# Analysis of Exonic Regions Involved in Nuclear Localization, Splicing Activity, and Dimerization of Muscleblind-like-1 Isoforms

Helene Tran, Nathalie Gourrier, Camille Lemercier-Neuillet, Claire-Marie Dhaenens, Audrey Vautrin, Francisco José Fernandez-Gomez, Ludovic Arandel, Celine Carpentier, Hélène Obriot, Sabiha Eddarkaoui, et al.

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


# Analysis of Exonic Regions Involved in Nuclear Localization, Splicing Activity, and Dimerization of Muscleblind-like-1 Isoforms*[S]

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

**Hélène Tran[‡§1,2], Nathalie Gourrier[¶1,3], Camille Lemercier-Neuillet[‖**‡‡1], Claire-Marie Dhaenens[‡§], Audrey Vautrin[¶4], Francisco José Fernandez-Gomez[‡§], Ludovic Arandel[‖**‡‡], Céline Carpentier[‡§], Hélène Obriot[‡§], Sabiha Eddarkaoui[‡§], Lucie Delattre[‡§], Edwige Van Brussels[‡§], Ian Holt[§§‖‖], Glenn E. Morris[§§‖‖], Bernard Sablonnière[‡§], Luc Buée[‡§5], Nicolas Charlet-Berguerand[¶¶], Susanna Schraen-Maschke[‡§], Denis Furling[‖**‡‡5], Isabelle Behm-Ansmant[¶5], Christiane Branlant[¶5,6], Marie-Laure Caillet-Boudin[‡§7], and Nicolas Sergeant[‡§7,8]**

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

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

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

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

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

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

[1] These authors contributed equally to this work.
[2] Fellow of the INSERM and the Nord/Pas de Calais Region.
[3] Association Française contre le Myopathies postdoctoral fellow.
[4] Fellow of the French Ministry for Youth, National Education and Research.
[5] Staff scientists at the CNRS.
[6] To whom correspondence may be addressed. E-mail: christiane.branlant@maem.uhp-nancy.fr.
[7] Staff scientists at the INSERM.
[8] To whom correspondence may be addressed: INSERM U837, Alzheimer and Tauopathies, Université Lille Nord de France, USDL, IMPRT, Batiment Biserte, 1, place de Verdun, rue Polonowski, F-59045 Lille, France. E-mail: nicolas.sergeant@inserm.fr.

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

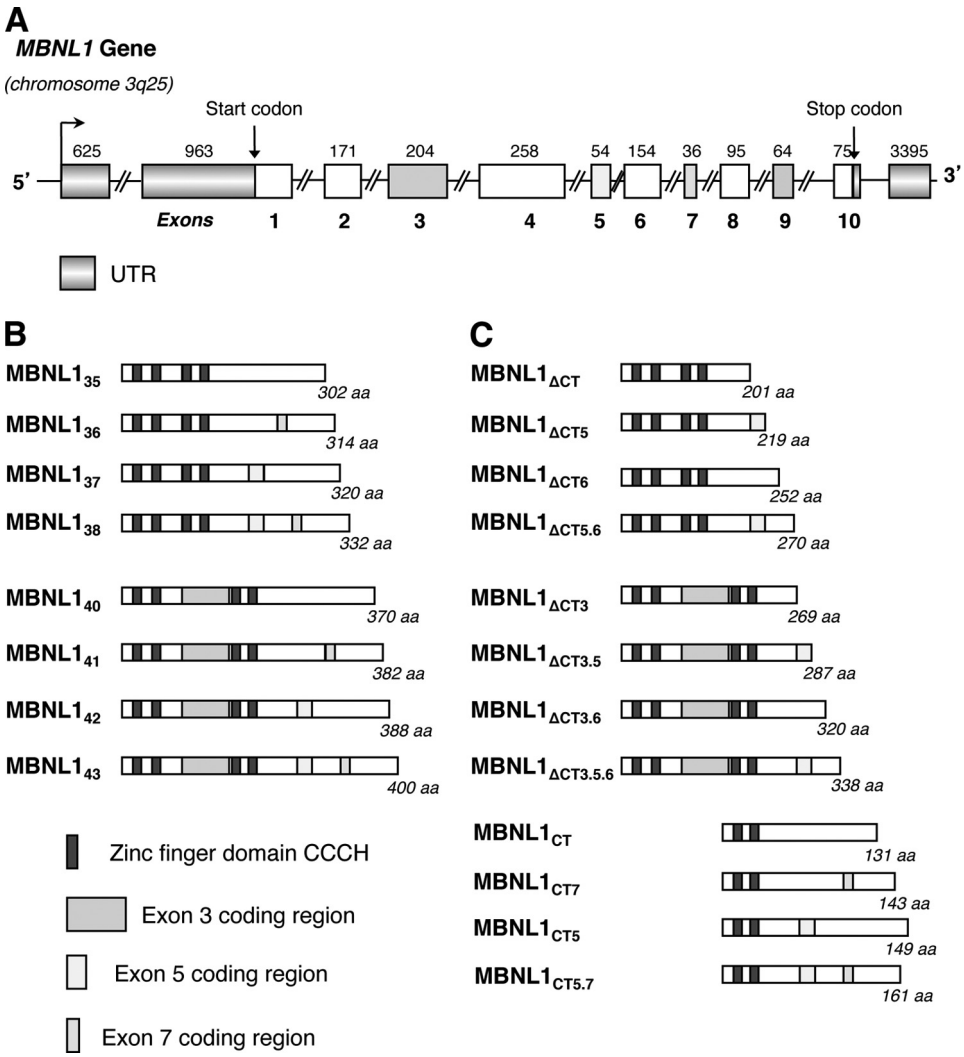



FIGURE 1. **Alternative splicing of MBNL1 generates protein diversity.** *Clear gray boxes* represent UTR. *Gray boxes* represent cassette exons. *Empty boxes* represent constitutive exons. *A,* genomic DNA organization of human *MBNL1* gene. Order and names of exons used in this work are shown. The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

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

MBNL1 is the best example of splicing regulatory factors known to be involved in development and splicing deregulation in disease. Originally identified in *Drosophila melanogaster*, MBNL1 is described as being essential for the terminal differentiation of photoreceptors and muscles (12, 13). In humans and in mice, the splicing factor MBNL1 has been shown to participate in the postnatal remodeling of skeletal muscle and of the developing heart by controlling the developmentally regulated switch of a key set of pre-mRNAs (9, 14, 15). MBNL1 binds to pre-mRNAs containing YGCY sequence elements (16, 17) and promotes either the inclusion or the exclusion of alternative exons depending on the 5′ or 3′ localization of *cis*-regulatory elements (17). For instance, the inclusion of exon 5 in human cardiac troponin T (*hcTNT/TNNT2*) mRNA is inhibited by MBNL1, whereas the inclusion of exon 11 in insulin receptor (*IR/INSR*) mRNA is promoted by MBNL1 (16). MBNL1 binds to the polypyrimidine tract of hcTNT intron 4 and competes with the splicing factor U2AF for its binding site, explaining the MBNL1 inhibition of exon 5 inclusion (18). Conversely, by binding to elements located 3′ to the alternative exon 11 in IR pre-mRNA (*IR/INSR*), MBNL1 enhances the inclusion of this exon (17, 19, 20).

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

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

In the present study, in an effort to define precisely the respective roles of the amino acid sequences encoded by the alternative cassette exons 3, 5, and 7 and constitutive exon 6 of human MBNL1 protein isoforms, we applied a large variety of *in vitro* and *in cellulo* approaches. This allowed us to determine the cellular localization, splicing activity, and RNA binding properties of the eight main normal MBNL1 isoforms as well as of a large number of MBNL1 variant proteins. Experiments were performed using both HeLa cells and human myoblasts that did or did not express CUG expansions. Our results provide a better definition of (i) the respective roles of the amino acid sequences encoded by exons 3 and 5 in the RNA binding property, nuclear localization and splicing regulatory property of MBNL1; (ii) the involvement of the exon 6-encoded sequence in MBNL1 nuclear retention and *in cellulo* splicing regulatory activity; (iii) the identification of a possible role of the exon 7-encoded sequence in MBNL1 self-dimerization.

## EXPERIMENTAL PROCEDURES

*Plasmids and Cloning*—Full-length MBNL1 variant constructs ($MBNL1_{35}$, $MBNL1_{36}$, $MBNL1_{37}$, $MBNL1_{38}$, $MBNL1_{40}$, $MBNL1_{41}$, $MBNL1_{42}$, and $MBNL1_{43}$) used in this study have been described previously (24). All truncated forms of MBNL1 were derived from the appropriate full-length MBNL1 cDNAs and cloned using standard techniques. Briefly, MBNL1 cDNA was amplified with DYNAzyme EXT™ Taq polymerase (Finnzymes, Espoo, Finland). The same forward primer sequence was used for all constructs with a truncated C-terminal tail: 5′-atggctgttagtgtcacacca-3′; the reverse primer sequences used were 5′-catggcagctgcggt-3′ for the $MBNL1_{\Delta CT}$ and $MBNL1_{\Delta CT3}$ constructs; 5′-caggtcaaaggttgcctc-3′ for $MBNL1_{\Delta CT5}$, and 5′-ctggtgggagaaatgctgt-3′ for $MBNL1_{\Delta CT6}$, $MBNL1_{\Delta CT3.6}$ and $MBNL1_{\Delta CT3.5.6}$, respectively. The cDNA of constructs with a truncated N-terminal tail were amplified using 5′-atgtaccaggtccctaagaagaa-3′ as the forward primer and 5′-ctacatctgggtaacatacttgtg-3′ as the reverse primer. PCR products were cloned into the PCR8/GW/TOPO TA cloning vector (Invitrogen) and recombined with LR Clonase in the expression plasmid pDEST53/GW, pcDNA3.1/nV5 plasmid for C-terminal truncated mutants, or in pACTII and pASIIΔΔ previously converted using the Gateway system according to the manufacturer's instructions. All plasmid DNA was double-stranded sequenced and purified using the Nucleobond® AX endotoxin free kit (Macherey Nagel, Düren, Germany). MBNL1 splicing activity was assessed using two minigenes described previously: the RTB300 minigene containing exon 5 *hcTNT* (26) and the *INSR* minigene containing exon 11 of the human IR (27).

