Human Molecular Genetics, 2004, Vol. 13, No. 5 ***495–507***
DOI: 10.1093/hmg/ddh056
Advance Access published on January 13, 2004

Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats

Yoshihiro Kino, Daisuke Mori, Yoko Oma, Yuya Takeshita, Noboru Sasagawa and Shoichi Ishiura*

Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

Received October 13, 2003; Revised and Accepted January 5, 2004

Myotonic dystrophy (DM) type 1 is caused by an expansion of a CTG repeat in the *DMPK* gene and type 2 by a CCTG repeat in the *ZNF9* gene. Previous reports have suggested that transcripts containing expanded CUG/CCUG repeats might have toxic gain-of-function effects, probably affecting the function of RNA-binding proteins in the pathogenesis of DM. Here, it was attempted to compare the RNA-binding properties of three proteins, CUG-BP, MBNL1/EXP and PKR, which have previously been suggested to interact with CUG repeats. MBNL1, but not CUG-BP or PKR, interacted with both CUG and CCUG repeats in a yeast three-hybrid system. By using various synthetic RNAs, it was found that MBNL1 specifically interacts with repetitive sequences summarized as CHHG and CHG repeats, where H is A, U or C. Interestingly, MBNL1 did not interact with a genuine double-stranded RNA comprising CAG/CUG repeats, suggesting that MBNL1 prefers bulge-containing double-stranded RNAs. Deletion analysis indicates a difference in RNA-binding abilities among splice variants of MBNL1. It was also found that MBNL1 can bind to repetitive motifs in *ZNF9*, which contain a minimal length of CCUG repeats with non-CCUG insertions.

INTRODUCTION

Myotonic dystrophy (DM) is an autosomal inherited disorder with multi-systemic and variable symptoms including muscle weakness, myotonia (delayed relaxation of muscle), cataracts, mental retardation and insulin resistance (1). DM can be categorized into two sub-types, DM1 and DM2, according to the kind of mutation. DM1 is caused by an expansion of a CTG repeat located in the 3′ untranslated region (3′-UTR) of the *DM protein kinase* (*DMPK*) gene (19q13.3) (2–4), while DM2 is caused by an expansion of a CCTG repeat located in the first intron of the *ZNF9* gene (3q21) (5). The clinical manifestations of DM1 and DM2 are similar, although there are some differences, such as the presence of a congenital form specific to DM1 (6). Since the *DMPK* and *ZNF9* genes seem to have nothing in common except for the expansion of repetitive sequences, the expansion itself rather than defects in these genes appears to be essential for the pathogenesis of DM. Indeed, no DMPK mutation other than the repeat expansion has ever been reported, suggesting that DM1 is not caused by a loss-of-function of *DMPK* itself. In addition, mice deficient in DMPK show mild myopathy and abnormalities in cardiac conductance, but do not recapitulate other symptoms of DM1 (7–9). On the other hand, transgenic mice with a CUG_{250} repeat expressed in the 3′-UTR of the muscle-specific actin gene show myotonia and progressive myopathy (10), suggesting that the expression of an expanded CUG repeat is sufficient for causing these symptoms independently from the genetic context around DMPK. Since the severity of phenotypes in model mice is related to the expression level of expanded CUG repeats (10), the expanded RNA itself might have toxic effects on cells. Notably, by using fluorescent *in situ* hybridization (FISH) techniques, several investigators have shown that CUG repeat RNAs form nuclear foci in cells or tissues of DM1 patients and DM1 model cells and mice expressing expanded CUG repeats (10–14). Furthermore, expanded CCUG repeat RNAs also form nuclear foci in cells of DM2 (5,14), suggesting that these expanded transcripts have some role in the pathogenesis. One reasonable explanation is that the expanded CUG/CCUG repeat RNA might affect the function of RNA-binding proteins in a *trans*-dominant manner, presumably by their sequestration or aberrant activation, leading to cytotoxic effects. To date, several proteins have been suggested to interact with CUG repeats in *in vitro* or *in vivo* experiments.

