'anglais «splicing» signifiant épissage). Cette machinerie d'épissage est composée de plusieurs particules ribonucléoprotéiques appelées snRNP et d'autres facteurs d'épissage auxiliaires non-snRNP. Il existe à l'heure actuelle deux types de spliceosomes : le spliceosome majeur qui épisse les introns de type U2 et le spliceosome mineur qui épisse les introns rares de type U12 (Sharp & Burge, 1997; Burge *et al.*, 1998). Par soucis de simplification, j'aborderai uniquement par la suite le spliceosome majeur.

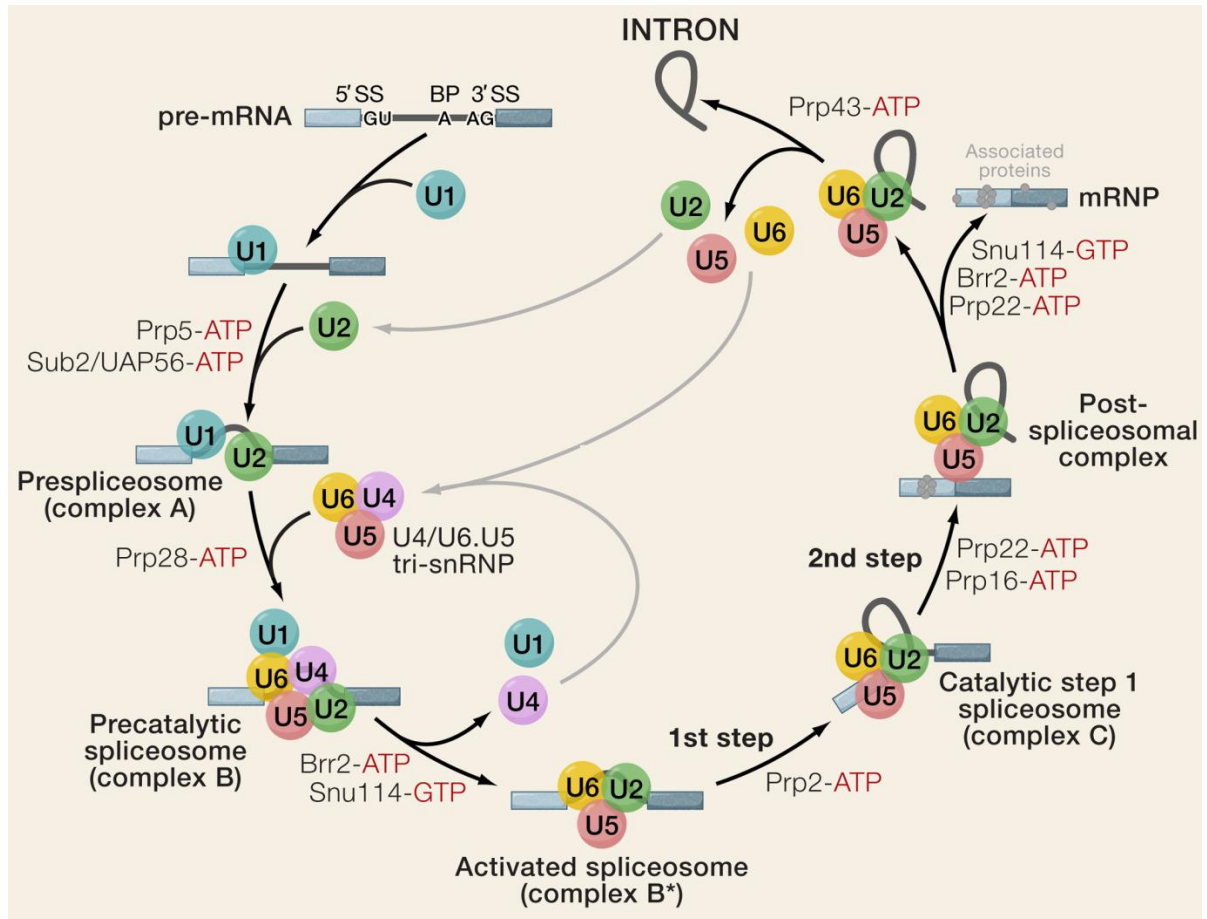


Figure 3 : *Etapes précoces d'assemblage et désassemblage du spliceosome majeur.* Les exons sont représentés par des rectangles bleus clairs et foncés. Au niveau de l'intron, le site 5' d'épissage est reconnu par des interactions ARN/ARN par la snRNP U1. Le site accepteur 3' qui contient un site riche en pyrimidines est reconnu par les protéines U2AF65/35 et SF1. Ces facteurs vont s'associer à d'autres protéines pour former un complexe précoce « E » qui pont l'intron et rapproche ainsi les sites donneurs et accepteurs. U2AF recrute la snRNP U2 et permet selon un mécanisme ATP-dépendant la fixation de cette snRNP au niveau du point de branchement situé en aval du site 3' d'épissage, formant ainsi le complexe « A ». La formation d'un complexe final est complétée par la fixation ATP-dépendante des snRNPs U4, U5 et U6. Le départ des snRNPs U1 et U4 conduit à la forme catalytique du spliceosome qui par transestérification, raccorde les sites accepteurs et donneurs, permettant la ligation des deux exons et l'excision de l'intron sous la forme d'un lasso (Wahl et al., 2009).

Les analyses protéomiques effectuées sur le spliceosome ont permis d'établir une liste non-exhaustive de plus de 300 protéines, illustrant la complexité de la machinerie d'épissage (Zhou *et al.*, 2002). Le spliceosome majeur se compose de 5 snRNPs : snRNP U1, U2, la di-snRNP U4/U6 et la snRNP U5. Ces particules ribonucléoprotéiques sont constituées d'un petit ARN (snRNA) auquel s'associent un nombre variable de protéines. Il existe deux familles de protéines associées :

- Les protéines Sm (SmB/B', D1, D2, E, F, G), au nombre de sept, communément appelées « Sm core proteins », qui s'associent en anneau heptamérique autour des snRNAs U1, U2, U4 et U5 (Walke *et al.*, 2001).
- Les protéines spécifiquement associées à un snRNA en particulier (comme U1A, U1C et 70K, caractéristiques de la snRNP U1 ou les protéines de type SF de la snRNP U2) (Will & Lührmann, 2010).

La machinerie d'épissage contient également un ensemble de protéines non snRNP telles-que les protéines U2AF65/35 (« U2 snRNP auxiliary factors »), SF1 (« splicing factor 1 »), les protéines SR (« serine-arginine rich proteins ») et les hnRNPs (« heterogeneous nuclear ribonucleoproteins »).

L'assemblage des complexes du spliceosome sur l'ARN pré-messager pour les réactions d'épissage se déroule en plusieurs étapes successives (Figure 3) (Wahl *et al.*, 2009). Cet assemblage s'opère grâce à des interactions ARN-ARN, protéines-ARN et protéine-protéine qui ont lieu au niveau des sites 5' et 3' d'épissage et du site de branchement (Reed, 2000).

La snRNP U1 intervient dans l'identification du site 5' d'épissage et dans la mise en place de la machinerie d'épissage formant ainsi le complexe E (Seraphin & Rosbash, 1989) par un appariement de quelques nucléotides entre le snRNA et le site d'épissage. Ces liaisons sont stabilisées par les protéines auxiliaires associées au snRNP U1 (Zhang & Rosbash, 1999; Puig *et al.*, 1999). La snRNP U1 facilite l'association du facteur d'épissage U2AF (U2 associated factor) avec la région riche en pyrimidine (Brow, 2002). U2AF est une protéine composée de deux sous-unités, U2AF65 qui se liera à la protéine SF1 et U2AF35 qui se liera au site 3' d'épissage. Ce duplexe U2AF est nécessaire pour la fixation ultérieure et la stabilisation de la snRNP U2 sur l'ARN pré-messager (Reed, 1996).

Lors de l'étape suivante, le snRNP U2 interagit de manière stable avec le site de branchement pour former le complexe A. La stabilité de l'interaction, en dehors du duplexe U2AF, dépend de la fixation de plusieurs facteurs d'épissage (SF3a, SF3b...) au niveau du site de branchement ou de la région riche en pyrimidines, formant le complexe A. Ensuite, le

complexe snRNP U4/U6 s'associe au snRNP U5 et ces trois snRNP vont s'assembler avec les snRNP U1 et U2 pour former le complexe B (Moore & Sharp, 1993).

Des réarrangements structuraux majeurs vont mener à la formation du complexe B*. L'interaction entre les snRNP U4 et U6 est rompue et la snRNP U6 s'apparie avec le site donneur via son snRNA déplaçant ainsi la snRNP U1. Une région du snRNA U6 va former une boucle interne qui est impliquée dans la réaction de catalyse de l'épissage.

Le snRNA U5 interagit avec des nucléotides de l'exon du site donneur mais aussi avec des nucléotides de l'exon proche du site accepteur (Madhani & Guthrie, 1994). Tous ces réarrangements provoquent la déstabilisation et la dissociation des snRNP U1 et U4 du spliceosome, rendant le complexe B* actif pour procéder aux réactions d'épissage (Wahl *et al.*, 2009). Suite à cela, le spliceosome effectue la première réaction de transestérification générant le complexe C, puis subit un autre réarrangement pour en effectuer la deuxième (Query & Konarska, 2006) Une fois les deux réactions opérées, le spliceosome se dissocie, libérant l'ARN messager et les différentes snRNPs.

Le remodelage dynamique des interactions ARN-ARN, ARN-protéine et protéine-protéine durant les étapes successives d'assemblage du spliceosome nécessitent l'intervention de facteurs externes permettant une régulation minutieuse des cascades enzymatiques aboutissant à la formation de l'ARN messager mûré. Plusieurs enzymes parmi lesquelles les DExD/H-box ATPases/helicases ARN-dépendant et les peptidyl-propyl cis/trans isomérases (PPIases) fonctionnent conjointement avec des modifications post-traductionnelles (par exemple des phosphorylations) de composants du spliceosome pour diriger ces remodelages. Huit de ces protéines DExD/H-box (Sub2/UAP56, Prp5, U5-100K/Prp28, Brr2/U5-200K, Prp2, Prp16, Prp22 et Prp43) sont conservées chez les levures et l'homme (Wahl *et al.*, 2009) et agissent sur certaines des étapes de l'épissage pour faciliter les transitions entre les réseaux d'interactions mutuellement exclusifs des complexes ribonucléoprotéiques (Madhani & Guthrie, 1994).

En délogeant des facteurs, ARN ou protéine, ces protéines pourraient permettre à une région particulière de l'ARN pré-messager de s'engager avec d'autres facteurs, permettant ainsi le déroulement des étapes successives. Bien que toutes les protéines préalablement citées soient conservées au cours de l'évolution démontrant la nature conservée des événements de remodelage qu'elles facilitent, au moins trois autres DExD/H-box, sans paralogues identifiés chez la levure, sont trouvées dans le spliceosome humain (KIAA0560, Abstrakt, DDX35) et n'ont à ce jour aucun rôle fonctionnel identifié (Hirose *et al.*, 2006; Wahl *et al.*, 2009). Il reste à déterminer si ces trois autres facteurs sont également impliqués

dans des étapes de remodelage qui pourraient être spécifiques au spliceosome humain.

Chez les mammifères, les protéines SR sont aussi impliquées dans l'étape d'assemblage du spliceosome car elles faciliteraient l'incorporation du complexe U4/U5/U6 dans les spliceosomes (Rosciigno & Garcia-Blanco, 1995) et stimuleraient l'appariement entre les bases nucléotidiques de snRNA U2 et de snRNA U6 (Tarn *et al.*, 1995).

3) Variabilités des séquences régulatrices et influence sur l'épissage alternatif

Les exons chez l'homme ont une taille d'environ 120 nucléotides et moins de 5% sont plus grands que 250 nucléotides. D'après des analyses statistiques d'exons alternatifs de type 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 (exons inférieurs à 30 bases) (Sorek *et al.*, 2004b). Cependant, la taille des exons n'est pas la seule caractéristique permettant l'identification des exons alternatifs et d'autres séquences régulatrices dites séquences « cis » ont un rôle majeur dans la reconnaissance de ceux-ci.

Les séquences « cis » se distinguent des séquences consensus définissant les sites d'épissage 5' et 3' et le point de branchement. Elles correspondent à l'ensemble des séquences régulatrices permettant à la machinerie d'épissage de faciliter la reconnaissance des sites d'épissage réels tout en évitant la reconnaissance de nombreux sites cryptiques (Figure 4A). Il existe deux types de séquences régulatrices qui influent sur l'inclusion ou l'exclusion de l'exon alternatif : les séquences activatrices et les séquences inhibitrices . Ces séquences peuvent être exoniques ou introniques et sont appelées ESE (Exonic Splicing Enhancer), ISE (Intronic Splicing Enhancer), ESS (Exonic Splicing Silencer) et ISS (Intronic Splicing Silencer).

Dans le cadre de l'épissage alternatif, il est nécessaire qu'un exon soit inclus dans une certaine condition physiologique et exclu dans une autre (Figure 4B pour les différentes formes d'épissage alternatif), ce qui illustre le besoin d'une balance délicate entre l'ensemble des éléments régulateurs.

Ces éléments d'épissage auxiliaires ont des séquences très variables et sont des sites d'interaction de protéines régulatrices (ou facteurs « trans ») qui régulent l'épissage alternatif.

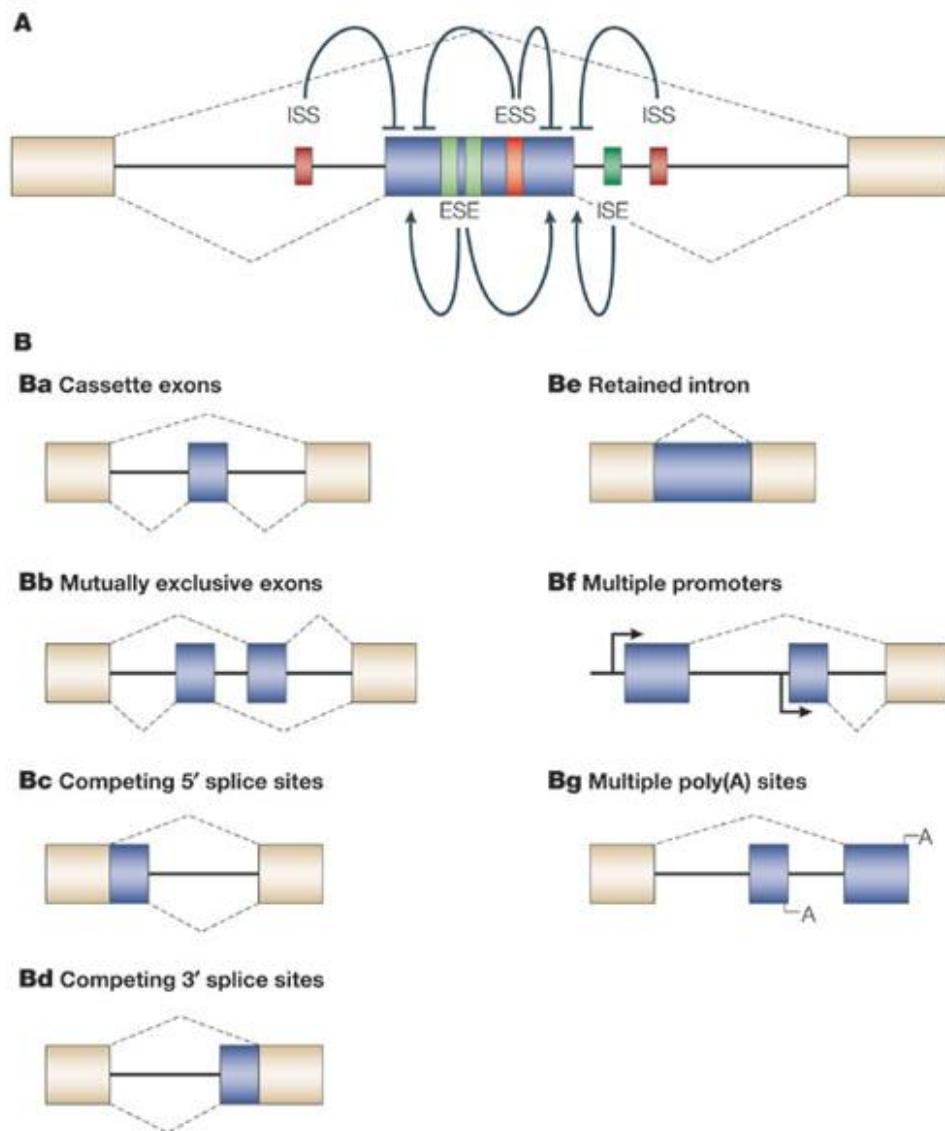


Figure 4: Eléments régulateurs et différents types d'épissage alternatif. *A.* Au côté des sites donneurs et accepteurs localisés au niveau des jonctions exon-intron, d'autres éléments localisés au niveau d'introns ou d'exons exercent un effet stimulateur (ISE/ESE) ou répresseur (ISS/ESS) sur l'épissage alternatif. Les séquences régulatrices activatrices peuvent activer un site adjacent ou à s'opposer à l'action des séquences suppressives et inversement. L'inclusion ou l'exclusion de l'exon alternatif est déterminée par une balance entre les concentrations relatives des facteurs activateurs et répresseurs. *B.* Les exons de type "cassette" (**Ba**) correspondent à des exons pouvant être indépendamment inclus ou exclus de l'ARN mature. Ils peuvent par la suite être subdivisés en deux catégories (constitutif ou cryptique). Les exons mutuellement exclusifs (**Bb**) impliquent la sélection d'un unique exon dans la forme mature parmi plusieurs exons. Les exons définis par des sites d'épissage 5' ou 3' alternatifs (**Bc** et **Bd**) contiennent plusieurs sites d'épissage différents qui permettent l'insertion d'un exon plus ou moins long selon le site utilisé. La rétention d'intron (**Be**) correspond à l'insertion d'un intron généralement épissé dans l'ARN mature. L'utilisation de promoteurs (**Bf**) ou de sites polyA (**Bg**) alternatifs permet l'expression de transcrits possédant des régions 5' et 3' non traduites spécifiques et qui peuvent ainsi modifier la stabilité du transcrit (Matlin et al., 2005).

En effet, les sites d'épissage alternatif sont souvent en compétition et ont de ce fait une séquence ou une architecture qui n'est pas aussi optimale que celle d'un exon constitutif qui sera inclus dans tous les cas. Cependant, pour que l'exon ne soit pas exclu en permanence, cette faiblesse de l'exon doit être compensée par des séquences de soutien. Par ailleurs, il peut exister des séquences spécifiques qui sont rencontrées uniquement dans des transcrits et des conditions physiologiques particulières.

Il n'est souvent pas aisé de faire une distinction formelle, au niveau du mécanisme, entre l'épissage constitutif et l'épissage alternatif. La force relative des sites d'épissage influence la fréquence avec laquelle un exon est sélectionné. Cependant, une paire de site fort d'épissage n'est pas suffisante pour définir un exon. En effet, de nombreux « pseudo-exons » qui ont des sites d'épissage à prédiction forte ne sont pas épissés. Des analyses récentes indiquent que l'enrichissement relatif en séquences « cis » auxiliaires, activatrices ou inhibitrices, aiderait à distinguer un exon « authentique » d'un pseudo-exon (Wang *et al.*, 2004; Zhang & Chasin, 2004).

C) Régulation de l'épissage alternatif

1) Séquences activatrices et protéines SR

Les protéines SR sont une famille de protéines qui possède un rôle important dans l'épissage tant constitutif qu'alternatif des ARN pré-messagers chez les métazoaires. Ces protéines sont caractérisées par le fait qu'elles possèdent un domaine RS (Arginine-Serine containing domain). Elles sont impliquées dans de nombreuses étapes de l'épissage, par leur fixation majoritaire sur des séquences activatrices exoniques (ESE) via leur motif de reconnaissance à l'ARN (RRM) (Kohtz *et al.*, 1994; Wu & Maniatis, 1993). Toutes les protéines SR possèdent des caractéristiques communes, notamment la présence de domaines hautement conservés (voir la liste des protéines SR en Tableau 1). La plupart des protéines SR interviennent également dans l'export des ARN, le Nonsense mRNA Mediated Decay (NMD) et la traduction (Huang *et al.*, 2003; Zhang & Krainer, 2004; Sanford *et al.*, 2004). Cependant, ces protéines ont une activité propre, en effet l'inactivation de chacune de ces protéines entraîne une létalité dans les souris knockout démontrant l'absence d'une complémentarité de leurs fonctions (Ring & Lis, 1994; Ding *et al.*, 2004a; Li *et al.*, 2005).

| Name* | Domains | Binding sequence | Target genes |
|------------------------------|-------------------------|--------------------|---|
| Canonical SR proteins | | | |
| SRp20 (SFRS3) | RRM and RS | GCUCUCUUC | SRP20, CALCA and INSR |
| SC35 (SFRS2) | RRM and RS | UGCUGUU | ACHE and GRIA1-GRIA4 |
| ASF/SF2 (SFRS1) | RRM, RRMH and RS | RGAAGAAC | HIPK3, CAMK2D, HIV RNAs and GRIA1-GRIA4 |
| SRp40 (SFRS5) | RRM, RRMH and RS | AGGAGAAGGA | HIPK3, PRKCB and FN1 |
| SRp55 (SFRS6) | RRM, RRMH and RS | GGCAGCACCU | TNNT2 and CD44 |
| SRp75 (SFRS4) | RRM, RRMH and RS | GAAGGA | FN1, E1A and CD45 |
| 9G8 (SFRS7) | RRM, zinc finger and RS | (GAC) _n | TAU, GNRH and SFRS7 |
| SRp30c (SFRS9) | RRM, RRMH and RS | CUGGAUU | BCL2L1, TAU and HNRNPA1 |
| SRp38 (FUSIP1) | RRM and RS | AAAGACAAA | GRIA2 and TRD |
| Other SR proteins | | | |
| SRp54 | RRM and RS | ND | TAU |
| SRp46 (SFRS2B) | RRM and RS | ND | NA |
| RNPS1 | RRM and Ser-rich | ND | TRA2B |
| SRp35 | RRM and RS | ND | NA |
| SRp86 (SRp508 and SFRS12) | RRM and RS | ND | NA |
| TRA2α | RRM and two Arg-rich | GAAARGARR | dsx |
| TRA2β | RRM and two RS | (GAA) _n | SMN1, CD44 and TAU |
| RBM5 | RRM and RS | ND | CD95 |
| CAPER (RBM39) | RRM and RS | ND | VEGF |

Tableau 1 : Liste des protéines SR. Les protéines SR sont caractérisées par des domaines protéiques de types RRM (RNA Recognition Motif), RS (Arginine-Serine repeat-containing) et RRMH (RRM Homologue). Les séquences consensus d'interaction, ainsi que des cibles de ces protéines sont indiquées dans les deux colonnes à droite (Chen & Manley, 2009).

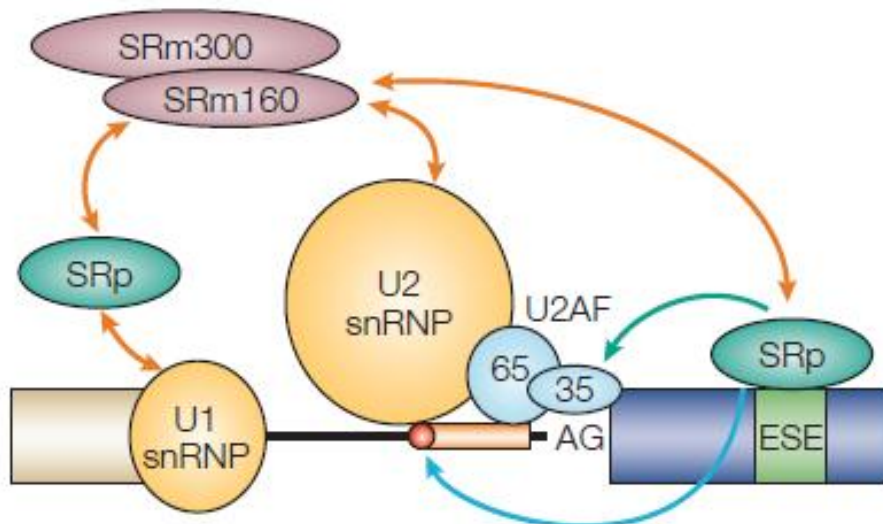


Figure 5 : Mécanismes de régulation de l'épissage par les protéines SR. Les protéines SR interagissent avec des séquences exoniques activatrices (ESE) et recrutent le duplex U2AF au niveau du site 3' d'épissage et la snRNP U1 au niveau du site 5' d'épissage. Elles interagissent également avec les co-activateurs SRm160 et SRm300 et participent au recrutement de la tri-snRNP U4/U5/U6 (non indiqué) (Matlin et al., 2005).

Concernant l'épissage, ces protéines ont un rôle primordial dans la facilitation de la reconnaissance des sites d'épissage (Figure 5). Elles aideraient par exemple au recrutement de la snRNP U1 au niveau du site 5' d'épissage et du complexe U2AF et de la snRNP U2 au niveau du site 3' d'épissage en interagissant directement avec ces protéines (Zuo & Maniatis, 1996; Bourgeois *et al.*, 1999; Feng *et al.*, 2008). Ces interactions sont régulées par les domaines RS (Arginine-Serine containing domain) qui doivent être correctement phosphorylés ou déphosphorylés (Wu & Maniatis, 1993; Kohtz *et al.*, 1994).

Les protéines SR fonctionnent également en coopération avec d'autres facteurs activateurs d'épissage, comme par exemple transformer 2 (Tra2) et les protéines de la matrice nucléaire SRm160 ou SRm300, formant des multi-complexes régulateurs sous la forme d'interactions entre les domaines SR (Tacke & Manle, 1999; Blencowe *et al.*, 2000). Enfin, comme nous l'avons vu précédemment, les protéines SR sont également impliquées dans l'assemblage du spliceosome de façon ARN-indépendante en facilitant l'incorporation du complexe U4/U5/U6 (Roscigno & Garcia-Blanco, 1995).

2) Séquences inhibitrices et hnRNP

Les séquences inhibitrices (ESS ou ISS) ont des séquences très variables et certaines interagissent avec la famille des heterogeneous nuclear ribonucleoproteines (hnRNP) dont une liste non-exhaustive est présentée dans le tableau 2. Ce groupe de protéines est très hétérogène au point de contenir des protéines impliquées dans des mécanismes aussi divers que la traduction ou la stabilité des télomères (Lynch *et al.*, 2005; Zhang *et al.*, 2006; Lee *et al.*, 2010a). Ces protéines contiennent des domaines de fixation à l'ARN de type RRM et KH et des domaines auxiliaires impliqués dans les interactions protéine-protéine et agissent majoritairement comme des répresseurs (Krecic & Swanson, 1999). D'une manière générale, les hnRNP fixées sur les ESS ou ISS empêchent par encombrement stérique l'intervention de facteurs d'épissage constitutif ou alternatif en rendant les exons, sites d'épissage ou ESE/ISE inaccessibles.

Plusieurs types de mécanismes ont été découverts et quatre d'entre eux sont représentés dans la Figure 6.

| Name | Other names | Domains* | Binding sequences | Target genes |
|----------|-------------|----------------------|-------------------|---|
| hnRNP A1 | NA | RRM, RGG and G | UAGGGA/U | SMN2 and RAS |
| hnRNP A2 | NA | RRM, RGG and G | (UUAGGG)n | HIV tat and IKBKAP |
| hnRNP B1 | | | | |
| hnRNP C1 | AUF1 | RRM | U rich | APP |
| hnRNP C2 | | | | |
| hnRNP F | NA | RRM, RGG and GY | GGGA and G rich | PLP, SRC and BCL2L2 |
| hnRNP G | NA | RRM and SRGY | CC(A/C) and AAGU | SMN2 and TMP1 |
| hnRNP H | DSEF1 | RRM, RGG, GYR and GY | GGGA and G rich | PLP, HIV tat and BCL2L1 |
| hnRNP H' | | | | |
| hnRNP I | PTB | RRM | UCUU and CUCUCU | PTB, nPTB, SRC, CD95, TNNT2, CALCA and GRIN3B |
| hnRNP L | NA | RRM | C and A rich | NOS and CD45 |
| hnRNP LL | SRRF | RRM | C and A rich | CD45 |
| hnRNP M | NA | RRM and GY | ND | FGFR2 |
| hnRNP Q | NA | RRM and RGG | ND | SMN2 |

Tableau 2 : Liste des protéines hnRNP. Les protéines hnRNP sont principalement caractérisées par des domaines protéiques de types RRM (RNA Recognition Motif) et RGG. Les séquences consensus d'interaction, ainsi que des cibles de ces protéines sont indiquées (Chen & Manley, 2009).

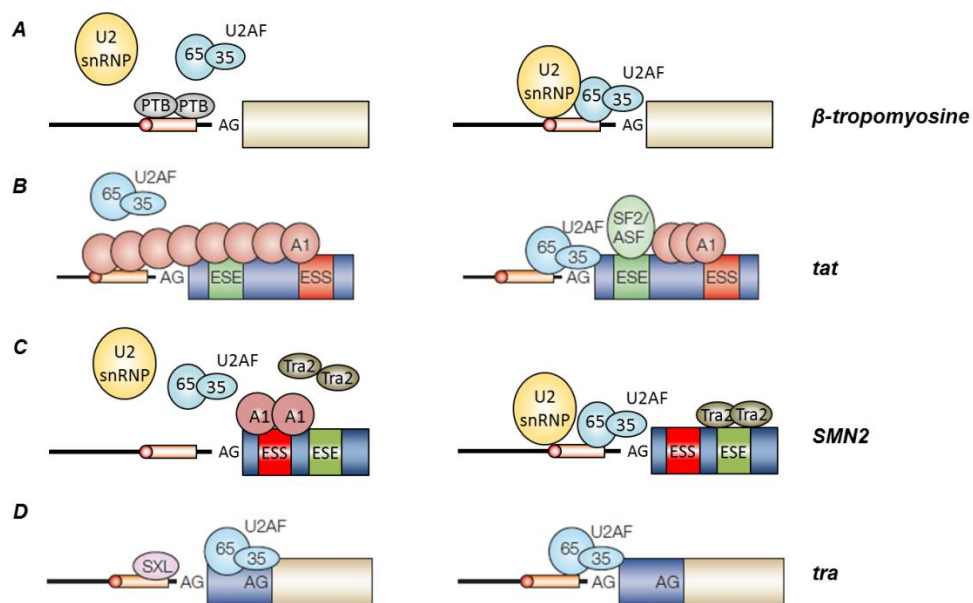


Figure 6 : Différents mécanismes d'inhibition de l'épissage par les protéines hnRNP. **A.** La protéine PTB interfère avec la fixation de U2AF65 sur la région riche en pyrimidines et empêche l'association de la snRNP U2 sur le point de branchement de l'exon 6B de l'ARN pré-messager codant la β -tropomyosine. **B.** L'inclusion de l'exon 3 de l'ARN pré-messager tat de HIV1 est déterminée par la multimérisation ou non de la hnRNP A1. Celle-ci est bloquée stériquement par l'interaction de SF2/ASF avec la région activatrice amont. **C.** La protéine hnRNP A1 interagit avec une séquence inhibitrice (ESS) localisée dans l'exon 7 de SMN 2 et inhibe la formation ou la stabilisation du complexe snRNP U2 directement ou indirectement par le blocage de l'interaction de Tra2 sur sa séquence spécifique localisée en aval. **D.** La répression du site 3' d'épissage du transcrit tra implique l'interaction de la protéine régulatrice SXL avec une séquence inhibitrice intronique (ISS) intégrée dans la région riche en pyrimidines et empêche la fixation du duplex U2AF. Ceci aboutit en la sélection du site 3' alternatif localisé en aval (Matlin et al., 2005).

L'un des exemples est la protéine polypyrimidine tract binding protein (PTB/hnRNP I) qui interagit avec la séquence riche en polypyrimidine, bloquant l'accès du duplex U2AF au niveau de l'exon alternatif 6B du transcrit β -tropomyosine (Sauliere *et al.*, 2006). D'autres types de régulations par les protéines hnRNP peuvent être trouvés dans les transcrits *tat* ou *SMN2*.

Il existe également des protéines hnRNP spécifique d'une espèce donnée, comme la protéine Sex Lethal (SXL) chez la drosophile. A la différence des mâles, les femelles produisent une protéine déterminant le sexe, la protéine SXL. Cette protéine se fixe sur la région riche en pyrimidine située en amont du site 3' d'épissage de l'ARN pré-messager tra, bloquant l'interaction de U2AF. Un autre site 3' en aval de l'exon alternatif, qui fixe U2AF et non SXL est alors sélectionné (Valcárcel *et al.*, 1993) entraînant la production de la protéine Tra exprimée uniquement chez la femelle. Chez le mâle, l'utilisation du site 3' amont entraîne l'insertion d'un codon stop prématuré et la dégradation de l'ARNm par le NMD, et par conséquent l'absence de production de la protéine Tra.

Bien qu'elles soient majoritairement considérées comme des répresseurs de l'épissage, il existe des cas où les hnRNP peuvent stimuler l'utilisation d'un site alternatif, principalement par des séquences inhibitrices exoniques. C'est le cas pour les sites 5' d'épissage faibles des transcrits eNOS ou FAS (Förch *et al.*, 2000; Hui *et al.*, 2003).

3) Effets combinatoires entre les activateurs et répresseurs

L'épissage d'un ARN pré-messager est fréquemment contrôlé par des effets combinatoires ou compétitifs des facteurs activateurs et répresseurs. La décision finale qui détermine si l'exon alternatif sera inclus ou exclu est déterminée par la concentration relative ou l'activité de ces facteurs antagonistes, qui sont souvent des protéines SR et des hnRNP (Mayeda *et al.*, 1993; Zhu & Krainer, 2000).

L'un des exemples majeurs concerne la régulation de l'épissage alternatif de l'exon 3 du pré-messager *Tat* de HIV par les facteurs antagonistes hnRNP A1 et ASF/SF2 (Figure 6C). La protéine inhibitrice hnRNP A1 se fixe sur un ESS, entraînant sa nucléation jusqu'aux séquences régulatrices introniques en amont de l'exon alternatif. Cette multimérisation empêche la fixation du duplexe U2AF et aboutit en l'exclusion de l'exon. La protéine

ASF/SF2 peut interagir avec un ESE en amont de l'ESS et bloquer la nucléation de hnRNP A1, permettant par la suite la fixation de U2AF65/35.

D'autres exemples peuvent être également trouvés dans des transcrits musculaires. En effet, la protéine SR 9G8 et les hnRNP F/H régulent l'épissage alternatif de l'exon 2 de l' α -tropomyosine par une compétition sur une même séquence régulatrice (Crawford & Patton, 2006). De plus, la hnRNP A1 et les protéines SR ASF/SF2 et SC35 ont des fonctions antagonistes sur la régulation de l'épissage alternatif de l'exon 6B du transcrit β -tropomyosine (Expert-Bezançon *et al.*, 2004).

Les protéines SR et les hnRNPs sont exprimées de façon ubiquitaire mais leur expression est variable suivant les tissus. Un changement dans leur stœchiométrie peut engendrer des différences d'épissage alternatif tissu-spécifique (Ge & Manley, 1990; Cáceres *et al.*, 1994). De plus, l'activité des protéines se liant à l'ARN peut être régulée par des régulations post-traductionnelles (phosphorylation ou séquestration dans un compartiment cellulaire en réponse au stress...) qui peuvent « transformer » un facteur répresseur en facteur activateur (Biamonti & Cáceres, 2009).

4) Les autres facteurs protéiques régulateurs de l'épissage alternatif

L'épissage alternatif a un rôle déterminant dans la définition de la spécificité tissulaire. De récentes études à haut-débit ont montré que plus de 50% des isoformes issues d'un épissage alternatif sont différemment exprimées suivant les tissus, signifiant que l'épissage alternatif est régulé de façon tissu-spécifique (Wang *et al.*, 2008a).

Cet épissage alternatif tissu-spécifique peut être expliqué en partie par une forte expression de certains facteurs d'épissage dans certains tissus (Castle *et al.*, 2008; David & Manley, 2008). Une liste non-exhaustive est représentée dans le Tableau 3.

Parmi tous les tissus humains, le cerveau est celui pour lequel il existe le plus d'isoformes tissu-spécifiques caractérisées. Parallèlement, plusieurs facteurs régulateurs spécifiques du cerveau ont été identifiés dont nPTB, NOVA 1, NOVA 2 et les protéines Hu/Elav (Okano & Darnell, 1997; Darnell, 2006; Makeyev *et al.*, 2007). PTB est exprimée dans les cellules progénitrices neuronales, mais son expression est grandement réduite dans les neurones différenciés, dans lesquels l'expression de nPTB est activée (Figure 7) (Boutz *et al.*, 2007b; Coutinho-Mansfield *et al.*, 2007; Makeyev *et al.*, 2007).

| Name | Other names | Binding domain | Binding motif | Tissue expression | Target genes |
|---------|--------------------|-------------------------|------------------|---|--|
| nPTB | brPTB and PTBP2 | RRM | CUCUCU | Neurons, myoblasts and testes | BIN1, GLYRA2, ATP2B1, MEF2, NASP, SPAG9 and SRC |
| NOVA1 | NA | KH | YCA Y | Neurons of the hindbrain and spinal cord | GABRG2, GLYRA2 and NOVA1 |
| NOVA2 | NA | KH | YCA Y | Neurons of the cortex, hippocampus and dorsal spinal cord | KCNJ, APLP2, GPHN, JNK2, NEO, GRIN1 and PLCB4 |
| FOX1 | A2BP1 | RRM | (U)GCAUG | Muscle, heart and neurons | ACTN, EWSR1, FGFR2, FN1 and SRC |
| FOX2 | RBM9 | RRM | (U)GCAUG | Muscle, heart and neurons | EWS, FGFR2, FN1 and SRC |
| RBM35a | ESRP1 | RRM | GU rich | Epithelial cells | FGFR2, CD44, CTNND1 and ENAH |
| RBM35b | ESRP2 | RRM | GU rich | Epithelial cells | FGFR2, CD44, CTNND1 and ENAH |
| TIA1 | mTIA1 | RRM | U rich | Brain, spleen and testes | MYPT1, CD95, CALCA, FGFR2, TIAR, IL8, VEGF, NF1 and COL2A1 |
| TIAR | TIAL1 and mTIAR | RRM | U rich | Brain, spleen, lung, liver and testes | TIA1, CALCA, TIAR, NF1 and CD95 |
| SLM2 | KHDRBS3 and TSTAR | KH | UAAA | Brain, tests and heart | CD44 and VEGFA |
| Quaking | QK and QKL | KH | ACUAAY[...]JUAAY | Brain | MAG and PLP |
| HUB | HUC, HUD and ELAV2 | RRM | AU rich | Neurons | CALCA, CD95 and NF1 |
| MBNL | NA | CCCH zinc finger domain | YGCU(U/G)Y | Muscles, uterus and ovaries | TNNT2, INSR, CLCN1 and TNNT3 |
| CELF1 | BRUNOL2 | RRM | U and G rich | Brain | TNNT2 and INSR |
| ETR3 | CELF2 and BRUNOL3 | RRM | U and G rich | Heart, skeletal muscle and brain | TNNT2, TAU and COX2 |
| CELF4 | BRUNOL4 | RRM | U and G rich | Muscle | MTMR1 and TNNT2 |
| CELF5 | BRUNOL5 and NAPOR | RRM | U and G rich | Heart, skeletal muscle and brain | ACTN, TNNT2 and GRIN1 |
| CELF6 | BRUNOL6 | RRM | U and G rich | Kidney, brain and testes | TNNT2 |

Tableau 3 : Liste des facteurs d'épissage alternatif tissu-spécifiques. Les domaines de fixation à l'ARN, les sites consensus d'interaction, le tissu dans lequel ces protéines sont majoritairement exprimées ainsi que des cibles régulées par ces facteurs sont indiqués (Chen & Manley, 2009).

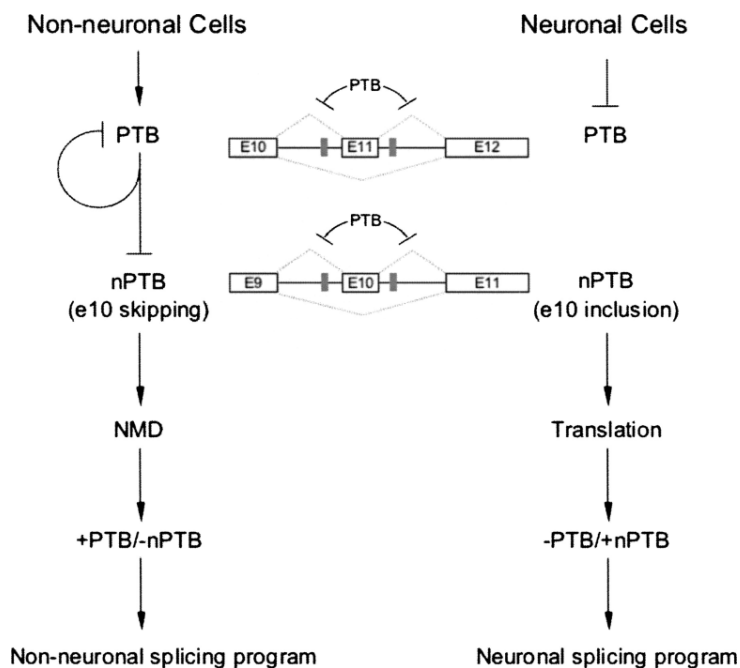


Figure 7 : Mécanisme de régulation des facteurs antagonistes PTB et nPTB. PTB est transcriptionnellement actif dans les cellules non neuronales et inactif dans les cellules neuronales. PTB réprime l'inclusion de l'exon alternatif 10 de nPTB en se fixant sur une séquence inhibitrice intronique (représentée en gris foncé). Ceci aboutit à l'exclusion de l'exon ce qui introduit un codon stop prématuré en phase et la dégradation du transcrit par le NMD. Dans les cellules neuronales, l'absence de PTB permet l'expression de nPTB (Coutinho-Mansfield et al., 2007).

De récentes expériences montrent que la transition PTB/nPTB est un mécanisme post-transcriptionnel important pour la différenciation neuronale et qu'il est à l'origine de près de 25% des changements d'épissage alternatif spécifiques du cerveau (Boutz *et al.*, 2007b).

Un autre exemple concerne la famille de protéine exclusivement neuronale Hu/ELAV. Ces protéines inhibent le recrutement de la snRNP U1 sur une séquence riche en AU localisée en aval du site 5' d'épissage de l'exon 23a du transcrite calcitonin-related polypeptide- α (CALCA), au profit de la protéine TIA1 (Zhou & Lou, 2008).

L'épissage alternatif d'un grand nombre de transcrits est également régulé durant la différenciation myogénique, principalement par deux facteurs antagonistes MBNL1 et CUGBP1, dont le rôle dans l'épissage alternatif et la Dystrophie Myotonique de type 1 sera abordé en détail dans la seconde partie de l'introduction (Bland *et al.*, 2010).

5) La définition d'un « code d'épissage »

Si la compréhension du « code génétique » fut relativement aisée, il est vite apparu que ce code n'autorisait qu'un décryptage fragmentaire de l'information génique. Depuis quelques années émerge l'idée que l'épissage alternatif serait également déterminé par un « code d'épissage » qui définirait le profil d'épissage d'un ARN pré-messager dans un tissu donné.

Plusieurs exemples de séquences régulatrices et de protéines spécifiques fortement exprimées dans un tissu et sans équivalence d'expression dans d'autres tissus ont donc été décrits. En général, il est admis que les concentrations relatives des facteurs activateurs ou répresseurs pourraient définir un « code d'épissage » qui définirait un modèle de régulation propre à chaque ARN pré-messager, afin que des mêmes groupes d'éléments régulateurs puissent être régulés de manière coordonnée (Matlin *et al.*, 2005; Keene, 2007; Wang & Burge, 2008).

Qu'est ce qui définit ce « code d'épissage » ? Est-il uniquement composé des séquences et facteurs régulateurs que nous avons préalablement décrits ? Ou existe-t-il de nouveaux paramètres qui entreraient dans la définition de ce code ? Une fois la totalité des informations compilées, pourrait-on prédire avec une grande efficacité les profils d'épissage ? Ce code, une fois déchiffré, permettrait-il d'identifier les bases moléculaires de nombreuses maladies génétiques dues à des dérégulations de l'épissage ?

Afin de répondre à ces questions, j'aborderai dans les deux prochaines parties les découvertes récentes de nouveaux éléments jouant un rôle dans la régulation de l'épissage alternatif et les dernières avancées sur le déchiffrement de ce « code d'épissage » qui représente le challenge du 21^{ème} siècle.

D) Nouveaux éléments régulateurs du « code d'épissage »

1) Influence de la transcription

L'épissage de l'ARN pré-messager et sa régulation *in vivo* ne peuvent être considérés comme un processus isolé, du fait que cette étape se déroule pour une grande partie de manière co-transcriptionnelle. Pendant que l'ARN polymérase II (ARNpol II) synthétise l'ARN pré-messager, la machinerie d'épissage est immédiatement recrutée sur le transcrit naissant et épisse les introns avant que la transcription ne soit terminée, suggérant des voies de régulation concomitantes entre la transcription et l'épissage (Moore & Proudfoot, 2009; Perales & Bentley, 2009). En effet, plusieurs équipes, dont celle de Kornblihtt ont démontré un lien fort entre la vitesse d'élongation de l'ARN polymérase II et la régulation de l'épissage alternatif (Kornblihtt *et al.*, 2004; Luco *et al.*, 2011). Un schéma représentatif est présenté dans la Figure 8.

L'étude de mutants de la ARNpol II ou l'insertion d'éléments de « pause » dans des minigènes rapporteurs ont montré que la pause ou la réduction de la vitesse d'élongation de l'ARN polymérase II favorise l'inclusion de l'exon alternatif dans les gènes de la fibronectine et du *fibroblast growth factor receptor 2* (FGFR2) (de la Mata *et al.*, 2003; Robson-Dixon & Garcia-Blanco, 2004). Par ailleurs, l'équipe de Smith a également observé qu'un retard de transcription au niveau d'une séquence inhibitrice de l'exon 3 du gène de l' α -tropomyosine humaine affecte l'épissage alternatif de cet exon (Roberts *et al.*, 1998).

Ces résultats suggèrent que l'épissage pourrait être régulé de façon cinétique. Le ralentissement de la vitesse ou la pause de la polymérase pourraient augmenter la fenêtre de temps au cours de laquelle un exon « faible », situé en amont, pourrait recruter la machinerie d'épissage avant qu'un exon « fort », situé en aval, ne prenne l'avantage.

D'autres études *in vivo* ont également indiqué un rôle de l'élongation dans l'épissage alternatif : l'inhibition de la phase d'élongation par l'ARN polymérase II dans des lymphocytes-T humains par la camptothécine modifie l'épissage alternatif d'un groupe d'ARN pré-messagers, impliqué dans le processing des ARN. Ceci favorise l'inclusion d'un exon alternatif contenant un codon stop prématuré, dégradé par le NMD (Ip *et al.*, 2011).

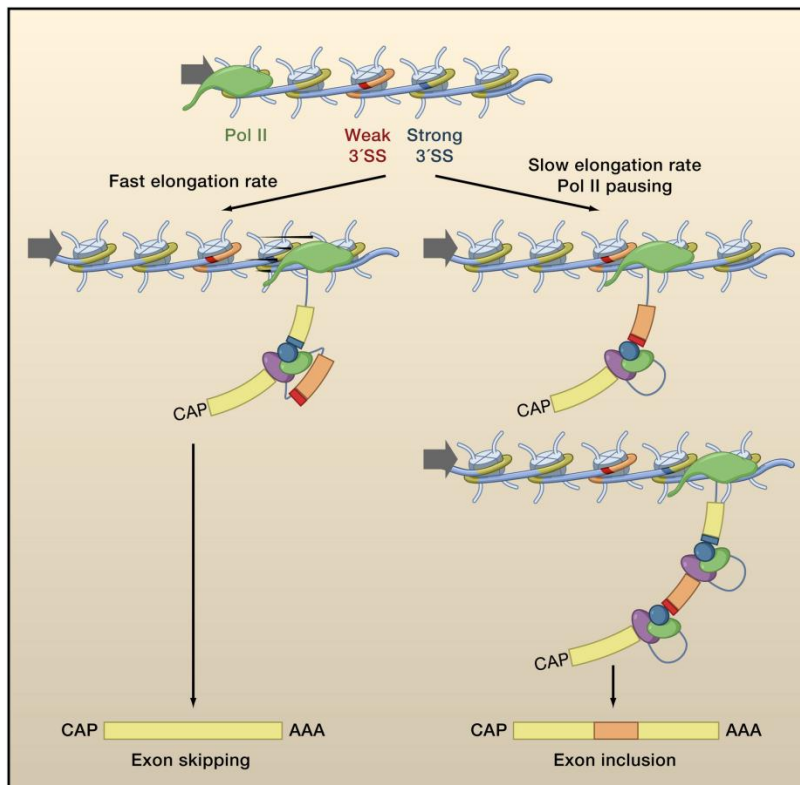


Figure 8 : Influence de la vitesse d'élongation par l'ARN polymérase II sur l'épissage alternatif. Lorsque le site 3' d'épissage de l'exon alternatif amont est plus faible que celui de l'exon en aval, une élongation rapide met les deux sites en compétition et favorise l'exclusion de l'exon alternatif. A l'inverse, une élongation lente permet au site 3' d'épissage faible d'être reconnu et favorise ainsi son inclusion (Luco *et al.*, 2011).

Les modifications post-traductionnelles de l'ARN polymérase II semblent également jouer un rôle dans la régulation de l'épissage alternatif. L'ARN polymérase II possède à son extrémité C-terminale un domaine (CTD) impliqué dans le recrutement de protéines accessoires qui ont un rôle majeur au cours de la transcription. Ce domaine consiste en une séquence heptamérique répétée (YSPTSPS) chez les eucaryotes. Afin de former un complexe d'initiation actif au niveau des promoteurs de gènes cibles, le CTD doit exister dans un état hyperphosphorylé (Oesterreich *et al.*, 2011). A la suite de l'assemblage du complexe d'initiation, le CTD montre une augmentation de la phosphorylation de la sérine 5 (S5-P). Cette phosphorylation est à l'origine du recrutement d'un complexe permettant l'ajout d'une coiffe méthylée sur l'extrémité 5' de l'ARN lors des premiers instants de la transcription. Au cours de l'élongation de la transcription, le profil de phosphorylation se retrouve modifié (Oesterreich *et al.*, 2011). Des expériences de ChIP ont démontré une perte de phosphorylation de la sérine 5 au profit de la phosphorylation de la sérine 2 (S2-P) au cours de la phase d'élongation de la transcription.

L'ensemble de ces expériences suggèrent que la forme S2-P soit plus active pour l'élongation que la forme S5-P. Ainsi, nous pouvons imaginer que des changements du ratio de ces phosphorylations au cours de l'élongation peuvent influencer sur la vitesse d'élongation de l'ARN polymérase II. Comme nous l'avons vu précédemment, l'hypothèse actuelle propose qu'une pause de l'élongation de la transcription puisse affecter l'épissage alternatif. L'une des hypothèses serait donc que le profil de phosphorylation pourrait être différent au niveau des jonctions exons-introns afin d'allouer plus de temps au spliceosome pour effectuer sa tâche.

Dans ce sens, l'équipe de Jean Beggs a montré une accumulation transitoire de l'ARNpol II (S5-P) au niveau des sites 3' d'épissage des introns chez *Saccharomyces cerevisiae* (Alexander *et al.*, 2010). L'équipe de Neugebauer a quant à elle montré que les gènes sans introns ou qui ne sont pas efficacement épissés ne montrent pas de pause de l'ARN polymérase II (Carrillo Oesterreich *et al.*, 2010). Cependant, cette étude a été faite chez un organisme ne possédant que peu d'introns et ne permet pas d'avoir une vue d'ensemble du mécanisme chez les eucaryotes en général.

Ceci suggère l'importance d'une pause de l'élongation avant l'exon pour qu'il soit reconnu par le spliceosome. La fin de cette pause est corrélée avec la phosphorylation de la sérine 2 et le recrutement de facteurs d'épissage. Ce mécanisme constituerait un « checkpoint transcriptionnel », dépendant de l'épissage qui inhiberait la transcription jusqu'à l'assemblage du spliceosome et permettrait de garantir une réaction d'épissage complète avant la fin de la transcription et le relargage du transcrit

2) Influence de la structure chromatinienne

Si la vitesse d'élongation de l'ARN polymérase II intervient dans la régulation de l'épissage alternatif, comment cette vitesse est-elle régulée ? L'hypothèse la plus évidente proviendrait sans aucun doute de la structure de la chromatine. Il est généralement admis que la vitesse de transcription est plus lente au niveau de régions génomiques riches en nucléosomes (hétérochromatine) qu'au niveau de régions pauvres en nucléosome (euchromatine). Une étude a montré que les nucléosomes *in vitro* se comportent comme des barrières qui augmentent localement la pause de la polymérase ou retardent la reprise de l'élongation (Hodges *et al.*, 2009). Concernant l'épissage, les études de la structure chromatinienne par digestion nucléasique chez les plantes, *Caenorhabditis elegans* ou chez l'humain ont révélé un enrichissement en nucléosomes au niveau des exons (Andersson *et al.*, 2009; Kolasinska-Zwierz *et al.*, 2009; Nahkuri *et al.*, 2009; Schwartz *et al.*, 2009; Spies *et al.*, 2009; Tilgner *et al.*, 2009; Chodavarapu *et al.*, 2010). Cette accumulation est indépendante du

niveau d'expression du gène (Andersson *et al.*, 2009; Tilgner *et al.*, 2009) et semble être déterminée par les séquences nucléiques définissant les introns (Kaplan *et al.*, 2009; Tillo *et al.*, 2010). Les exons dits « isolés » (entourés de part et d'autre par de très longs introns) et les exons alternatifs peu inclus sont hautement enrichis en nucléosomes par rapport aux exons constitutifs (Spies *et al.*, 2009; Tilgner *et al.*, 2009). Cependant, ce dernier point semble controversé, l'équipe de Gil Ast ayant récemment trouvé un résultat contradictoire (Schwartz *et al.*, 2009).

Il est tentant d'envisager l'enrichissement spécifique des nucléosomes au niveau des exons, et en particulier des exons contenant des sites faibles d'épissage, comme une barrière permettant la pause de l'ARN polymérase II assurant ainsi le recrutement et l'assemblage du spliceosome.

En plus des nucléosomes, les modifications des histones ainsi que la méthylation de l'ADN (Figure 9) sont distribuées de façon spécifique au niveau des exons et pourraient ainsi participer à la régulation de l'épissage alternatif (Schwartz *et al.*, 2009; Spies *et al.*, 2009; Kolasinska-Zwierz *et al.*, 2009; Chodavarapu *et al.*, 2010).

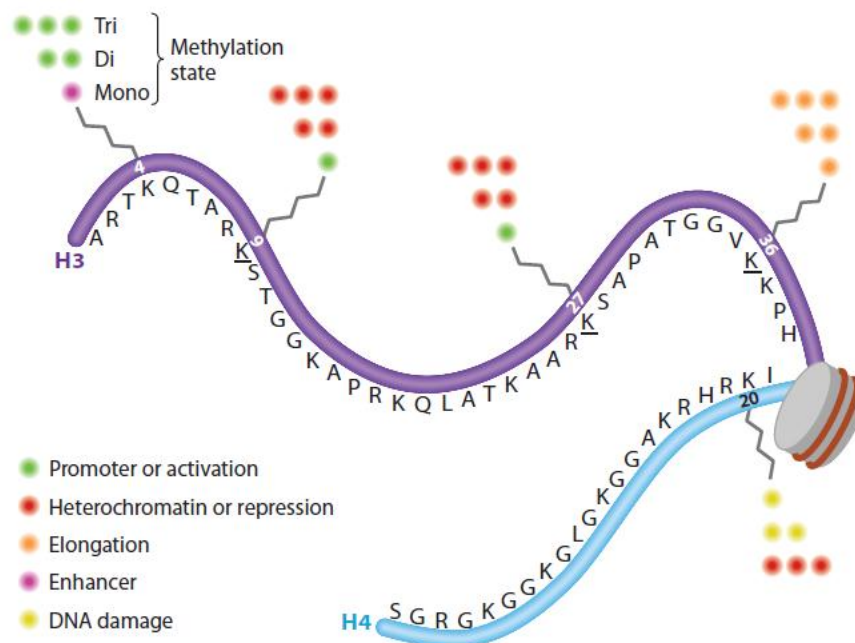


Figure 9 : Marques épigénétiques principales correspondant à la méthylation des lysines sur l'extrémité N-terminale des histones H3 et H4. Les nombres inscrits correspondent à la position des lysines pour chaque histone. La fonction générale de chaque état de méthylation (mono-, di- ou triméthylation) est représenté par des points et une couleur particulière (Mosammaparast & Shi, 2010).

Plusieurs groupes ont étudié la répartition de 42 marques d'histones dans des lymphocytes T humains et ont révélé une distribution préférentielle de plusieurs marques au niveau des exons (Schwartz *et al.*, 2009; Spies *et al.*, 2009). Plus particulièrement, la marque H3K36me3 se retrouve au niveau des exons hautement inclus et la marque H3K27me2 se retrouve au niveau des exons très peu inclus (Spies *et al.*, 2009). De manière intéressante, H3K36me3 est enrichie dans les exons constitutifs par rapport aux exons alternatifs, ce qui pourrait être un mécanisme pour distinguer ces deux types d'exons (Kolasinska-Zwierz *et al.*, 2009). En accord avec cette hypothèse, différents types de modifications des histones ont été corrélées avec des types d'épissage particuliers. L'un des exemples concerne le gène neuronal *NCAM*. La dépolarisation de la membrane plasmique augmente localement l'acétylation de l'histone H3K9 et la méthylation de l'histone H3K36 autour de l'exon alternatif 18, favorisant son inclusion (Schor *et al.*, 2009).

Dans le cas du gène de la fibronectine, l'induction d'hétérochromatine via un siARN au niveau de l'exon alternatif EDI augmente les marques épigénétiques H3K9 et H3K27, ce qui favorise l'inclusion de l'exon alternatif (Alló *et al.*, 2009). Les exons alternatifs des gènes *CD44* et *PKN2* sont marqués par des niveaux élevés de marques H3K9me3 et sont associés à HP1 γ , une protéine de la chromatine aboutissant à une inclusion de l'exon alternatif (Saint-André *et al.*, 2011).

Enfin, pour plusieurs gènes régulés par le facteur d'épissage PTB (dont la β -tropomyosine), l'inclusion de l'exon alternatif est corrélée avec une augmentation des marques H3K27me3 et H3K4me3 alors que l'exclusion de l'exon est corrélée avec un taux élevé de la marque H3K36me3 (Luco *et al.*, 2010). Il est à noter que l'altération de ces marques épigénétiques par des inhibiteurs ou l'inactivation des protéines modifiant les histones provoquent des changements au niveau de l'épissage alternatif, suggérant un rôle direct des histones dans la régulation de ce dernier. Au sujet de ce dernier point, une étude très récente suggère que l'épissage est mécaniquement couplé au recrutement de la H3K36 méthyltransférase HYPB au niveau de l'ARN polymérase II lors de l'élongation (de Almeida *et al.*, 2011).

3) **Influence des ARN non codants**

Les ARN non codants (ARNnc) ont récemment émergés comme de potentiels régulateurs de l'épissage alternatif. Ces ARNnc peuvent réguler l'expression ou modifier localement les concentrations relatives des facteurs régulateurs de l'épissage alternatif.

Durant le développement et la différenciation cellulaire, des ARNnc peuvent être spécifiquement exprimés afin de réguler l'expression de facteurs d'épissage.

L'un des exemples les plus emblématiques est le microARN mir-124, spécifique des neurones, qui provoque la dégradation de l'ARN codant PTB par interférence à ARN (Makeyev *et al.*, 2007). La réduction du taux intracellulaire de PTB déclenche un programme d'épissage spécifique aboutissant à la différenciation neuronale. A l'opposé, durant la différenciation myoblastique, le microARN muscle spécifique mir-133 est surexprimé, aboutissant à la réduction de l'expression du facteur neurone spécifique nPTB ce qui permet une différenciation myogénique fonctionnelle (Boutz *et al.*, 2007a). Enfin, le microARN mir-23a/b inhibe le taux protéique de la famille de facteurs d'épissage CELF qui régulent une majorité des inversions d'épissage alternatif lors de la différenciation cardiaque (Kalsotra *et al.*, 2010).

Mis à part les microARN, les siARN sont également impliqués dans la régulation de l'épissage alternatif. L'équipe de Kornbliht a montré qu'un siARN complémentaire de la région intronique située en aval de l'exon alternatif EDI du gène de la fibronectine provoque une compaction locale de la chromatine qui augmente l'inclusion de l'exon alternatif (Alló *et al.*, 2009).

Les ARNnc peuvent également utiliser d'autres méthodes moins directes pour participer à la régulation de l'épissage alternatif. Un exemple saisissant est celui de l'ARNnc MALAT-1, qui séquestre les facteurs d'épissage de la famille des protéines SR afin de réguler leur concentration relative dans le nucléoplasme (Figure 10). Lorsque MALAT-1 est exprimé, il séquestre les protéines SR au niveau des speckles nucléaires, qui ne peuvent plus jouer leur rôle régulateur. La dérégulation de l'expression de MALAT-1 relargue les protéines SR actives dans le nucléoplasme permettant de réguler l'épissage alternatif d'un grand nombre de gènes (Tripathi *et al.*, 2010). Nous verrons dans la seconde partie de l'introduction que l'ARN mutant DMPK séquestre également des facteurs d'épissage et est à l'origine d'une pathologie : la Dystrophie Myotonique de type 1.

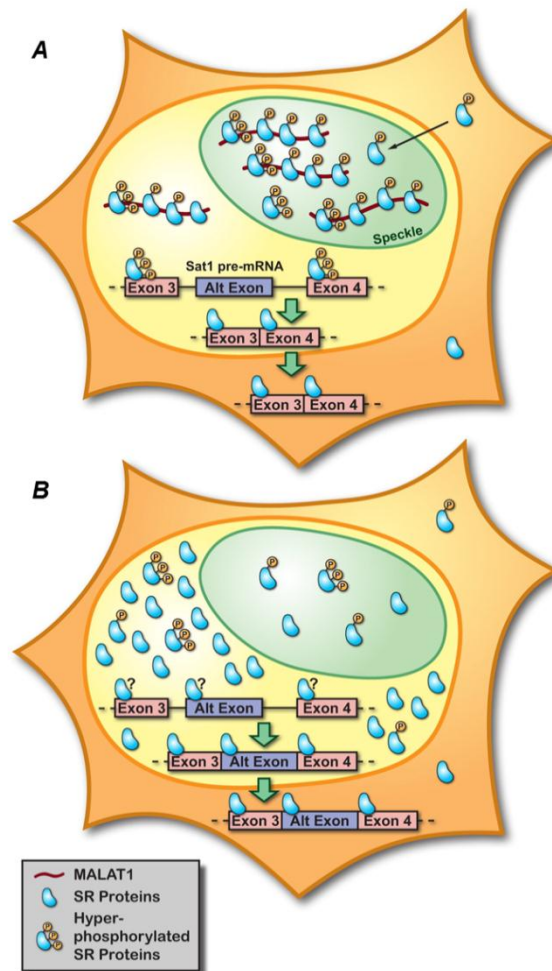


Figure 10 : Rôle de MALAT1 dans la régulation de l'épissage alternatif. **A.** Dans les cellules normales, MALAT1 s'associe avec des protéines SR dans les speckles nucléaires et le nucléoplasme. Ceci régule le recrutement de ces protéines sur l'ARN pré-messager afin de réguler l'épissage alternatif. **B.** Dans les cellules dont le gène codant MALAT1 est inactivé, le niveau cellulaire des protéines SR augmente sous une forme majoritairement non-phosphorylées, ce qui est à l'origine des altérations d'épissage alternatif. Le transcrit *Sat1*, dont l'épissage est altéré lorsque MALAT1 est inactivé est indiqué dans les deux conditions (Tripathi *et al.*, 2010).

4) Influence des structures secondaires ARN

Les molécules d'ARN peuvent adopter de nombreuses conformations en solution grâce aux appariements intramoléculaires et aux interactions hydrophobes. Par exemple, les ARN de transfert (tRNA) adoptent une structure secondaire et tertiaire bien particulière à quatre tiges ou bras, appelée «feuille de trèfle» (Holbrook, 2005). Cependant, la plupart des ARNm ne possèdent que des structures secondaires locales (Graveley *et al.*, 1998). Ces structures peuvent être reconnues par d'autres molécules, par exemple, l'aminoacétyl-tRNA synthétase qui interagit avec le bras accepteur du tRNA (Sankaranarayanan *et al.*, 1999) ce qui illustre l'importance fonctionnelle des structures ARN. Cependant, la plupart des facteurs qui régulent la sélection des sites d'épissage reconnaissent uniquement des ARN simple brin

qui possèdent peu ou pas d'appariements. En effet, les domaines KH (hnRNP K homology), les RRM (RNA recognition motifs) ou les motifs à doigts de zinc se fixent sur les séquences simple brin de 2 à 10 nucléotides (Auweter *et al.*, 2006). Fréquemment, dans une structure secondaire ARN, le site d'interaction avec une protéine est localisé au niveau de la boucle d'une tige-boucle (Buckanovich & Darnell, 1997; Skrisovska *et al.*, 2007). La spécificité de séquence de l'interaction est permise par les interactions hydrophobes entre les bases de l'ARN et les acides aminés à la surface de la protéine, ce qui explique le fait que les protéines se fixent à de l'ARN non structuré.

Suivant les conditions thermodynamiques, la stabilité de ces structures peut interférer ou moduler la sélection des sites d'épissage. En principe, une structure secondaire peut activer ou inhiber l'assemblage du spliceosome, puisque la reconnaissance des sites activateurs ou répresseurs dépend d'une interaction entre des facteurs protéiques et une séquence simple brin d'ARNm (Maris *et al.*, 2005; Hiller *et al.*, 2007). Ainsi, la structure locale de l'ARN peut interférer avec l'assemblage du spliceosome si elle contient une hélice stable comportant un site d'épissage ou un site de fixation à un facteur activateur. Une structure secondaire peut au contraire faciliter la formation du spliceosome en masquant un site répresseur qui ne pourra pas être reconnu par une protéine régulatrice (Hertel, 2008).

L'importance de la stabilité d'une structure secondaire ARN dans la modulation de l'épissage a été démontrée dans plusieurs études. En effet, il est estimé que 4% des événements d'épissage alternatif conservés au cours de l'évolution sont associés à une structure secondaire conservée ayant pour but de moduler l'accès à un site d'épissage (Shepard & Hertel, 2008).

D'après les nombreux exemples de la littérature, les structures ARN peuvent jouer sur la régulation de l'épissage de deux façons : de façon directe en modifiant l'accessibilité du spliceosome pour des séquences régulatrices ou de façon indirecte en nécessitant l'intervention de protéines capables de moduler la structure secondaire.

a) Régulation directe

Il existe deux classes de structures : les structures qui inhibent l'épissage et les structures qui le favorisent.

Les structures qui inhibent l'épissage

Les premières étapes du recrutement du spliceosome ont lieu par l'appariement de séquences entre les snRNA contenus dans les snRNP et l'ARN pré-messager. Nous pouvons donc imaginer que les structures secondaires puissent inhiber l'épissage en empêchant l'accès à ces séquences. Cependant, ces structures secondaires ne sont pas toujours extrêmement stables et une compétition peut exister entre une structure interne et l'appariement du snRNP avec l'ARN pré-messager. Dans plusieurs cas, les structures secondaires séquestrent les séquences impliquées dans l'épissage, comme par exemple les sites d'épissage 5' et 3', le point de branchement ou la région riche en pyrimidines. Lorsque le site 5' d'épissage ou le point de branchement est structuré, le recrutement des snRNP U1 ou U2 est inhibé (Figure 11).

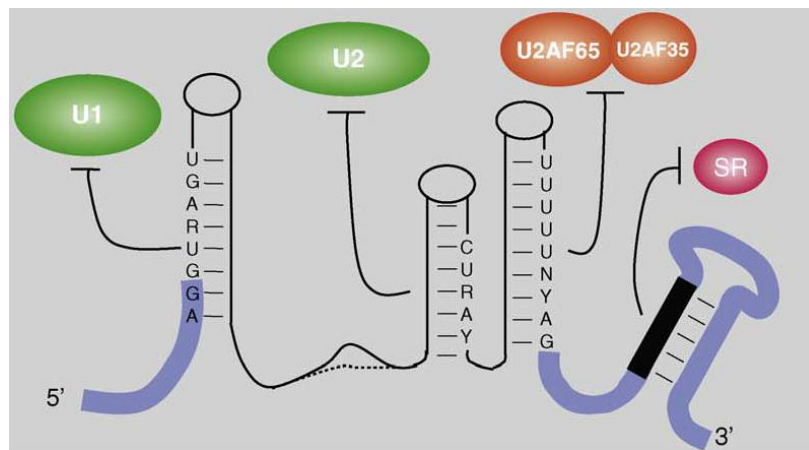


Figure 11 : Les structures qui inhibent l'épissage. Sont représentées les structures en tige-boucle qui répriment la fixation de la snRNP U1 sur le site 5' d'épissage, de la snRNP U2 sur le point de branchement, de U2AF65/35 sur le site 3' d'épissage et une protéine SR sur une séquence régulatrice au niveau d'un exon. (Warf & Berglund, 2010).

L'un des tout premiers articles illustrant ce mécanisme a été publié par le laboratoire d'Edward Brody et Marc Fizman. Le modèle d'étude était l'ARN pré-messager de la β -tropomyosine aviaire, un transcrite régulé de façon tissu-spécifique. Le gène possède dans sa région centrale deux exons mutuellement exclusifs selon le type cellulaire. L'exon 6A est inclus dans les ARNm des cellules non-musculaires et les cellules de muscles lisses tandis que l'exon 6B est inclus dans les ARNm des cellules musculaires différenciées (Libri *et al.*, 1990). L'intron en aval de l'exon 6B possède une structure en tige-boucle qui participe à la régulation de l'épissage alternatif de cet exon en empêchant l'interaction du facteur U1snRNP sur le site 5' d'épissage de l'exon 6B (Clouet d'Orval *et al.*, 1991; Libri *et al.*, 1991; Sirand-Pugnet *et al.*, 1995). Un rôle inhibiteur de la structure secondaire de l'ARN a également été

démontré pour la reconnaissance du site d'épissage de l'exon 7 du gène SMN2. La formation d'une tige-boucle ARN proche du site 5' d'épissage de l'exon interfère avec la fixation de la snRNP U1, aboutissant à la réduction du taux d'inclusion de cet exon (Singh *et al.*, 2007).

Dans d'autres cas, ces structures peuvent inhiber l'interaction de protéines au niveau de l'ARN pré-messager. L'hétérodimère U2AF consiste en deux protéines, U2AF65 et U2AF35 qui interagissent avec la région riche en pyrimidine et le site 3' d'épissage respectivement, et aident au recrutement de la snRNP U2 au niveau du point de branchement (Wahl *et al.*, 2009). Des études structurales et biochimiques ont démontré que U2AF65 reconnaît la région riche en pyrimidine uniquement lorsque celle-ci est en conformation simple brin (Sickmier *et al.*, 2006). Lorsque cette région est structurée, l'interaction avec U2AF65 n'a pas lieu et l'épissage est bloqué (Warf *et al.*, 2009). Les structures secondaires peuvent également moduler l'interaction de U2AF35 avec le site 3' d'épissage (Loeb *et al.*, 2002).

Les structures qui activent l'épissage

L'un des mécanismes les plus représentés concerne le fait que les structures secondaires dans l'ARN pré-messager peuvent rapprocher des séquences régulatrices qui sont éloignées au niveau de la structure primaire (Figure 12). Dans la majorité des ARN pré-messagers, le site 3' d'épissage et le point de branchement sont généralement proches (Black, 2003). Cependant pour certains introns de grande taille, ces deux régions peuvent être éloignées de plusieurs centaines de nucléotides mais cette distance ne perturbe pas l'efficacité d'épissage (Deshler & Rossi, 1991).

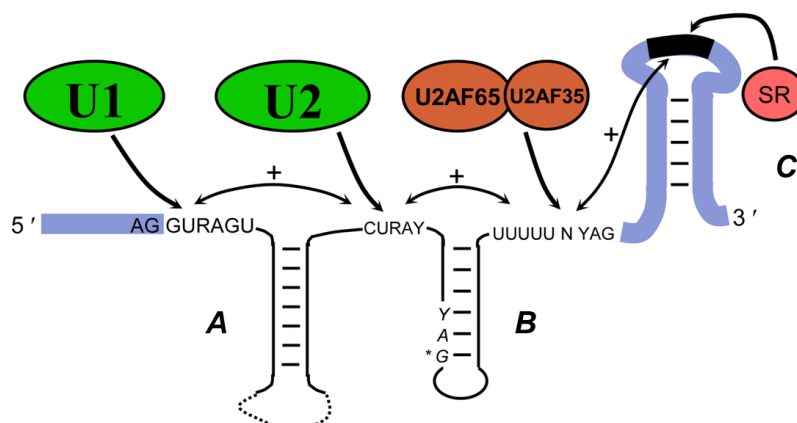


Figure 12 : *Les structures qui activent l'épissage.* Ce schéma représente une structure secondaire formée de trois tige-boucle qui permet le rapprochement des sites d'épissage 5' et 3'. La première tige-boucle (A) rapproche le site 3' d'épissage et le point de branchement. La deuxième tige-boucle (B) masque un site 3' d'épissage cryptique (noté YAG*). La troisième tige-boucle (C) permet l'accès à une séquence régulatrice exonique à une protéine SR. (Warf & Berglund, 2010).

Dans plusieurs de ces introns, des régions structurées localisées entre le point de branchement et le site 3' d'épissage ont été identifiées et servent à rapprocher ces deux séquences, permettant à un point de branchement éloigné d'être utilisé. L'un des exemples les plus intéressants est celui observé pour le gène *Down syndrome cell adhesion molecule* (Dscam) chez la drosophile. Deux classes d'éléments conservés ont été découvertes dans le cluster de l'exon 6 de l'ARNm *Dscam*, qui contient 48 exons alternatifs : un site d'encrage commun à chaque exon et une séquence sélectrice propre à chaque variant de l'exon 6. Chaque séquence sélectrice peut s'apparier avec le site d'encrage afin de former une structure secondaire, ayant pour conséquence la sélection d'exons mutuellement exclusifs (Graveley, 2005).