*Cell Culture and Transfection*—HeLa cells (ECACC 93021013) were grown in monolayer cultures in 6-well plates in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen) supplemented with 10% fetal calf serum, 50 units/ml penicillin, 50 μg/ml streptomycin, and 4 mmol/liter glutamine, at 37 °C in a humidified (5% $CO_2$) incubator. Cells were plated in 6-wells plates, grown to ~70% confluence, and transiently co-transfected with 1 μg of minigene plasmid DNA and 3 μg of MBNL1 plasmid DNA, in triplicate, using the FuGENE HD transfection reagent (Roche Diagnostics) according to the manufacturer's instructions. Human muscle cells were isolated from skeletal muscle biopsies or autopsies as described (28), in accordance with French legislation on ethics. Wild-type (WT) and DM1 myoblasts were grown in Ham's F10 medium supplemented with 20% fetal calf serum and 5 μg/ml gentamycin (Invitrogen), at 37 °C and under 5% $CO_2$.

*Immunofluorescence Microscopy*—HeLa cells and myoblasts were transiently transfected with 2 μg of the different MBNL1 constructs, using the transfection reagent FuGENE HD according to the manufacturer's instructions. After 48 h, cells were fixed with 4% paraformaldehyde in 0.1 M phosphate buffer (PBS) for 15 min at room temperature. Cells were washed three times with PBS. The coverslips were mounted onto slides using the mounting medium Vectashield (Vector Laboratories) with DAPI. Images were acquired with a Zeiss apotome microscope or a Leica confocal microscope. All data were analyzed using Photoshop Element 6 software (Adobe).

*Fluorescent in Situ Hybridization (FISH)*—FISH was performed as described (29) using a Cy3-labeled peptide nucleic acid $(CAG)_7$ probe. Images were captured using a Leica confocal microscope and software (Leica microsystems) and processed with Adobe Photoshop software.

*Protein Extraction, Nuclear Fractionation, and Western Blot Analysis*—HeLa cells were lysed 48 h after transfection in radioimmune precipitation assay buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Nonidet P-40, 0.5% deoxycholate) supplemented with protease inhibitors (Complete mini; Roche Applied Science). Lysates were sonicated, homogenized at 4 °C, and centrifuged at 12,000 × *g* at 4 °C for 20 min. Myoblasts were washed in 20 mM HEPES at pH 7.9 with 5 mM NaF, 1 mM $Na_2MoO_4$, and 0.1 mM EDTA and lysed with the

addition of 0.5% Nonidet P-40. Nuclear and cytoplasmic fractions were separated by centrifugation (30 s at 16,000 × *g*). Each fraction was lysed in radioimmune precipitation assay buffer supplemented with protease inhibitors (Complete mini) and sonicated. Protein concentration was determined with the BCA protein assay kit (Pierce). HeLa samples were prepared under reducing conditions (NuPage sample buffer with sample reducing agent; Invitrogen) and heated to 100 °C for 10 min. 10 μg of proteins were loaded and separated on a 4–12% BisTris polyacrylamide NuPage gel (Invitrogen) and transferred onto Hybond nitrocellulose membranes (GE Healthcare) using the XCell™ II blot module (Invitrogen). Membranes were blocked with 5% skimmed milk in TBS-T (Tris-buffered saline Tween 20: 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 0.1% Tween 20) or with 2% fetal calf serum (FCS) in PBS-T (PBS + 0.1% Tween 20) and incubated overnight at 4 °C with a primary antibody raised against the GFP tag (anti-GFP B-2; Santa Cruz Biotechnology) or with a primary antibody raised against MBNL1 (MB1a) (30, 31), β-tubulin (Sigma T4026), or H3 histone (Cell Signaling 9715). Horseradish peroxidase-conjugated antibodies (Vector) were used as secondary antibodies, and horseradish peroxidase activity was detected with the ECL™ detection kit on Hyperfilms (GE Healthcare) according to the manufacturer's instructions. Western blots were reproduced at least three times.

*RNA Extraction and Semiquantitative Analysis*—Total RNA was isolated from cells 48 h after transfection using a total RNA extraction kit (Nucleospin® RNA II kit; Macherey Nagel, Düren, Germany). RNA concentration was determined by measuring the absorbance at 260 nm using Nanodrop ND1000 technology (Labtech). RT-PCR was performed in triplicate with 1 μg of total RNA using random hexamers (5 μM/liter) and the M-MLV reverse transcriptase (Invitrogen) according to standard protocols. No DNA amplification was observed in RT controls. PCR was carried out in a final volume of 25 μl, with a 10 or 15 pM concentration of each primer as described previously (26, 27), 1.5 mM $MgCl_2$, and 1 unit of Taq polymerase (Invitrogen), under the following conditions: 5 min at 94 °C, 22–30 cycles of a 1-min denaturation step at 94 °C, annealing for 2 min at 65 °C, 2 min of extension, and 7 min of final extension at 72 °C. 18 S rRNA was used as an internal control. Reaction products were resolved by electrophoresis using a 5% or 8% polyacrylamide gel, and bands were stained with SYBR Gold (Invitrogen). The intensity of SYBR Gold luminescence was measured using a FluoroImager scanner (Claravision). PCR experiments were repeated at least three times. For more details of MBNL1 RT-PCR and two-dimensional gel electrophoresis, see supplemental Experimental Procedures.

*Yeast Double- and Triple-hybrid Assays*—For yeast double-hybrid assays, the appropriate pACT-II and pAS2ΔΔ plasmids were introduced into haploid *Saccharomyces cerevisiae* test strains (CG929 or Y190 and Y187), which were then crossed. Diploid cells were plated on double- or triple-selection media (-Leu, -Trp, or -Leu, -Trp, and -His), and growth was assessed 3 days later. For yeast triple-hybrid assays, *S. cerevisiae* L40-coat strains carrying various pACTII::MBNL1 plasmids were mated with R40-coat strains carrying the plasmid pIIIMS2-2::5′hcTNT or pIIIMS2–2::$CUG_{21}$ as a control. Diploids were selected on -Trp, -Leu, and -Ura media and were tested using -Leu -Ura -His media. For both yeast double-hybrid and yeast triple-hybrid assays, the strength of the interactions tested was evaluated by the addition of different concentrations of 3-amino-1,2,4-triazole (3AT), which competes with the *HIS3* gene product.

*Recombinant Protein Preparation*—The plasmids pGEX-6P1-$MBNL1_{40}$, pGEX-6P1-$MBNL1_{\Delta CT3}$, and pGEX-6P1-$MBNL1_{\Delta CT}$ were constructed for the overexpression of recombinant GST-$MBNL1_{40}$, GST-$MBNL1_{\Delta CT3}$, and GST-$MBNL1_{\Delta CT}$ proteins in *E. coli*. To this end, MBNL1 ORFs were PCR-amplified from pET28-MBNL1 plasmids using oligonucleotides that generated a BamHI and an XhoI restriction site, respectively. The amplified DNA fragments were digested and inserted into the pGEX-6P1 plasmid (Novagen) cut by the same two nucleases. The resulting constructs were used to transform *E. coli* BL21 CodonPlus (DE3) RIL cells (Stratagene). Recombinant GST-$MBNL1_{40}$, GST-$MBNL1_{\Delta CT3}$, and GST-$MBNL1_{\Delta CT}$ proteins were purified from *E. coli* BL21 cells under native conditions, using glutathione-Sepharose 4B as recommended by the manufacturer (Pharmacia). The GST-MBNL1 proteins bound to glutathione-Sepharose 4B beads were cleaved on the beads using 80 units of PreScission protease (GE Healthcare)/ml of glutathione-Sepharose bead suspension. Cleavage was carried out overnight at 4 °C. The purified proteins were dialyzed against buffer D (150 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 20 mM HEPES, pH 7.9, 10% glycerol), and aliquots were stored at −80 °C.

*Electrophoresis Mobility Shift Assays*—Cold hcTNT RNA was synthesized by *in vitro* transcription of a PCR fragment corresponding to the 50 nucleotides upstream of exon 5 in hcTNT pre-mRNA, using T7 RNA polymerase. The hcTNT RNA substrate used for complex formation was 5′ end-labeled using [γ-$^{32}$P]ATP and T4 polynucleotide kinase and was gel-purified. For complex formation, 1.2 fmol of 5′ end-labeled hcTNT RNA was preincubated in 7 μl of buffer D (150 mM KCl, 4.5 mM $MgCl_2$, 20 mM HEPES, pH 7.9, 10% glycerol) for 10 min at 65 °C, in the presence of 2 μg of tRNA, followed by slow cooling. Then, 4 μl of MBNL1 recombinant protein was added to the renaturated RNA for a final protein concentration of between 0 and 2,150 nM. The binding reactions were incubated for 30 min at room temperature and loaded onto an 8% (38/2) nondenaturing polyacrylamide gel containing 44.5 mM borate, 1 mM EDTA, 44.5 mM Tris borate, pH 8.3, and 5% glycerol. Gels were run for 45 min at 15 mA. The amount of radioactivity in the bands, corresponding to free and bound complexed RNA, was estimated using a PhosphorImager and ImageQuant software. Using these values, apparent $K_d$ values were determined with SigmaPlot software (SPSS Science Software). The overall $K_d$ and the number of binding sites were determined by Hill plotting; log [PR/R] *versus* log[P] (where P = protein concentration; PR = bound RNA, and R = free RNA).