One such protein is CUG–binding protein 1 (CUG–BP), first identified as a binding protein for a CUG_8 probe in gel

*To whom correspondence should be addressed. Tel/Fax: +81 354546739; Email: cishiura@mail.ecc.u-tokyo.ac.jp

retardation analysis (15,16). CUG–BP is the first discovered member of the CELF (CUG–BP and ETR3-like factors) proteins and acts as a regulator of alternative splicing (17–20), translation (21,22) and deadenylation (23). Remarkably, several reports have shown the involvement of this protein in the symptoms of DM1, particularly myotonia and insulin resistance, by aberrantly regulating the alternative splicing of chloride channel 1 (CLC-1) and insulin receptor (IR), respectively (19,20). Cardiac troponin T (cTNT) is also a target of CUG–BP, and its splicing pattern is altered in DM1 patients (18). It has also been suggested that a loss of translational induction by CUG–BP in DM1 cells results in the reduction of p21, a key factor in the cell cycle and differentiation, and leads to the defect in myogenic progression (24). Therefore, it is almost certain that CUG–BP is a very important factor in the pathogenesis of DM.

As indicated by its name, the sequestration of CUG–BP by expanded CUG repeats was first assumed. However, whether this protein is actually sequestered by the expanded repeats is controversial, because it does not co-localize with nuclear RNA foci in DM1 or DM2 cells (25,26). The interaction between CUG–BP and CUG repeat RNA is not length-dependent (27). Moreover, contrary to its name, CUG–BP specifically binds to UG motifs rather than to CUG repeats (28), and its binding to CUG repeats seems to be weak (28,29). Nevertheless, a recent paper has shown that CUG–BP binds to an expanded CUG repeat *in vitro* (30), making the situation more complicated. Thus it remains elusive whether CUG–BP actually interacts with CUG/CCUG repeats.

Another group of candidates is the muscleblind protein family, consisting of MBNL1/EXP, MBNL2/MBLL and MBNL3/MBXL/MBLX/CHCR. All these proteins have been shown to co-localize with RNA foci in the cells of both DM1 and DM2 patients, suggesting that they interact with both CUG and CCUG repeats (14,25,26). MBNL1 interacts with CUG repeats in a length-dependent manner (31). In contrast to CUG–BP, however, it is still unclear whether MBNL proteins are involved in the symptoms of DM. Although, *muscleblind*, the fly ortholog of MBNLs, has been reported to play some role in the differentiation of the eye and muscle (32,33), the molecular function and characteristics of MBNL1 are currently unknown.

The third candidate is PKR, a protein kinase activated by double-stranded RNA. Since CUG repeats can form stable double-stranded hairpin structures (34), it is conceivable that CUG repeats are a potential target of PKR. Previous *in vitro* experiments have shown that PKR binds to CUG repeat RNAs in a length-dependent manner and the activation of PKR by CUG repeats has been observed to also be length-dependent (34). Since PKR plays multiple roles in various cellular events such as the anti-viral response, signal transduction, cell growth and apoptosis (35), the abnormal activation of PKR might cause impairments in several cellular functions. However, the involvement of PKR in the pathogenesis of DM is not clear as in the case of MBNL proteins.

To understand the mechanism of DM, the characterization of proteins interacting with CUG/CCUG repeats is important, because this may be related to events downstream of the repeat expansion and its expression. In this study, we tried to reveal the RNA-binding properties of three proteins, CUG–BP, MBNL1 and PKR, to compare their interactions with CUG/CCUG repeats. These proteins have not previously been analyzed together in the same experimental system. We found that MBNL1, but not CUG–BP or PKR, interacts with both CUG and CCUG repeats in an *in vivo* system. Next, we characterized the binding specificity of MBNL1 since its binding target might also be involved in the pathogenesis of DM. For this purpose, we generated an array of repetitive sequences to profile the RNA-binding properties. Our system clearly shows differences in the RNA-binding properties of the above proteins. We determined the target sequence of MBNL1, suggesting that this protein has a novel RNA-binding property.