Un autre exemple concerne les sites cryptiques d'épissage, ces séquences homologues aux séquences consensus des sites 5' et 3' d'épissage mais qui ne sont pas utilisées par le spliceosome, et qui peuvent être séquestrées par des structures secondaires. Si ces structures sont modifiées, la machinerie d'épissage utilise ces sites d'épissage incorrect aboutissant à un transcrit abortif (Deshler & Rossi, 1991; Domenjoud *et al.*, 1991). Sachant que beaucoup d'introns possèdent de nombreux sites cryptiques, nous pouvons imaginer que ces structures pourraient masquer ces sites d'épissage non-fonctionnels, suggérant un rôle indispensable pour la sélection des sites d'épissage fonctionnels. Les séquences régulatrices auxiliaires dites « enhancer » et « silencer » peuvent également être retrouvées dans des structures secondaires. L'un des exemples les mieux caractérisés est une ESE localisée dans la boucle d'une tige-boucle au niveau de l'exon EDI du gène de la fibronectine (Buratti *et al.*, 2004). Lorsque la structure est altérée et que l'ESE n'est plus en conformation simple brin, les protéines SR ne se fixent plus aboutissant en l'exclusion de l'exon alternatif. Cette étude apporte une nouvelle preuve qu'au-delà de l'importance de la séquence primaire pour les séquences régulatrices, la structure secondaire est également indispensable pour permettre l'accessibilité des facteurs d'épissage.

b) Régulation indirecte

L'un des exemples majeurs de régulation indirecte concerne la protéine MBNL1, un facteur d'épissage alternatif qui se fixe sur des régions structurées de l'ARN et qui régule de manière positive ou négative l'épissage alternatif d'un groupe important d'exon de type cassette (Pascual *et al.*, 2006; Ranum & Cooper, 2006). L'un des sites les mieux caractérisés

pour ce facteur est localisé dans l'ARN pré-messager de la troponine T cardiaque (*TNNT2*) (Figure 13). MBNL1 inhibe l'inclusion de l'exon alternatif 5 par son interaction avec une structure en tige-boucle en amont de l'exon alternatif (Warf & Berglund, 2007; Warf *et al.*, 2009). La fixation de MBNL1 sur cette structure bloque la fixation de U2AF65 sur la région riche en pyrimidines provoquant l'exclusion de l'exon alternatif. Cependant, de récentes études de cristallographie ont montré que MBNL se fixe sur des fragments d'ARN simple brin, localisés dans une structure en tige-boucle composée de répétitions CUG (Teplova & Patel, 2008). Ainsi, l'une des questions serait de savoir comment MBNL peut interagir avec des résidus localisés dans la tige d'une tige-boucle et si elle nécessite le renfort d'une autre protéine afin de moduler cette structure secondaire.

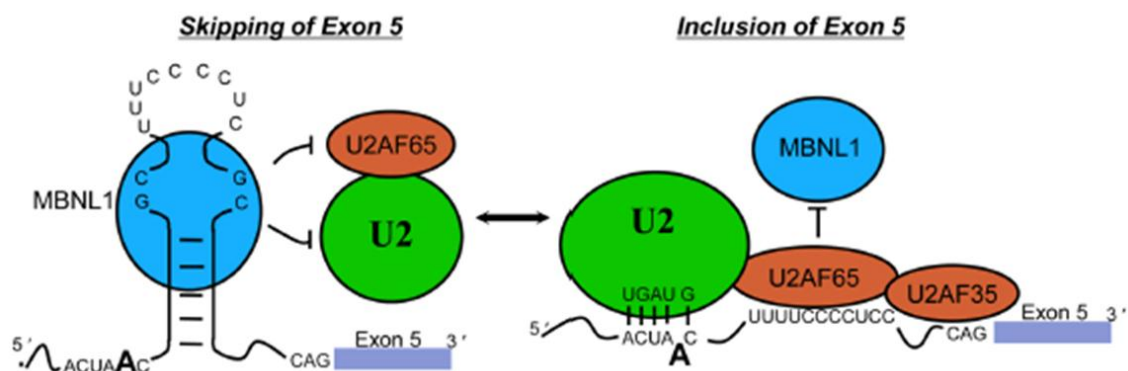


Figure 13 : Mécanisme de la régulation de l'épissage alternatif de l'exon 5 du transcrite codant la troponine T cardiaque (*TNNT2*). L'interaction du facteur d'épissage MBNL1 avec une structure secondaire dans l'intron 4 de *TNNT2* inhibe par encombrement stérique la fixation de la protéine U2AF65 sur la région polypyrimidique et le recrutement de la snRNP U2 sur le site 3' d'épissage. (Warf *et al.*, 2009; Warf & Berglund, 2010).

Les ARN hélicases sont importantes pour l'épissage puisqu'elles aident au recyclage des snRNPs (Wahl *et al.*, 2009). Cependant, trois exemples récents indiquent que les ARN hélicases pourraient aussi intervenir dans l'épissage alternatif en déstabilisant des structures inhibitrices au niveau de l'ARN pré-messager. L'un des premiers exemples a été publié en 2002. La DEAD-box ARN hélicase p72 (également nommée DDX17) augmente l'inclusion de l'exon alternatif 4 du minigène CD44 (Hönig *et al.*, 2002). Une tige-boucle séquestre le site 5' d'épissage en aval de l'exon 4, et l'hypothèse proposée est que p72 augmenterait l'inclusion de l'exon en déstabilisant cette tige (Figure 14). Bien qu'un rôle direct de p72 sur la structure de l'ARN n'ait pu être démontré, des mutations insérées dans cette structure secondaire ou au niveau des domaines protéique de p72 impliqués dans l'activité hélicase augmentent également l'inclusion de l'exon alternatif (Hönig *et al.*, 2002). De façon

intéressante, la protéine paralogue à p72, l'ARN hélicase p68 ne régule pas l'épissage alternatif de l'exon 4, suggérant des activités spécifiques pour chacune de ces protéines.

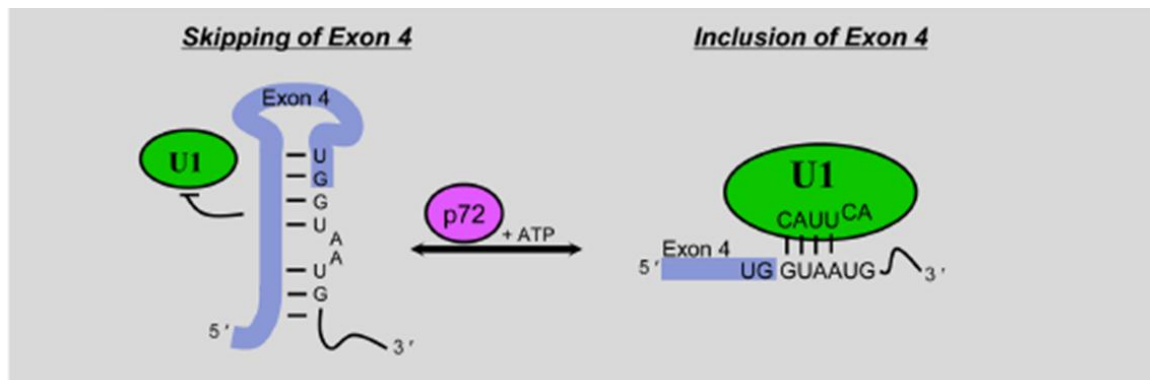


Figure 14 : Mécanisme putatif de la régulation de l'épissage alternatif de l'exon 4 du transcrit CD44. La protéine DEAD-box ARN hélicase p72 augmenterait l'inclusion de l'exon alternatif 4 dans l'ARNm CD44 par une déstabilisation ATP-dépendante de la structure tige-boucle contenant le site 5' d'épissage en aval de l'exon 4. (Hönig *et al.*, 2002; Warf & Berglund, 2010).

Au sujet de cette dernière, la DEAD-box ARN hélicase p68 a également été rapportée. p68 induit le changement conformationnel d'une tige boucle située en aval du site 5' de l'exon 10 du gène Tau, augmentant l'accès de la snRNP U1 sur le site 5' d'épissage (Kar *et al.*, 2011). Dans une autre étude, il a été proposé que p68 altère la structure secondaire d'une tige-boucle située en aval de l'exon alternatif IDX dans l'ARN pré-messager c-H-ras, bloquant l'accès de hnRNP H (Camats *et al.*, 2008).

De nombreuses structures secondaires ARN inhibitrices pourraient être régulées par d'autres ARN hélicases, indiquant un rôle plus direct de la famille des ARN hélicases dans la régulation de l'épissage.

E) Analyse globale du « code d'épissage » et prédiction

Un nombre important de nouvelles technologies à haut-débit a récemment été développé pour analyser la régulation de l'épissage alternatif en recherchant les différentes isoformes d'épissage à un niveau génomique. La première technique utilisée fut les puces à ADN, qui permettaient l'utilisation de sondes oligonucléotides ciblant les jonctions entre exons constitutifs (Johnson *et al.*, 2003), ou les jonctions exon alternatif-exon constitutif (Clark *et al.*, 2002; Pan *et al.*, 2004; Blanchette *et al.*, 2005). Ces technologies ont facilité l'analyse de l'épissage alternatif à l'échelle du génome chez l'humain, la souris ou le chimpanzé ainsi que la détection de l'impact de facteurs d'épissage spécifiques ou de stimuli environnementaux sur l'épissage (Srinivasan *et al.*, 2005; Calarco *et al.*, 2007; Boutz *et al.*,

2007b; Ip *et al.*, 2007; Makeyev *et al.*, 2007; Kalsotra *et al.*, 2008; de la Grange *et al.*, 2010; Dutertre *et al.*, 2010b).

L'identification systématique des cibles ARN pour différents facteurs d'épissage a pu également être effectuée par l'utilisation de nouvelles approches biochimiques, telles que le CLIP (cross-linking immunoprécipitation) (Ule *et al.*, 2003), le RIP (RNA immunoprécipitation) (Keene *et al.*, 2006) ou le SELEX (Systematic Evolution of Ligands by EXponential enrichment) (Lorenz *et al.*, 2006). Ces analyses peuvent également être couplées aux techniques d'étude à haut-débit comme les puces à ADN/ARN ou le séquençage. L'analyse des séquences d'interaction aide à définir une séquence consensus et à identifier des régulations coopératives ou antagonistes entre les différents facteurs.

Le but ultime pour les équipes étudiant l'épissage alternatif et sa régulation est la détermination d'un ensemble de règles qui prédisent le profil d'épissage de n'importe quel transcrit primaire, que l'on pourrait qualifier de « code d'épissage » (Fu, 2004; Matlin *et al.*, 2005; Barash *et al.*, 2010). Du fait du large nombre de séquences « cis » régulatrices identifiées et de la détermination de l'identité et de la fonction des facteurs « trans », l'étape ultime est l'assemblage de toutes ces données pour définir un algorithme prédictif de la reconnaissance des exons et introns durant l'épissage. L'une des premières contributions fut la mise en place d'Exon-scan, un algorithme impliquant la recherche dans l'ARN pré-messager de paires de sites d'épissage 5' et 3' proches, la notation de la force prédite de ces sites d'après les homologies avec les séquences consensus et la détection de séquences « cis » de régulation introniques ou exoniques (Wang *et al.*, 2004). Cette étude a également permis la mise en évidence de la fonction *in vivo* des ESS.

Une étude récente parue en 2010 représente une percée dans la modélisation de l'épissage alternatif tissu-spécifique (Barash *et al.*, 2010). Cet article proposa le premier déchiffrement d'un code d'épissage permettant la prédiction de profil d'épissage alternatif de quatre classes de tissus adultes et embryonnaires. Les auteurs ont récolté des données de puces sur l'épissage alternatif de plus de 3000 exons dans quatre tissus (système nerveux central, muscle, système digestif et cellules souches/embryonnaires) ainsi que des informations sur les séquences de régulation et les structures secondaires ARN dans le but de compiler une liste de plusieurs milliers de caractéristiques qui joueraient un rôle dans la régulation de l'épissage alternatif. Enfin, ils ont utilisé une approche bio-informatique pour sélectionner les caractéristiques qui une fois cumulées expliquent le mieux les profils observés dans les quatre types de tissus analysés.

Cette méthode atteint un succès remarquable dans la prédiction de l'épissage alternatif tissu-spécifique et permet la mise en lumière de nouvelles relations agonistes ou antagonistes entre les facteurs d'épissage (Figure 15). Les auteurs révèlent également que l'inclusion d'exon possédant un codon stop prématuré est un type de régulation génique très répandu dans la transition entre les tissus embryonnaires et adultes (Barash *et al.*, 2010). Pour être complet, ce code nécessite des améliorations, notamment l'ajout d'informations en relation avec la structure de la chromatine, les marques épigénétiques et les sites ou structures secondaires d'interaction de nouveaux facteurs « trans ».

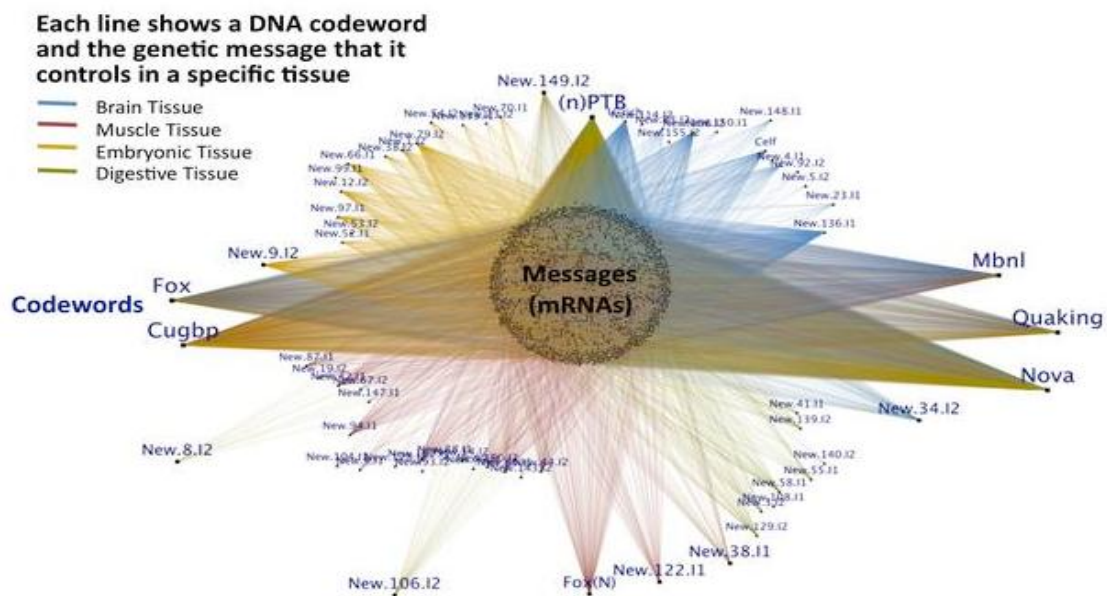


Figure 15 : Représentation schématique des résultats obtenus par Barash *et al.* concernant la redondance des sites consensus de facteurs d'épissage suivant les tissus. Chaque ligne correspond à une séquence consensus correspondant au site d'interaction d'un facteur d'épissage. La ligne bleue représente le cerveau, la ligne rouge représente le muscle, la ligne jaune correspondant au tissu embryonnaire et la ligne marron au tissu digestif. Certaines séquences (codewords) ou protéines semblent spécifiques d'un tissu donné. (Barash *et al.*, 2010).

Le déchiffrement de ce code d'épissage est primordial pour la compréhension globale de ce mécanisme ainsi que pour des raisons médicales. En effet, il est estimé que 15% à 50% des mutations à l'origine de maladies affectent la sélection d'un ou plusieurs sites d'épissage. Le déchiffrement de ce code permettrait donc potentiellement de définir les voies de régulations défectueuses pour un grand nombre de maladies et d'imaginer des stratégies thérapeutiques pour contrer ces altérations.

L'un des exemples les plus caractéristiques de maladies génétiques dues à un défaut d'épissage est la Dystrophie Myotonique de type 1 qui sera abordée dans la deuxième partie de l'introduction.

II. La Dystrophie Myotonique de type I : une maladie caractérisée par un gain de fonction de l'ARN

A) Généralités

La Dystrophie Myotonique de type 1 (DM1), décrite indépendamment par Steinert et Batten & Gibbs, représente la forme la plus courante de dystrophie chez l'adulte (Steinert, 1909; Batten & Gibb, 1909). En plus d'une atteinte musculaire, caractérisée par une myotonie associée à une faiblesse et une atrophie musculaire, cette pathologie présente un phénotype multisystémique complexe avec des symptômes qui, en apparence, ne présentent pas de liens évidents et incluent des atteintes cardiaques, oculaires, de l'appareil digestif, du système nerveux central et du système endocrinien (Ranum & Day, 2004; Day & Ranum, 2005).

La distribution géographique de la Dystrophie Myotonique est universelle et sa prévalence est d'environ 1 cas pour 8 000 habitants (Peter S Harper, 1989), à l'exception de certaines régions, notamment au Québec où elle atteint 1/530 (Yotova *et al.*, 2005).

La Dystrophie Myotonique est une maladie héréditaire autosomique dominante, de pénétrance complète mais d'expression très variable d'un individu à l'autre.

La mutation associée à la Dystrophie Myotonique de type 1 a été identifiée en 1992 et correspond à une expansion instable d'un triplet CTG localisé au sein de la région 3'-UTR (*untranslated region*) au niveau de l'exon 15 du gène codant la DMPK (*Dystrophia Myotonica Protein Kinase*), situé sur le chromosome 19q13.3 (Brook *et al.*, 1992; Buxton *et al.*, 1992; Fu *et al.*, 1992; Mahadevan *et al.*, 1992). Un schéma est représenté Figure 16.



Figure 16 : Représentation schématique du gène DMPK contenant la mutation responsable de la pathologie. Le gène DMPK se situe au niveau du locus 19q13.3 et est composé de 14 exons (représentés en bleu) et d'une région 5'UTR et 3'UTR. Le nombre de répétitions CUG dans la partie 3'UTR est déterminant pour l'apparition de la maladie. Chez les personnes saines, le nombre de répétitions est inférieur à 37 CTG (en gris). Chez les patients, le nombre de répétitions peut aller de 50 jusqu'à 3000 répétitions CTG (en rose).

Le nombre de répétitions CTG varie entre 5 et 37 chez les individus sains (avec une moyenne se situant entre 5 et 10 dans la population générale) et de 50 à plusieurs milliers de répétitions chez les patients atteints de la pathologie. La sévérité des symptômes cliniques ainsi que l'âge de manifestation des premiers symptômes sont corrélés avec la taille des répétitions, un nombre de CTG supérieur à 1500 résultant généralement en des formes congénitales (Ranum & Cooper, 2006).

Deux caractéristiques intéressantes de cette maladie ont été mises en évidence. Une instabilité somatique marquée, avec des différences dans le nombre de répétitions entre les différents tissus, mais également au sein d'un même tissu est régulièrement observée (Seznec *et al.*, 2000). Ce phénomène semble prédominant dans le muscle, mais est également présent dans d'autres tissus comme le cerveau. De plus, un phénomène d'anticipation, autrement dit une aggravation de la maladie au cours des générations est associé à une expansion du nombre de répétitions (Monckton *et al.*, 1995; Martorell *et al.*, 1998). Les mécanismes physiopathologiques associés à l'instabilité des répétitions CTG entraînant l'expansion des répétitions CTG sont encore peu connus mais impliqueraient des protéines de réparation des mésappariements de l'ADN (Savouret *et al.*, 2004; Foiry *et al.*, 2006; Tomé *et al.*, 2009).

Par ailleurs, une seconde mutation, suspectée par des études de liaisons génétiques a été identifiée en 1998 (Ranum *et al.*, 1998) et consiste en une expansion instable du quadruplet CCTG au niveau de l'intron 1 du gène ZNF 9 sur le chromosome 3q21. Les patients atteints de cette deuxième forme de Dystrophie Myotonique (DM2) développent des symptômes équivalents à ceux de la DM1 mais globalement moins sévères malgré un nombre de répétitions bien plus élevés atteignant près de 11000 répétitions pour les formes les plus graves (Liquori *et al.*, 2001).

B) Aspects cliniques

On distingue quatre formes cliniques de la DM1 : la forme congénitale, la forme juvénile, la forme adulte et la forme tardive (Machuca-Tzili *et al.*, 2005). Il n'existe cependant pas de distinctions absolues entre ces différentes formes qui peuvent plutôt être vues comme formant un continuum. Les nombres de répétitions CTG les plus importants conduisent aux formes les plus graves de la maladie sans qu'un nombre précis de répétitions puisse être corrélé avec un seuil entre deux formes, ceci étant expliqué par la variabilité importante dans les manifestations de la pathologie.

Un récapitulatif des symptômes est visible dans le Tableau 4 et la Figure 17.

| Système / Organe | Atteinte |
|---------------------|--|
| Muscle squelettique | <ul style="list-style-type: none"> • Myotonie • Faiblesse des muscles distaux et à long terme proximaux • Faiblesse des muscles de la face, de la mâchoire, de la langue et des muscles fléchisseurs du cou • Atrophie du sternocléidomastoidien |
| Cœur | <ul style="list-style-type: none"> • Défaut de conduction • Arythmie cardiaque |
| Muscle lisse | <ul style="list-style-type: none"> • Constipation et diarrhées • Douleurs abdominales de type coliques • Syndrome pseudo-occlusif • Contractions non-ordonnées de l'utérus durant l'accouchement |
| 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 |
| Yeux | <ul style="list-style-type: none"> • Cataracte • Ptose • Dégénérescence de la rétine |
| Endocrinien | <ul style="list-style-type: none"> • Résistance à l'insuline • Atrophie testiculaire • Taux élevés de FSH |
| Peau | <ul style="list-style-type: none"> • Calvitie prématurée |

Tableau 4 : Liste des symptômes présents chez les patients DMI adulte suivant les tissus. Adapté de (Peter S Harper, 1989; Delaporte, 1998).

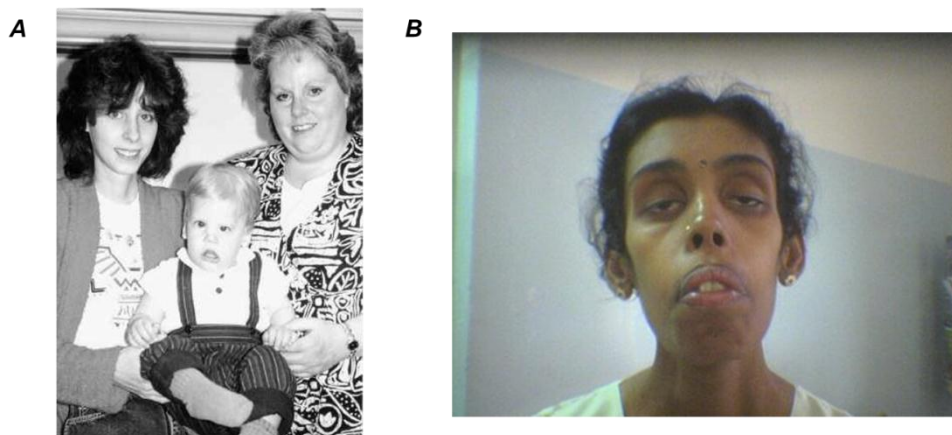


Figure 17 : Patients atteints de Dystrophie Myotonique de type 1. A. Une famille composée de trois générations affectée par la DMI. Le degré de sévérité augmente entre les générations : la grand-mère (100 répétitions) est légèrement affectée, la mère (200 répétitions) présente une faiblesse musculaire au niveau du visage et une expression faciale limitée. Le bébé (>1000 répétitions) est plus sévèrement atteint et présente des altérations musculaires au niveau du visage ainsi qu'une bouche ouverte en triangle, caractéristiques de la forme congénitale de la DMI. B. Patiente présentant une forme adulte de la pathologie (nombre de répétitions inconnu). (Jorde LB, 2000).

1) **Forme bénigne**

Cette forme se limite à une cataracte et parfois une calvitie frontale précoce avec éventuellement un déficit musculaire minime. Il existe également un risque cardiaque non-négligeable. Si cette forme n'entraîne que des symptômes superficiels, elle constitue une réserve de pré-mutation qui représente la source de transmission vers les formes plus graves.

2) **Forme adulte**

Les symptômes de la forme adulte de la maladie débutent entre 30 et 40 ans et sont très variables, regroupant de nombreux troubles d'ordre musculaire associés à une calvitie précoce et de grandes oreilles décollées donnant au visage un aspect particulier. Outre les symptômes musculaires prédominants, il existe également des troubles cardiaques, respiratoires, digestifs, psychiatriques et endocriniens.

3) **Forme juvénile**

Cette forme est caractérisée par une faiblesse des muscles du visage. L'atteinte musculaire peut prendre la forme classique ou être paradoxalement plus discrète. En revanche, les troubles du comportement sont au premier plan avec des difficultés d'apprentissage, un retard scolaire constant et des problèmes d'insertion sociale. L'atteinte cardiaque représentée par des troubles rythmiques de conduction favorisés par l'effort peut être révélatrice dès l'âge de 10 ans avec un risque de mort subite imposant un suivi cardiologique fréquent (Machuca-Tzili *et al.*, 2005).

4) **Forme congénitale**

La forme congénitale est la forme la plus sévère. Elle est associée à une transmission principalement maternelle, par opposition aux autres formes de la DM1 où la transmission peut être maternelle comme paternelle (Peter S Harper, 2001) et est caractérisée lors de la grossesse par une diminution de la motilité fœtale. A la naissance, l'enfant présente différents symptômes tels qu'une hypotonie généralisée, une détresse respiratoire, une faiblesse des muscles de la face, des difficultés d'alimentation pouvant être associées à des troubles de la succion et de la déglutition, mais également un retard du développement psychomoteur (Peter S Harper, 2001). La myotonie ne fait pas partie des symptômes durant la première année de vie (Machuca-Tzili *et al.*, 2005) et les cataractes ne se manifestent qu'à l'âge adulte.

C) Mécanismes physiopathologiques proposés dans DM1

Trois modèles physiopathologiques reliant l'expansion des triplets CTG aux symptômes cliniques de DM1 ont été proposés : (1) l'haplo-insuffisance de la DMPK ; (2) la diminution d'expression des gènes adjacents au gène DMPK ; (3) le gain de fonction toxique de l'ARNm mutant. Un récapitulatif de ces modèles est représenté dans la Figure 18.

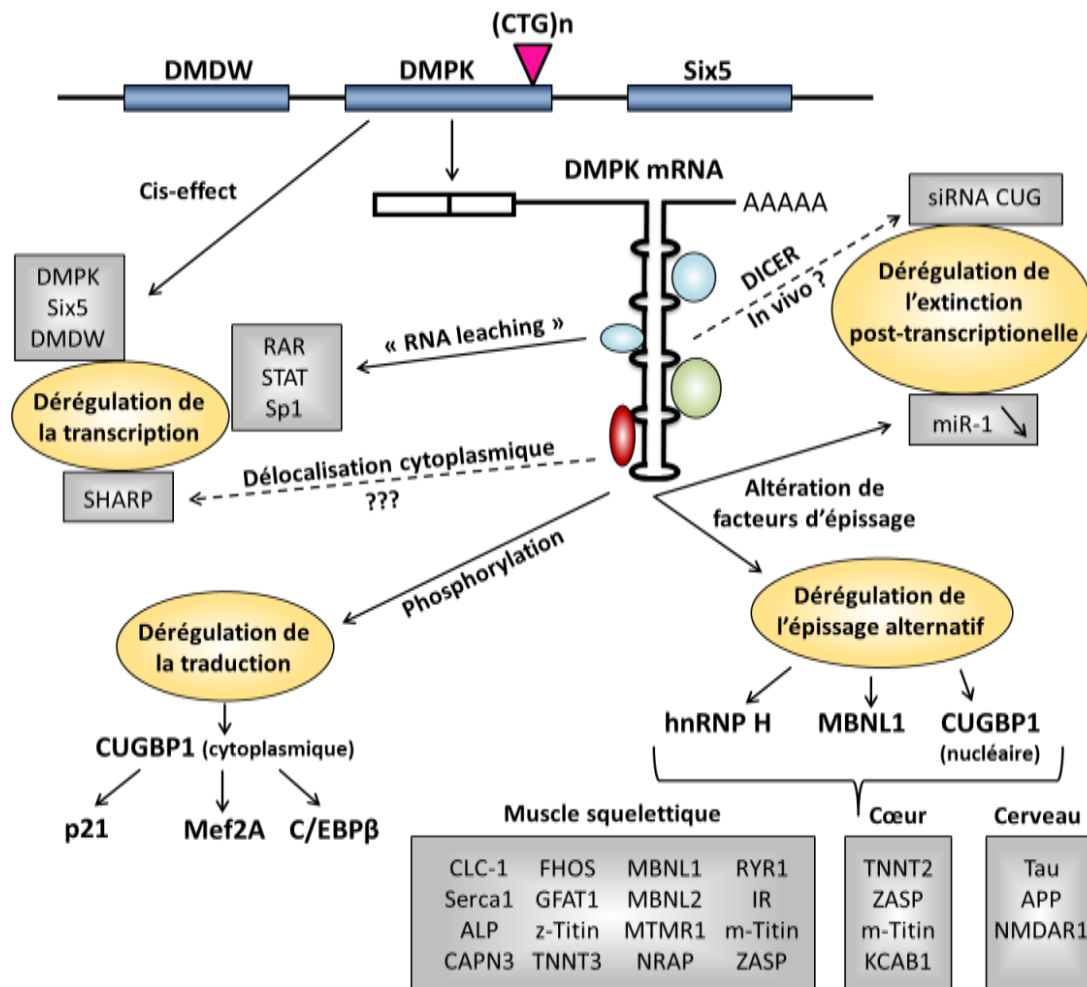


Figure 18 : Les expansions CTG interfèrent avec la régulation de l'expression génique à plusieurs niveaux. Au niveau de la dérégulation de la transcription (haploinsuffisance de DMPK, gènes adjacents dont l'expression est diminuée, délocalisation des facteurs de transcription RAR, STAT, Sp1 et SHARP), de l'extinction post-transcriptionnelle (les expansions CUG sont la cible de DICER et l'expression de miR-1 est diminuée chez les patients), de l'épissage alternatif (altération d'une vingtaine de transcrits dans les muscles et le cerveau) et de la traduction. Ce schéma est inspiré de l'article de Junghans et mis à jour avec les dernières découvertes de l'année 2011. (Llamusi & Artero, 2008; Rau et al., 2011; Dansithong et al., 2011).

1) Haploinsuffisance de la DMPK

Comme nous l'avons vu précédemment, la mutation associée à la pathologie DM1 est localisée dans la région 3'UTR du gène codant la protéine DMPK. Il s'agit d'une sérine/thréonine protéine kinase qui présente des homologies de séquences avec des kinases de la famille Rho qui sont impliquées dans le trafic intracellulaire et l'organisation du cytosquelette (Ridley, 2001; Furukawa *et al.*, 2005). Le gène codant cette protéine est également épissé de façon alternative et donne naissance à différentes isoformes dont la plus abondante est une isoforme de 80 kDa exprimée majoritairement dans les muscles squelettiques, lisses et cardiaque (Kaliman & Llagostera, 2008).

L'une des caractéristiques de la pathologie DM1 est l'haplo-insuffisance de la protéine DMPK chez les patients. En effet, les premières études tentant d'établir un lien entre l'expansion de triplets et la pathologie révélèrent une diminution de la quantité d'ARNm de la DMPK dans différents tissus provenant de patients DM1 (Carango *et al.*, 1993; Hofmann-Radvanyi *et al.*, 1993; Novelli *et al.*, 1993; Koga *et al.*, 1994; Wang *et al.*, 1995). La quantité de la protéine DMPK était également diminuée dans le tissu musculaire squelettique et le myocarde mais cette diminution semblait indépendante de la taille des expansions (Salvatori *et al.*, 2005). D'après ce modèle, les grandes répétitions CUG empêchent l'export du transcrit mutant hors du noyau ce qui aboutit en la diminution de la quantité d'ARNm et de la protéine entraînant une perte de fonction (Salvatori *et al.*, 2005). Afin d'analyser si cette haplo-insuffisance pourrait être à l'origine de plusieurs symptômes, des souris dont le gène *DMPK* a été invalidé ont été générées. Ces souris présentent des anomalies de conduction des canaux calciques et sodiques dans les muscles squelettiques et cardiaques et une myopathie d'intensité modérée d'apparition tardive. On peut également noter l'absence de myotonie (Reddy *et al.*, 1996; Jansen *et al.*, 1996; Berul *et al.*, 1999; Furling *et al.*, 2001).

D'autres études ont tenté de mieux comprendre la fonction de la DMPK afin d'associer sa perte de fonction partielle due à l'haplo-insuffisance, aux éventuels mécanismes participant à la physiopathologie de la DM1. Peu d'éléments sont connus sur les voies de signalisation de la protéine DMPK, elle semblerait cependant être activée par la voie des Rho-GTPase (Rac-1 et Raf-1) (Shimizu *et al.*, 2000). Récemment, une équipe a analysé les partenaires protéiques de la protéine DMPK par GST pull-down et spectrométrie de masse (Figure 19) et a identifié des molécules de choc thermique comme HSB5, HSP25/HSPB1 et l' α -B cristallin mais le rôle de ces protéines dans la fonction de la DMPK reste inconnu (Forner *et al.*, 2010). De plus, certains substrats de la protéine DMPK, dont plusieurs sont communs avec la protéine

kinase C, tels que le phospholamban et le phospholemman (tous deux impliqués dans la conductance électrique cardiaque) ou la CUGBP1 (dont le rôle sera développé ultérieurement) ont été identifiés (Wansink *et al.*, 2003).

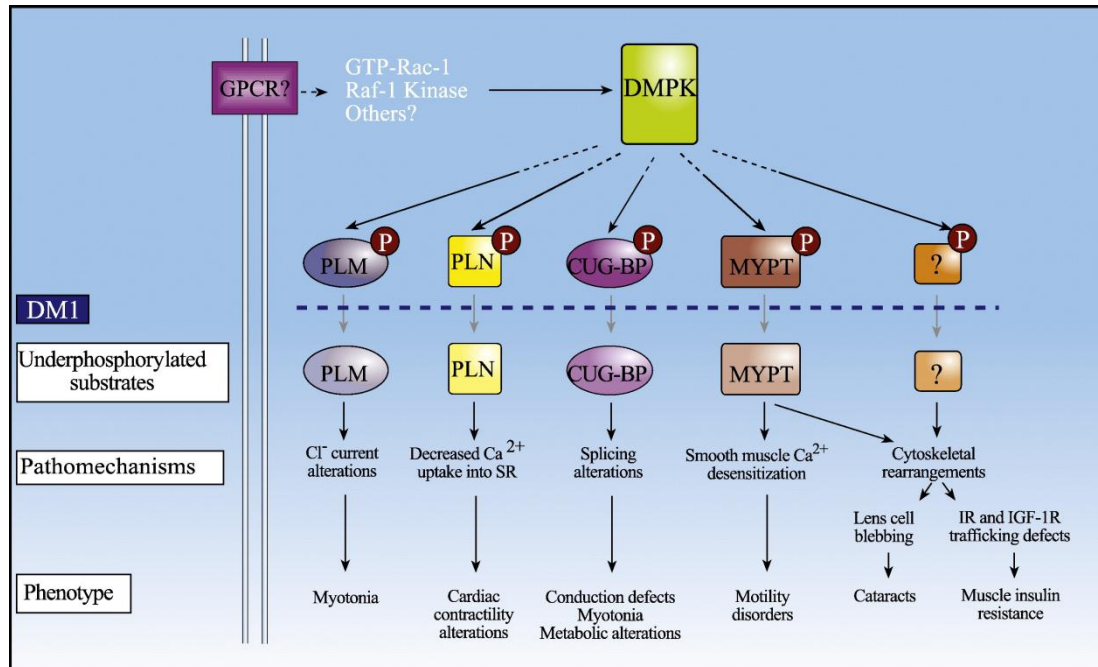


Figure 19 : Possibles implications pathologiques de l'activité kinase de DMPK dans la Dystrophie Myotonique de type 1. L'activité kinase de DMPK est induite par la liaison des GTPases Rac-1 et de la kinase Raf-1. Les récepteurs membranaires impliqués dans cette signalisation restent encore inconnus. La phosphorylation des cibles de DMPK peut être corrélée aux diverses altérations physiologiques rencontrées dans la pathologie. PLM = phospholemman ; PLN = phospholamban ; MYPT1, Myosin Phosphatase-1 (Kaliman & Llagostera, 2008).

Plus récemment, un article et une revue provenant de l'équipe de Kaliman a démontré un rôle de l'haploinsuffisance de la protéine DMPK dans la régulation de l'homéostasie du glucose en découvrant des défauts d'assimilation des sucres associés à un défaut de translocation du récepteur de l'insuline (Llagostera *et al.*, 2007, 2009) sans pour autant conclure sur le rôle de DMPK dans l'insulino-résistance caractéristique de la pathologie.

Pour conclure, les données issues de l'étude des souris dont le gène DMPK a été invalidé et les travaux effectués sur DMPK indiquent que l'haplo-insuffisance ne joue probablement pas un rôle majeur dans la physiopathologie de la DM1, suggérant que d'autres voies de signalisation pourraient être à l'origine des symptômes de la maladie.

2) Diminution d'expression des gènes adjacents au gène DMPK

Les répétitions de trinuécléotides CTG dans l'ADN forment des structures stables en épingle à cheveux *in vivo* (Freudenreich *et al.*, 1997). La structure chromatinienne à proximité des répétitions CUG est hautement perturbée et anormalement condensée par une concentration locale de nucléosomes (Otten & Tapscott, 1995; Wang, 2007). Cette région hautement condensée semble altérer l'efficacité de la transcription du gène DMPK, ainsi que les gènes directement adjacents au gène DMPK sur le chromosome 19, DMWD en amont et SIX5 en aval (Figure 20) (Frisch *et al.*, 2001). En effet, l'expression de ces deux gènes est relativement diminuée à cause des grandes expansions CUG (Thornton *et al.*, 1997; Alwazzan *et al.*, 1999; Frisch *et al.*, 2001). Alors que la protéine DMWD pourrait être impliquée dans les symptômes cérébraux ainsi que l'atrophie testiculaire et les problèmes d'infertilité observés dans la pathologie (Sarkar *et al.*, 2004), une variation de l'expression du gène SIX5, impliqué dans le développement de l'œil chez la drosophile, pourrait expliquer les anomalies oculaires. En effet, les souris KO pour SIX5 présentent une cataracte sous leurs formes hétérozygotes ou homozygotes (Klesert *et al.*, 2000; Sarkar *et al.*, 2000). Cependant, ce type de cataracte est différent de celui retrouvé chez les patients DM1, ne présentant ni opacité iridescente, ni localisation sous-capsulaire postérieure. Bien que ces modèles de souris soient en accord avec un effet cis de la mutation, ils ne présentent aucune altération musculaire (Klesert *et al.*, 2000; Sarkar *et al.*, 2000).

En plus de l'effet cis des répétitions CTG sur l'expression des gènes adjacents s'ajoutent des modifications des marques génétiques de part et d'autre de la mutation. En effet, les répétitions CTG sont ancrées dans un large îlot CpG de 3,5 kb s'étendant de l'extrémité 3' du gène DMPK jusqu'au promoteur du gène SIX5 (Boucher *et al.*, 1995). Lorsque cette région n'est pas méthylée, la protéine CTCF se fixe sur son domaine de reconnaissance et constitue ainsi une barrière physique, également surnommée "insulateur", qui isole les gènes DMPK et SIX5, permettant ainsi une régulation propre via leur promoteur spécifique. Lorsque cette région est méthylée, la protéine CTCF ne peut plus se fixer, conduisant à l'inefficacité de l'insulateur et donc à des anomalies d'expression génique. De récentes études suggèrent que les méthylations des îlots CpG et la fixation de la protéine CTCF pourraient contribuer à l'instabilité somatique caractérisée chez les patients (Gorbunova *et al.*, 2004; Libby *et al.*, 2008; Cleary *et al.*, 2010).

Une étude sur le profil de méthylation des îlots CpG a révélé que bien qu'il n'y ait pas de méthylation allèle-spécifique ou de marques épigénétiques parentales, une hyperméthylation était présente dans les cellules provenant de patients atteints de forme congénitale (Steinbach *et al.*, 1998; Filippova *et al.*, 2001). Dernièrement, il a été observé que cette hyperméthylation n'est pas restreinte à la forme congénitale et est également retrouvée pour les formes moins sévères et ne semble pas être corrélée avec la taille des expansions CTG (Spits *et al.*, 2010; López Castel *et al.*, 2011a).

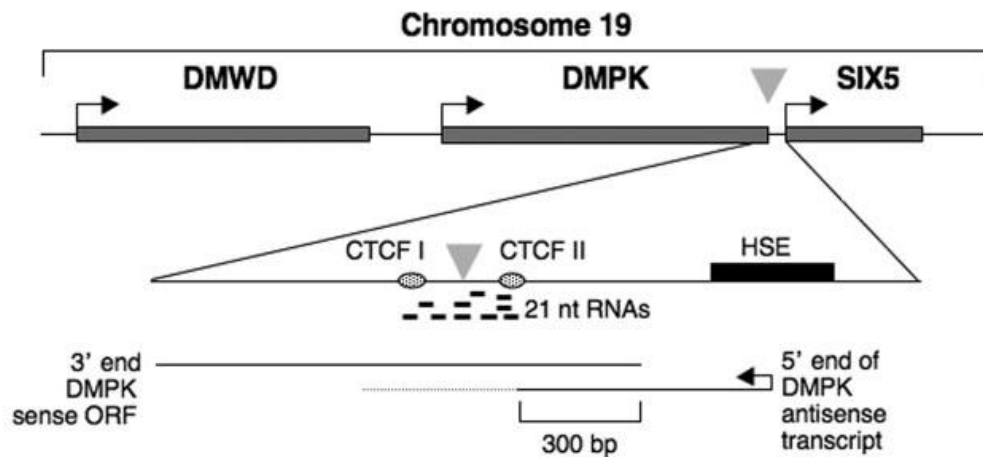


Figure 20 : Organisation du locus DMPK humain. La région localisée entre la fin de la phase codante de DMPK et le début du promoteur du gène SIX5. La localisation des répétitions CTG_CAG est indiquée par un triangle, ainsi que les sites CTCF adjacents aux répétitions (Kumari & Usdin, 2009).

Dans un article récent, l'équipe de Christopher Pearson a montré que les grandes répétitions CTG démarquent une barrière pour la méthylation des îlots CpG. En effet, les séquences situées en amont, incluant le site CTCF, sont méthylées alors que les répétitions elles-mêmes ou les séquences en aval ne le sont pas (López Castel *et al.*, 2011b). Par ailleurs, ce mécanisme pourrait également être commun à d'autres pathologies à triplets comme par exemple l'ataxie de Friedreich, et le syndrome de l'X fragile (Kumari & Usdin, 2009).

Pour finir, il existe dans cette région un gène anti-sens, avec son propre promoteur, qui chevauche la région contenant les répétitions CTG en l'encadrant de deux sites de reconnaissance pour la protéine à doigt de zinc CTCF (Filippova *et al.*, 2001; Cho *et al.*, 2005). Ce gène pourrait être transcrit et ainsi contenir chez les patients DM1 de longues expansions CAG dans la phase codante pouvant donner lieu à une protéine contenant des expansions d'acides aminés glutamine (Q) retrouvées dans d'autres maladies génétiques (maladie de Huntington, ataxie spinocérébelleuse). Cependant, aucune trace de cette protéine n'a été identifiée chez les patients DM1.

3) Gain de fonction de l'ARN et altération de mécanismes liés au métabolisme de l'ARN

La découverte en 1998, d'un second type de Dystrophie Myotonique (DM2), cliniquement proche de la DM1, mais causée par des répétitions CCTG dans un autre gène sur un autre chromosome que le gène *DMPK*, permis l'émergence de l'hypothèse en faveur d'un effet « trans » de ces expansions, indépendamment du gène affecté. La première étude confirmant le rôle de l'ARN mutant dans la physiopathologie fut une expérience dans laquelle une construction de 200 CTG, incluse dans la séquence 3'UTR du gène *DMPK*, a été exprimée de façon transitoire dans des myoblastes sains. L'expression du transcrite mutant fut suffisante pour déclencher un effet toxique se manifestant par un effet inhibiteur de la myogenèse, empêchant la fusion des myoblastes en myotubes (Amack *et al.*, 1999). Cette hypothèse fut confirmée par le développement de différents modèles murins de la pathologie. La première étude fut reportée par le laboratoire de Thornton qui démontra que l'expression spécifiquement musculaire d'un transcrite codant l'alpha-actine humaine, contenant 250 répétitions CTG au niveau de la partie 3' UTR aboutissait au développement de la myotonie chez les souris (Mankodi *et al.*, 2000). Ainsi, l'effet pathologique pourrait être dû au transcrite ARN et non à l'ADN. Ces observations ont été confirmées dans plusieurs autres modèles murins (Mankodi *et al.*, 2000; Seznec *et al.*, 2001; Kanadia *et al.*, 2006; Mahadevan *et al.*, 2006; Gomes-Pereira *et al.*, 2007; Orengo *et al.*, 2008; Yadava *et al.*, 2008; Dansithong *et al.*, 2008; Vignaud *et al.*, 2010) chez *Drosophila melanogaster* (de Haro *et al.*, 2006), chez *Caenorhabditis elegans* (Chen *et al.*, 2007; Wang *et al.*, 2008b) ou encore le poisson zèbre (Machuca-Tzili *et al.*, 2011). Ces modèles qui se caractérisent par une expression particulière (tissu spécifique ou inductible) des répétitions dans un contexte déterminé (promoteur, gène, 3'UTR) montrent que les expansions CTG permettent d'induire un phénotype pathologique multisystémique au niveau des muscles squelettiques, du cœur ainsi que du système nerveux central.

L'analyse par microscopie électronique a révélé que les répétitions CUG forment des structures en tige-boucles composées d'ARN double-brin (Michalowski *et al.*, 1999) qui ont par ailleurs la capacité d'interagir avec la protéine kinase PKR (Tian *et al.*, 2000) et les protéines de la famille Muscleblind (Miller *et al.*, 2000). Des données récentes de cristallographie ont confirmé que les répétitions CUG forment des hélices double-brin antiparallèles similaires à la forme A de l'ARN à l'exception des mésappariements U-U (Mooers *et al.*, 2005). Cette structure particulière est à l'origine d'un modèle qui confère à

l'ARN mutant un gain de fonction par lequel les ARN contenant de longues répétitions CUG pourraient séquestrer certaines protéines se liant à l'ARN altérant leur localisation cellulaire normale. Bien que les transcrits mutants soient correctement coiffés, épissés et polyadénylés, cette conformation empêche leur export vers le cytoplasme entraînant leur accumulation dans le noyau des cellules (Davis *et al.*, 1997) sous forme d'inclusions ribonucléiques intranucléaires appelées foci (Taneja *et al.*, 1995) représenté en Figure 21.

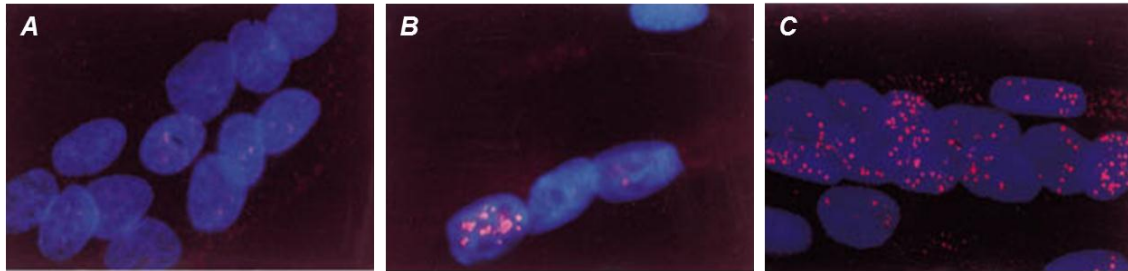


Figure 21 : Visualisation des foci nucléaires dans des myoblastes humains. Les répétitions CUG de l'ARN DMPK mutant, observable par RNA-FISH à l'aide d'une sonde $(CAG)_{30}$ -cy3 sont représentées en rouge. Les noyaux des cellules sont visualisés par marquage au DAPI. A=sujet sain, B=patient DMI faiblement atteint (150 répétitions CTG) et C=patient DMI fortement atteint. (Davis *et al.*, 1997).

Le mécanisme par lequel l'ARNm est retenu dans le noyau n'est pas encore élucidé. Cependant ces structures sont associées à des structures nucléaires appelées «nuclear speckles», connues pour être des sites de stockage des petits ARNs et un site d'épissage des pré-ARN (Spector, 1993). L'allèle normal transite par ces speckles nucléaires puis par le pore nucléaire pour être traduit. A l'opposé, l'ARN DMPK mutant reste bloqué à l'entrée des nuclear speckles, suggérant un rôle des expansions de triplets dans la rétention nucléaire du transcrit mutant (Holt *et al.*, 2007; Smith *et al.*, 2007). Ces foci nucléaires peuvent être retrouvés dans les modèles cellulaires, les modèles animaux ainsi que dans les tissus provenant de patients atteints de la pathologie.

Ces inclusions, identifiables par hybridation in situ (RNA-FISH) en utilisant une sonde fluorescente complémentaire aux répétitions (Taneja *et al.*, 1995; Davis *et al.*, 1997) sont également responsables de la séquestration de protéines dont certaines sont connues pour être impliquées dans le processus d'épissage alternatif, tel que la protéine muscleblind1 (MBNL1) ou certains facteurs de transcription à l'origine de la réduction de l'expression de certains gènes. De façon intéressante, l'expression de la mutation DM2 est également associée à la formation de foci nucléaires qui séquestrent MBNL1 (Fardaei *et al.*, 2001; Mankodi *et al.*, 2001; Fardaei *et al.*, 2002).

a) Effet trans des répétitions CUG sur la transcription ou "RNA leaching".

Un certain nombre de facteurs de transcription dont la localisation nucléaire normale au niveau de la chromatine est altérée dans les modèles cellulaires DM1 au profit d'une séquestration au niveau des foci ribonucléoprotéiques a été identifié (Ebraldidze *et al.*, 2004). Ce processus, appelé le "RNA leaching", fut en effet prouvé par l'observation d'une baisse importante de la fraction protéique associée à la chromatine des facteurs Specificity protein 1 (Sp1), Signal Transducer and Activator of Transcription (STAT1 and STAT3) et Reinoic Acid Receptor gamma (RAR γ), après expression de l'ARN DMPK pendant trois semaines dans un modèle cellulaire. La séquestration de ces facteurs est également à l'origine de la baisse de transcription de plusieurs gènes, dont le gène codant le canal chlore (CLCN1), qui est impliqué dans la myotonie (Ebraldidze *et al.*, 2004). Le facteur MyoD, un facteur de transcription ayant un rôle-clé dans la détermination et la différenciation des cellules musculaires striées est également altéré par les répétitions. Son niveau protéique est réduit dans les myoblastes qui expriment la partie 3'UTR du gène DMPK contenant de longues répétitions CUG (Amack *et al.*, 2002). Les modifications d'activité de facteurs de transcription entraînent une altération de la transcription de certain gènes-cibles; la comparaison des profils d'expression génique par puces à ADN provenant de biopsies musculaires de patients atteints de DM1 et DM2 par rapport à des muscles contrôles a révélé que six gènes, requis pour le métabolisme des canaux calciques et potassiques et les fonctions mitochondriales, étaient dérégulés dans la DM1 et la DM2, renforçant l'hypothèse d'une cause commune entre ces deux types de maladie (Botta *et al.*, 2007).

La Dystrophie Myotonique de type 1 a également été associée avec la surexpression du facteur de transcription *NK2 transcription factor related, locus 5 (Drosophila) (Nkx2-5)* dans les tissus cardiaques de modèles de souris transgéniques corrigées pour la pathologie (Yadava *et al.*, 2008). Dans les tissus humains, cette augmentation est spécifique à la DM1 car non présente dans les autres dystrophies musculaires suggérant une contribution de ce facteur aux symptômes de défaut de conduction cardiaque présents dans la maladie (Yadava *et al.*, 2008).

Récemment, l'équipe de Sita Reddy a montré que la séquestration de ces facteurs de transcription n'était pas le seul mécanisme pouvant causer des altérations d'expression géniques dans la pathologie (Dansithong *et al.*, 2011). Le facteur de transcription *SHARP* (SMART/HDAC1-associated repressor protein) appartient à de nombreux complexes multi-protéiques qui peuvent fonctionner comme activateur ou répresseur de la transcription (Shi *et*

al., 2001). La protéine homologue de SHARP chez la drosophile est le facteur SPEN qui est impliqué dans l'apparition d'un phénotype neuro-dégénératif lors de l'expression d'ARN contenant des expansions CUG (Mutsuddi *et al.*, 2004). L'équipe de Sita Reddy a observé que la localisation cellulaire de SHARP, normalement nucléaire dans des cellules musculaires contrôles, est altérée chez les patients DM1 avec une prédominance pour une localisation cytoplasmique (Dansithong *et al.*, 2011). Cette délocalisation cellulaire est à l'origine d'une baisse du niveau de transcription d'une vingtaine d'ARN messagers, impliqués dans le développement du muscle squelettique, parmi lesquels des facteurs myogéniques (CDH15), des protéines impliquées dans le métabolisme de l'ARN (EIF1AX, DDX3Y) ou du cytosquelette (DMD, PDLIM3). Les défauts d'expression de ces facteurs pourraient être corrélés à certains symptômes de la pathologie (Dansithong *et al.*, 2011). Ceci suggère que les thérapies cellulaires pour guérir la DM1 devront non-seulement rétablir un épissage alternatif approprié mais également corriger les défauts d'expression d'un nombre important de transcrits. Pour conclure, la modification de l'expression de facteurs de transcription par "RNA leaching" ou d'autres mécanismes comme ceux présentés ci-dessus confère une explication convaincante à l'origine des changements d'expression, et ceux-ci peut être corrélé avec les altérations de l'épissage alternatif identifiées dans la pathologie, considérant le lien étroit qu'il existe entre transcription et épissage alternatif .

b) Réexpression des isoformes fœtales à l'origine des symptômes cliniques

L'identification de facteurs protéiques fixées aux expansions de répétitions CUG suggère fortement un modèle de gain de fonction de l'ARN mutant DMPK (Miller *et al.*, 2000) et apporte une explication au niveau moléculaire pour les similarités phénotypiques entre la DM1 et la DM2 vu que des facteurs protéiques similaires pourraient être séquestrés par les répétitions CUG ou CCUG. Plusieurs études ont établi que l'une des raisons à l'origine de la toxicité de l'ARN mutant DMPK est qu'il interfère avec l'activité de facteurs d'épissage parmi lesquels la famille Muscleblind, la CUG-Binding Protein 1 (CUGBP1) et la hnRNP H. La voie métabolique plus généralement admise pour expliquer les symptômes de la pathologie est l'effet trans-dominant des répétitions CUG sur l'épissage alternatif d'un certain nombre de transcrits principalement exprimés au niveau du muscle (squelettique et cardiaque) et du cerveau.

Une liste des ARN pré-messagers dont l'épissage alternatif est dérégulé dans la pathologie peut être trouvée dans le Tableau 5.

| Tissu | Gènes | Exon/intron | Inclusion/exclusion | Références |
|--|---|-------------|---------------------|------------------------|
| Muscle squelettique | Insulin Receptor (INSR) | Exon 11 | Exclusion | Savkur et al., 2001 |
| | Chloride Channel (CLCN-1) | Intron 2 | Inclusion | Charlet-B et al., 2002 |
| | | Exon 7a | Inclusion | |
| | Fast skeletal troponin T (TNNT3) | Foetal exon | Inclusion | Kanadia et al., 2003 |
| | Amphiphysin II (BIN1) | Exon 11 | Inclusion | Fugier et al., 2011 |
| | Ryanodine Receptor (RyR) | Exon 70 | Exclusion | Kimura et al., 2005 |
| | Sarcoplasmic/endoplasmic reticulum Ca ²⁺ ATPase 1 (SERCA1) | Exon 22 | Exclusion | |
| | Sarcoplasmic/endoplasmic reticulum Ca ²⁺ ATPase 2 (SERCA2) | Intron 19 | Inclusion | |
| | LIM domain binding 3 (LBB3, ZASP) | Exon 11 | Inclusion | Lin et al., 2006 |
| | Titin (TTN) | Zr4 | Inclusion | |
| | | Zr5 | Inclusion | |
| | Nebulin-related anchoring protein (NRAP) | Exon 12 | Inclusion | |
| | Calpain 3 (CAPN3) | Exon 16 | Exclusion | |
| | Attractin-like 1 (ATRNL 1, ALP) | Exon 5a | Exclusion | |
| | | Exon 5b | Exclusion | |
| | Formin homology 2 domain containing 1 (FHOD1, FHOS) | Exon 11a | Exclusion | |
| | Glutamine-fructose-6-phosphate transaminase 1 (GFPT1, GFAT1) | Exon 10 | Exclusion | |
| | Muscleblind 1 (MBNL1) | Exon 7 | Inclusion | |
| | Muscleblind 2 (MBNL2) | Exon 7 | Inclusion | |
| | Nuclear factor I/X (CCAAT-binding transcription factor (NFIX) | Exon 123 | Inclusion | Du et al., 2010 |
| SET and MYND domain containing 1 (SMYD1) | Exon 39 | Inclusion | | |
| Sperm associated antigen 9 (SPAG9) | Exon 39 | exclusion | | |
| Myomesin 1 (MYOM1) | Exon 17a | inclusion | Koebis et al., 2011 | |
| Muscle squelettique et coeur | Myotubularin-related protein 1 (MTMR1) | Exon 2.1 | exclusion | Ho et al., 2005 |
| | | Exon 2.3 | exclusion | |
| | α-dystrobrevin (DTNA) | Exon 11a | Inclusion | Nakamori et al., 2008 |
| | | Exon 12 | Inclusion | |
| Coeur | Cardiac Troponin T (TNNT2) | Exon 5 | Inclusion | Philips et al., 1998 |
| Cerveau | Tau (MAPT) | Exon 2 | exclusion | Sergeant et al., 2001 |
| | | Exon3 | exclusion | |
| | | Exon 10 | exclusion | |
| | N-methyl-D-aspartate receptor (NMDAR1) | Exon 5 | Inclusion | Jiang et al., 2004 |
| | Amyloid precursor protein (APP) | Exon 7 | exclusion | |

Tableau 5 : Liste des gènes dont l'épissage alternatif est dérégulé dans la DM1 suivant les tissus. Le numéro de l'exon (ou intron) alternatif est indiqué ainsi que le type d'évènement d'épissage (inclusion ou exclusion).

La particularité de la dérégulation de l'épissage alternatif des transcrits dans la Dystrophie Myotonique est le fait que le profil d'épissage de ces transcrits à l'âge adulte correspond dans la grande majorité à un profil d'épissage alternatif de type fœtal (Buj-Bello *et al.*, 2002; Charlet-B *et al.*, 2002; Jiang *et al.*, 2004; Kalsotra *et al.*, 2008). L'hypothèse proposée actuellement est que l'expression d'isoformes fœtales, incapable de remplir les besoins physiologiques des tissus adultes, serait à l'origine des divers symptômes observés dans la pathologie.

Gènes musculaires et anomalies musculaires/cardiaques :

En 1998, l'équipe de Tom Cooper publia la première étude qui démontra que l'ARN pré-messager codant la troponine T cardiaque (*TNNT2*) est anormalement épissé dans les muscles striés des patients DM1 ainsi que dans des modèles cellulaires exprimant les répétitions CTG (Philips *et al.*, 1998). L'épissage alternatif de la troponine T est régulé au cours du développement cardiaque par l'inclusion ou non de l'exon alternatif 5 qui détermine une forme protéique adulte ou fœtale. L'augmentation accrue de l'inclusion de l'exon 5 (Figure 22) aboutit à la production d'une isoforme fœtale chez l'adulte, incapable de remplir son rôle physiologique, qui semble être l'une des origines des troubles de la conduction cardiaque chez les patients DM1 (Philips *et al.*, 1998).

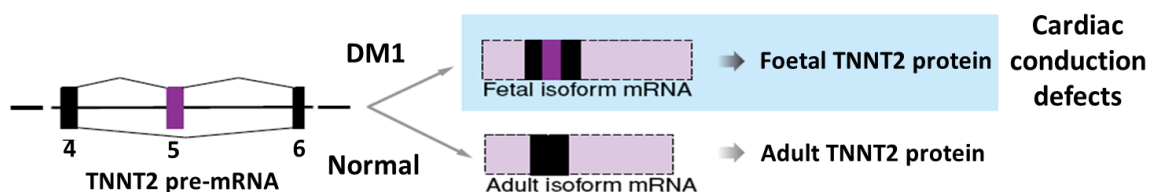


Figure 22 : Dérégulation de l'épissage alternatif de *TNNT2* dans la DM1. L'inclusion de l'exon alternatif 5 chez les patients DM1 entraîne la production d'une protéine fœtale qui pourrait être à l'origine des troubles de conduction cardiaque (Ranum & Cooper, 2006).

Il existe également des anomalies d'épissage d'autres gènes musculaires comme le récepteur à la Ryanodine RYR1 (Kimura *et al.*, 2005, 2009), le récepteur calcique SERCA2 (Kimura *et al.*, 2005) ou la myotubularine MTMR1 (Buj-Bello *et al.*, 2002). Ceux-ci sont proposés comme intervenant dans la physiopathologie DM1 soit dans l'atrophie/faiblesse musculaire soit dans des anomalies de conductions cardiaques.

Enfin très récemment, l'équipe de Nicolas Charlet-Berguerand a identifié un nouveau gène dérégulé dans les pathologies DM1 et DM2. Le gène *Bridging integrator-1* (*BIN1*) code pour l'amphiphysine II, une protéine membranaire dont la fonction est régulée par épissage alternatif. L'exon alternatif 11 de ce gène code un domaine essentiel dans la biogénèse des

tubules T, structure membranaire fondamentale pour le couplage excitation-contraction dans le muscle squelettique. L'exclusion de cet exon dans la pathologie DM (Figure 23) aboutit à la production d'une isoforme protéique non fonctionnelle à l'origine d'une altération structurelle des tubules T et d'une faiblesse musculaire chez la souris (Fugier *et al.*, 2011). Des mutations ponctuelles du gène BIN1 sont également retrouvées dans d'autres pathologies, dont la myopathie centro-nucléaire, suggérant qu'une altération de l'expression correcte de BIN1 pourrait, au-delà de la Dystrophie Myotonique, être une caractéristique commune à de nombreuses maladies musculaires.



Figure 23 : *Dérégulation de l'épissage alternatif de BIN1 dans la DM1. L'exclusion de l'exon alternatif 11 chez les patients DM1 entraîne la production d'une isoforme protéique inactive entraînant une altération structurelle des tubules T et à l'origine d'une faiblesse musculaire. Inspiré de (Ranum & Cooper, 2006) d'après les travaux de (Fugier *et al.*, 2011).*

Récepteur à l'insuline et diabète :

Les patients DM1 développent entre autres un diabète de type 2 caractérisé par une résistance à l'insuline. Grâce à l'utilisation d'études génétiques, il fut reporté que l'épissage alternatif de l'ARN pré-messager codant le récepteur de l'insuline (INSR) était également dérégulé dans la pathologie. La cause provient d'une diminution de l'inclusion de l'exon alternatif 11 chez les patients DM1 (Figure 24) (Savkur *et al.*, 2001).



Figure 24 : *Dérégulation de l'épissage alternatif du gène codant le récepteur à l'insuline dans la DM1. L'exclusion de l'exon alternatif 11 chez les patients DM1 entraîne la production d'une isoforme protéique fœtale moins performant chez l'adulte que l'isoforme adulte à l'origine de la résistance à l'insuline (Ranum & Cooper, 2006).*

Il existe deux isoformes de l'INSR produites à partir de l'épissage alternatif du gène. Ces deux isoformes sont spécifiques de tissus et diffèrent par l'inclusion ou non de l'exon 11 composé de 36 nucléotides (Moller *et al.*, 1989). L'IR-B, l'isoforme incluant l'exon 11, est

principalement exprimée au niveau des tissus responsables de l'homéostasie du glucose incluant le foie, les tissus adipeux ou les muscles squelettiques et présente une forte réponse à l'insuline. En revanche, l'isoforme non musculaire, l'IR-A (absence de l'exon 11), présente deux fois plus d'affinité pour l'insuline, un temps d'internalisation et de recyclage plus rapide que l'IR-B (McClain, 1991; Vogt *et al.*, 1991) mais possède une plus faible activité (Kellerer *et al.*, 1992). L'exclusion de l'exon alternatif dans la DM1 aboutit ainsi à la formation d'un récepteur à l'insuline moins fonctionnel et peut être corrélé avec une résistance à l'insuline observée chez les patients (Stuart *et al.*, 1983).

Canal chlore et myotonie :

Nous pouvons citer le transcrite codant le canal chlore muscle-spécifique dont la majorité des ARNm détectés par RT-PCR dans les muscles squelettiques de patients DM1 contiennent un codon stop prématuré due à l'inclusion de l'exon alternatif 7a entre les exons 7 et 8, conduisant à l'absence de protéine, condition suffisante pour expliquer la myotonie dans la pathologie (Figure 25) (Charlet-B *et al.*, 2002; Mankodi *et al.*, 2002; Lueck *et al.*, 2007a, 2007b).



Figure 25 : Dérégulation de l'épissage alternatif du canal chlore dans la DM1. L'inclusion de l'exon alternatif 7a dans l'ARNm chez les patients DM1 entraîne l'apparition d'un codon stop prématuré. Le transcrite est dégradé par le Nonsense mediated mRNA decay, résultant en l'absence de protéine et le développement d'une myotonie (Ranum & Cooper, 2006).

Un autre argument en faveur de l'importance de cette dérégulation dans les symptômes cliniques est venu de l'équipe de Charles Thornton qui montra qu'un oligonucléotide ciblant le site 3' d'épissage de l'exon 7a corrigeait le défaut d'épissage alternatif dans deux modèles de souris DM1, restaurant le cadre de lecture correct et l'expression de la protéine à la surface membranaire, et diminuant progressivement la myotonie (Wheeler *et al.*, 2007).

Symptômes cérébraux et anomalies neuronales:

Outre le muscle squelettique et le cœur, l'épissage alternatif de plusieurs transcrits au niveau du cerveau est également dérégulé chez les patients DM1 ce qui pourrait expliquer en partie certaines anomalies du système nerveux central. Nous pouvons citer les ARN pré-messagers Tau (MAPT) (Sergeant *et al.*, 2001; Jiang *et al.*, 2004; Leroy *et al.*, 2006; Dhaenens *et al.*, 2008; Ghanem *et al.*, 2009; Dhaenens *et al.*, 2011) N-méthyl-D-aspartate receptor (NMDAR1) (Jiang *et al.*, 2004) et amyloid precursor protein (APP) (Jiang *et al.*, 2004).

Comme nous allons le voir, la dérégulation de l'épissage alternatif dans la pathologie est principalement due à l'altération de deux facteurs d'épissage alternatifs antagonistes, MBNL1 et CUGBP1, qui participent aux changements d'épissage au cours du développement post-natal (Figure 26) (Kalsotra *et al.*, 2008).

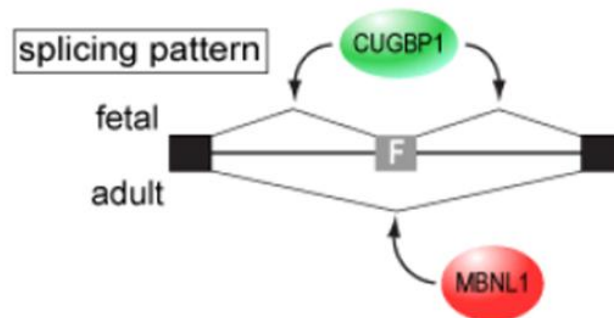


Figure 26 : Effet antagoniste des protéines *CUGBP1* et *MBNL1* dans la régulation de l'épissage alternatif. (Lee & Cooper, 2009).

c) MBNL1 et sa dérégulation dans la pathologie

Généralités

Les protéines Muscleblind-like (MBNL) font partie d'une famille de protéines régulatrices des ARN très conservées chez une grande variété d'espèces : chez l'humain (Miller *et al.*, 2000), la souris et le rat (Kanadia *et al.*, 2003b; Hao *et al.*, 2008) chez la drosophile *Drosophila melanogaster* (Begemann *et al.*, 1997; Artero *et al.*, 1998; Houseley *et al.*, 2005, 2006; Machuca-Tzili *et al.*, 2006; Goers *et al.*, 2008; Fernandez-Costa & Artero, 2010), le poulet (Huang *et al.*, 2008), le ver *Caenorhabditis elegans* (Wang *et al.*, 2008b;

Sasagawa *et al.*, 2009; Wang *et al.*, 2011), le *zebrafish* (Liu *et al.*, 2008; Machuca-Tzili *et al.*, 2011).

Chez l'homme la famille Muscleblind-like est composée de trois protéines homologues : MBNL1, MBNL2 et MBNL3.

MBNL1, dont le gène est localisé sur le chromosome 3, a une expression ubiquitaire avec une prédominance dans le muscle, le cœur et le cerveau. Par ailleurs, son expression augmente à l'âge adulte. MBNL2, dont le gène est localisé sur le chromosome 13, a également une expression ubiquitaire et une prédominance semblable à MBNL1. Enfin, MBNL3, dont le gène est localisé sur le chromosome X est exprimé très majoritairement dans le placenta (Fardaei *et al.*, 2002). De plus, son expression diminue au cours de la différenciation cellulaire.

Ces trois gènes sont soumis à l'épissage alternatif ce qui provoque la production de dizaines d'isoformes protéiques dont l'expression est régulée de façon tissu-spécifique et selon le stade de développement (Pascual *et al.*, 2006; Tran *et al.*, 2011). Par mesure de clarté, j'aborderai dans la suite de cette introduction uniquement le rôle de la protéine MBNL1 sur laquelle la majorité des travaux scientifiques s'est concentré.

Structure/fonction de la protéine MBNL1

Les protéines MBNL contiennent quatre motifs en doigts de zinc très conservés avec une région de jonction entre les doigts de zinc 2 et 3 (Fardaei *et al.*, 2002; Squillace *et al.*, 2002). Les doigts de zinc 1 et 3 sont de type CX7CX6CX3H alors que les 2 et 4 sont de type CX7CX4CX3H. Les quatre doigts de zinc présents au niveau de la région N terminale sont nécessaires à la liaison à l'ARN (Kino *et al.*, 2004). En ce qui concerne la région intermédiaire entre les doigts de zinc 2 et 3, celle-ci semble agir sur la structure de la protéine, permettant aux quatre doigts de zinc d'agir comme une plate-forme flexible permettant la fixation de la protéine sur ces cibles ARN (Kino *et al.*, 2004; Teplova & Patel, 2008; Cass *et al.*, 2011). Enfin, la région C-terminale n'est pas impliquée dans la fixation à l'ARN mais requise pour l'interaction avec d'autres protéines dont MBNL1 elle-même, formant ainsi des homodimères ou la protéine YB-1 (Yuan *et al.*, 2007; Onishi *et al.*, 2008).

La protéine MBNL1 a tout d'abord été décrite chez la drosophile comme un facteur régulateur de la différenciation des cellules photo-réceptrices et musculaires (Begemann *et al.*, 1997; Artero *et al.*, 1998). Depuis quelques années, MBNL1 est également décrit comme facteur d'épissage, à la fois activateur ou répresseur d'un nombre important de transcrits

musculaires et cérébraux (Philips *et al.*, 1998; Sergeant *et al.*, 2001; Charlet-B *et al.*, 2002; Ho *et al.*, 2004; Savkur *et al.*, 2001; Kimura *et al.*, 2005; Goers *et al.*, 2010; Fugier *et al.*, 2011).

Plusieurs études, notamment issues de SELEX ont pu identifier un motif particulier. Les protéines MBNL se fixent aux motifs CHHG et CHG répétés, où H représente les nucléotides A, U ou C (Miller *et al.*, 2000; Kino *et al.*, 2004) et les motifs YGCU(U/G)Y, où Y représente les pyrimidines (Ho *et al.*, 2004; Goers *et al.*, 2010) présents dans une conformation en double-brin d'ARN et possédant un mésappariement au niveau de la tige (Miller *et al.*, 2000; Warf & Berglund, 2007).

Dans le but d'identifier de nouveaux sites putatifs pour MBNL1, l'équipe de Berglund a analysé les régions introniques bordant l'exon alternatif de 24 gènes humains dérégulés dans la pathologie. Même si ces sites ne sont que putatifs et n'ont pas tous été vérifiés expérimentalement, tous les ARN pré-messagers contiennent un ou plusieurs motifs YGCY dans un intron, et même parfois au niveau de l'exon (Goers *et al.*, 2010). Mis à part le fait que ce motif est dégénéré et qu'il puisse être retrouvé en grande quantité par des outils de prédictions bioinformatiques, il semble y avoir une corrélation entre le facteur d'épissage alternatif MBNL1 et la dérégulation de l'épissage alternatif dans la DM1.

Le laboratoire de Berglund a montré que la fixation de MBNL1 sur ces motifs stabiliserait la structure secondaire en tige-boucle de l'ARN pré-messager de *TNNT2* ce qui pourrait influencer la formation du spliceosome au niveau de l'exon alternatif (Warf *et al.*, 2009)

La capacité de MBNL1 à inclure ou exclure l'exon alternatif dépend de la localisation de son site d'interaction par rapport à cet exon. En effet, la fixation de MBNL1 en amont de l'exon alternatif empêche le recrutement du spliceosome au niveau du site 3' d'épissage aboutissant en l'exclusion de l'exon alternatif. Au contraire, la présence d'un motif de fixation de MBNL en aval de l'exon alternatif favoriserait son inclusion (Goers *et al.*, 2010).

Séquestration de MBNL1 et conséquence sur l'épissage alternatif

En 2000, grâce à des expériences d'immunoprécipitation à partir d'extraits fractionnés de cellules exprimant des répétitions (CUG)_n, le groupe de Swanson a montré que le facteur d'épissage Muscleblind-like 1 (MBNL1) pouvait interagir avec ces expansions (Miller *et al.*, 2000).

Parmi les protéines de la famille MBNL, MBNL1 est la plus abondante (Pascual *et al.*, 2006; Holt *et al.*, 2009).

Dans une ancienne étude, Kino et al. ont montré que les seules isoformes qui contiennent l'exon alternatif 3 sont capables de reconnaître la structure en épingle à cheveux des ARNm mutants contenant des expansions CUG et se retrouvent co-localisées avec les foci dans les noyaux des cellules de patients DM1 (Kino *et al.*, 2004). Cependant, une étude très récente du laboratoire de Nicolas Sergeant contredit ce résultat et montre que si les isoformes sans l'exon 3 ont une affinité réduite pour les expansions CUG in vitro, elles colocalisent tout de même avec les foci in vivo (Figure 27) (Mankodi *et al.*, 2001; Fardaei *et al.*, 2001; Jiang *et al.*, 2004; Tran *et al.*, 2011).