*Statistical Analyses*—Statistical analyses were performed using unpaired *t* tests with two tails, with the help of Prism software (GraphPad Software, San Diego, CA).

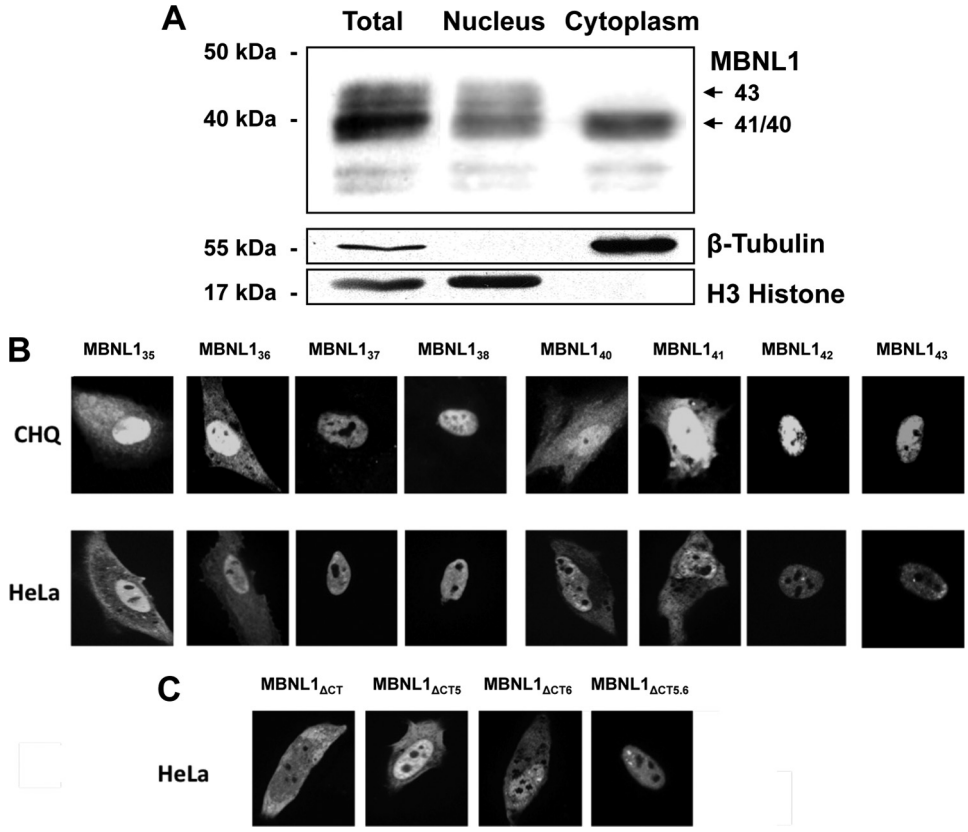



FIGURE 2. **Protein sequence encoded by exon 5 modulates MBNL1 nuclear localization.** *A*, nucleocytoplasmic localization of endogenous MBNL1 isoforms in human myoblasts. Nuclear fractionation was performed on human myoblasts that endogenously expressed several isoforms of MBNL1. The total fraction (*Total*) is in the *left lane*, the nuclear fraction (*Nucleus*) is in the *middle*, and the cytoplasmic fraction (*Cytoplasm*) is in the *right lane*. Molecular masses are indicated on the *left*, and antibodies used are on the *right*. To visualize MBNL1 isoforms, we used a monoclonal antibody raised against the region encoded by exon 3 (31). β-Tubulin was used as a cytoplasmic marker and H3 histone as a nuclear marker to assess the different fractions. *B*, immunofluorescence of GFP-tagged MBNL1 isoforms in human myoblast and HeLa cells. GFP-tagged MBNL1 isoforms were transiently expressed in human myoblasts and HeLa cells. Fluorescence appears in *gray* and indicates transfected cells and reveals the subcellular localization of MBNL1. The name of the isoform used is indicated at the *top* of each cell, using the same nomenclature as in Fig. 1. *C*, nucleocytoplasmic localization of truncated MBNL1 mutants in HeLa cells. GFP-tagged mutants were transiently expressed in HeLa cells. Fluorescence in *gray* indicates transfected cells and reveals the subcellular localization of truncated MBNL1. The name of the construct used is indicated at the *top* of each cell.

## RESULTS

*Nuclear Retention of MBNL1 Protein Depends on Segments Encoded by Both Exons 5 and 6*—To analyze the subcellular distribution of the various MBNL1 isoforms, nucleocytoplasmic fractionation was performed using human primary muscle cells (Fig. 2*A*). The fractionation protocol was assessed using histone H3 as a nuclear protein marker and β-tubulin as a marker for the cytosolic fraction. MBNL1 isoforms were detected with a monoclonal antibody that specifically recognizes the protein segment encoded by exon 3 (31). After gel electrophoresis, six bands containing proteins ranging from 30 kDa to 45 kDa were observed in the whole cell lysate (Fig. 2*A*, *Total lane*). Based on the epitope recognized by the antibody used, MBNL1 polypeptides detected must contain the exon 3-encoding sequence. According to their apparent molecular masses, the two top bands resolved at 45 kDa were expected to correspond to MBNL1$_{43}$ and MBNL1$_{42}$ isoforms, whereas the two middle bands (a major and a minor) at 40 kDa were expected to correspond to the MBNL1$_{41}$ and MBNL1$_{40}$ isoforms. The two bottom bands resolved at less than 35 kDa were expected to correspond to proteolytic products because the shortest identified MBNL1 isoform containing exon 3 is MBNL1$_{40}$. Based on mRNA analysis, the MBNL1 isoforms 40 S and 41 S were described previously (22); however, they were never identified at the protein level. To support our protein assignment, additional RT-PCR and two-dimensional gel electrophoresis analysis were performed. First, RT-PCR amplification of the MBNL1 mRNAs followed by DNA sequencing identified five distinct mRNAs encoding MBNL1 isoforms 43, 41, 40, 40 S, and 41 S (supplemental Fig. 1*A*). A second Western blot analysis of the two-dimensional gel using MBNL1 antibody revealed three series of four spots aligned in the same range of isoelectric points (supplemental Fig. 1*B*). These spots correspond to MBNL1 isoforms 40, 41, and 43 because they have very close theoretical isoelectric points (pI of 8.9–9.1). The isoforms 40 S and 41 S have theoretical pI values of 8.92–8.96 and a molecular mass of 35–37 kDa, indicating that the isovariants with a molecular mass below 30 kDa and pI values lower than 8.0 were therefore most probably catabolic products of MBNL1 rather than 40 S or 41 S MBNL1 isoforms. Altogether, our results indicate that MBNL1 isoforms 40, 41 and 43 are expressed in human myoblasts.

The MBNL1 isoforms detected in the whole cell lysate were also found in the nuclear fraction (Fig. 2*A*, *Nucleus lane*). How-

ever, the upper bands assigned as isoform 43 and containing the exon 5 region were not detected in the cytoplasm. In contrast, the proteins that we assigned to isoforms 41 and 40, which do not contain the exon 5 region, were present in both the cytoplasmic and nuclear fraction. In agreement with this subcellular localization, the exon 5 region was previously proposed to be involved in nuclear localization of MBNL1. Therefore, these data bring additional support to our isoform assignment and extend to human muscle cells the observations previously made in mice (9) and chickens (15).

To complete the analysis of the sequence dependence of MBNL1 nuclear retention, we investigated the nucleocytoplasmic distribution of all WT MBNL1 isoforms, using GFP-tagged constructs (Fig. 1*B*). First, the eight-tagged isoforms were ectopically expressed in myoblasts and HeLa cells (Fig. 2*B*). As expected, all of the MBNL1 isoforms containing exon 5 ($MBNL1_{37,\ 38,\ 42,\ and\ 43}$) were detected in the nucleus but not in the cytoplasm. Conversely, isoforms lacking this sequence ($MBNL1_{35,\ 36,\ 40,\ and\ 41}$) showed both nuclear and cytoplasmic distributions. The localization of MBNL1 isoforms was not modified by the inclusion of exon 3 and/or 7. These results further support the idea that the region encoded by exon 5 regulates the nuclear retention of MBNL1. However, it did not exclude the possibility that other regions of the protein might have contributed to MBNL1 subcellular localization. To test this hypothesis, we generated four GFP-tagged truncated MBNL1 constructs: $MBNL1_{\Delta CT}$, $MBNL1_{\Delta CT5}$, $MBNL1_{\Delta CT6}$, and $MBNL1_{\Delta CT5.6}$, respectively (Fig. 1*C*). The proteins expressed by these constructs all contained the four zinc finger domains (exons 1, 2, and 4 were present in all of them), with or without exon 5. The truncated protein $MBNL1_{\Delta CT6}$ contained, in addition, the entire exon 6 region, but not the exon 5 region, whereas $MBNL1_{\Delta CT5.6}$ contained both the exon 5 and 6 regions. These truncated MBNL1 proteins were expressed in HeLa cells, and their subcellular localizations were visualized using GFP fluorescence. Surprisingly, $MBNL1_{\Delta CT}$ and $MBNL1_{\Delta CT5}$ showed both nuclear and cytoplasmic distribution, even though $MBNL1_{\Delta CT5}$ contained the exon 5 region (Fig. 2*C*). In contrast, the $MBNL1_{\Delta CT5.6}$ protein containing both the exon 5 and 6 regions had an exclusive nuclear localization. Furthermore, its variant lacking the exon 5 region had both a nuclear and cytoplasmic distribution. Taken together, our results show that the exon 5 region is necessary but not sufficient for the nuclear retention of MBNL1. The downstream region encoded by exon 6 is necessary to mediate the nuclear retention of MBNL1 isoforms. Therefore, the nuclear retention of MBNL1 depends on both the exon 5 and exon 6 regions.