RESULTS

Comparative analysis of RNA-binding proteins

The RNA-binding proteins assayed in this study are shown in Figure 1A. The isoform of CUG–BP used here does not have an LYLQ insertion (28). $MBNL1_{40}$ is a 40 kDa isoform of MBNL1. A p20 PKR fragment corresponds to its double-stranded RNA binding domain (36,37), and was previously suggested to interact with CUG repeats in a gel retardation assay (34). We used a yeast three-hybrid system, in which the interaction between RNA and protein can be detected by the activity of reporter genes (38,39) (Fig. 1B). Two kinds of assay were performed: an *HIS3* assay and a β-galactosidase (β-gal) assay. In the former assay, yeast transformants are picked up onto a plate lacking histidine, on which *HIS3* activity is necessary for histidine synthesis. The *HIS3* gene product is competitively inhibited by 3-amino-1,2,4-triazole (3-AT) in a dose-dependent manner. Therefore, the expression of *HIS3*, which is activated by the interaction between RNA and protein, can be detected by the viability of yeast transformants on plates containing 3-AT but no histidine. By varying the concentration of 3-AT, the activation can be classified according to the viability of yeast. The β-gal assay is more quantitative in simply measuring the β-gal activity in cell lysates of yeast transformants.

The results of the *HIS3* assay are shown in Table 1. Neither CUG–BP nor PKR interacted strongly with CUG/CCUG repeats, although these proteins are not inactive in yeast (see below). Figure 2A also shows that the activation of β-gal was almost negligible with CUG and CCUG repeats compared with a UG repeat when CUG–BP was used as a prey. As shown in Figure 2C, the activation induced by PKR and CUG/CCUG repeats was also undetectable. By contrast, MBNL1 showed apparent interactions with both CUG and CCUG repeats. This was also confirmed by the β-gal assay as shown in Figure 2B. We could not reproduce the length-dependent interaction between CUG repeats and MBNL1, which preferably interacts with CUG_{21} but not so much with longer repeats. The reasons why MBNL1 preferred CUG_{21} might be attributable to the limitation of the yeast three-hybrid system rather than indicating an actual preference of MBNL1 for CUG_{21}. Since hybrid RNAs are transcribed by RNA polymerase III, which produces relatively small RNAs in general, longer repetitive sequences might not be efficiently transcribed (39). Another possibility is that the GAL4 activation domain fused to RNA-binding proteins that interact with long RNAs might be located at a distance from the promoter region, thus leading to inefficient activation. Taking the principal of this system into