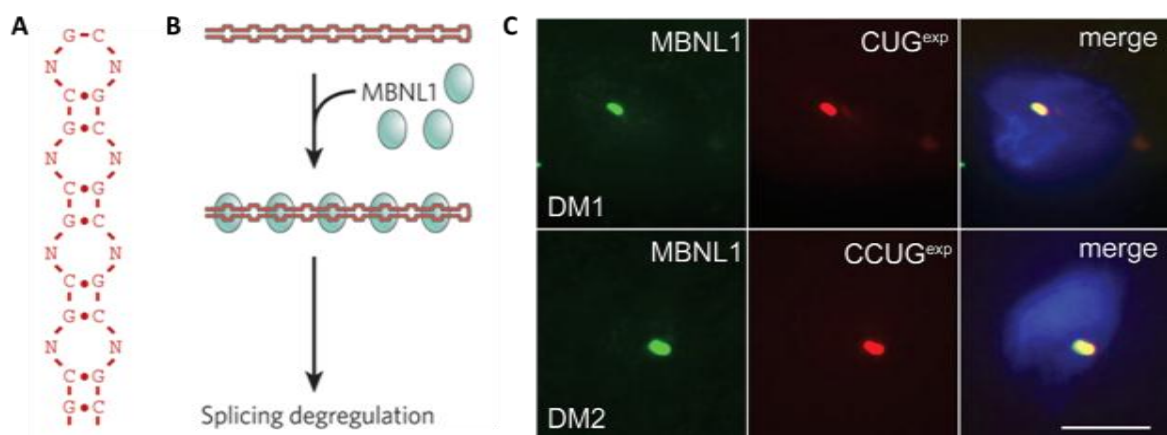


Figure 27 : Dérégulation du facteur d'épissage MBNL1 dans la pathologie DM1. **A.** Les expansions CUG (tout comme les expansions de type CXG) forment une structure en épingle à cheveux stable. **B.** Le modèle de la séquestration de la protéine MBNL1 par les expansions CUG à l'origine des défauts d'épissage alternatif dans la DM1. **C.** Colocalisation de MBNL1 avec les expansions CUG et CCUG dans des tissus musculaires de patients DM1. MBNL1 est observable par immunofluorescence avec un anticorps dirigé contre MBNL1 et un anticorps secondaire couplé à un fluorochrome. Les expansions sont visualisées par RNA-FISH (RNA-Fluorescent In Situ Hybridization) en utilisant une sonde complémentaire couplée à un fluorochrome. Le noyau est marqué au DAPI. (Osborne & Thornton, 2006; Mirkin, 2007).

Le fait que cette protéine soit impliquée dans la différenciation myogénique dans un modèle de drosophile, suggérait un rôle de cette protéine dans la physiopathologie musculaire. Le mécanisme mis en jeu proposé était alors une captation de MBNL1 dans les foci entraînant la diminution de sa concentration libre et finalement une perte de fonction à l'origine de certains défauts d'épissage alternatif (Fardaei *et al.*, 2001, 2002). Par la suite, il fut confirmé que cette protéine jouait un rôle central dans la physiopathologie de la maladie et que sa séquestration par les ARNm entraînait une dérégulation de l'épissage alternatif de nombreux gènes.

Un modèle de souris knock-out (KO) pour MBNL1 a permis d'étudier les répercussions de sa perte de fonction (Kanadia *et al.*, 2003b) et de mimer le contexte des cellules DM1 dans lesquelles la protéine MBNL1 est séquestrée dans le noyau et donc non fonctionnelle. Les souris KO *Mbnl1* présentent différents symptômes caractéristiques de la DM1 telles qu'une myotonie et une cataracte. Sur un plan moléculaire, une modification de l'épissage du gène codant *Tnnt3* (fast skeletal muscle troponin T3) similaire à celle observée chez les patients a été reportée. Cette implication de la déplétion fonctionnelle de MBNL1 dans le phénotype pathologique a été renforcée par l'observation que la surexpression de cette protéine dans les modèles de souris DM1 HSALR permettait de reverser la myotonie ainsi que des défauts d'épissage (Kanadia *et al.*, 2006). Afin d'évaluer l'implication de la perte de fonction de MBNL1 dans le phénotype DM1, des études récentes ont comparé le transcriptome du muscle squelettique de souris KO *mbnl1* et de souris HSALR (exprimant le gène de l'actine humaine contenant des expansions CTG). Plusieurs groupes ont ainsi mis en évidence un effet sur l'expression de gènes touchant des fonctions telles que la signalisation du calcium, l'homéostasie et affectant également des gènes du réticulum sarcoplasmique ainsi que de la matrice (Osborne *et al.*, 2009; Du *et al.*, 2010). L'équipe de Manuel Ares a également souligné l'importance de l'implication de la perte de fonction de MBNL1 dans les défauts d'épissage observés dans la DM1 (Du *et al.*, 2010). En effet, le modèle de souris KO *mbnl1* expliquerait 80 % des défauts d'épissage observés dans le modèle murin HSALR (Du *et al.*, 2010).

d) Dérégulation de CUGBP1

Les protéines CELF (CUG-BP and ETR-3 Like Factors), initialement nommées BRUNOL (Bruno-Like) pour leur homologie avec le facteur régulateur de la traduction Bruno décrit chez la *Drosophila* (Good *et al.*, 2000), forment une famille de protéines relativement conservée qui intervient dans la régulation post-transcriptionnelle des ARN (Ladd *et al.*, 2001). Les membres de la famille CELF sont des facteurs d'épissage alternatif (Philips *et al.*, 1998; Ladd *et al.*, 2001, 2004) et certains membres de cette famille ont également un rôle dans la régulation de la traduction (Timchenko *et al.*, 1999; Mukhopadhyay *et al.*, 2003; Iakova *et al.*, 2004), la déadénylation et la dégradation des ARN (Paillard *et al.*, 2003; Vlasova *et al.*, 2008; Le Tonquèze *et al.*, 2010; Lee *et al.*, 2010b; Rattenbacher *et al.*, 2010) ainsi que dans l'édition des ARN (Anant *et al.*, 2001).

La famille CELF est caractérisée par 6 membres présentant de fortes homologies de séquence : CUG-Binding Protein 1 (CUGBP1), ETR-3 (également surnommée CUGBP2),

CELF3, CELF4, CELF5 et CELF6. Ces facteurs sont caractérisés par trois domaines de liaison à l'ARN de type RRM leur permettant d'interagir spécifiquement avec des motifs riches en U/G (Philips *et al.*, 1998; Savkur *et al.*, 2001; Charlet-B *et al.*, 2002; Mori *et al.*, 2008) et, en particulier, les motifs UGUU (Faustino & Cooper, 2005; Marquis *et al.*, 2006).

L'expression tissulaire des 6 protéines CELF est très variable (Ladd *et al.*, 2001). CELF-3 et -5 sont exclusivement exprimées dans le cerveau. En revanche, CUGBP1 est fortement exprimée dans le muscle dans lequel les autres CELF sont faiblement représentées et sont régulées au cours du développement (Ladd *et al.*, 2001)

La protéine CUGBP1 a été identifiée à l'origine pour sa capacité à se fixer au niveau de répétitions CUG d'ARN sous forme de structure simple brin (Timchenko *et al.*, 1996) mais elle n'interagit pas avec l'ARN CUG sous forme double-brin et ne colocalise pas avec les foci (Figure 28), que ce soit dans les cellules surexprimant des répétitions CUG, les modèles cellulaires DM1 issus de biopsie de patients, les tissus musculaires ou les modèles de souris reproduisant la maladie (Kanadia *et al.*, 2003a).

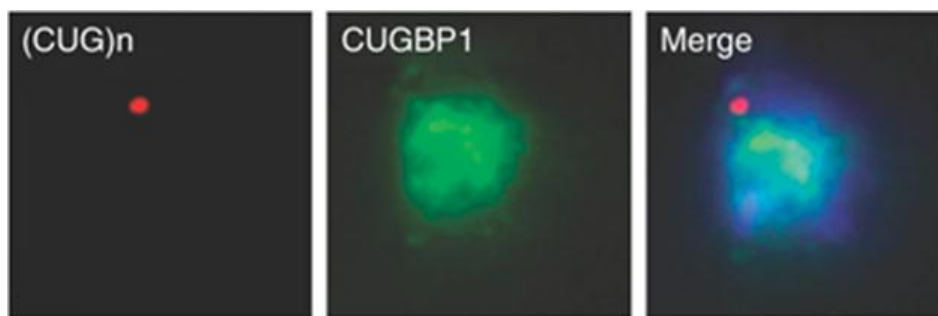


Figure 28 : Absence de colocalisation de CUGBP1 dans les tissus musculaires de patients DM1. La présence de foci CUG n'altère pas la localisation de la protéine CUGBP1 qui reste uniforme dans le nucléoplasme (Mankodi *et al.*, 2003).

Ceci suggère ainsi que la CUGBP1 n'est pas séquestrée dans la pathologie. Cependant en présence de l'ARNm DMPK muté, la protéine CUGBP1 se trouve stabilisée au sein d'un complexe ribonucléoprotéique. En effet, sa demi-vie est augmentée d'un facteur cinq dans des cellules COS surexprimant des expansions CUG (Timchenko *et al.*, 2001; Savkur *et al.*, 2001) ce qui a pour effet l'augmentation de sa concentration relative chez les patients DM1 par rapport à des personnes non-atteintes (Figure 29). Le mécanisme à l'origine de cette stabilisation, ainsi que le lien direct entre les expansions CUG et celle-ci demeure inconnu à l'heure actuelle. Le laboratoire de Tom Cooper a cependant montré que la CUGBP1 serait stabilisée par une hyperphosphorylation par la Protéine Kinase C (PKC) dont une activation a

été observée dans le tissu cardiaque de patients DM1 ainsi que dans des modèles de souris reproduisant la maladie. Cette hyperphosphorylation de CUGBP1 a été observée chez des patients DM1, dans des modèles cellulaires après induction de l'expression des répétitions CTG (Timchenko *et al.*, 2001; Savkur *et al.*, 2001; Mahadevan *et al.*, 2006) ainsi que dans plusieurs modèles murins exprimant les répétitions CTG (Mahadevan *et al.*, 2006; Wang *et al.*, 2007; Dansithong *et al.*, 2008; Orengo *et al.*, 2008; Yadava *et al.*, 2008).

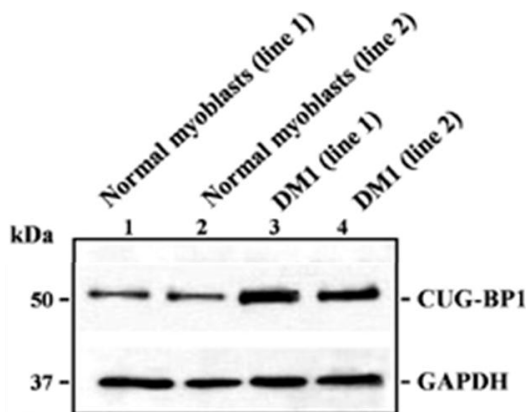


Figure 29 : Augmentation de la quantité relative de CUGBP1 chez les patients DM1. La quantité relative de CUGBP1 analysée par western-blot est en moyenne deux fois plus importante chez les patients DM1 que chez les contrôles (Paul *et al.*, 2006).

De manière logique, l'accumulation de CUGBP1 conduit à une augmentation de son activité dans différents modèles cellulaires et murins (Timchenko *et al.*, 2004; Ho *et al.*, 2005; Kalsotra *et al.*, 2008; Koshelev *et al.*, 2010). La protéine CUGBP1 intervient dans le maintien du profil d'épissage foetal de plusieurs gènes et son activité diminue à la naissance. Aussi, la stabilisation de la CUGBP1 dans des cellules adultes a pour effet d'augmenter sa concentration relative et entraîne une dérégulation de l'épissage de plusieurs gènes dont la troponine cardiaque (*TNNT2*), du récepteur de l'insuline (*INSR*), le canal chlore musculaire (*CLCN1*), la myotubularine *MTMR1*, le N-méthyl-D-aspartate (*NMDAR1*) et *Tau* (Philips *et al.*, 1998; Savkur *et al.*, 2001; Charlet-B *et al.*, 2002; Faustino & Cooper, 2005; Leroy *et al.*, 2006). L'équipe de Tom Cooper a récemment développé un modèle murin surexprimant la CUGBP1 de façon inductible uniquement dans le cœur. L'induction de l'expression de ce transgène est de l'ordre de 4 à 8 fois le niveau d'expression de la protéine endogène et aboutit à l'apparition de nombreux symptômes (Kalsotra *et al.*, 2008; Koshelev *et al.*, 2010; Ward *et al.*, 2010). Les souris présentent un cœur plus volumineux, avec présence de nécrose et perte de fibres musculaires, ainsi qu'une cardiomyopathie associée à des troubles de la conduction cardiaque comparables à ceux observés dans les modèles murins exprimant des répétitions CUG960 et chez les patients DM1. Ces souris présentent également de nombreuses

dérégulations de l'épissage alternatif, parmi lesquelles les ARN pré-messagers *Chromosome-10 open reading frame-97* (un gène suppresseur de tumeur), *Mfn2* (code la mitofusine II, une protéine membranaire mitochondriale), *Capzb* (impliqué dans les filaments d'actine) et *Ppfibp1* (qui code la liprine) qui sont des cibles non-régulées par MBNL1 et dérégulées dans la DM1, apportant une nouvelle preuve de l'implication de la CUGBP1 dans la pathologie. Par ailleurs, l'induction au-delà de deux semaines entraîne une létalité chez ces souris (Kalsotra *et al.*, 2008; Koshelev *et al.*, 2010).

e) Dérégulation de la hnRNP H

La protéine hnRNP H a également été mise en évidence comme étant associée à l'ARNm mutant de la DMPK (Kim *et al.*, 2005). L'équipe de Sita Reddy a observé que l'expression de hnRNP H est deux fois plus élevée dans des myoblastes de patients DM1 par rapport à des myoblastes contrôles (Paul *et al.*, 2006). De plus, hnRNP H semble également interagir de manière directe avec MBNL1, et la surexpression de MBNL1 est nécessaire pour observer une colocalisation de hnRNP H avec les foci CUG dans des myoblastes DM1 (Paul *et al.*, 2006). Enfin, la hnRNP H régule l'épissage alternatif du récepteur à l'insuline (INSR) et une interaction directe entre MBNL1 et hnRNP H semble contribuer à inhiber l'effet répresseur de MBNL 1 sur l'inclusion de l'exon alternatif 11 (Paul *et al.*, 2006).

Aucunes données ultérieures sur l'importance de ce facteur dans la pathologie n'ont été publiées, ce qui explique le manque de modèle mécanistique quant à l'implication de hnRNP H dans la DM1.

f) Dérégulation des micro ARN

Les microARNs (miRNAs) émergent comme des régulateurs clés du transcriptome cellulaire de par leur capacité à réguler finement l'expression génique. Une altération de certains microARNs a par ailleurs été répertoriée dans certaines maladies génétiques, par exemple le syndrome de DiGeorge, ainsi que certains cancers (Meola *et al.*, 2009). Les raisons de ces dérégulations peuvent avoir plusieurs origines : une mutation dans la séquence mature du miRNA, dans la séquence de la cible du miRNA ou de manière plus globale par la mutation ou la perturbation d'une protéine impliquée dans la voie de régulation des miRNAs (Meola *et al.*, 2009).

La Dystrophie Myotonique de type 1 étant caractérisée par une dérégulation du métabolisme de l'ARN, l'une des hypothèses actuelles est que certains microARNs, impliqués dans le développement musculaire ou neuronal, pourraient être dérégulés dans la pathologie et expliquer une partie des symptômes observés. La première étude fut publiée par le laboratoire de Botta, qui analysa l'expression de plusieurs microARNs dans des tissus musculaires provenant de biopsies de patients DM1 en comparaison avec des tissus musculaires sains (Gambardella *et al.*, 2010). D'après leurs résultats, seul miR-206, un microARN spécifiquement exprimé durant la myogenèse dans le tissu cardiaque et le muscle squelettique, est surexprimé environ 5 fois chez les patients DM1 par rapport au groupe contrôle (Gambardella *et al.*, 2010). Dans une autre étude, l'équipe de Martelli a montré que les microARNs miR-1 et miR-335 étaient surexprimés dans les muscles de patients DM1, et au contraire miR-29b, miR-29c et miR-33 sous-exprimés (Perbellini *et al.*, 2011). Ils ont également noté que la distribution cellulaire de miR-1, miR-133b et miR-206 était altérée (avec une localisation principalement perinucléaire chez les patients) et que l'expression de certaines cibles de ces miRNAs était modifiée (Perbellini *et al.*, 2011). Cependant, aucune démonstration n'a été faite concernant le mécanisme à l'origine des variations d'expression de ces miRNAs dans la DM1 et leur possible relation à un ou plusieurs symptômes.

Dernièrement, l'équipe de Nicolas Charlet-Berguerand a démontré que la maturation du microARN miR-1 était altérée dans les tissus cardiaques de patients DM1 (Figure 30) (Rau *et al.*, 2011).

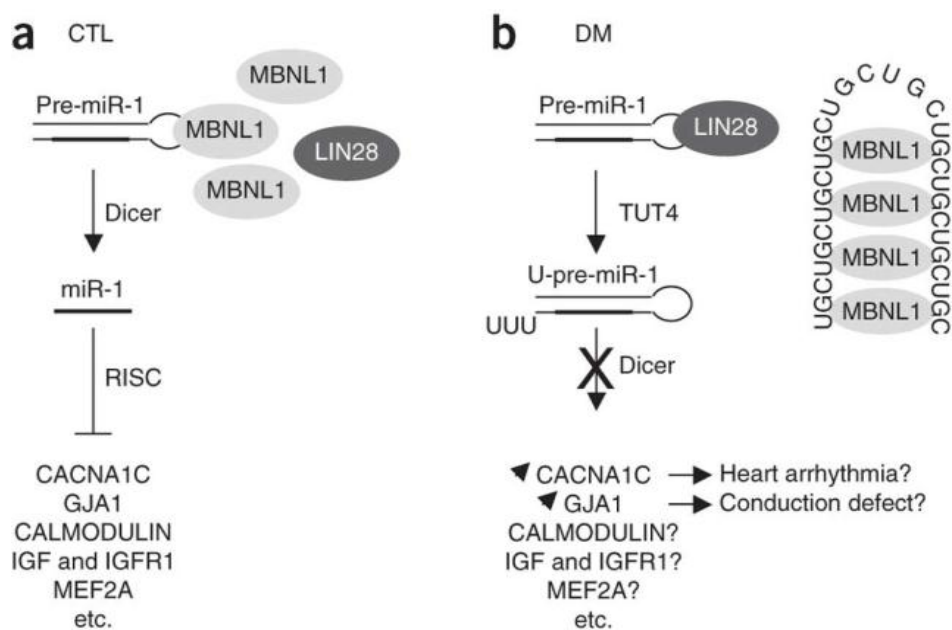


Figure 30 : Modèle de la dérégulation du processing du pre-miR-1 dans la DM1. **A.** MBNL1 participe au processing du pre-miR-1 dans un contexte sain. **B.** MBNL1 est séquestré par les expansions CUG et le processing de pre-miR-1 est altéré dans la DM1. (Rau *et al.*, 2011).

MBNL1, au-delà de son activité de facteur d'épissage, est un régulateur de la biogénèse du pre-miR-1 via sa fixation sur un motif UGC localisé dans la boucle. L'équipe a également montré qu'il existe une compétition d'interaction sur ce site entre MBNL1 et LIN28, une protéine favorisant l'uridylation par le facteur TUT4 de ce microRNA le rendant incapable d'être pris en charge par Dicer (Rau *et al.*, 2011). La réduction de miR-1 augmente le niveau d'expression de deux de ses cibles (GJA1 et CACNA1C) dans le tissu cardiaque de patients DM1 et DM2 (Rau *et al.*, 2011). L'équipe propose que la surexpression de ces deux protéines, qui codent un canal calcique et une jonction-communicante maintenant l'homéostasie cellulaire, pourrait contribuer aux altérations cardiaques observées dans la pathologie.

Cette étude est la première à montrer un lien possible entre la dérégulation d'un microARN et un symptôme clinique de la maladie. Ceci suggère que d'autres microARN pourraient également être dérégulés dans cette pathologie, ainsi que plus généralement dans d'autres maladies musculaires.

g) Les expansions CUG sont des cibles de Dicer

En 2007, l'équipe de Krzyzosiak a montré que les transcrits contenant de longues tige-boucle composées de répétitions CNG (où N représente n'importe quel nucléotide) sont des cibles de Dicer. Le découpage de cet ARN double-brin induit la production de petits fragments CUG issus de la digestion par Dicer qui pourraient agir comme siRNA et être utilisés par la voie de l'interférence à ARN pour éteindre l'expression génique de transcrits contenant des séquences complémentaires, contribuant ainsi aux altérations d'expression de plusieurs gènes dans la pathologie (Krol *et al.*, 2007). Cependant, aucune étude postérieure n'a montré une dérégulation de l'expression de gènes due à l'utilisation de ces « siRNA CUG » par la machinerie cellulaire.

h) Effet trans sur la traduction

La protéine CUGBP1 est comme nous l'avons vu précédemment une protéine multifonctionnelle se fixant à l'ARN et qui régule le métabolisme de l'ARN à différentes étapes parmi lesquelles la traduction coiffe-dépendante et coiffe-indépendante, la stabilité des ARN et l'épissage alternatif (Ladd *et al.*, 2001; Salisbury *et al.*, 2008). Bien qu'il fut précédemment montré que la forme cytoplasmique de la CUGBP1 pouvait activer la

traduction de p21 (Iakova *et al.*, 2004), C/EBP β (Karagiannides *et al.*, 2006) et Mef2A (Timchenko *et al.*, 2004) dans les cellules en différenciation par l'interaction avec le facteur d'initiation eIF2, le mécanisme à l'origine du défaut de traduction de ces facteurs dans les cellules musculaires DM1 n'est pas connu. En effet, le niveau de l'ARNm codant CUGBP1 n'est pas altéré chez les patients par rapport à des myoblastes sains (Nezu *et al.*, 2007). En 2008, l'équipe de Timchenko a montré que deux événements de phosphorylation de la CUGBP1 étaient dérégulés chez les patients DM1 durant la myogenèse (Salisbury *et al.*, 2008). Ces phosphorylations sont requises pour l'interaction de la CUGBP1 sur les ARNm de p21, C/EBP β et Mef2A et le recrutement du facteur d'initiation eIF2 pour la traduction effective de ces transcrits. De plus, l'absence d'expression dans les myoblastes DM1 de la protéine p21, impliquée dans le cycle cellulaire, est corrélée avec l'absence d'arrêt du cycle cellulaire, étape nécessaire pour la différenciation myogénique (Timchenko *et al.*, 2001).

D) Remise en cause du modèle actuel

Comme tout modèle proposé en biologie, le dogme établi qui propose que la séquestration de MBNL1 par les foci soit la cause principale de la pathologie peut être remis en question. L'une des premières controverses est venue du constat que la séquestration de MBNL1 dans des foci CAG n'entraîne pas de dérégulation de l'épissage alternatif (Ho *et al.*, 2005). Le laboratoire de Tom Cooper a ainsi démontré que la colocalisation de MBNL1 au niveau des foci peut être séparée de la dérégulation de l'épissage alternatif. De plus, l'équipe de Sita Reddy a montré que l'inactivation de MBNL1 dans des myoblastes DM1 réduit le nombre de foci de l'ordre de 70%. Si l'on peut facilement imaginer que la protéine MBNL1 elle-même puisse être impliquée dans la structuration des foci, qu'en est-il du devenir de l'ARN mutant DMPK ? En effet, le taux d'expression du gène DMPK mutant n'est pas modifié lorsque MBNL1 est inactivé, suggérant que la quantité de transcrits mutants est la même dans des cellules contrôles. Si la quantité d'ARN DMPK mutant est la même, comment expliquer que le nombre de foci diminue ? L'ARN mutant pourrait-il adopter diverses structures, dont certaines ne seraient pas observables par RNA-FISH ? Ceci pourrait ainsi relancer le débat sur l'apparente contradiction existant entre les données biochimiques indiquant que CUGBP1 puisse se fixer *in vitro* sur les répétitions (CUG) $_n$ (Timchenko *et al.*, 2001) et l'absence de colocalisation de cette protéine avec les foci. La CUGBP1 pourrait ainsi interagir avec l'ARN mutant sans pour autant se retrouver localisée dans les foci.

Ainsi, la localisation d'une protéine au niveau des expansions CUG ne signifie pas forcément que cette protéine puisse être déplétée. A l'inverse, une protéine qui n'est pas présente au niveau des foci CUG ne signifie pas qu'elle n'est pas dérégulée (comme la CUGBP1 ou plus récemment SHARP). L'équipe de Mani Mahadevan a montré que des souris transgéniques surexprimant à un haut niveau et de manière réversible des séquences CTG de taille normale (CTG5) présentent également des anomalies cardiaques et une myotonie ressemblant au phénotype DM1 (Mahadevan *et al.*, 2006). De plus, ce modèle murin ne présente pas de formation de foci suggérant la possibilité d'un effet toxique de l'ARN mutant indépendant de sa séquestration nucléaire. Dans ce cas, ce serait la surexpression à haut niveau de CTG sans expansions qui reproduirait le même effet que la séquence répétée.

Dans un article récent, R.P. Junghans propose un modèle (Figure 31) qui fait intervenir la présence, au-delà des foci insoluble, d'une seconde forme de l'ARN DMPK mutant soluble.

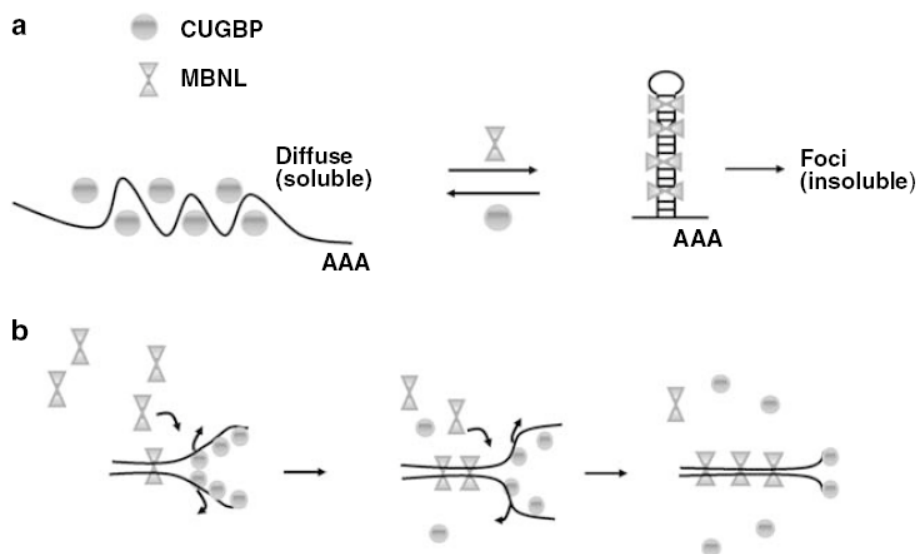


Figure 31 : Représentation schématique du modèle proposé par R.P. Junghans. A. Les protéines ayant une affinité pour les répétitions en conformation simple brin (ex. CUGBP1) se fixent et favorisent la forme simple brin de l'ARNm DMPK mutant alors que les protéines ayant une affinité pour les structures en double-brin (ex. MBNL1) se fixent et favorisent une forme double-brin. **B.** La structure en épingle à cheveux au niveau des répétitions CUG est instable et permet le passage d'une forme à l'autre. (Junghans, 2009).

Ce modèle envisage la fixation de diverses protéines, qui possèdent une affinité soit pour l'ARNm simple brin, soit pour l'ARN double-brin, et qui pourraient donc intervenir suivant l'état structurel de l'ARNm mutant (Junghans, 2009). La concentration relative entre ces différentes protéines définirait ainsi la structure majoritaire simple ou double-brin de l'ARNm mutant. La structure en épingle à cheveux formée par les répétitions CUG est

thermodynamiquement modulable et pourrait permettre la transition d'un état simple brin à double-brin selon un mécanisme qui pourrait être illustré selon le principe d'une « fermeture éclair ». La fixation des protéines parmi lesquelles MBNL1 favoriserait la structure double-brin au détriment d'une structure simple brin. L'agrégation de protéines sur cet ARN pourrait rendre le complexe protéine-ARN insoluble dans la cellule, ce qui pourrait être observable sous forme de foci dans les cellules. Au contraire, la fixation de la CUGBP1 ou des facteurs de transcription (Ebraldidze *et al.*, 2004), sur une structure simple brin de l'ARNm mutant pourrait conduire à un effet stabilisateur sur ces protéines. Il pourrait ainsi exister une fraction de l'ARN messager DMPK mutant à la fois simple brin et double-brin.

Ce nouveau modèle proposé par Junghans soulève également d'autres questions. L'une d'entre elles pourrait être le mécanisme à l'origine de la transition entre ARN CUG simple brin et ARN CUG double-brin. Comme nous l'avons précédemment abordé, les ARN ne sont pas nus dans la cellule mais associés à de nombreux complexes protéiques. Le modèle de « fermeture éclair » proposé par Junghans implique que les complexes protéiques sur les expansions CUG pourraient rentrer en compétition, et que leur encombrement stérique suffirait à modifier la structure de l'ARN (Junghans, 2009). Cependant, nous pourrions imaginer que ces transitions nécessiteraient une nouvelle classe de protéines dont le rôle serait de modifier les structures secondaires de l'ARN messager DMPK mutant pour assembler et/ou désassembler ces complexes ribonucléoprotéiques.

Les modifications structurelles de l'ARN ont fait l'objet de nombreuses études dans toutes les étapes du métabolisme de l'ARN (transcription, épissage, traduction...) et une classe de protéine, les DEAD-box ARN hélicases a été identifiée comme requise pour la modulation des structures secondaires des ARNs.

III. p68 et p72 : DEAD-box ARN hélicases

A) Généralités sur les hélicases

L'ARN, par exemple chez de nombreux virus, peut exister dans une forme double-brin, mais l'ARN simple brin peut également établir des régions double-brin par la formation d'interaction intra ou intermoléculaires. Un exemple parmi tant d'autre est la structure secondaire des ARN de transfert ou des ARN ribosomiques (Onoa & Tinoco, 2004). De plus, l'appariement de bases est important pour de nombreuses étapes du métabolisme de l'ARN tels que l'épissage (comme précédemment vu dans l'introduction) ou l'ARN interférence, prouvant que le désappariement autant que la formation de structure double-brin est un mécanisme clé de la cellule (Bleichert & Baserga, 2007).

Les ARN hélicases sont des enzymes capables de coupler l'hydrolyse de nucléotides triphosphates (NTP) à des dissociations d'acides nucléiques et sont requises pour la synthèse des ribosomes, la transcription, la maturation, l'épissage ainsi que l'export nucléaire des ARN (Jankowsky, 2011).

B) Les six superfamilles d'hélicases ADN ou ARN

Les hélicases et translocases constituent une famille de "moteurs moléculaires" ADN-ou ARN-dépendant qui consomme les nucleosides triphosphates (préférentiellement l'ATP) comme source d'énergie afin de voyager le long du brin. Les translocases se déplacent le long d'un substrat ADN ou ARN afin d'ammener l'acide nucléique vers une autre localisation ou pour le libérer de ces protéines. Les hélicases quant à elles se déplacent également le long des acides nucléiques mais ont pour rôle de séparer les deux brins appariés. Toutes les hélicases et translocases contiennent un ou deux domaines catalytiques fixant un NTP : les modules RecA et les modules AAA+.

Les hélicases à ADN ou ARN ont été classées sur la base de la présence dans leur séquence de motifs conservés. A l'origine, les hélicases ont été répertoriées comme étant des NTPases c'est-à-dire des enzymes capables d'hydrolyser des NTP et impliquées dans la réplication de l'ADN et la transcription (Hodgman, 1988). Quelques années plus tard, des analyses bioinformatiques d'alignements de séquences ont permis de révéler sept à neuf motifs conservés dont l'ensemble forme un domaine appelé « core ».

| Superfamily/Class | Protein Fold | Oligomeric State | Polarity | Function | Example Members |
|---------------------------|--------------------|--------------------------|--|--|---|
| Superfamily I (SF-I) | RecA (tandem pair) | Monomer (dimer/multimer) | 3'-5' (SF-IA), 5'-3' (SF-IB) | DNA unwinding, repair and degradation | Bacterial PciA, Rep, UvrD, RecBCD, Dda; eukaryotic Rrm3, Pif1, Dna2 |
| Superfamily II (SF-II) | RecA (tandem pair) | Monomer (dimer/multimer) | 3'-5' (SF-IIA), 5'-3' (SF-IIB), some dsDNA translocases | RNA melting, RNA-binding protein displacement; DNA or RNA unwinding; chromatin remodeling; DNA/RNA translocation; melting and migration of Holliday junctions or branched structures | DEXD/H-box proteins (eukaryotic eIF4A, Prp2, Ski2, Vasa, Dpbs, NS3 of hepatitis C); Snf2/SWI proteins (eukaryotic Snf2, ISWI, Rad54, archaeal Hei308); bacterial RecQ, RecG, UvrB |
| Superfamily III (SF-III) | AAA+ | Hexamer (dodecamer?) | 3'-5' | DNA unwinding/replication | Papilloma virus E1, simian virus 40 Large T-antigen, adeno-associated virus Rep40 |
| Superfamily IV (SF-IV) | RecA | Hexamer (other states?) | 5'-3' (dsDNA?) | DNA unwinding/replication; ssRNA packaging | Bacterial DnaB; phage T7 gp4, T4 gp41, SPP1 G40P; pRSF1010 RepA; phage ϕ -12 P4; mitochondrial protein Twinkle |
| Superfamily V (SF-V) | RecA | Hexamer | 5'-3' | RNA translocation, RNA/DNA heteroduplex unwinding; transcription termination | Bacterial Rho |
| Superfamily VI (SF-VI) | AAA+ (PS-II clade) | Hexamer (other states?) | 3'-5' (dsDNA?) | DNA unwinding/replication | Eukaryotic/archaeal MCMs |
| Superfamily VII? (SF-VII) | AAA+ (new clade?) | Hexamer | 5'-3' | Chromatin remodeling | Eukaryotic Tip48/49, Reptin/pontin |

Helicases

Tableau 6 : Liste des superfamilles d'hélicases. Ces superfamilles sont classées selon le domaine RecA ou AAA+ qui les définit, leur état oligomérique, leur polarité, leur fonction connue et une liste non-exhaustive d'hélicases appartenant à chaque superfamille (Berger, 2008).

Les hélicases peuvent être regroupées en 6 Super Familles (SFI, II, III, IV, V et VI), définie par leur structure primaire et leur activité cellulaire (Tableau 6) (Gorbalenya & Koonin, 1993; Singleton *et al.*, 2007).

Il existe pour l'ensemble de ces protéines une structure tridimensionnelle conservée et par conséquent une mécanistique moléculaire partagée. En effet, des motifs conservés se retrouvent à l'intérieur d'un module catalytique dont la topologie est conservée pour l'ensemble des enzymes des six familles.

L'ARN hélicase p68, identifiée au cours de ma thèse appartenant à la Super Famille II, j'aborderai uniquement cette superfamille au cours de mon introduction.

La superfamille II est la plus grande famille des ARN hélicases, regroupant des protéines impliquées dans divers processus cellulaires. Cette famille inclue les ARN hélicases à boîte DEXD/H (Cordin *et al.*, 2006) parmi lesquelles celles du type Ski2-like ou les protéines proches de l'hélicase RecQ (Bennett & Keck, 2004). Les membres de cette famille sont capables de se déplacer sur des ARN simple ou double-brin et dans le sens 5' → 3' et 3' → 5' (à l'inverse des ARN hélicases de la superfamille I qui ne peuvent se déplacer que dans un seul sens). Ces enzymes possèdent également des domaines ultra-conservés qui confèrent à ces protéines leur fonction.

C) Motifs conservés et fonctions associées

Les différents motifs conservés des hélicases sont bien caractérisés et leur fonction ainsi que leurs interactions intramoléculaires sont de mieux en mieux comprises. A l'inverse des motifs conservés, les séquences qui séparent ces différents motifs sont divergentes mais leurs longueurs sont relativement bien conservées (Gorbalenya & Koonin, 1993). Dans l'ensemble, ces motifs incluent des résidus conservés impliqués dans la liaison et l'hydrolyse de NTP, équivalents aux boîtes Walker A et Walker B retrouvées dans beaucoup d'ATPases (Smith & Rayment, 1996; Leipe *et al.*, 2003) et ne suffisent donc pas à définir une ARN hélicase. Comme nous le verrons dans cette partie, ces motifs ne sont pas indépendants. Il existe en effet de fortes coopérations entre-eux par des liaisons intramoléculaires qui peuvent évoluer lors de l'interaction avec l'ATP ou l'ARN. Ces motifs sont requis pour la liaison et l'hydrolyse de l'ATP (Q, I et II), la liaison au substrat ARN (IV et V) et le couplage de l'hydrolyse à l'activité hélicase (III, VI). L'ensemble de ces activités et les changements conformationnels lors de ce processus enzymatique aboutissent à une activité hélicase permettant le remodelage des complexes ribonucléoprotéiques (Figure 32).

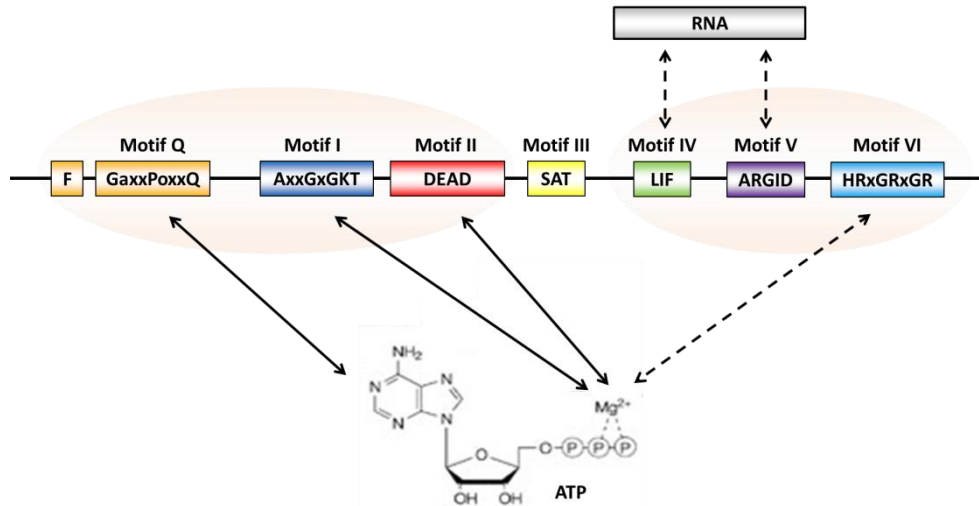


Figure 32 : Représentation schématique des différents motifs conservés des DEAD-box ARN hélicases et de leur interaction avec l'ATP et leur substrat ARN. Les lettres correspondent à l'acide aminé d'après le code génétique, « o » correspond à une sérine (S) ou une thréonine (T), « a » est un acide aminé aromatique et « x » n'importe quel acide aminé. Adapté de (Rocak & Linder, 2004).

1) Motif Q (G-a-x-x-P-o-x-x-Q)

Le motif Q est le motif dont la découverte est la plus récente. Le laboratoire de Patrick Linder est à l'origine de l'identification de ce motif par la caractérisation d'une glutamine (Q) caractérisée comme essentielle *in vivo*, dans cinq protéines du groupe des ARN hélicases de type DEAD : Dbp1, Dbp2, Fal1, Prp5 et eIF4 (Tanner *et al.*, 2003). Cette glutamine fait partie d'une séquence conservée de 9 acides aminés à laquelle s'ajoute une phénylalanine (F) localisée 17 acides aminés en amont (Tanner *et al.*, 2003). Par la suite, ce motif a été retrouvé dans l'ensemble des protéines appartenant aux ARN hélicases de type DEAD, ce qui en fait un motif caractéristique. De plus, la conservation de la glutamine s'étend à d'autres hélicases parmi lesquelles Rad3 (famille des ChII), Mtr4/Dob1 (famille des Ski2), Dna2 (famille des Upf1) (Tanner *et al.*, 2003).

Diverses analyses comprenant des alignements de séquences, des expériences de mutagenèse et des mesures de l'activité ATPase montrent que ce motif comprenant la glutamine ainsi que la phénylalanine en amont sont hautement conservés et essentiels pour la viabilité chez *Saccharomyces cerevisiae*. De plus, ces résidus interviennent dans la reconnaissance de l'adénine de l'ATP et dans son hydrolyse pour le facteur d'initiation eIF4A chez la levure (Tanner *et al.*, 2003), l'ATPase NS3 du virus Powassan (Gallivan & McGarvey, 2003) et une ARN hélicase chez *P. falciparum* homologue à Dbp5 et DDX19, chez la levure et l'homme respectivement (Mehta & Tuteja, 2011). Des études

supplémentaires, en particulier sur l'ARN hélicase Ded1 ont montré que le motif Q régulait également l'affinité de la protéine pour ses substrats ARN et ainsi son rôle dans l'activité hélicase (Tanner *et al.*, 2003; Cordin *et al.*, 2004) et pourrait être requis dans la stabilisation d'une conformation catalytiquement compétente de plusieurs ARN hélicases (Strohmeier *et al.*, 2011).

La découverte de ce motif dans les ARN hélicases de type DEAD a depuis été élargie au-delà des ARN hélicases et translocases. En effet, des alignements de séquences entre divers domaines d'ARN hélicases de la superfamille SF2 et des translocases impliquées dans l'empaquetage de l'ADN viral montrent également une glutamine conservée. La mutation de ce résidu aboutit à une forte réduction de la rapidité de compaction de l'ADN de manière ATP-dépendante, confirmant l'importance de motif analogue au motif Q pour la fixation et l'hydrolyse de l'ATP (Tsay *et al.*, 2009).

2) Motif I (A/G-X-X-G-X-G-K-T)

Ce motif, également connu comme le motif Walker A, est composé d'une séquence consensus AxTGxGKT caractéristique de la famille SF2 et se retrouve dans certaines ATPases, GTPases et hélicases à ADN (Gorbalenya *et al.*, 1989). Des mutations au niveau de ce motif, notamment de la première alanine (A), la lysine (K) ou la dernière thréonine (T), abolissent les fonctions ATPase et hélicase par la réduction de l'affinité pour l'ATP et de l'efficacité de son hydrolyse (Pugh *et al.*, 1999; Caruthers & McKay, 2002; Tanner *et al.*, 2003; Cordin *et al.*, 2004). Des analyses biochimiques et structurales ont montré que la lysine (K) du motif GKT forme une liaison hydrogène avec le phosphate du complexe Mg-ATP ou Mg-ADP et que la thréonine (T) interagit directement avec l'ion magnésium Mg²⁺ (Caruthers & McKay, 2002).

3) Motif II (DEAD)

Ce motif correspond au domaine Walker B, présent dans toutes les ATPases et caractérisé en 1982 par Walker (Walker *et al.*, 1982). Le premier acide aspartique (D) et le glutamate (E) présents dans ce motif sont ultra-conservés et permettent de distinguer les différentes familles d'ARN hélicases comme les DEAD ou les DEAH lorsque ceux-ci sont associés à d'autres acides aminés. C'est par ailleurs ce motif II qui est à l'origine du nom de ces hélicases. Le premier acide aspartique (D) et le glutamate (E) interagissent avec l'ion magnésium complexé avec l'ATP, ce qui pourrait entraîner un changement conformationnel

du core de l'hélicase et le rapprochement du deuxième acide aspartique (D) du motif II avec l'histidine (H) conservée du motif VI (Dillingham *et al.*, 1999; Story *et al.*, 2001; Shi *et al.*, 2004; Cordin *et al.*, 2006). Ce rapprochement permettrait ainsi la formation d'une poche, qui grâce aux interactions intramoléculaires entre les motifs I et II, les motifs II et III et les motifs II et VI établiraient un site actif enzymatique requis pour l'hydrolyse de l'ATP (Shi *et al.*, 2004; Cordin *et al.*, 2006). Des mutations ponctuelles des résidus de ce motif, notamment du premier acide aspartique (D) abolissent l'activité ATPase et hélicase des protéines DEAD-box ARN hélicases p68 et p72 (Ogilvie *et al.*, 2003).

4) Motif III (SAT)

La fonction du domaine III n'est pas très bien définie, en effet plusieurs travaux contradictoires ont été publiés. Les premières études sur ce motif, sur la base de mutations ponctuelles dans la protéine eIF4 montrèrent que des mutations dans ce motif découplent l'hydrolyse de l'ATP et l'activité hélicase (Pause & Sonenberg, 1992; Graves-Woodward *et al.*, 1997; Schwer & Meszaros, 2000). Cependant, des mutations semblables dans la protéine Has1 chez la levure aboutissent à une diminution de l'affinité pour l'ARN et l'activité ATPase, conduisant à une réduction de l'activité hélicase (Rocak *et al.*, 2005). Enfin, une étude récente effectuée sur la protéine Ded1 chez la levure montre que des mutations de chacun des acides aminés du motif SAT affectent l'hydrolyse de l'ATP et l'affinité de la protéine pour les ARN simple brin mais pas l'affinité pour l'ATP (Banroques *et al.*, 2010). L'inactivation de l'activité hélicase ne serait donc qu'une conséquence indirecte de l'abolition de la fixation à l'ARN (Banroques *et al.*, 2010).

5) Motif IV

Etant donné la faible conservation de ce motif, peu d'études ont été menées sur ce dernier. Il est composé de neuf résidus avec une séquence consensus V/L/I-V/I-F et des résidus chargés de part et d'autre (Gorbalenya & Koonin, 1993; Banroques *et al.*, 2008). Ce motif est requis pour la liaison à l'ARN pour la protéine eIF4A, du fait de la forte affinité des résidus arginines (R) dans le motif IV (VIFCNTRR) de cette protéine pour l'ARN simple brin (Caruthers & McKay, 2002). Récemment, l'équipe de Josette Banroques et Kyle Tanner a démontré l'importance de la phénylalanine (F), résidu conservé à près de 100% chez les ARN hélicases de la superfamille II (Banroques *et al.*, 2008). La mutation de cette phénylalanine dans la protéine Ded1 chez la levure, aboutit à un phénotype thermosensible avec retard de

croissance, suggérant l'implication de ce motif dans le maintien et la stabilité de la conformation structurale des ARN hélicases. De plus, la mutation de ce résidu résulte entre autre en une réduction importante de la capacité à lier l'ARN, que ce soit en absence ou présence d'ATP (Banroques *et al.*, 2008), suggérant qu'un changement conformationnel, dû à la fixation et/ou l'hydrolyse de l'ATP est nécessaire pour une interaction optimale avec son substrat. Ce type de mécanisme, dans une forme différente dans laquelle la protéine Dbp5 complexée à l'ADP adopte une structure favorable à la fixation de l'ARN, a également été décrit par l'équipe de Susan Wentz (Tran *et al.*, 2007).

6) Motif V

Le motif V, dont la séquence consensus est TDVuARGID (où u correspond à une Alanine ou Glycine) joue un rôle dans la liaison à l'ARN de façon conjointe avec les motifs auxiliaires Ia, Ib et IV (Cordin *et al.*, 2006). Cependant, un rôle dans le couplage de l'activité ATPase et l'activité hélicase ne peut être exclu (Caruthers *et al.*, 2000). En effet, l'équipe de Liu a montré qu'une mutation du domaine RGLD en LGLD abolit l'activité ATPase ARN-dépendant et l'activité unwinding de la DEAD-Box ARN hélicase p68 (Lin *et al.*, 2005). Ce motif est d'autant plus important que des mutations de l'arginine (R) et de l'acide aspartique (D) dans la protéine Prp28p ont un effet sévère sur la survie cellulaire (Chang *et al.*, 1997). Enfin, le motif V de la protéine Prp22p chez la levure est requis pour la stimulation de l'activité ATPase suite à l'association avec l'ARN (Schneider *et al.*, 2004).

7) Motif VI (X-R-X-G-R-X-G-R)

Ce motif, caractérisé par la séquence consensus « HRxGRxGR » est localisé à l'interface des deux domaines structurellement fonctionnels composés d'un côté par les motifs (Q, I, Ia, Ib et II) et de l'autre côté par les motifs (IV, V et VI) chez les hélicases DEAD. Ce motif est requis pour l'activité ATPase et par conséquent joue un rôle dans liaison à l'ARN (Pause *et al.*, 1993; Rogers *et al.*, 2002; Lin *et al.*, 2005). D'après le modèle proposé, la deuxième arginine (R), serait essentielle pour la liaison à l'ATP et son hydrolyse en agissant comme un « doigt arginine » qui stabiliserait la liaison Mg^{2+}/α -phosphate pour l'hydrolyse de l'ATP. De plus, il a été montré que ce motif pouvait avoir un rôle dans le couplage de l'hydrolyse de l'ATP à l'activité hélicase. En effet, des mutations de ce domaine dans la protéine eIF4A réduisent l'activité ATPase ainsi que l'activité hélicase (Rogers *et al.*, 2002; Lin *et al.*, 2005).

D) Les ARN hélicases p68/DDX5 et p72/DDX17

1) Généralités

Il y a trente ans, il fut montré qu'un anticorps dirigé contre l'antigène T du Simian Virus SV40 cross-réagissait avec une protéine nucléaire de 68kDa (Lane & Hoeffler, 1980). Cet anticorps fut par la suite utilisé pour « screener » une banque de cDNA qui identifia le gène *DDX5* codant cette protéine (Ford *et al.*, 1988).

Des alignements de séquences révélèrent une homologie avec le facteur d'initiation de la transcription eucaryote eIF4AIII, la première protéine identifiée pour sa capacité à dérouler l'ARN de manière ATP-dépendante (Ford *et al.*, 1988) ce qui facilita l'identification de *DDX5* comme ATPase ARN-dépendante possédant une activité hélicase (Hirling *et al.*, 1989; Iggo & Lane, 1989). Par la suite, le gène *DDX17* fut cloné et il fut démontré qu'il code deux protéines différentes : les ARN hélicases p72 et p82 (Uhlmann-Schiffler *et al.*, 2002). L'isoforme p82 est due à un site alternatif d'initiation de la traduction sur un codon non-AUG en amont du codon AUG qui définit le début de la traduction de l'isoforme p72 (Lamm *et al.*, 1996; Uhlmann-Schiffler *et al.*, 2002; Ogilvie *et al.*, 2003). Quant à leurs fonctions, des études biochimiques suggèrent que p68, p72 et p82 ont des propriétés très proches, notamment au niveau de l'activité hélicase (Uhlmann-Schiffler *et al.*, 2002) sans pour autant exclure de possibles fonctions physiologiques spécifiques (Chen *et al.*, 2008). Enfin, p72/p82 et p68 peuvent former des homodimères ou hétérodimères, suggérant qu'elles appartiennent à des voies moléculaires identiques (Ogilvie *et al.*, 2003).

Des alignements des séquences (Figure 33) montrent que 439 acides aminés sur 614, soit 71.5%, sont identiques entre les protéines p68 et p72/82. Cette homologie est encore plus grande lorsque l'on compare uniquement le « core » de ces hélicases avec près de 90% identité. Cependant, les parties N-terminale (acides aminés 1-95 de p68) et C-terminale (acides aminés 437-614 de p68) ne sont identiques qu'à 71% et 54% respectivement aux domaines de p72/82, suggérant que les protéines p68 et p72/82 pourrait avoir des fonctions propres, notamment par l'interaction de protéines spécifiques au niveau de leurs extrémités.


```

      10           20           30           40           50
DDX5  MSGYSSDRDRGRDRG-FGAPRFGGSRAGPLSGKKFGNPGKLVKKKWNLDLDPKFEKNF
      :  . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DDX17 MRGGGFGDRDRDRDRGGFGA-RGGGG----LPPKKFGNPGERLRKKKWDLSELPKFEKNF
      10           20           30           40           50

      60           70           80           90           100          110
DDX5  YQEHPLARRTAQEVETYSRSKEITVRGHN-CPKPVLNFYEAHFANVMDVIARQNFTPE
      :  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DDX17 YVEHPEVARLTPYEVDLRRKKEITVRGGDVCPKPVFAFHANFPQYVMDVLMDDQHFTEP
      60           70           80           90           100          110

Motif Q
      130           140           150           160           170
DDX5  TAIQAQGWVVALSGLDMVGVAQTGSGKTLSYLLPAIVHINHQPFLERGDGPICLVLAPTR
      :  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DDX17 TPIQCQGFPLALSGRDMVGIAQTGSGKTLAYLLPAIVHINHQPYLERGDGPICLVLAPTR
      120           130           140           150           160           170

      180           190           200           210           220
DDX5  ELAQVQVVAEYCRACRLKSTCIYGGAPKGPQIRDLERGVEICIAATPGRLIDFLECGKT
      :  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DDX17 ELAQVQVVAADDYGKCSRLKSTCIYGGAPKGPQIRDLERGVEICIAATPGRLIDFLESGBT
      180           190           200           210           220           230

Motif I

      240           250           260           270           280           290
DDX5  NLRRTTYLVIDEADRMLDMGFEPQIRKIVDQIRPDRQTLMSATWPKEVRQLAEDFLKDY
      :  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DDX17 NLRRTTYLVIDEADRMLDMGFEPQIRKIVDQIRPDRQTLMSATWPKEVRQLAEDFLRDY
      240           250           260           270           280           290

Motif II           Motif III

      300           310           320           330           340           350
DDX5  IHINIGALELSANHNILQIVDVCHDVEKDEKLIRLMEEIMSEKENKTIIVFVETKRRCDL
      :  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DDX17 TQINVGNLELSANHNILQIVDVCMSEKDHKLIQLMEEIMAEKENKTIIVFVETKRRCDL
      300           310           320           330           340           350

Motif IV

      360           370           380           390           400           410
DDX5  TRKMRDGGWPAMGIHGDKSQQERDWWLNEFKHGKAPILIAATDVASRGLLVEDVKFVINYD
      :  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DDX17 TRRMRDGGWPAMCIHGDKSQPERDWWLNEFRSGKAPILIAATDVASRGLLVEDVKFVINYD
      360           370           380           390           400           410

Motif V

      420           430           440           450           460           470
DDX5  YPNSSEDIYHRIGRTARSTKTGTAYTFFTPNNIKQVSDLISVLREANQAINPKLLQLVED
      :  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DDX17 YPNSSEDIYHRIGRTARSTNKGTAAYTFFTPGNLQARELIKVLEEANQAINPKMLQLVDH
      420           430           440           450           460           470

      480           490           500           510
DDX5  RGS-----RSRGRGG-----MKDDRRDRYSAG-KRGGF---NTFRDRENYDR-----
      :  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DDX17 RGGGGGGGGRSRYRTTSSANNPNLMYQDECRRRLRGVKDGGRRDSASYRDRSETDRAGYA
      480           490           500           510           520           530

      520           530           540           550           560
DDX5  ---GYSSLLKRDFGAKT-----QNGVYSAANYTNGSFGSNFVSAGIQTSFRTGNPTGTYQ
      :  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DDX17 NGSYGSP-NSAFGAQAGQYTYGQGTGAAAYGTSSYTAQEYAG---TYGASSTTSTGR
      540           550           560           570           580           590

      570           580           590           600           610
DDX5  NGYDSTQQY---GSNPNMHNMGMNQQAYAYPATAAAPMIGYPMPTGYSQ
      :  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DDX17 SSQSSSQQFSGIGRSGQQPQLMSQQ-FAQPP-GATNMIGYMGQTAYQYPPPPPPPPPSR
      600           610           620           630           640

```

Figure 33 : Alignement des séquences protéiques DDX5 et DDX17 humaines et caractérisation des motifs conservés.

D'une manière générale, les ARN hélicases p68 et p72 sont décrites comme des protéines essentiellement nucléaires (Lane & Hoeffler, 1980; Iggo & Lane, 1989; Lamm *et al.*, 1996) bien que certaines analyses révèlent qu'une quantité significative de protéine est également présente dans le cytoplasme dans plusieurs types cellulaires (Kahlina *et al.*, 2004; Yang *et al.*, 2006). Récemment, l'équipe de Liu a montré que p68 peut transiter entre le noyau et le cytoplasme, par l'intermédiaire de deux signaux d'export nucléaire et deux signaux de localisation nucléaire, cependant le mécanisme régulant ce changement de localisation n'a pas été identifié (Wang *et al.*, 2009). De plus, la distribution intranucléaire de p68 semble être régulée au cours du cycle cellulaire ; elle serait en effet localisée dans le nucléole durant la télophase mais absente durant l'interphase (Iggo *et al.*, 1991; Nicol *et al.*, 2000). De manière intéressante, p72 est identifiée parmi les protéines nucléolaires des cellules HeLa par spectrométrie de masse (Andersen *et al.*, 2005).

L'expression des ARNm codant DDX5 et DDX17 est ubiquitaire, mais leur expression peut varier significativement entre différents organes. Suivant les publications, elle est majoritaire dans le cœur et le muscle squelettique (Janknecht, 2010) ou le poumon et le placenta (Rössler *et al.*, 2000).

Une analyse Oncomine publiée en 2007 récapitule les voies de signalisation dans lesquelles p68 et p72 sont impliquées (Wilson & Giguère, 2007). La base de données Oncomine est une collection contenant des données provenant d'analyses de puces (micro-arrays ou multi-arrays) effectuées majoritairement dans des cellules cancéreuses. La mise en commun de ces bases de données permet l'établissement d'une liste de gènes co-exprimés avec p68 et p72 qui pourraient ainsi appartenir aux mêmes voies de signalisation et/ou à un même complexe ribonucléoprotéique. Les résultats de cette étude confirment des co-expressions avec des gènes impliqués dans l'épissage et la transcription, comme précédemment décrits dans la littérature, mais également de nouvelles voies de signalisations telles que le remodelage du cytosquelette ou le repliement protéique (chaperoning). De plus, *MBNL1* est le deuxième gène le plus souvent co-exprimé avec *DDX5* et le treizième avec *DDX17* (Wilson & Giguère, 2007), suggérant que p68, p72 et *MBNL1* pourraient appartenir à une voie de signalisation commune. Nous verrons par la suite que mes travaux de thèse confirment cette hypothèse.

2) Les différents rôles de p68 et p72

a) Activité ATPase et hélicase de p68 et p72

Les ARN hélicases p68 et p72 se lient à l'ARN simple brin et double-brin, bien que l'affinité pour ce dernier semble bien plus majoritaire. De plus, la fixation à l'ARN stimule leur activité ATPase, ce qui leur confère l'énergie nécessaire pour dérouler les duplex ARN dans le sens 5'→3' et 3'→5' (Huang & Liu, 2002; Lee, 2002). p68 et p72 possèdent également une activité « annealing » qui, en coopération avec l'activité hélicase, catalyse les réarrangement des structures secondaires de l'ARN (Rössler *et al.*, 2001). Cette activité confèrent aux DEAD-box la possibilité d'assembler et désassembler les complexes ribonucléoprotéiques sur l'ARN, également déterminée sous le terme d'activité « chaperonne » (Rössler *et al.*, 2001; Lund & Guthrie, 2005; Uhlmann-Schiffler *et al.*, 2006; Linder, 2006; Jankowsky & Bowers, 2006; Tran *et al.*, 2007; Jarmoskaite & Russell, 2011). Cette fonction modulatrice est dépendante de la présence de l'ARN, vu que l'activité ATPase de la plupart des protéines DEAD-box est dépendante ou largement stimulée par la présence d'ARN. Cependant, certaines DEAD-box, dont la protéine p68 humaine ou DbpA chez *E. coli*, possèdent un site pour l'ARN simple brin ATP-dépendant et un site pour l'ARN double brin ATP-indépendant (Böddeker *et al.*, 1997; Yang *et al.*, 2004). Afin de restreindre cette activité au niveau de certaines structures ou étapes du métabolisme de l'ARN, les ARN hélicases peuvent s'associer uniquement de façon transitoire avec les complexes ribonucléoprotéiques, ou appartenir à des complexes qui eux sont requis de façon transitoire, comme par exemple les particules pré-ribosomales intermédiaires dans lesquelles p68 et p72 ont été identifiées par spectrométrie de masse (Saveanu *et al.*, 2003; Jalal *et al.*, 2007). Dans ce cas, un changement conformationnel induit par l'association ou la dissociation d'une autre sous-unité à ce complexe pourrait changer la position de l'ARN et activer l'activité ATPase de la protéine DEAD-box, qui à son tour induirait un changement conformationnel de tout le complexe RNP. Ce changement modifierait ainsi la structure secondaire de l'ARN d'une manière telle qu'il ne serait plus un substrat pour l'ARN hélicase, ce qui décrocherait le complexe protéique de l'ARN permettant à un autre complexe d'interagir avec l'ARN cible. Ce mécanisme permettrait de limiter l'activité des protéines DEAD-box en régulant de manière unidirectionnelle l'enchaînement des complexes ribonucléoprotéiques sur l'ARN (Linder, 2006; Jarmoskaite & Russell, 2011).

b) Rôle dans la transcription

La transcription peut être divisée en trois volets : l'initiation, l'élongation et la terminaison. Peu d'évidences existent quant à l'implication de p68 et p72 dans ces deux derniers volets, cependant ces deux ARN hélicases ont été maintes fois impliquées dans l'initiation de la transcription (Nicol & Fuller-Pace, 2010). En effet, p68 et p72 interagissent avec l'ARN polymérase II et des facteurs de transcription ubiquitaires tels que CBP, p300 et PCAF (Rossow & Janknecht, 2003; Shin & Janknecht, 2007), qui sont des acétyltransférases dont un des rôles connus est l'acétylation des lysines des histones et de nombreux facteurs de transcription (Janknecht, 2002). p68 et p72 interagissent également avec les histones déacétylases HDAC1, HDAC2 et HDAC3 (Wilson *et al.*, 2004; Mooney *et al.*, 2010) qui ont une action antagoniste à CBP, p300 et PACF. Il existe ainsi très probablement une balance entre les acétyltransférases et les histones déacétylases, qui déterminent au final si p68 et p72 activent ou répriment la transcription.

La première indication que p68 pourrait intervenir dans la transcription est venue d'une étude montrant que p68 interagit et co-active spécifiquement la transcription du gène codant le gène *oestrogen receptor α* (Era) (Endoh *et al.*, 1999; Watanabe *et al.*, 2001). L'activité hélicase de p68 ne semble pas requise pour la co-activation de Era, suggérant un rôle distinct de celui découvert dans le processing des ARN pour lequel l'activité unwinding est requise. Une étude ultérieure démontra que p72 interagit et co-active également Era ; cette co-activation est décuplée par le facteur *steroid receptor co-activator 1* (SRC1) et est dépendante de l'interaction au co-activateur steroid receptor RNA activator (SRA) suggérant que p68 et p72 pourraient stimuler l'activité de Era en agissant comme lien direct entre Era et SRA/SRC1 (Watanabe *et al.*, 2001).

Enfin, dans un article très récent, il a été montré que le complexe p68/SRA stabilise l'interaction de la cohesine avec CTCF en agissant comme intermédiaire protéique entre ces deux facteurs (Yao *et al.*, 2010). De plus, p68/SRA semblent être requis pour la fonction insulatrice de CTCF, notamment dans la région IGF2/H19 (Yao *et al.*, 2010). De manière intéressante, des sites CTCF sont également retrouvés de part et d'autre des expansions de triplets CTG dans le gène DMPK (López Castel *et al.*, 2011b). Cependant le rôle de p68 ou p72 dans l'insulation des expansions CUG dans la Dystrophie Myotonique de type 1 n'a pas encore été analysé.

c) Rôle dans l'épissage

L'épissage alternatif implique l'association et la dissociation de l'ARN pré-messager avec les snRNA, qui pourraient être facilitées par les ARN hélicases. En effet, comme nous l'avons vu précédemment dans la première partie de l'introduction, les ARN hélicases aident non seulement au recyclage des snRNPs (Wahl *et al.*, 2009) mais de plus la protéine hélicase p68 est retrouvée par expériences de protéomiques dans certains complexes du spliceosome (Neubauer *et al.*, 1998). Il fut premièrement démontré que p68 et p72 interagissaient spécifiquement avec la snRNP U1 qui reconnaît le site 5' d'épissage (Liu *et al.*, 1998; Lee, 2002; Liu, 2002). De plus, des versions mutantes de p68, dépourvues d'activité ATPase ou hélicase, inhibent la dissociation de la snRNP U1 du site 5' d'épissage (Lin *et al.*, 2005). De la même façon, l'inactivation de p68 entraîne l'accumulation d'ARN non-épissé (Lin *et al.*, 2005). Les protéines p68 et p72 ont également été impliquées dans l'épissage alternatif d'au moins 3 transcrits : CD44, H-ras et tau (Hönig *et al.*, 2002; Guil *et al.*, 2003; Camats *et al.*, 2008; Kar *et al.*, 2011). Le mécanisme d'activation n'est pas totalement élucidé, cependant ces trois transcrits ont en commun la présence d'une structure secondaire de l'ARN en tige-boucle au niveau de séquences cis régulatrices. L'équipe de Jane Wu est par ailleurs la seule à montrer par des analyses de « footprint » de digestion à la RNase H et de protection par la snRNP U1 que p68 induit un changement conformationnel de la structure en tige-boucle, augmentant ainsi l'accessibilité de la snRNP U1 sur le site 5' de l'exon alternatif 10 de l'ARN pré-messager tau (Kar *et al.*, 2011). Ce changement conformationnel est effectué par l'activité ATPase et l'interaction à l'ARN, et donc par conséquent nécessite l'activité hélicase. En effet, des mutations ponctuelles au niveau des domaines conservés du core des ARN hélicases p68 et p72 abolissent la régulation de l'épissage (Hönig *et al.*, 2002; Kar *et al.*, 2011).

En conclusion, ces résultats suggèrent que les ARN hélicases p68 et p72 sont des facteurs cruciaux pour la régulation de l'épissage alternatif de plusieurs transcrits. Ce nombre de transcrits reste encore restreint d'après la littérature, mais des données non publiées provenant de l'analyse de puces Affymetrix Exon Array développées par le laboratoire de Didier Auboeuf (EASANA - <http://www.fast-db.com/cgi-bin/easana/index.pl>) montrent que l'inactivation conjointe de p68 et p72 dans des cellules myogénique murines C2C12 dérégule l'épissage alternatif d'un très grand nombre de transcrits qui restent à être validés expérimentalement (Données non publiées). Enfin, mes travaux de thèse apportent la preuve que p68 et p72 régulent l'épissage alternatif du gène *TNNT2* codant la troponine T cardiaque, qui est dérégulé dans la Dystrophie Myotonique de type 1.

d) Rôle de p68/72 dans les miRNA

Une autre fonction des ARN hélicases p68 et p72 concerne le processing des microARN (miRNA). Le miRNA primaire (ou pri-miRNA) est processé dans le noyau cellulaire par le complexe Drosha en un pré-miRNA, un transcrite intermédiaire comportant une structure en tige-boucle. Des analyses par spectrométrie de masse ont révélé que p68 et p72 font parties du complexe Drosha (Gregory *et al.*, 2004; Davis *et al.*, 2010). De plus, l'inactivation des gènes DDX5 ou DDX17 résulte en des altérations du processing de certains pri-miRNA ainsi que le précurseur de l'ARN ribosomique 5.8S (Fukuda *et al.*, 2007). Le fait que l'activité ATPase soit requise pour le rôle de p72 sur le processing des miRNA suggère que p68 et p72 dérouleraient l'ARN pour le rendre accessible aux coupures enzymatiques dans le complexe Drosha (Fukuda *et al.*, 2007). Cette hypothèse est en partie confirmée par des études *in vitro* qui montrent que p68 peut débobiner le duplex du miRNA let-7 (Salzman *et al.*, 2007). Plusieurs microARN sont dérégulés dans la Dystrophie Myotonique de type 1, notamment mir-1 (Rau *et al.*, 2011). Le laboratoire de Nicolas Charlet-Berguerand a également montré que ce microARN contient des sites de fixation de MBNL1, et que la séquestration de MBNL1 dans la pathologie est à l'origine d'une altération du processing de ce microARN et de défauts cardiaques (Rau *et al.*, 2011). Mes travaux de thèse montrent que p68, de par son activité hélicase, favorise la fixation de la protéine MBNL1 sur les expansions CUG ainsi que sur une structure en tige-boucle localisée sur l'ARN pré-messager *TNNT2*. Ainsi, p68 pourrait-il être un facteur important pour le processing de mir-1 ? p68 pourrait en effet moduler la structure secondaire de ce microARN afin de faciliter la fixation de MBNL1. Cette hypothèse sera abordée plus en détail dans la partie Discussion de cette thèse.

D'autre part, l'équipe de Fuller-Pace a publié cette année un article concernant la découverte d'un microARN dans un intron conservé du gène *DDX5* (Moore *et al.*, 2011). Ce microARN (appelé *p68 miRNA*) subit un processus de maturation par la voie classique DICER/Risc et s'associe avec la protéine Argonaute Ago2 (Moore *et al.*, 2011). L'équipe montre par ailleurs que ce microARN est fonctionnel et peut inactiver l'expression d'un transcrite contenant une séquence complémentaire. Cependant, aucune donnée physiologique n'a pu être apportée. Ce microARN régule-t-il l'expression de plusieurs gènes ? S'agit-il d'une autorégulation, à savoir que ce microARN régulerait l'expression de son propre gène et donc de la protéine p68 ? Sachant que p68 est impliquée dans la maturation des microARN, p68 serait-elle requise pour la maturation de *p68 miRNA* ? Enfin, quel est son rôle

physiologique, notamment dans les cancers où p68 et p72 sont surexprimées ? Des analyses complémentaires devront être effectuées pour répondre à ces questions.

Plusieurs autres voies de signalisation impliquant l'ARN nécessitent l'activité hélicase, et ainsi les ARN hélicases p68 et p72. Par exemple les orthologues de p68 chez la levure et la drosophile affectent le Nonsense Mediated mRNA Decay (NMD) et la libération de l'ARN de la chromatine après sa transcription (Bond *et al.*, 2001; Buszczak & Spradling, 2006). p68 et p72 interviennent également dans les défenses anti-virales et/ou peuvent être déroutées de leur rôle au profit de la réplication virale. La protéine ZAP (pour Zinc-finger Antiviral Protein), qui interagit avec les ARNm viraux et est impliquée dans leur dégradation, interagit avec p72, dont l'activité enzymatique permettrait un changement structural de l'ARN viral et ainsi faciliterait sa dégradation (Chen *et al.*, 2008). D'autre part, p68 est utilisée par l'ARN polymérase ARN-dépendante du virus de l'hépatite C pour faciliter la production du brin antisens à partir du brin sens de l'ARN viral (Goh *et al.*, 2004).

e) Rôle dans la différenciation myogénique et ostéoblastique

Une récente étude a montré que p68 et p72 s'associent avec l'ARN non-codant *Steroid Receptor RNA* (SRA) pour initier l'activité transcriptionnelle de MyoD, un facteur de transcription ayant un rôle-clé dans la détermination et la différenciation des cellules musculaires striées (Caretto *et al.*, 2006). p68 et p72 co-immunoprécipitent avec MyoD et SRA, suggérant que ces facteurs appartiennent à un même complexe d'initiation de la transcription (Caretto *et al.*, 2007). Plusieurs modèles ont été proposés et suggèrent que p68 et p72 pourraient être importantes pour le recrutement de facteurs de transcription spécifiques de la machinerie de transcription, incluant des facteurs de remodelage de la chromatine, qui pourraient faciliter la formation et la stabilisation du complexe d'initiation.

Plusieurs expériences ex-vivo ont également montré le rôle important de p68 et p72 dans la différenciation myogénique (Figure 34). Ainsi, la surexpression de p68 et p72 avec l'ARN non-codant SRA induit la différenciation myogénique de fibroblastes murins C3H10T1/2, alors qu'à l'inverse l'inactivation conjointe de p68 et p72 par shRNA (small-hairpin RNA) inhibe la différenciation myogénique de cellules musculaires murines C2C12 (Caretto *et al.*, 2006). De manière intéressante, ces défauts de différenciation sont corrigés par la surexpression de p68 (Caretto *et al.*, 2006), mais l'expérience n'a pas été tenté avec p72, ce qui ne permet pas de conclure quant au rôle de p72 dans la différenciation myogénique.

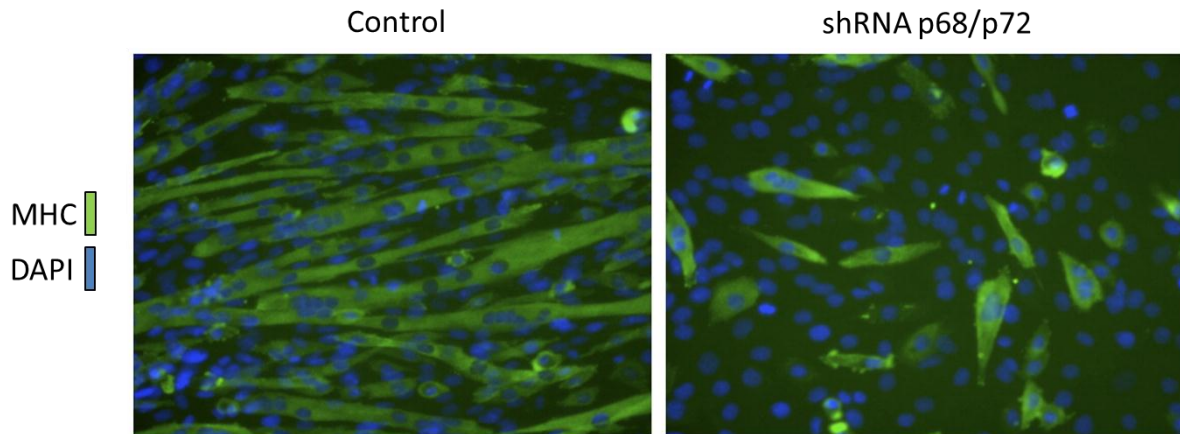


Figure 34 : *L'inactivation conjointe de p68/p72 altère l'expression de gènes musculaires et la différenciation myogénique.* Des cellules musculaires murines (C2C12) transduites avec un rétrovirus contrôle ou exprimant un shRNA dirigé contre p68 et p72. Les cellules ont été différenciées pendant 36 heures et la protéine Myosin Heavy Chain (MHC) est détectée par immunofluorescence. (Caretti *et al.*, 2006).

Des expériences d'immunoprécipitation de la chromatine (ChIP) ont montré que p68 était recrutée spécifiquement au niveau de régions régulatrices de gènes exprimés dans les myotubes, ce qui est en accord avec l'idée que p68 est recrutée au niveau du promoteur de gènes hautement transcrits sans être pour autant un co-activateur général de la transcription. Il n'existe à l'heure actuelle pas de données pour p72 et son recrutement au niveau de promoteurs de gènes musculaires.