*MBNL1 Exon 3-encoded Region Is Necessary to Modulate Alternative Splicing of Both hcTNT and Human IR Pre-mRNAs*—We next investigated whether the splicing function of MBNL1 was isoform-dependent. To assess the splicing activity of the various MBNL1 isoforms, we took advantage of two well characterized MBNL1 targets, hcTNT exon 5 and IR exon 11, and used their respective minigenes (26, 27). The eight MBNL1 isoforms (including or not the exon 3, 5, and 7 regions) were coexpressed with hcTNT or IR minigenes in HeLa cells. The transient expression of MBNL1 constructs yielded very similar expression levels of MBNL1 proteins (Fig. 3*A*): only the two largest isoforms ($MBNL1_{42}$ and $MBNL1_{43}$) were expressed at slightly lower levels. The splicing products of the hcTNT exon 5 and IR exon 11 minigenes were analyzed by RT-PCR, and the data obtained demonstrated that the presence or absence of the MBNL1 exon 5 and 7 regions had no detectable effect on the MBNL1 splicing regulatory activity (Fig. 3*B*). In contrast, the presence of the exon 3 region in MBNL1 isoforms was found to be essential for the repression of hcTNT exon 5 inclusion and the activation of IR exon 11 inclusion. Therefore, we concluded that the exon 5 and 7 regions have no obvious splicing regulatory function, in sharp contrast to the exon 3 region (Fig. 3*B*). To rule out potential bias introduced by the GFP tag, we generated similar constructs in which the GFP tag was replaced by five amino acids (V5 tag). Similar results were obtained with these constructs (data not shown).

Our data suggest that the sequence encoded by exon 3 is essential to the splicing regulatory activity of protein MBNL1. This amino acid sequence links the second and third zinc finger domains (Fig. 1*B*), and this linker sequence has previously been shown to be required for the efficient binding of MBNL1 to RNAs containing CUG repeats (22). Therefore, we hypothesized that this linker sequence also increases MBNL1 affinity for its target sites in pre-mRNAs. To address this question, a yeast triple-hybrid assay was developed (Fig. 3*D*). In this assay, the interaction between RNA and protein was detected by the activity of the *HIS3* reporter gene. The strength of the interaction was assessed by the addition of increasing concentrations of 3AT to the medium. 3AT is a competitive inhibitor of the activity of the *HIS3* gene product. The MBNL1-Gal4 transcriptional activation domain fusion proteins used in these assays corresponded to: (i) the $MBNL1_{40}$ isoform, (ii) its homolog lacking exon 3, $MBNL1_{35}$, (iii) the $MBNL1_{41}$ isoform, (iv) its homolog lacking exon 3, $MBNL1_{36}$, (v) $MBNL1_{\Delta CT3}$ including the exon 3 region, and (vi) $MBNL1_{\Delta CT}$ lacking the exon 3 region. The RNA used in the tests corresponded to the 3′ end of the fourth intron of hcTNT pre-mRNA (hcTNT RNA). This RNA has been previously shown to interact with the $MBNL1_{40}$ isoform (16, 19). The truncation of the C-terminal region in $MBNL_{40}$ did not change the affinity for the hcTNT RNA (Fig. 3*D*). Conversely, this affinity was strongly reduced in the absence of the exon 3 region (proteins $MBNL1_{35}$, $MBNL1_{36}$, and $MBNL1_{\Delta CT}$). Indeed, almost no interaction of $MBNL1_{35}$ and $MBNL1_{\Delta CT}$ was detected at a 3AT concentration of 0.5 mM (Fig. 3*D*). Interestingly, at 1 mM 3AT concentration, a stronger interaction was observed for $MBNL1_{41}$ including the exon 7 region compared with $MBNL1_{40}$ missing this region. In addition, at this 3AT concentration, some remaining interaction was also detected for $MBNL1_{36}$ compared with $MBNL1_{\Delta CT}$. Therefore, the exon 7 region may partially compensate for the absence of exon 3 (Fig. 3*D*). One possible explanation will be given in a following paragraph describing properties of the exon 7 region.

We then confirmed the role of exon 3 region in MBNL1 binding to its target sites in hcTNT pre-mRNAs by gel shift experiments using a 5′ end-labeled hcTNT RNA (1.2 fmol) and the recombinant proteins $MBNL1_{40}$, $MBNL1_{\Delta CT3}$, and $MBNL1_{\Delta CT}$ at concentrations ranging from 100 to 2,150 nM

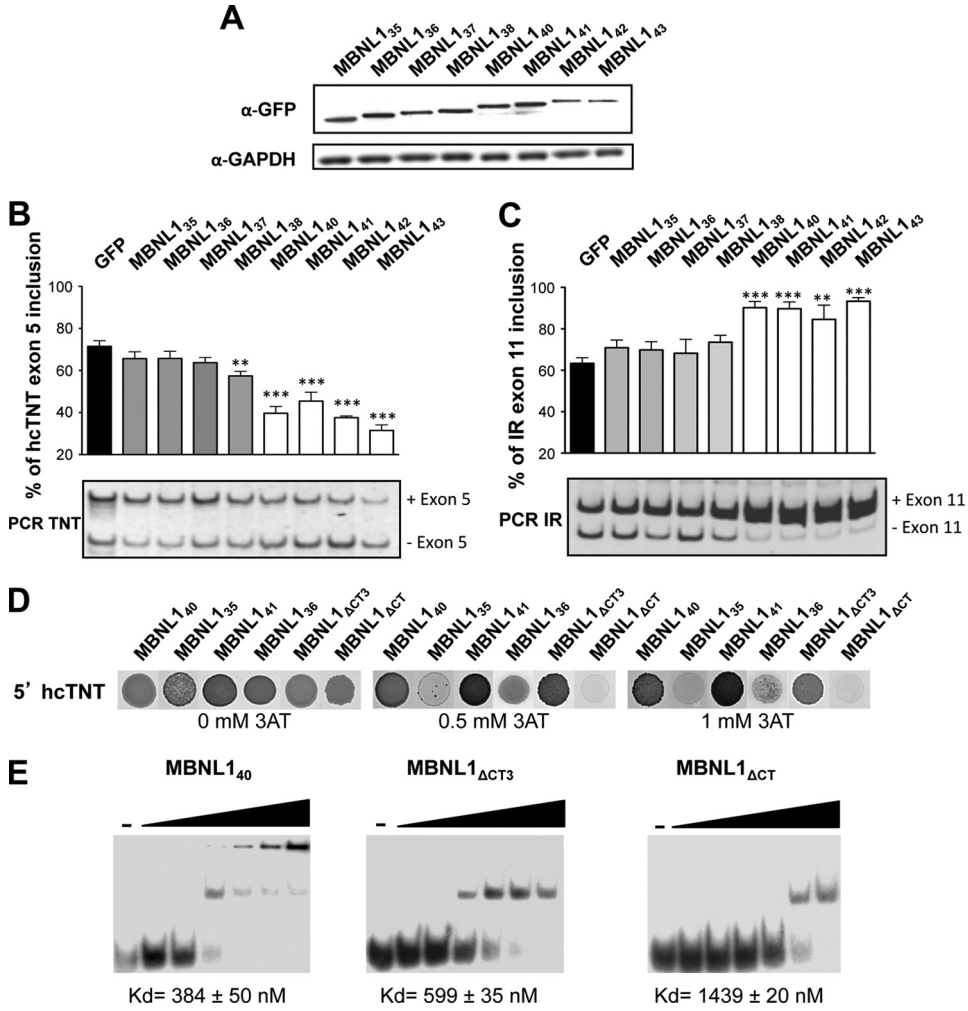



FIGURE 3. **Protein sequence encoded by exon 3 modulates MBNL1 splicing and RNA binding properties.** *A*, eight GFP-tagged MBNL1 isoforms were transiently expressed in HeLa cells. The level of expression among the different isoforms was verified to be the same by Western blotting, using an antibody raised against the GFP tag. GAPDH was used as a loading control. *B* and *C*, MBNL1 isoforms were co-expressed in HeLa cells in the presence of hcTNT exon 5 (*B*) or IR exon 11 (*C*) minigenes. The name of each isoform is indicated at the *top*. Cells were lysed 48 h after transfection, and splicing products were assayed by RT-PCR. Data are expressed as means ± S.E. (*error bars*). The percentage of inclusion is calculated as follows: ((mRNA + exon)/(mRNA + exon) + (mRNA − exon)) × 100. Results are derived from at least three independent experiments. An unpaired *t* test with two tails was used to compare values. Statistically significant results are indicated by ** ($p < 0.001$) and *** ($p < 0.0001$). *D*, triple-hybrid assays were performed using $MBNL1_{40}$, $MBNL1_{35}$, $MBNL1_{41}$, $MBNL1_{36}$, $MBNL1_{\Delta CT}$, and $MBNL1_{\Delta CT3}$ fusion proteins and 5′ hcTNT as the bait RNA. The strength of the interaction was evaluated by the addition of 3AT at a concentration of 0.5 or 1 mM. *E*, ribonucleoprotein complexes were formed by the incubation of 1.2 fmol of 5′-labeled hcTNT RNA in the presence of $MBNL1_{40}$, $MBNL1_{\Delta CT}$, or $MBNL1_{\Delta CT3}$ recombinant proteins at a concentration of 100, 215, 430, 750, 1,000, or 2,150 nM. All complexes were formed at room temperature in the presence of 2 μg of yeast tRNAs under the conditions described under "Experimental Procedures." *Lanes* (−) are control experiments in the absence of protein. Complexes were fractionated by electrophoresis in a nondenaturing 8% polyacrylamide gel.