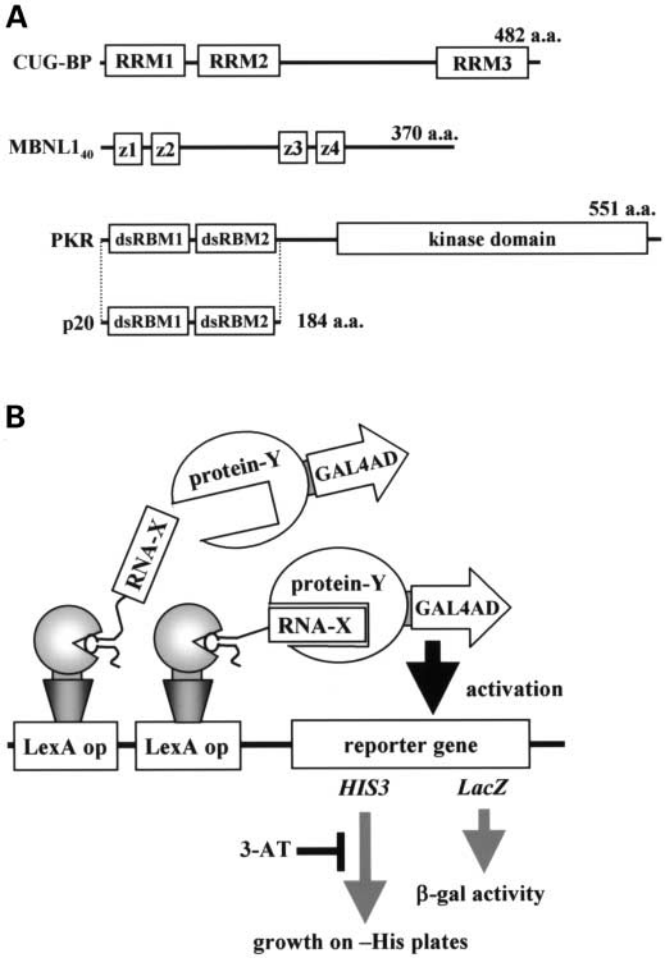

Figure 1. Yeast three-hybrid analysis. (**A**) Structures of the RNA-binding proteins examined in this work. Motifs involved in RNA-binding are indicated by boxes. RRM, RNA recognition motif. z, CCCH-type zinc finger motif. dsRBM, double-stranded RNA binding motif. (**B**) Scheme of the yeast three-hybrid system. Reporter genes (*HIS3* and *LacZ*) downstream of the LexA operator (LexA op) are integrated in the genome of yeast strain *L40-coat.* A fusion protein of LexA and MS2 coat protein (dotted shape) binds to the operator. RNA-X is expressed as a hybrid RNA with MS2 recognition sites, through which the hybrid RNA binds to LexA-MS2 coat fusion protein, and connected to a LexA operator. Protein-Y is expressed as a fusion with GAL4 activation domain (GAL4AD). When RNA-X binds to protein-Y, GAL4AD activates the expression of the reporter genes. Since the gene product of *HIS3*, which is required for cell growth on plates lacking histidine, is competitively inhibited by 3-AT. The concentration of 3-AT at which yeast transformants can grow represents the activity of *HIS3.*

account, it seems natural that there is a threshold length of each RNA for efficient activation, rather than no limitation of RNA length. Nevertheless, the interaction between MBNL1 and CUG/CCUG repeats was obvious and consistent with the results of previous FISH analyses.

Characterization of RNA-binding properties using an array of repetitive sequences

To characterize the binding specificities of the RNA-binding proteins described above, we generated a variety of repetitive sequences, including tri-, tetra- and penta-nucleotide repeats, and examined them in the *HIS3* assay. The results are shown in Table 2. As previously reported (28), CUG–BP interacted strongly with UG-containing sequences such as UG_{12}, UG_{24}, $UCUG_{32}$ and $UAUG_7$. It also interacted with a UCCG repeat, suggesting that CUG–BP prefers GU motifs as well as UG motifs. Although both $CCUG_{22}$ and $CUGG_{50}$ repeats contain UG motifs, CUG–BP did not interact with them. This might be due to their secondary or higher structures.

Table 1. Results of *HIS3* assay of CUG and CCUG repeats

	CUG–BP	MBNL1	PKR (p20)
MS2-2	–	–	–
UG 24	+++++	–	–
CA 24	–	–	–
CUG 7	–	–	–
CUG 16	–	++	–
CUG 21	–	+++	–
CUG 37	–	++	–
CUG 41	+	++	–
CUG 47	–	++	–
CUG 70	–	++	+
CCUG 7	+	–	–
CCUG 22	–	++++	–
CCUG 50	–	++++	–
CAGG 22	–	–	–
CAGG 50	+	–	–

PKR strongly interacted with $UAUG_7$+$CAUA_7$, CUG_{16}+CAG_{16} and $CCUG_{22}$+$CAGG_{22}$ repeats, which might form double-stranded hairpins. Although we have no direct evidence for the structures of these RNAs in cells, it seems natural that they are dsRNA in cells. Indeed, CUG–BP did not interact with $UAUG_7$+$CAUA_7$, although it interacted preferentially with the UAUG repeat alone, suggesting that the former sequence might have a different structure from the latter. Interestingly, PKR interacted with CAG repeats to some extent, consistent with a previous report in which PKR bound to transcripts of huntingtin with expanded CAG repeats (40). This might be because CAG repeat can form double-stranded structures containing bulges (mispairs). Surprisingly, PKR did not interact with CUG repeat, although this RNA can be expected to have a structure similar to CAG repeats.