Dans une autre étude, il a été montré que p68 interagit avec le facteur de transcription Runx2. p68 est recrutée au niveau du promoteur du gène codant l'ostéocalcine, une protéine spécifique des tissus osseux sécrétée par les ostéoblastes qui favorise la fixation du calcium sur les tissus conjonctifs osseux. De plus, ce promoteur contient un site d'interaction de Runx2 (Jensen *et al.*, 2008). Tout comme pour MyoD, l'extinction génique de p68 et p72 dans des C2C12 inhibe l'activité transcriptionnelle de Runx2, protéine essentielle pour la différenciation ostéoblastique. Cependant, l'inactivation de p68/p72 dans ces cellules accélère la maturation des ostéoblastes, suggérant qu'au-delà de son rôle de co-activateur de Runx2, p68 elle-même pourrait inhiber la différenciation (Jensen *et al.*, 2008).

f) Rôle dans la régulation du cycle cellulaire

Les ARN hélicases p68 et p72 sont également impliquées dans la régulation du cycle cellulaire par action sur la transcription de diverses cyclines et facteurs associés. En effet, p68 active la transcription de la cycline D1, protéine permettant à la cellule de passer un

checkpoint afin de rentrer dans le cycle cellulaire, par l'intermédiaire de la β -caténine (Yang *et al.*, 2006; Shin *et al.*, 2007; Yang *et al.*, 2007). De manière intéressante, il semble exister deux modes d'action par lesquels p68 et p72 affectent la β -caténine. Premièrement, le facteur *platelet-derived growth factor* (PDGF) stimule la phosphorylation de la tyrosine Y593 de p68 qui à son tour facilite la translocation de la β -caténine du cytoplasme au noyau (Yang *et al.*, 2006). L'autre mode de régulation proposé est que p68 et/ou p72 agiraient comme co-activateur transcriptionnel de la β -caténine dans le noyau et interagiraient ainsi avec les promoteurs de gènes-cibles de la β -caténine parmi lesquels la cycline D1 ou c-myc (Shin *et al.*, 2007; Yang *et al.*, 2007).

g) p68, p72 et leurs modifications post-traductionnelles

La première démonstration que p68 peut subir des modifications post-traductionnelle fut par la découverte que p68 pouvait être phosphorylée par la *protéine kinase C* (PKC), une protéine kinase partenaire de p68 (Buelt *et al.*, 1994; Rosenberger *et al.*, 2002) et qui est impliquée dans l'hyperphosphorylation de la CUGBP1 dans la Dystrophie Myotonique (Kuyumcu-Martinez *et al.*, 2007) comme précédemment vu dans la première partie de l'introduction. Cette phosphorylation inhibe l'activité ATPase de p68 (Buelt *et al.*, 1994) ainsi que la capacité de p68 à fixer l'ARN simple brin (Yang *et al.*, 2004), suggérant que la phosphorylation de p68 serait un moyen de réguler la fonction de cette ARN hélicase au cours de la vie cellulaire.

D'autres facteurs, comme par exemple PDGF ou tumor necrosis factor- α (TNF- α) peuvent phosphoryler les tyrosines de p68 (Yang *et al.*, 2005a, 2005b). Ces phosphorylations apparaissent comme étant prédominantes dans plusieurs lignées de cellules cancéreuses mais absentes de cellules non tumorales, suggérant que ces phosphorylations puissent être un marqueur moléculaire de certains types de cellules tumorales. Il fut montré par la suite que la stimulation de PDGF dans des cellules cancéreuses du colon active le proto-oncogène c-Alb, qui à son tour phosphoryle p68 au niveau de la tyrosine Y593 (Lin & Arlinghaus, 2008). Ce type de régulation semble concerner uniquement p68 du fait que ni p72 ni p82 ne possèdent de résidus tyrosine homologues aux tyrosines Y593 et Y595.

D'autres types de modifications post-traductionnelles existent, notamment l'ubiquitinylation (Causevic *et al.*, 2001; Mooney *et al.*, 2010) ou la sumoylation. Concernant cette dernière, une lysine (K) ultra-conservée localisée dans la partie N-terminale de p68 et p72 (K53 pour p68, K50 pour p72) est sumoylée *in vivo* (Fuller-Pace *et al.*, 2007; Jacobs *et*

al., 2007; Mooney *et al.*, 2010). Cette lysine conservée est contenue dans un domaine correspondant à une séquence consensus de sumoylation ψ KXE (où ψ correspond à un acide aminé aliphatique et X n'importe quel acide aminé) (Geiss-Friedlander & Melchior, 2007). La sumoylation de p68 et p72 augmente la stabilité de la protéine en doublant sa demi-vie, favorise son interaction avec HDAC1 sans pour autant modifier son affinité pour HDAC2, HDAC3, p300 ou ER α et affecte le potentiel co-activateur de p68 et p72 (Jacobs *et al.*, 2007; Mooney *et al.*, 2010).

h) Rôle dans le développement

La protéine orthologue de p68/72 chez la levure est la protéine Dbp2p. Bien qu'elle fut au préalable montrée comme essentielle pour la viabilité, une étude postérieure démontra que l'inactivation du gène DBP2 n'est pas létale, mais réduit grandement la croissance et conduit à un phénotype thermosensible (Iggo *et al.*, 1991; Barta & Iggo, 1995). De plus, des mutations de cette protéine chez la drosophile aboutissent à de multiples phénotypes allant de la stérilité à la mort (Buszczak & Spradling, 2006) suggérant l'importance de cette protéine au cours du développement. Cette protéine est également essentielle chez la souris, en effet des souris KO pour le gène DDX5 ou DDX17 présentent un phénotype létal (Fukuda *et al.*, 2007). Cependant, ce phénotype est différent suivant le gène inactivé : les souris KO DDX5 meurent au stade embryonnaire alors que les souris KO DDX17 survivent jusqu'à deux jours après la naissance (Fukuda *et al.*, 2007).

i) Implication de p68 et p72 dans les cancers

Il est aujourd'hui établi que p68 et p72 sont impliquées dans plusieurs types de cancers, elles sont notamment surexprimées dans des tumeurs colorectales (Causevic *et al.*, 2001). Cette surexpression pourrait intervenir dans le développement de cancers colorectaux et il existe une corrélation entre l'activation aberrante de la β -caténine dans ce type de cancer et le fait que p68 et p72 interagissent avec celle-ci et stimulent la transcription de c-myc, la cycline D1 et c-jun, qui sont des cibles transcriptionnelle de la β -caténine (Yang *et al.*, 2007). De plus, l'inactivation de p68/p72 dans des lignées cellulaires de cancer colorectal réduit la prolifération et la formation de tumeurs (Yang *et al.*, 2006; Shin *et al.*, 2007; Yang *et al.*, 2007).

De plus, p68 et p72 sont surexprimées dans les tumeurs du cancer du sein, de 30% à 70% suivant les études (Wortham *et al.*, 2009; Mooney *et al.*, 2010). Près de 70% des tumeurs sont également positives pour la présence de ER α . Sachant que p68 et p72 sont deux co-activateurs de ER α , leur surexpression pourrait favoriser l'oncogénicité de ER α , ce qui constituerait un mécanisme par lequel p68 et p72 contribuent au développement de tumeurs. p68 et p72 sont également surexprimées dans les tumeurs de la prostate et sont des co-activateurs du récepteur à l'androgène (Clark *et al.*, 2008), l'un des facteurs malins les plus caractéristiques des tumeurs prostatiques (Attard *et al.*, 2009). Enfin, l'équipe de Didier Auboeuf a très récemment montré que p68 et p72 étaient des médiateurs de la sélection de promoteurs alternatifs dans des gènes induits par l'œstradiol (Dutertre *et al.*, 2010a). Certaines cibles dont le gène NET1 possèdent des sites d'interaction pour ER α à proximité des promoteurs, et l'inactivation conjointe de p68 et p72 dans une lignée de cellules cancéreuses MCF-7 abolit l'effet de l'œstradiol sur le choix du promoteur alternatif, suggérant un rôle direct de ces protéines dans cette régulation (Dutertre *et al.*, 2010a). En accord avec le rôle de co-activateur de p68 et p72 dans la transcription, l'inactivation de ces protéines entraîne une baisse du recrutement de la polymérase II et de l'acétylation de l'histone H3 au niveau du promoteur exclu. A l'inverse, le recrutement de la polymérase II est favorisé au niveau du promoteur inclus (Dutertre *et al.*, 2010a).

L'équipe de Fuller-Pace a également montré que p68 est un co-activateur du facteur p53 (Bates *et al.*, 2005). En effet, p68 et p72 interagissent avec p53 *in vitro* et *in vivo* et p68 stimule la transcription de gènes régulés de façon transcriptionnelle par p53. L'inactivation de p68 par siRNA inhibe la transcription de gènes-cibles de p53 en réponse aux dommages de l'ADN et à l'apoptose (Bates *et al.*, 2005). Sachant que près de 50% des cancers humains sont caractérisés par une isoforme p53 mutée qui a perdu ses capacités anti prolifératives et apoptotiques, le complexe p68/p53 pourrait être un futur axe de recherche pour comprendre les mécanismes moléculaires à l'origine du développement de cancers.

Plusieurs questions restent cependant non élucidées. Quels sont les mécanismes moléculaires à l'origine de la surexpression de p68 et p72 dans les cancers ? Pourquoi p68 et p72 ne sont surexprimées qu'uniquement dans certains types de cancers ? Des modifications post-traductionnelles (sumoylation, phosphorylation...) sont-elles à l'origine d'une augmentation de la stabilité de p68 et p72 dans certains cancers ? D'autres ARN hélicases sont-elles également surexprimées dans les cellules cancéreuses ?

RESULTATS

RESULTATS

La purification des complexes protéiques assemblés sur les répétitions CUG a permis l'identification de la protéine DEAD-box ARN hélicase p68/DDX5. Nos travaux montrent que cette protéine interagit avec les expansions CUG *in vitro* et colocalise avec les foci CUG dans un modèle cellulaire reproduisant la maladie. Si cette protéine n'est pas séquestrée comme l'est MBNL1, elle facilite l'interaction de MBNL1 au niveau des répétitions CUG, par un mécanisme faisant intervenir son activité hélicase. Ces résultats suggèrent que p68 pourrait être importante pour l'accès de MBNL1 au niveau des expansions CUG *in vivo* et ainsi jouer un rôle dans sa séquestration. D'autre part, nous avons montré que p68 régule l'épissage alternatif du transcrit codant la troponine T cardiaque (*TNNT2*), en facilitant l'interaction de MBNL1 au niveau de ses séquences régulatrices sur l'ARN pré-messager. Nous avons identifié que son effet sur l'épissage alternatif est dépendant de MBNL1, signifiant que p68 n'est pas un facteur d'épissage « classique ». Nos résultats suggèrent que p68 pourrait être une ARN hélicase spécifique de MBNL1 qui lui faciliterait l'accès, par modulation locale de la structure secondaire, sur ses cibles normales ou pathologiques.

Ces résultats ont fait l'objet d'un article publié dans le journal *Nucleic Acids Research*. Cette publication est intitulée « New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats ».

New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats

François-Xavier Laurent¹, Alain Sureau¹, Arnaud F Klein², François Trouslard¹, Erwan Gasnier², Denis Furling² and Joëlle Marie^{1,*}

¹Centre de Génétique Moléculaire, CNRS, UPR 3404, Avenue de la Terrasse, 91198 Gif-sur-Yvette, Université Paris-Sud, Orsay and ²UPMC University Paris 06, Um76, Institut de Myologie and CNRS, UMR7215, and INSERM, U974, F-75013, Paris, France

Received August 22, 2011; Revised and Accepted November 23, 2011

ABSTRACT

Myotonic Dystrophy type I (DM1) is caused by an abnormal expansion of CTG triplets in the 3' UTR of the *dystrophia myotonica protein kinase (DMPK)* gene, leading to the aggregation of the mutant transcript in nuclear RNA foci. The expanded mutant transcript promotes the sequestration of the MBNL1 splicing factor, resulting in the misregulation of a subset of alternative splicing events. In this study, we identify the DEAD-box RNA helicase p68 (DDX5) in complexes assembled onto *in vitro*-transcribed CUG repeats. We showed that p68 colocalized with RNA foci in cells expressing the 3'UTR of the *DMPK* gene containing expanded CTG repeats. We found that p68 increased MBNL1 binding onto pathological repeats and the stem-loop structure regulatory element within the cardiac Troponin T (TNNT2) pre-mRNA, splicing of which is misregulated in DM1. Mutations in the helicase core of p68 prevented both the stimulatory effect of the protein on MBNL1 binding and the colocalization of p68 with CUG repeats, suggesting that remodeling of RNA secondary structure by p68 facilitates MBNL1 binding. We also found that the competence of p68 for regulating TNNT2 exon 5 inclusion depended on the integrity of MBNL1 binding sites. We propose that p68 acts as a modifier of MBNL1 activity on splicing targets and pathogenic RNA.

INTRODUCTION

Myotonic dystrophy type 1 (DM1) is a dominant autosomal neuromuscular disorder, characterized by multisystemic defects affecting muscle, heart, brain and

endocrine systems (1). DM1 is one of the most frequent forms of muscular dystrophy in adults that is caused by an expansion of CTG triplets in the 3' untranslated region of the *dystrophia myotonica protein kinase (DMPK)* gene (2–4). The mutant allele is transcribed, correctly spliced and polyadenylated but it is not translated into a protein. Instead, expanded transcripts are retained in the nucleus and accumulate as discrete RNA foci, located at the boundary of nuclear speckles (5–7). However, nuclear/cytoplasmic distribution of mutant DMPK remains controversial as cytoplasmic aggregates have also been detected (8). It is now well established that the expanded CUG repeats are toxic for the cells and play a major role in DM1 pathogenesis. Based on structural studies on short CUG repeats, it has been proposed that expanded CUG repeats fold into an imperfect hairpin structure that interferes with the activities of RNA binding proteins and alters their normal cellular function. The muscleblind-like 1 protein (MBNL1) was identified by its ability to bind to CUG repeats (9). MBNL1 is a splicing factor that interacts with RNA through its zinc-finger domains. Recently it has been shown that MBNL1 binds to single-stranded and structured RNA, which raises the question of how the protein recognizes different RNA targets (10). In the DM1 context, MBNL1 colocalizes with nuclear foci of CUG repeats and its sequestration by the expanded repeats leads to a loss-of-function of the protein [review in refs (11,12)]. Another RNA-binding protein involved in DM1 is CUGBP1. This factor plays an important role in several post-transcriptional processing steps, including translation, RNA stability and alternative splicing. Instead of being sequestered by the repeats, the steady-state level of CUGBP1 is increased in DM1 tissues due to its hyperphosphorylation by protein kinase C, leading to a gain of activity of the protein (13). MBNL1 and CUGBP1 are antagonistic splicing factors that are involved in the reprogramming of splicing events during fetal to post-natal development in muscle (14). The sequestration

*To whom correspondence should be addressed. Tel: +33 1 69 82 38 00; Fax: +33 1 69 82 38 77; Email: marie@cgm.cnrs-gif.fr

of MBNL1 and the upregulation of CUGBP1 in DM1 are proposed to trigger the misregulation of alternative splicing of a subset of muscle and brain-specific transcripts, leading to the re-expression of fetal isoforms in adult tissues. Several of these abnormal splicing events correlate with clinical symptoms such as myotonia, insulin resistance and heart conduction defects, which are caused by the misregulation of the muscle-chloride channel, insulin receptor (INSR) and cardiac Troponin T (TNNT2) pre-mRNA, respectively (15–18). In addition to MBNL1 and CUGBP1, other factors such as transcription factors and hnRNP F/H have been proposed to be involved in DM1 (19,20). However, their function in the disease is unclear. Studies using the CUG RNA fly model have identified modifiers of CUG toxicity phenotypes (21), suggesting that factors or signaling pathway other than MBNL1 and CUGBP1 could be involved in DM1 pathogenesis. Recently, the mislocalization of the transcription factor SHARP in DM1 has been associated with alteration of steady-state levels of numerous mRNAs that are important for muscle development (22). Moreover, a recent study has demonstrated that the processing of the pre-miR1 is altered in DM1, reinforcing the idea that other mechanisms are involved in DM1 pathophysiology (23).

The aim of this work was to isolate new factors that bind to CUG repeats. Using an affinity chromatography strategy with an RNA containing 95 pure CUG repeats, we identified the RNA helicase p68 (DDX5). p68 is a prototype of DEAD-box RNA helicase proteins. This family is characterized by a conserved core, consisting of nine conserved motifs including the DEAD signature, which gives rise to the name to these proteins (24). p68 is involved in many aspects of RNA metabolism including transcription, RNA processing, RNA export, translation and mRNA degradation (25,26). Here, we show that p68 colocalizes with RNA foci in different cell types that overexpressed expanded CUG repeats. *In vitro* experiments suggest that p68 promotes a conformational change of CUG repeats that favors MBNL1 binding or stabilizes the complex through additional interactions. We also found that p68 regulates alternative splicing of TNNT2 exon 5 and we demonstrate that the competence for regulation depends on MBNL1. From these results, we propose that p68 acts as a modifier of MBNL1 activity on splicing targets and pathogenic RNAs.

MATERIALS AND METHODS

Plasmids and constructions

Plasmid CTG 95 was constructed by cloning polymerase chain reaction (PCR) fragment containing CTG repeats in plasmid pSP72 that was digested with PvuII. CTG repeats were obtained by PCR using a sense oligonucleotide containing 7 CTG repeats and an antisense oligonucleotide containing 7 CAG repeats. PCR fragments containing varying lengths of CTG repeats were purified on agarose gel and cloned into the pSP72 plasmid. Several clones were subjected to sequence analysis. According to the orientation of the repeats in the plasmid, different lengths of

CUG or CAG repeats were obtained. Plasmid CTG 95 contains 95 CTG repeats. Plasmid CAG 61 contains 61 CAG repeats. Plasmids containing 14 CTG repeats or 16 CAG repeats were chosen for gel retardation assays. Plasmid-containing 62 CCTG repeats were obtained by PCR using a sense oligonucleotide containing 6 CCTG repeats and an antisense oligonucleotide containing 6 CAG repeats. PCR fragments were cloned into the pSP72 plasmid and proceed as for CTG and CAG plasmids. Plasmids containing the 3'UTR of *DMPK* gene with 5 or 200 pure CTG repeats were described previously (27). Plasmid-expressing *DMPK* exons 11–15 containing 960 interrupted CUG repeats in exon 15 were described previously (28). The 3' UTR of *DMPK* gene containing 960 interrupted CTG was also cloned into a Tet-on inducible lentiviral construct as previously described (29). Mutations in the helicase core domains II and IV of human p68/DDX5 were made by reverse PCR from pcDNA4 p68/Ha-Myc-His provided by Dr D. Auboeuf. For mutation in domain II (p68 mt2), sense and antisense oligonucleotides are 5'-AGATAGA ATGCTTGATATGGGC-3' and 5'-GCTTCATTAAGG ACAAGGTAGG-3', respectively. For mutation in domain IV (p68 mt4), sense and antisense oligonucleotides are 5'-GCTGTGGAAACCAAAAGAAGA-3' and 5'-AA CAATGGTTTTATTCTCCTTCTC-3', respectively. The wild-type and mutant 4G cardiac Troponin T (TNNT2) minigenes were described previously (30). Mutant TNNT2, CUG stem (CUGS) and GUF were described previously (31). The TNNT2 RNA used for UV-cross-linking experiments was generated by PCR from the wild-type TNNT2 plasmid using sense and antisense oligonucleotides (5'-ACACATACGATTTAGGTGACA CTATAGAACCCAGACTAACCTGT-3' and 5'-CTGA GGTTCAGGGAGTGG-3', respectively). Plasmid p68 for expression in *Escherichia coli* was generated by PCR from pcDNA4 p68/Ha-Myc-His using a sense oligonucleotide (5'-ATCTAGGATCCATGTCGGGTTATTC GA-3') and an antisense oligonucleotide (5'-AGATCTC GAGTTGGGAATATCCTGT-3'). The PCR product after digestion with NdeI and XhoI was cloned in a variant of pet28 (gift from Dr H. Le Hir) that was digested by NdeI and XhoI. This vector contained in the N-terminal a CBP tag followed by a TEV protease and in the C-terminal a His-tag (32). The deletion of a C-terminal part of p68 giving rise to p68 Δ Ct2 was made by reverse PCR using a sense oligonucleotide (5'-CTCGAGACCA CCACCACCACCACTGA-3') and an antisense oligonucleotide (5'-ATAATTTTCCCTGTCTCTAA-3') using pet28/p68 as a template. p68 Δ Ct2 containing mutations in the helicase core domains II and IV were made by reverse PCR from wild-type p68 Δ Ct2. Plasmid pGEX-6P1-MBNL1/40 kDa isoform was described previously (33).

Affinity capture of protein complexes and MALDI analysis

HeLa cell nuclear extracts were purchased from A. Miller (Cilbiotech, B-7000 Mons, Belgium). Myoblast and myotube nuclear extracts from C2C12 were prepared as

described (34). Biotinylated *in vitro*-transcribed CUG95 repeats bound to streptavidin agarose were incubated with 40% nuclear extracts under splicing conditions as described (35). Proteins eluted from the beads were separated by electrophoresis and detected by Coomassie staining. Fourteen protein bands were excised from the gel and identified using a nanoLC MS/MS-Orbitrap spectrometer (platform of proteomic of South-West Paris, INRA).

Purification of recombinant proteins

Recombinant protein GST-MBNL1 and GST-UAP56 (DDX39) were expressed in *E. coli* BL21 (DE3) and purified using Glutathione Uniflow resin (BD biosciences Clontech) according to standard procedure. GST-UAP56 was treated by thrombin to eliminate the GST tag according to standard procedure. Recombinant protein p68 Δ Ct2 and eIF4A3 were expressed in *E. coli* BL21 (DE3) and successively purified on calmodulin resin (Stratagene) and on nickel sepharose (GE-healthcare) as previously described (32). Proteins were dialyzed against buffer D (36).

Western blot analysis and antibodies

Western blots were performed according to standard procedure with the following antibodies: anti-MBNL1 monoclonal antibodies [MB1a(4A8), gift from Dr G. Morris (7)], anti-p68 /DDX5 monoclonal antibodies (1/200, SantaCruz, sc 81350), anti-p72/DDX17 polyclonal antibodies (1/1000, Bethyl laboratories A300-509 A), anti-Myc monoclonal antibodies (1/500, sigma), anti-Flag monoclonal antibodies (1/500), anti-Emerin (1/500, SantaCruz, sc25284). The bands were detected with the SuperSignal West Pico detection kit (Pierce) and quantified with a Fuji LAS 3000.

UV-cross-linking, immunoprecipitation and gel retardation experiments

RNAs for gel shift experiments and UV-cross-linking experiments were transcribed using a cap analogue and SP6 or T7 RNA polymerase in the presence of [α - 32 P]UTP or [α - 32 P]CTP. Uniformly labeled RNA (10 fmol) were incubated for 15 min at 30°C, in 10 μ l with recombinant MBNL1 and p68 as indicated in the figures under splicing conditions with 0.5 mM ATP, 1 mM MgCl₂ and 20 mM creatine-phosphate. Samples were UV irradiated for 15 min and treated with 0.3 μ g RNase A and 10 U of RNase T1 at 37°C for 30 min. Immunoprecipitation experiments were carried out as described with anti-MBNL1 (clone 3A4, sc47740 SantaCruz) or p68 monoclonal antibodies immobilized on A/G agarose (34). The cross-linked proteins were resolved on a 10% sodium dodecyl sulfate (SDS) polyacrylamide gel and visualized by a PhosphorImager. RNA mobility shift assays were performed as described with recombinant MBNL1 and p68 Δ Ct2 as indicated in the figures (37). Recombinant proteins were diluted in Dignam buffer D containing 0.5 mg/ml bovine serum albumin (BSA). The reaction was incubated for 15 min at 30°C and the protein complexes were resolved by electrophoresis on a 4% non-denaturing polyacrylamide gel

(39/1), 0.5 \times Tris/Borate/EDTA buffer (TBE). RNA bands were quantified using a PhosphorImager (Molecular Dynamics).

Cell culture

Cos, HeLa and N1E115 cells were grown at 37°C in Dulbecco's-modified Eagle's medium with 4500 mg/l D-glucose (DMEM) supplemented with 10% fetal bovine serum and antibiotics. C2C12 cells were grown in the same medium supplemented with 20% fetal bovine serum. Human muscle cells were derived from primary human satellite cells and immortalized as described (38). Cells were grown in Ham's F10 medium supplemented with 20% fetal bovine serum and antibiotics. Cultures of human myoblasts were also infected with lentiviral vector expressing the Tet-on inducible 960 CTG construct. Transduction was carried out overnight in the presence of polybrene (4 μ g/ml) as previously described (39). Cells were grown in the same medium as above but with 50 μ g/ml Gentamicin (Invitrogen). For immunofluorescence cells were seeded in 24-well vessel and grown to ~90% confluence. Cells were co transfected with 0.4 μ g of plasmids expressing the repeats and 0.4 μ g of plasmids encoding proteins using lipofectamin 2000 (Invitrogen) according to the manufacturer's recommendations. For plasmid DNA transfection, human muscle cells were seeded in six-well vessel, and grown to ~90% confluence. Cells were cotransfected with 0.5 μ g minigenes and protein expression vectors with lipofectamine LTX plus reagent (Invitrogen, ratio 3:1), according to the manufacturer's recommendations. Cells were harvested 24 h later. RNA and proteins were prepared from the same plates using NucleoSpin RNA/protein kit (Macherey-Nagel) and splicing patterns were analyzed by RT-PCR in the presence of [32 P]dCTP as previously described (37). PCR products were resolved on 6% denaturing polyacrylamide gels and quantified using a PhosphorImager. siRNAs and primers are listed in [Supplementary Table S3](#). siRNA duplex was transfected at 50 nM with lipofectamine RNAiMAX (Invitrogen) in growth medium without antibiotics. Twenty-four hours later, the cells were cotransfected with 50 nM siRNA and 0.5 μ g of minigene and p68 or MBNL1 expression plasmid using lipofectamine 2000 (Invitrogen). The media was replaced with growth medium lacking antibiotics 3 h later. Cells were harvested 24 h after transfection of the minigene. RNA and proteins were prepared as described above.

RNA-FISH combined with immunofluorescence

FISH was done as described using (CAG)₈-Cy3 DNA oligonucleotide probe (40). Glass coverslips containing plated cells were fixed in 4% paraformaldehyde in PBS (pH 7.4) for 20 min and washed three times with PBS. The coverslips were incubated for 5 min in PBS/0.5% Triton X-100 and washed three times with PBS before prehybridization in 30% formamide, 2 \times SCC for 10 min. The coverslips were hybridized for 2 h at 37°C in 30% formamide, 2 \times SCC, 2 mM vanadyl ribonucleoside, 10 μ g/ml BSA and 0.5 μ g (CAG)₈-Cy3 DNA oligonucleotide probe (Sigma). The coverslips were washed twice in

2× SCC/30% formamide at 42°C and twice in 1× SCC at room temperature. Following FISH, the coverslips were washed twice in PBS. The coverslips were incubated with primary antibodies DDX5 (1/2000 dilution, Abcam ab21696), c-myc (1/500 dilution, clone 9E10; Santa Cruz), M2-FLAG (1/200 dilution, clone A2220, Sigma) and polyclonal antibodies DDX17 (1/400 dilution, Bethyl laboratories A300-509 A) in PBS/0.1% BSA at room temperature for 60 min. The coverslips were washed four times with PBS/0.1% Tween20 before incubation with a goat anti-rabbit or rabbit anti-mouse secondary antibody conjugated with Alexa-Fluor 488 (1/1000 dilution; Molecular probes, Invitrogen) in PBS/0.1% BSA for 60 min. Then, the coverslips were incubated for 10 min in PBS/0.1% BSA/DAPI (1/10 000 dilution) and rinsed twice in PBS before mounting in Fluorescent Mounting Medium (DAKO). Slides were examined using either a simple fluorescence microscope (Leica) or a Leica DM4000 B confocal microscope, equipped with a Leica 100× HCX Plan Apo CS 1.40 objective, in 1- μ m optical sections.

RESULTS

p68/DDX5 and p72/DDX17 form aggregates that colocalize with CUG repeats

In an attempt to find novel factors involved in DM1 pathophysiology, protein complexes assembled onto biotinylated CUG repeats were purified by affinity chromatography on streptavidin-agarose beads. We verified that MBNL1 from HeLa nuclear extracts became cross-linked to a biotinylated RNA containing 95 pure CUG repeats (Figure 1A). Therefore, streptavidin-agarose beads coupled to the biotinylated (CUG)95 RNA was used to capture protein complexes from HeLa or myogenic nuclear extracts. After several washing steps,

proteins were eluted and separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Numerous proteins were detected (Figure 1B). Some are common to the three nuclear extracts tested while others seemed to be specific to the cell type or to the differentiation status of muscle cells (Figure 1B). Fourteen bands were selected, from which about 100 proteins were identified by mass-spectrometry (Figure 1B and Supplementary Table S1). We focused on proteins that presented interesting features regarding their properties (double-stranded RNA proteins, splicing factors) or because they are involved in myogenesis (Figure 1B). To test the putative implication of the candidates in DM1, we investigated the localization of the proteins by RNA-FISH/immunofluorescence experiments in HeLa cells that had been cotransfected with a plasmid encoding the 3'UTR of the *DMPK* gene containing either 960 interrupted CTG repeats (28) or non-expanded repeats (5 CTG) as a control. Most of the candidates did not colocalize with RNA foci as exemplified by the RNA helicase UAP56/DDX39 and the splicing factor CDC5L (Supplementary Figure S1 and Supplementary Table S2). In contrast, the RNA helicase p68 (DDX5) colocalized with RNA foci (Figure 2). In the absence of expanded repeats, myc-tagged p68 showed a diffuse localization in the nucleoplasm. However in the presence of expanded repeats, p68-myc formed aggregates that colocalized with the foci as shown by FISH/immunofluorescence using anti-Myc monoclonal antibodies (Figure 2A). The colocalization of exogenous p68 was not cell-line dependent, because aggregates were observed in different cell types such as myogenic (C2C12) and neuronal (NIE115) cell lines (Supplementary Figure S2). We also showed that p68 accumulates with CUG RNA foci in cells that had been cotransfected with a 3'UTR *DMPK* minigene containing 200 pure CUG repeats (41) (Figure 2A). We next tested whether endogenous p68 colocalizes with CUG

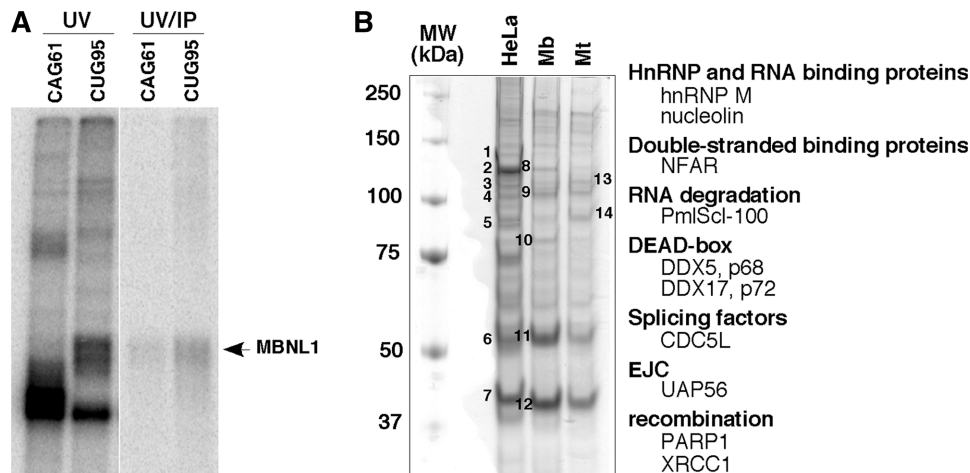


Figure 1. Affinity capture of protein complexes from HeLa and myogenic nuclear extracts by an *in vitro* transcribed RNA containing 95 CUG repeats. (A) MBNL1 is cross-linked to 32 P-labeled CUG95 repeats. *In vitro* transcribed (CUG)95 or (CAG)61 RNAs labelled with [α - 32 P]CTP were incubated with 40% HeLa nuclear extracts. After UV cross-linking an aliquot was precipitated with anti-MBNL1 antibodies. The cross-linked proteins were separated by SDS-PAGE. (B) Proteins from HeLa nuclear extracts, myoblast (Mb) and myotube (Mt) nuclear extracts from C2C12 were separated onto a 10% polyacrylamide denaturing gel and detected by Coomassie Brilliant Blue staining. Numbers 1 through 14 refer to bands that have been cut out from the gel and identified by mass-spectrometry. A subset of proteins identified by mass spectrometry is indicated.

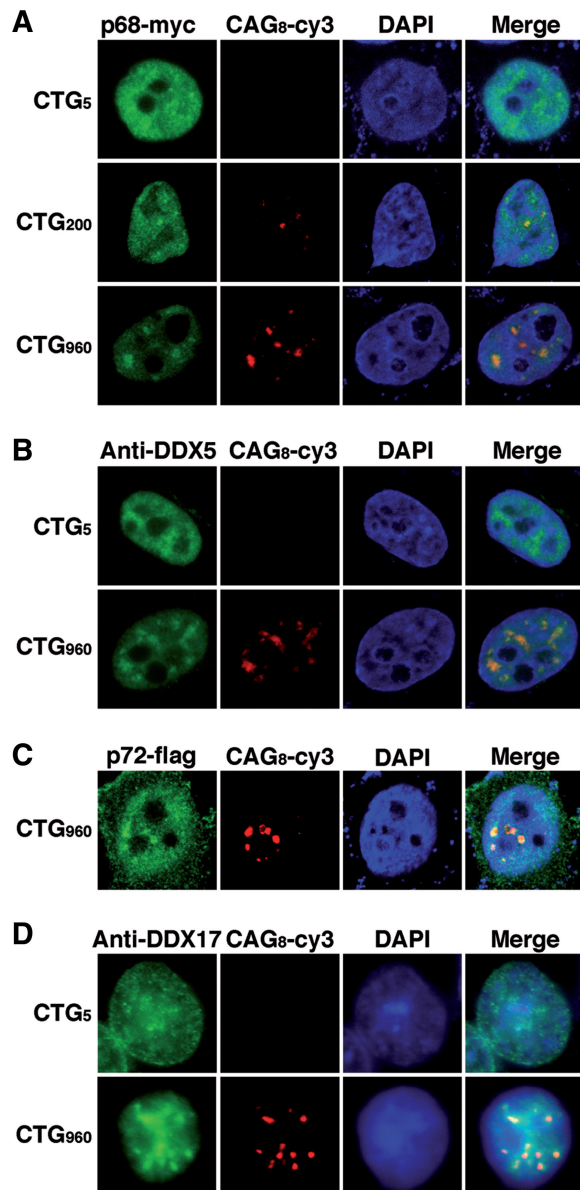


Figure 2. p68 and p72 colocalized with CUG RNA foci in HeLa cells. (A) Cells were cotransfected with a plasmid encoding p68-myc and plasmids expressing either 5, 200 or 960 CTG repeats. Expression of p68-myc was detected by immunofluorescence using monoclonal anti-Myc antibodies, followed by rabbit anti-mouse A488 antibodies. Foci were detected by RNA-FISH using Cy3-labeled (CAG)₈ probes. Nuclei were stained with DAPI. (B) Endogenous p68 colocalizes with CUG repeats. p68/DDX5 was revealed with a DDX5 monoclonal antibody. (C) Same experiment as in (A) except that cells were cotransfected with a Flag-p72 expression vector. Expression of Flag-p72 was detected by immunofluorescence using anti-Flag antibodies. (D) Endogenous p72 colocalizes with CUG repeats. P72/DDX17 was revealed with DDX17 polyclonal antibodies.

repeats. Immunofluorescence with monoclonal p68 antibodies confirmed that endogenous p68 colocalized with expanded CUG repeats (Figure 2B). p72 (DDX17) is a paralog of p68 that is highly similar to p68 (42). p72 was also found in complexes associated with *in vitro* transcribed 95 CUG repeats. We thus investigated whether p72 colocalized with the 960 CUG repeats. The

results showed that p72 tagged with a flag epitope and endogenous p72 formed aggregates that colocalized with RNA foci (Figure 2C and D). In conclusion, our results suggest that the DEAD-box RNA helicases p68 and p72 are novel factors that localize in CUG RNA aggregates that recruit MBNL1.

p68 is a modifier of MBNL1 binding on CUG repeats

Structural analyses have shown that MBNL1 binds to CUG repeats via its Watson-Crick interface, suggesting that the long A-form stem of the CUG repeats must be partially unfolded (43). One function ascribed to DEAD-box RNA helicases is to separate strands of short RNA duplexes to promote the remodeling of RNA-protein complexes (25). As RNA chaperones, these proteins facilitate the correct folding of structured-RNA molecules (44). Thus, we hypothesized that RNA helicase p68 and/or p72 could locally disrupt the G-C pairs of the CUG helix allowing the recruitment or the stabilization of MBNL1 to the repeats. To test this hypothesis, recombinant protein MBNL1 was cross-linked to *in vitro* transcribed ³²P RNA containing 95 CUG repeats in the absence or presence of recombinant p68 protein under splicing conditions. We focused on p68 because recombinant p72 was insoluble in our hands. To increase the solubility of the recombinant p68 in *E. coli*, 100 amino-acids of p68 were deleted from the C-terminus, keeping NLS and NES signals intact (45). This truncated p68 protein (p68 Δ Ct2) was still able to colocalize with the RNA foci of expanded CUG repeats and to regulate splicing (see below) as did the full-length p68 (Supplementary Figure S3). As previously observed in other studies (9), MBNL1 was cross-linked to CUG repeats even in the absence of p68 (Figure 3A). However, in the presence of increasing concentrations of recombinant p68, the amount of MBNL1 cross-linked to the repeats was significantly increased (Figure 3A and supplementary Figure S4). Immunoprecipitation with monoclonal antibodies against MBNL1 confirmed that the cross-linked band was indeed MBNL1 (Figure 3B). As already reported, p68 was not efficiently cross-linked to RNA using UV-cross-linking (46). However, we confirmed the interaction of p68 with CUG repeats by gel retardation assays (Supplementary Figure S5). It has been shown that MBNL1 was recruited to RNA foci in cells expressing CAG repeats (28). Thus, we tested whether p68 had an effect on MBNL1 binding on these repeats. Results shown in Figure 3C indicated that p68 was able to increase the binding of MBNL1 on CAG repeats in a similar range to those obtained on CUG repeats. Myotonic dystrophy type 2 (DM2) is the second form of myotonic dystrophy (47). It has been shown that expanded CCUG repeats in DM2 also form RNA foci that recruit MBNL1 (48). Interestingly, the stimulating effect of p68 was less efficient on CCUG repeats than on the other repeats (Figure 3C). This suggests that the effect of p68 on MBNL1 binding might depend on particular structural features embedded in the repeats. We next tested whether the stimulating effect on MBNL1 binding was specific to p68. UV-cross-linking experiments were

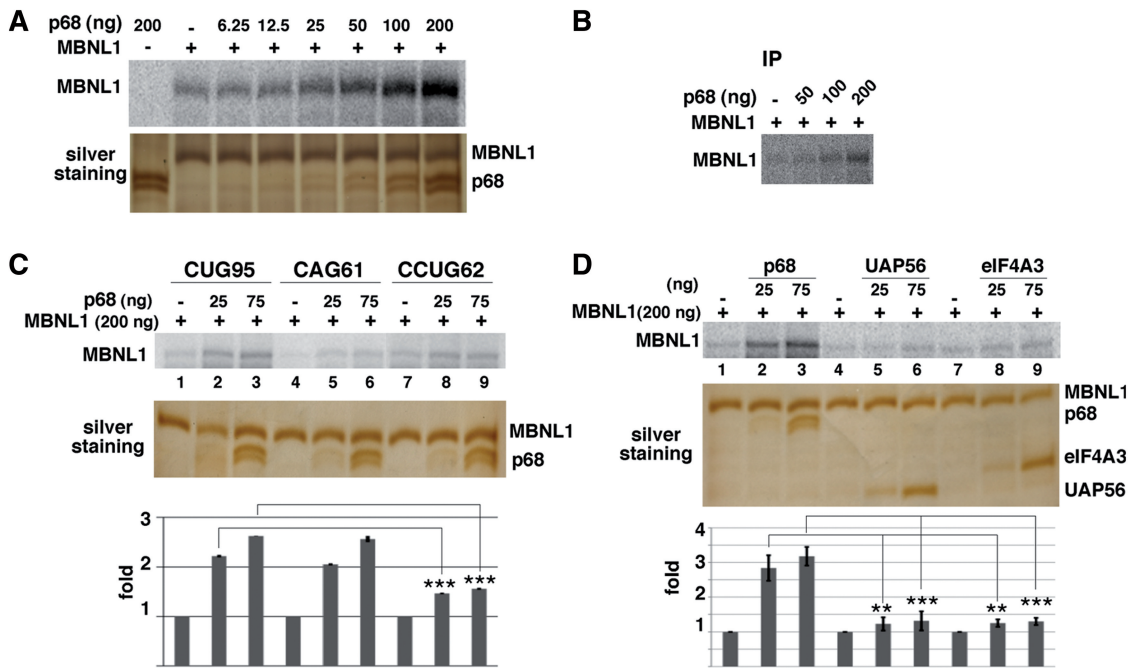


Figure 3. p8 increases the binding of MBNL1 to 95 CUG repeats. (A) Plasmid (CUG)95 was *in vitro* transcribed in the presence of [α - 32 P]UTP. Labeled (CUG)95 RNA was incubated with a constant amount of recombinant MBNL1 protein (200 ng) and increasing amounts of recombinant p68 Δ Ct2 under splicing conditions with ATP. Proteins cross-linked to labeled RNA were separated on a 10% SDS-PAGE. Bottom image shows a silver stain of a gel run in parallel. Note that p68 Δ Ct2 migrates as two bands that are recognized by anti-p68 antibodies (data not shown). (B) The cross-linked proteins shown in (A) were immunoprecipitated with anti-MBNL1 antibodies and separated by SDS-PAGE. (C) (CUG)95, (CAG)61 and (CCUG)62 RNAs were labeled with [α - 32 P]CTP and used for UV-cross-linking experiments. (D) The increase of cross-linked MBNL1 to CUG repeats is specific to p68. Labeled (CUG)95 was incubated with 200 ng of recombinant MBNL1 protein and 25 or 75 ng of recombinant p68 Δ Ct2, UAP56 or eIF4A3 proteins. Quantifications result from three independent experiments, with error bars indicating standard deviation. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

performed in the presence of UAP56/DDX39, one of the candidates found in complexes assembled onto (CUG)95 repeats and another DEAD-box protein, eIF4A3. The stimulation was markedly reduced in the presence of either UAP56 or eIF4A3, suggesting that p68 exhibits specificity toward the repeats despite the fact that the catalytic core of the DEAD-box helicase family is highly conserved (Figure 3D). To further characterize the requirement of p68 for MBNL1 binding, we tested whether mutations that affected ATPase/helicase activity and RNA binding altered the stimulatory effect of p68 on MBNL1 binding. Mutations D248N of the DEAD-box motif and mutation F346A in motif IV, which have been shown to be important for these activities in other DEAD-box proteins (24,49) reduced by 2-fold the ability of p68 to stimulate MBNL1 binding (Figure 4A). Consistent with this *in vitro* analysis, mutations in both the DEAD motif and RNA binding motif IV abolish colocalization of p68 to RNA foci (Figure 4B). Altogether, these results suggest that p68 acts as a specific factor to facilitate MBNL1 binding onto CUG repeats. Previous experiments have proposed that MBNL1 is a primary determinant of RNA foci formation (50). To examine whether knockdown of both p68 and p72 by siRNA has an effect on RNA foci formation, human muscle cells were transduced with a 960 CTG construct under the control of a Tet-on inducible promoter. RNA foci were detected as early as 1 h following the addition of

doxycycline to the culture medium (Figure 5). In contrast, the number of nuclei showing RNA foci was significantly reduced in induced cells treated by siRNA against p68 and p72, particularly at the early time points (1, 3 and 6 h) following doxycycline addition. No difference could be observed at 24 h between induced-cells that were treated or non-treated by siRNA against both p68 and p72 (Figure 5). Altogether, our results suggest that p68 and p72 play a role in the formation of RNA foci by facilitating the binding of MBNL1 and that the interaction of these proteins with RNA foci is transient.

p68 regulates alternative splicing of TNNT2 that is misregulated in DM1

One of the characteristic features of DM1 is the misregulation of alternative splicing of a subset of transcripts that resulted in the re-expression of fetal isoforms in adult tissues. It was shown that some of these events are responsive to MBNL1. In this study, we have found that p68 modifies MBNL1 binding on CUG repeats. Therefore, we wondered whether p68 played a role in the regulation of alternative splicing events that are regulated by MBNL1 and misregulated in DM1. To test this hypothesis we first inactivated both p68 and p72 by siRNA interference using a siRNA that targeted both proteins. Western blot analysis showed that about 95% of p68 was depleted from the cells (Figure 6A). The level of p72 depletion was also around 95% (Supplementary

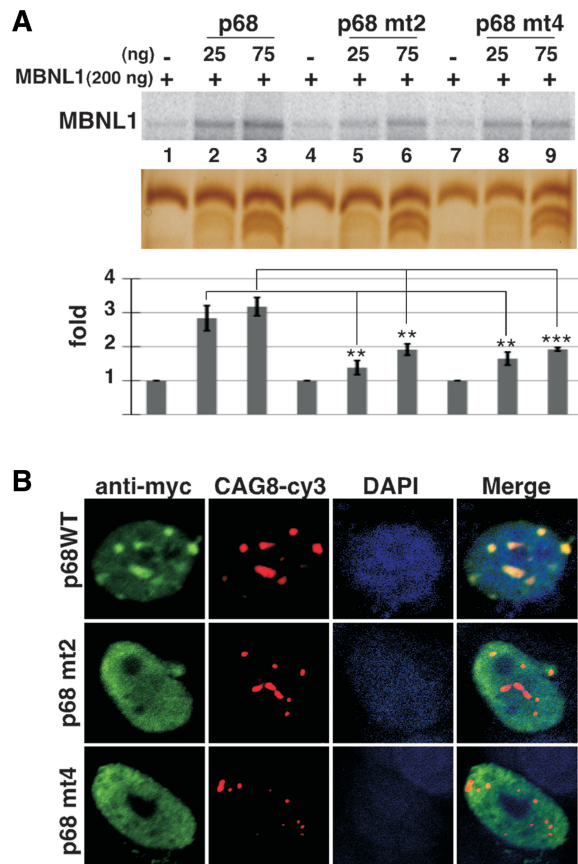


Figure 4. (A) Mutations in the helicase core domain of p68 modify its ability to increase MBNL1 binding. UV-cross-linking experiments with labeled (CUG)₉₅ were performed with 200 ng of recombinant MBNL1 and 25 or 75 ng of wild-type recombinant p68 Δ Ct2 or p68 Δ Ct2 mutated in domain II (p68 mt2, mutation D248N), or in domain IV (p68 mt4, mutation F346A). Quantifications result from three independent experiments, with error bars indicating standard deviation. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. (B) Mutations in the helicase core domain of p68 strongly affect colocalization of the protein with CUG repeats. RNA/FISH experiments were performed as described in Figure 2 with wild-type p68-myc expressing vector or p68-myc expressing vectors mutated in domain II (DEAD) or IV.

Figure S6). The knockdown of both p68 and p72-activated splicing of endogenous TNNT2 alternative exon 5 as a MBNL1 knockdown did, but it had no effect on alternative splicing of INSR exon 11 or on either sarco/endoplasmic reticulum Ca²⁺-ATPase (ATP2A1) exon 22 or MBNL1 exon 7 (Figure 6B). Thus, these results suggest that p68 and/or p72 regulates a subset of alternative splicing events. It has been found that p68 and p72 could display distinct functions. Notably, it has been shown that p68 and p72 have different abilities to activate splicing of the CD44 alternative exons (51). To test whether the increase of TNNT2 exon 5 inclusion upon inactivation of p68 and p72 was specific to one of these two proteins, we performed individual p68 and p72 knockdown experiments. All of the siRNAs tested that targeted different regions of p68 or p72 were able to increase TNNT2 exon 5 inclusion (Supplementary Figure S6). This suggests that both p68 and p72 are regulators of splicing TNNT2 exon 5. Recently, it has been

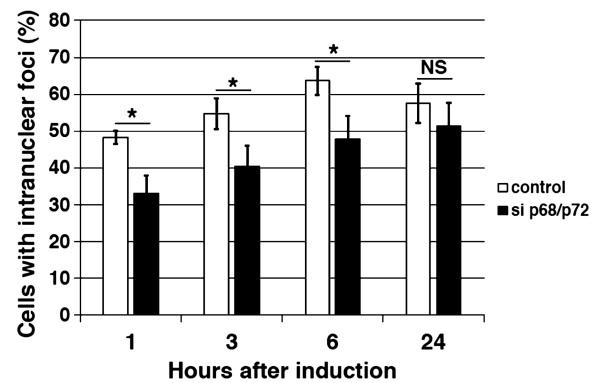


Figure 5. P68 and p72 facilitate RNA foci formation. Human muscle cells expressing 960-interrupted CTG repeats under the control of a Tet-on inducible promoter was used to attend the early steps of RNA foci formation. The number of nuclei containing foci was determined at different times after doxycycline addition (1, 3, 6 and 24 h) in non-treated or treated cells by siRNA to inhibit p68/p72. SiRNA duplexes were transfected 48 h before CUG repeats induction by doxycycline. Quantification results from three independent experiments. Approximately 300 cells were counted for each condition, with error bars indicating standard deviation * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

proposed that MBNL1 binding sites identified in TNNT2 are embedded in a stem-loop structure containing pyrimidine-pyrimidine mismatches that is similar to the CUG repeats, suggesting that MBNL1 recognizes these structures through a common mechanism (31). Thus, we hypothesized that RNA helicases p68 and p72 might facilitate MBNL1 binding in TNNT2 pre-mRNA as it does for the CUG repeats. To test this proposal, we performed UV-cross-linking experiments with an RNA encompassing the regulatory stem-loop element. Figure 7A showed that indeed p68 increased the binding of MBNL1 onto the regulatory element. Then, we asked whether mutations that reduce MBNL1 binding would impair regulation of TNNT2 exon 5 by p68. To answer this question, we used several TNNT2 minigenes (obtained from the groups of T. Cooper and A. Berglund), in which several mutations had been introduced to disrupt MBNL1 binding and/or the stem-loop RNA structure (Figure 7B) (30,31). Human myoblasts were cotransfected with wild-type and mutant TNNT2 minigenes that include exon 4–6 and with vectors overexpressing p68 or MBNL1. In agreement with siRNA experiments, overexpression of p68 repressed exon 5 inclusion as did MBNL1 (Figure 7B). As previously reported by Ho *et al.* and Warf *et al.*, the 4G minigene in which mutations disrupted the RNA structure and MBNL1 binding did not respond to MBNL1 overexpression (Figure 7B). Interestingly, the 4G mutant was no longer able to respond to p68. In contrast, the CUGS mutant that still responded to MBNL1 also remained responsive to p68 regulation (Figure 7B). Finally, the GUF mutant that resulted in complete inclusion of exon 5 and became insensitive to MBNL1 regulation was also insensitive to p68 regulation (Figure 7B). These results suggested that the competence for regulating alternative splicing of TNNT2 by p68 depended on the presence of MBNL1 binding sites and the ability of MBNL1 to regulate this

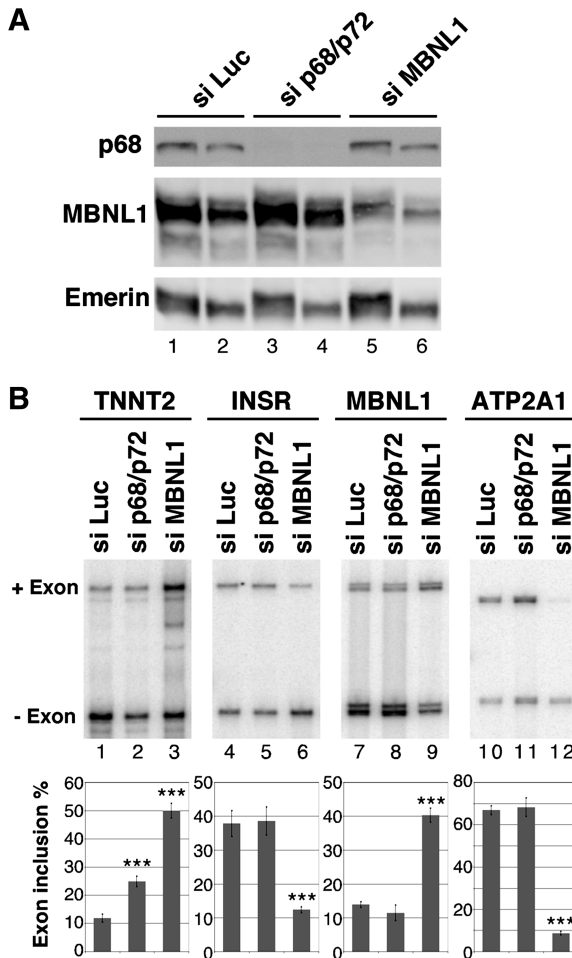


Figure 6. p68 regulates splicing of TNNT2 alternative exon 5. (A) Western blot analysis of human myoblasts treated with a control luciferase siRNA, (siLuc), sip68/p72 or siMBNL1. About 15 μ g (lanes 1, 3 and 5) or 7.5 μ g (lanes 2, 4 and 6) of proteins were loaded per lane. The blot was probed with antibodies against p68, MBNL1 and emerlin as a control. (B) RT-PCR analysis of endogenous transcripts in human myoblast cells transfected with a control siRNA (siLuc), siRNA against p68/p72 or siRNA against MBNL1. The histogram shows the quantification of exon inclusion. Results are derived from six independent experiments for TNNT2 and four independent experiments for INSR, MBNL1 and ATP2A1 with error bars indicating standard deviation. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

transcript. Therefore, we can anticipate that the inactivation of MBNL1 would result in the inability of p68 to regulate TNNT2 exon 5. To validate this hypothesis, we first cotransfected human myoblasts with the 3'DMPK UTR containing 5 or 960 CTG repeats and the TNNT2 minigene in the presence or absence of a vector coding for p68 (Figure 8A). We used an expression vector encoding PTB that is a repressor of exon 5 inclusion as a control (52). As previously observed in other studies and consistent with sequestration of MBNL1 by the repeats, expressing CUG repeats resulted in an increase of TNNT2 exon 5 inclusion (12) (Figure 8A). Interestingly, p68 was not able to regulate inclusion of TNNT2 exon 5 when 960 CUG repeats were expressed. In contrast, p68 was still able to regulate TNNT2 exon 5 in control experiments where only

A 5' GAACCCAGACUAACCGUCUCGCUUUUCCC
CUCGUCGCGGCCACUCCUGAACCCUCAG 3'

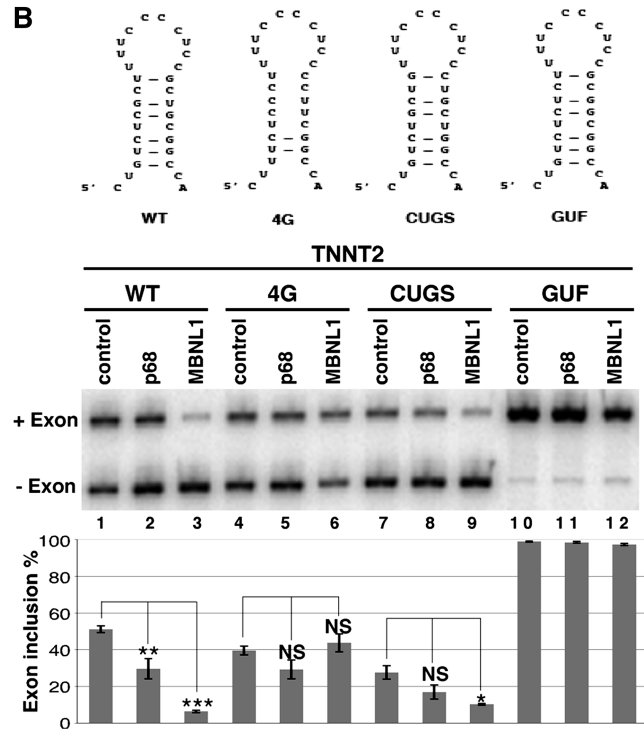
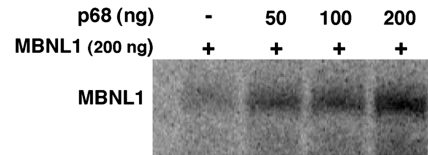


Figure 7. RNA stem-loop mutations that abolish MBNL1 binding impair the regulation by p68. (A) p68 stimulates binding of MBNL1 onto the stem-loop regulatory element of TNNT2. Labeled stem-loop regulatory RNA was incubated with a constant amount of recombinant MBNL1 protein (200 ng) and increasing amounts of recombinant p68 Δ Ct2. Proteins cross-linked to labeled RNA were separated on a 10% SDS-PAGE. The TNNT2 RNA used for cross-linking experiments is indicated. (B) RT-PCR analysis of splicing of TNNT2 alternative exon 5 in wild-type and mutant minigenes. Schematic representation of wild-type and mutant stem-loop structure within TNNT2 intron 4 [adapted from ref. (31)] is shown. Human myoblast cells were cotransfected with wild-type or mutant TNNT2 minigenes and expression vector coding for p68 (lanes 2, 5, 8 and 11) or MBNL1 (lanes 3, 6, 9 and 12). The histogram shows the quantification of exon 5 inclusion. Results are derived from at least three experiments, except for lanes 4, 6, 7, 9, 10 and 12, which are derived from two experiments. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. NS: non-specific.

five CUG repeats were expressed. This finding was not due to a general effect of the repeats on splicing, because repression of exon 5 inclusions by PTB was the same regardless of the expressed CUG repeats length. We then tested the ability of p68 to regulate TNNT2 exon 5 inclusions after inactivation of MBNL1 by siRNA (Figure 8B). In agreement with previous studies, knockdown of MBNL1 increased splicing of TNNT2 exon 5 (Figure 8B). As already shown (Figure 8B), cotransfecting cells with a plasmid coding p68 resulted in a 2-fold decrease of

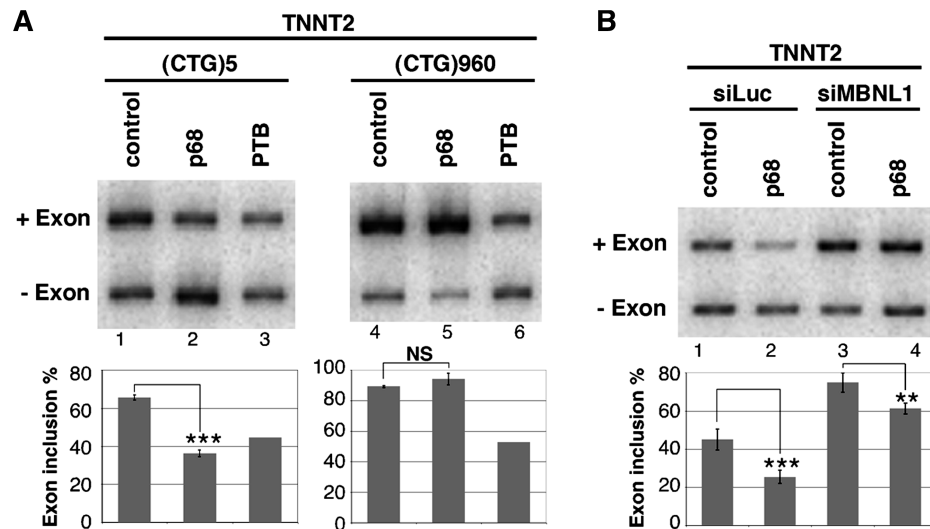


Figure 8. Sequestration of MBNL1 by CUG repeats or MBNL1 knockdown by siRNA makes splicing of TNNT2 exon 5 non-responsive to p68. (A) RT-PCR analysis of TNNT2 minigene in human myoblast cells cotransfected with the 3' UTR *DMPK* minigene containing 5 (lanes 1–3) or 960 (lanes 4 to 6) CUG repeats and a p68 expression vector (lanes 2 and 5) or as a control a PTB expression vector (lanes 3 and 6). The histogram shows the quantification of exon 5 inclusion. Results are from at least two experiments except lanes 3 and 6, for which a single experiment was performed. (B) RT-PCR analysis of TNNT2 minigene in human myoblast cells transfected with a control siRNA (lanes 1 and 2) or siRNA against MBNL1 (lanes 3 and 4) and a p68 expression vector (lanes 2 and 4). The histogram shows the quantification of exon 5 inclusion. Results are from four experiments with error bars indicating standard deviation. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TNNT2 exon 5 splicing. In contrast, splicing of TNNT2 exon 5 became less sensitive to p68 when cells were deprived of MBNL1. Consistent with these results, splicing of TNNT2 exon 5 became unresponsive to the knockdown of both p68 and p72 in DM1 myoblasts (Supplementary Figure S7). Altogether, these results demonstrate that p68 regulates the splicing of TNNT2 exon 5 alternative exon and suggest that p68-mediated regulation depends on MBNL1 binding.

DISCUSSION

DEAD-box proteins are a well-characterized class of RNA helicases that play essential roles in cellular processes involving RNA (25,26). Their basic functions are to remodel RNA and RNA-protein complexes. In this study, we identified DEAD-box RNA helicases p68 and p72 in complexes that assembled onto pure CUG repeats. We showed that these proteins colocalized with RNA foci formed by the 3'UTR of *DMPK* mRNA-containing 200 or 960 CUG repeats. In particular, we found that p68 increases the binding of MBNL1 onto *in vitro* transcribed CUG repeats. Moreover, we showed that TNNT2-mediated splicing regulation by p68 is dependent on MBNL1 binding sites. On the basis of our results we suggest that p68 acts as a modifier of MBNL1 activity by facilitating the binding of MBNL1 onto the repeats and splicing target.

Affinity chromatography is a powerful method that has been used to decode the composition of protein complexes that assemble onto RNA. Using pure CUG repeats as bait, about 100 proteins were identified by mass-spectrometry. Many of them are hnRNP and are involved in several aspect of RNA metabolism (Figure 1 and Supplementary Table 1). Surprisingly, MBNL1 was

not found in MS spectra. A similar observation was made by Kim *et al.* (53), who also did not detect MBNL1 in complexes purified using 46CUG repeats. However, in accordance with Miller *et al.* (9), we showed that MBNL1 is one of the major proteins from nuclear extracts that becomes cross-linked to the 95 CUG repeats (Figure 1A). In addition, MBNL1 was detected in the elution fraction from the affinity chromatography by western blot analysis (data not shown). Thus, the reason for the absence of MBNL1 from MS-spectra is not completely understood. One possible explanation is that, unfortunately, the band containing MBNL1 was not excised from the gel for identification by mass-spectrometry. Another possibility is that in our experimental conditions other proteins compete with MBNL1 for binding to the repeats. Among the proteins that we identified, only the DEAD-box proteins p68 and p72 form aggregates that colocalize with expanded CUG repeats. Interestingly, a study utilizing coexpression frequency shows that MBNL1 is coexpressed with p68 and p72, suggesting that these proteins could belong to the same signaling pathway (54). Additionally, another report shows that p68 is found in complexes associated with MBNL1 in myoblasts (55). The colocalization of p68 and p72 with RNA foci was observed with overexpressed RNA containing interrupted 960 CUG repeats or pure 200 CUG repeats and was independent of the cell lines. We also found that endogenous p68 and p72 colocalized with the repeats, as did exogenous proteins, suggesting that p68 and p72 are indeed recruited to RNA foci. However, in contrast to MBNL1 our results also show that only a fraction of p68 and p72 aggregates with RNA foci. p68 and p72 proteins are involved in multiple biological pathways and play a crucial role during development (25,26). It has been shown that the

disruption of either p68 or p72 genes in mice results in a very severe phenotype that leads to early lethality (56). In addition, the same study found that the processing of several microRNAs involved in cell survival is affected in p68 and p72 homozygous mice (56). Thus, according to these findings, a sequestration mechanism of p68 and p72 that leads to a deficiency of protein function similar to that of MBNL1 seems unlikely. In fact, no colocalization of p68 with RNA foci could be observed in muscle tissues from DM1 patients or mice expressing large CTG expansions (data not shown). As recently reported, we never observed recruitment of p68 in RNA foci of DM1 myoblasts (55). However, taking advantage of an inducible system in which the expression of expanded CUG repeats is under the control of an inducible promoter, we were able to show that the knockdown of both p68 and p72 has an effect on RNA foci formation (Figure 5). The effect on the number of newly formed foci was transient following doxycycline addition suggesting that p68 and p72 act at an early stage of RNA aggregation and then are released from the RNA foci. Interestingly a recent model has been proposed by Junghans to conciliate some conflicting results of DM1 pathobiology. This model poises the existence of insoluble and soluble RNA foci with distinct binding properties (57). Thus, one possibility would be that p68 and p72 intervene in the transition between these two forms of RNA foci.

One general task of DEAD-box RNA helicase is to modify RNA structure and/or RNA protein complexes (25,26). A recent crystal analysis of two zinc-finger domains of MBNL1 with a short CGCUGU RNA shows that MBNL1 targets GC steps, indicating that G–C base pairing within the helix has to be distorted (43). On the basis of our results, we propose that the recruitment of p68 to the CUG repeats would promote structural changes of the RNA that would facilitate and/or stabilize the binding of the splicing factor MBNL1 to the repeats. Consistent with that proposal, mutations in the helicase core domain of p68 strongly decreased both the stimulation of p68 on MBNL1 binding and the colocalization of p68 with CUG repeats (Figure 4). Although we focused on the role of p68, data in other publications suggest that p68 and p72 can exist as heterodimer in the cell (58). Indeed, our results show that p68 and p72 have the same function, suggesting that at least a fraction of p68 and p72 acts functionally as heterodimers (Figure 2 and Supplementary Figure S6). We also observed that the stimulatory effect operates on CAG repeats and on the regulatory intronic sequence of TNNT2 pre-mRNA (Figures 3C and 7A). Interestingly, several biochemical and structural studies predicted that pathogenic CUG repeats and TNNT2 RNA adopt similar RNA secondary structure (31,59). For CAG repeats, recent structural analysis also suggests that CAG and CUG duplexes share structural similarities (60). Thus, we can hypothesize that these common structural elements compose the basis for p68 recognition and subsequently for the increase of MBNL1 on the RNA. In contrast, p68 only slightly increases the binding of MBNL1 onto CCUG repeats. Interestingly, MBNL1 binds to CCUG RNA with an affinity 2-fold stronger than CUG RNA

(31). However, it has been shown that CCUG has a decreased stability compared to CUG (31). Thus, it is tempting to speculate that the requirement for p68 might depend on the stability of the RNA structure. We can suggest that the less an RNA is structured, the less p68 will be needed to distort the structure. Our results are consistent with a recent study that shows that MBNL1 is able to bind to both single-stranded and double-stranded RNAs (10). The mechanism by which p68 favors the binding of MBNL1 to the CUG repeats remains unknown. Time course experiments show that there is a huge increase of MBNL1 binding in the first few minutes following incubation with p68 (Supplementary Figure S4). It is known that the DEAD-box RNA helicases as RNA chaperones, modify the folding of RNA molecules. Thus, we can speculate that p68 accelerates the transition from misfolded RNA conformation to correct folding RNA structure, suitable for MBNL1 binding. Several studies point to the importance of U–U mismatches for MBNL1 binding (59,61). Thus, we propose that the mismatches could provide an anchoring site for p68 from which local strand separation could be initiated which would allow the interaction of MBNL1 to the repeats. It also has been shown that MBNL1 forms a ring-like structure with CUG repeats. From this finding we can suggest that *in vivo* p68 accelerates formation of this structure that is proposed to be the basis for MBNL1 sequestration (59).

In this study, we provide evidence that the alternative splicing of TNNT2, which is misregulated in DM1 is regulated by p68. We also found that alternative splicing of MBNL1 exon 7, ATP2A1 exon 22 or INSR exon 11, which are regulated by MBNL1 and misregulated in DM1, are not sensitive to p68 and/or p72, suggesting that these proteins regulate specific classes of transcripts. This finding raises an intriguing question. Why does alternative splicing of some transcripts regulated by MBNL1 require p68? The answer is not yet understood, but could be related to the architecture of the MBNL1 binding sites. Indeed, a recent study of Cass and coworkers suggests that different modes of MBNL1 recognition must operate for single-stranded and structured RNA targets (10). In the case of TNNT2, MBNL1 binds to a short stem-loop RNA structure, and we show that p68 increases the binding of MBNL1 onto this RNA structure. In contrast, at least for the INSR and MBNL1 transcripts no potential RNA secondary structure has been identified within the MBNL1 response element (62,63). Thus, we can propose that p68 would be required only for splicing RNA targets in which MBNL1 binding sites are embedded into a stem-loop structure. The modification of RNA stem-loop structure is a common theme for regulating alternative splicing by p68. However, the mechanism by which p68 mediates its effect is different according to the different targets. In the case of alternative splicing of Tau exon 10 and IDX alternative exon of c-H-ras pre-mRNA, it has been shown that p68 promotes the access of U1snRNP to the 5' splice site (64) or block the binding of hnRNP H (65), respectively. Our findings provide a new mode for how p68 can regulate alternative splicing by increasing the activity of a splicing factor. We also found that preventing the binding of MBNL1 onto the RNA, either by

sequestration of the protein by expanded CUG repeats or by mutations in the MBNL1 binding sites, made TNNT2 exon 5 insensitive to p68 regulation. This suggests that the repression of TNNT2 exon 5 by p68 is dependent on MBNL1 binding to the regulatory element, and it supports the idea that p68 would not act as a splicing factor per se, but rather, would be a coregulator or a coadaptator of MBNL1 binding. Thus, we favor a model according to which p68 modifies the stem-loop RNA structure of TNNT2 as it does on the toxic CUG repeats, which allows the stabilization of MBNL1.

In conclusion, our results have identified a modifier of MBNL1 activity. We suggest that through a common mechanism, the DEAD-box RNA helicase p68 modifies the RNA secondary structures of splicing targets and pathological RNAs to facilitate or stabilize MBNL1 binding that may contribute to the pathogenesis of DM1.

SUPPLEMENTARY DATA

Supplementary data is available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–7, Supplementary Reference [66].

ACKNOWLEDGEMENTS

The authors are grateful to T. Cooper (Baylor University) for DMPK CTG 960 plasmid, TNNT2 and mutant TNNT2 4G minigenes; M. Mahadevan (University of Virginia) for DMPK CTG 5 and CTG 200 plasmids; A. Berglund (University of Oregon) for mutant TNNT2 CUGS and GUF plasmids; S. Kato (University of Tokyo), F. Fuller-Pace (University of Dundee) and D. Auboeuf (Centre Léon Bérard, Lyon) for p68 and p72 plasmids; G. Morris (Centre for Inherited Neuromuscular Disease; Robert Jones and Agnes Hunt Orthopaedic Hospital) for the MBNL1 (MB1a) monoclonal antibodies; H. Le Hir (Ecole Normale Supérieure, Paris) for recombinant eIF4A3 protein; O. Cordin and J. Banroques (Institut de Biologie Physico-chimique, Paris) for recombinant UAP56 protein. N. Gourrier and C. Branlant (Laboratoire AREMS, Nancy) for GST-MBNL1 plasmid; N. Sergeant for N1E115 cells. They acknowledge Kyle Tanner and Katriona Laurent-Price for their constant interest and for carefully reading the manuscript. They also thank all the members of the French DM1 network for helpful discussions.

FUNDING

Centre National de la Recherche Scientifique (CNRS); Association Française contre les Myopathies (AFM); Paris-Sud University, the AFM and Fondation des Treilles fellowships (to F.X.L.).

Conflict of interest statement. None declared.

REFERENCES

- Harper,P.S. (2001) *Myotonic Dystrophy*. W. B. Saunders, London.
- 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) 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**, 799–808.
- Mahadevan,M., Tsilfidis,C., Sabourin,L., Shutler,G., Amemiya,C., Jansen,G., Neville,C., Narang,M., Barcelo,J., O'Hoy,K. *et al.* (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science*, **255**, 1253–1255.
- 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) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science*, **255**, 1256–1258.
- 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**, 995–1002.
- 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. USA*, **94**, 7388–7393.
- 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**, 1035–1048.
- Schoser,B. and Timchenko,L. (2010) Myotonic dystrophies 1 and 2: complex diseases with complex mechanisms. *Curr. Genomics*, **11**, 77–90.
- 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**, 4439–4448.
- Cass,D.M., Hotchko,R., Barber,P., Jones,K., Gates,D.P. and Berglund,J.A. (2011) The four Zn fingers of MBNL1 provide a flexible platform for recognition of its RNA binding elements. *BMC Mol. Biol.*, **12**, 20.
- O'Rourke,J.R. and Swanson,M.S. (2009) Mechanisms of RNA-mediated disease. *J. Biol. Chem.*, **284**, 7419–7423.
- Ranum,L.P. and Cooper,T.A. (2006) RNA-mediated neuromuscular disorders. *Annu. Rev. Neurosci.*, **29**, 259–277.
- 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**, 68–78.
- Lin,X., Miller,J.W., Mankodi,A., Kanadia,R.N., Yuan,Y., Moxley,R.T., Swanson,M.S. and Thornton,C.A. (2006) Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.*, **15**, 2087–2097.
- Charlet,B.N., Savkur,R.S., Singh,G., Philips,A.V., Grice,E.A. and Cooper,T.A. (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol. Cell*, **10**, 45–53.
- 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**, 35–44.
- 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**, 737–741.
- 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**, 40–47.
- Ebralidze,A., Wang,Y., Petkova,V., Ebralidze,K. and Junghans,R.P. (2004) RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. *Science*, **303**, 383–387.
- 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**, 4271–4283.