(for incubation conditions, see "Experimental Procedures"). No gel shift experiments were performed with the $MBNL_{41}$, $MBNL1_{36}$, and $MBNL1_{35}$ recombinant proteins because of their very low solubility during the purification steps. In agreement with the above triple-hybrid data, protein $MBNL1_{\Delta CT}$, lacking the exon 3 region, bound to the hcTNT RNA with a lower affinity (apparent $K_d$ 1,439 ± 20 nM) compared with protein $MBNL1_{\Delta CT3}$, containing the exon 3 region (apparent $K_d$ 599 ± 35 nM) (Fig. 3*E*). The $MBNL1_{40}$ protein isoform, containing both the exon 3 region and additional C-terminal sequences, had a slightly lower apparent $K_d$ (384 ± 50 nM). In addition, due to its aggregation at high concentration, RNP complexes formed with this protein were retained in the loading well at high protein concentrations.

For the three proteins studied, the curves obtained by plotting the ratio of bound *versus* bound plus unbound RNA (PR/PR+R) against the protein concentration [P] (where PR = bound RNA and R = free RNA) indicated the presence of multiple binding sites that might be filled either individually, or in a co-operative manner because no binding intermediates were visible (supplemental Fig. 2*A*). By determining the overall $K_d$ and studying the binding co-operativity by Hill plots (log (PR/R) *versus* log [P]) (supplemental supplementary Fig. 2), we concluded that each of the studied proteins ($MBNL1_{40}$, $MBNL1_{\Delta CT3}$, and $MBNL1_{\Delta CT}$) binds co-operatively on two MBNL1 binding sites which are present in the hcTNT RNA.

Taken together, these results indicate that the exon 3 region is not absolutely required for MBNL1 RNA binding activity. However, in its absence, the affinity of MBNL1 for its target site in hcTNT pre-mRNA is reduced by a factor of about 2. Although we cannot exclude a possible interaction of the exon 3 region with splicing factors, the decreased RNA affinity of

MBNL1 isoforms lacking this region likely strongly participates in their reduced capabilities to regulate splicing.

*Exon 6 Region Is Also Needed for the Splicing Regulatory Property of MBNL1*—Having shown the importance of the alternative MBNL1 exon 3 region for MBNL1 splicing regulatory activity and the absence of the involvement of the amino acid regions coded by the two other alternative exons (5 and 7) in this process, we decided to investigate whether the amino acid regions encoded by some of the constitutive exons are also involved in splicing regulation. The contribution of the regions encoded by exons 1, 2, and 4 is obvious because they correspond to the two pairs of zinc finger domains needed for RNA binding (25). Therefore, we focused our study on the possible role of the exon 6 region, which was present in all of the isoforms tested for their capabilities to regulate splicing in the above experiments (Figs. 1*B* and 3*A*). To test for a possible effect of this region in both the presence or and the absence of the exon 3 or 5 region, we prepared a series of constructs expressing proteins lacking exon 3 and/or 5 region(s) ($MBNL1_{\Delta CT}$, $MBNL1_{\Delta CT3}$, $MBNL1_{\Delta CT6}$, $MBNL1_{\Delta CT5.6}$, $MBNL1_{\Delta CT3.6}$, and $MBNL1_{\Delta CT3.5.6}$ (Fig. 1*C*). These constructs were co-transfected in HeLa cells together with the hcTNT or IR minigenes, and the expression of the truncated proteins was verified by Western blot analysis (Fig. 4*A*). Minigene splicing products were assessed by RT-PCR 48 h after transfection. As expected based on the above results, MBNL1 variants lacking the exon 3 region had no detectable effect on the alternative splicing of the hcTNT exon 5 and IR exon 11 minigenes (Fig. 4, *B* and *C*). Interestingly, the presence of the exon 3 region conferred some splicing regulatory activity on the MBNL1 protein even in the absence of the exon 5 and 6 regions, revealing some splicing regulatory properties of the ΔCT3 N-terminal part of MBNL1. However, the simultaneous presence of both exons 3 and 6 regions was needed for recovering a full splicing activity (Fig. 4, *B* and *C*). In agreement with the above results, the deletion or addition of the exon 5 region had no significant effect on MBNL1 splicing regulatory properties. Similar results were obtained with V5-tagged MBNL1 isoforms (data not shown). Combined together, our results reveal the essential role of both the alternative exon 3 region and the constitutive exon 6 region for full MBNL1 splicing regulation activity.

*All MBNL1 Isoforms Are Trapped in DM1 Foci*—One of the primary causes of DM1 pathophysiology is the abnormal sequestration of MBNL1 by CUG repeats. To determine whether the trapping of MBNL1 in nuclear aggregates that is triggered by expanded CUG repeats was dependent on the presence of the exon 3-encoded region, we examined the ability of each MBNL1 isoform to co-localize with *foci* in primary muscle cells isolated from a muscle biopsy of a DM1 patient. CUG-RNA *foci* were visualized by FISH using a $(CAG)_7$-Cy3 probe. Various MBNL1 isoforms were ectopically expressed in their wild-type form and visualized using their GFP tag.

The whole GFP-tagged MBNL1 isoforms co-localized with discrete *foci* containing CUG-RNA repeats, indicating that MBNL1 isoforms with or without the exon 3 region were sequestered by CUG repeats (Fig. 5). This was surprising, considering the previously described low affinity of MBNL1 isoforms lacking the exon 3 region for CUG repeats (22). Noteworthy, the dual nucleocytoplasmic localization of the $MBNL1_{35,36,40,41}$ isoforms was maintained even in presence of *foci* (Fig. 5).

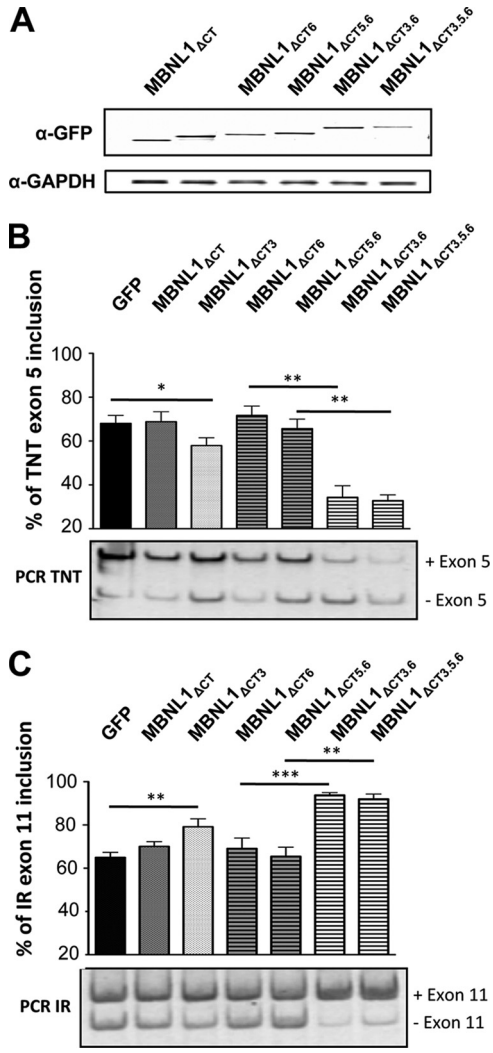



FIGURE 4. **MBNL1 splicing function requires the region encoded by MBNL1 exon 6.** *A*, truncated MBNL1 constructs were transiently expressed in HeLa cells. The name of each construct is indicated at the *top*. The regions expressed are indicated, and the presence of alternative exons is shown in *subscripts*. The level of expression of the different isoforms was verified to be the same by Western blotting using an antibody raised against the GFP tag. GAPDH was used as a loading control. *B* and *C*, truncated MBNL1 constructs were co-expressed in HeLa cells in the presence of hcTNT exon 5 (*B*) or IR exon 11 (*C*) minigenes. The name of each construct is indicated at the *top* and the regions expressed with the presence of alternative exons in *subscripts*. Cells were lysed 48 h after transfection, and splicing products were assayed by RT-PCR. Data are expressed as indicated for Fig. 3. An unpaired *t* test with two tails was used to compare values. Statistically significant results are indicated by * ($p < 0.05$), ** ($p < 0.001$), and *** ($p < 0.0001$). *Error bars*, S.E.