MBNL1 interacted strongly with $CCCG_{21}$, modestly with $CUUG_{38}$ and $CAAG_{38}$, weakly with $CCAG_{50}$, CAG repeats, CCG_{30}, and $CAUG_7$, but not at all with $CGGG_{20}$, $CUGG_{50}$, and CAGG repeats. The deduced consensus MBNL1 binding sequence is CHHG or CHG, where H corresponds to A, C or U (not G). Importantly, both of these sequences form double-stranded structures with bulges. Interestingly, MBNL1 did not interact with CUG_{16}+CAG_{16} or $CCUG_{22}$+$CAGG_{22}$, with which PKR interacted strongly. This indicates that MBNL1 seems to prefer double-stranded RNAs containing bulges. Unlike CUG–BP, which can bind to both UG and GU motifs, MBNL1 did not interact with the UCCG repeat, the reverse (but not complementary) sequence of the CCUG repeat. UCCG appears to form a double-stranded structure, but not suitable for the consensus of CHHG, suggesting that there is some kind of sequence selectivity in the recognition by MBNL1. Alternatively, UCCG does not have a hairpin structure in the cell, as CUG–BP interacted well with this sequence. We have examined several pentanucleotide repeats, some of which

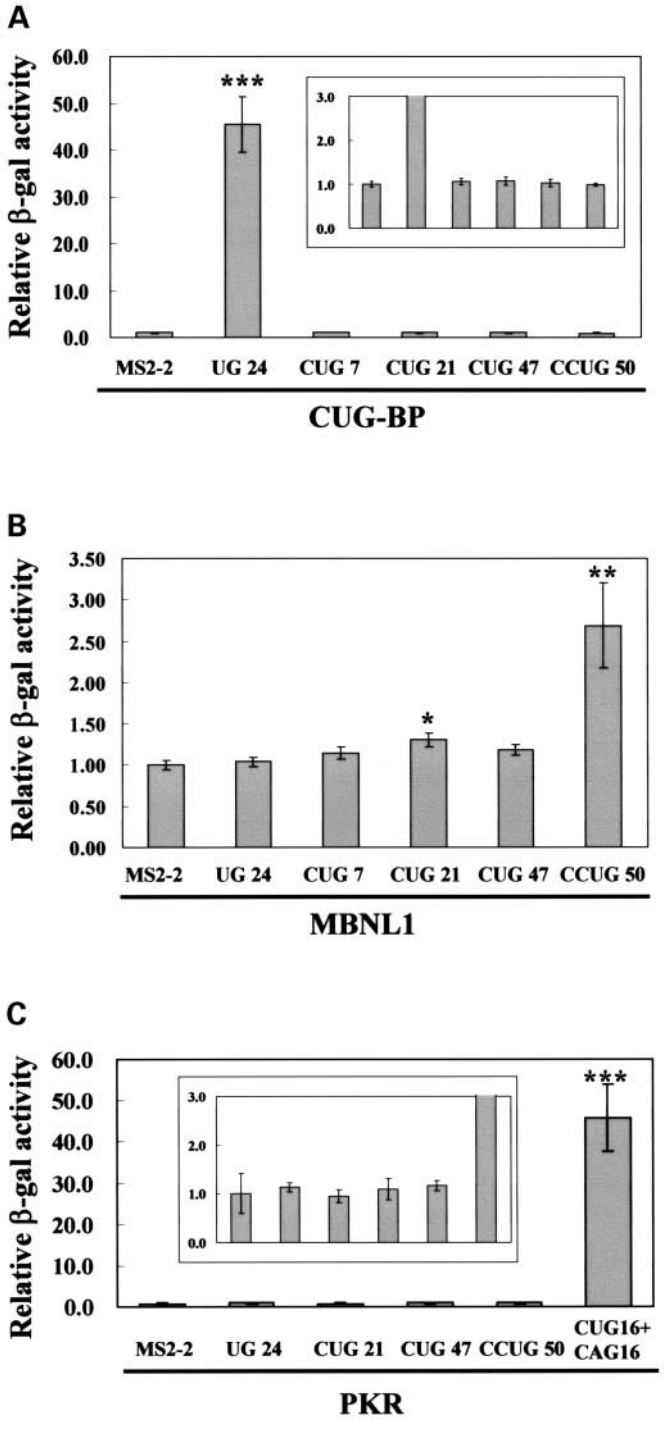

Figure 2. Results of β-galactosidase assay. Bars represent the average value of relative β-galactosidase activities of four independent transformants. The RNA sequences used are shown below the bars. Error bars correspond to standard deviations. MS2-2 is a negative control with no repetitive insertion. Insets shows the same bars with magnified scales. (**A**) Results of the interaction of CUG–BP with repetitive RNAs. The β-galactosidase activity is very strong with UG repeats, while it is almost not detectable with CUG and CCUG repeats. (**B**) Results with MBNL1. A relatively strong activation was observed with $CCUG_{50}$ and a small amount activation was observed with CUG_{21}. (**C**) Results with PKR. An intense activation was observed with CUG_{16}+CAG_{16}, while no detectable activation was observed with other RNAs. Significant difference from MS2-2: $***P < 0.001$, $**P < 0.01$, $*P < 0.05$, in one-way ANOVA, followed by Tukey type multiple comparison.

Table 2. RNA-binding specificities of CUG–BP, MBNL1 and PKR in *HIS3* assay