21. Garcia-Lopez,A., Monferrer,L., Garcia-Alcover,I., Vicente-Crespo,M., Alvarez-Abril,M.C. and Artero,R.D. (2008) Genetic and chemical modifiers of a CUG toxicity model in *Drosophila*. *PLoS One*, **3**, e1595.
22. 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**, 735–742.
23. Rau,F., Freyermuth,F., Fugier,C., Villemin,J.P., Fischer,M.C., Jost,B., Dembele,D., Gourdon,G., Nicole,A., Duboc,D. *et al.* (2011) Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat. Struct. Mol. Biol.*, **18**, 840–850.
24. Cordin,O., Banroques,J., Tanner,N.K. and Linder,P. (2006) The DEAD-box protein family of RNA helicases. *Gene*, **367**, 17–37.
25. Fuller-Pace,F.V. (2006) DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Res.*, **34**, 4206–4215.
26. Janknecht,R. (2010) Multi-talented DEAD-box proteins and potential tumor promoters: p68 RNA helicase (DDX5) and its paralog, p72 RNA helicase (DDX17). *Am. J. Transl. Res.*, **2**, 223–234.
27. 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**, 1975–1984.
28. Ho,T.H., Savkur,R.S., Poulos,M.G., Mancini,M.A., Swanson,M.S. and Cooper,T.A. (2005) Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy. *J. Cell Sci.*, **118**, 2923–2933.
29. Chaouch,S., Mouly,V., Goyenvall,A., Vulin,A., Mamchaoui,K., Negroni,E., Di Santo,J., Butler-Browne,G., Torrente,Y., Garcia,L. *et al.* (2009) Immortalized skin fibroblasts expressing conditional MyoD as a renewable and reliable source of converted human muscle cells to assess therapeutic strategies for muscular dystrophies: validation of an exon-skipping approach to restore dystrophin in Duchenne muscular dystrophy cells. *Hum. Gene Ther.*, **20**, 784–790.
30. 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**, 3103–3112.
31. 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**, 2238–2251.
32. Ballut,L., Marchadier,B., Baguet,A., Tomasetto,C., Seraphin,B. and Le Hir,H. (2005) The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.*, **12**, 861–869.
33. Tran,H., Gourrier,N., Lemerrier,C., Dhaenens,C.M., Vautrin,A., Fernandez-Gomez,F.J., Arandel,L., Carpentier,C., Obriot,H., Eddarkaoui,S. *et al.* (2011) Analysis of exonic-regions involved in nuclear localization, splicing activity and dimerization of muscleblind-like-1 isoforms. *J. Biol. Chem.*, **286**, 16435–16446.
34. Sauliere,J., Sureau,A., Expert-Bezancon,A. and Marie,J. (2006) The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. *Mol. Cell Biol.*, **26**, 8755–8769.
35. 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**, 16614–16623.
36. Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.
37. 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**, 94–106.
38. Zhu,C.H., Mouly,V., Cooper,R.N., Mamchaoui,K., Bigot,A., Shay,J.W., Di Santo,J.P., Butler-Browne,G.S. and Wright,W.E. (2007) Cellular senescence in human myoblasts is overcome by human telomerase reverse transcriptase and cyclin-dependent kinase 4: consequences in aging muscle and therapeutic strategies for muscular dystrophies. *Aging Cell*, **6**, 515–523.
39. Francois,V., Klein,A.F., Beley,C., Jollet,A., Lemerrier,C., Garcia,L. and Furling,D. (2011) Selective silencing of mutated mRNAs in DM1 by using modified hU7-snRNAs. *Nat. Struct. Mol. Biol.*, **18**, 85–87.
40. Taneja,K.L. (1998) Localization of trinucleotide repeat sequences in myotonic dystrophy cells using a single fluorochrome-labeled PNA probe. *Biotechniques*, **24**, 472–476.
41. Amack,J.D. and Mahadevan,M.S. (2001) The myotonic dystrophy expanded CUG repeat tract is necessary but not sufficient to disrupt C2C12 myoblast differentiation. *Hum. Mol. Genet.*, **10**, 1879–1887.
42. Lamm,G.M., Nicol,S.M., Fuller-Pace,F.V. and Lamond,A.I. (1996) p72: a human nuclear DEAD box protein highly related to p68. *Nucleic Acids Res.*, **24**, 3739–3747.
43. 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**, 1343–1351.
44. Woodson,S.A. (2010) Taming free energy landscapes with RNA chaperones. *RNA Biol.*, **7**, 677–686.
45. Wang,H., Gao,X., Huang,Y., Yang,J. and Liu,Z.R. (2009) P68 RNA helicase is a nucleocytoplasmic shuttling protein. *Cell Res.*, **19**, 1388–1400.
46. Liu,Z.R., Sargueil,B. and Smith,C.W. (1998) Detection of a novel ATP-dependent cross-linked protein at the 5' splice site-U1 small nuclear RNA duplex by methylene blue-mediated photo-cross-linking. *Mol. Cell Biol.*, **18**, 6910–6920.
47. 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**, 864–867.
48. Day,J.W. and Ranum,L.P. (2005) RNA pathogenesis of the myotonic dystrophies. *Neuromuscul Disord*, **15**, 5–16.
49. Banroques,J., Cordin,O., Doere,M., Linder,P. and Tanner,N.K. (2008) A conserved phenylalanine of motif IV in superfamily 2 helicases is required for cooperative, ATP-dependent binding of RNA substrates in DEAD-box proteins. *Mol. Cell Biol.*, **28**, 3359–3371.
50. 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**, 5773–5780.
51. Honig,A., Auboeuf,D., Parker,M.M., O'Malley,B.W. and Berget,S.M. (2002) Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72. *Mol. Cell Biol.*, **22**, 5698–5707.
52. Charlet,B.N., Logan,P., Singh,G. and Cooper,T.A. (2002) Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. *Mol. Cell*, **9**, 649–658.
53. Kim,D.H., Langlois,M.A., Lee,K.B., Riggs,A.D., Puymirat,J. and Rossi,J.J. (2005) HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. *Nucleic Acids Res.*, **33**, 3866–3874.
54. Wilson,B.J. and Giguere,V. (2007) Identification of novel pathway partners of p68 and p72 RNA helicases through OncoPrint meta-analysis. *BMC Genomics*, **8**, 419.
55. Paul,S., Dansithong,W., Jog,S.P., Holt,L., Mittal,S., Brook,D., Morris,G.E., Comai,L. and Reddy,S. (2011) Expanded CUG repeats dysregulate RNA splicing by altering the stoichiometry of the muscleblind 1 complex. *J. Biol. Chem.*, **286**, 38427–38438.
56. Fukuda,T., Yamagata,K., Fujiyama,S., Matsumoto,T., Koshida,I., Yoshimura,K., Mihara,M., Naitou,M., Endoh,H., Nakamura,T. *et al.* (2007) DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat. Cell Biol.*, **9**, 604–611.
57. Junghans,R.P. (2009) Dystrophin myotonia: why focus on foci? *Eur. J. Hum. Genet.*, **17**, 543–553.
58. Ogilvie,V.C., Wilson,B.J., Nicol,S.M., Morrice,N.A., Saunders,L.R., Barber,G.N. and Fuller-Pace,F.V. (2003)

- The highly related DEAD box RNA helicases p68 and p72 exist as heterodimers in cells. *Nucleic Acids Res.*, **31**, 1470–1480.
59. 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**, 5474–5486.
 60. Kiliszek, 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**, 8370–8376.
 61. Warf, M.B., Diegel, J.V., von Hippel, P.H. and Berglund, J.A. (2009) The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc. Natl Acad. Sci. USA*, **106**, 9203–9208.
 62. Grammatikakis, I., Goo, Y.H., Echeverria, G.V. and Cooper, T.A. (2011) Identification of MBNL1 and MBNL3 domains required for splicing activation and repression. *Nucleic Acids Res.*, **39**, 2769–2780.
 63. 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.*, **286**, 34224–34233.
 64. Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S. and Wu, J.Y. (2011) RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol. Cell Biol.*, **31**, 1812–1821.
 65. Camats, M., Guil, S., Kokolo, M. and Bach-Elias, M. (2008) P68 RNA helicase (DDX5) alters activity of cis- and trans-acting factors of the alternative splicing of H-Ras. *PLoS One*, **3**, e2926.
 66. Uhlmann-Schiffler, H., Rossler, O.G. and Stahl, H. (2002) The mRNA of DEAD box protein p72 is alternatively translated into an 82-kDa RNA helicase. *J. Biol. Chem.*, **277**, 1066–1075.

Legends to supplementary figures

Supplementary Figure S1.

UAP56 and CDC5L do not co-localize with RNA foci. HeLa cells were co-transfected with GFP-UAP56 or GFP-CDC5L expression plasmids and minigenes containing the 3'UTR of *DMPK* gene with 960 CTG repeats (CUG960). Foci were detected by RNA-FISH using Cy3-labeled (CAG)₈ probes. Nuclei were stained with DAPI.

Supplementary Figure S2.

p68-myc colocalizes with CUG repeats in different cell lines. Cells were co-transfected with a myc-p68 expression vector and a minigene containing the 3'UTR of *DMPK* gene with 960 CTG repeats (CUG960). Expression of p68-myc was detected by immunofluorescence using anti-myc antibodies, followed by rabbit anti-mouse A488 antibodies. Foci were detected by RNA-FISH using Cy3-labeled (CAG)₈ probes. Nuclei were stained with DAPI.

Supplementary Figure S3

(A) Schematic representation of wild-type p68 and p68 Δ Ct2. 100 aminoacids from the C-terminus have been deleted. (B) p68-myc Δ Ct2 colocalizes with RNA foci. HeLa cells were co-transfected with p68-myc Δ Ct2 expression plasmid and minigene containing the 3'UTR of *DMPK* gene with 5 CTG (CUG5) or 960 CTG repeats (CUG960). Expression of p68-myc was detected by immunofluorescence using anti-myc antibodies, followed by rabbit anti-mouse A488 antibodies. Foci were detected by RNA-FISH using Cy3-labeled (CAG)₈ probes. Nuclei were stained with DAPI. (C) Overexpressed p68 Δ Ct2 protein has the same splicing activity on TNNT2 exon 5 inclusion as p68 wt. Human myoblast cells were cotransfected with the TNNT2 minigene and with expression vector coding for p68 or p68 Δ Ct2. The histogram shows the quantification of exon 5 inclusion. Results are derived from three experiments with error bars indicating standard deviation. ** $P < 0.01$

Supplementary Figure S4

p68 accelerates the binding of MBNL1 to 95 CUG repeats. (A) Time courses of MBNL1 binding to CUG repeats in the absence or presence of p68. Labeled (CUG)₉₅ RNA were incubated for indicated times with different amounts of recombinant MBNL1 protein with or without a constant amount of recombinant p68 Δ Ct2 (100ng) under splicing conditions with ATP. Proteins crosslinked to labeled RNA were separated on a 10% SDS-PAGE. The results are representative of three independent experiments. (B) Progress curves of MBNL1 binding for one experiment are shown. The radioactivity from labelled crosslinked band was quantified using a PhosphoImager. Open symbols show MBNL1 binding incubated without p68. Closed symbols show MBNL1 binding incubated with p68.

Supplementary Figure S5.

Gel shift experiments on (CUG)₁₄, (CAG)₁₆ and TNNT2 stem-loop structure element. In vitro transcribed (CUG)₁₄, (CAG)₁₆ and TNNT2 RNAs labeled with [α -³²P]CTP were incubated with the indicated amount of recombinant proteins. RNA-protein complexes were separated on native gel electrophoresis. Short CUG and CAG repeats are preferred to the long (CUG)₉₅ and (CAG)₆₂ repeats because RNA-MBNL1 complexes from the long repeats remain in the wells.

Supplementary Figure S6.

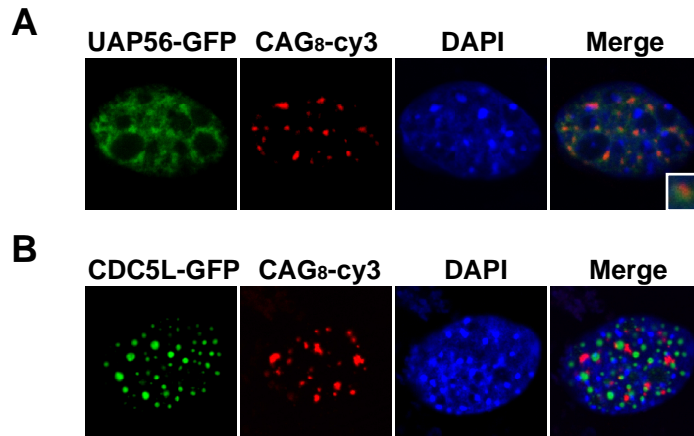
Several different siRNAs that target various regions of p68 or p72 have the same effect on alternative splicing of TNNT2 exon 5. Co-transfection of the TNNT2 minigene in human myoblasts treated with 50 nM of a control siRNA (lane 1), siRNA targeting different regions of p68 or p72 (lanes 3 to 5 and lanes 6 to 8, respectively), the siRNA targeting a common region of p68 and p72 (p68/p72, lane 2) or the siRNA targeting MBNL1 (lane 9). **(A)** Western blot analysis of cells treated with the different siRNAs. Blots were probed with antibodies against p68/DDX5, p72/DDX17, MBNL1 and emerin as a control. Note that polyclonal antibodies against DDX17 recognize the alternative polypeptide p82, which is the result of upstream translation initiation from a non-AUG codon (66). **(B)** RT-PCR analysis of TNNT2 exon 5 inclusion in cells treated with the different siRNAs. The results are from two independent experiments with error bars indicating standard deviation. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Supplementary Figure S7

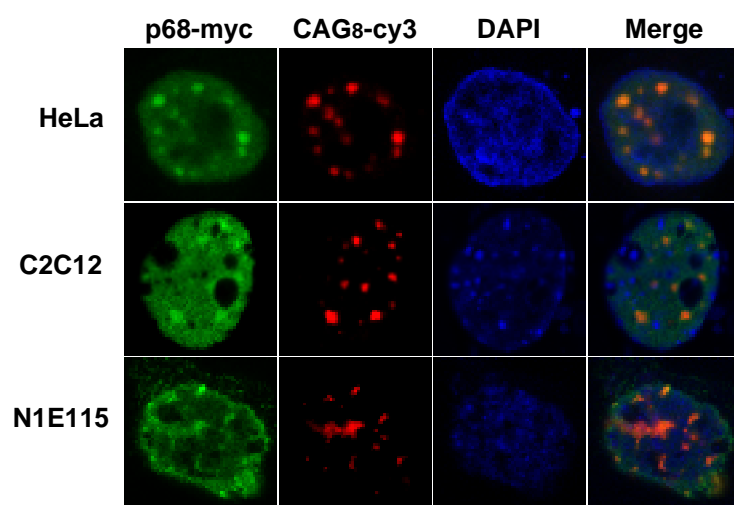
The splicing of TNNT2 exon 5 in DM1 myoblasts is unaffected by the knockdown of p68/p72. **(A)** RT-PCR analysis of endogenous transcripts in DM1 cells transfected with a control siRNA (siLuc) or siRNA against p68 and p72. **(B)** The histogram shows the quantification of exon 5 inclusion. Results are from two independent experiments. **(C)** Western blot analysis of cells treated with the different siRNAs. Blots were probed with antibodies against p68/DDX5, p72/DDX17 and β -actin as a control

66. Uhlmann-Schiffler, H., Rossler, O.G. and Stahl, H. (2002) The mRNA of DEAD box protein p72 is alternatively translated into an 82-kDa RNA helicase. *J Biol Chem*, **277**, 1066-1075.

Supplementary Figure S1

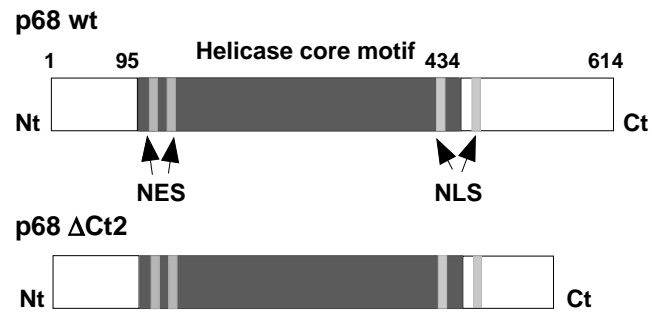


Supplementary Figure S2

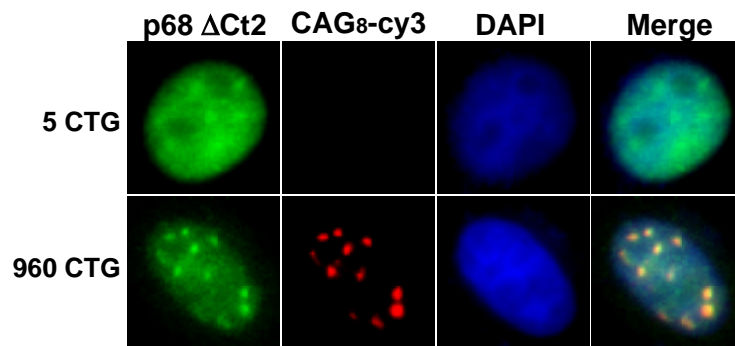


Supplementary Figure S3

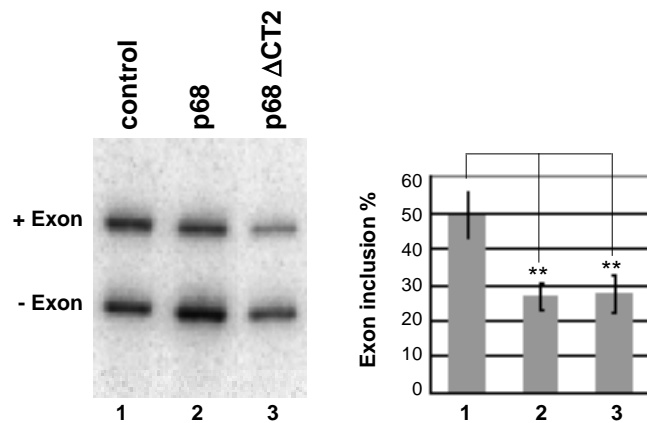
A



B

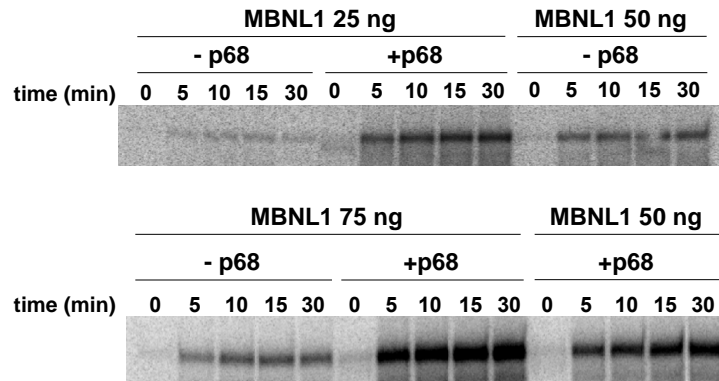


C

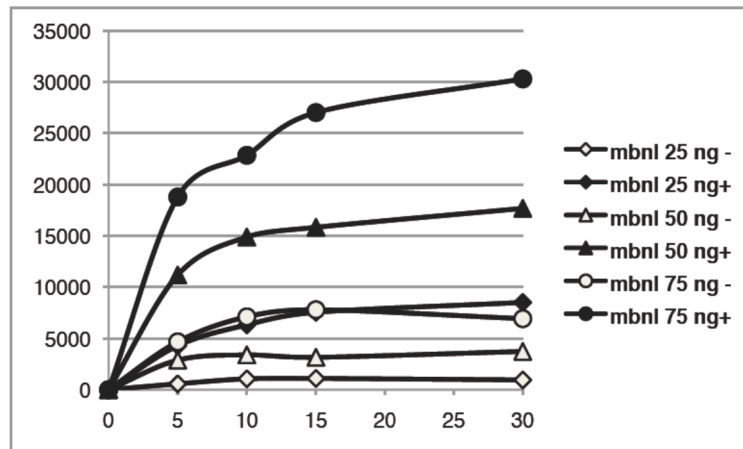


Supplementary Figure S4

A



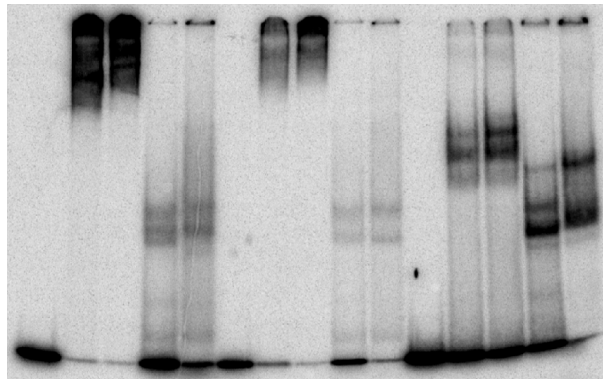
B



Supplementary Figure S5

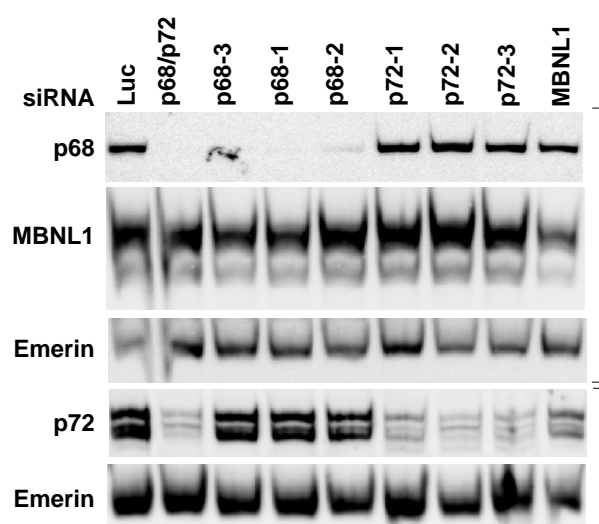
| | CUG14 | | | CAG16 | | | TNNT2 | | | | | | | | |
|-------------------------|-------|-----|-----|-------|-----|---|-------|-----|-----|-----|---|-----|-----|-----|-----|
| r P68 Δ Ct2 (ng) | - | - | - | 200 | 400 | - | - | - | 200 | 400 | - | - | - | 200 | 400 |
| r MBNL1 (ng) | - | 300 | 900 | - | - | - | 300 | 900 | - | - | - | 300 | 900 | - | - |

free RNA

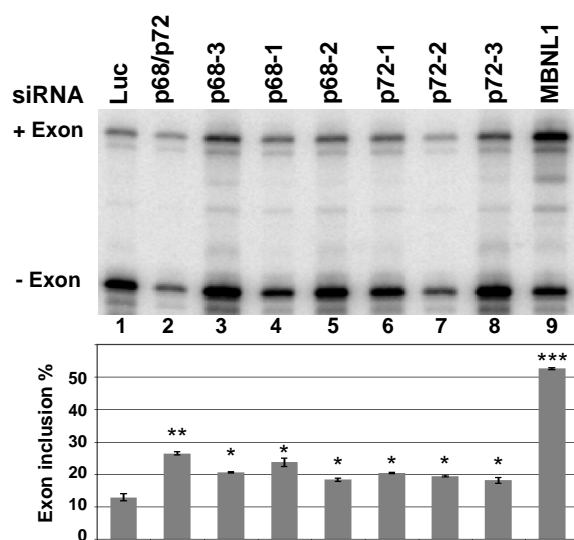


supplementary Figure S6

A

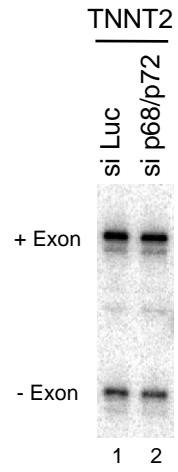


B

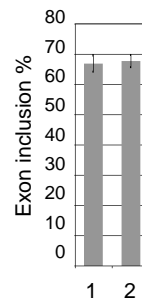


Supplementary Figure 7

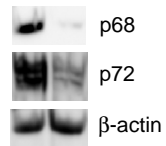
A



B



C



Legends to supplementary tables

Supplementary Table S1.

Proteins identified by mass-spectrometry. Proteins from the fourteen bands are indicated.

Supplementary Table S2.

Proteins tested for colocalization with CUG repeats by FISH/immunofluorescence. Note that some proteins not found in complexes assembled onto CUG repeats have been tested for colocalization because a member identified by mass-spectrometry has known interacting proteins (like proteins for the EJC complex).

Supplementary Table S3.

Sequences of siRNAs and oligonucleotides used for RT-PCR.

Supplementary Table S2

| Proteins tested by FISH/immunofluorescence | colocalization |
|---|----------------|
| AIMP1 | - |
| cullin 4A | - |
| CDC5L | - |
| DDX17 | + |
| DDX5 | + |
| DNA ligase III | - |
| EML3 | - |
| hnRNP M | - |
| Ifi204 | - |
| Magoh | - |
| MBNL | + |
| NF110c | - |
| NF110L | - |
| NF90c | - |
| NF90L | - |
| Nucleolin | - |
| p43 | - |
| PARP-1 | - |
| Pml Scl-100 | - |
| SC35 | - |
| UAP56/DDX39 | - |
| XRCC1 | - |
| Y14 | - |

Supplementary Table S3

SiRNA (Eurogentec)

| | |
|--------------|--------------------------|
| siRNA si Luc | CGUACGCGGAAUACUUCGAdTdT |
| sip68-3 | CCGCAACCAUUGACGCCAUdTdT |
| sip68/p72 | GGCUAGAUGUGGAAGAUGUdTdT |
| sip68-1 | UCAAUUAUGACUACCCUAAAdTdT |
| sip68-2 | AUGAAUUCAAACAUGGAAAdTdT |
| siMBNL1 | CACGGAAUGUAAAUUUGCAdTdT |
| sip72-1 | CAAGGGUACCGCCUAUACCDdTdT |
| sip72-2 | CAGCAGACUAAAUACAUUdTdT |
| sip72-3 | GAAAGAGAUUGGGUACUUAAdTdT |

Oligonucleotides against endogenous genes

TNNT2

| | |
|-----------|-----------------------------|
| sens | 5' AAGAGGTGGTGGGAAGAGTAC 3' |
| antisense | 5' AGCCTCCTTTGCTTCCTCTT 3' |

INSR

| | |
|-----------|----------------------------|
| sens | 5' AATGCTGCTCCTGTCCAAAG 3' |
| antisense | 5' ACCGTCACATTCCCAACATC 3' |

ATP2A1

| | |
|-----------|----------------------------|
| sens | 5'TGGTCCTCAAGATCTCACTG 3' |
| antisense | 5' TAAACAGACACATGACACGG 3' |

MBNL1

| | |
|-----------|----------------------------|
| sens | 5' TTGCAAGCCAAGATCAAGGC 3' |
| antisense | 5' CTGTTGTAAGTGCATGTTGG 3' |

Minigene

TNNT2 :

sens 5' CATTACCCACATTGGTGTGC 3'

antisense 5' AGGTGCTGCCGCCGGGCGGT 3'

RESULTATS COMPLEMENTAIRES

RESULTATS COMPLEMENTAIRES

Au cours de ma thèse, j'ai développé plusieurs axes de recherches qui n'ont pas été intégrés à l'article final que nous avons soumis. Ces résultats peuvent servir de compléments aux expériences décrites dans cette publication ainsi que proposer de nouvelles pistes de recherche et de nouveaux axes qu'il serait intéressant d'élucider.

A) Expression de p68 dans la DM1

Comme nous l'avons abordé dans l'introduction, la Dystrophie Myotonique de type 1 est caractérisée par la dérégulation des facteurs d'épissage MBNL1 et CUGBP1. L'équipe de Cooper a notamment montré que la quantité relative de la CUGBP1 était augmentée dans les cellules musculaires de patients DM1 et dans les tissus cardiaques de modèles de souris reproduisant la pathologie (Kuyumcu-Martinez *et al.*, 2007). Cette élévation est due à une hyperphosphorylation par la Protéine Kinase C (PKC) qui a pour effet d'augmenter la demi-vie de la protéine et ainsi sa concentration cellulaire. Par ailleurs, une étude montre que p68 est une cible de la PKC, notamment au niveau de l'extrémité C-terminale de la protéine (Rosenberger *et al.*, 2002) suggérant que p68 pourrait également être hyperphosphorylée dans la DM1. Afin d'étudier une possible altération de la quantité relative de p68 dans la pathologie, nous avons analysé l'expression protéique de p68 par western-blot sur des extraits protéiques totaux provenant de myoblastes contrôles et DM1 (Figure 35).

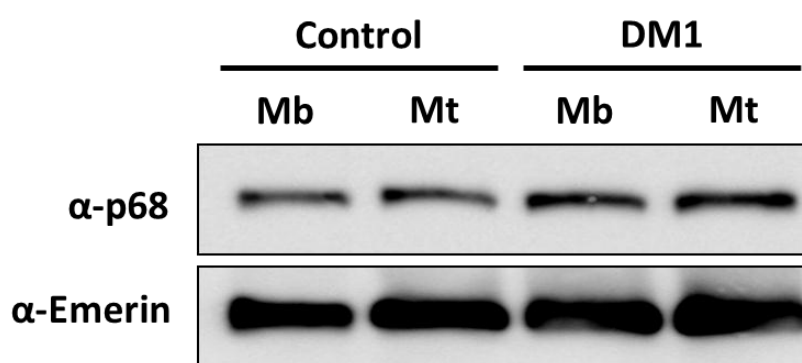


Figure 35 : Expression de la protéine p68 dans des extraits protéiques totaux de myoblastes et myotubes contrôles et DM1. Les pistes « Mb » correspondent à des myoblastes et les pistes « Mt » correspondent à des myotubes différenciés pendant 96h. la protéine p68 est révélée avec un anticorps monoclonal par western-blot. L'émerine est une protéine ubiquitaire localisée sous la membrane nucléaire utilisée comme contrôle de dépôt.

D'après nos résultats, il ne semble pas y avoir de différences significatives entre la quantité de p68 dans les myoblastes contrôles et dans les myoblastes DM1. Ces résultats sont également semblables dans des myotubes (différenciés 96h), suggérant que le retard de différenciation myogénique des cellules DM1 ne perturberait pas la quantité de p68.

D'après les études faites par Cooper, l'expression des expansions CTG960 augmente la phosphorylation de CUGBP1 par l'activation de la PKC. Dans ce même but, nous avons testé l'effet de la surexpression transitoire de CTG960 sur la quantité cellulaire de p68 par transfection transitoire dans des myoblastes humains contrôles. Les résultats n'ont pas montré de différences significatives de la quantité de p68 par western-blot, ni de changement de modifications post-traductionnelles (phosphorylation...) en gel à 2-dimensions (résultats non montrés).

En conclusion, nos résultats montrent que la colocalisation de p68 ne semble pas être à l'origine ou la conséquence d'une altération d'expression de p68 dans la pathologie. Ceci conforte un modèle selon lequel seule une fonction particulière (facilitation de la fixation de MBNL1 sur les transcrits) serait altérée, sans pour autant atteindre d'autres voies de signalisation.

B) Absence de colocalisation de p68 dans les cellules DM1 et tissus de souris SXL

Nos résultats indiquent que la protéine p68 colocalise avec les expansions CUG960 dans un modèle cellulaire. Ce modèle reproduit un certain nombre de défauts observés chez les patients DM1, tels qu'une dérégulation de l'épissage alternatif de plusieurs transcrits, un retard de différenciation ou encore l'hyperphosphorylation de la CUGBP1. Cependant, ce modèle ne semble pas être physiologique. En effet, l'expression du transgène, le nombre de foci ainsi que leur taille sont bien supérieurs à ceux observés chez les patients. Afin de tester la colocalisation de p68 dans un modèle plus physiologique, nous avons analysé la localisation cellulaire de p68 dans des cellules musculaires provenant de biopsies de patients (fournies par Denis Furling), ainsi que dans des coupes de tissus (cerveau et cœur) de souris DM1 SXL contenant de longues répétitions CTG (fournies par Geneviève Gourdon) (Gomes-Pereira *et al.*, 2007). Ces résultats sont présentés dans les figures 36 et 37 et montrent que p68 ne colocalise pas avec les foci CUG dans des myoblastes DM1 ou dans des tissus de souris DM1 SXL. Plusieurs hypothèses peuvent être envisagées pour expliquer ces résultats et seront exposées dans la discussion de la thèse.

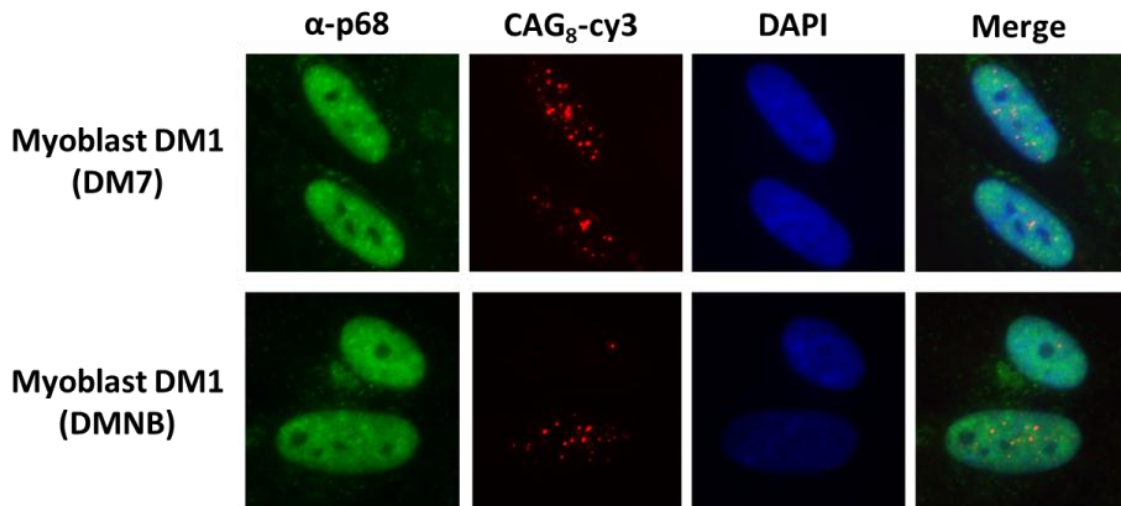


Figure 36 : Absence de colocalisation de p68 avec les foci CUG dans des myoblastes DM1. Deux types de cellules, provenant de patients DM1 et ayant un nombre de répétitions CTG différents ont été utilisés. p68, révélée en immunofluorescence avec un anticorps monoclonal ne colocalise pas dans ces deux types de cellules avec les foci CUG observés par RNA-FISH (rouge).

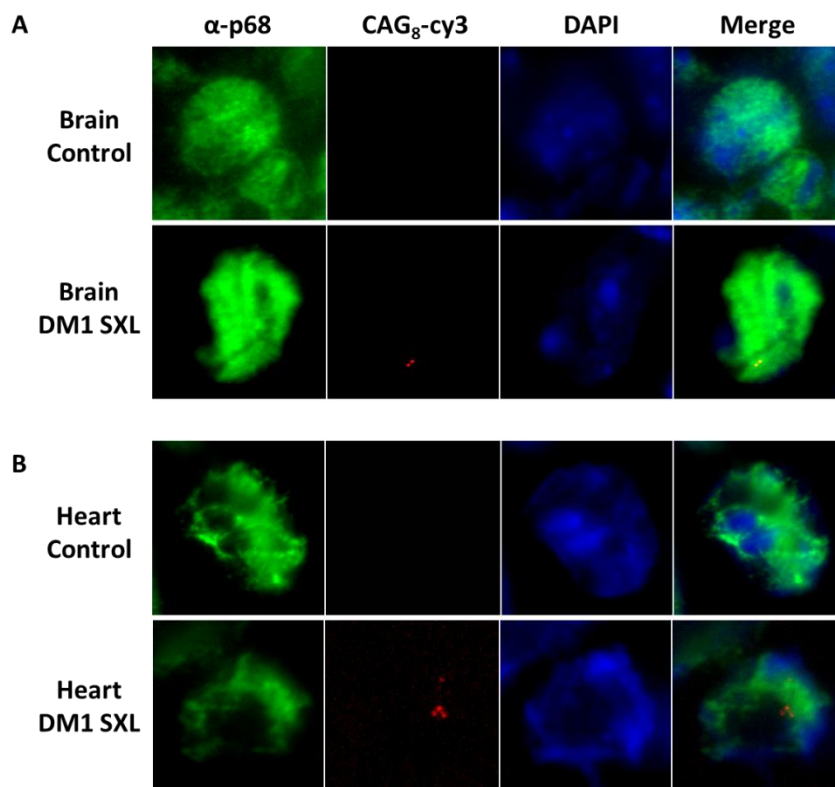


Figure 37 : Absence de colocalisation de p68 avec les foci CUG dans tissus provenant de souris DM1 SXL. La localisation de p68, révélée en immunofluorescence avec un anticorps monoclonal, n'est pas altérée entre les souris contrôles et les souris DM1 SXL. p68 ne colocalise pas avec les expansions CUG dans des coupes de cerveau (A) ou de cœur (B) provenant de modèles de souris DM1.

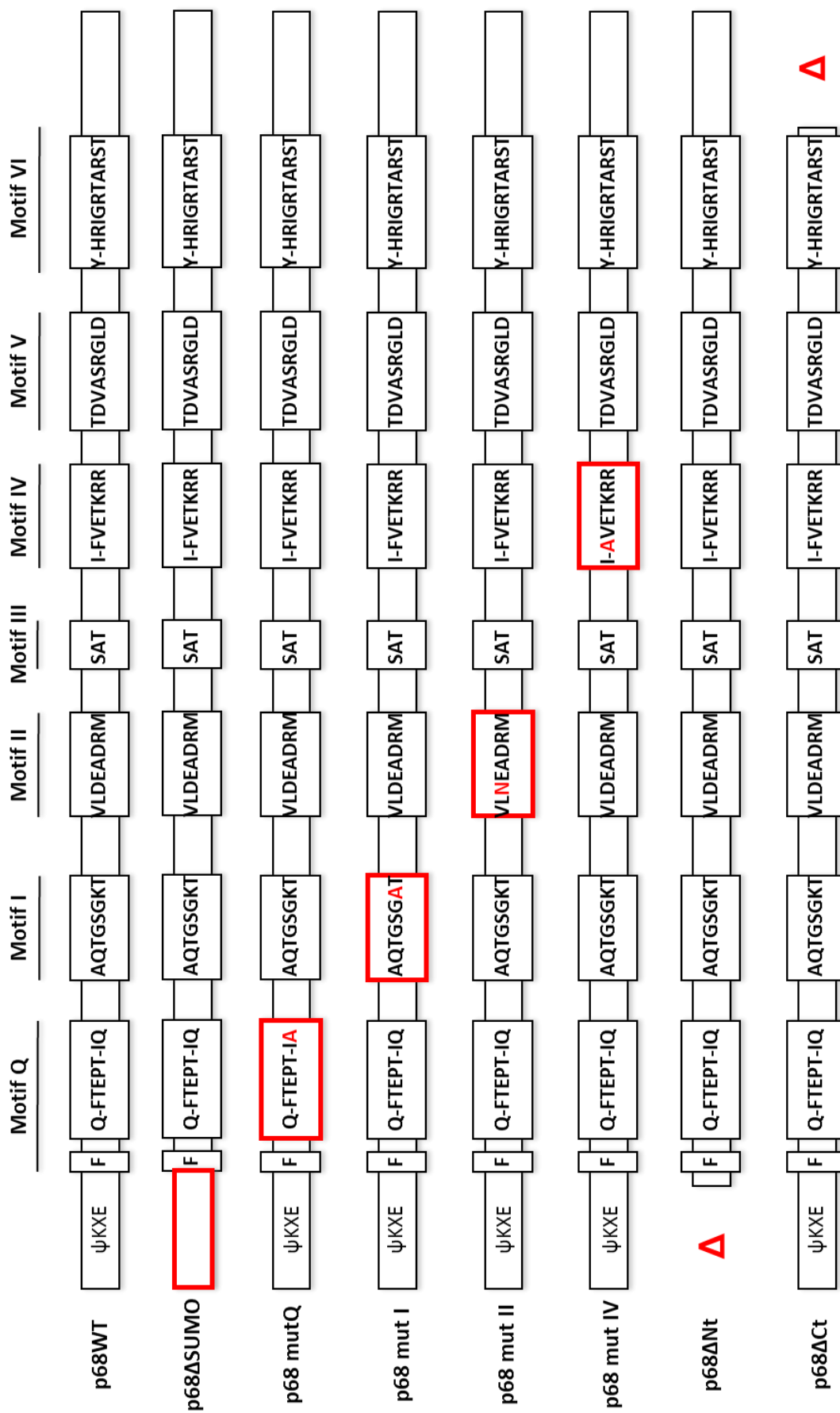


Figure 38 : Liste des différents mutants principaux de p68 utilisés au cours de ma thèse. Les séquences des différents motifs conservés sont indiquées et les mutations sont marquées en rouge. Le signe Δ rouge signifie une délétion. Le motif ψKXE correspond au motif consensus de sumoylation.

C) Analyse structure-fonction de p68

p68, comme tous les autres membres des protéines DEAD-box contient 9 motifs hautement conservés impliqués dans les fonctions des ARN hélicases (hydrolyse de l'ATP, domaine de fixation à l'ARN...). Afin d'identifier les domaines ou résidus requis pour la capacité qu'a p68 à colocaliser avec les expansions CUG, nous avons développé plusieurs mutants contenant des mutations ponctuelles (mutation d'un acide aminé par un autre) ou une délétion des extrémités N-terminale et C-terminale (Figure 38).

Dans l'article présenté dans la première partie des résultats, nous avons montré que la mutation d'un résidu dans les motifs II et IV abolit la colocalisation de p68 avec les expansions CUG, suggérant que l'activité hélicase soit requise pour l'interaction avec les répétitions CUG *in vivo*. Nous avons également montré que la mutation d'acides aminés dans les motifs Q et I réduit, voir abolit la colocalisation de p68 (Figure 39).

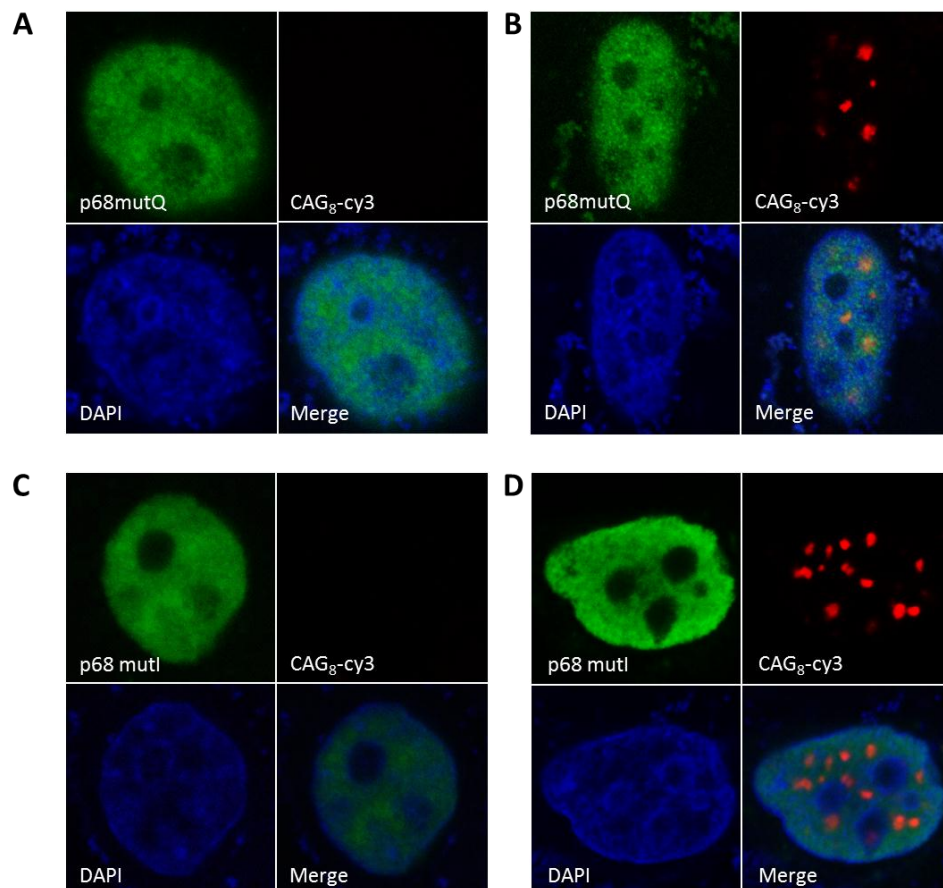


Figure 39 : La mutation d'acides aminés dans les motifs Q et I abolit la colocalisation de p68 avec les foci. Des cellules HeLa sont cotransfectées avec un plasmide exprimant p68 mutée et un plasmide exprimant 5CTG (A et C) ou 960CTG (B et D).

La délétion des extrémités N-terminale et C-terminale a également montré des résultats intéressants (Figure 40). La délétion de l'extrémité C-terminale altère la localisation nucléaire de la protéine, que ce soit en présence de CTG5 (Figure 40C) ou CTG960 (Figure 40D) ce qui semble être la conséquence de la suppression d'un signal de localisation nucléaire (Wang *et al.*, 2009). La délétion de l'extrémité N-terminale n'altère pas la localisation de la protéine en présence de CTG5 (Figure 40E) qui est identique à celle de la protéine sauvage (Figure 40A). De manière surprenante, la coexpression de CTG960 entraîne la formation d'agrégats qui ne colocalisent pas avec les foci mais se trouvent à leur périphérie.

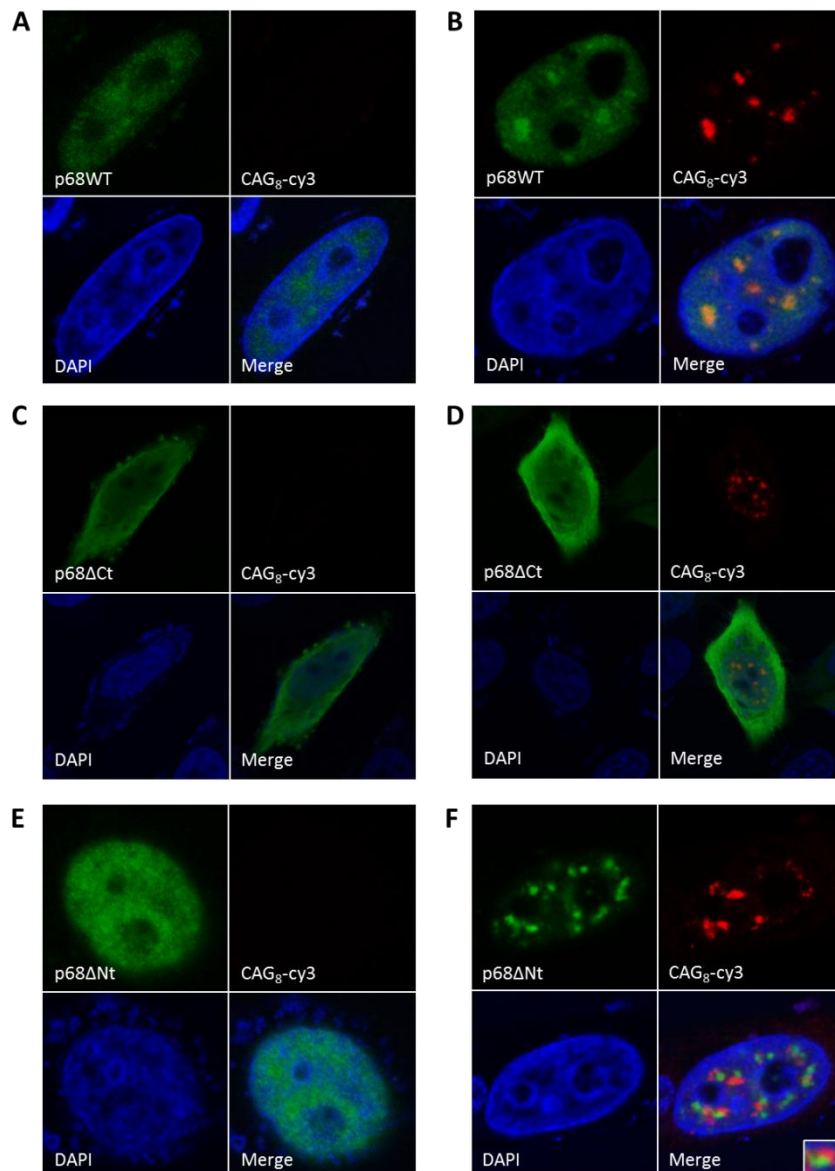


Figure 40 : La délétion des extrémités N-terminale et C-terminale de p68 altère la colocalisation avec les foci CUG. Les cellules sont cotransfectées avec p68WT (A et B), p68ΔCt (C et D), p68ΔNt (E et F), CTG5 (A, C et E) ou CTG960 (B, D, F).

Afin d'identifier avec plus de précision le domaine impliqué dans ce phénotype particulier, nous avons procédé à des délétions plus courtes. La délétion ΔNt (100aa) a été divisée en 4 délétions de taille équivalente (25aa) et la localisation de ces mutants en présence de répétitions CUG960 a été analysée (Figure 41).

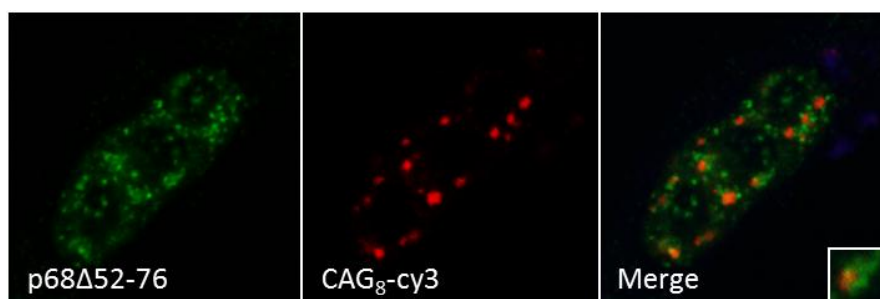


Figure 41 : La délétion de la région 52-76 suffit à reproduire la localisation de la protéine p68 dont la totalité de l'extrémité N-terminale est délétée. En présence de répétitions CUG960, p68 Δ 52-76 forme des agrégats localisés à la périphérie des foci CUG.

Seule la délétion des acides aminés 52 à 76 (noté p68 Δ 52-76) provoque le même type de localisation que la délétion N-terminale totale, suggérant qu'une région ou un acide aminé particulier serait requis pour la localisation au niveau des foci. Malheureusement, l'identité de cette séquence n'a pu être identifiée. En effet, des délétions de plus en plus réduites (8 acides aminés) au niveau de la zone 52-76 ne reproduisent pas le phénotype (résultats non montrés). De plus, la mutation ponctuelle de la lysine 53 identifiée dans le site consensus de sumoylation ψKXE ou la délétion complète de ce motif n'a aucun effet sur la colocalisation de p68 avec les foci CUG. De manière intéressante, la délétion de la partie N-terminale ou 52-76 abolit l'effet de p68 sur la régulation de l'épissage alternatif de *TNNT2* (résultats non montrés), alors que les délétions 1-24, 25-51 et 77-99 n'altèrent pas la fonction de p68. Ces résultats intrigants seront discutés plus en détail dans la discussion.

D) Interaction de p68 sur la tige-boucle de *TNNT2*

Afin de confirmer un rôle de p68 dans la facilitation de la fixation de MBNL1, nous avons vérifié par des expériences de gel retard que la protéine recombinante p68 Δ Ct2 interagit avec un ARN transcrit in vitro contenant la tige-boucle de *TNNT2*. En utilisant une concentration croissante de protéine recombinante, nous avons montré une diminution du signal correspondant à l'ARN libre au profit de la formation de complexes ARN-protéines de haut poids moléculaires (Figure 42).

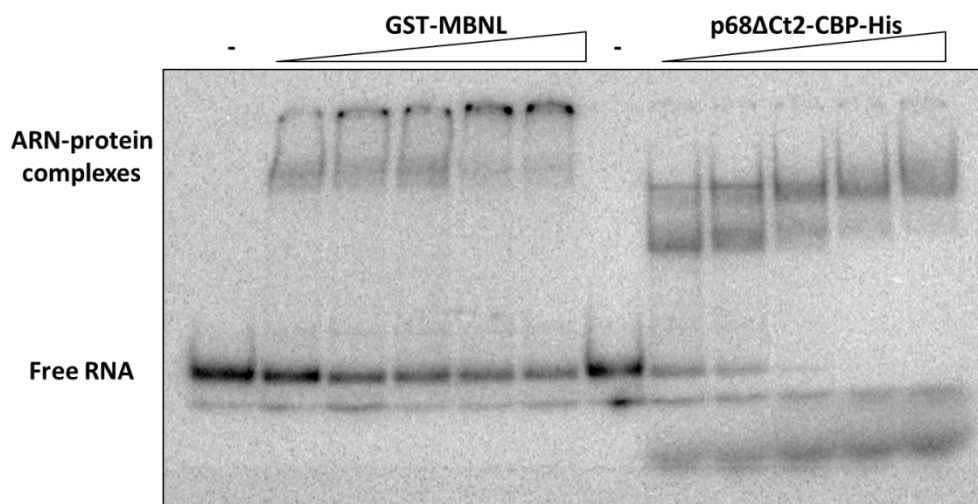


Figure 42 : *p68ΔCt2-CBP-His* et *GST-MBNL* interagissent avec la tige-boucle de *TNNT2* en gel retard. La gamme de *GST-MBNL* utilisée est 50, 250, 500, 750 et 1000ng. La gamme de *p68ΔCt2* utilisée est 25, 50, 100, 200 et 400ng.

Dans l'article principal, nous avons également testé l'effet de mutations affectant les sites de fixation à MBNL ou la structure secondaire de la tige boucle sur la régulation de l'épissage alternatif de *TNNT2* par p68. Nous avons montré que les mutations 4G, CUGS et GUF abolissent l'effet de p68 et réduisent drastiquement l'effet de MBNL1. Cependant, la perte d'effet pourrait être due au fait que p68 n'interagisse plus avec l'ARN à cause des mutations introduites. Nous avons donc analysé l'interaction de p68 avec un ARN composé de la tige-boucle de *TNNT2* avec et sans mutations (Figure 43).

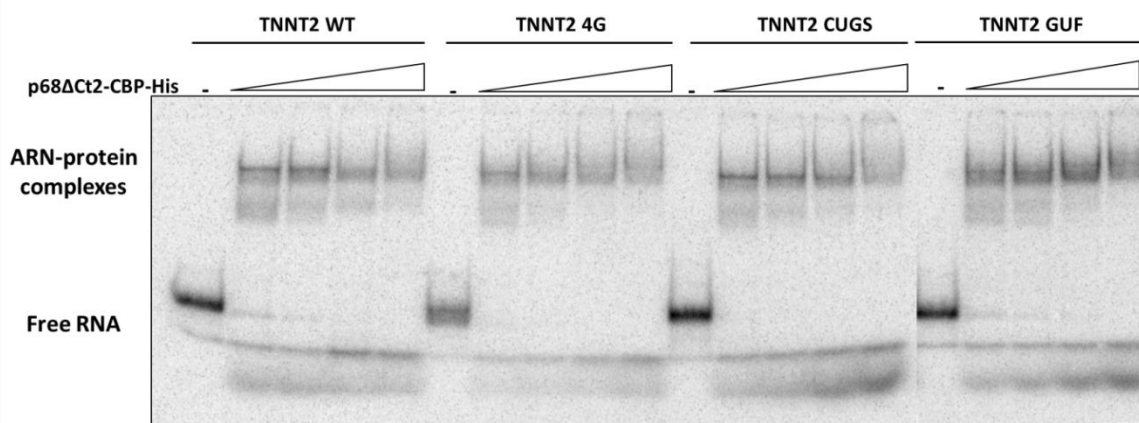


Figure 43 : *p68ΔCt2-CBP-His* semble interagir avec la tige-boucle de *TNNT2* sauvage ou mutante avec une affinité comparable. La gamme de *p68ΔCt2* utilisée est 50, 100, 200 et 400ng.

Ces résultats suggèrent que les mutations insérées dans la tige-boucle de *TNNT2* n'influent pas sur l'interaction avec p68. Il semble donc que p68 ne se fixe pas

spécifiquement au niveau des sites MBNL1, mais plutôt au niveau d'une structure, confirmant l'hypothèse développée dans l'article. Ce dernier point sera abordé dans la discussion.

E) Purification des partenaires protéiques de MBNL1

Le deuxième projet de ma thèse, outre l'identification des protéines interagissant avec les expansions CUG, consistait en la purification en tandem des partenaires protéiques de MBNL1. L'une des questions soulevée dans la communauté travaillant sur cette pathologie concerne le devenir des partenaires de ce facteur d'épissage. Sont-ils également séquestrés ? Sont-ils dérégulés ? Par cette approche, nous souhaitons identifier les différents complexes ribonucléoprotéiques contenant la protéine MBNL1 afin de cartographier leurs fonctions cellulaires et leurs rôles possibles dans la DM1.

Au cours de ma thèse, j'ai cloné le cDNA codant MBNL1 dans un vecteur d'expression eucaryote fusionné à deux types d'étiquettes (Figure 44A) : le TAP-tag (composé d'un peptide protéine A, un site de clivage TEV, et une étiquette Calmodulin-Binding Protein) et le HB-tag (6 histidines, un site de clivage TEV et une séquence de biotonylation *in vivo*). Ces deux étiquettes permettent la purification des partenaires par deux étapes de purification successives en utilisant chacune des étiquettes associées.

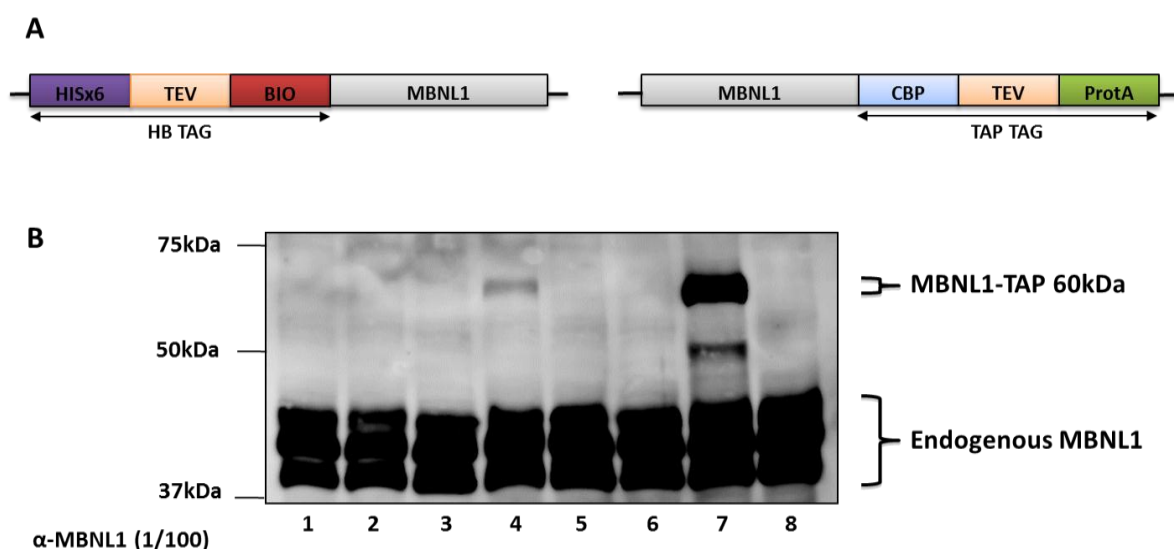


Figure 44 : Mise au point de la purification des partenaires protéiques de MBNL1. **A.** Constructions plasmidiques contenant le cDNA de MBNL1 avec un tag HB-tag ou TAP-tag. **B.** Expression de MBNL1 taguée dans des clones C2C12 stables (1-6), en transfection transitoire (7) et avec le vecteur vide (8).

Comme nous pouvons l'observer dans la Figure 44B, l'expression transitoire du plasmide fonctionne bien et permet l'expression d'une protéine MBNL1-TAP de 60kDa (ligne 7) absente du contrôle contenant le vecteur vide (ligne 8). Ce résultat était également identique pour la construction HB-MBNL1 (non montré).

Afin d'obtenir assez de matériel biologique pour visualiser les partenaires protéiques de MBNL1 sur gel SDS-PAGE, j'ai établi des clones stables de cellules murines C2C12 exprimant de façon stable la protéine MBNL1-TAP et HB-MBNL1. Comme observable sur la Figure 44 (ligne 1-6), les clones stables n'expriment que faiblement la protéine exogène.

L'absence d'expression pouvant être due à la trop longue expression de la protéine exogène au cours de l'établissement des lignées, nous avons opté pour un système inductible. Des cellules HeLa tet-on ont été transfectées avec un plasmide contenant les constructions ci-dessus sous le contrôle d'un promoteur inductible à la doxycycline, permettant une expression contrôlée et limitée dans le temps. Malheureusement, les clones stables issus de ces lignées présentent le même problème avec un taux d'expression très faible qui diminue au cours des passages cellulaires. Au bout de quelques mois, je me suis concentré sur la deuxième partie du projet et laissé inachevé cette partie.

Depuis quelques mois, une lignée stable 293T-Rex MBNL1-ProteinC a été établie au laboratoire en collaboration avec le Laboratoire d'Enzymologie et de Biologie Structurale (LEBS, Gif-sur-Yvette). Ces cellules contiennent une seule copie de la séquence codante de MBNL fusionnée à une étiquette histidine et protéine C, intégrée dans un site unique du génome et sous le contrôle d'un promoteur inductible, ce qui facilite grandement les étapes de sélections des clones stables. Des expériences pilotes ont été réalisées au laboratoire et ont permis l'identification d'un candidat intéressant, impliqué notamment dans le Non-sense mediated mRNA decay (NMD). Des analyses sont en cours pour déterminer le type d'interaction entre ce facteur et MBNL1 et son rôle potentiel dans la pathologie.

DISCUSSION

DISCUSSION

La Dystrophie Myotonique de type 1 est une maladie génétique caractérisée par une dérégulation de l'épissage alternatif de nombreux transcrits neuromusculaires avec la réexpression d'isoformes fœtales dans les tissus adultes. L'expression d'un ARNm DMPK mutant comportant une expansion de triplets CUG aboutit à la formation d'inclusions ribonucléoprotéiques nucléaires qui perturbent de nombreuses étapes du métabolisme de l'ARN (transcription, épissage, traduction, miRNA...). La séquestration de la famille de protéines Muscleblind par les expansions CUG semble être la cause principale de la dérégulation de l'épissage alternatif (Du *et al.*, 2010), mais plusieurs points restent obscurs...

Que connaît-on vraiment du mécanisme de la séquestration par les foci dans la DM1 ? En effet, la séquestration de MBNL1 ne semble pas être une condition déterminante pour expliquer la dérégulation de l'épissage alternatif. Le laboratoire de Tom Cooper a montré que les expansions de répétitions CAG pouvaient former des foci et séquestrer MBNL1, sans pour autant entraîner de dérégulation de l'épissage alternatif (Ho *et al.*, 2005), ces expériences ont par ailleurs été confirmées par l'absence de dérégulation de l'épissage alternatif des transcrits *TNNT2*, *Clc-1* et *MBNL1* dans des tissus musculaires de souris transgéniques contenant 200 répétitions CAG (Hsu *et al.*, 2011). De plus, la découverte de la Dystrophie Myotonique de Type 2, causée par l'expansion d'un quadruplet CCUG entraîne également la séquestration de MBNL1. Bien que le nombre de répétitions CCUG soit en moyenne près de 10 fois supérieur au nombre de CUG, ce qui devrait entraîner une séquestration plus importante et donc une perte d'activité de ce facteur plus drastique, les symptômes sont moins sévères (Liquori *et al.*, 2001). Enfin, d'autres facteurs semblent également interagir avec les expansions CUG (Jiang *et al.*, 2004) et plusieurs modèles récents suggèrent que ces expansions pourraient exister sous plusieurs formes : soluble (simple brin) ou insoluble (double-brin). Des facteurs encore inconnus pourraient donc être impliqués dans la pathologie DM1 et leur découverte permettrait une meilleure compréhension des bases moléculaires de la maladie.

C'est dans ce but que j'ai démarré mon projet de thèse, qui a porté sur l'identification de nouveaux candidats pouvant interagir avec les répétitions CUG.

A) Identification de nouvelles protéines interagissant avec les expansions CUG caractéristiques de la Dystrophie Myotonique de type 1

1) Affinité spécifique pour les expansions CUG

Nous avons montré, par des expériences de pontage aux UV, que le répertoire de protéines pouvant interagir avec les expansions de trinuécléotides était spécifique selon le type de répétitions utilisé. En effet, alors que MBNL1 semble être la protéine majoritaire pouvant être pontée sur les expansions CUG (Figure 1 de l'article), les protéines majoritaires se fixant sur les expansions CAG sont de poids moléculaires plus faibles. Ceci est en accord avec les données publiées dans différents laboratoires qui montrent que les différentes maladies à triplets ne séquestrent pas les mêmes facteurs (Wojciechowska & Krzyzosiak, 2011). Nous pouvons donc imaginer que certains phénotypes ou dérégulations caractéristiques d'une pathologie mais absentes dans d'autres maladies à triplets pourraient être reliés à la séquestration d'une ou plusieurs protéines spécifiques. Ceci est notamment le cas pour le syndrome du Fragile X-associated tremor/ataxia (FXTAS) dans lequel la protéine Sam68 est séquestrée par les répétitions CGG et non par les expansions CUG caractéristiques de la DM1 (Sellier *et al.*, 2010). Si la très grande majorité des maladies à triplets est à l'origine de symptômes neuromusculaires, on peut déceler des différences, comme la sévérité des symptômes ou le type de tissus dans lesquels ces défauts sont rencontrés. La séquestration d'un facteur particulier, comme c'est le cas pour FXTAS, entraîne des dérégulations de l'épissage alternatif qui ne sont pas rencontrées dans d'autres pathologies dues à des expansions de triplets (Sellier *et al.*, 2010). Ces observations démontrent que l'un des enjeux principaux pour la compréhension des mécanismes moléculaires dans ce type de pathologie passe par l'identification des complexes protéiques spécifiquement associés à ces expansions.

2) Affinité suivant le type cellulaire et de l'état de différenciation

Grace à des expériences de chromatographie d'affinité, en utilisant un ARN contenant 95 répétitions CUG comme appât, nous avons purifié les protéines ayant une affinité pour ce type d'expansions. Nous avons montré que le profil de protéines est spécifique du type cellulaire utilisé ; en effet des bandes sont spécifiques des extraits de cellules HeLa et d'autres bandes sont spécifiques des extraits myogéniques. De plus, il existe des différences de profil entre des extraits C2C12 myoblastes et myotubes. Ces résultats suggèrent ainsi que les

complexes protéiques interagissant ou étant séquestrés par les expansions CUG puissent être spécifiques suivant le contexte cellulaire. Ceci pourrait être à l'origine du développement de syndromes tissu-spécifiques. De plus, il a été montré que les cellules issues de biopsies de patients DM1 montrent des anomalies du métabolisme musculaire, parmi lesquels un retard de différenciation myogénique (Furling *et al.*, 2001). Ce retard pourrait s'expliquer par la séquestration de facteurs protéiques spécifiquement exprimés au cours de la différenciation.

L'identification à grande échelle des complexes ribonucléoprotéiques associés à tous les types d'expansions devrait permettre d'identifier les protéines spécifiques associées à ces répétitions afin de comprendre les bases moléculaires propres à chaque pathologie.

3) Identification et test des candidats putatifs

L'analyse par spectrométrie de masse des bandes issues de gel de purification a permis l'identification d'une centaine de protéines. Nous avons observé que la très grande majorité de ces protéines est impliquée dans le métabolisme de l'ARN (facteurs de transcription, facteur d'épissage, facteur d'export des ARNm, ARN hélicases...). Plusieurs de ces protéines possèdent des fonctions qui sont altérées dans la pathologie DM1. C'est le cas de la protéine DEAD-box UAP56 qui est impliquée dans l'export des ARN messagers et qui a été retrouvée dans les complexes assemblés sur les répétitions CUG. Sachant que l'ARN mutant DMPK n'est pas exporté dans le cytoplasme, il pourrait être intéressant d'étudier l'implication putative de cette protéine dans la pathologie. Certaines protéines, comme par exemple XRCC1, Ku70/80 ou Ligase III, toutes les trois impliquées dans la réparation de l'ADN pourraient être reliées avec les défauts de réparations de la voie MSH2/MSH3 observées dans les modèles murins (Tomé *et al.*, 2009).

Afin de tester leur possible intervention dans la pathologie, nous avons analysé la localisation des protéines candidates dans des cellules HeLa exprimant de façon transitoire la partie 3'UTR du gène DMPK humain contenant 960 répétitions CTG. Ce modèle cellulaire mime la formation des foci CUG, les défauts d'épissage alternatif et le retard de différenciation. C'est par ailleurs cette méthode qui a permis la découverte du facteur MBNL1 et sa séquestration dans la pathologie (Fardaei *et al.*, 2001). Nous avons montré que la très grande majorité des candidats testés ne colocalise pas avec les foci. Bien que ce crible soit extrêmement efficace, il ne permet pas d'exclure complètement la possibilité que ces protéines puissent jouer un rôle dans la pathologie. En effet, la CUGBP1 est impliquée dans la DM1 sans pour autant colocaliser avec les répétitions CUG (Jiang *et al.*, 2004). De même, le facteur de transcription SHARP, ne colocalise pas avec les foci mais se trouve délocalisé dans

le nucléoplasme (Dansithong *et al.*, 2011). Dans notre cas, à l'exception de p68 et p72, nous n'avons pas observé de changement de localisation pour les candidats testés en présence de répétitions CUG960. L'étude de la localisation a été effectuée après 24h de transfection, temps qui a été choisi afin de limiter la forte expression des expansions CUG qui pourrait à la longue laisser entrevoir un phénotype non physiologique. Cependant, nous ne pouvons pas exclure que la localisation de ces protéines puisse être altérée après 24h. En effet, l'équipe de Nicolas Charlet-Berguerand a montré que les expansions CGG recrutent progressivement différents facteurs au cours du temps (Sellier *et al.*, 2010). Si Sam68 est séquestrée dans les premières heures après la transfection, d'autres protéines, parmi lesquelles MBNL sont recrutées au bout de 72h. De plus, l'équipe montre que les foci augmentent en taille au cours du temps ce qui entraîne des phénotypes non physiologiques et l'apoptose (Sellier *et al.*, 2010). Si les foci CUG n'ont pas cette capacité d'expansion au cours du temps, il nous a semblé préférable de regarder la localisation des protéines candidates à un temps relativement court après la transfection.

Une modification de la localisation n'est pas le seul paramètre qui pourrait être altéré dans la DM1. Une altération de la quantité protéique, à l'instar de CUGBP1 pourrait intervenir dans certains symptômes de la pathologie. Par conséquent, il pourrait être intéressant d'analyser l'expression protéique par western-blot de ces candidats dans des myoblastes DM1 en comparaison avec des myoblastes contrôles.

B) Identification de p68 comme nouveau facteur impliqué dans la DM1 au niveau des expansions CUG

1) Colocalisation de p68 avec les répétitions CUG dans un modèle cellulaire mimant la Dystrophie Myotonique de type 1

La purification par chromatographie d'affinité d'extraits nucléaires sur des expansions CUG a permis l'identification des protéines DEAD-box ARN hélicases p68 et p72. Ces deux protéines appartiennent à la famille des DEAD-box et possèdent une activité hélicase permettant le réarrangement de complexes protéines sur l'ARN. Nous avons montré que les ARN hélicases p68 et p72 colocalisent avec les expansions CUG dans des cellules HeLa surexprimant la partie 3'UTR du gène DMPK contenant 960 répétitions CTG. En effet, la localisation très majoritairement nucléoplasmique de p68 dans des cellules contrôles surexprimant uniquement 5 répétitions CTG est altérée au profit d'une formation d'agrégats

nucléaires qui colocalisent avec les foci CUG observable par RNA-FISH. Les expansions CTG960 étant interrompues (le clonage de ces séquences ayant nécessité la présence de polylinker toutes les 30 répétitions), nous avons également montré que p68 colocalise avec des expansions CTG200 non interrompues, tout comme MBNL1. Nous avons par ailleurs observé que la protéine exogène p68 étiquetée HAmcHis ou la protéine endogène p68 réagissent de la même façon en présence de répétitions CUG, ce qui permet d'écartier tout biais quant à la surexpression de la protéine.

Nous avons également montré que p68 colocalise avec les répétitions CUG dans des cellules musculaires (C2C12) et neuronales (N1E115) murines. Ceci permet d'envisager que la colocalisation de p68 dans la DM1 pourrait avoir une conséquence pathologique au niveau du muscle et du cerveau.

2) Analyse structure-fonction de la colocalisation de p68

Nous avons observé que la mutation de résidus conservés dans les motifs Q, I et II, (impliqués dans la liaison et l'hydrolyse de l'ATP) et IV (impliqué dans la fixation à l'ARN) réduit ou abolit la colocalisation de p68 avec les foci. Ces résultats suggèrent que l'activité hélicase est requise pour la colocalisation avec les foci ex vivo.

La mutation du site de sumoylation, localisée au niveau de la lysine K53, ainsi que plusieurs sites de phosphorylation dans la partie N-terminale n'ont pas d'effet sur la colocalisation, suggérant, du moins pour ces sites, qu'une modification post-traductionnelle n'est pas requise pour l'interaction avec les foci.

La délétion de l'extrémité N-terminale, tout en conservant le core de l'hélicase, induit un phénotype très particulier. En effet, en absence de répétitions, p68 Δ Nt est localisée uniformément dans le nucléoplasme, tout comme la protéine sauvage. Cependant, en présence de répétitions CTG960, p68 Δ Nt forme des agrégats qui ne colocalisent plus avec les foci, mais qui au contraire sont localisés à la périphérie de ceux-ci, et qui sembleraient correspondre aux speckles nucléaires. La délétion de l'extrémité N-terminale étant assez importante (100 acides aminés) nous avons effectué de nouvelles constructions en réduisant la taille de la délétion. Nous avons observé que la délétion des résidus 52-76 est suffisante pour reproduire le phénotype. De plus, des délétions plus courtes dans cette zone ne reproduisent pas ce phénotype, suggérant que l'intégrité de cette séquence est requise pour la colocalisation. A part le site de sumoylation en position 53, aucun résidu ou motif conservé n'a pu être prédit par divers logiciels d'analyses bioinformatiques. L'importance de cette

région dans la localisation de p68 avec les foci reste donc mystérieuse. L'une des hypothèses serait que cette région corresponde à un site d'interaction avec une protéine intermédiaire qui permettrait l'adressage de p68 au niveau des foci. De manière intéressante, cette région 52-76 est totalement conservée dans p72 et pourrait de ce fait indiquer la présence d'un domaine important. Nous avons analysé la localisation d'une protéine fusion correspondant à la partie 52-76 de p68 fusionnée à la GFP. Cette courte protéine ne colocalise pas ni ne forme d'agrégats à la périphérie des foci, suggérant que ce domaine est requis mais pas suffisant pour la colocalisation avec les expansions CUG *in vivo*.

Enfin, la délétion de la partie C-terminale perturbe la localisation de la protéine qui se retrouve exclusivement dans le compartiment cytoplasmique. Ces constructions ont été effectuées avant une publication de l'équipe de Liu qui a identifiée deux signaux de localisation nucléaire (NLS) dont un se situe au niveau de la délétion, et qui sont requis pour la localisation correcte de la protéine (Wang *et al.*, 2009). Du fait de son altération en l'absence de répétitions, nous n'avons pas poursuivi l'étude sur cette forme mutée de p68.

3) **Absence de colocalisation de p68 dans des lignées cellulaires issues de biopsies de patients DM1, ni dans des tissus provenant de modèles de souris DM1**

Une grande partie des altérations d'épissage dans la pathologie DM1 est due à la séquestration du facteur d'épissage MBNL1 par les expansions CUG. MBNL1 colocalise avec les foci CUG des modèles cellulaires exprimant les répétitions CTG960, dans des lignées cellulaires musculaires issues de biopsies de patient et dans les tissus humains ou murins provenant de modèles de souris DM1. Sachant que p68 colocalise avec les expansions CUG960, nous avons étudié la localisation de p68 dans ces différents modèles. Nous avons montré par immunofluorescence que la localisation de p68 n'est pas altérée dans les lignées cellulaires DM1, ni dans des coupes de tissus de souris DM1. Plusieurs protocoles de fixation ou de démasquage des épitopes ont par ailleurs été testés, sans résultats. Ces observations suggèrent donc que p68 ne serait pas séquestrée dans la pathologie.