Our detection of all isoforms of MBNL1 in *foci* containing CUG expansions could be explained by the capability of all isoforms to interact directly with the CUG RNA expansions or by their capability to establish protein-protein interactions. To get insight into this question, we first compared the affinity for CUG repeats of all MBNL1 isoforms, using triple-hybrid assays in the same conditions as in Fig. 3*D*, but by replacing the hcTNT RNA by an RNA containing 21 successive CUG triplets ($CUG_{21}$ RNA; Fig. 6*A*). All MBNL1 isoforms interacted with

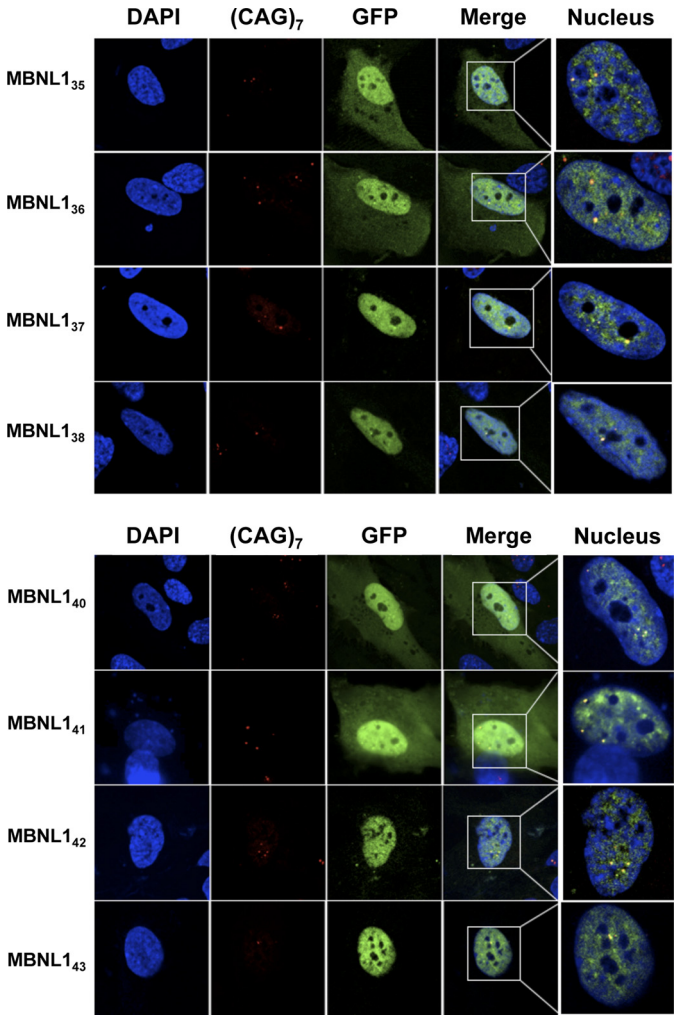



FIGURE 5. **All MBNL1 isoforms are trapped by the CUG repeats.** Eight GFP-tagged MBNL1 isoforms ($MBNL1_{35, 36, 37, 38, 40, 41, 42, 43}$) were transiently expressed in DM1 myoblasts. *Green fluorescence* indicates transfected cells. CUG repeats appear in *red* after FISH using a $(CAG)_7$-Cy3 probe. *Merged images* show that all MBNL1 isoforms co-localize with the *foci* of expanded DMPK transcripts. A higher magnification of nuclei is also showed in the *fifth column*.

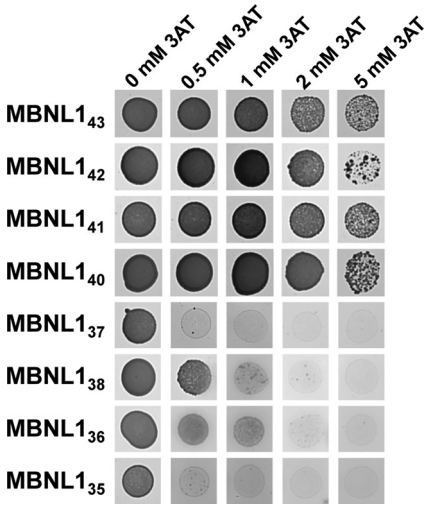


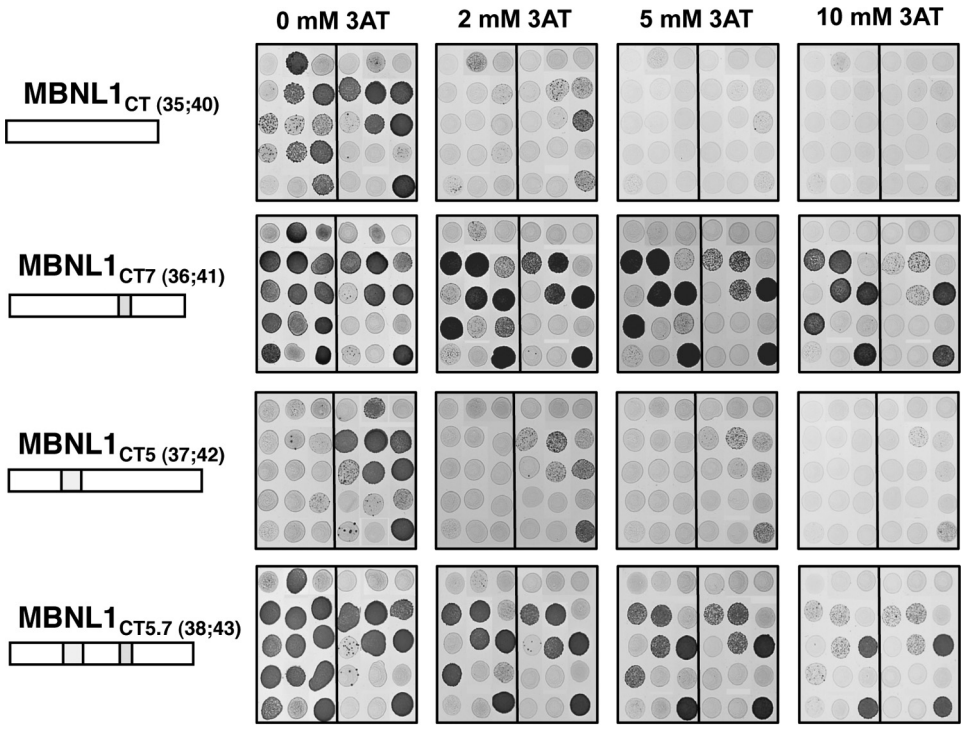

FIGURE 6. **MBNL1 isoforms that include the exon 3 region bind to $CUG_{21}$ RNA with higher affinity than those lacking this region.** Triple-hybrid assays were performed using $CUG_{21}$ RNA as the bait RNA and all MBNL1 isoforms as fusion proteins. The strength of the interaction was evaluated by the addition of 3AT at a concentration of 0.5, 1, 2, or 5 mM.

this RNA in the absence of 3AT. In agreement with data in Fig. 3*D*, all isoforms $MBNL1_{40, 41, 42, 43}$ containing the exon 3 region had a stronger affinity for the $CUG_{21}$ RNA than the ones missing this region (MBNL1 isoforms 35–37). Indeed, an interaction was detected up to 5 mM 3AT for MBNL1 isoforms 40–43 whereas this interaction was abolished at 2 mM 3AT for MBNL1 isoforms 35–37 (Fig. 6*A*). Similar to data obtained for the hcTNT RNA (Fig. 3*D*), the presence of the exon 7 region was found to increase the capability of MBNL1 isoforms missing the exon 3 region ($MBNL1_{38}$ and $MBNL1_{36}$) to interact with $CUG_{21}$ RNA (Fig. 6*A*). One possible explanation for this latter observation was that by enhancing MBNL1-MBNL1 interactions, the exon 7 region reinforces the co-operative binding of MBNL1 to the RNA. We therefore undertook a study of homo- and heterodimerization of the WT MBNL1 isoforms.

*Exon 7-Encoded Sequence Contributes to MBNL1 Dimerization*—Previous studies performed in yeast using two-hybrid assays or in mammalian cells have shown that MBNL1 self-interacts via its C-terminal region (amino acids 239–382) (32). The C-terminal domains of MBNL1 isoforms mainly differ one from the other by the presence or absence of the regions encoded by exons 5 and 7. To test whether these two amino acid sequences or constitutive amino acid sequences in the MBNL1 C-terminal region are implicated in protein dimerization, we performed two-hybrid assay experiments in yeast.

Each of the eight full-length MBNL1 isoforms studied above and each of the four different MBNL1 C-terminal regions were used in the assays. We also used the truncated $MBNL1_{\Delta CT}$ and $MBNL1_{\Delta CT3}$, versions of MBNL1, as controls of MBNL1 isoforms lacking the C-terminal domain. The four C-terminal domains all contained part of the exon 4 region located downstream to the zinc finger 4 and the exon 6, 8, and 10 regions with or without the exon 5 and/or 7 regions ($MBNL1_{CT}$, $MBNL1_{CT5}$, $MBNL1_{CT7}$, and $MBNL1_{CT5.7}$) (Figs. 1 and 7). To test the interaction capabilities of each protein pair in the two possible orientations, recombinant two-hybrid pAS2 and pACT2 plasmids, allowing the expression of the Gal4 DNA binding domain or Gal4 transcriptional activation domain in fusion with each of the tested proteins, were produced (Fig. 7). As for the three-hybrid experiments, protein-protein interactions were detected by the activity of the *HIS3* reporter gene, and their interaction strengths were challenged by the addition of increasing concentrations of 3AT. We tested the capability of homo- and heterodimerization of (i) the C-terminal domains and (ii) the C-terminal domain with full-length MBNL1 isoforms and $MBNL1_{\Delta CT}$ and $MBNL1_{\Delta CT3}$ proteins.