	CUG–BP	MBNL1	PKR (p20)
UA 30	−	−	+++
CAG 16	−	++	++
CAG 41	+	++	++
CAG 70	−	++	++
CCG 30	++	++	−
CGG 30	+	−	−
AAG 30	−	−	−
CCU 37	−	−	−
UCUG 32	++++	−	+
CAGA 32	++	−	+
UCCG 27	++++	−	+
CGGA 30	++	−	+
CUGG 50	−	−	−
CCAG 50	−	++	−
CAAG 38	−	+++	−
CUUG 38	+	+++	−
CGGG 20	−	−	−
CCCG 21	++	++++	−
UAUG 7	+++	−	−
CAUA 7	−	−	−
CAUG 7	++	++	++
CUGAG 19	−	−	++++
CUCAG 19	−	−	++
CCUAG 13	−	+	−
CUAGG 13	−	−	+
CAUUG 16	−	−	+
CAAUG 16	−	−	+++
UAUG7+CAUA7	−	−	+++++
CUG16+CAG16	−	−	+++++
CCUG22+CAGG22	−	−	+++++

presumably form bulge-containing double-stranded structures. Although PKR interacted with some of them, for example, $CUGAG_{19}$ and $CAAUG_{16}$, MBNL1 showed almost no interaction with any of these sequences. Together with the observation that MBNL1 seems to prefer tetranucleotide repeats over trinucleotide repeats, these results suggest that the number of nucleotides in each repetitive unit and/or the ratio of bulges in the duplex might also be important for the binding of MBNL1.

At last, two points should be noted about three-hybrid assays. First, the strength of maximum reporter activity is varied among proteins, since fusing them with GAL4 activation domain might affect their RNA-binding activity and/or transcriptional activity. This effect may depend on the structure of each proteins. Therefore, it is not strange that the maximum activities of each protein (i.e. CUG–BP–UG_{24}, PKR–CUG_{16}+CAG_{16} and MBNL1–$CCUG_{50}$) differ largely. However, it is also possible that some unknown sequences show higher reporter activities than CCUG repeats in combination with MBNL1. Next, these large differences in β-gal assay do not appear to be reproduced in the *HIS3* assays shown in Tables 1 and 2, in which CUG–BP–UG_{24} and PKR–CUG_{16} + CAG_{16} were classified as (+++++), while MBNL1 was (++++). We examined the former two combinations on the selection plates containing much higher concentration of 3-AT than 1 mM, resulting in the growth of these transformants in the presence of at least 10 mM 3-AT (data not shown). Therefore, the tendencies of reporter activities in these two assays do not seem to be considerably different.