L'absence de séquestration de p68 dans les foci CUG soulève de nombreuses questions, notamment celle du rôle de p68 dans la pathologie. S'il on admet que la déplétion de MBNL1 due à sa séquestration dans les cellules DM1 est à l'origine de la dérégulation de l'épissage de plusieurs transcrits, le fait que p68 ne soit pas séquestrée suggère que p68 ne serait pas déplétée dans la cellule et qu'ainsi son activité ne serait globalement pas altérée

dans la DM1. Comment expliquer alors le fait que p68 colocalise dans ce modèle cellulaire surexprimant les expansions CUG960 ?

p68 et p72 sont toutes les deux impliquées dans de nombreuses voies de signalisation et fonctions cellulaires, comme nous l'avons abordé dans l'introduction. De plus, des analyses phénotypiques de souris dont les gènes codant p68 et p72 ont été inactivées montrent le développement d'un phénotype très sévère entraînant une létalité embryonnaire (Fukuda *et al.*, 2007). Il semble donc difficile d'imaginer une déplétion nucléaire de p68 dans la DM1 qui entraînerait probablement une non-viabilité des patients.

L'une des hypothèses que nous avons formulées serait que la colocalisation de p68 et p72 avec les foci CUG soit transitoire. Des résultats préliminaires obtenus au laboratoire montrent que la colocalisation de p68 serait maximale dans des temps courts après la transfection des répétitions CTG (18-24h) plutôt que dans des temps long (48-96h). Ceci suggère que p68 pourrait interagir avec les expansions CUG dans les premières heures de la formation des foci et être relarguée par la suite après avoir rempli une certaine fonction. Ceci pourrait expliquer l'absence de colocalisation dans des lignées cellulaires DM1 ou dans les modèles murins, qui pour leur part expriment continuellement l'ARN mutant. L'équipe de Denis Furling a récemment développé une lignée cellulaire exprimant des expansions CTG960 sous le contrôle d'un promoteur inductible. Il serait donc judicieux de tester la localisation de p68 et p72 dans ce modèle au cours du temps et des différentes étapes de la formation des foci.

Une autre hypothèse est que cette colocalisation pourrait être due au niveau d'expression des expansions CUG qui est relativement élevée lors de la transfection transitoire. D'une manière générale, les foci CUG dans des cellules transfectées de façon transitoire sont plus nombreux et plus larges en comparaison avec des lignées cellulaires de patient. On peut imaginer qu'il est plus facile d'observer un événement transitoire dans des cellules lorsque l'expression du transgène est maximale. Ainsi, une colocalisation de p68 avec les foci ne pourrait être observable que dans certaines conditions.

Enfin, la dernière hypothèse fait suite au modèle récemment proposé qui suggère que les expansions CUG pourraient exister sous deux formes : soluble et non soluble (Junghans, 2009). Les formes insolubles de l'ARN *DMPK* mutant pourraient être visualisées par RNA-FISH grâce à son agrégation nucléaire. Au contraire, les formes solubles pourraient être localisées de manière uniforme dans le nucléoplasme et interagir avec des complexes protéiques. Nous pouvons supposer que p68 aurait une affinité pour l'une de ces structures et que le ratio entre forme soluble et forme insoluble serait modifié suivant les lignées cellulaires

utilisées. De plus, nous pouvons aussi imaginer que p68 aurait un rôle dans la transition de ces formes, de par son activité hélicase, ce qui expliquerait sa colocalisation transitoire avec les foci CUG.

Nous avons montré que p68 interagit de façon directe avec les expansions CUG par expériences de retard sur gel et pontage aux UV. Ces expériences *in vitro* pourraient être renforcées par des expériences *ex vivo*. Outre les analyses de localisation cellulaire, nous pourrions utiliser des méthodes de RNA immunoprécipitation (RIP) avec un anticorps dirigé contre p68 et regarder si l'ARNm DMPK est associé. Ceci confirmerait de manière claire une interaction, aussi transitoire soit-elle, entre p68 et les expansions CUG.

C) p68 est un modificateur de l'interaction de MBNL1 avec les expansions CUG

1) p68 facilite l'interaction de MBNL1 sur les expansions CUG

L'étude cristallographique faite sur deux domaines à doigts de zinc de MBNL1 et un ARN de séquence CGCUGU montre que MBNL interagit avec le dinucléotide G-C en disposant ses domaines à doigts de zinc de façon antiparallèle, telle une pince (Teplova & Patel, 2008). Sachant que les expansions CUG sont stabilisées par des liaisons hydrogènes G-C entre les deux brins de la tige, les dinucléotides G-C sont engagés dans une liaison et ne sont donc pas ou peu accessibles pour MBNL1.

D'après les fonctions connues des DEAD-box ARN hélicases dans le remodelage des complexes ribonucléoprotéiques et les résultats obtenus au cours de ma thèse (montrant une colocalisation de p68 avec les foci CUG960 et une interaction avec les expansions CUG *in vitro*), nous avons proposé l'hypothèse selon laquelle p68 pourrait moduler la structure secondaire des expansions CUG pour faciliter la fixation de MBNL sur ses sites d'interaction.

Nous avons montré que p68 augmente la quantité de MBNL1 fixée sur un ARN contenant 95 répétitions CUG par pontage aux UV. De plus, nous avons observé que l'effet sur la fixation de MBNL1 est propre à p68; deux autres ARN hélicases testées (UAP56 et eIF4AIII) n'ont en effet pas cette capacité.

Nous avons montré que p68 est capable de faciliter d'une manière semblable la fixation de MBNL1 sur les expansions CAG et CUG. Cette observation est en accord avec une étude récente qui propose que les expansions CUG et CAG partageraient des similitudes

structurales (Kiliszek *et al.*, 2010). De plus, des analyses de FRAP ont montré que la GFP-MBNL1 possède une demi-vie de redistribution semblable après photo-blanchiment sur les foci CAG et CUG (Ho *et al.*, 2005) suggérant un même mécanisme de fixation de MBNL1 semblable entre ces deux structures. p68 n'augmente que faiblement la quantité de MBNL1 sur les expansions CCUG. Des analyses biochimiques ont montré que MBNL1 a une affinité deux fois plus forte pour les CCUG par rapport aux expansions CUG, alors que la structure des expansions CCUG est moins stable que les CUG (Warf & Berglund, 2007). Il semblerait donc que la stabilité de la structure secondaire soit déterminante quant à l'effet de p68 sur la fixation de MBNL1. Nous pouvons supposer que MBNL1 accède à ses sites d'interaction plus facilement sur une structure peu stable. A l'inverse, si une structure est stable, p68 pourrait être requise pour ouvrir localement la structure secondaire afin de faciliter l'accès de MBNL1 sur ses sites d'interaction.

2) **L'activité hélicase de p68 est requise pour faciliter la fixation de MBNL1**

La plupart des études effectuées sur les activités de remodelage des protéines DEAD-box ont montré que l'activité hélicase ATP-dépendante est requise pour l'ouverture locale de l'ARN. Nous avons montré que l'utilisation d'AMP-PNP, un analogue non-hydrolysable de l'ATP, réduit de moitié l'effet de p68 sur la fixation de MBNL1. De plus, la mutation ponctuelle de résidus conservés dans les motifs II et IV, qui abolissent l'activité ATPase et la fixation à l'ARN ATP-dépendante, réduit également de moitié l'effet de p68. Ceci est en accord avec la bibliographie, notamment avec une étude récente de l'équipe de Jane Wu qui montre que l'activité hélicase est requise pour le remodelage d'une structure secondaire sur l'ARN pré-messager codant la protéine Tau (Kar *et al.*, 2011). De plus, comme nous l'avons vu précédemment, une mutation dans ces deux domaines abolit la colocalisation de p68 avec les expansions CUG960. Nos résultats suggèrent donc que l'activité hélicase de p68 est requise pour moduler la structure secondaire des expansions CUG afin de faciliter la fixation de MBNL1. Cependant, des analyses supplémentaires doivent être effectuées pour démontrer de façon claire l'effet de p68 sur la structure. Nous pourrions effectuer des empreintes à ARN sur des expansions CUG en présence et absence de p68 afin d'observer de possibles changements structuraux, soit par des méthodes classiques de « footprint », soit par de nouvelles techniques comme SHAPE (Selective 2'-hydroxyacylation analyzed by primer extension) qui permet une analyse quantitative de la structure de l'ARN avec une résolution au nucléotide près (Wilkinson *et al.*, 2006).

3) Quel est le mécanisme d'action de p68 sur la fixation de MBNL1 ?

Plusieurs questions restent à éclaircir, notamment le mécanisme d'action de p68 et son rôle dans la séquestration de MBNL1. Il semble clair que p68, via son activité hélicase, puisse se fixer de façon directe aux expansions CUG. Maintenant, quel pourrait être l'enchaînement des étapes qui conduisent à une augmentation de la fixation de MBNL1 ? Plusieurs hypothèses peuvent être proposées :

- p68 et MBNL1 sont dans des complexes séparés. p68 pourrait cibler les expansions CUG et modifier leur structure secondaire. MBNL1 pourrait ensuite être recruté au niveau de ses sites d'interaction. Le changement conformationnel dû à la fixation sur l'ARN pourrait entraîner une perte d'affinité de p68 pour l'ARN qui serait alors relargué. Sachant que p68 ne colocalise pas avec les foci CUG dans les cellules de patients DM1, il semblerait que p68 se détache des foci une fois son action terminée.

- p68 et MBNL1 pourraient appartenir à un même complexe protéique. D'après une récente étude Oncomine, *MBNL1* est le deuxième gène le plus souvent co-exprimé avec *DDX5* et le treizième avec *DDX17* (Wilson & Giguère, 2007), suggérant que p68, p72 et MBNL1 pourraient appartenir à une voie de signalisation commune. Cependant, de nombreuses expériences de co-immunoprécipitation, de GST-pull down ou d'interaction de protéines recombinantes traduites in-vitro ont été effectuées au laboratoire et indiquent uniquement une très faible interaction entre p68 et MBNL1. Le fait qu'une interaction directe entre p68 et MBNL1 ne semble exister ne signifie pas pour autant que les deux protéines ne puissent exister dans un complexe commun. En faveur de cette hypothèse, la purification des partenaires protéiques de MBNL1 au laboratoire à partir d'une lignée stable de cellules 293T-rex a permis l'identification de p68. De plus, plusieurs partenaires intermédiaires entre MBNL1 et p68 pourraient intervenir. Selon cette hypothèse, p68 pourrait cibler les expansions CUG et moduler leur structure secondaire. Le fait que MBNL1 soit stériquement proche de ses sites, étant dans un même complexe protéique que p68 lui permettrait d'être immédiatement recruté et ainsi séquestré. p68 pourrait être ensuite relarguée de l'ARNm pour remplir d'autres fonctions.

Au sujet de la séquestration de MBNL1, quel est le rôle exact de p68 ? Plusieurs expériences ont été effectuées au laboratoire pour tenter de répondre à cette question. Nous avons tout d'abord inactivé conjointement p68 et p72 dans des cellules DM1 et observé le nombre de foci. Aucune différence significative sur le nombre de foci par rapport à des

cellules DM1 non inactivées n'a pu être observée, ce qui suggère que p68 n'est pas une protéine requise pour l'intégrité des foci. D'autres analyses sont en cours au laboratoire pour déterminer si p68 est impliqué dans le recrutement in vivo de MBNL1 au foci CUG par des méthodes de FRAP en inactivant p68/p72. Nous testons également si l'inactivation de p68/p72 entraîne une diminution de la colocalisation de MBNL1-GFP et MBNL1 endogène dans les foci CUG960 par RNA-FISH.

Les conclusions de ces études permettront de confirmer un rôle in vivo de p68 dans le recrutement de MBNL1 au niveau des expansions CUG. Si tel est le cas, p68 pourrait être la première protéine identifiée comme modificatrice d'inclusion ribonucléoprotéiques ou foci dans des maladies à triplets. Son implication dans la DM1 pourrait être un élément aggravant de la séquestration de MBNL1 et ce qui pourrait constituer une piste intéressante pour le développement d'outils thérapeutiques.

D) p68 régule l'épissage alternatif de l'exon 5 du transcrit *TNNT2*, dérégulé dans la DM1

1) p68 régule l'épissage alternatif d'un certain type de transcrits

Comme nous l'avons précédemment abordé, la Dystrophie Myotonique de type 1 est caractérisée par une dérégulation de l'épissage alternatif de plusieurs transcrits. Cette dérégulation semble principalement due à la séquestration de MBNL1 par les expansions CUG, puisque l'inactivation de MBNL1 dans un modèle murin récapitule plus de 80% des défauts d'épissage observés dans les tissus musculaires de souris DM1 (Du *et al.*, 2010). Comme nous l'avons abordé dans l'introduction, l'une des fonctions de p68 et p72 est de réguler certains événements d'épissage alternatif par l'ouverture de structures secondaires permettant l'accès de facteurs d'épissage à leur site d'interaction (Hönig *et al.*, 2002; Camats *et al.*, 2008; Kar *et al.*, 2011). Les résultats obtenus au cours de ma thèse suggérant une coopérativité entre MBNL1 et p68 nous ont conduit à proposer que p68 pourrait réguler l'épissage alternatif de gènes régulés par MBNL1 et dérégulés dans la pathologie DM1. Quatre transcrits dont l'épissage alternatif est dérégulé dans la DM1 ont été testés. Nous avons observé que l'inactivation conjointe ou séparée des protéines p68 et p72 par siARN dans des myoblastes humains active l'exclusion de l'exon alternatif 5 de l'ARN pré-messager codant la troponine-T cardiaque (*TNNT2*). En revanche, l'inactivation de p68/p72 n'a pas d'effet sur l'épissage alternatif des ARN pré-messagers *INSR* (exon 11), *SERCA* (exon 22) et

MBNL1 (exon 7). Ces expériences ont été confirmées par la cotransfection de minigènes codant l'exon alternatif entouré de deux exons constitutifs adjacents et d'un vecteur exprimant p68. Seul l'épissage alternatif de *TNNT2* est régulé par p68. Ces résultats suggèrent donc que p68 et p72 pourraient réguler l'épissage alternatif d'une certaine classe de transcrits régulés par MBNL1.

2) **p68 facilite la fixation de MBNL1 sur des sites intégrés dans une structure secondaire de type tige-boucle**

Quels déterminants pourraient expliquer cette différence de sensibilité à p68 ? Ces transcrits contiennent de multiples sites putatifs ou confirmés biochimiquement pour MBNL1 (Goers *et al.*, 2010). Ces sites sont localisés avant ou après l'exon alternatif et il ne semble pas y avoir de corrélation entre le fait que MBNL1 se comporte comme un activateur ou un répresseur de l'épissage. En effet, MBNL1 agit en tant que répresseur pour les transcrits *TNNT2* et MBNL1 et comme un activateur pour INSR et SERCA. Ces données suggèrent qu'un ou plusieurs éléments, autre que la position du site d'interaction à MBNL1, rentrent en jeu dans la régulation de l'épissage par p68.

Dans une étude récente, l'équipe de Berglund suggère qu'il existe différents modes de reconnaissance de MBNL1 pour ses cibles, suivant la conformation simple brin ou double-brin de ses sites (Cass *et al.*, 2011). Ainsi l'une des hypothèses que nous proposons est que la particularité de *TNNT2* pourrait provenir de la conformation et l'architecture des sites d'interaction à MBNL1 dans l'ARN pré-messager. En effet, les sites d'interaction de MBNL1 en amont de l'exon alternatif de *TNNT2* sont intégrés dans une structure secondaire de type tige-boucle contenant des mésappariements pyrimidines-pyrimidines similaires à ceux contenus dans les expansions CUG, suggérant que MBNL1 reconnaît ces structures par un mécanisme commun (Warf & Berglund, 2007). D'autre part, l'équipe de Cooper a récemment montré que les sites d'interaction de MBNL1 sur le transcrit INSR ne sont pas structurés (Grammatikakis *et al.*, 2011) et aucune étude ne fait état d'une structuration particulière au niveau des sites d'interaction de MBNL1 sur les transcrits SERCA et MBNL1. En accord avec l'hypothèse que nous proposons, nous avons montré par pontage aux UV que p68 augmente la fixation de MBNL1 sur la tige-boucle de *TNNT2*, de la même manière que sur les expansions CUG. Ce résultat suggère que p68 pourrait modifier la structure secondaire de la tige-boucle contenant les sites d'interaction de MBNL1 afin de les rendre plus accessibles.

3) L'effet de p68 sur l'épissage de *TNNT2* est dépendant de MBNL1

Bien que de nombreux arguments suggèrent que les DEAD-box ARN hélicases régulent l'épissage alternatif, ces mêmes données proposent également que cette famille de protéines ne se comporte pas comme un facteur d'épissage « classique ». L'hypothèse actuelle propose que certaines protéines DEAD-Box, de par leur activité hélicase pourrait assembler et désassembler les différents complexes régulateurs de l'épissage en modulant les structures secondaires aux environs des exons alternatifs. Bien que p68 favorise l'accès de la snRNP U1 sur le site 5' d'épissage dans le transcrit Tau (Kar *et al.*, 2011) et bloque l'interaction de hnRNP H avec ses sites sur le transcrit h-Ras (Camats *et al.*, 2008), aucune étude n'a démontré à l'heure actuelle qu'une ARN hélicase puisse faciliter la fixation d'un facteur d'épissage alternatif spécifique. Mes travaux rapportent le premier exemple suggérant que les ARN hélicases agissent comme cofacteur ou co-activateur de l'activité d'épissage d'un facteur particulier. Les arguments en faveur de cette hypothèse sont les suivants : (1) des mutations ponctuelles qui altèrent la stabilité de la structure secondaire de la tige-boucle en amont de l'exon 5 ou l'intégrité des sites d'interaction de MBNL1 abolissent la capacité de régulation de p68 sur l'épissage alternatif de *TNNT2*. Ces mutations n'empêchent pas l'interaction de p68 sur ces séquences, comme illustré dans la figure 42 par des expériences de retard sur gel, signifiant que l'abolition de la régulation par p68 n'est pas due à une incapacité de p68 à se fixer au niveau de ces structures secondaires. (2) la cotransfection du plasmide codant 960 répétitions CTG qui séquestre MBNL1, ou l'inactivation de MBNL1 par siRNA suffit à réduire de façon drastique l'effet de p68. Ces résultats indiquent que l'activité de p68 est dépendante de celle de MBNL1 et que p68 est un co-activateur de MBNL1 dans la régulation de l'épissage alternatif de l'exon 5 de *TNNT2*.

Dans son modèle proposé en 2007, Berglund propose que la fixation de MBNL1 sur la structure secondaire de *TNNT2* stabiliserait cette tige-boucle ce qui inhiberait la reconnaissance du site 3' d'épissage par le spliceosome et résulterait en une exclusion de l'exon alternatif (Warf & Berglund, 2007). D'après ce modèle, en l'absence de formation de la tige-boucle, MBNL n'est pas présent et la machinerie reconnaît le site 3' d'épissage et l'exon est inclus (Warf & Berglund, 2007). Il existe cependant un point critique concernant le fait que cette région puisse être ou non structurée suivant les conditions cellulaires. Il me semble peu probable qu'une structure secondaire d'un ARN puisse se déformer sans l'intervention d'une ou plusieurs protéines. D'après nos résultats, nous pouvons intégrer le

rôle de la protéine p68 dans le modèle de régulation de l'épissage alternatif de *TNNT2* proposé par Berglund (Warf & Berglund, 2007). Ce modèle est présenté en Figure 45.

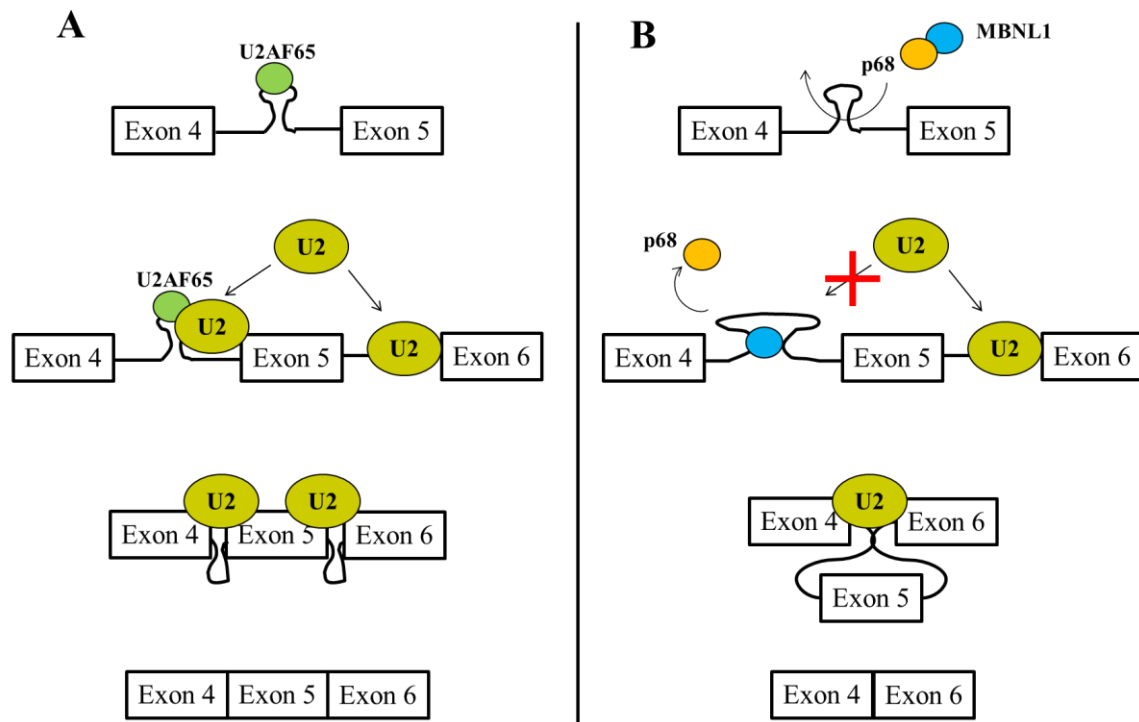


Figure 45 : *Modèle proposé par Berglund (Warf et al., 2009) revisité avec l'intervention de p68. A. Mécanisme de régulation pour l'inclusion de l'exon alternatif 5. U2AF65 interagit avec la région riche en pyrimidines localisée dans la boucle de la structure secondaire. Elle permet le recrutement de la snRNP U2 au niveau du site 3' d'épissage. L'intervention du reste du spliceosome permet l'inclusion de l'exon alternatif dans l'ARN messenger. B. Le complexe p68/MBNL1 interagit avec la tige-boucle contenant des sites régulateurs de MBNL1. p68 ouvre localement la structure secondaire pour permettre la fixation de MBNL1, ce qui empêche la fixation de U2AF65. La snRNP U2 ne peut se fixer et l'exon alternatif n'est pas reconnu par le spliceosome et sera épissé.*

E) Modèle de régulation

Au cours de ma thèse, nous avons identifié une nouvelle fonction pour la DEAD-box ARN hélicase p68 qui consiste à faciliter la fixation du facteur d'épissage alternatif MBNL1 sur deux types de transcrits : une cible d'épissage alternatif (*TNNT2*) et un ARN non-codant pathologique (DMPK) à l'origine de la Dystrophie Myotonique de type 1. Sachant que dans la pathologie, MBNL1 est séquestré par les répétitions et n'est donc plus accessible pour réguler l'épissage alternatif de certains transcrits, quel pourrait être le rôle de p68 dans la DM1 ?

D'après nos résultats, la fonction de p68 ne semble pas être altérée de manière globale dans la pathologie DM1 :

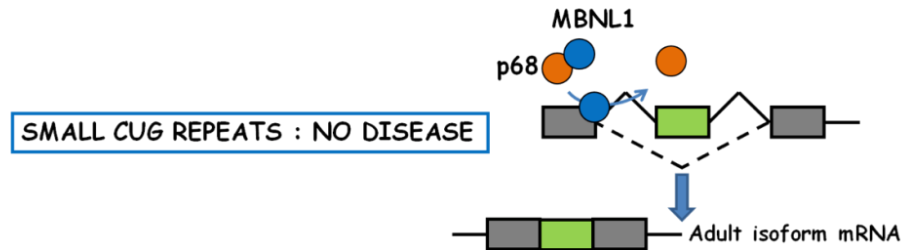
- 1) Il n'existe pas de différences significatives d'expression de p68 entre des myoblastes contrôles et des myoblastes DM1 en western blot.
- 2) Plusieurs autres processus (transcription, traduction...) impliquant p68 n'ont pas été reportés comme étant dérégulés dans la pathologie. Il faut toutefois être prudent sachant que tous ces processus n'ont pas été testés.
- 3) Enfin, l'inactivation de p68 et p72 dans des souris KO provoque une létalité embryonnaire, incompatible avec la survie à l'état adulte des patients atteints de DM1.

La dérégulation de p68 dans la DM1 pourrait donc être plus subtile, et affecter uniquement certaines de ces fonctions. D'après nos résultats et les données de la littérature, le rôle de p68 est multiple. p68 appartient à plusieurs complexes ribonucléoprotéiques, chacun affilié à une fonction. S'il pourrait exister un complexe « A » p68/MyoD qui serait impliqué dans la régulation transcriptionnelle de gènes musculaires (probablement par la modulation de la structure secondaire de promoteur contenant des sites d'interaction de MyoD), un complexe « B » p68/Drosha/DGCR8 impliqué dans le processing des microARN (probablement par la modulation des interactions intramoléculaires au niveau du pré-miRNA, comme suggéré dans un article récent (Salzman *et al.*, 2007)), il pourrait exister un complexe « C » composé d'au moins p68 et MBNL1.

Si l'on raisonne en termes de complexes ribonucléoprotéiques, nous pouvons suggérer qu'un complexe p68/MBNL1 pourrait avoir plusieurs cibles. Il pourrait en effet permettre de moduler les structures secondaires stables qui masquent les sites d'interaction de MBNL1 sur de nombreux transcrits, parmi lesquels *TNNT2*. Dans une cellule saine, le complexe p68/MBNL1 serait ainsi actif et régulerait de manière fonctionnelle l'épissage alternatif de nombreuses cibles. Par ailleurs, les cibles ne possédant pas de structures secondaires fortes seraient régulées par un complexe composé d'au moins MBNL1 en l'absence de p68. Dans des cellules de patients atteints de DM1, on pourrait imaginer que seul le complexe p68/MBNL1 serait altéré. A cause du nombre élevé de sites d'interaction de MBNL1 potentiels sur les expansions CUG contenues dans le transcrit mutant DMPK, en comparaison avec les quelques sites localisés en amont ou aval des exons alternatifs, nous pourrions imaginer que l'équilibre serait déplacé vers la séquestration de MBNL1 sur les répétitions CUG au détriment de la modulation des structures secondaires des sites d'interaction de MBNL1 sur les ARN pré-messagers. p68 pourrait alors être un facteur « aggravant » de la

dérégulation de MBNL1, et a fortiori un nouveau facteur impliqué dans les mécanismes moléculaires altérés dans la Dystrophie Myotonique de Type 1. Un modèle récapitulatif est présenté en Figure 46.

p68 facilite la fixation de MBNL1 sur des transcrits par la modification de structures secondaires pour réguler l'épissage alternatif



p68 facilite l'interaction de MBNL1 avec les expansions CUG entrainant sa séquestration dans la Dystrophie Myotonique de type 1

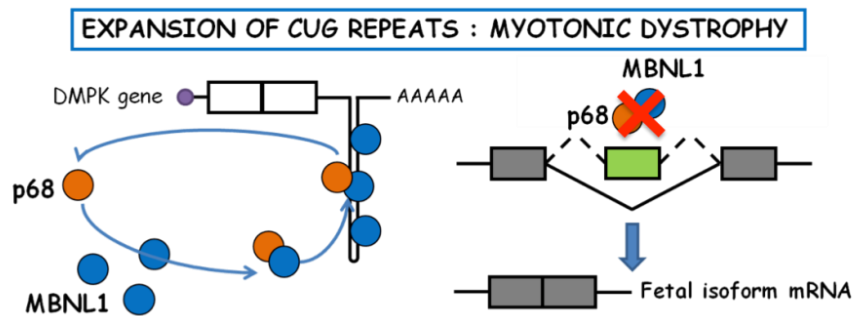


Figure 46 : Modèle proposé sur le rôle de p68 dans la Dystrophie Myotonique de type 1. Dans les cellules saines, p68 forme un complexe avec MBNL1, ouvre localement la structure secondaire de l'ARN afin de faciliter son interaction avec ses sites régulateurs entrainant un épissage alternatif correct et la formation d'isoformes adultes. Dans un contexte DM1, le complexe p68/MBNL1 facilite la fixation de MBNL1 sur les paires de nucléotides G-C par l'ouverture locale des expansions CUG entrainant sa séquestration dans la pathologie. L'absence de complexe p68/MBNL1 au niveau des cibles d'épissage aboutit à une altération de l'épissage alternatif de nombreux transcrits avec une réexpression des isoformes fœtales chez l'adulte à l'origine des symptômes cliniques.

F) Perspectives... Quelques points à élucider !

Mes travaux de thèse ont permis d'identifier de nouvelles fonctions pour p68 :

- **p68 est le premier facteur identifié ayant un rôle de facilitateur spécifique dans la fixation d'un facteur d'épissage alternatif.**
- **p68 est la première protéine modulant la structure secondaire des ARN pathologiques issus de maladies à expansions de triplets.**

Cependant, de nombreuses questions restent à éclaircir...

1) Existe-t-il d'autres cibles régulées par p68/MBNL1 et dérégulées dans la pathologie ?

L'une des questions les plus évidentes est de savoir si le cas de *TNNT2* dans la pathologie est isolé ou si au contraire une classe de transcrits bien définie nécessiterait l'activité de p68 pour permettre l'accessibilité de MBNL1. Il est difficile de répondre à cette question puisque nos travaux n'ont porté que sur l'étude de 4 transcrits dérégulés dans la pathologie. L'une des solutions pourrait venir de la comparaison du transcriptome de cellules musculaires inactivées pour MBNL1 avec le transcriptome de cellules inactivées pour p68/p72. De telles données ont été recueillies par l'équipe de Didier Auboeuf qui a développé des puces exon-array qui permettent d'établir le profil d'épissage de cibles régulées par de nombreux facteurs d'épissage dans des cellules musculaires murine C2C12. L'analyse de ces données devrait permettre l'identification d'une nouvelle classe de transcrits co-régulés par MBNL1 et p68 ce qui renforcerait le modèle que nous proposons.

Un autre point qu'il serait intéressant d'étudier concerne l'épissage alternatif de Tau. L'équipe de Nicolas Sergeant a très récemment reporté que l'épissage alternatif de l'exon 10 de *tau* était dérégulé dans le cerveau de certains patients DM1 (Dhaenens *et al.*, 2011) et que cette dérégulation n'est pas liée à la perte de fonction de MBNL1. Par ailleurs, l'équipe de Jane Wu a montré que p68, de par son activité hélicase, ouvre la structure secondaire à proximité de cet exon alternatif permettant l'accès de la snRNP U1 au site 5' d'épissage par un mécanisme faisant également intervenir le facteur d'épissage RBM4, (Kar *et al.*, 2011). Si l'effet de p68 n'est pas de faciliter la fixation de MBNL1 lors de la régulation de l'épissage alternatif de *tau*, peut-être que p68 faciliterait d'autres facteurs ? Des expériences de pontage

aux UV permettront de tester lequel de ces facteurs, identifiés comme régulateurs de l'épissage alternatif de l'exon 10, pourrait voir son accessibilité à l'ARN augmentée par p68.

2) p68 agit-il dans tous les cas comme un facilitateur de la fixation de MBNL1 ?

Dans les deux cas que nous avons répertoriés, p68 agit comme un facilitateur de la fixation de MBNL1. Est-ce un mécanisme général ou pourrions-nous penser que dans certains cas, p68 pourrait au contraire réduire la fixation de MBNL1 sur une de ces cibles ?

L'équipe de Bach-Elias a montré que p68 inhibe la fixation de hnRNP H en amont de l'exon alternatif IDX de h-Ras (Camats *et al.*, 2008), suggérant que dans certains cas, p68 pourrait agir à l'opposé de la facilitation de fixation d'un facteur d'épissage.

L'équipe de Hans Stahl a montré *in vitro* que p68 posséderait, conjointement à une activité hélicase, une activité « annealing » qui permet l'appariement de duplex ARN (Rössler *et al.*, 2001). D'après ces résultats, cette activité « annealing » serait prédominante à des concentrations saturantes de protéines et impliquerait la formation de molécules assemblées par cross-hybridation (Rössler *et al.*, 2001).

Nous pouvons donc supposer que dans certains cas, p68 et p72 pourrait favoriser la fermeture de la tige-boucle ce qui rendrait inaccessibles les sites d'interactions de MBNL1.

3) L'altération des complexes p68/MBNL1 pourrait-elle concerner d'autres voies que l'épissage alternatif dans la DM1?

D'après notre hypothèse, la facilitation de la fixation de MBNL1 par p68 sur les expansions CUG au détriment des cibles d'épissage pourrait entraîner l'aggravation de la séquestration. D'autres voies de signalisation que l'épissage alternatif pourraient être altérées par la perturbation du complexe p68/MBNL1.

L'équipe de Nicolas Charlet-Berguerand a récemment montré que MBNL1 est un régulateur de la biogénèse du pre-miR-1 via sa fixation sur un motif UGC localisé dans la boucle d'une structure secondaire en tige-boucle, représentée dans la Figure 47 (Rau *et al.*, 2011).

Il est proposé que la séquestration de MBNL1 dans la DM1 abolit son interaction avec le pre-miR-1 ce qui empêche sa maturation par Dicer. De plus, plusieurs travaux ont montré que p68 interagit avec Drosha et DGCR8 et que son activité hélicase permet de dérouler *in vivo* le duplex ARN de let-7 (Gregory *et al.*, 2004; Salzman *et al.*, 2007). Il serait donc intéressant de tester si p68 serait impliqué dans la fixation de MBNL1 par la modification de la structure secondaire de ce pre-miRNA. L'altération des complexes p68/MBNL1 pourrait ainsi participer à la dérégulation de maturation de miR-1 et être de ce fait impliqué dans les symptômes cardiaques de la DM1.

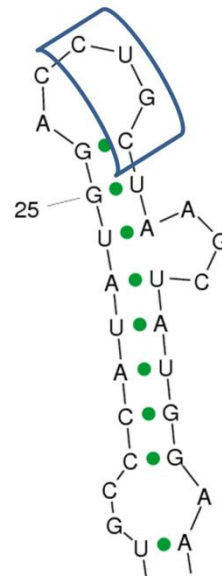


Figure 47 : Prédiction d'une partie de la structure secondaire du pre-miR-1 humain par *Sfold*. Ce pre-miR-1 contient un site d'interaction à MBNL1 (encadré en bleu) conforme au site consensus identifié par SELEX localisé dans une tige-boucle.

4) L'ARN non-codant SRA pourrait-il être le chaînon manquant entre MBNL1 et p68 ?

Comme nous l'avons abordé dans l'introduction, p68 interagit avec l'ARN non codant SRA (Steroid receptor RNA activator). Cet ARN non codant est produit par un gène dont l'épissage alternatif du premier intron détermine la production de cet ARN non codant SRA ou d'une protéine fonctionnelle appelée SRAP (Hatchell *et al.*, 2006). L'ARN non codant SRA est impliqué dans l'augmentation de l'activité de facteur de transcription parmi lesquels MyoD (Caretto *et al.*, 2006).

Lors de mes travaux de thèse, j'ai tenté de caractériser une interaction protéique entre MBNL1 et p68 par diverses méthodes (co-IP, GST Pull-down...). Seule une très faible interaction a été observée, suggérant que l'interaction ne soit pas directe et qu'un partenaire intermédiaire soit requis

L'un des candidats potentiels pourrait justement être l'ARN non-codant SRA. En effet, deux sites putatifs de MBNL1 (YGCY) peuvent être identifiés en utilisant l'outil de prédiction SFmap (Paz *et al.*, 2010). De plus la prédiction de la structure secondaire par Sfold (Ding *et al.*, 2004b) montre que ces deux sites seraient intégrés dans une tige-boucle (Figure 48). Nous pourrions ainsi supposer que l'ARN non-codant SRA puisse interagir avec MBNL1 de façon stable ainsi qu'avec p68, formant un complexe ribonucléoprotéique. Afin de tester cette hypothèse, nous pourrions inactiver l'ARN non-codant SRA par siRNA et voir si p68 perdrait son activité dans l'épissage alternatif de *TNNT2*. Il serait également intéressant de voir si SRA est requis pour la colocalisation de p68 avec les expansions CUG960.

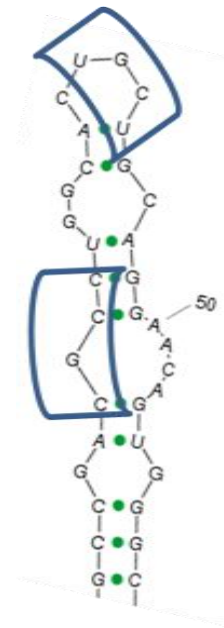


Figure 48 : Prédiction de la sous-structure STR7 de l'ARN non-codant SRA par Sfold. Cette région possède deux sites d'interaction à MBNL1 (encadré en bleu) conforme au site consensus identifié par SELEX localisé dans une tige-boucle.

L'équipe de Claire Francastel a récemment montré que la balance entre l'ARN non-codant SRA et sa version protéique SRAP est dérégulée dans des cellules satellites de patients DM1 et que ceci participerait au défaut de différenciation myogénique observé dans ces modèles cellulaires (Hubé *et al.*, 2011). La diminution de présence de l'ARN non-codant SRA pourrait altérer la fonction de p68 dans la DM1 et par conséquent sa capacité à réguler la fixation de MBNL1 sur ses cibles. Des expériences complémentaires sont nécessaires pour établir clairement l'implication de SRA sur la fonction de p68 dans la pathologie DM1.

CONCLUSION

CONCLUSION

En conclusion, les travaux effectués au cours de ma thèse permettent d'apporter de nouveaux éléments en vue d'une meilleure compréhension de la physiopathologie de la Dystrophie Myotonique de type 1, et plus particulièrement sur les mécanismes permettant la fixation de MBNL1 sur ces cibles normales et pathologiques par l'intervention de la DEAD-box ARN hélicase p68 qui modifie la structure secondaire de l'ARN contenant ses séquences régulatrices. Ces résultats suggèrent également que la fonction de p68 en tant que facilitateur de la fixation de MBNL1 pourrait être spécifique à ce facteur, à condition que sa séquence régulatrice soit intégrée dans une structure secondaire particulière. Ainsi, cette découverte permet d'envisager un moyen supplémentaire pour tenter de déchiffrer le « code d'épissage ».

Nos résultats suggèrent que seuls les événements d'épissage alternatif régulés conjointement par p68 et MBNL1, et non par p68 seul, soient dérégulés dans la pathologie DM1.

Pouvons-nous imaginer que d'une manière générale les DEAD-box ARN hélicases facilitent ou empêchent l'interaction d'un facteur d'épissage alternatif spécifique sur ses séquences régulatrices ? Peut-on envisager la régulation de certains types d'évènement d'épissage alternatif en termes de « paire DEAD-box ARN hélicase/facteur d'épissage » ?

Les analyses à grande échelle de prédiction du « code d'épissage » suivant les tissus effectuées par Blencowe tendent à montrer que les structures secondaires de l'ARN interviennent dans la régulation de l'épissage alternatif... sans pour autant qu'une « structure consensus » n'ait été définie.

Cette « structure consensus » pourrait-elle être spécifique à un certain type d'ARN hélicase, dont la fonction serait de la moduler pour permettre l'accès à un facteur d'épissage alternatif spécifique ?

Le regroupement des données issues d'analyses à haut-débit sur des puces analysant les variations d'épissage alternatif suivant l'inactivation d'un facteur d'épissage particulier, d'une ARN hélicase associée aux prédictions structurales de l'ARN permettra l'identification de nouvelles classes d'exons alternatifs régulés par ces différents éléments.

La prédiction du « splicing code » avec ces nouvelles données aurait un double intérêt : comprendre les mécanismes moléculaires à l'origine de la régulation de l'épissage alternatif par les ARN hélicases (dont p68) et prédire les transcrits pouvant être dérégulés dans une pathologie impliquant la DEAD-box ARN hélicase p68, comme la Dystrophie Myotonique de type 1 ou certains cancers dans lesquels p68 est surexprimée.

REFERENCES

BIBLIOGRAPHIQUES

REFERENCES BIBLIOGRAPHIQUES

- Alekseyenko AV, Kim N & Lee CJ (2007) Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA* **13**: 661-670
- Alexander RD, Innocente SA, Barrass JD & Beggs JD (2010) Splicing-dependent RNA polymerase pausing in yeast. *Mol. Cell* **40**: 582-593
- Alló M, Buggiano V, Fededa JP, Petrillo E, Schor I, de la Mata M, Agirre E, Plass M, Eyraas E, Elela SA, Klinck R, Chabot B & Kornblihtt AR (2009) Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat. Struct. Mol. Biol* **16**: 717-724
- de Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, Levezinho H, Gut M, Eick D, Gut I, Andrau J-C, Ferrier P & Carmo-Fonseca M (2011) Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat Struct Mol Biol*
- Alwazzan M, Newman E, Hamshere MG & Brook JD (1999) Myotonic dystrophy is associated with a reduced level of RNA from the DMWD allele adjacent to the expanded repeat. *Hum. Mol. Genet* **8**: 1491-1497
- Amack JD, Paguio AP & Mahadevan MS (1999) Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model. *Hum. Mol. Genet* **8**: 1975-1984
- Amack JD, Reagan SR & Mahadevan MS (2002) Mutant DMPK 3'-UTR transcripts disrupt C2C12 myogenic differentiation by compromising MyoD. *J. Cell Biol* **159**: 419-429
- Anant S, Henderson JO, Mukhopadhyay D, Navaratnam N, Kennedy S, Min J & Davidson NO (2001) Novel role for RNA-binding protein CUGBP2 in mammalian RNA editing. CUGBP2 modulates C to U editing of apolipoprotein B mRNA by interacting with apobec-1 and ACF, the apobec-1 complementation factor. *J. Biol. Chem* **276**: 47338-47351
- Andersen JS, Lam YW, Leung AKL, Ong S-E, Lyon CE, Lamond AI & Mann M (2005) Nucleolar proteome dynamics. *Nature* **433**: 77-83
- Andersson R, Enroth S, Rada-Iglesias A, Wadelius C & Komorowski J (2009) Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res* **19**: 1732-1741
- Artero R, Prokop A, Paricio N, Begemann G, Pueyo I, Mlodzik M, Perez-Alonso M & Baylies MK (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**: 131-143
- Ast G (2004) How did alternative splicing evolve? *Nat. Rev. Genet* **5**: 773-782
- Attard G, Cooper CS & de Bono JS (2009) Steroid hormone receptors in prostate cancer: a hard habit to break? *Cancer Cell* **16**: 458-462
- Auweter SD, Oberstrass FC & Allain FH-T (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res* **34**: 4943-4959
- Banroques J, Cordin O, Doère M, Linder P & Tanner NK (2008) A conserved phenylalanine

of motif IV in superfamily 2 helicases is required for cooperative, ATP-dependent binding of RNA substrates in DEAD-box proteins. *Mol. Cell. Biol* **28**: 3359-3371

Banroques J, Doère M, Dreyfus M, Linder P & Tanner NK (2010) Motif III in superfamily 2 « helicases » helps convert the binding energy of ATP into a high-affinity RNA binding site in the yeast DEAD-box protein Ded1. *J. Mol. Biol* **396**: 949-966

Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ & Frey BJ (2010) Deciphering the splicing code. *Nature* **465**: 53-59

Barrass JD & Beggs JD (2003) Splicing goes global. *Trends Genet* **19**: 295-298

Barta I & Iggo R (1995) Autoregulation of expression of the yeast Dbp2p « DEAD-box » protein is mediated by sequences in the conserved DBP2 intron. *EMBO J* **14**: 3800-3808

Bates GJ, Nicol SM, Wilson BJ, Jacobs A-MF, Bourdon J-C, Wardrop J, Gregory DJ, Lane DP, Perkins ND & Fuller-Pace FV (2005) The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. *EMBO J* **24**: 543-553

Batten FE & Gibb HP (1909) Two Cases of Myotonia Atrophica, showing a peculiar Distribution of Muscular Atrophy. *Proc. R. Soc. Med* **2**: 32-33

Begemann G, Paricio N, Artero R, Kiss I, Pérez-Alonso M & Mlodzik M (1997) muscleblind, a gene required for photoreceptor differentiation in Drosophila, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. *Development* **124**: 4321-4331

Bennett RJ & Keck JL (2004) Structure and function of RecQ DNA helicases. *Crit. Rev. Biochem. Mol. Biol* **39**: 79-97

Berger JM (2008) SnapShot: nucleic acid helicases and translocases. *Cell* **134**: 888-888.e1

Berul CI, Maguire CT, Aronovitz MJ, Greenwood J, Miller C, Gehrmann J, Housman D, Mendelsohn ME & Reddy S (1999) DMPK dosage alterations result in atrioventricular conduction abnormalities in a mouse myotonic dystrophy model. *J. Clin. Invest* **103**: R1-7

Biamonti G & Caceres JF (2009) Cellular stress and RNA splicing. *Trends Biochem. Sci* **34**: 146-153

Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem* **72**: 291-336

Blanchette M, Green RE, Brenner SE & Rio DC (2005) Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. *Genes Dev* **19**: 1306-1314

Bland CS, Wang ET, Vu A, David MP, Castle JC, Johnson JM, Burge CB & Cooper TA (2010) Global regulation of alternative splicing during myogenic differentiation. *Nucleic Acids Res* **38**: 7651-7664

Bleichert F & Baserga SJ (2007) The long unwinding road of RNA helicases. *Mol. Cell* **27**: 339-352

Blencowe BJ, Baurén G, Eldridge AG, Issner R, Nickerson JA, Rosonina E & Sharp PA (2000) The SRm160/300 splicing coactivator subunits. *RNA* **6**: 111-120

Böddeker N, Stade K & Franceschi F (1997) Characterization of DbpA, an Escherichia coli DEAD box protein with ATP independent RNA unwinding activity. *Nucleic Acids Res* **25**: 537-545

Bond AT, Mangus DA, He F & Jacobson A (2001) Absence of Dbp2p alters both nonsense-mediated mRNA decay and rRNA processing. *Mol. Cell. Biol* **21**: 7366-7379

Botta A, Vallo L, Rinaldi F, Bonifazi E, Amati F, Biancolella M, Gambardella S, Mancinelli E, Angelini C, Meola G & Novelli G (2007) Gene expression analysis in myotonic dystrophy:

indications for a common molecular pathogenic pathway in DM1 and DM2. *Gene Expr* **13**: 339-351

Boucher CA, King SK, Carey N, Krahe R, Winchester CL, Rahman S, Creavin T, Meghji P, Bailey ME & Chartier FL (1995) A novel homeodomain-encoding gene is associated with a large CpG island interrupted by the myotonic dystrophy unstable (CTG)_n repeat. *Hum. Mol. Genet* **4**: 1919-1925

Bourgeois CF, Popielarz M, Hildwein G & 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**: 7347-7356

Boutz PL, Chawla G, Stoilov P & Black DL (2007a) MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes Dev* **21**: 71-84

Boutz PL, Stoilov P, Li Q, Lin C-H, Chawla G, Ostrow K, Shiue L, Ares M Jr & Black DL (2007b) A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev* **21**: 1636-1652

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion J-P, Hudson T, Sohn R, Zemelman B, Snell RG, Rundle SA, Crow S, Davies J, Shelbourne P, Buxton J, Jones C, Juvonen V, *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**: 799-808

Brow DA (2002) Allosteric cascade of spliceosome activation. *Annu. Rev. Genet* **36**: 333-360

Buckanovich RJ & Darnell RB (1997) The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol. Cell. Biol* **17**: 3194-3201

Buelt MK, Glidden BJ & Storm DR (1994) Regulation of p68 RNA helicase by calmodulin and protein kinase C. *J. Biol. Chem* **269**: 29367-29370

Buj-Bello A, Furling D, Tronchère H, Laporte J, Lerouge T, Butler-Browne GS & Mandel J-L (2002) Muscle-specific alternative splicing of myotubularin-related 1 gene is impaired in DM1 muscle cells. *Hum. Mol. Genet* **11**: 2297-2307

Buratti E, Muro AF, Giombi M, Gherbassi D, Iaconcig A & Baralle FE (2004) RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol. Cell. Biol* **24**: 1387-1400

Burge CB, Padgett RA & Sharp PA (1998) Evolutionary fates and origins of U12-type introns. *Mol. Cell* **2**: 773-785

Buszczak M & Spradling AC (2006) The *Drosophila* P68 RNA helicase regulates transcriptional deactivation by promoting RNA release from chromatin. *Genes Dev* **20**: 977-989

Buxton J, Shelbourne P, Davies J, Jones C, Tongeren TV, Aslanidis C, de Jong P, Jansen G, Anvret M, Riley B, Williamson R & Johnson K (1992) Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* **355**: 547-548

Cáceres JF, Stamm S, Helfman DM & Krainer AR (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* **265**: 1706-1709

Calarco JA, Xing Y, Cáceres M, Calarco JP, Xiao X, Pan Q, Lee C, Preuss TM & Blencowe BJ (2007) Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev* **21**: 2963-2975

Camats M, Guil S, Kokolo M & Bach-Elias M (2008) P68 RNA Helicase (DDX5) Alters

Activity of Cis- and Trans-Acting Factors of the Alternative Splicing of H-Ras. *PLoS ONE* **3**: Carango P, Noble JE, Marks HG & Funanage VL (1993) Absence of myotonic dystrophy protein kinase (DMPK) mRNA as a result of a triplet repeat expansion in myotonic dystrophy. *Genomics* **18**: 340-348

Caretti G, Lei EP & Sartorelli V (2007) The DEAD-box p68/p72 proteins and the noncoding RNA steroid receptor activator SRA: eclectic regulators of disparate biological functions. *Cell Cycle* **6**: 1172-1176

Caretti G, Schiltz RL, Dilworth FJ, Di Padova M, Zhao P, Ogryzko V, Fuller-Pace FV, Hoffman EP, Tapscott SJ & Sartorelli V (2006) The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. *Dev. Cell* **11**: 547-560

Carmel I, Tal S, Vig I & Ast G (2004) Comparative analysis detects dependencies among the 5' splice-site positions. *RNA* **10**: 828-840

Carrillo Oesterreich F, Preibisch S & Neugebauer KM (2010) Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol. Cell* **40**: 571-581

Caruthers JM & McKay DB (2002) Helicase structure and mechanism. *Curr. Opin. Struct. Biol* **12**: 123-133

Cass D, Hotchko R, Barber P, Jones K, Gates DP & Berglund JA (2011) The four Zn fingers of MBNL1 provide a flexible platform for recognition of its RNA binding elements. *BMC Mol. Biol* **12**: 20

Castle JC, Zhang C, Shah JK, Kulkarni AV, Kalsotra A, Cooper TA & Johnson JM (2008) Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat Genet* **40**: 1416-1425

Causevic M, Hislop RG, Kernohan NM, Carey FA, Kay RA, Steele RJ & Fuller-Pace FV (2001) Overexpression and poly-ubiquitylation of the DEAD-box RNA helicase p68 in colorectal tumours. *Oncogene* **20**: 7734-7743

Chang TH, Latus LJ, Liu Z & Abbott JM (1997) Genetic interactions of conserved regions in the DEAD-box protein Prp28p. *Nucleic Acids Res* **25**: 5033-5040

Charlet-B N, Savkur RS, Singh G, Philips AV, Grice EA & Cooper TA (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol. Cell* **10**: 45-53

Chen G, Guo X, Lv F, Xu Y & Gao G (2008) p72 DEAD box RNA helicase is required for optimal function of the zinc-finger antiviral protein. *Proceedings of the National Academy of Sciences* **105**: 4352-4357

Chen K-Y, Pan H, Lin M-J, Li Y-Y, Wang L-C, Wu Y-C & Hsiao K-M (2007) Length-dependent toxicity of untranslated CUG repeats on *Caenorhabditis elegans*. *Biochem. Biophys. Res. Commun* **352**: 774-779

Chen M & Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**: 741-754

Cho DH, Thienes CP, Mahoney SE, Analau E, Filippova GN & Tapscott SJ (2005) Antisense Transcription and Heterochromatin at the DM1 CTG Repeats Are Constrained by CTCF. *Molecular Cell* **20**: 483-489

Chodavarapu RK, Feng S, Bernatavichute YV, Chen P-Y, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, Clark AT, Krämer U, Merchant SS, Zhang

- X, Jacobsen SE & Pellegrini M (2010) Relationship between nucleosome positioning and DNA methylation. *Nature* **466**: 388-392
- Clark EL, Coulson A, Dalgliesh C, Rajan P, Nicol SM, Fleming S, Heer R, Gaughan L, Leung HY, Elliott DJ, Fuller-Pace FV & Robson CN (2008) The RNA helicase p68 is a novel androgen receptor coactivator involved in splicing and is overexpressed in prostate cancer. *Cancer Res* **68**: 7938-7946
- Clark TA, Sugnet CW & Ares M Jr (2002) Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* **296**: 907-910
- Cleary JD, Tomé S, López Castel A, Panigrahi GB, Foiry L, Hagerman KA, Sroka H, Chitayat D, Gourdon G & Pearson CE (2010) Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus. *Nat. Struct. Mol. Biol* **17**: 1079-1087
- Clouet d'Orval B, d' Aubenton Carafa Y, Sirand-Pugnet P, Gallego M, Brody E & Marie J (1991) RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. *Science* **252**: 1823-1828
- Cordin O, Banroques J, Tanner NK & Linder P (2006) The DEAD-box protein family of RNA helicases. *Gene* **367**: 17-37
- Cordin O, Tanner NK, Doère M, Linder P & Banroques J (2004) The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J* **23**: 2478-2487
- Coutinho-Mansfield GC, Xue Y, Zhang Y & Fu X-D (2007) PTB/nPTB switch: a post-transcriptional mechanism for programming neuronal differentiation. *Genes & Development* **21**: 1573 -1577
- Crawford JB & Patton JG (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**: 8791-8802
- Dansithong W, Jog SP, Paul S, Mohammadzadeh R, Tring S, Kwok Y, Fry RC, Marjoram P, Comai L & Reddy S (2011) RNA steady-state defects in myotonic dystrophy are linked to nuclear exclusion of SHARP. *EMBO Rep* **12**: 735-742
- Dansithong W, Wolf CM, Sarkar P, Paul S, Chiang A, Holt I, Morris GE, Branco D, Sherwood MC, Comai L, Berul CI & Reddy S (2008) Cytoplasmic CUG RNA foci are insufficient to elicit key DM1 features. *PLoS ONE* **3**: e3968
- Darnell RB (2006) Developing global insight into RNA regulation. *Cold Spring Harb. Symp. Quant. Biol* **71**: 321-327
- David CJ & Manley JL (2008) The search for alternative splicing regulators: new approaches offer a path to a splicing code. *Genes Dev* **22**: 279-285
- Davis BM, McCurrach ME, Taneja KL, Singer RH & Housman DE (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**: 7388-7393
- Davis BN, Hilyard AC, Nguyen PH, Lagna G & Hata A (2010) Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha. *Mol. Cell* **39**: 373-384
- Day J & Ranum L (2005) RNA pathogenesis of the myotonic dystrophies. *Neuromuscular*

Disorders **15**: 5-16

Delaporte C (1998) Personality patterns in patients with myotonic dystrophy. *Arch. Neurol* **55**: 635-640

Deshler JO & Rossi JJ (1991) Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron. *Genes Dev* **5**: 1252-1263

Dhaenens CM, Schraen-Maschke S, Tran H, Vingtdeux V, Ghanem D, Leroy O, Delplanque J, Vanbrussel E, Delacourte A, Vermersch P, Muraige CA, Gruffat H, Sergeant A, Mahadevan MS, Ishiura S, Buée L, Cooper TA, Caillet-Boudin ML, Charlet-Berguerand N, Sablonnière B, *et al.* (2008) Overexpression of MBNL1 fetal isoforms and modified splicing of Tau in the DM1 brain: two individual consequences of CUG trinucleotide repeats. *Exp. Neurol* **210**: 467-478

Dhaenens CM, Tran H, Frandemiche M-L, Carpentier C, Schraen-Maschke S, Sistiaga A, Goicoechea M, Eddarkaoui S, Van Brussels E, Obriot H, Labudeck A, Gevaert MH, Fernandez-Gomez F, Charlet-Berguerand N, Deramecourt V, Muraige CA, Buée L, de Munain AL, Sablonnière B, Caillet-Boudin ML, *et al.* (2011) 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**: 732-742

Dillingham MS, Soultanas P & Wigley DB (1999) Site-directed mutagenesis of motif III in PcrA helicase reveals a role in coupling ATP hydrolysis to strand separation. *Nucleic Acids Res* **27**: 3310-3317

Ding J-H, Xu X, Yang D, Chu P-H, Dalton ND, Ye Z, Yeakley JM, Cheng H, Xiao R-P, Ross J, Chen J & Fu X-D (2004a) Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J* **23**: 885-896

Ding Y, Chan CY & Lawrence CE (2004b) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* **32**: W135-141

Domenjoud L, Gallinaro H, Kister L, Meyer S & Jacob M (1991) Identification of a specific exon sequence that is a major determinant in the selection between a natural and a cryptic 5' splice site. *Mol. Cell. Biol* **11**: 4581-4590

Du H, Cline MS, Osborne RJ, Tuttle DL, Clark TA, Donohue JP, Hall MP, Shiue L, Swanson MS, Thornton CA & Ares M Jr (2010) Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat. Struct. Mol. Biol* **17**: 187-193

Dutertre M, Gratadou L, Dardenne E, Germann S, Samaan S, Lidereau R, Driouch K, de la Grange P & Auboeuf D (2010a) Estrogen Regulation and Physiopathologic Significance of Alternative Promoters in Breast Cancer. *Cancer Research* **70**: 3760 -3770

Dutertre M, Lacroix-Triki M, Driouch K, de la Grange P, Gratadou L, Beck S, Millevoi S, Tazi J, Lidereau R, Vagner S & Auboeuf D (2010b) Exon-based clustering of murine breast tumor transcriptomes reveals alternative exons whose expression is associated with metastasis. *Cancer Res* **70**: 896-905

Ebraldize A, Wang Y, Petkova V, Ebraldize K & Junghans RP (2004) RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. *Science* **303**: 383-387

Edgell DR, Belfort M & Shub DA (2000) Barriers to Intron Promiscuity in Bacteria. *J. Bacteriol.* **182**: 5281-5289

Endoh H, Maruyama K, Masuhiro Y, Kobayashi Y, Goto M, Tai H, Yanagisawa J, Metzger D, Hashimoto S & Kato S (1999) Purification and identification of p68 RNA helicase acting

as a transcriptional coactivator specific for the activation function 1 of human estrogen receptor alpha. *Mol. Cell. Biol* **19**: 5363-5372

Expert-Bezançon A, Sureau A, Durosay P, Salesse R, Groeneveld H, Lecaer JP & Marie J (2004) hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B. *J. Biol. Chem* **279**: 38249-38259

Fardaei M, Larkin K, Brook JD & Hamshere MG (2001) In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res* **29**: 2766-2771

Fardaei M, Rogers MT, Thorpe HM, Larkin K, Hamshere MG, Harper PS & Brook JD (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**: 805-814

Faustino NA & Cooper TA (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**: 879-887

Feng Y, Chen M & Manley JL (2008) Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. *Nat Struct Mol Biol* **15**: 1040-1048

Fernandez-Costa JM & Artero R (2010) A conserved motif controls nuclear localization of Drosophila Muscleblind. *Mol. Cells* **30**: 65-70

Filippova GN, Thienes CP, Penn BH, Cho DH, Hu YJ, Moore JM, Klesert TR, Lobanekov VV & Tapscott SJ (2001) CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat. Genet* **28**: 335-343

Foiry L, Dong L, Savouret C, Hubert L, te Riele H, Junien C & Gourdon G (2006) Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Hum. Genet* **119**: 520-526

Förch P, Puig O, Kedersha N, Martínez C, Granneman S, Séraphin B, Anderson P & Valcárcel J (2000) The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Mol. Cell* **6**: 1089-1098

Ford MJ, Anton IA & Lane DP (1988) Nuclear protein with sequence homology to translation initiation factor eIF-4A. *Nature* **332**: 736-738

Forner F, Furlan S & Salvatori S (2010) Mass spectrometry analysis of complexes formed by myotonic dystrophy protein kinase (DMPK). *Biochim. Biophys. Acta* **1804**: 1334-1341

Freudenreich CH, Stavenhagen JB & Zakian VA (1997) Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell. Biol* **17**: 2090-2098

Frisch R, Singleton KR, Moses PA, Gonzalez IL, Carango P, Marks HG & Funanage VL (2001) Effect of triplet repeat expansion on chromatin structure and expression of DMPK and neighboring genes, SIX5 and DMWD, in myotonic dystrophy. *Mol. Genet. Metab* **74**: 281-291

Fu X-D (2004) Towards a Splicing Code. *Cell* **119**: 736-738

Fu Y, Pizzuti A, Fenwick R, King J, Rajnarayan S, Dunne P, Dubel J, Nasser G, Ashizawa T, de Jong P & et al. (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**: 1256-1258

Fugier C, Klein AF, 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 AL, Sergeant N, et al. (2011)

Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. *Nat. Med* **17**: 720-725

Fukuda T, Yamagata K, Fujiyama S, Matsumoto T, Koshida I, Yoshimura K, Mihara M, Naitou M, Endoh H, Nakamura T, Akimoto C, Yamamoto Y, Katagiri T, Foulds C, Takezawa S, Kitagawa H, Takeyama K-ichi, O'Malley BW & Kato S (2007) DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat. Cell Biol* **9**: 604-611

Fuller-Pace FV, Jacobs A-MF & Nicol SM (2007) Modulation of transcriptional activity of the DEAD-box family of RNA helicases, p68 (Ddx5) and DP103 (Ddx20), by SUMO modification. *Biochem. Soc. Trans* **35**: 1427-1429

Furling D, Lemieux D, Taneja K & Puymirat J (2001) Decreased levels of myotonic dystrophy protein kinase (DMPK) and delayed differentiation in human myotonic dystrophy myoblasts. *Neuromuscul. Disord* **11**: 728-735

Furukawa N, Ongusaha P, Jahng WJ, Araki K, Choi CS, Kim H-J, Lee YH, Kaibuchi K, Kahn BB, Masuzaki H, Kim JK, Lee SW & Kim Y-B (2005) Role of Rho-kinase in regulation of insulin action and glucose homeostasis. *Cell Metab* **2**: 119-129

Gallivan J-P & McGarvey MJ (2003) The importance of the Q motif in the ATPase activity of a viral helicase. *FEBS Lett* **554**: 485-488

Gambardella S, Rinaldi F, Lepore SM, Viola A, Loro E, Angelini C, Vergani L, Novelli G & Botta A (2010) Overexpression of microRNA-206 in the skeletal muscle from myotonic dystrophy type 1 patients. *J Transl Med* **8**: 48

Ge H & Manley JL (1990) A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell* **62**: 25-34

Geiss-Friedlander R & Melchior F (2007) Concepts in sumoylation: a decade on. *Nat. Rev. Mol. Cell Biol* **8**: 947-956

Ghanem D, Tran H, Dhaenens C-M, Schraen-Maschke S, Sablonnière B, Buée L, Sergeant N & Caillet-Boudin M-L (2009) Altered splicing of Tau in DM1 is different from the foetal splicing process. *FEBS Lett* **583**: 675-679

Goers ES, Purcell J, Voelker RB, Gates DP & Berglund JA (2010) MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res* **38**: 2467-2484

Goers ES, Voelker RB, Gates DP & Berglund JA (2008) RNA binding specificity of Drosophila muscleblind. *Biochemistry* **47**: 7284-7294

Goh P-Y, Tan Y-J, Lim SP, Tan YH, Lim SG, Fuller-Pace F & Hong W (2004) Cellular RNA helicase p68 relocalization and interaction with the hepatitis C virus (HCV) NS5B protein and the potential role of p68 in HCV RNA replication. *J. Virol* **78**: 5288-5298

Gomes-Pereira M, Foiry L, Nicole A, Huguet A, Junien C, Munnich A & Gourdon G (2007) CTG trinucleotide repeat « big jumps »: large expansions, small mice. *PLoS Genet* **3**: e52

Good PJ, Chen Q, Warner SJ & Herring DC (2000) A family of human RNA-binding proteins related to the Drosophila Bruno translational regulator. *J. Biol. Chem* **275**: 28583-28592

Gorbalenya AE & Koonin EV (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology* **3**: 419-429

Gorbalenya AE, Koonin EV, Donchenko AP & Blinov VM (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res* **17**: 4713-4730

Gorbunova V, Seluanov A, Mittelman D & Wilson JH (2004) Genome-wide demethylation destabilizes CTG.CAG trinucleotide repeats in mammalian cells. *Hum. Mol. Genet* **13**: 2979-2989

Grammatikakis I, Goo Y-H, Echeverria GV & Cooper TA (2011) Identification of MBNL1 and MBNL3 domains required for splicing activation and repression. *Nucleic Acids Res* **39**: 2769-2780

de la Grange P, Gratadou L, Delord M, Dutertre M & Auboeuf D (2010) Splicing factor and exon profiling across human tissues. *Nucleic Acids Research* **38**: 2825 -2838

Graveley BR (2005) Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**: 65-73

Graveley BR, Hertel KJ & Maniatis T (1998) A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO J* **17**: 6747-6756

Graves-Woodward KL, Gottlieb J, Challberg MD & Weller SK (1997) Biochemical analyses of mutations in the HSV-1 helicase-primase that alter ATP hydrolysis, DNA unwinding, and coupling between hydrolysis and unwinding. *J. Biol. Chem* **272**: 4623-4630

Gregory RI, Yan K-P, Amuthan G, Chendrimada T, Doratotaj B, Cooch N & Shiekhattar R (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**: 235-240

Guil S, Gattoni R, Carrascal M, Abián J, Stévenin J & Bach-Elias M (2003) Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Mol. Cell. Biol* **23**: 2927-2941

Hao M, Akrami K, Wei K, De Diego C, Che N, Ku J-H, Tidball J, Graves MC, Shieh PB & Chen F (2008) Muscleblind-like 2 (Mbnl2) -deficient mice as a model for myotonic dystrophy. *Dev. Dyn* **237**: 403-410

de Haro M, Al-Ramahi I, De Gouyon B, Ukani L, Rosa A, Faustino NA, Ashizawa T, Cooper TA & Botas J (2006) MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1. *Hum. Mol. Genet* **15**: 2138-2145

Hatchell EC, Colley SM, Beveridge DJ, Epis MR, Stuart LM, Giles KM, Redfern AD, Miles LEC, Barker A, MacDonald LM, Arthur PG, Lui JCK, Golding JL, McCulloch RK, Metcalf CB, Wilce JA, Wilce MCJ, Lanz RB, O'Malley BW & Leedman PJ (2006) SLIRP, a small SRA binding protein, is a nuclear receptor corepressor. *Mol. Cell* **22**: 657-668

Hertel KJ (2008) Combinatorial control of exon recognition. *J. Biol. Chem* **283**: 1211-1215

Hiller M, Zhang Z, Backofen R & Stamm S (2007) Pre-mRNA secondary structures influence exon recognition. *PLoS Genet* **3**: e204

Hirling H, Scheffner M, Restle T & Stahl H (1989) RNA helicase activity associated with the human p68 protein. *Nature* **339**: 562-564

Hirose T, Ideue T, Nagai M, Hagiwara M, Shu M-D & Steitz JA (2006) A Spliceosomal Intron Binding Protein, IBP160, Links Position-Dependent Assembly of Intron-Encoded Box C/D snoRNP to Pre-mRNA Splicing. *Molecular Cell* **23**: 673-684

Ho TH, Charlet-B N, Poulos MG, Singh G, Swanson MS & Cooper TA (2004) Muscleblind proteins regulate alternative splicing. *EMBO J* **23**: 3103-3112

Ho TH, Savkur RS, Poulos MG, Mancini MA, Swanson MS & Cooper TA (2005) Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy. *J. Cell. Sci* **118**: 2923-2933

Hodges C, Bintu L, Lubkowska L, Kashlev M & Bustamante C (2009) Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* **325**: 626-628

Hodgman TC (1988) A new superfamily of replicative proteins. *Nature* **333**: 22-23

Hofmann-Radvanyi H, Lavedan C, Rabès JP, Savoy D, Duros C, Johnson K & Junien C (1993) Myotonic dystrophy: absence of CTG enlarged transcript in congenital forms, and low expression of the normal allele. *Hum. Mol. Genet* **2**: 1263-1266

Holbrook SR (2005) RNA structure: the long and the short of it. *Curr. Opin. Struct. Biol* **15**: 302-308

Holt I, Jacquemin V, Fardaei M, Sewry CA, Butler-Browne GS, Furling D, Brook JD & Morris GE (2009) Muscleblind-like proteins: similarities and differences in normal and myotonic dystrophy muscle. *Am. J. Pathol* **174**: 216-227

Holt I, Mittal S, Furling D, Butler-Browne GS, Brook JD & Morris GE (2007) Defective mRNA in myotonic dystrophy accumulates at the periphery of nuclear splicing speckles. *Genes Cells* **12**: 1035-1048

Hönig A, Auboeuf D, Parker MM, O'Malley BW & Berget SM (2002) Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72. *Mol. Cell. Biol* **22**: 5698-5707

Houseley JM, Garcia-Casado Z, Pascual M, Paricio N, O'Dell KMC, Monckton DG & Artero RD (2006) Noncanonical RNAs from transcripts of the *Drosophila* muscleblind gene. *J. Hered* **97**: 253-260

Houseley JM, Wang Z, Brock GJR, Soloway J, Artero R, Perez-Alonso M, O'Dell KMC & Monckton DG (2005) Myotonic dystrophy associated expanded CUG repeat muscleblind positive ribonuclear foci are not toxic to *Drosophila*. *Hum. Mol. Genet* **14**: 873-883

Hsu R-J, Hsiao K-M, Lin M-J, Li C-Y, Wang L-C, Chen L-K & Pan H (2011) Long Tract of Untranslated CAG Repeats Is Deleterious in Transgenic Mice. *PLoS ONE* **6**: e16417

Huang H, Wahlin KJ, McNally M, Irving ND & Adler R (2008) Developmental regulation of muscleblind-like (MBNL) gene expression in the chicken embryo retina. *Dev. Dyn* **237**: 286-296

Huang Y & Liu Z-R (2002) The ATPase, RNA unwinding, and RNA binding activities of recombinant p68 RNA helicase. *J. Biol. Chem* **277**: 12810-12815

Huang Y, Gattoni R, Stévenin J & Steitz JA (2003) SR Splicing Factors Serve as Adapter Proteins for TAP-Dependent mRNA Export. *Molecular Cell* **11**: 837-843

Hubé F, Velasco G, Rollin J, Furling D & Francastel C (2011) Steroid receptor RNA activator protein binds to and counteracts SRA RNA-mediated activation of MyoD and muscle differentiation. *Nucleic Acids Res* **39**: 513-525

Hui J, Stangl K, Lane WS & Bindereif A (2003) HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat Struct Mol Biol* **10**: 33-37

Iakova P, Wang G-L, Timchenko L, Michalak M, Pereira-Smith OM, Smith JR & Timchenko NA (2004) Competition of CUGBP1 and calreticulin for the regulation of p21 translation determines cell fate. *EMBO J* **23**: 406-417

Iggo RD & Lane DP (1989) Nuclear protein p68 is an RNA-dependent ATPase. *EMBO J* **8**: 1827-1831

Iggo RD, Jamieson DJ, MacNeill SA, Southgate J, McPheat J & Lane DP (1991) p68 RNA helicase: identification of a nucleolar form and cloning of related genes containing a

conserved intron in yeasts. *Mol. Cell. Biol* **11**: 1326-1333

Ip JY, Schmidt D, Pan Q, Ramani AK, Fraser AG, Odom DT & Blencowe BJ (2011) Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res* **21**: 390-401

Ip JY, Tong A, Pan Q, Topp JD, Blencowe BJ & Lynch KW (2007) Global analysis of alternative splicing during T-cell activation. *RNA* **13**: 563-572

Jacobs A-MF, Nicol SM, Hislop RG, Jaffray EG, Hay RT & Fuller-Pace FV (2007) SUMO modification of the DEAD box protein p68 modulates its transcriptional activity and promotes its interaction with HDAC1. *Oncogene* **26**: 5866-5876

Jalal C, Uhlmann-Schiffler H & Stahl H (2007) Redundant role of DEAD box proteins p68 (Ddx5) and p72/p82 (Ddx17) in ribosome biogenesis and cell proliferation. *Nucleic Acids Res* **35**: 3590-3601

Janknecht R (2002) The versatile functions of the transcriptional coactivators p300 and CBP and their roles in disease. *Histol. Histopathol* **17**: 657-668

Janknecht R (2010) Multi-talented DEAD-box proteins and potential tumor promoters: p68 RNA helicase (DDX5) and its paralog, p72 RNA helicase (DDX17). *Am J Transl Res* **2**: 223-234

Jankowsky E (2011) RNA helicases at work: binding and rearranging. *Trends Biochem. Sci* **36**: 19-29

Jankowsky E & Bowers H (2006) Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res* **34**: 4181-4188

Jansen G, Groenen PJ, Bächner D, Jap PH, Coerwinkel M, Oerlemans F, van den Broek W, Gohlsch B, Pette D, Plomp JJ, Molenaar PC, Nederhoff MG, van Echteld CJ, Dekker M, Berns A, Hameister H & Wieringa B (1996) Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. *Nat. Genet* **13**: 316-324

Jarmoskaite I & Russell R (2011) DEAD-box proteins as RNA helicases and chaperones. *WIREs RNA* **2**: 135-152

Jensen ED, Niu L, Caretti G, Nicol SM, Teplyuk N, Stein GS, Sartorelli V, van Wijnen AJ, Fuller-Pace FV & Westendorf JJ (2008) p68 (Ddx5) interacts with Runx2 and regulates osteoblast differentiation. *J. Cell. Biochem* **103**: 1438-1451

Jiang H, Mankodi A, Swanson MS, Moxley RT & Thornton CA (2004) Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Human Molecular Genetics* **13**: 3079 -3088

Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R & Shoemaker DD (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**: 2141-2144

Jorde LB (2000) Medical Genetics 2^e éd.