No interaction of the $MBNL1_{\Delta CT}$ and $MBNL1_{\Delta CT3}$ proteins was detected with any of the C-terminal domains or full-length MBNL1 proteins, even in the absence of 3AT. This result demonstrated the absence of interaction of the MBNL1 Zinc finger domains with the C-terminal domain (Fig. 7). In contrast, in the absence of 3AT, several of the C-terminal regions formed homo- or heterodimers with C-terminal domains or full-length proteins. The weakest interactions were detected for the all MBNL1 isoforms, such as $MBNL1_{CT}$ and $MBNL1_{CT5}$, which lack the exon 7 region (Fig. 7, *blue squares* compared with *orange squares*). These interactions almost disappeared at a 2




| pACT2 | 37 | ΔCT | pAS2 | 37 | ΔCT |
|---|---|---|---|---|---|
| 43 | 38 | CT | 43 | 38 | CT |
| 42 | 36 | CT7 | 42 | 36 | CT7 |
| 41 | 35 | CT5 | 41 | 35 | CT5 |
| 40 | ΔCT3 | CT5.7 | 40 | ΔCT3 | CT5.7 |

FIGURE 7. **Protein sequence encoded by exon 7 modulates MBNL1 dimerization properties.** The interaction of the four MBNL1 C-terminal tails (*A–D*) with all MBNL1 isoforms and with themselves was assessed by yeast two-hybrid assay. cDNAs encoding the proteins of interest were PCR-amplified and cloned into both pAS2 and pACT2 vectors, containing the GAL4 DNA binding domain and the GAL4 activation domain, respectively. Protein-protein interactions were detected by the activity of the *HIS3* reporter gene and the strength of the interaction assessed by the addition of incremental concentrations of 3AT. In each *panel*, the *three left columns* correspond to the MBNL1 C-terminal tail cloned into the pAS2 vector, whereas the *three right columns* correspond to the pACT2 clone. The presence of exon 5- and exon 7-encoded sequences in the C-terminal tails is schematically indicated by *gray boxes*, and the name of the full-length isoforms containing a given C-terminal tail is indicated in *brackets*. The identity of bait proteins is indicated in the scheme on the *bottom*, and the presence of the exon 7-encoded sequence corresponds to *squares CT5.7*, *CT7*, *36*, *37*, *38*, *41*, and *43* whereas the isoforms lacking the exon 7 correspond to *squares CT*, *CT5*, *35*, *37*, *40*, and *42*. The controls are indicated in the *white squares*.

mM 3AT concentration. At 2 mM 3AT, these two $MBNL1_{CT}$ and $MBNL1_{CT5}$ C-terminal regions only showed a low level of interaction with the C-terminal regions containing the exon 7 or both the exon 5 and 7 regions. In contrast, some of the interactions formed by the $MBNL1_{CT7}$ and $MBNL1_{CT5.7}$ C-terminal domains (CT7 or CT5.7) containing the exon 7 region and the full-length proteins containing one of these two C-terminal domains (36, 38, 41, and 43) were highly stable. They were detected at 3AT concentrations up to 5 or 10 mM (Fig. 7). Altogether, these data confirmed the fact that only the C-terminal domain of MBNL1 is involved in MBNL1 dimerization and revealed the previously unknown importance of the exon 7 region in this process.

## DISCUSSION

Alternative splicing produces several MBNL1 isoforms, differing mainly in the inclusion or exclusion of the in-frame sequences encoded by exons 3, 5, and 7. The exon 3 region has already been proposed to modulate RNA binding affinity, and here we provide new information as to how the presence or absence of the exon 3 region modifies the binding of MBNL1 to its target sites in pre-mRNA and to CUG expansions. We show for the first time the crucial role of this region in the splicing regulatory property of MBNL1. The exon 5 region has previously been proposed to play a critical role in the nuclear retention of the MBNL1 protein. Here, we show that this nuclear retention also depends on amino acid sequences encoded by

exon 6 and that the exon 6 region is required for splicing regulation. Finally, we have identified the role of the alternative exon 7 region in MBNL1 dimerization and showed also that exon 8 is alternative and not constitutive. We will now discuss the implications of our findings for the normal and pathological activities of the MBNL1 splicing factor.

*Nuclear Localization of MBNL1 and Splicing Activity*—MBNL1 isoforms displayed different nucleocytoplasmic localization patterns in HeLa cells and in human myoblasts. Isoforms lacking the exon 5 region were distributed in both the nucleus and the cytoplasm whereas the subcellular localization of those including the exon 5 region was restricted to the nucleus. Cassette exon 5 therefore probably encodes a region involved in the nuclear localization of MBNL1. This corroborates previously published data (9, 15, 24) and extends them to human myoblasts.

Because all MBNL1 isoforms were located in the nucleus, with or without the inclusion of the alternative exon 5 region, we can rule out the idea that this region is solely driving the nuclear localization of MBNL1. This observation was further confirmed using truncated proteins with or without the exon 5 region and lacking the C-terminal region. Indeed, diffuse staining for the truncated proteins $MBNL1_{\Delta CT}$ and $MBNL1_{\Delta CT5}$ was present in both the nucleus and the cytoplasm. The apparent molecular size of these truncated proteins fused to the GFP tag was around 60 kDa, excluding passive diffusion through the nuclear pore complex. Thus, the RNA binding domain composed of four CCCH zinc finger motifs was sufficient to target MBNL1 to the nucleus. Furthermore, this result also suggests that the MBNL1 exon 5 region co-operates with another region to maintain MBNL1 in the nucleus. In line with our hypothesis, we observed that both exons 5 and 6 were required to induce a nuclear localization of MBNL1 exclusively, as seen in Fig. 2*C*. Interestingly, a new type of highly conserved nuclear localization signal motif (the KRAEK motif) has recently been identified in *D. melanogaster* Muscleblind-like C and E isoforms (33). This motif has been shown to be sufficient to target Muscleblind-like to the nucleus of *D. melanogaster* Schneider 2 cells. Although this motif is conserved in the exon 6 region of human MBNL1, but not in exon 5, our data strongly suggest that additional regions might be required for MBNL1 nuclear retention. Thus, the nuclear retention of MBNL1, according to our results, necessitates a bipartite signal with one motif in the exon 5 and the second in the exon 6, but separately, these motifs are nonfunctional. We suggest that the MBNL1 exon 6 region contains a nuclear localization signal that is either functional in the presence of the exon 5 region or is regulated by an adjacent upstream sequence. Importantly, we cannot exclude the possibility that post-translational modification around the exon 5 sequence also regulates the subcellular localization of MBNL1.

Because splicing occurs in the nucleus, we asked whether there was a relationship between strict nuclear localization and the splicing activity of MBNL1. Surprisingly, with a similar level of expression, we observed no difference in the splicing activity of the different MBNL1 isoforms on hcTNT exon 5 or IR exon 11 splicing, as assayed using minigenes. This result raises the question of the interest of a restrictive nuclear localization. One possibility is that nuclear retention is a cellular mechanism that impedes MBNL1 cytoplasmic activity while maintaining its splicing activity when located in the nucleus. Few studies have reported the potential cytoplasmic functions of MBNL1; judging by its similarity to MBNL2, another member of the MBNL protein family, MBNL1 could be involved in mRNA transport and/or stability (34). In line with this observation, some recent studies have demonstrated a global change in gene expression in *MBNL1* knock-out mice (11, 35). Stability and translatability may also be regulated by MBNL1 under conditions of stress, in which MBNL1 co-localizes with stress granules (36). In this latter study, the authors used the isoform $MBNL1_{40}$, which displayed both nuclear and cytoplasmic localization at base line. However, this property cannot be extrapolated to other isoforms of MBNL1, especially to MBNL1 isoforms having the exon 5 and 6 sequences. Additional work is needed to determine whether MBNL1 isoforms located only in the nucleus also co-localize with stress granules under conditions of stress. Accordingly, the function of MBNL1 isoforms lacking exon 3 also remains to be established.

*MBNL1 Splicing Activity Depends on Both Exon 3- and Exon 6-Encoded Regions*—MBNL1 splicing factor can enhance or repress the inclusion of alternative exons (16). Based on our data, when an isoform is active in splicing, it can have a silencing or an enhancing activity on splicing, depending on the targeted splicing site. Our results exclude the possibility that some isoforms may be dedicated to splicing inhibition, whereas others would have splicing activation properties. As suggested previously, MBNL1 splicing activity might rather depend on the localization of its RNA-binding elements (17). Indeed, the MBNL1 exon 3 region is located between the zinc finger 2 and zinc finger 3 motifs and acts as a linker between the zinc finger 1/2 and zinc finger 3/4 domains that structure the MBNL1 RNA binding domain (22, 25) and probably modulate its RNA binding affinity. Triple-hybrid assays and gel shift experiments confirmed that the presence of the exon 3 region was essential for MBNL1 efficient binding to its pre-mRNA target. Moreover, we observed that all MBNL1 isoforms that included the exon 3 region ($MBNL1_{40}$, $MBNL1_{41}$, $MBNL1_{42}$, and $MBNL1_{43}$) strongly modified MBNL1 splicing activity with respect to hcTNT exon 5 and IR exon 11, whereas those that lacked exon 3 ($MBNL1_{35}$, $MBNL1_{36}$, $MBNL1_{37}$, and $MBNL1_{38}$) barely did. Also in accordance with the need for an RNA binding domain for MBNL1 activity, experiments performed with a truncated protein revealed that the splicing activity of the exon 3 region was weak but accurate (Fig. 4, $MBNL1_{\Delta CT3}$). The recovery of full splicing activity was observed with proteins that expressed both the exon 3 and exon 6 regions (Fig. 4, $MBNL1_{\Delta CT3}$). This result suggests that MBNL1 splicing activity mainly resides in the exon 6 region. Furthermore, it also suggests that MBNL1 splicing activity likely depends on its affinity for its RNA targets and the presence of both exon 3 and 6 regions.