Junghans RP (2009) Dystrophia myotonia: why focus on foci? *Eur. J. Hum. Genet* **17**: 543-553

Kahlina K, Goren I, Pfeilschifter J & Frank S (2004) p68 DEAD box RNA helicase expression in keratinocytes. Regulation, nucleolar localization, and functional connection to proliferation and vascular endothelial growth factor gene expression. *J. Biol. Chem* **279**: 44872-44882

Kaliman P & Llagostera E (2008) Myotonic dystrophy protein kinase (DMPK) and its role in

the pathogenesis of myotonic dystrophy 1. *Cell. Signal* **20**: 1935-1941

Kalsotra A, Wang K, Li P-F & Cooper TA (2010) MicroRNAs coordinate an alternative splicing network during mouse postnatal heart development. *Genes Dev* **24**: 653-658

Kalsotra A, Xiao X, Ward AJ, Castle JC, Johnson JM, Burge CB & Cooper TA (2008) A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc. Natl. Acad. Sci. U.S.A* **105**: 20333-20338

Kanadia RN, Johnstone KA, Mankodi A, Lungu C, Thornton CA, Esson D, Timmers AM, Hauswirth WW & Swanson MS (2003a) A muscleblind knockout model for myotonic dystrophy. *Science* **302**: 1978-1980

Kanadia RN, Shin J, Yuan Y, Beattie SG, Wheeler TM, Thornton CA & Swanson MS (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**: 11748-11753

Kanadia RN, Urbinati CR, Crusselle VJ, Luo D, Lee Y-J, Harrison JK, Oh SP & Swanson MS (2003b) Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3. *Gene Expr. Patterns* **3**: 459-462

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J & Segal E (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**: 362-366

Kar A, Fushimi K, Zhou X, Ray P, Shi C, Chen X, Liu Z, Chen S & Wu JY (2011) RNA Helicase p68 (DDX5) Regulates tau Exon 10 Splicing by Modulating a Stem-Loop Structure at the 5' Splice Site. *Mol. Cell. Biol.* **31**: 1812-1821

Karagiannides I, Thomou T, Tchkonina T, Pirtskhalava T, Kypreos KE, Cartwright A, Dalagiorgou G, Lash TL, Farmer SR, Timchenko NA & Kirkland JL (2006) Increased CUG triplet repeat-binding protein-1 predisposes to impaired adipogenesis with aging. *J. Biol. Chem* **281**: 23025-23033

Keene JD (2007) RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet* **8**: 533-543

Keene JD, Komisarow JM & Friedersdorf MB (2006) RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* **1**: 302-307

Kellerer M, Lammers R, Ermel B, Tippmer S, Vogt B, Obermaier-Kusser B, Ullrich A & Häring HU (1992) Distinct alpha-subunit structures of human insulin receptor A and B variants determine differences in tyrosine kinase activities. *Biochemistry* **31**: 4588-4596

Kiliszek A, Kierzek R, Krzyzosiak WJ & Rypniewski W (2010) Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucleic Acids Res* **38**: 8370-8376

Kim D-H, Langlois M-A, Lee K-B, Riggs AD, Puymirat J & Rossi JJ (2005) HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. *Nucleic Acids Res* **33**: 3866-3874

Kimura T, Lueck JD, Harvey PJ, Pace SM, Ikemoto N, Casarotto MG, Dirksen RT & Dulhunty AF (2009) Alternative splicing of RyR1 alters the efficacy of skeletal EC coupling. *Cell Calcium* **45**: 264-274

Kimura T, Nakamori M, Lueck JD, Pouliquin P, Aoike F, Fujimura H, Dirksen RT, Takahashi MP, Dulhunty AF & 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**: 2189-2200

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

Klesert TR, Cho DH, Clark JI, Maylie J, Adelman J, Snider L, Yuen EC, Soriano P & Tapscott SJ (2000) Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy. *Nat. Genet* **25**: 105-109

Koga R, Nakao Y, Kurano Y, Tsukahara T, Nakamura A, Ishiura S, Nonaka I & Arahata K (1994) Decreased myotonin-protein kinase in the skeletal and cardiac muscles in myotonic dystrophy. *Biochem. Biophys. Res. Commun* **202**: 577-585

Kohtz JD, Jamison SF, Will CL, Zuo P, Luhrmann R, Garcia-Blanco MA & Manley JL (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368**: 119-124

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS & Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet* **41**: 376-381

Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ & Nogues G (2004) Multiple links between transcription and splicing. *RNA* **10**: 1489-1498

Koshelev M, Sarma S, Price RE, Wehrens XHT & Cooper TA (2010) Heart-specific overexpression of CUGBP1 reproduces functional and molecular abnormalities of myotonic dystrophy type 1. *Hum. Mol. Genet* **19**: 1066-1075

Krecic AM & Swanson MS (1999) hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol* **11**: 363-371

Krol J, Fiszler A, Mykowska A, Sobczak K, de Mezer M & Krzyzosiak WJ (2007) Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Mol. Cell* **25**: 575-586

Kumari D & Usdin K (2009) Chromatin remodeling in the noncoding repeat expansion diseases. *J. Biol. Chem* **284**: 7413-7417

Kuyumcu-Martinez NM, Wang G-S & Cooper TA (2007) Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. *Mol. Cell* **28**: 68-78

Ladd AN, Charlet N & Cooper TA (2001) The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol. Cell. Biol* **21**: 1285-1296

Ladd AN, Nguyen NH, Malhotra K & Cooper TA (2004) CELF6, a member of the CELF family of RNA-binding proteins, regulates muscle-specific splicing enhancer-dependent alternative splicing. *J. Biol. Chem* **279**: 17756-17764

Lamm GM, Nicol SM, Fuller-Pace FV & Lamond AI (1996) p72: a human nuclear DEAD box protein highly related to p68. *Nucleic Acids Res* **24**: 3739-3747

Lane DP & Hoeffler WK (1980) SV40 large T shares an antigenic determinant with a cellular protein of molecular weight 68,000. *Nature* **288**: 167-170

Lareau LF, Inada M, Green RE, Wengrod JC & Brenner SE (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**: 926-929

Lee C-G (2002) RH70, a bidirectional RNA helicase, co-purifies with U1snRNP. *J. Biol.*

Chem **277**: 39679-39683

Lee EK, Kim HH, Kuwano Y, Abdelmohsen K, Srikantan S, Subaran SS, Gleichmann M, Mughal MR, Martindale JL, Yang X, Worley PF, Mattson MP & Gorospe M (2010a) hnRNP C promotes APP translation by competing with FMRP for APP mRNA recruitment to P bodies. *Nat Struct Mol Biol* **17**: 732-739

Lee JE & Cooper TA (2009) Pathogenic mechanisms of myotonic dystrophy. *Biochem. Soc. Trans* **37**: 1281-1286

Lee JE, Lee JY, Wilusz J, Tian B & Wilusz CJ (2010b) Systematic analysis of cis-elements in unstable mRNAs demonstrates that CUGBP1 is a key regulator of mRNA decay in muscle cells. *PLoS ONE* **5**: e11201

Leipe DD, Koonin EV & Aravind L (2003) Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol* **333**: 781-815

Leroy O, Dhaenens C-M, Schraen-Maschke S, Belarbi K, Delacourte A, Andreadis A, Sablonnière B, Buée L, Sergeant N & 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**: 852-859

Li X, Wang J & Manley JL (2005) Loss of splicing factor ASF/SF2 induces G2 cell cycle arrest and apoptosis, but inhibits internucleosomal DNA fragmentation. *Genes Dev* **19**: 2705-2714

Libby RT, Hagerman KA, Pineda VV, Lau R, Cho DH, Baccam SL, Axford MM, Cleary JD, Moore JM, Sopher BL, Tapscott SJ, Filippova GN, Pearson CE & La Spada AR (2008) CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. *PLoS Genet* **4**: e1000257

Libri D, Goux-Pelletan M, Brody E & Fiszman MY (1990) Exon as well as intron sequences are cis-regulating elements for the mutually exclusive alternative splicing of the beta tropomyosin gene. *Mol. Cell. Biol* **10**: 5036-5046

Libri D, Piseri A & Fiszman MY (1991) Tissue-specific splicing in vivo of the beta-tropomyosin gene: dependence on an RNA secondary structure. *Science* **252**: 1842-1845

Lin C, Yang L, Yang JJ, Huang Y & Liu Z-R (2005) ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome. *Mol. Cell. Biol* **25**: 7484-7493

Lin J & Arlinghaus R (2008) Activated c-Abl tyrosine kinase in malignant solid tumors. *Oncogene* **27**: 4385-4391

Linder P (2006) Dead-box proteins: a family affair—active and passive players in RNP-remodeling. *Nucleic Acids Research* **34**: 4168 -4180

Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW & Ranum LPW (2001) Myotonic Dystrophy Type 2 Caused by a CCTG Expansion in Intron 1 of ZNF9. *Science* **293**: 864 -867

Liu Y-F, Liu H-Y, Tu L-C, Lin C-W, Hsiao K-M & Pan H (2008) Zebrafish muscleblind-like genes: identification, structural features and expression. *Comp. Biochem. Physiol. B, Biochem. Mol. Biol* **151**: 118-124

Liu Z-R (2002) p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5' splice site duplex. *Mol. Cell. Biol* **22**: 5443-5450

Liu ZR, Sargueil B & Smith CW (1998) Detection of a novel ATP-dependent cross-linked

protein at the 5' splice site-U1 small nuclear RNA duplex by methylene blue-mediated photo-cross-linking. *Mol. Cell. Biol* **18**: 6910-6920

Llagostera E, Carmona MC, Vicente M, Escorihuela RM & Kaliman P (2009) High-fat diet induced adiposity and insulin resistance in mice lacking the myotonic dystrophy protein kinase. *FEBS Lett* **583**: 2121-2125

Llagostera E, Catalucci D, Marti L, Liesa M, Camps M, Ciaraldi TP, Kondo R, Reddy S, Dillmann WH, Palacin M, Zorzano A, Ruiz-Lozano P, Gomis R & Kaliman P (2007) Role of myotonic dystrophy protein kinase (DMPK) in glucose homeostasis and muscle insulin action. *PLoS ONE* **2**: e1134

Llamusi B & Artero R (2008) Molecular Effects of the CTG Repeats in Mutant Dystrophia Myotonica Protein Kinase Gene. *Curr. Genomics* **9**: 509-516

Loeb DD, Mack AA & Tian R (2002) A secondary structure that contains the 5' and 3' splice sites suppresses splicing of duck hepatitis B virus pregenomic RNA. *J. Virol* **76**: 10195-10202

López Castel A, Nakamori M, Thornton CA & Pearson CE (2011a) Identification of restriction endonucleases sensitive to 5-cytosine methylation at non-CpG sites, including expanded (CAG)_n/(CTG)_n repeats. *Epigenetics* **6**: 416-420

López Castel A, Nakamori M, Tomé S, Chitayat D, Gourdon G, Thornton CA & Pearson CE (2011b) Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Hum. Mol. Genet* **20**: 1-15

Lorenz C, von Pelchrzim F & Schroeder R (2006) Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat Protoc* **1**: 2204-2212

Luco RF, Allo M, Schor IE, Kornblihtt AR & Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. *Cell* **144**: 16-26

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM & Misteli T (2010) Regulation of alternative splicing by histone modifications. *Science* **327**: 996-1000

Lueck JD, Lungu C, Mankodi A, Osborne RJ, Welle SL, Dirksen RT & Thornton CA (2007a) Chloride channelopathy in myotonic dystrophy resulting from loss of posttranscriptional regulation for CLCN1. *Am. J. Physiol., Cell Physiol* **292**: C1291-1297

Lueck JD, Mankodi A, Swanson MS, Thornton CA & Dirksen RT (2007b) Muscle chloride channel dysfunction in two mouse models of myotonic dystrophy. *J. Gen. Physiol* **129**: 79-94

Lund MK & Guthrie C (2005) The DEAD-Box Protein Dbp5p Is Required to Dissociate Mex67p from Exported mRNPs at the Nuclear Rim. *Molecular Cell* **20**: 645-651

Lynch M, Chen L, Ravitz MJ, Mehtani S, Korenblat K, Pazin MJ & Schmidt EV (2005) hnRNP K binds a core polypyrimidine element in the eukaryotic translation initiation factor 4E (eIF4E) promoter, and its regulation of eIF4E contributes to neoplastic transformation. *Mol. Cell. Biol* **25**: 6436-6453

Machuca-Tzili L, Brook D & Hilton-Jones D (2005) Clinical and molecular aspects of the myotonic dystrophies: A review. *Muscle & Nerve* **32**: 1-18

Machuca-Tzili L, Thorpe H, Robinson TE, Sewry C & Brook JD (2006) Flies deficient in Muscleblind protein model features of myotonic dystrophy with altered splice forms of Z-band associated transcripts. *Hum. Genet* **120**: 487-499

Machuca-Tzili LE, Buxton S, Thorpe A, Timson CM, Wigmore P, Luther PK & Brook JD

(2011) Zebrafish deficient for Muscleblind-like 2 exhibit features of myotonic dystrophy. *Dis Model Mech* **4**: 381-392

Madhani HD & Guthrie C (1994) Dynamic RNA-RNA interactions in the spliceosome. *Annu. Rev. Genet* **28**: 1-26

Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barcelo J, O'Hoy K & et al. (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**: 1253 -1255

Mahadevan MS, Yadava RS, Yu Q, Balijepalli S, Frenzel-McCardell CD, Bourne TD & Phillips LH (2006) Reversible model of RNA toxicity and cardiac conduction defects in myotonic dystrophy. *Nat. Genet* **38**: 1066-1070

Makeyev EV, Zhang J, Carrasco MA & Maniatis T (2007) The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* **27**: 435-448

Mankodi A, Logigian E, Callahan L, McClain C, White R, Henderson D, Krym M & Thornton CA (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* **289**: 1769-1773

Mankodi A, Takahashi MP, Jiang H, Beck CL, Bowers WJ, Moxley RT, Cannon SC & Thornton CA (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**: 35-44

Mankodi A, Teng-Umuay P, Krym M, Henderson D, Swanson M & Thornton CA (2003) Ribonuclear inclusions in skeletal muscle in myotonic dystrophy types 1 and 2. *Ann. Neurol* **54**: 760-768

Mankodi A, Urbinati CR, Yuan QP, Moxley RT, Sansone V, Krym M, Henderson D, Schalling M, Swanson MS & Thornton CA (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet* **10**: 2165-2170

Maris C, Dominguez C & Allain FH-T (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* **272**: 2118-2131

Marquis J, Paillard L, Audic Y, Cosson B, Danos O, Le Bec C & Osborne HB (2006) CUG-BP1/CELF1 requires UGU-rich sequences for high-affinity binding. *Biochem. J* **400**: 291-301

Martorell L, Monckton DG, Gamez J, Johnson KJ, Gich I, Lopez de Munain A & Baiget M (1998) Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Hum. Mol. Genet* **7**: 307-312

de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D & Kornbliht AR (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* **12**: 525-532

Matlin AJ, Clark F & Smith CWJ (2005) Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol* **6**: 386-398

Mayeda A, Helfman DM & Krainer AR (1993) Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Mol. Cell. Biol* **13**: 2993-3001

McClain DA (1991) Different ligand affinities of the two human insulin receptor splice variants are reflected in parallel changes in sensitivity for insulin action. *Mol. Endocrinol* **5**: 734-739

Mehta J & Tuteja R (2011) A novel dual Dbp5/DDX19 homologue from Plasmodium falciparum requires Q motif for activity. *Mol. Biochem. Parasitol* **176**: 58-63

Meola N, Gennarino VA & Banfi S (2009) microRNAs and genetic diseases. *Pathogenetics* **2**: 7

Michalowski S, Miller JW, Urbinati CR, Paliouras M, Swanson MS & 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**: 3534-3542

Miller JW, Urbinati CR, Teng-Umuay P, Stenberg MG, Byrne BJ, Thornton CA & Swanson MS (2000) Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J* **19**: 4439-4448

Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* **447**: 932-940

Moller DE, Yokota A, Caro JF & Flier JS (1989) Tissue-specific expression of two alternatively spliced insulin receptor mRNAs in man. *Mol. Endocrinol* **3**: 1263-1269

Monckton DG, Wong LJ, Ashizawa T & Caskey CT (1995) Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum. Mol. Genet* **4**: 1-8

Mooers BHM, Logue JS & Berglund JA (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc. Natl. Acad. Sci. U.S.A* **102**: 16626-16631

Mooney SM, Grande JP, Salisbury JL & Janknecht R (2010) Sumoylation of p68 and p72 RNA helicases affects protein stability and transactivation potential. *Biochemistry* **49**: 1-10

Moore HC, Johnston M, Nicol SM, Bourdon J-C, Thompson AM, Hutvagner G & Fuller-Pace FV (2011) An evolutionarily conserved, alternatively spliced, intron in the p68/DDX5 DEAD-box RNA helicase gene encodes a novel miRNA. *RNA* **17**: 555-562

Moore MJ & Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**: 688-700

Moore MJ & Sharp PA (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365**: 364-368

Mori D, Sasagawa N, Kino Y & Ishiura S (2008) Quantitative analysis of CUG-BP1 binding to RNA repeats. *J. Biochem* **143**: 377-383

Mosammaparast N & Shi Y (2010) Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annu. Rev. Biochem.* **79**: 155-179

Moss L (2002) What Genes Can't Do The MIT Press

Mukhopadhyay D, Jung J, Murmu N, Houchen CW, Dieckgraefe BK & Anant S (2003) CUGBP2 plays a critical role in apoptosis of breast cancer cells in response to genotoxic injury. *Ann. N. Y. Acad. Sci* **1010**: 504-509

Mutsuddi M, Marshall CM, Benzow KA, Koob MD & Rebay I (2004) The spinocerebellar ataxia 8 noncoding RNA causes neurodegeneration and associates with staufen in Drosophila. *Curr. Biol* **14**: 302-308

Nagy E & Maquat LE (1998) A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci* **23**: 198-199

Nahkuri S, Taft RJ & Mattick JS (2009) Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle* **8**: 3420-3424

Neubauer G, King A, Rappsilber J, Calvio C, Watson M, Ajuh P, Sleeman J, Lamond A & Mann M (1998) Mass spectrometry and EST-database searching allows characterization of

the multi-protein spliceosome complex. *Nat. Genet* **20**: 46-50

Nezu Y, Kino Y, Sasagawa N, Nishino I & Ishiura S (2007) Expression of MBNL and CELF mRNA transcripts in muscles with myotonic dystrophy. *Neuromuscul. Disord* **17**: 306-312

Nicol SM & Fuller-Pace FV (2010) Analysis of the RNA helicase p68 (Ddx5) as a transcriptional regulator. *Methods Mol. Biol* **587**: 265-279

Nicol SM, Causevic M, Prescott AR & Fuller-Pace FV (2000) The nuclear DEAD box RNA helicase p68 interacts with the nucleolar protein fibrillarin and colocalizes specifically in nascent nucleoli during telophase. *Exp. Cell Res* **257**: 272-280

Novelli G, Gennarelli M, Menegazzo E, Mostacciolo ML, Pizzuti A, Fattorini C, Tessarolo D, Tomelleri G, Giacanelli M & Danieli GA (1993) (CTG)_n triplet mutation and phenotype manifestations in myotonic dystrophy patients. *Biochem. Med. Metab. Biol* **50**: 85-92

Oesterreich FC, Bieberstein N & Neugebauer KM (2011) Pause locally, splice globally. *Trends Cell Biol* **21**: 328-335

Ogilvie VC, Wilson BJ, Nicol SM, Morrice NA, Saunders LR, Barber GN & Fuller-Pace FV (2003) The highly related DEAD box RNA helicases p68 and p72 exist as heterodimers in cells. *Nucleic Acids Res* **31**: 1470-1480

Okano HJ & Darnell RB (1997) A hierarchy of Hu RNA binding proteins in developing and adult neurons. *J. Neurosci* **17**: 3024-3037

Onishi H, Kino Y, Morita T, Futai E, Sasagawa N & Ishiura S (2008) MBNL1 associates with YB-1 in cytoplasmic stress granules. *J. Neurosci. Res* **86**: 1994-2002

Onoa B & Tinoco I Jr (2004) RNA folding and unfolding. *Curr. Opin. Struct. Biol* **14**: 374-379

Orengo JP, Chambon P, Metzger D, Mosier DR, Snipes GJ & Cooper TA (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**: 2646-2651

Osborne & Thornton (2006) RNA-dominant diseases. *Human Molecular Genetics* **15**: R162-R169

Osborne RJ, Lin X, Welle S, Sobczak K, O'Rourke JR, Swanson MS & Thornton CA (2009) Transcriptional and post-transcriptional impact of toxic RNA in myotonic dystrophy. *Hum. Mol. Genet* **18**: 1471-1481

Otten AD & Tapscott SJ (1995) Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc. Natl. Acad. Sci. U.S.A* **92**: 5465-5469

Pagani F & Baralle FE (2004) Opinion: Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* **5**: 389-396

Paillard L, Legagneux V & Beverley Osborne H (2003) A functional deadenylation assay identifies human CUG-BP as a deadenylation factor. *Biol. Cell* **95**: 107-113

Pan Q, Shai O, Lee LJ, Frey BJ & Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**: 1413-1415

Pan Q, Shai O, Misquitta C, Zhang W, Saltzman AL, Mohammad N, Babak T, Siu H, Hughes TR, Morris QD, Frey BJ & Blencowe BJ (2004) Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* **16**: 929-941

Pascual M, Vicente M, Monferrer L & Artero R (2006) The Muscleblind family of proteins:

- an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation* **74**: 65-80
- Paul S, Dansithong W, Kim D, Rossi J, Webster NJG, Comai L & Reddy S (2006) Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. *EMBO J* **25**: 4271-4283
- Pause A & Sonenberg N (1992) Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J* **11**: 2643-2654
- Pause A, Méthot N & Sonenberg N (1993) The HRIGRXXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol. Cell. Biol* **13**: 6789-6798
- Paz I, Akerman M, Dror I, Kosti I & Mandel-Gutfreund Y (2010) SFmap: a web server for motif analysis and prediction of splicing factor binding sites. *Nucleic Acids Res* **38**: W281-285
- Perales R & Bentley D (2009) « Cotranscriptionality »: the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell* **36**: 178-191
- Perbellini R, Greco S, Sarra-Ferraris G, Cardani R, Capogrossi MC, Meola G & Martelli F (2011) Dysregulation and cellular mislocalization of specific miRNAs in myotonic dystrophy type 1. *Neuromuscul. Disord* **21**: 81-88
- Peter S Harper (1989) Myotonic Dystrophy 2^e éd.
- Peter S Harper (2001) Myotonic Dystrophy, 3rd Edition
- Philips AV, Timchenko LT & Cooper TA (1998) Disruption of Splicing Regulated by a CUG-Binding Protein in Myotonic Dystrophy. *Science* **280**: 737 -741
- Pugh GE, Nicol SM & Fuller-Pace FV (1999) Interaction of the Escherichia coli DEAD box protein DbpA with 23 S ribosomal RNA. *Journal of Molecular Biology* **292**: 771-778
- Puig O, Gottschalk A, Fabrizio P & Séraphin B (1999) Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. *Genes & Development* **13**: 569 -580
- Query CC & Konarska MM (2006) Splicing fidelity revisited. *Nat. Struct. Mol. Biol* **13**: 472-474
- Ranum LP, Rasmussen PF, Benzow KA, Koob MD & Day JW (1998) Genetic mapping of a second myotonic dystrophy locus. *Nat. Genet* **19**: 196-198
- Ranum LPW & Cooper TA (2006) RNA-Mediated Neuromuscular Disorders. *Annu. Rev. Neurosci* Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16704370> [Consulté avril 29, 2011]
- Ranum LPW & Day JW (2004) Myotonic Dystrophy: RNA Pathogenesis Comes into Focus. *Am J Hum Genet* **74**: 793-804
- Rattenbacher B, Beisang D, Wiesner DL, Jeschke JC, von Hohenberg M, St Louis-Vlasova IA & Bohjanen PR (2010) Analysis of CUGBP1 targets identifies GU-repeat sequences that mediate rapid mRNA decay. *Mol. Cell. Biol* **30**: 3970-3980
- 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 JW, Fujimura H, Takahashi MP, Auboeuf D, Dreumont N, Furling D & Charlet-Berguerand N (2011) Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat. Struct. Mol. Biol* **18**: 840-845
- Reddy S, Smith DB, Rich MM, Leferovich JM, Reilly P, Davis BM, Tran K, Rayburn H,

- Bronson R, Cros D, Balice-Gordon RJ & Housman D (1996) Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy. *Nat. Genet* **13**: 325-335
- Reed R (1996) Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr. Opin. Genet. Dev* **6**: 215-220
- Reed R (2000) Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol* **12**: 340-345
- Ridley AJ (2001) Rho family proteins: coordinating cell responses. *Trends Cell Biol* **11**: 471-477
- Ring HZ & Lis JT (1994) The SR protein B52/SRp55 is essential for Drosophila development. *Mol. Cell. Biol* **14**: 7499-7506
- Roberts GC, Gooding C, Mak HY, Proudfoot NJ & Smith CW (1998) Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res* **26**: 5568-5572
- Robson-Dixon ND & Garcia-Blanco MA (2004) MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb. *J. Biol. Chem* **279**: 29075-29084
- Rocak S & Linder P (2004) DEAD-box proteins: the driving forces behind RNA metabolism. *Nat. Rev. Mol. Cell Biol* **5**: 232-241
- Rocak S, Emery B, Tanner NK & Linder P (2005) Characterization of the ATPase and unwinding activities of the yeast DEAD-box protein Has1p and the analysis of the roles of the conserved motifs. *Nucleic Acids Res* **33**: 999-1009
- Rogers GW Jr, Komar AA & Merrick WC (2002) eIF4A: the godfather of the DEAD box helicases. *Prog. Nucleic Acid Res. Mol. Biol* **72**: 307-331
- Roscigno RF & Garcia-Blanco MA (1995) SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA* **1**: 692-706
- Rosenberger U, Lehmann I, Weise C, Franke P, Hucho F & Buchner K (2002) Identification of PSF as a protein kinase Calpha-binding protein in the cell nucleus. *J. Cell. Biochem* **86**: 394-402
- Rössler OG, Hloch P, Schütz N, Weitzenegger T & Stahl H (2000) Structure and expression of the human p68 RNA helicase gene. *Nucleic Acids Res* **28**: 932-939
- Rössler OG, Straka A & Stahl H (2001) Rearrangement of structured RNA via branch migration structures catalysed by the highly related DEAD-box proteins p68 and p72. *Nucleic Acids Res* **29**: 2088-2096
- Rossow KL & Janknecht R (2003) Synergism between p68 RNA helicase and the transcriptional coactivators CBP and p300. *Oncogene* **22**: 151-156
- Saint-André V, Batsché E, Rachez C & Muchardt C (2011) Histone H3 lysine 9 trimethylation and HP1 γ favor inclusion of alternative exons. *Nat. Struct. Mol. Biol* **18**: 337-344
- Salisbury E, Sakai K, Schoser B, Huichalaf C, Schneider-Gold C, Nguyen H, Wang G-L, Albrecht JH & Timchenko LT (2008) Ectopic expression of cyclin D3 corrects differentiation of DM1 myoblasts through activation of RNA CUG-binding protein, CUGBP1. *Exp. Cell Res* **314**: 2266-2278
- Salvatori S, Fanin M, Trevisan CP, Furlan S, Reddy S, Nagy JI & Angelini C (2005) Decreased expression of DMPK: correlation with CTG repeat expansion and fibre type composition in myotonic dystrophy type 1. *Neurol. Sci* **26**: 235-242

Salzman DW, Shubert-Coleman J & Furneaux H (2007) P68 RNA Helicase Unwinds the Human let-7 MicroRNA Precursor Duplex and Is Required for let-7-directed Silencing of Gene Expression. *Journal of Biological Chemistry* **282**: 32773 -32779

Sanford JR, Gray NK, Beckmann K & Cáceres JF (2004) A novel role for shuttling SR proteins in mRNA translation. *Genes & Development* **18**: 755 -768

Sankaranarayanan R, Dock-Bregeon A-C, Romby P, Caillet J, Springer M, Rees B, Ehresmann C, Ehresmann B & Moras D (1999) The Structure of Threonyl-tRNA Synthetase-tRNA^{Thr} Complex Enlightens Its Repressor Activity and Reveals an Essential Zinc Ion in the Active Site. *Cell* **97**: 371-381

Sarkar PS, Appukuttan B, Han J, Ito Y, Ai C, Tsai W, Chai Y, Stout JT & Reddy S (2000) Heterozygous loss of Six5 in mice is sufficient to cause ocular cataracts. *Nat. Genet* **25**: 110-114

Sarkar PS, Paul S, Han J & Reddy S (2004) Six5 is required for spermatogenic cell survival and spermiogenesis. *Hum. Mol. Genet* **13**: 1421-1431

Sasagawa N, Ohno E, Kino Y, Watanabe Y & Ishiura S (2009) Identification of *Caenorhabditis elegans* K02H8.1 (CeMBL), a functional ortholog of mammalian MBNL proteins. *J. Neurosci. Res* **87**: 1090-1097

Sauliere J, Sureau A, Expert-Bezancon A & Marie J (2006) The Polypyrimidine Tract Binding Protein (PTB) Represses Splicing of Exon 6B from the {beta}-Tropomyosin Pre-mRNA by Directly Interfering with the Binding of the U2AF65 Subunit. *Mol. Cell. Biol.* **26**: 8755-8769

Saveanu C, Namane A, Gleizes P-E, Lebreton A, Rousselle J-C, Noaillac-Depeyre J, Gas N, Jacquier A & Fromont-Racine M (2003) Sequential Protein Association with Nascent 60S Ribosomal Particles. *Mol. Cell. Biol.* **23**: 4449-4460

Savkur RS, Philips AV & Cooper TA (2001) Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat. Genet* **29**: 40-47

Savouret C, Garcia-Cordier C, Megret J, te Riele H, Junien C & Gourdon G (2004) MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. *Mol. Cell. Biol* **24**: 629-637

Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE & Zipursky SL (2000) *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**: 671-684

Schneider S, Campodonico E & Schwer B (2004) Motifs IV and V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8. *J. Biol. Chem* **279**: 8617-8626

Schor IE, Rascovan N, Pelisch F, Alló M & Kornblihtt AR (2009) Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl. Acad. Sci. U.S.A* **106**: 4325-4330

Schwartz S, Meshorer E & Ast G (2009) Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol* **16**: 990-995

Schwer B & Meszaros T (2000) RNA helicase dynamics in pre-mRNA splicing. *EMBO J* **19**: 6582-6591

Sellier C, Rau F, Liu Y, Tassone F, Hukema RK, Gattoni R, Schneider A, Richard S,

Willemsen R, Elliott DJ, Hagerman PJ & Charlet-Berguerand N (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J* **29**: 1248-1261

Seraphin B & Rosbash M (1989) Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**: 349-358

Sergeant N, Sablonnière B, Schraen-Maschke S, Ghestem A, Maurage CA, Wattez A, Vermersch P & Delacourte A (2001) Dysregulation of human brain microtubule-associated tau mRNA maturation in myotonic dystrophy type 1. *Hum. Mol. Genet* **10**: 2143-2155

Seznec H, Agbulut O, Sergeant N, Savouret C, Ghestem A, Tabti N, Willer JC, Ourth L, Duros C, Brisson E, Fouquet C, Butler-Browne G, Delacourte A, Junien C & Gourdon G (2001) Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities. *Hum. Mol. Genet* **10**: 2717-2726

Seznec H, Lia-Baldini AS, Duros C, Fouquet C, Lacroix C, Hofmann-Radvanyi H, Junien C & Gourdon G (2000) Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Hum. Mol. Genet* **9**: 1185-1194

Sharp PA (1994) Split genes and RNA splicing. *Cell* **77**: 805-815

Sharp PA & Burge CB (1997) Classification of introns: U2-type or U12-type. *Cell* **91**: 875-879

Shepard PJ & Hertel KJ (2008) Conserved RNA secondary structures promote alternative splicing. *RNA* **14**: 1463-1469

Shi H, Cordin O, Minder CM, Linder P & Xu R-M (2004) Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proc. Natl. Acad. Sci. U.S.A* **101**: 17628-17633

Shi Y, Downes M, Xie W, Kao HY, Ordentlich P, Tsai CC, Hon M & Evans RM (2001) Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev* **15**: 1140-1151

Shimizu M, Wang W, Walch ET, Dunne PW & Epstein HF (2000) Rac-1 and Raf-1 kinases, components of distinct signaling pathways, activate myotonic dystrophy protein kinase. *FEBS Lett* **475**: 273-277

Shin S & Janknecht R (2007) Concerted activation of the Mdm2 promoter by p72 RNA helicase and the coactivators p300 and P/CAF. *J. Cell. Biochem* **101**: 1252-1265

Shin S, Rossow KL, Grande JP & Janknecht R (2007) Involvement of RNA helicases p68 and p72 in colon cancer. *Cancer Res* **67**: 7572-7578

Sickmier EA, Frato KE, Shen H, Paranawithana SR, Green MR & Kielkopf CL (2006) Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol. Cell* **23**: 49-59

Singh NN, Singh RN & Androphy EJ (2007) Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res* **35**: 371-389

Singleton MR, Dillingham MS & Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem* **76**: 23-50

Sirand-Pugnet P, Durosay P, Clouet d'Orval BC, Brody E & Marie J (1995) beta-Tropomyosin pre-mRNA folding around a muscle-specific exon interferes with several steps

of spliceosome assembly. *J. Mol. Biol* **251**: 591-602

Skrisovska L, Bourgeois CF, Stefl R, Grellscheid S-N, Kister L, Wenter P, Elliott DJ, Stevenin J & Allain FH-T (2007) The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep* **8**: 372-379

Smith CA & Rayment I (1996) Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biophys. J* **70**: 1590-1602

Smith KP, Byron M, Johnson C, Xing Y & Lawrence JB (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**: 951-964

Sorek R, Lev-Maor G, Reznik M, Dagan T, Belinky F, Graur D & Ast G (2004a) Minimal conditions for exonization of intronic sequences: 5' splice site formation in alu exons. *Mol. Cell* **14**: 221-231

Sorek R, Shamir R & Ast G (2004b) How prevalent is functional alternative splicing in the human genome? *Trends Genet* **20**: 68-71

Spector DL (1993) Nuclear organization of pre-mRNA processing. *Curr. Opin. Cell Biol* **5**: 442-447

Spies N, Nielsen CB, Padgett RA & Burge CB (2009) Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* **36**: 245-254

Spits C, Seneca S, Hilven P, Liebaers I & Sermon K (2010) Methylation of the CpG sites in the myotonic dystrophy locus does not correlate with CTG expansion size or with the congenital form of the disease. *J. Med. Genet* **47**: 700-703

Squillace RM, Chenault DM & Wang EH (2002) Inhibition of muscle differentiation by the novel muscleblind-related protein CHCR. *Dev. Biol* **250**: 218-230

Srinivasan K, Shiue L, Hayes JD, Centers R, Fitzwater S, Loewen R, Edmondson LR, Bryant J, Smith M, Rommelfanger C, Welch V, Clark TA, Sugnet CW, Howe KJ, Mandel-Gutfreund Y & Ares M Jr (2005) Detection and measurement of alternative splicing using splicing-sensitive microarrays. *Methods* **37**: 345-359

Stamm S, Zhang MQ, Marr TG & Helfman DM (1994) A sequence compilation and comparison of exons that are alternatively spliced in neurons. *Nucleic Acids Res* **22**: 1515-1526

Steinbach P, Gläser D, Vogel W, Wolf M & 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**: 278-285

Steinert H (1909) Myopathologische Beiträge : über das klinische und anatomische Bild des Muskelschwunds der Myotoniker. *Deutsche Z. für Nervenheilkunde* **XXXVII**:

Story RM, Li H & Abelson JN (2001) Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*. *Proc. Natl. Acad. Sci. U.S.A* **98**: 1465-1470

Strohmeier J, Hertel I, Diederichsen U, Rudolph MG & Klostermeier D (2011) Changing nucleotide specificity of the DEAD-box helicase Hera abrogates communication between the Q-motif and the P-loop. *Biol. Chem* **392**: 357-369

Stuart CA, Armstrong RM, Provow SA & Plishker GA (1983) Insulin resistance in patients with myotonic dystrophy. *Neurology* **33**: 679-685

Sureau A, Gattoni R, Dooghe Y, Stévenin J & Soret J (2001) SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J* **20**: 1785-1796

Tacke R & Manle JL (1999) Functions of SR and Tra2 Proteins in Pre-mRNA Splicing Regulation. *Proceedings of the Society for Experimental Biology and Medicine* **220**: 59-63

Taneja KL, McCurrach M, Schalling M, Housman D & Singer RH (1995) Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol* **128**: 995-1002

Tanner NK, Cordin O, Banroques J, Doère M & Linder P (2003) The Q Motif: A Newly Identified Motif in DEAD Box Helicases May Regulate ATP Binding and Hydrolysis. *Molecular Cell* **11**: 127-138

Tarn WY, Yario TA & Steitz JA (1995) U12 snRNA in vertebrates: evolutionary conservation of 5' sequences implicated in splicing of pre-mRNAs containing a minor class of introns. *RNA* **1**: 644-656

Teplova M & Patel DJ (2008) Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat Struct Mol Biol* **15**: 1343-1351

Thornton CA, Wymer JP, Simmons Z, McClain C & Moxley RT 3rd (1997) Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nat. Genet* **16**: 407-409

Tian B, White RJ, Xia T, Welle S, Turner DH, Mathews MB & Thornton CA (2000) Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA* **6**: 79-87

Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcárcel J & Guigó R (2009) Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol* **16**: 996-1001

Tillo D, Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Field Y, Lieb JD, Widom J, Segal E & Hughes TR (2010) High nucleosome occupancy is encoded at human regulatory sequences. *PLoS ONE* **5**: e9129

Timchenko LT, Miller JW, Timchenko NA, DeVore DR, Datar KV, Lin L, Roberts R, Caskey CT & Swanson MS (1996) Identification of a (CUG)_n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res* **24**: 4407-4414

Timchenko NA, Cai ZJ, Welm AL, Reddy S, Ashizawa T & Timchenko LT (2001) RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. *J. Biol. Chem* **276**: 7820-7826

Timchenko NA, Patel R, Iakova P, Cai Z-J, Quan L & Timchenko LT (2004) Overexpression of CUG triplet repeat-binding protein, CUGBP1, in mice inhibits myogenesis. *J. Biol. Chem* **279**: 13129-13139

Timchenko NA, Welm AL, Lu X & Timchenko LT (1999) CUG repeat binding protein (CUGBP1) interacts with the 5' region of C/EBPbeta mRNA and regulates translation of C/EBPbeta isoforms. *Nucleic Acids Res* **27**: 4517-4525

Tomé S, Holt I, Edelmann W, Morris GE, Munnich A, Pearson CE & Gourdon G (2009) MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice. *PLoS Genet* **5**: e1000482

Le Tonquèze O, Gschloessl B, Namanda-Vanderbeken A, Legagneux V, Paillard L & Audic Y (2010) Chromosome wide analysis of CUGBP1 binding sites identifies the tetraspanin CD9 mRNA as a target for CUGBP1-mediated down-regulation. *Biochem. Biophys. Res. Commun* **394**: 884-889

Tran EJ, Zhou Y, Corbett AH & Wente SR (2007) The DEAD-Box Protein Dbp5 Controls mRNA Export by Triggering Specific RNA:Protein Remodeling Events. *Molecular Cell* **28**: 850-859

Tran H, Gourrier N, Lemercier-Neuillet C, Dhaenens C-M, Vautrin A, Fernandez-Gomez FJ, Arandel L, Carpentier C, Obriot H, Eddarkaoui S, Delattre L, Van Brussels E, Holt I, Morris GE, Sablonnière B, Buée L, Charlet-Berguerand N, Schraen-Maschke S, Furling D, Behm-Ansmant I, *et al.* (2011) Analysis of exonic regions involved in nuclear localization, splicing activity, and dimerization of Muscleblind-like-1 isoforms. *J. Biol. Chem* **286**: 16435-16446

Tress ML, Bodenmiller B, Aebersold R & Valencia A (2008) Proteomics studies confirm the presence of alternative protein isoforms on a large scale. *Genome Biol* **9**: R162

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG & Prasanth KV (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**: 925-938

Tsay JM, Sippy J, Feiss M & Smith DE (2009) The Q motif of a viral packaging motor governs its force generation and communicates ATP recognition to DNA interaction. *Proc. Natl. Acad. Sci. U.S.A* **106**: 14355-14360

Uhlmann-Schiffler H, Jalal C & Stahl H (2006) Ddx42p—a human DEAD box protein with RNA chaperone activities. *Nucleic Acids Res* **34**: 10-22

Uhlmann-Schiffler H, Rössler OG & Stahl H (2002) The mRNA of DEAD box protein p72 is alternatively translated into an 82-kDa RNA helicase. *J. Biol. Chem* **277**: 1066-1075

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A & Darnell RB (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**: 1212-1215

Valcárcel J, Singh R, Zamore PD & Green MR (1993) The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* **362**: 171-175

Vignaud A, Ferry A, Huguet A, Baraibar M, Trollet C, Hyzewicz J, Butler-Browne G, Puymirat J, Gourdon G & Furling D (2010) Progressive skeletal muscle weakness in transgenic mice expressing CTG expansions is associated with the activation of the ubiquitin-proteasome pathway. *Neuromuscul. Disord* **20**: 319-325

Vlasova IA, Tahoe NM, Fan D, Larsson O, Rattenbacher B, Sternjohn JR, Vasdewani J, Karypis G, Reilly CS, Bitterman PB & Bohjanen PR (2008) Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1. *Mol. Cell* **29**: 263-270

Vogt B, Carrascosa JM, Ermel B, Ullrich A & Häring H-U (1991) The two isoforms of the human insulin receptor (HIR-A and HIR-B) follow different internalization kinetics. *Biochemical and Biophysical Research Communications* **177**: 1013-1018

Wahl MC, Will CL & Lührmann R (2009) The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **136**: 701-718

Walke S, Bragado-Nilsson E, Séraphin B & Nagai K (2001) Stoichiometry of the Sm proteins in yeast spliceosomal snRNPs supports the heptamer ring model of the core domain. *J. Mol. Biol* **308**: 49-58

Walker JE, Saraste M, Runswick MJ & Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**: 945-951

Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP & Burge CB (2008a) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470-476

Wang G-S, Kearney DL, De Biasi M, Taffet G & Cooper TA (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**: 2802-2811

Wang H, Gao X, Huang Y, Yang J & Liu Z-R (2009) P68 RNA helicase is a nucleocytoplasmic shuttling protein. *Cell Res* **19**: 1388-1400

Wang J, Pegoraro E, Menegazzo E, Gennarelli M, Hoop RC, Angelini C & Hoffman EP (1995) Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation. *Hum. Mol. Genet* **4**: 599-606

Wang L-C, Chen K-Y, Pan H, Wu C-C, Chen P-H, Liao Y-T, Li C, Huang M-L & Hsiao K-M (2011) Muscleblind participates in RNA toxicity of expanded CAG and CUG repeats in *Caenorhabditis elegans*. *Cell. Mol. Life Sci* **68**: 1255-1267

Wang L-C, Hung W-T, Pan H, Chen K-Y, Wu Y-C, Liu Y-F & Hsiao K-M (2008b) Growth-dependent effect of muscleblind knockdown on *Caenorhabditis elegans*. *Biochem. Biophys. Res. Commun* **366**: 705-709

Wang Y-H (2007) Chromatin structure of repeating CTG/CAG and CGG/CCG sequences in human disease. *Front. Biosci* **12**: 4731-4741

Wang Z & Burge CB (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**: 802-813

Wang Z, Rolish ME, Yeo G, Tung V, Mawson M & Burge CB (2004) Systematic identification and analysis of exonic splicing silencers. *Cell* **119**: 831-845

Wansink DG, van Herpen REMA, Coerwinkel-Driessen MM, Groenen PJTA, Hemmings BA & Wieringa B (2003) Alternative splicing controls myotonic dystrophy protein kinase structure, enzymatic activity, and subcellular localization. *Mol. Cell. Biol* **23**: 5489-5501

Ward AJ, Rimer M, Killian JM, Dowling JJ & Cooper TA (2010) CUGBP1 overexpression in mouse skeletal muscle reproduces features of myotonic dystrophy type 1. *Hum. Mol. Genet* **19**: 3614-3622

Warf MB & Berglund JA (2007) MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA* **13**: 2238-2251

Warf MB & Berglund JA (2010) Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci* **35**: 169-178

Warf MB, Diegel JV, von Hippel PH & Berglund JA (2009) The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc. Natl. Acad. Sci. U.S.A* **106**: 9203-9208

Watanabe M, Yanagisawa J, Kitagawa H, Takeyama K-ichi, Ogawa S, Arao Y, Suzawa M, Kobayashi Y, Yano T, Yoshikawa H, Masuhiro Y & Kato S (2001) A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor [alpha] coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. *EMBO J* **20**: 1341-1352

Watanabe Y-ichi, Yokobori S-ichi, Inaba T, Yamagishi A, Oshima T, Kawarabayasi Y, Kikuchi H & Kita K (2002) Introns in protein-coding genes in Archaea. *FEBS Lett* **510**: 27-30

Wheeler TM, Lueck JD, Swanson MS, Dirksen RT & Thornton CA (2007) Correction of CLC-1 splicing eliminates chloride channelopathy and myotonia in mouse models of myotonic

dystrophy. *J. Clin. Invest* **117**: 3952-3957

Wilkinson KA, Merino EJ & Weeks KM (2006) Selective 2[prime]-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protocols* **1**: 1610-1616

Will CL & Lührmann R (2010) Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*

Wilson BJ & Giguère V (2007) Identification of novel pathway partners of p68 and p72 RNA helicases through Oncomine meta-analysis. *BMC Genomics* **8**: 419

Wilson BJ, Bates GJ, Nicol SM, Gregory DJ, Perkins ND & Fuller-Pace FV (2004) The p68 and p72 DEAD box RNA helicases interact with HDAC1 and repress transcription in a promoter-specific manner. *BMC Mol. Biol* **5**: 11

Wojciechowska M & Krzyzosiak WJ (2011) Cellular toxicity of expanded RNA repeats: Focus on RNA foci. *Hum Mol Genet*

Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA & Smith CWJ (2004) Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* **13**: 91-100

Wood V, Gwilliam R, Rajandream M-A, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A, *et al.* (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**: 871-880

Wortham NC, Ahamed E, Nicol SM, Thomas RS, Periyasamy M, Jiang J, Ochocka AM, Shousha S, Huson L, Bray SE, Coombes RC, Ali S & Fuller-Pace FV (2009) The DEAD box protein p72 regulates ER α -/Estrogen-dependent transcription and cell growth, and is associated with improved survival in ER α positive breast cancer. *Oncogene* **28**: 4053-4064

Wu JY & Maniatis T (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**: 1061-1070

Yadava RS, Frenzel-McCardell CD, Yu Q, Srinivasan V, Tucker AL, Puymirat J, Thornton CA, Prall OW, Harvey RP & Mahadevan MS (2008) RNA toxicity in myotonic muscular dystrophy induces NKX2-5 expression. *Nat. Genet* **40**: 61-68

Yang L, Lin C & Liu Z-R (2005a) Phosphorylations of DEAD box p68 RNA helicase are associated with cancer development and cell proliferation. *Mol. Cancer Res* **3**: 355-363

Yang L, Lin C & Liu Z-R (2005b) Signaling to the DEAD box--regulation of DEAD-box p68 RNA helicase by protein phosphorylations. *Cell. Signal* **17**: 1495-1504

Yang L, Lin C & Liu Z-R (2006) P68 RNA helicase mediates PDGF-induced epithelial mesenchymal transition by displacing Axin from beta-catenin. *Cell* **127**: 139-155

Yang L, Lin C, Zhao S, Wang H & Liu Z-R (2007) Phosphorylation of p68 RNA helicase plays a role in platelet-derived growth factor-induced cell proliferation by up-regulating cyclin D1 and c-Myc expression. *J. Biol. Chem* **282**: 16811-16819

Yang L, Yang J, Huang Y & Liu ZR (2004) Phosphorylation of p68 RNA helicase regulates RNA binding by the C-terminal domain of the protein. *Biochem. Biophys. Res. Commun* **314**: 622-630

Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD & Felsenfeld G (2010) Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes & Development* **24**: 2543 -2555

- Yotova V, Labuda D, Zietkiewicz E, Gehl D, Lovell A, Lefebvre J-F, Bourgeois S, Lemieux-Blanchard É, Labuda M, Vézina H, Houde L, Tremblay M, Toupance B, Heyer E, Hudson TJ & Laberge C (2005) Anatomy of a founder effect: myotonic dystrophy in Northeastern Quebec. *Hum Genet* **117**: 177-187
- Yuan Y, Compton SA, Sobczak K, Stenberg MG, Thornton CA, Griffith JD & Swanson MS (2007) Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res* **35**: 5474-5486
- Zhang D & Rosbash M (1999) Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes & Development* **13**: 581 -592
- Zhang Q-S, Manche L, Xu R-M & Krainer AR (2006) hnRNP A1 associates with telomere ends and stimulates telomerase activity. *RNA* **12**: 1116 -1128
- Zhang XH-F & Chasin LA (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev* **18**: 1241-1250
- Zhang Z & Krainer AR (2004) Involvement of SR Proteins in mRNA Surveillance. *Molecular Cell* **16**: 597-607
- Zhou H-L & Lou H (2008) Repression of prespliceosome complex formation at two distinct steps by Fox-1/Fox-2 proteins. *Mol. Cell. Biol* **28**: 5507-5516
- Zhou Z, Licklider LJ, Gygi SP & Reed R (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**: 182-185
- Zhu J & Krainer AR (2000) Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes & Development* **14**: 3166 -3178
- Zuo P & Maniatis T (1996) The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes & Development* **10**: 1356 -1368

ANNEXES

ANNEXES

Au cours de ma thèse, j'ai établi une collaboration avec le laboratoire de Jean Cohen au Centre de Génétique Moléculaire à Gif sur Yvette.

Le laboratoire a identifié la protéine Bug22, une nouvelle protéine ciliaire chez *Paramecium tetraurelia*. Ce facteur est localisée au niveau des corps basaux ciliaires chez *P. tetraurelia* et intervient dans la motilité des cils. Une protéine humaine et orthologue à Bug22 existe chez l'humain et l'équipe souhaitait analyser la localisation cellulaire de cette protéine dans des cellules humaines. Ayant développé l'étude de la localisation protéique dans les cellules de mammifères au cours de ma thèse, j'ai aidé à la construction d'un plasmide eucaryote contenant la séquence cDNA de *Bug22* fusionnée à la GFP et analysé la localisation de cette protéine chimère par immunofluorescence dans des cellules HeLa.

Ces résultats ont fait l'objet d'un article soumis à *Eucaryotic Cell*. Cette publication est intitulée « Bug22p, a Conserved Centrosomal/Ciliary Protein Also Present in Higher Plants, Is Required for an Effective Ciliary Stroke in Paramecium ».

Bug22p, a Conserved Centrosomal/Ciliary Protein Also Present in Higher Plants, Is Required for an Effective Ciliary Stroke in *Paramecium*^{∇†}

C. Laligné,^{1,2} C. Klotz,^{1,2} N. Garreau de Loubresse,^{1,2} M. Lemullois,³ M. Hori,⁴
F. X. Laurent,^{1,2} J. F. Papon,^{5,6,7,8} B. Louis,⁵ J. Cohen,^{1,2} and F. Koll^{1,2*}

Centre de Génétique Moléculaire, CNRS, 91198 Gif-sur-Yvette Cedex, France¹; Université Paris-Sud, 91405 Orsay, France²; Laboratoire de Biologie Cellulaire 4, Université Paris-Sud, 91405 Orsay Cedex, France³; Division of Environmental Science and Engineering, Graduate School of Science and Engineering, Yamaguchi University, Yamaguchi, Japan⁴; INSERM Unit U955, Créteil, F-94010, France⁵; INSERM Unit U933, Paris, F-75012, France⁶; AP-HP, Groupe Henri Mondor-Albert Chenevier, Service d'ORL et de Chirurgie Cervico-Faciale, Créteil, F-94010, France⁷; and Hôpital Intercommunal, service d'ORL et de Chirurgie Cervico-Faciale, Créteil, F-94010, France⁸

Received 11 December 2009/Accepted 22 January 2010

Centrioles, cilia, and flagella are ancestral conserved organelles of eukaryotic cells. Among the proteins identified in the proteomics of ciliary proteins in *Paramecium*, we focus here on a protein, Bug22p, previously detected by cilia and basal-body high-throughput studies but never analyzed *per se*. Remarkably, this protein is also present in plants, which lack centrioles and cilia. Bug22p sequence alignments revealed consensus positions that distinguish species with centrioles/cilia from plants. In *Paramecium*, antibody and green fluorescent protein (GFP) fusion labeling localized Bug22p in basal bodies and cilia, and electron microscopy immunolabeling refined the localization to the terminal plate of the basal bodies, the transition zone, and spots along the axoneme, preferentially between the membrane and the microtubules. RNA interference (RNAi) depletion of Bug22p provoked a strong decrease in swimming speed, followed by cell death after a few days. High-speed video microscopy and morphological analysis of Bug22p-depleted cells showed that the protein plays an important role in the efficiency of ciliary movement by participating in the stroke shape and rigidity of cilia. The defects in cell swimming and growth provoked by RNAi can be complemented by expression of human Bug22p. This is the first reported case of complementation by a human gene in a ciliate.

The centriole, a microtubule-based cylindrical structure, is an ancient organelle traced to primitive eukaryotes and present in numerous phyla but lost in some others, such as fungi and higher plants (24). Centriolar function is dual: organization of the centrosome and, as a basal body, nucleation of an axoneme, the backbone of cilia and flagella. The ultrastructure of centrioles and axonemes appears to be highly conserved. In the majority of cases, the centriole is composed of 9 triplets of microtubules, and when the centriole acts as a basal body, the triplets are elongated into 9 doublets of microtubules forming the axoneme, with occurrence of a central pair of microtubules (9-plus-2 pattern) (reviewed by Satir and Christensen [30]). Renewal of interest in cilia in recent years has been prompted by their implication in numerous functions in development and cell physiology through their motile and sensory properties, so that many human genetic diseases, now called ciliopathies, appear to arise from primary defects in ciliary proteins (14). The ultrastructural conservation of cent-

rioles and cilia in evolution is correlated with good conservation of the protein constituents.

High-throughput studies (proteomics, transcriptome analysis, and comparative genomics) have been conducted in several species to identify ciliary and centriolar proteins or proteins involved in their biogenesis. The set of proteins identified comprises both proteins specific to cilia (such as outer- and inner-arm dynein subunits or radial-spoke proteins) or to their biogenesis (specific transcription factors and intraflagellar-transport [IFT] proteins) and ubiquitous proteins (such as tubulins).

In order to identify novel conserved proteins and to specify their roles in ciliary function, we built a new knowledge base, Cildb (1) (<http://cildb.cgm.cnrs-gif.fr>), which integrates orthology links between the whole proteomes of 18 species and high-throughput studies identifying ciliary proteins, as well as the OMIM database. Cildb includes species for which no ciliary studies are available and species that lack centrioles or cilia in order to distinguish proteins specific to these organelles from ubiquitous proteins present in them.

In the present work, starting from the ciliary proteome analyzed in *Paramecium* (1), the Bug22p protein, highly conserved through evolution, was selected. Apart from its identification in high-throughput studies of various species (6, 8, 17, 21, 23, 25, 32, 36), no analysis dedicated to Bug22p exists to date for any organism. This protein has amazing sequence

* Corresponding author. Mailing address: Centre de Génétique Moléculaire, CNRS, Ave. de la Terrasse, 91198 Gif-sur-Yvette Cedex, France. Phone: 33169823145. Fax: 33169823181. E-mail: koll@cgm.cnrs-gif.fr.

† Supplemental material for this article may be found at <http://ec.asm.org/>.

[∇] Published ahead of print on 29 January 2010.

conservation, indicating a key role in cellular processes and/or the existence of numerous interacting partners. Intriguingly, the protein is also present in higher plants (e.g., *Arabidopsis*), a phylum in which centriolar and ciliary structures are absent, which induced us to analyze the relevance of its ciliary function.

We report here the localization of Bug22p in HeLa and *Paramecium* cells and its functional study in *Paramecium*. The choice of this organism, beyond the power of functional analysis by RNA interference (RNAi) (13), stems from the fact that we can specifically analyze the motility aspect of ciliary function.

MATERIALS AND METHODS

Strains and culture conditions. Stock d4-2 of *Paramecium tetraurelia*, the wild-type reference strain, was used in all RNAi experiments. The mutant nd7-1 (31), which carries a recessive monogenic mutation preventing trichocyst discharge, a dispensable function under laboratory conditions, was used for transformations. Cells were grown at 27°C in wheat grass infusion, BHB (L'arbre de vie, Luçay Le Mâle, France), inoculated with *Klebsiella pneumoniae* and supplemented with 0.8 µg/ml β-sitosterol according to standard procedures (33).

Expression vectors. In order to visualize the cellular localization of Bug22p, the *Bug22a* *paramecium* gene, as well as the human (Hs) Bug22 cDNA, were cloned into the *Paramecium* expression vector pPXV-GFP. The human Bug22 cDNA was also cloned into the plasmid pD2EGFP for expression in HeLa cells. In all three constructs, the green fluorescent protein (GFP) sequence was fused at the N terminus of the *BUG22* gene.

(i) **pPXV-HsBUG22 expression vector for *Paramecium*.** The human *BUG22* sequence was amplified by PCR from a cDNA library kindly provided by François Lacroute using the primers 5'-GGTGCTCGAGATGTTCAAAAACACGT TCCAG-3' and 5'-GTGGGGTACCTCATTGCTTTGCCTTGTCTTG-3' and cloned into the XhoI/KpnI sites of the *Paramecium* expression vector pPXV.

(ii) **GFP-PtBUG22 fusion gene to be expressed in *Paramecium*.** The pPXV-GFP expression vector was used as previously described. The *Paramecium* (Pt) *BUG22c* gene was amplified by PCR with primers 5'-GTGGGGTACCGGAG GAGGAATGTTCAAGAACTCATTCCAAAG-3' and 5'-CCACGGTACCTC ATCCTTATTTTTAGATGGGTAAG-3' and cloned into the KpnI site of pPXV-GFP.

(iii) **GFP-HsBUG22 fusion gene to be expressed in *Paramecium*.** The human sequence of *BUG22* was amplified by PCR with primers 5'-GTGGGGTACCG GAGGAGGAATGTTCAAAAACACGTTCAGAG-3' and 5'-CCACGGTACCTCCTCATTGCTTTGCCCTTGTCTGAAC-3' and cloned into the KpnI site of pPXV-GFP.

RNAi vectors. The sequences of genes to be knocked down were amplified by PCR and cloned into a modified Litmus 28i plasmid (New England Biolabs) in which the EcoRI site had been replaced by an SrfI site, using the method developed for PCRscript cloning by Stratagene (La Jolla, CA), and transferred to the feeding vector L4440 (37) into the XhoI/HindIII sites of the polylinker between two T7 promoters.

(i) **Constructs for RNAi of the *Paramecium* *BUG22* genes.** Two constructs were used to knock down the two subfamilies of *BUG22* genes. For *BUG22a* and *-b* knockdowns, *BUG22a* was amplified from positions 75 to 391 with the primers 5'-GGATAAGCAAAGTAATTTATGATTGG-3' and 5'-CTTTATACTAATT ACTTAATAATTCGATGC-3'. For *BUG22c* and *-d* knockdowns, *BUG22d* was amplified from positions 75 to 394 with the primers 5'-TCAAAGCCCTTGTA GATTG-3' and 5'-GTGTTTTAAGATGTACTTGATAATTAG-3'. A short sequence was also used for RNAi by direct cloning of a pair of complementary oligonucleotides corresponding to the sequence 5'-ATGTTCAAGAACTCATT CCAAAGTGGATTCTTTCAATTTTATATTAATAGGATCAAAGCC-3'.

(ii) **Constructs for RNAi of human *BUG22* transgenes in *Paramecium*.** The HsBUG22 sequence was transferred from the pPXV-HsBUG22 construct by digestion and cloning into the XhoI/HindIII sites of the L4440 vector.

(iii) **Constructs for RNAi of the *IFT172* genes in *Paramecium*.** One of the two close paralogous *IFT172* genes (GSPATG00033191001) was amplified by PCR using the primers 5'-GCAGTCATTAGATCATCATAGG-3' and 5'-CAATA- ACTTCACTTTCAGGAAC-3' and cloned into the L4440 vector as the *BUG22* sequences. Due to high sequence conservation, this single plasmid is supposed to knock down both paralogs.

In all cases, plasmid DNA was prepared using the plasmid midi kit according

to the protocol of the manufacturer (Qiagen, Courtaboeuf, France). The sequences of *Paramecium* genes were retrieved from *ParameciumDB* (2) (<http://paramecium.cgm.cnrs-gif.fr/>) and verified by resequencing after cloning.

Transfection. (i) **HeLa cells.** HeLa cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin (all from Gibco) at 37°C in 5% CO₂. The day prior to transfection, HeLa cells (1 × 10⁵) were plated on a 24-well plate containing presterilized coverslips in DMEM supplemented with 10% FBS. The cells were transfected with 0.5 µg of Hs-Bug22-GFP using Lipofectamine 2000 (Invitrogen) according to the manufacturer's directions.

(ii) ***Paramecium* transformation.** nd7-1 cells were transformed by microinjection into their macronuclei of filtered and concentrated plasmid DNA (5 µg/µl) containing a mixture at a ratio of 10:1 of the plasmids of interest (GFP fusion genes and HsBug22p expression) and a plasmid directing the expression of the *ND7* gene (31). Microinjection was done under an inverted Nikon phase-contrast microscope, using a Narishige micromanipulation device and an Eppendorf air pressure microinjector. Successfully transformed cells were screened for the ability to discharge trichocysts on picric acid stimulation and further analyzed.

Antibodies. The polyclonal GTL3 anti-Bug22p antibody was purchased from Aviva Systems Biology (San Diego, CA). The monoclonal antibody 20H5, raised against *Chlamydomonas* centrin (29), was used to label centrioles in HeLa cells. The monoclonal antibody ID5 (38), which essentially labels basal bodies and a few microtubule arrays (10), was used to visualize the basal-body pattern. The polyclonal anti-GFP antibody was purchased from Interchim (Montluçon, France). The polyclonal anti-*Paramecium* ciliary tubulin (9) and anti-ciliary rootlet (34) antibodies were homemade antibodies.

The secondary antibodies used in immunofluorescence were Alexa Fluor 488 or 594 anti-mouse or -rabbit IgG antibodies from Invitrogen, the antibody used in Western blot experiments was an anti-rabbit IgG-alkaline phosphatase (AP) conjugate (Promega), and the antibodies used in immunoelectron microscopy were colloidal-gold-conjugated anti-rabbit immunoglobulins (Gar G5 and Gar G10; Aurion).

Protein extraction, SDS gels, and Western blots. *Paramecium* protein extracts were prepared from a pellet of 2 × 10⁸ cells homogenized in a potter with tight clearance in 10 mM Tris HCl, 1 mM EDTA, 250 mM sucrose containing a protease inhibitor cocktail (Set1; Calbiochem). Low-speed pellets and supernatants were separated by centrifugation at 10,000 × *g*. Total protein extracts of bacteria were obtained by sonication of a bacterial pellet in lysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na deoxycholate). Gel electrophoresis was performed using Criterion XT gels at neutral pH and 12% Bis-Tris (Bio-Rad) according to the manufacturer's recommendations. The primary antibodies used were anti-GTL3 (1:500) and anti-GFP (1:750). The secondary antibody was anti-rabbit-AP (1:500). Western blots were revealed using the 5-bromo-4-chloro-3-indolylphosphate (BCIP)-nitroblue tetrazolium (NBT) liquid substrate system from Sigma.

Immunofluorescence microscopy. (i) **HeLa cells.** After 24 h of culture, cells were washed with phosphate-buffered saline (PBS), fixed for 6 min at -20°C in 100% methanol, and blocked in 3% bovine serum albumin (BSA) and 0.1% Tween 20 in PBS for 30 min. The cells were incubated at room temperature with either the 20H5 (1:1,000) or the GTL3 (1:200) antibody for 1 h and then washed three times in 1% BSA and 0.1% Tween 20 and incubated for 1 h with the secondary antibody before three additional washes.

(ii) ***Paramecium*.** Immunostaining of cells was carried out as previously described (28) using the anti-ciliary tubulin (1:500), the monoclonal ID5 (1:350), or the anti-ciliary rootlet (1:400) antibody. For both HeLa cells and *paramecium*, appropriate fluorescent secondary antibodies were applied at 1:500 dilutions, nuclei were stained with Hoechst-33258, and slides were mounted in Cityfluor (London). Preparations were observed under a Zeiss Axioskop 2-plus fluorescence microscope equipped with a Roper Coolsnap-CF intensifying camera with GFP filters. Images were processed with Metamorph software (Universal Imaging).

Electron microscopy. For morphological observations, cell pellets were processed as previously described (5) using 1% or 2% glutaraldehyde for fixation. For postembedding immunolocalization, cell pellets were processed as described previously (18), using either 2% paraformaldehyde, 0.15% glutaraldehyde, or 3% paraformaldehyde and 0.5% glutaraldehyde for fixation, and embedded in LR White (London Resin). Thin sections were collected on nickel grids and treated with double-distilled water or 0.1 M NH₄Cl in 0.1 M PBS and then saturated and processed with 3% BSA in PBS. Primary polyclonal antibodies were diluted 1:50 to 1:75 (anti-GTL3) and 1:10, 1:20, or 1:500 (anti-GFP). After being washed, the sections were incubated with, respectively, 1:40, 1:75, or 1:100 dilutions of 5-nm (GTL3 labeling) or 10-nm (GFP labeling) gold particles. The sections were examined with a Philips CMI0 or a Jeol transmission electron microscope.

RNAi by feeding. RNAi gene knockdown was performed according to the method of Galvani and Sperling (13). HT115, an RNase III-deficient strain of *Escherichia coli* with an isopropyl- β -D-thiogalactopyranoside (IPTG)-inducible T7 polymerase, was transformed by the desired constructs into Litmus or L4440 plasmids. Wild-type paramecia were incubated in double-stranded-RNA (dsRNA)-expressing bacteria and were transferred daily into fresh feeding medium. Control cells were fed with bacteria carrying the complete coding region of the *ND7* gene.

Swimming analysis. Two methods were used for swimming analysis. First, we measured the swimming speed of cells under dark-field microscopy using a Zeiss Axioskop 2-plus microscope with a 10 \times objective by recording tracks during 650-ms pauses. Second, we used a Zeiss Steni 2000-C dissecting microscope with 1-min time-lapse acquisitions at 7 frames per second with a Roper Coolsnap-CF camera and Metamorph software (Universal Imaging). The speed means and standard errors were calculated by measuring the lengths of the tracks using ImageJ and subtracting a mean length of the paramecia for the first method and using automated analysis of the swimming tracks currently under development for the second method (J. Ferracci, O. Arnaiz, and J. Cohen, unpublished data).

High-speed video microscopy. All of the observations were made under an inverted microscope (Axiocvert 200; Carl Zeiss S.A.S., Le Pecq, France) using a 100 \times objective on 5 μ l of a suspension of paramecia immobilized between a cover glass (diameter, 50 mm; thickness, 0.13 to 0.16 mm) from Menzel-Glaser and a 22- by 22-mm coverslip, with a few hundred calibrated 10-, 15-, or 20- μ m polystyrene beads (Fluka), during overall periods shorter than 5 min to avoid changes in ciliary activity. A final dilution of ca. 1/10 of India ink was added to the medium to be visualized as particles that materialized the fluid flow during the acquisitions. Cilium movements were digitally recorded with a camera (Pixelink A741; Ottawa, Canada) at a rate of 355 images per second. A movie was constituted of 1,800 images with a definition of 256 by 192 pixels. The size of a pixel was 0.13 by 0.13 μ m². In each sequence, we selected a cilium that we were able to follow during an entire cycle of beating. For this, the video sequences were played back frame by frame using Streampix software (Norpix Inc., Montreal, Canada). The angle of beating was calculated by trigonometric calculations in the triangle made by the anchoring point and the two extreme positions of the cilium extremities during the power stroke.

Swimming-behavior assays. Cells to be assayed were first equilibrated for a few minutes in resting solution (2 mM Na citrate, 1 mM NaH₂PO₄, 1 mM Na₂HPO₄, 1.5 mM CaCl₂) and then individually transferred to test solutions that consisted of either 30 mM KCl or 5 mM tetraethylammonium, 10 mM NaCl in resting solution (19). Observation of immediate swimming behavior was done during the first minute after transfer under the dissecting microscope.

RESULTS

Conservation of Bug22p in eukaryotes. Bug22p, a small basic protein of 22 kDa, was one of the abundant proteins identified in a high-throughput proteomics analysis of *Paramecium* cilia (9 distinct peptides), and its mRNA level in cells regenerating cilia was increased by a 2- to 3-fold induction (reference 1 and unpublished data). A BLAST search for similar proteins in other species for which whole-genome sequence data are available revealed that it is widespread in eukaryotes harboring centrioles/cilia at some stage of their life cycles. Among eukaryotes devoid of centrioles/cilia, plants possess *BUG22* genes whereas fungi do not. Interestingly, fungi, such as *Batrachomyces dendrobatis*, which form flagellated gametes also have a *BUG22* gene. The protein sequence of Bug22p proteins is highly conserved through evolution (Fig. 1), except for the presence of C-terminal extensions in *Trypanosoma* and *Caenorhabditis*. Such lack of sequence variation during evolution suggests a conserved role in different organisms, including but not restricted to ciliary function, since the sequence conservation is also high in plants. Despite high overall amino acid sequence conservation, diagnostic positions in the sequence allow us to distinguish *BUG22* genes of acentriolar organisms (higher plants) from the others, as shown in a global alignment (Fig. 1).

The whole Bug22p protein constitutes by itself the Interpro domain DUF667 (for domain of unknown function) that occurs in some proteins other than Bug22p. However, the sequence conservation of the DUF667 domain in these proteins is much lower than within the Bug22p subfamily, suggesting that the functions of the less conserved DUF667 domains may be divergent with respect to the Bug22p domain.

Four Bug22 genes for a single Bug22p protein in Paramecium. Four paralogous *BUG22* genes encoding four 100% identical proteins are present in the *Paramecium* genome. These paralogous genes arose from the last two whole-genome duplications (WGDs) that occurred in this organism (3). The four genes, with the identifiers (IDs) GSPATG00007012001, GSPATG00016658001, GSPATG00012810001, and GSPATG00013199001, were, respectively, called *BUG22a*, *-b*, *-c*, and *-d*. The DNA sequences are almost identical (96.1% identity between *BUG22a* and *-b* and 96.9% between *BUG22c* and *-d* [both pairs are paralogs from the most recent WGD] and 83.2% to 83.8% between the pair *BUG22a* plus *-b* and the pair *BUG22c* plus *-d*). Most of the divergence occurs in intron sequences, as visualized by multiple alignments (see Fig. S1 in the supplemental material). The few differences in exons are all silent.

Localization of Bug22p in HeLa cells and Paramecium. HeLa cells transfected with the GFP-HsBUG22 cDNA displayed labeling of two dots close to the nucleus, which colocalized with centrin, a centriolar marker (11), as revealed by the 20H5 monoclonal antibody (see Fig. S2A in the supplemental material), indicating that HsBug22p is present at the centrosome. This observation is similar to images published by Keller et al. (16), who transfected U2OS cells with a C-terminal GFP-Bug22p fusion protein in a systematic study of centriolar proteins. Our results thus confirm the centriolar localization of Bug22p.

Both the *HsBUG22* and the *PtBUG22* GFP fusion genes were used to transform *Paramecium* by DNA microinjection into the macronucleus, and the fluorescence was observed in the clonal descendants of the injected cells. Labeling of basal bodies and the proximal two-thirds of cilia was observed with both GFP fusions (Fig. 2), showing that, in agreement with the proteomics of cilia in several species, Bug22p localizes to basal bodies and cilia and that the human ortholog HsBug22p is targeted to these organelles, as well.

GTL3 is a commercial antibody raised against the C-terminal portion of the human Bug22p protein. As the primary sequence is highly conserved between *Paramecium* and humans (89% identity in the amino acid sequences) (Fig. 1), we tested the GTL3 antibody by Western blotting (Fig. 3) and found that it specifically recognized a 22-kDa peptide in *Paramecium* extracts (Fig. 3A), as well as the GFP-HsBug22p fusion expressed in *E. coli* (Fig. 3B and C). In HeLa cells, the GTL3 labeling coincided with the localization of centrin (see Fig. S2B in the supplemental material) and GFP-HsBug22p (see Fig. S2C in the supplemental material), confirming its Bug22p specificity.