*Relevance to DM1 Pathology*—Myotonic dystrophy type 1 belongs to the triplet repeat disorder family. The unstable CTG expansions are transcribed into RNA and retained in the nucleus. These toxic RNAs have been shown to sequester MBNL1 proteins (8, 37). There are various MBNL1 protein isoforms, which in our study differed in the presence or absence of the exon 3, 5, and 7 regions. All of them co-localized with the CUG repeats expressed in DM1 myoblasts. The MBNL1 RNA

binding domain alone ($MBNL1_{\Delta CT}$, $MBNL1_{\Delta CT3}$) was sufficient to mediate MBNL1 binding to the CUG repeats, as previously suggested (19), with or without the exon 3 region. Furthermore, using the yeast triple-hybrid system, we observed that the presence of the exon 3 region strongly enhanced the binding affinity of MBNL1 to the CUG repeats. Therefore, MBNL1 isoforms that lack the exon 3 region likely interact with CUG repeats very weakly. We thus suggest that their sequestration is not direct but mediated by the dimerization of MBNL1 with itself. The self-dimerization of MBNL1 has been previously reported and shown to occur via the C-terminal tail of MBNL1 (32). In our study, double-hybrid assays performed with the MBNL1 C-terminal tail revealed that these C-terminal fragments only dimerized with the full-length MBNL1. The presence of exon 7 also improved MBNL1 binding on both CUG repeats and hcTNT mRNA, most probably by promoting MBNL1 dimerization. Interestingly, the expression of MBNL1 protein isoforms, including the exon 7, region is greatly increased with the development of DM1 (9, 24). These isoforms could therefore contribute to the focal localization of all MBNL1 isoforms, with or without the exon 3 region. Altogether, our results suggest that the binding of MBNL1 proteins to the CUG repeats could either be direct or indirect, through the dimerization of MBNL1 with itself. Thus, the MBNL1 missplicing observed in DM1 could enhance its own sequestration in the *foci*.

## CONCLUSION

Altogether, our study reveals the importance of the alternative splicing of MBNL1 in the post-transcriptional regulation of the subcellular localization and function of MBNL1. We attempted to understand better the function of the various MBNL1 isoforms. Our results show that (i) the presence of the exon 3 region is essential to ensure a high binding affinity of MBNL1 for pre-mRNA, and thus its splicing activity; (ii) the presence of the exon 5 region ensures a restrictive nuclear localization; (iii) exons 3 and 5 are fully functional only in the presence of the exon 6 region; and (iv) the exon 7 region modulates the dimerization properties of MBNL1. Our data therefore indicate that association of MBNL1 with CUG repeats depends both on RNA-protein interactions and on self-dimerization. During fetal development and in DM1 pathophysiology, both MBNL1 exons 5 and 7 are included. This could contribute to the exclusively nuclear localization of MBNL1 and to its self-dimerization, two key features for the sequestration within *foci*. However, further studies will be needed for a better understanding of all the molecular consequences of MBNL1 sequestration in DM1. This knowledge is important to define precisely the mechanism for a potential therapeutic development.

*Acknowledgments—We thank Dr. Tom Cooper for the hcTNT minigene and Dr. S Reddy for the IR minigene. We acknowledge the Muscular Dystrophy Association and the Monoclonal Antibody Resource.*

## REFERENCES


1. Black, D. L. (2003) *Annu. Rev. Biochem.* **72,** 291–336
2. Ranum, L. P., and Cooper, T. A. (2006) *Annu. Rev. Neurosci.* **29,** 259–277
3. Lee, J. E., and Cooper, T. A. (2009) *Biochem. Soc. Trans.* **37,** 1281–1286
4. Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., *et al.* (1992) *Cell* **69,** 385
5. Fu, Y. H., Pizzuti, A., Fenwick, R. G., Jr., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., de Jong, P., *et al.* (1992) *Science* **255,** 1256–1258
6. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barceló, J., and O'Hoy, K. (1992) *Science* **255,** 1253–1255
7. Mooers, B. H., Logue, J. S., and Berglund, J. A. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102,** 16626–16631
8. Fardaei, M., Larkin, K., Brook, J. D., and Hamshere, M. G. (2001) *Nucleic Acids Res.* **29,** 2766–2771
9. Lin, X., Miller, J. W., Mankodi, A., Kanadia, R. N., Yuan, Y., Moxley, R. T., Swanson, M. S., and Thornton, C. A. (2006) *Hum. Mol. Genet.* **15,** 2087–2097
10. Kanadia, R. N., Johnstone, K. A., Mankodi, A., Lungu, C., Thornton, C. A., Esson, D., Timmers, A. M., Hauswirth, W. W., and Swanson, M. S. (2003) *Science* **302,** 1978–1980
11. Du, H., Cline, M. S., Osborne, R. J., Tuttle, D. L., Clark, T. A., Donohue, J. P., Hall, M. P., Shiue, L., Swanson, M. S., Thornton, C. A., and Ares, M., Jr. (2010) *Nat. Struct. Mol. Biol.* **17,** 187–193
12. Begemann, G., Paricio, N., Artero, R., Kiss, I., Pérez-Alonso, M., and Mlodzik, M. (1997) *Development* **124,** 4321–4331
13. Artero, R., Prokop, A., Paricio, N., Begemann, G., Pueyo, I., Mlodzik, M., Perez-Alonso, M., and Baylies, M. K. (1998) *Dev. Biol.* **195,** 131–143
14. Kalsotra, A., Xiao, X., Ward, A. J., Castle, J. C., Johnson, J. M., Burge, C. B., and Cooper, T. A. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 20333–20338
15. Terenzi, F., and Ladd, A. N. (2010) *RNA Biol.* **7,** 43–55
16. Ho, T. H., Charlet, B. N., Poulos, M. G., Singh, G., Swanson, M. S., and Cooper, T. A. (2004) *EMBO J.* **23,** 3103–3112
17. Goers, E. S., Purcell, J., Voelker, R. B., Gates, D. P., and Berglund, J. A. (2010) *Nucleic Acids Res.* **38,** 2467–2484
18. Warf, M. B., Diegel, J. V., von Hippel, P. H., and Berglund, J. A. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106,** 9203–9208
19. Warf, M. B., and Berglund, J. A. (2007) *RNA* **13,** 2238–2251
20. Sen, S., Talukdar, I., Liu, Y., Tam, J., Reddy, S., and Webster, N. J. (2010) *J. Biol. Chem.* **285,** 25426–25437
21. Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshere, M. G., Harper, P. S., and Brook, J. D. (2002) *Hum. Mol. Genet.* **11,** 805–814
22. Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N., and Ishiura, S. (2004) *Hum. Mol. Genet.* **13,** 495–507
23. Pascual, M., Vicente, M., Monferrer, L., and Artero, R. (2006) *Differentiation* **74,** 65–80
24. Dhaenens, C. M., Schraen-Maschke, S., Tran, H., Vingtdeux, V., Ghanem, D., Leroy, O., Delplanque, J., Vanbrussel, E., Delacourte, A., Vermersch, P., Maurage, C. A., Gruffat, H., Sergeant, A., Mahadevan, M. S., Ishiura, S., Buée, L., Cooper, T. A., Caillet-Boudin, M. L., Charlet-Berguerand, N., Sablonnière, B., and Sergeant, N. (2008) *Exp. Neurol.* **210,** 467–478
25. Teplova, M., and Patel, D. J. (2008) *Nat. Struct. Mol. Biol.* **15,** 1343–1351
26. Cooper, T. A. (1998) *Mol. Cell. Biol.* **18,** 4519–4525
27. Kosaki, A., Nelson, J., and Webster, N. J. (1998) *J. Biol. Chem.* **273,** 10331–10337
28. Edom, F., Mouly, V., Barbet, J. P., Fiszman, M. Y., and Butler-Browne, G. S. (1994) *Dev. Biol.* **164,** 219–229
29. Taneja, K. L. (1998) *BioTechniques* **24,** 472–476
30. Holt, I., Mittal, S., Furling, D., Butler-Browne, G. S., Brook, J. D., and Morris, G. E. (2007) *Genes Cells* **12,** 1035–1048
31. Holt, I., Jacquemin, V., Fardaei, M., Sewry, C. A., Butler-Browne, G. S., Furling, D., Brook, J. D., and Morris, G. E. (2009) *Am. J. Pathol.* **174,** 216–227
32. Yuan, Y., Compton, S. A., Sobczak, K., Stenberg, M. G., Thornton, C. A., Griffith, J. D., and Swanson, M. S. (2007) *Nucleic Acids Res.* **35,** 5474–5486
33. Fernandez-Costa, J. M., and Artero, R. (2010) *Mol. Cells* **30,** 65–70
34. Adereth, Y., Dammai, V., Kose, N., Li, R., and Hsu, T. (2005) *Nat. Cell Biol.* **7,** 1240–1247
35. Osborne, R. J., Lin, X., Welle, S., Sobczak, K., O'Rourke, J. R., Swanson, M. S., and Thornton, C. A. (2009) *Hum. Mol. Genet.* **18,** 1471–1481
36. Onishi, H., Kino, Y., Morita, T., Futai, E., Sasagawa, N., and Ishiura, S. (2008) *J. Neurosci. Res.* **86,** 1994–2002
37. Miller, J. W., Urbinati, C. R., Teng-Umnuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., and Swanson, M. S. (2000) *EMBO J.* **19,** 4439–4448