In immunofluorescence, no labeling of *Paramecium* by GTL3 was observed. This negative result may be due to the solubility of Bug22p molecules in the medium after detergent permeabilization or to lack of accessibility to the antigen under the fixation and labeling procedures used. The protein local-

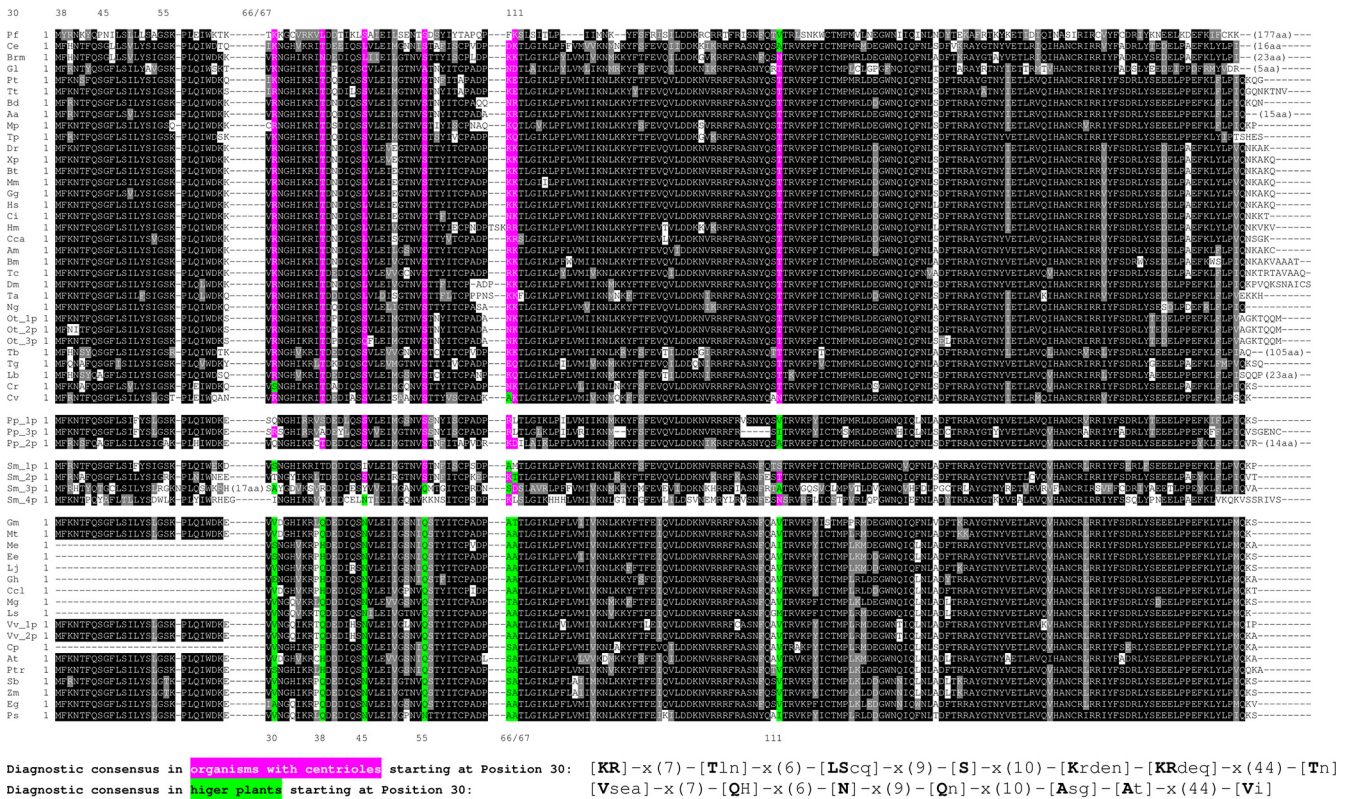


FIG. 1. Sequence conservation of Bug22p in eukaryotes and positions that distinguish ciliary versus nonciliary species. The ClustalW alignment of 58 Bug22p sequences, representative of all the clades of eukaryotic species in which it can be found, was visualized using Boxshade and illustrates the strong conservation of the Bug22p protein through evolution. Careful examination of the divergences at seven specific positions (amino acids 30, 38, 45, 55, 66, 67, and 111 according to the numbering of the *Homo sapiens* Bug22p sequence) revealed that species having cilia or centrioles in at least some stage of their life cycle have a consensus at these positions (denoted in purple) that dramatically differs from the consensus found in plants (denoted in green), as displayed at the bottom of the figure. Note that both the moss *Physcomitrella patens* and the fern *Selaginella moellendorffii* (in the middle of the alignment), which both have several Bug22 genes in their genomes and are plants with occurrence of flagellated cells in their life cycles, display a patchwork of the plant consensus and the ciliary consensus at these positions. In addition, the occurrence of a *BUG22* gene with ciliary consensus in the genome of *Chlorella vulgaris*, harboring a centriole that never develops a cilium (39), suggests that Bug22p has a centriolar function. The species studied and the accession numbers of the proteins are as follows: Aa, *Aedes Aegypti* XP_001663334.1; At, *Arabidopsis thaliana* AAG51069.1; Am, *Apis mellifera* XP_395824.1; Bd, *Bartrachochytrium dendrobadiis* jgi |Batde5| 32931; Bm, *Bombyx mori* NP_001040392.1; Brm, *Brugia malayi* XP_001894987.1; Bt, *Bos taurus* NP_032213.2; Cca, *Capitella capitata* jgi |Capca1| 77038; Ccl, *Citrus clementina* EST:FC926856.1, GI:218791863; Ci, *Ciona-intestinalis* XP_002131975.1; Ce, *Caenorhabditis elegans* NP_871706.1; Cv, *Chlorella vulgaris* jgi |Chlvu1|29644|e; Cr, *Chlamydomonas reinhardtii* XP_001692489; Cp, *Carica papaya* EST: EX271746.1, GI:186814042; Cq, *Culex quinquefasciatus* XP_001869024.1; Dm, *Drosophila melanogaster* NP_609402.1; Dr, *Danio rerio* NP_957105.1; Ee, *Euphorbia esula* EST: DV114861.1; Eg, *Elaeis guineensis* ACF06559; Gg, *Gallus gallus* NP_001005833.1; Gh, *Gossypium hirsutum* EST: EX168390.1 GI:164311590; G1, *Giardia lamblia* XP_001708694.1; Gm, *Glycine max* gb:ACU16157.1; Hm, *Hydra magnipapillata* XP_002163101.1; Hs, *Homo sapiens* NP_037374.1; Lb, *Leishmania braziliensis* XP_001564754.1; Lj, *Lotus japonicus* EST: GO028159.1, GI:223439164; Ls, *Lactuca serriola* EST:BO986870.1, GI:22404395; Me, *Manihot esculenta* EST:DV448997; Mg, *Mimulus guttatus* EST:GR148812.1, GI:238398755; Mm, *Mus musculus* NP_032213.2; Mp, *Micromonas pusilla* jgi |Micpu2|70275|AZW; Mt, *Medicago truncatula* ACJ85502.1; Ng, *Naegleria gruberi* JGI Protein ID: 81674; Ot_1p, *Oxytricha trifallax* OXADaaa05e09; Ot_2p, *Oxytricha trifallax* OXAEaaf21g04; Ot_3p, *Oxytricha trifallax* OXAEaaa06d08; Pf, *Plasmodium falciparum* XP_001351642.1; Pp_1, *Physcomitrella patens* XP_001775608.1; Pp_2, *Physcomitrella patens* XP_001760380.1; Pp_3, *Physcomitrella patens* XP_001784194.1; Ps, *Picea sitchensis* ABK26040.1; Pt, *Paramecium teraurelia* (this article); Ptr, *Populus trichocarpa* XP_002304809.1; Sb, *Sorghum bicolor* XP_002460501.1; Sm_1, *Selaginella moellendorffii* jgi |Selmo1|172529; Sm_2, *Selaginella moellendorffii* jgi |Selmo1|96514; Sm_3, *Selaginella moellendorffii* jgi |Selmo1|130480; Sm_4, *Selaginella moellendorffii* jgi |Selmo1|117290; Ta, *Trichophlox adherens* jgi |Triad11| 23937; Tb, *Trypanosoma brucei* XP_822473.1; Tc, *Tribolium castaneum* XP_975505.1; Tg, *Toxoplasma gondii* XP_002364709.1; Tt, *Tetrahymena thermophila* XP_001007313.2; Tp, *Thalassiosira pseudonana* XP_002286290.1; Vv_1p, *Vitis vinifera* XP_002268386.1; Vv_2p, *Vitis vinifera* XP_002265223.1; Xp, *Xenopus tropicalis* NP_001004835.1; Zm, *Zea mays* NP_001148848.1.

ization was then examined by postembedding immunoelectron microscopy, both with the GTL3 antibody on wild-type cells and with an anti-GFP antibody on GFP-Bug22p-expressing cells (Fig. 4). Examination of 142 sections of cilia and basal bodies using the GTL3 antibody and 115 sections with the anti-GFP antibody indicated that both antibodies gave the same labeling in the basal bodies and cilia: gold particles oc-

curred preferentially in the terminal plate of the basal body and the transition zone to the cilium, as well as in the cilium, in the vicinity of the outer doublets of the axoneme, and often between the axoneme and the membrane. In GFP-expressing cells, additional labeling was often found within the basal body, close to its proximal part (Fig. 4G). In GFP-expressing cells, 34 longitudinal sections of cilia revealed an apparent periodicity

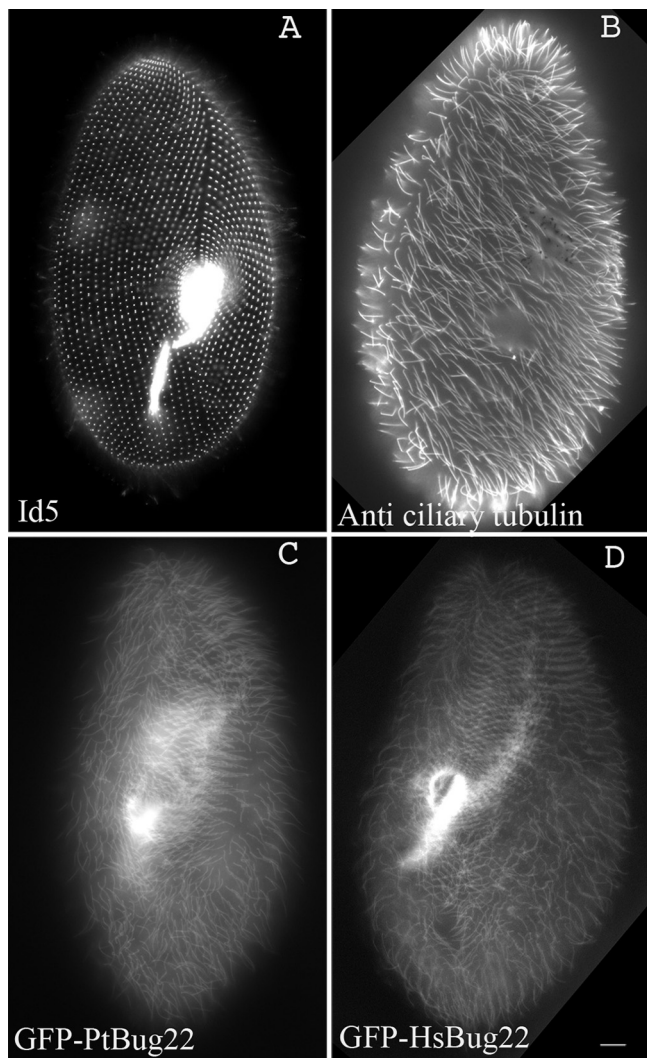


FIG. 2. Ciliary and basal-body localization of *Paramecium* and human Bug22p GFP fusions in *Paramecium*. (A and B) Wild-type paramecia immunofluorescently labeled with the ID5 monoclonal antibody to visualize basal bodies (A) and with the anti-ciliary tubulin to show cilia (B). (C) *Paramecium* expressing GFP-PtBug22p showing clear labeling of basal bodies and cilia. (D) Same as panel C for a *Paramecium* expressing GFP-HsBug22p. Bar = 10 μ m.

of the labeling as a regular disposition of the gold particles at 500- to 550-nm intervals (Fig. 4J and K) and, less often, at 250- and 800-nm intervals.

Growth and swimming defects provoked by Bug22p depletion in *Paramecium* can be complemented by the human ortholog. The function of Bug22p was investigated in *Paramecium* using the RNAi feeding method. Since the four *Paramecium* *BUG22* genes encode the same protein, we would expect not qualitative but quantitative differences by inactivating each paralog of the gene family. We used *BUG22a* and *BUG22c* sequences for double-stranded-RNA production. The two dsRNAs were expected to silence all four paralogs, based on sequence conservation (see Fig. S1 in the supplemental material). Double silencing using the two sequences was needed to improve the knockdown efficiency on all four genes.

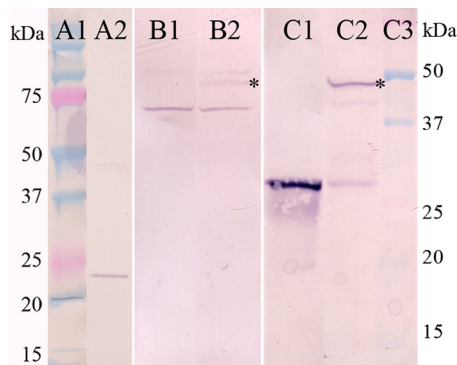


FIG. 3. Western blot evidence that Bug22p is the antigen of a commercial anti-GTL3 antibody. Lanes: A1, molecular mass markers for lane A2; A2, anti-GTL3 labeling of a 10,000 \times g supernatant of whole homogenized paramecia showing a positive band between 20 and 25 kDa; B1 and C1, total extracts of *E. coli* expressing GFP in the pPXV plasmid normally used for expression in *Paramecium* but which allows low expression in the bacteria; B2 and C2, *E. coli* expressing the GFP-Bug22p fusion; C3, molecular mass markers for lanes B1 to C2. The B series lanes were revealed with the anti-GTL3 antibody and the C series with the anti-GFP antibody. The asterisk indicates the level of the fusion protein that is recognized by both GTL3 and GFP antibodies.

We first evaluated the efficiency of the RNAi procedure on fluorescent *Paramecium* cells from a clone transformed with the *GFP-PtBUG22a* fusion gene: the fluorescence was completely abolished by the double *BUG22a-BUG22c* knockdown, showing that the RNAi efficiency depleted Bug22p (data not shown). Phenotypic analyses obtained in further RNAi experiments were performed on nontransformed wild-type cells.

Knocking down each pair of *BUG22a-BUG22b* or *BUG22c-BUG22d* genes resulted in a growth rate of 1 or 2 fissions per day compared to 4 or 5 in control cells and led to lethality in 2 to 3 days. When all 4 genes products were knocked down, cells divided 1 or 2 times on the first day and 0 or 1 time on the second day and died within 3 days. Lethality might be the result of starvation owing to impaired ciliary beating (see below) that prevented food capture in the oral apparatus, as observed by defective take-up of India ink (not shown).

Swimming speed was measured by recording swimming tracks under dark-field microscopy, which showed that, after 1 day of silencing, knocking down each pair of gene products decreased the rate of motility by 75%, while knocking down all four gene products decreased the rate of motility by 90% (Fig. 5).

To test whether a human Bug22 protein could complement RNAi depletion of *Paramecium* orthologs, we transformed *Paramecium* for heterolog expression of the human *BUG22* cDNA and submitted the cells to RNAi against the *BUG22a* and *BUG22c* *Paramecium* sequences, which do not share nucleotide identity stretches sufficient to induce silencing of the human gene, and reciprocally to RNAi against the human *BUG22* cDNA, which did not cross-react with *Paramecium* sequences (see Fig. S1 in the supplemental material). A control with double human and *Paramecium* *BUG22* RNAi was also performed. We observed that neither HsBUG22 nor PtBug22 RNAi alone could alter the growth rate or the swimming speed, whereas the double Hs/Pt RNAi could (Fig. 6 and

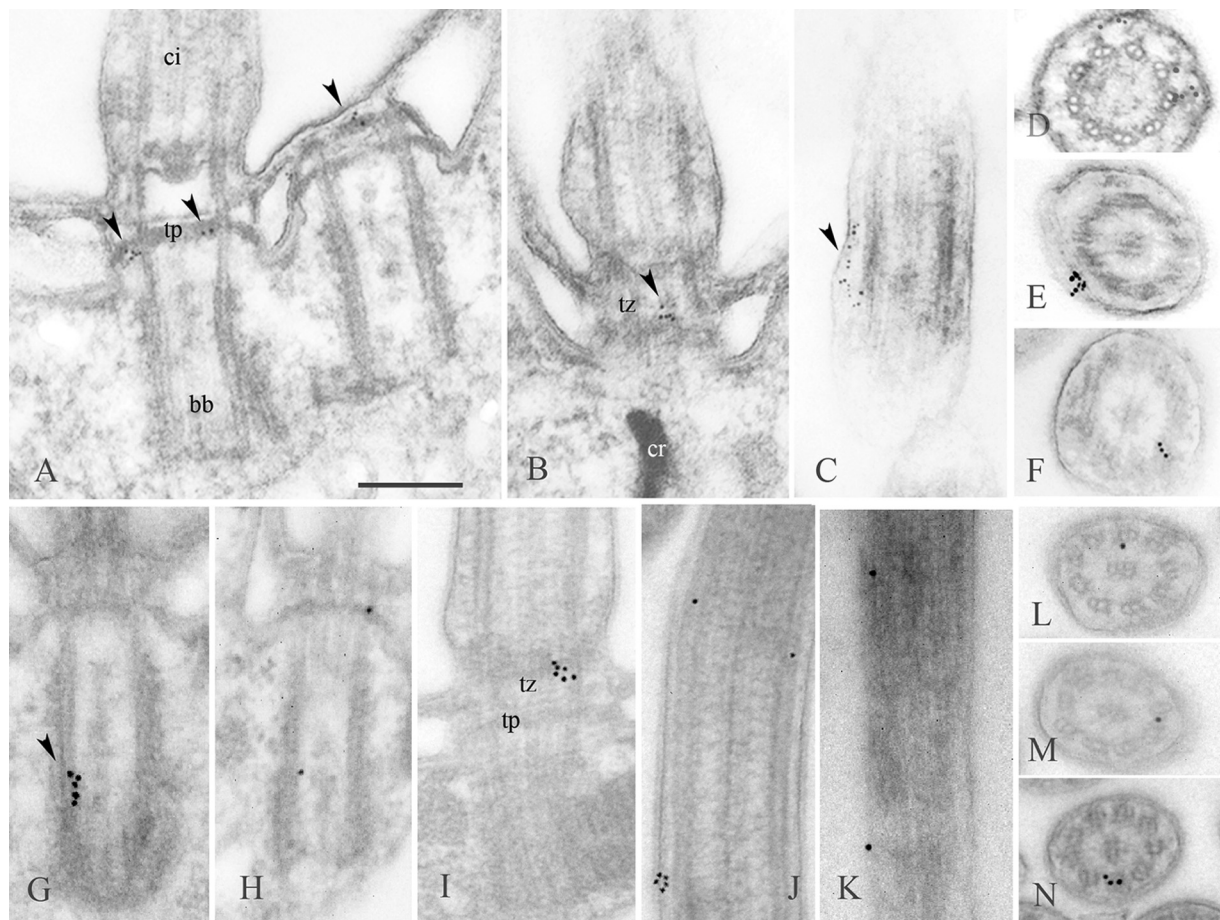


FIG. 4. Localization of Bug22p in *Paramecium* by postembedding immunoelectron microscopy. (A to F) Labeling of wild-type cells with the GTL3 antibody using 5-nm gold particles. (G to N) Labeling of GFP-Bug22p-expressing cells with an anti-GFP antibody using 10-nm gold particles. The arrowheads point to gold particles. (A and H) Labeling of the terminal plates of basal bodies. (B, D, and I) Labeling of the transition zone to the cilium. (F and J to N) Labeling in the vicinity of the outer doublets of the axoneme. (C, E, and J) Labeling between the axoneme and the membrane. (G) Labeling within the basal body close to its proximal part. (J and K) Examples of regular disposition of the gold particles along the axoneme. ci, cilium; bb, basal body; tp, terminal plate; tz, transition zone. Bar = 250 nm.

its legend provide details). This demonstrates the ability of the human protein to rescue depletion of *Paramecium* Bug22p.

Bug22p depletion affects swimming behavior. When carefully observed under the dissecting microscope, live Bug22p-depleted cells seem to undergo less frequent backward swimming than control cells. Since these events are rare under control conditions, we tested swimming behavior in ion solutions, as described for the identification of Mendelian mutants (19). Control cells immersed in 30 mM KCl or in 5 mM tetraethylammonium, 10 mM NaCl (see Materials and Methods) immediately underwent continuous backward swimming for more than 20 s and then whirled for a few seconds and resumed normal forward swimming. In contrast, Bug22p-depleted cells showed a very much weakened reaction, with very short or no backward swimming and whirling for less than 5 s. This suggests abnormal ciliary-channel activity or abnormal response of the axonemal motors to ion currents, induced by RNAi knockdown.

Bug22p depletion disturbs basal-body organization in the cortex of *Paramecium*. Since the Bug22p protein is present in basal bodies, in addition to cilia, we wondered whether its

depletion causes basal body defects. We therefore examined Bug22p-depleted cells using antibody labeling that revealed, not only the basal bodies themselves (antibody ID5), but also the ciliary rootlets (anti-KD), which provide good markers of the orientation of individual basal bodies in the global pattern of the polarized cell cortex. By the second day under *BUG22* RNAi conditions, a sporadic disorganization started to appear, with ciliary rootlets showing diverse erratic orientations (see Fig. S3 in the supplemental material). This effect on basal-body organization and polarity could be a direct consequence of Bug22p depletion or an indirect effect secondary to alteration of ciliary activity.

Bug22p depletion alters ciliary beating efficiency and stroke shape in *Paramecium*. To investigate the cause of the decrease in swimming speed induced by RNAi depletion of Bug22p, we first checked the presence of cilia by immunofluorescence using an anti-tubulin antibody. Cilia appeared unaffected in number, density, shape, or length, in contrast to what is obtained by silencing IFT172, a component of the intraflagellar-transport machinery involved in ciliary growth and maintenance (26), used as a control (see Fig. S4 in the supplemental material). In

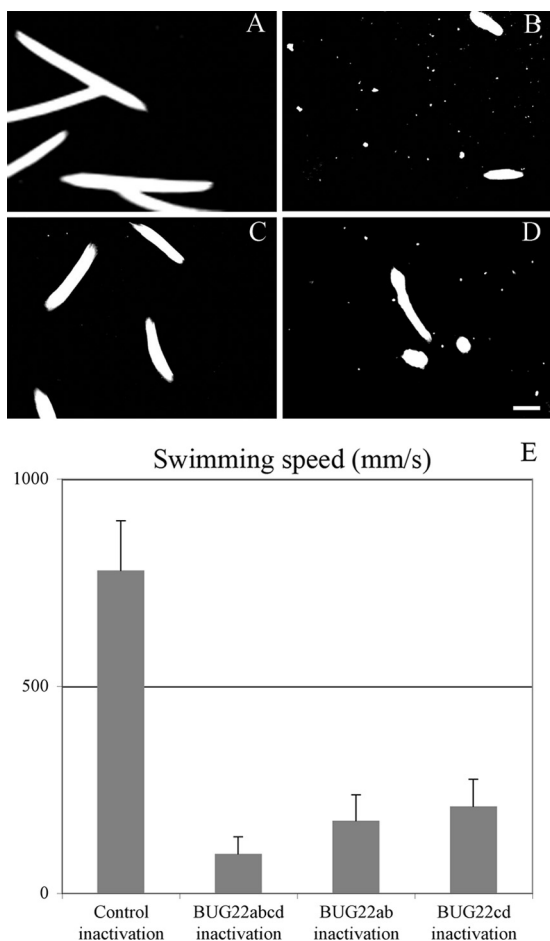


FIG. 5. Swimming-speed diminution in Bug22p-depleted cells. Shown is a track recording of parametia swimming under a dark-field microscope equipped with a 10 \times objective during a 650-ms pause. (A) Control Nd7p-depleted cells. (B) Double RNAi targeting all four *BUG22* genes. (C and D) Simple RNAi targeting *BUG22a* and *BUG22b*, and *BUG22c* and *BUG22d*, respectively. Bar = 100 μ m. (E) Histograms of measured speeds under all conditions showing quantization of the cumulative effects of double RNAi compared to simple RNAi. The error bars indicate standard errors.

living cells observed under phase-contrast microscopy at high magnification, the cilia appeared to beat vigorously, in apparent contradiction of the strong decrease in the cell swimming speed. This indicates that other factors than beating frequency can affect swimming speed. We therefore studied the ciliary beating of Bug22p-depleted cells by high-speed video microscopy (Fig. 7; see Videos S1 and S2 in the supplemental material). The recorded images revealed that although the cilia beat at an even higher frequency than normal (42 Hz versus 14 Hz), they beat with reduced efficiency, since (i) the angle of beating was narrower (55 $^\circ$ versus 97 $^\circ$) and (ii) the apparent distinction between power and recovery strokes, as well as the coordination of beating over the cell surface, was lost. In addition, cilia appeared “floppy” when they moved, as though they were less rigid, which could explain their altered beating. The global impairment in the beating efficiency of fluid movement could be visualized by particles in the medium moving in a disorga-

nized manner, more or less randomly, compared to the control, where particles were rapidly swept in a dominant direction.

Examination of Bug22p-depleted cells under transmission electron microscopy did not reveal particular defects in the general organization of the basal body or cilium (9-fold symmetry, a central pair of microtubules, and radial spokes in the axoneme). However, the overall shape of the cilia was strongly altered in 30% of the longitudinal sections observed, with sharp bending that was never observed in controls, accompanied by shrinking of the axoneme diameter in the region of the bend (Fig. 8). This indicated that the natural rigidity of cilia was greatly impaired by the depletion of Bug22p, in good agreement with the observation of the shape of beating cilia in live cells.

DISCUSSION

Bug22p, a novel ciliary protein. The present study of Bug22p was stimulated by finding the protein in proteomic analysis of *Paramecium* cilia (1) and by the fact that we could extract it by mining the results of high-throughput studies of various species. In organisms with centrioles and cilia, high-throughput studies provided strong evidence that Bug22p has a localization/function related to these organelles. In *Chlamydomonas*, Bug22p has been detected in proteomic analysis of flagella (25) and basal bodies (17), and *BUG22* mRNA is upregulated during flagellar biogenesis (17, 36). The protein has also been found in *Trypanosoma* flagella (8), in *Tetrahymena* cilia (32), in human cilia (23), and in mouse photosensory receptors (21). The promoter of the *BUG22* gene in *Caenorhabditis* displays an X box (6), characteristic of genes encoding proteins related to ciliary biogenesis and regulated by the daf19 transcription factor. The mouse ortholog of *BUG22*, *GTL3*, has been identified in a gene trap screen (transcript T10-2A2) as expressed in many tissues, but preferentially in the testis (27), an organ where flagellated sperm cells develop. A large-scale cDNA screen by *in situ* hybridization of zebra fish embryos (35; <http://zfin.org/>) localized major expression of *BUG22* in tissues where cilia are known to play an important role in embryonic development—Kupffers vesicle (equivalent to the embryonic node of mammals), neural tube, spinal cord, and pronephric duct—which is compatible with preferential expression in ciliated tissues. *In situ* hybridization in *Drosophila* (Berkeley *Drosophila* Genome Project [<http://www.fruitfly.org/>]) revealed preferential expression of *BUG22* in sensory complexes during embryogenesis, again compatible with expression in ciliary tissues. The high sequence conservation and apparently ubiquitous presence of Bug22p in centrosomes, cilia, and flagella stimulated its further study.

Potential roles of Bug22p in motile cilia. Cilia and flagella are organelles composed of hundreds of proteins whose assembly and function, apart from main components of the backbone of the axoneme, are not well understood. Even the most abundant proteins found in high-throughput proteomics studies of cilia remain to be studied, and we focused on one of them, Bug22p.

In *Paramecium*, four genes issuing from two successive whole-genome duplications encode four identical proteins, so that only one Bug22p protein is synthesized in the cell although the nucleotide sequences had time to diverge at silent posi-

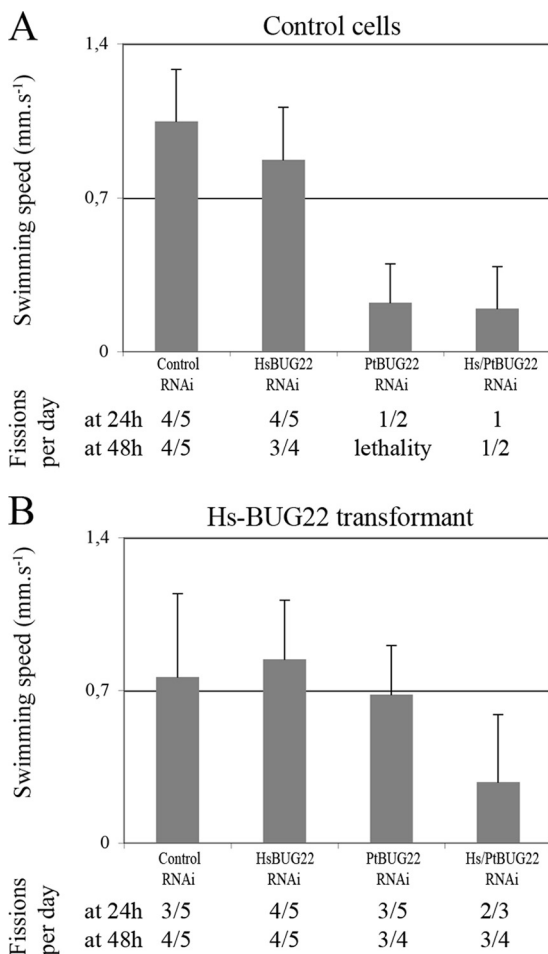


FIG. 6. Complementation of the effects of Bug22p depletion by expression of the human gene in *Paramecium*. To test possible rescue of Bug22p-depleted paramecia by expression of human Bug22p, cells were first transformed by a plasmid containing the human *BUG22* cDNA under *Paramecium* regulatory sequences and then subjected to RNAi against the *BUG22* endogene (*PtBUG22*), the human *BUG22* transgene (*HsBUG22*), or both (*HsBUG22* and *PtBUG22*) to follow the phenotypes. Transformation of *Paramecium* nd7-1 cells (unable to undergo trichocyst exocytosis) was performed by microinjection of a mixture of two plasmids, one driving the expression of the wild-type *ND7* gene to monitor the efficiency of the transformation and the other for expression of the human *BUG22* cDNA. The clones derived from transformed cells were cultured and submitted to RNAi, with the *ND7* gene used as a target for control RNAi. The phenotypic analysis consisted of monitoring the growth rate by cell counting and the swimming speed by automated analysis of recorded tracks. Each measurement was performed on accumulations of tracks of ca. 20 cells during 1-min acquisitions and was repeated 3 times on independent samples. Shown are the swimming speed (histograms and standard errors) and growth rate (values below the histograms) of an HsBug22 transformant (B) compared to a control clone (A). This transformant appeared to be resistant to RNAi targeting the endogenous *Paramecium* *BUG22* genes (*PtBUG22* RNAi) compared to the control clone under the same conditions, as shown by the heights of the bars labeled *PtBUG22* RNAi in both histograms, and to be resistant to the human *BUG22* transgene (*HsBUG22* RNAi), as was the control clone, but not to double RNAi (*Hs/PtBUG22* RNAi). This indicates that expression of the human cDNA is able to replace the depleted *Paramecium* protein, thus showing complementation ability, but does not represent a nonspecific effect, since the double RNAi inactivates any *BUG22* function in these cells. Altogether, six independently transformed clones have been extensively analyzed with similar results. Some variation was observed in the levels of the complementation: *PtBUG22*-

tions. An explanation is that the function of Bug22p is so constrained that there is no latitude for mutation without function impairment. Thus, Bug22p is likely to have an essential role in the cell, interacting with many different partners.

The localization experiments described here using a GFP-PtBug22p fusion *in vivo*, or using anti-Bug22p or anti-GFP antibodies on wild-type or GFP-PtBug22p-expressing cells in immunoelectron microscopy, confirmed the basal-body and ciliary localization of the protein in *Paramecium*. More precisely, the labeling showed a preferential concentration of the gold particles close to the triplet (basal bodies) or doublet (cilia) microtubules, with frequent labeling between the axoneme and the ciliary membrane. Bug22p could therefore be a protein more or less stably associated with microtubules. Its basic isoelectric point (~ 10) is compatible with possible interaction with tubulin, although this remains to be demonstrated biochemically.

The localization of Bug22p in the transition zone between the basal body and cilium, as well as in foci along the cilia, suggests a role in IFT. Such a view could be supported by the facts that high-throughput 2-hybrid screens revealed CLUAP1/qilin/dyf-3 as a potential interactant of Bug22p in *Drosophila melanogaster* (15) and that dyf-3 has been shown to belong to the IFT in *Caenorhabditis elegans* (12, 22). More precisely, dyf-3 is a component of the IFTB particle that is involved in the anterograde movements in cilia that control ciliary growth and turnover. However, knocking down Bug22p in *Paramecium* does not yield lack-of-cilia phenotypes found in IFTB knockdowns in other species, and in *Paramecium*, as verified by IFT172 RNAi experiments (see Fig. S4 in the supplemental material). The question of an interaction of Bug22p with dyf-3 in *Paramecium*, or its potential role in the IFT itself, or as IFT cargo, is still open.

The most obvious effect of Bug22p depletion is a reduction in cell swimming speed resulting from inefficient ciliary beating (Fig. 7; see Videos S1 and S2 in the supplemental material) and apparent loss of ciliary rigidity (Fig. 8). Little is known about the factors that rigidify membrane-bound microtubular complexes. Ciliary bending results from multiple interactions within the axoneme, where the nine microtubular doublets slide over neighboring doublets through dynein-tubulin interactions coordinated by the central pair and radial spokes (30). Interestingly, it has recently been shown that hydin, a central-pair-associated protein, is critical for regulation of ciliary bending during beating strokes and of beat frequency in a manner opposite to our observations for Bug22p. Indeed, hydin depletion in mutants produces excessive ciliary rigidity and reduces beat frequency (20), in contrast to Bug22p-depleted cilia, which display excessive flexibility and beating frequency.

depleted cells swam at 10% of the normal speed, while cells rescued by transformation swam at 35% to 90% of the normal speed. Such variation is routinely observed in transformation of *Paramecium*, depending on the amount of transforming DNA introduced via microinjection. In three cases where the amount of injected DNA was optimal, the rescue was abolished when both *Paramecium* and human sequences were targeted together by RNAi.

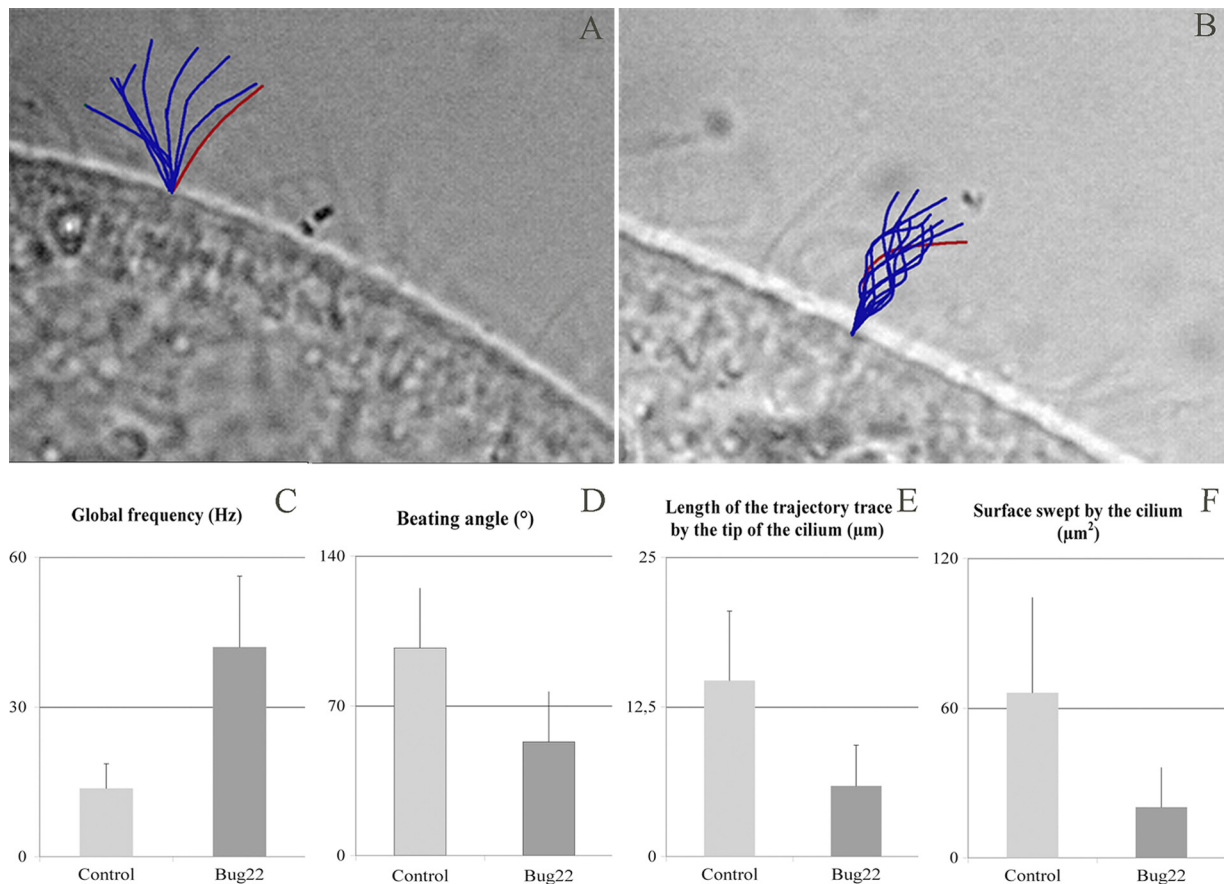


FIG. 7. Effects of Bug22p depletion on ciliary beating viewed by high-speed video microscopy. (A and B) Images of control Nd7p- and Bug22p-depleted cells, respectively, captured from high-speed video microscopy movies (see Videos S1 and S2 in the supplemental material) showing the ciliary beating, on which are superimposed the representation of a single cilium (in red) plus the representation of the same cilium in the 8 adjacent frames (blue), representing one power stroke in the control (25 ms) and erratic beating in the Bug22p-depleted cell. (C to F) Quantization of four parameters manually extracted from the movies on 14 beating cilia for BUG22 RNAi and 12 cilia for the control, with standard deviations. (C) Frequency of beating. (D) Angle between extreme positions taken by beating cilia. (E) Length of the trajectory of the tips of beating cilia. (F) Triangular surface swept by beating cilia. Bug22p depletion increased the frequency but decreased all the other parameters.

Functional conservation of Bug22p between *Paramecium* and humans. Bug22p sequence conservation from *Paramecium* to humans suggests conservation of its biological function, tested here by functional complementation. The expression of the human HsBug22p protein can rescue the depletion of the PtBug22p protein under RNAi conditions. Moreover, a balance can be found in which the expression level of the HsBug22p protein is in a physiological range, rescuing the PtBUG22 RNAi phenotype, while being susceptible to RNAi silencing itself. This is the first example of functional complementation by a human gene in a ciliate, which opens the use of *Paramecium* as a model to study human proteins by comparing the rescue properties of wild-type and mutated versions associated with diseases.

Different statuses for BUG22 genes through evolution and potential roles of the proteins. The Bug22p protein is a highly conserved eukaryotic protein; however, it is not found in all eukaryotes. All species that display centrioles and cilia, be they motile or not, at some stages of their life cycle have one or several BUG22 genes. With a few exceptions, plants and fungi do not have centrioles and cilia. Fungi do possess a centro-

some-like structure, the spindle pole body, while plants have MTOC structures to organize the microtubules that govern chromosome separation at mitosis. Interestingly, only plants, not fungi, have BUG22 genes in their genomes. Examination of the aligned sequences reveals a consensus at critical positions in Bug22p proteins of species with centriolar/ciliary function (Fig. 1). Plant Bug22p proteins display another consensus at the same positions and not only a lack of the ciliary consensus, suggesting that there is selective pressure to prevent divergence from this consensus. Plants that do possess centrioles and cilia/flagella, such as *Physcomitrella* and *Selaginella*, have several Bug22p proteins encoded in their genomes that display different combinations of ciliary and plant consensus residues at the critical positions, indicating that both putative centriolar/ciliary and plant functions of Bug22p are needed in these plants.

In the present study of the motile cilia of *Paramecium*, we have found a striking role of Bug22p in ciliary rigidity and beating efficiency, as well as a possible role in the cortical organization of basal bodies. Bug22p proteins seem to exist in other types of cilia, sensory and primary cilia. Indeed, a BUG22

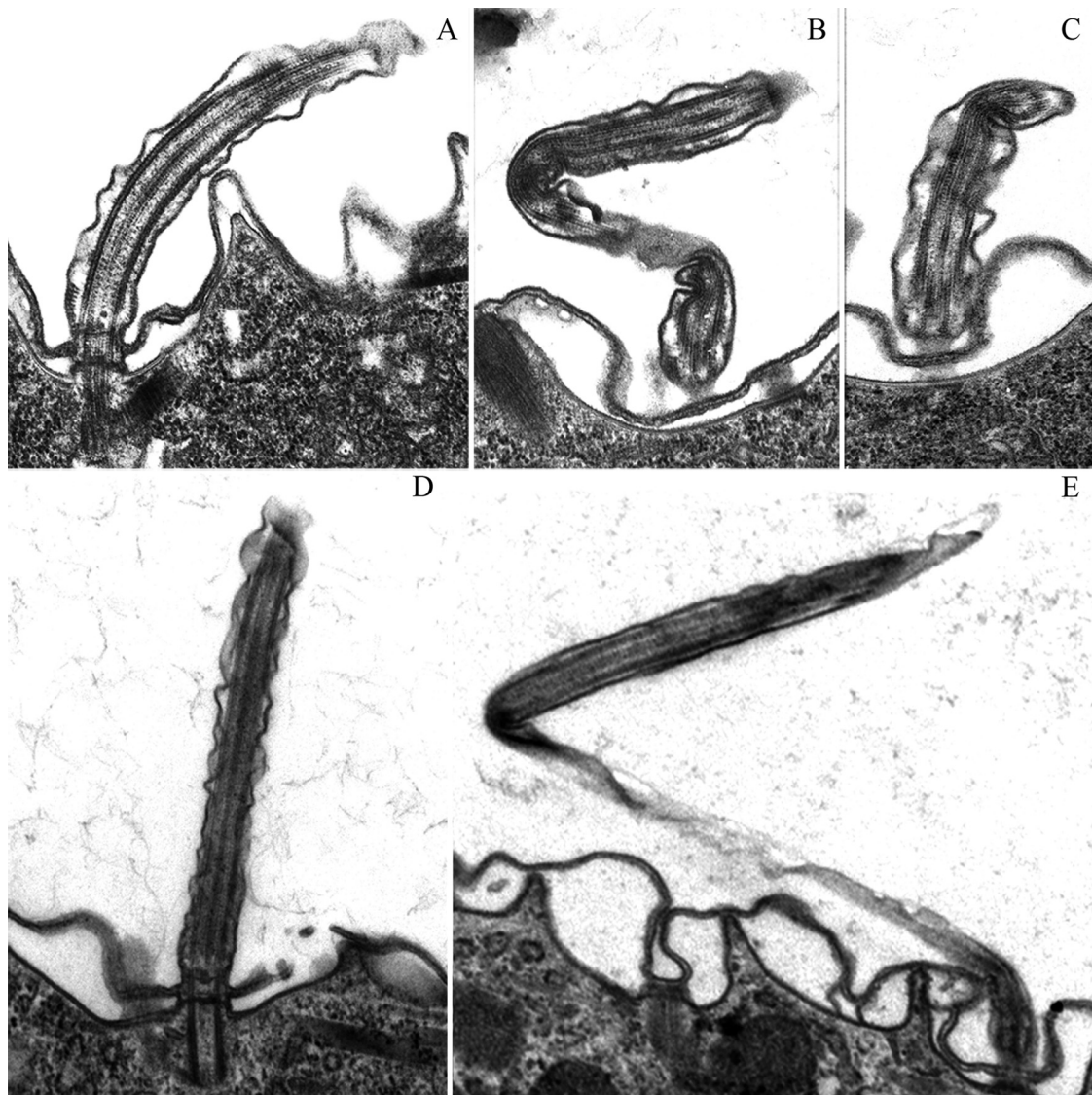


FIG. 8. Ciliary defects in Bug22p-depleted paramecia observed in electron microscopy. (A to C) Observation of RNAi after 24 h. Magnification, $\times 22,000$. (D and E) Observations of RNAi after 48 h. Magnification, $\times 16,000$. (A and D) Control ND7 RNAi. (B, C, and E) BUG22 RNAi. Note the unusual bending of cilia in Bug22p-depleted cells, as viewed in the longitudinal axis of cilia in panels B and E and with contiguous longitudinal and cross sections of the same cilium in panel C. Images of exaggerated ciliary bending were never observed earlier and were detected for the first time in Bug22p-depleted cells. Among 140 longitudinal sections of cilia from 15 different cells, 43 cilia displayed such exaggerated bending, whereas none were observed in a similar number of sections in the control.

gene with a cilium-specific X box in its promoter is found in *C. elegans* (6), which possesses only sensory cilia. Bug22p peptides have also been detected in the primary cilium-derived photo-sensory rod outer segments of mouse retina (21). Although the role of Bug22p in sensory and primary cilia is unknown, we can infer from its localization in the motile cilia of *Paramecium* that Bug22p plays a role associated with axonemal doublets and the ciliary membrane, potentially important for sensory and signaling functions.

The presence of *BUG22* genes in higher plants devoid of centrioles and cilia, with such sequence conservation, suggests that functions other than those associated with centrioles and cilia also exist for Bug22p. However, although higher plants have lost centrioles, they have kept microtubule-organizing

centers, as well as proteins with conserved domains characteristic of centrosomes, such as TOF, LisH, and PLL motif-containing proteins (TONNEAU1 in plants, FOP and OFD in animals) and centrin 2 (4). The prominent interactions of microtubules with the plasma membrane in these cells could be a clue to the function of Bug22p in plants, by extrapolation of what we have found in motile cilia.

Interestingly, coevolution of the centrin 2 and Bug22p proteins seems to have occurred in eukaryotes. Among the different centrin genes that exist in eukaryotes, the centrin 2 form is found, not only in species with centrioles and cilia, in functional association with these organelles, but also, as a derived form called centrin 2', in higher plants lacking centrioles and cilia (7). No Bug22p or centrin 2 sequences are present in fungi

devoid of centrioles and cilia, which rely on centrin 3 isoforms only for centrosomal-like function of the spindle pole body. If we call BUG22 the form found with the ciliary consensus and BUG22' the form with the plant consensus, a good parallel thus appears between the evolution of the centrin 2 and BUG22 families. A functional relationship between Bug22 and centrin 2 proteins may exist and should now be investigated to understand the essential role of Bug22p.

ACKNOWLEDGMENTS

We thank Joëlle Marie for hosting our HeLa cell cultures; François Lacroute for sharing his human cDNA libraries; J. Salisbury for the generous gift of the 20H5 antibody; Serge Amselem, Estelle Escudier, and Denise Escalier for helpful advice on ciliary-beating measurements; and Jean-Pierre Denizot and the Imagif platform for providing electron microscopy facilities. We are grateful to Janine Beisson and Linda Sperling for critical reading of the manuscript.

A fellowship to C.L. from the Ministère de la Recherche and grant NT05_1522 from the Agence Nationale de la Recherche are gratefully acknowledged.

REFERENCES

1. Arnaiz, O., A. Malinowska, C. Klotz, L. Sperling, M. Dadlez, F. Koll, and J. Cohen. 2009. Cildb: a knowledgebase for centrosomes and cilia. Database doi:10.1093/database/bap022.
2. Arnaiz, O., S. Cain, J. Cohen, and L. Sperling. 2007. *ParameciumDB*: a community resource that integrates the *Paramecium tetraurelia* genome sequence with genetic data. *Nucleic Acids Res.* 35:D439–D444.
3. Aury, J. M., O. Jaillon, L. Duret, B. Noël, C. Jubin, B. M. Porcel, B. Ségurens, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444:171–178.
4. Azimzadeh, J., P. Nacry, A. Christodoulidou, S. Drevensek, C. Camilleri, N. Amiou, F. Parcy, M. Pastuglia, and D. Bouchez. 2008. *Arabidopsis* TONNEAU1 proteins are essential for preprophase band formation and interact with centrin. *Plant Cell* 20:2146–2159.
5. Beisson, J., J. C. Clérot, A. Fleury-Aubusson, N. Garreau de Loubresse, F. Ruiz, and C. Klotz. 2001. Basal body-associated nucleation center for the centrin-based cortical cytoskeletal network in *Paramecium*. *Protist* 152:339–354.
6. Blaque, O. E., E. A. Perens, K. A. Borojevich, P. N. Inglis, C. Li, A. Warner, J. Khattri, et al. 2005. Functional genomics of the cilium, a sensory organelle. *Curr. Biol.* 15:935–941.
7. Bornens, M., and J. Azimzadeh. 2007. Origin and evolution of the centrosome. *Adv. Exp. Med. Biol.* 607:119–129.
8. Broadhead, R., H. R. Dawe, H. Farr, S. Griffiths, S. R. Hart, N. Portman, M. K. Shaw, et al. 2006. Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* 440:224–227.
9. Cohen, J., A. Adoutte, S. Grandchamp, L. M. Houdebine, and J. Beisson. 1982. Immunocytochemical study of microtubular structures throughout the cell cycle of *Paramecium*. *Biol. Cell* 44:35–44.
10. Dupuis-Williams, P., C. Klotz, H. Mazarguil, and J. Beisson. 1996. The tubulin gene family of *Paramecium*: characterization and expression of the alpha PT1 and alpha PT2 genes which code for alpha-tubulins with unusual C-terminal amino acids, GLY and ALA. *Biol. Cell* 87:83–93.
11. Errabolu, R., M. A. Sanders, and J. L. Salisbury. 1994. Cloning of a cDNA encoding human centrin, an EF-hand protein of centrosomes and mitotic spindle poles. *J. Cell Sci.* 107:9–16.
12. Follit, J. A., F. Xu, B. T. Keady, and G. J. Pazour. 2009. Characterization of mouse IFT complex B. *Cell Motil. Cytoskeleton* 66:457–468.
13. Galvani, A., and L. Sperling. 2002. RNA interference by feeding in *Paramecium*. *Trends Genet.* 18:11–12.
14. Gerdes, J. M., E. E. Davis, and N. Katsanis. 2009. The vertebrate primary cilium in development, homeostasis, and disease. *Cell* 137:32–45.
15. Giot, L., J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. L. Hao, et al. 2003. A protein interaction map of *Drosophila melanogaster*. *Science* 302:1727–1736.
16. Keller, L. C., S. Geimer, E. Romijn, J. Yates, I. Zamora, and W. F. Marshall. 2009. Molecular architecture of the centriole proteome: the conserved WD40 domain protein POC1 is required for centriole duplication and length control. *Mol. Biol. Cell* 20:1150–1166.
17. Keller, L. C., E. P. Romijn, I. Zamora, J. R. Yates, and W. F. Marshall. 2005. Proteomic analysis of isolated *Chlamydomonas* centrioles reveals orthologs of ciliary-disease genes. *Curr. Biol.* 15:1090–1098.
18. Klotz, C., N. Garreau de Loubresse, F. Ruiz, and J. Beisson. 1997. Genetic evidence for a role of centrin-associated proteins in the organization and dynamics of the infraciliary lattice in *Paramecium*. *Cell Motil. Cytoskeleton* 38:172–186.
19. Kung, C., S. Y. Chang, Y. Satow, J. van Houten, and H. Hansma. 1975. Genetic dissection of behavior in *Paramecium*. *Science* 188:898–904.
20. Lechtreck, K. F., P. Delmotte, M. L. Robinson, M. J. Sanderson, and G. B. Witman. 2008. Mutations in Hydin impair ciliary motility in mice. *J. Cell Biol.* 180:633–643.
21. Liu, Q., G. Tan, N. Levenkova, T. Li, E. N. Pugh, J. J. Rux, D. W. Speicher, and E. A. Pierce. 2007. The proteome of the mouse photoreceptor sensory cilium complex. *Mol. Cell. Proteomics* 6:1299–1317.
22. Murayama, T., Y. Toh, Y. Ohshima, and M. Koga. 2005. The dyf-3 gene encodes a novel protein required for sensory cilium formation in *Caenorhabditis elegans*. *J. Mol. Biol.* 346:677–687.
23. Ostrowski, L. E., K. Blackburn, K. M. Radde, M. B. Moyer, D. M. Schlatter, A. Moseley, and R. C. Boucher. 2002. A proteomic analysis of human cilia: identification of novel components. *Mol. Cell. Proteomics* 1:451–465.
24. Pazour, G. J. 2004. Comparative genomics: prediction of the ciliary and basal body proteome. *Curr. Biol.* 14:R575–R577.
25. Pazour, G. J., N. Agrin, J. Leszyk, and G. B. Witman. 2005. Proteomic analysis of a eukaryotic cilium. *J. Cell Biol.* 170:103–113.
26. Pedersen, L. B., M. S. Miller, S. Geimer, J. M. Leitch, J. L. Rosenbaum, and D. G. Cole. 2005. *Chlamydomonas* IFT172 is encoded by FLA11, interacts with CREB1, and regulates IFT at the flagellar tip. *Curr. Biol.* 15:262–266.
27. Rijkers, T., and U. Rütger. 1996. Sequence and expression pattern of an evolutionarily conserved transcript identified by gene trapping. *Biochim. Biophys. Acta* 1307:294–300.
28. Ruiz, F., A. Krzywicka, C. Klotz, A. Keller, J. Cohen, F. Koll, G. Balavoine, and J. Beisson. 2000. The *SM19* gene, required for duplication of basal bodies in *Paramecium*, encodes a novel tubulin, eta-tubulin. *Curr. Biol.* 10:1451–1454.
29. Sanders, M. A., and J. L. Salisbury. 1994. Centrin plays an essential role in microtubule severing during flagellar excision in *Chlamydomonas reinhardtii*. *J. Cell Biol.* 124:795–805.
30. Satir, P., and S. T. Christensen. 2008. Structure and function of mammalian cilium. *Histochem. Cell Biol.* 129:687–693.
31. Skouri, F., and J. Cohen. 1997. Genetic approach to regulated exocytosis using functional complementation in *Paramecium*: identification of the *ND7* gene required for membrane fusion. *Mol. Biol. Cell* 8:1063–1071.
32. Smith, J. C., J. G. B. Northey, J. Garg, R. E. Pearlman, and K. W. M. Siu. 2005. Robust method for proteome analysis by MS/MS using an entire translated genome: demonstration on the ciliome of *Tetrahymena thermophila*. *J. Proteome Res.* 4:909–919.
33. Sonneborn, T. M. 1970. Methods in *Paramecium* research. *Methods Cell Physiol.* 4:241–339.
34. Sperling, L., G. Keryer, F. Ruiz, and J. Beisson. 1991. Cortical morphogenesis in *Paramecium*: a transcellular wave of protein phosphorylation involved in ciliary rootlet disassembly. *Dev. Biol.* 148:205–218.
35. Sprague, J., L. Bayraktaroglu, D. Clements, T. Conlin, D. Fashena, K. Frazer, M. Haendel, et al. 2006. The Zebrafish Information Network: the zebrafish model organism database. *Nucleic Acids Res.* 34:D581–D585.
36. Stolc, V., M. P. Samanta, W. Tongprasit, and W. F. Marshall. 2005. Genome-wide transcriptional analysis of flagellar regeneration in *Chlamydomonas reinhardtii* identifies orthologs of ciliary disease genes. *Proc. Natl. Acad. Sci. U. S. A.* 102:3703–3707.
37. Timmons, L., D. L. Court, and A. Fire. 2001. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* 263:103–112.
38. Wehland, J., and K. Weber. 1987. Turnover of the carboxy-terminal tyrosine of alpha-tubulin and means of reaching elevated levels of detyrosination in living cells. *J. Cell Sci.* 88:185–203.
39. Woodland, H. R., and A. M. Fry. 2008. Pix proteins and the evolution of centrioles. *PLoS One* 3:e3778.

SUPPLEMENTAL FIGURES

Figure S1. Alignment of the four *Paramecium* *BUG22* gene DNA sequences and the human cDNA. The four *Paramecium* *BUG22* genes, which encode the same protein, as well as the human cDNA, are aligned to show up their differences and provide divergent sequences to design specific RNAi probes. Lines in red and green represent short RNAi probes respectively specific for inactivation of the *Paramecium* and human genes. Lines in purple and blue represent long RNAi probes to respectively inactivate the *Paramecium* *BUG22* paralogs *a/b* and *c/d*. The *Paramecium* introns are highlighted in pale blue.

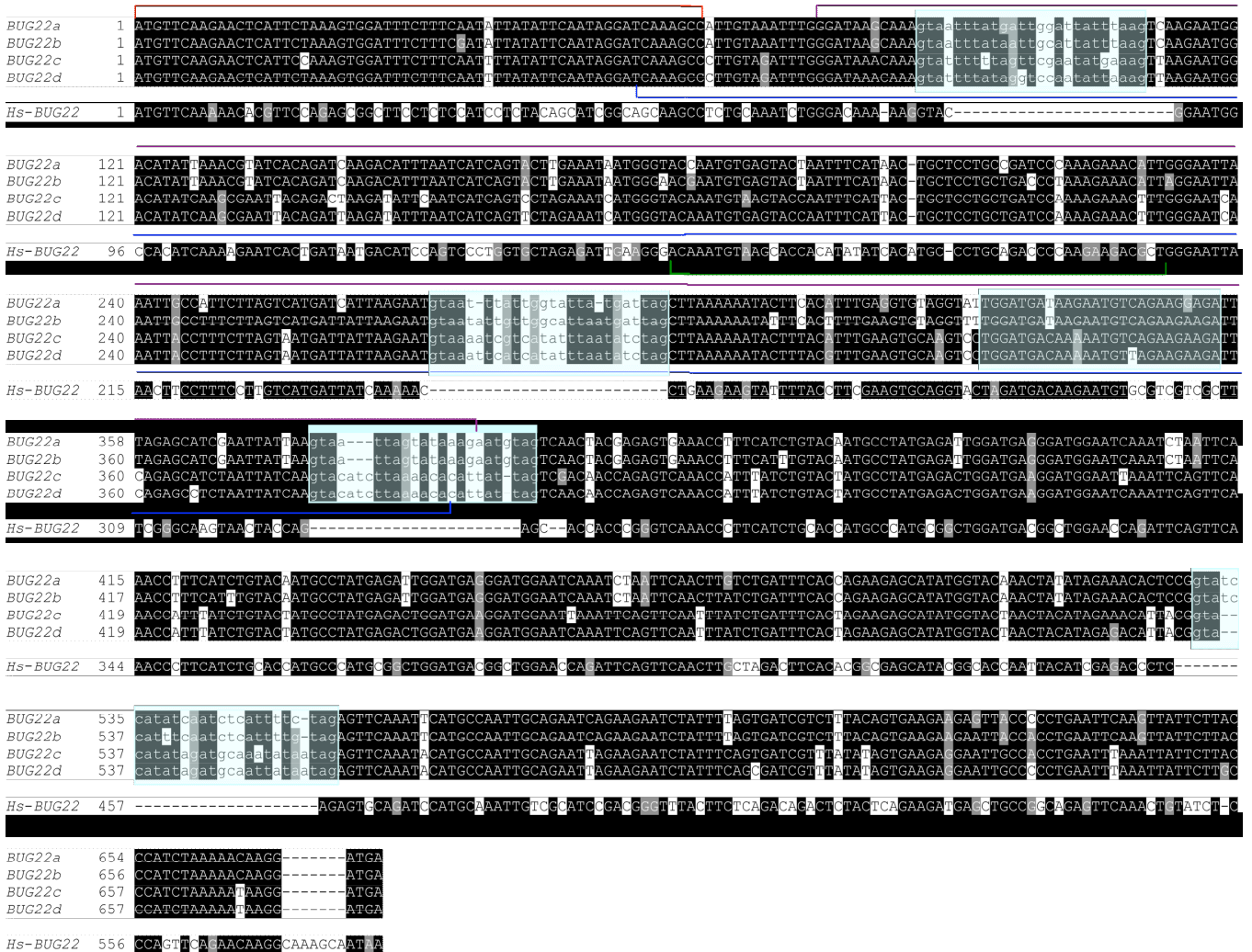


Figure S2. Centriolar localization of a human Bug22p in HeLa cells. A. HeLa cells expressing the GFP-HsBug22p protein were immuno-labeled with the 20H5 monoclonal antibody that recognizes centrin, which is a cytological marker of centrioles. Note the co-localization of Bug22p with centrin. B: Untransformed HeLa cells double labeled with GTL3 and the 20H5 anti-centrin showing co-localization of the markers. C. HeLa cells expressing the GFP-HsBug22p protein were immuno-labeled with the anti-GTL3 antibody that recognizes Bug22p. The immuno-labeling perfectly coincides with the GFP fluorescence. Bar = 10 μ m.

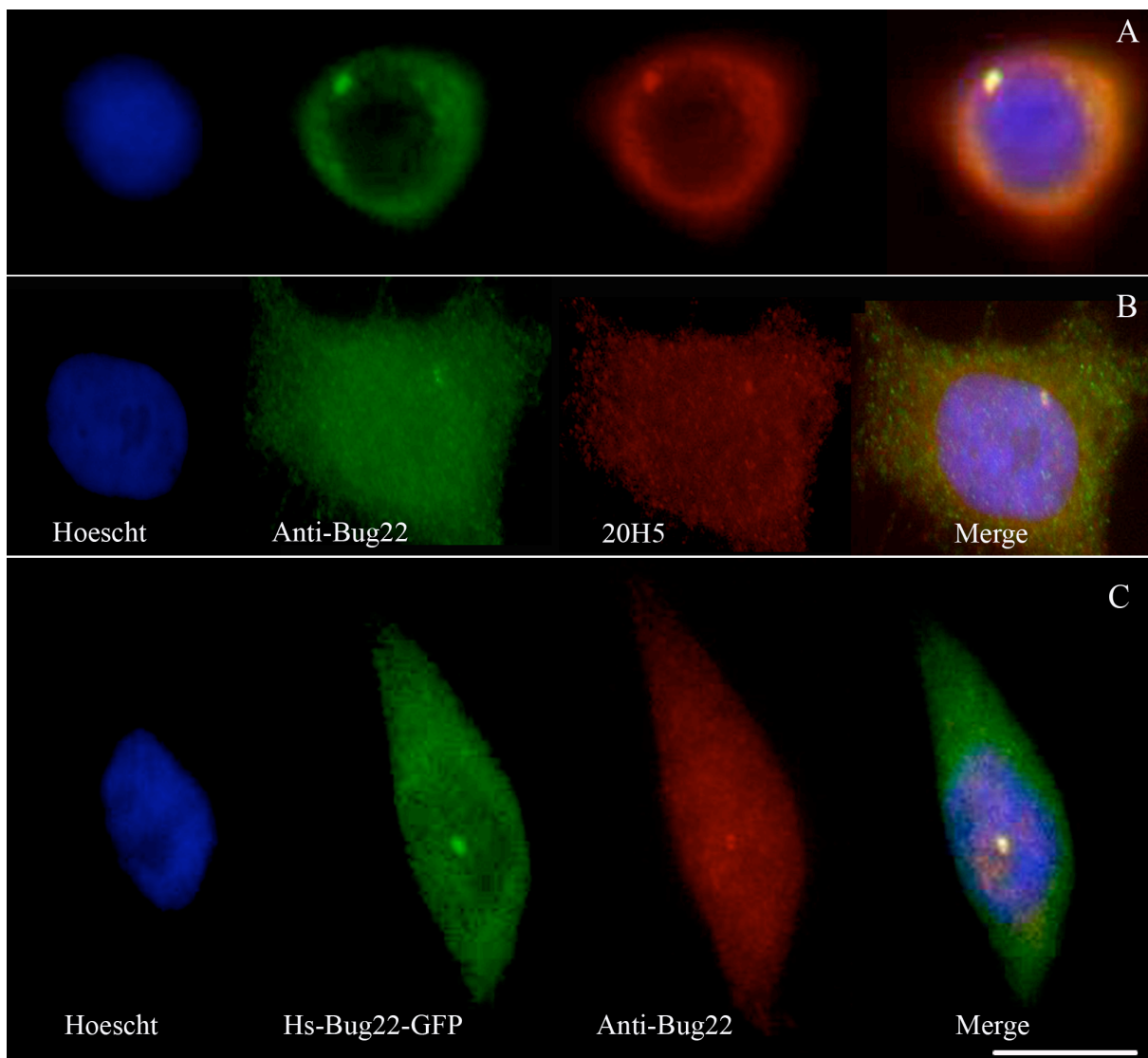


Figure S3. Effect of BUG22 RNAi on basal body and ciliary rootlet organization at two days of RNAi. A and D. Control Nd7p-depleted cells. B, C, E and F. Bug22p-depleted cells. A-C. ID5 anti-tubulin labeling basal bodies. D-F. Anti-ciliary rootlet labeling. Note the disorganization of basal body arrangement revealed by the observation of erratic ciliary rootlet orientations in Bug22p-depleted cells (highlighted in insets). Bar = 10 μ m.

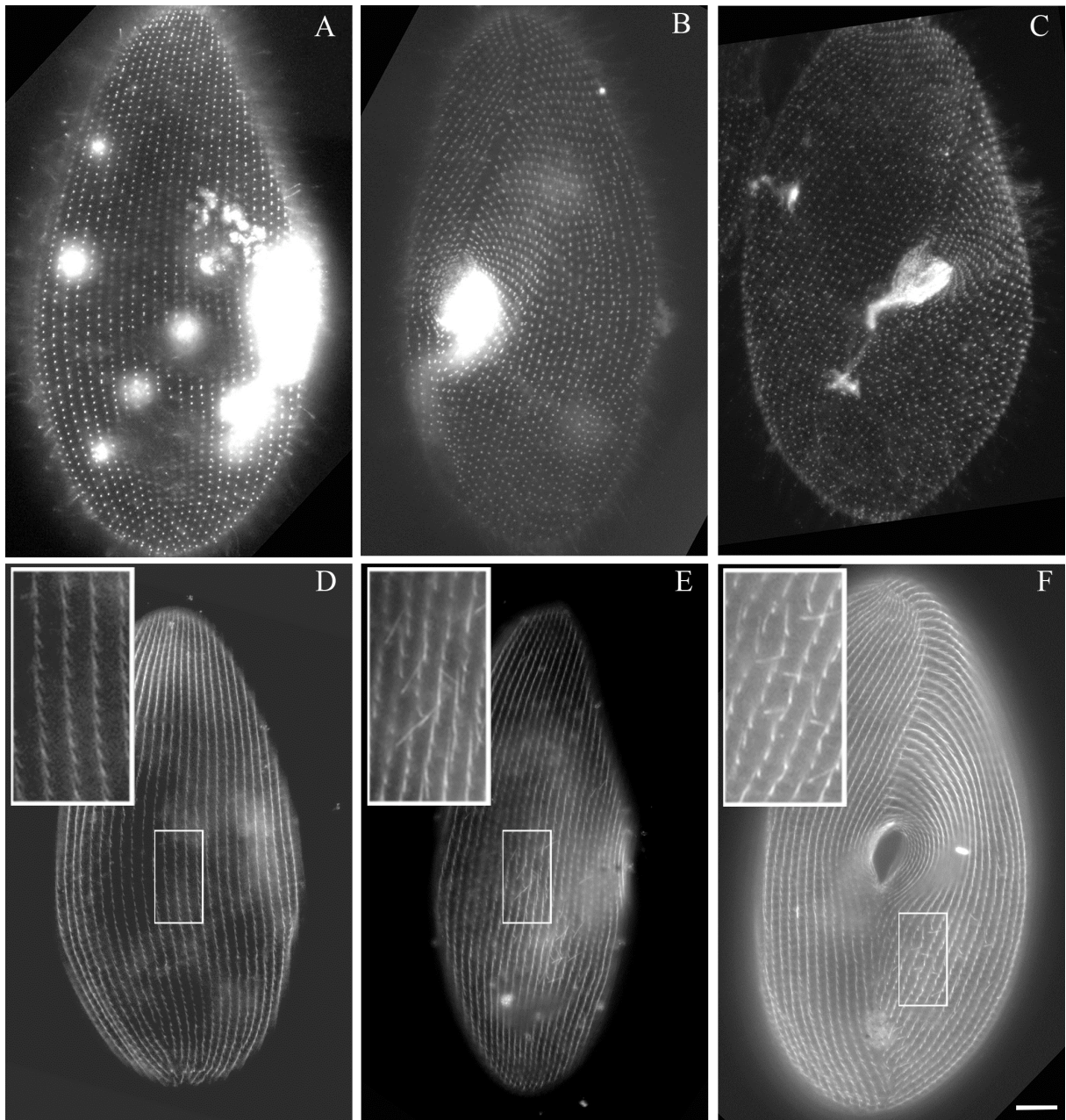
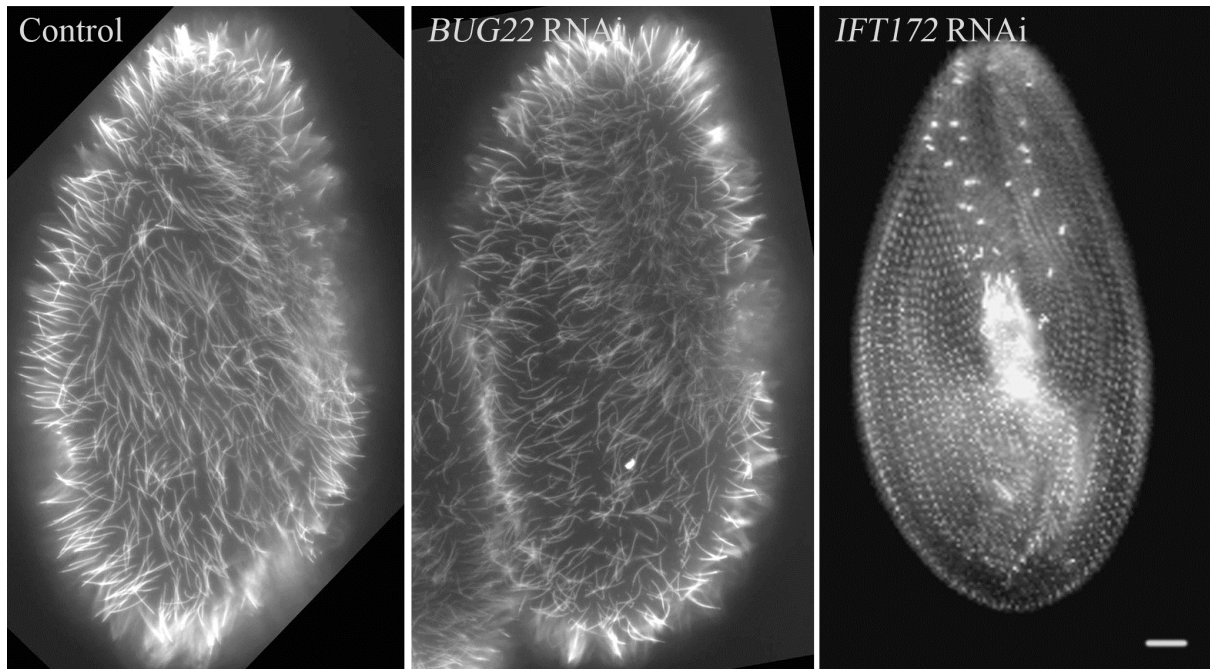


Figure S4. Comparison of ciliatures of Bug22p-depleted and Ift172p-depleted cells at two days of RNAi. A. Control cell depleted in its Nd7p protein with normal cilia as viewed in immunofluorescence using an anti-Paramecium tubulin. B. Bug22p-dpleted cell harboring normal cilia. C. In contrast, Ift172p-depleted cells have lost all their cilia and display only some remnants on their anterior ventral field (arrow). Bar = 10 μ m.



La Dystrophie Myotonique de type 1 (DM1) est causée par l'expansion anormale d'un triplet CTG dans la partie 3'UTR du gène *DMPK*, entraînant l'agrégation du transcrite mutant dans des inclusions ribonucléoprotéiques appelées foci. D'après plusieurs études structurales sur des courtes répétitions CUG, il a été proposé que les expansions CUG se replient en une structure en tige-boucle qui interfère avec l'activité de plusieurs facteurs liés au métabolisme de l'ARN et altère leur fonction cellulaire. Le facteur d'épissage muscleblind-like 1 (MBNL1) a été identifié par sa capacité à interagir avec les répétitions CUG. In vivo, ces répétitions entraînent la séquestration de cette protéine aboutissant en une déplétion nucléaire. Un autre facteur d'épissage, la CUG Binding protein (CUGBP1), est également impliqué dans la pathologie. Au lieu d'être séquestré par les répétitions, la stabilité protéique de CUGBP1 est augmentée dans les tissus DM1 entraînant un gain d'activité pour ce facteur. La séquestration de MBNL1 et la stabilisation de CUGBP1 résultent en la dérégulation de l'épissage alternatif de plusieurs transcrits musculaires et du cerveau et la réexpression d'isoformes protéiques fœtales dans les tissus adultes. Cependant, de récentes études suggèrent que d'autres facteurs ou voies de signalisation que celles faisant intervenir MBNL1 et CUGBP1 pourraient être impliquées dans la pathologie DM1.

Le but de mon travail de thèse a été d'identifier de nouveaux facteurs ayant la capacité d'interagir avec les répétitions CUG. À l'aide d'une purification sur chromatographie d'affinité utilisant un ARN contenant 95 répétitions CUG comme appât, nous avons identifié l'ARN hélicase p68/DDX5. p68 fait partie de la famille des protéines DEAD-box, caractérisée par un core protéique conservé constitué de neuf domaines hautement conservés, dont le motif DEAD, à l'origine du nom de ces protéines. p68 est impliquée dans de nombreux aspects du métabolisme de l'ARN, dont la transcription, l'épissage, l'export, la traduction et la dégradation des ARN. Nous avons montré que p68 colocalise avec les foci CUG dans un modèle cellulaire exprimant la partie 3'UTR du gène *DMPK* contenant de longues répétitions CTG. Nous avons identifié que p68 augmente l'interaction de MBNL1 sur les répétitions CUG et une structure secondaire particulière d'un élément régulateur de l'ARN pré-messager *cardiac Troponin T (TNNT2)*, dont l'épissage est dérégulé dans la pathologie. L'insertion de mutations dans le core de l'hélicase de p68 abolit l'effet de p68 sur la fixation de MBNL1 ainsi que la colocalisation de p68 avec les expansions CUG in vivo, suggérant que le remodelage des structures secondaires ARN de manière ATP-dépendante par p68 facilite l'interaction de MBNL1. Nous trouvons également que la compétence de p68 pour réguler l'inclusion de l'exon alternatif 5 de *TNNT2* dépend de l'intégrité des sites de fixation de MBNL1.

Nous proposons que p68 agit comme un modificateur de l'activité de MBNL1 sur ces cibles d'épissage ainsi que sur les expansions CUG à l'origine de la pathologie.

Myotonic Dystrophy type I (DM1) is caused by an abnormal expansion of CTG triplets in the 3' UTR of the *DMPK* gene, leading to the aggregation of the mutant transcript in nuclear RNA foci. Based on structural studies on short CUG repeats, it has been proposed that expanded CUG repeats fold into an imperfect hairpin structure that interferes with the activities of RNA binding proteins and alters their normal cellular function. The muscleblind-like 1 protein (MBNL1) was identified by its ability to bind to CUG repeats. It has been shown that the expanded mutant transcript promotes the sequestration of the MBNL1 splicing factor in nuclear RNA foci. CUGBP1 is another splicing factor that is involved in DM1. Instead of being sequestered by the repeats, the steady-state level of CUGBP1 is increased in DM1 tissues, leading to a gain of activity of the protein. The sequestration of MBNL1 and the up-regulation of CUGBP1 in DM1 results in the misregulation of alternative splicing of a subset of muscle and brain-specific transcripts, leading to the re-expression of fetal isoforms in adult tissues. However, several recent studies suggest that factors or signaling pathways other than MBNL1 and CUGBP1 could be involved in DM1 pathogenesis.

The aim of my thesis was to isolate new factors that bind to CUG repeats. Using an affinity chromatography strategy with an RNA containing 95 pure CUG repeats, we identified the RNA helicase p68 (DDX5). p68 is a prototype of DEAD-box RNA helicase proteins. This family is characterized by a conserved core, consisting of nine conserved motifs including the DEAD signature, which gives rise to the name of these proteins. p68 is involved in many aspects of RNA metabolism including transcription, RNA processing, RNA export, translation and mRNA degradation. We showed that p68 colocalized with RNA foci in cells expressing the 3'UTR of the *DMPK* gene containing expanded CTG repeats. We found that p68 increased MBNL1 binding onto pathological repeats and the stem-loop structure regulatory element within the cardiac Troponin T (*TNNT2*) pre-mRNA, splicing of which is misregulated in DM1. Mutations in the helicase core of p68 prevented both the stimulatory effect of the protein on MBNL1 binding and the colocalization of p68 with CUG repeats, suggesting that remodeling of RNA secondary structure through an ATP-dependent manner by p68 facilitates MBNL1 binding. We also found that the competence of p68 for regulating *TNNT2* exon 5 inclusion depended on the integrity of MBNL1 binding sites.

We propose that p68 acts as a modifier of MBNL1 activity on splicing targets and pathogenic RNA.