Deletion analysis of MBNL1

MBNL1 has at least nine splice variants, which we tentatively name here as shown in Figure 3B. Seven of them, $MBNL1_{42}$, $MBNL1_{41}$, $MBNL1_{40}$, $MBNL1_{41s}$, $MBNL1_{40s}$, $MBNL1_{36}$ and $MBNL1_{35}$, have been reported previously (25,31), while $MBNL1_n$ and MBNL1 delta N ($MBNL1_{\Delta N}$) are novel isoforms. The existence of the latter two is suggested by ESTs, AL562860 (and others) and BC050535, respectively. $MBNL1_n$ is an isoform containing only the N-terminus short region of MBNL1 with a single zinc finger motif. This isoform has a distinct C-terminus and 3′-UTR from other isoforms. On the other hand, $MBNL1_{\Delta N}$ lacks the N-terminus region and starts in the second zinc finger motif, resulting in the loss of this motif. Thus, this isoform has only two intact zinc finger motifs out of four. Since MBNL1 has a variety of C-terminus regions, $MBNL1_{\Delta N}$ also has more splice variation in the C-terminus other than as shown in Figure 3B.

To determine the contribution of each zinc finger motif to RNA-binding, deletion analysis of MBNL1 was performed using a yeast three-hybrid system. We generated 10 deletion constructs (del 1–10) as shown in Figure 3C, and their affinities for $CCUG_{22}$ and CUG_{21} are indicated. Del 1 has a deletion in the C-terminus and corresponds to the maximum common region among $MBNL1_{42}$, $MBNL1_{41}$, $MBNL1_{40}$, $MBNL1_{41s}$ and $MBNL1_{40s}$. Del 2 lacks the alanine-rich region in the C-terminus of del 1. Del 3–6 are deletion constructs with various zinc finger motifs. Del 7–9 lack a region (amino acids 116–183) that is also missing from $MBNL1_{36}$ and $MBNL1_{35}$. Del 10 mimics the deletion of the N-terminus, $MBNL1_{\Delta N}$. As can be seen in Figure 3C, deletions of the C-terminus and flanking alanine-rich region cause no reduction in the RNA binding activity in *HIS3* assay (del 1 and del 2, respectively). Rather, these two mutants appear to have increased binding activity, probably because of the structural facilitation of RNA binding or the activation of the reporter gene due to the lack of the C-terminus region. The deletions of each zinc finger motif suggest that all four zinc fingers are necessary for full RNA-binding ability, although the deletion of the fourth zinc finger (del 6) still allowed some binding to $CCUG_{22}$. Interestingly, del 7, which has all four zinc fingers, lost most of its binding abilities, suggesting that the linker region between the second and third zinc fingers is also needed for binding to the CCUG/CUG repeats, or that this long linker is necessary for the structural integrity of MBNL1. As del 1 and del 2 correspond to the common region among $MBNL1_{42}$, $MBNL1_{41}$, $MBNL1_{41s}$, $MBNL1_{40}$ and $MBNL1_{40s}$, these splice isoforms might interact with CCUG and CUG repeats. On the other hand, other variants, $MBNL1_{36}$, $MBNL1_{35}$, $MBNL1_{\Delta N}$ and $MBNL1_n$, might not interact with these repeats, since they lack either zinc fingers or the linker region necessary for the binding. If this is the case, there might be a difference among MBNL1 isoforms in their interactions with expanded repeats.

In vitro binding experiments of GST-MBNL1

To confirm the results of the three-hybrid analyses, we performed gel retardation analysis. First, we fused $MBNL1_{40}$ with glutathione *S*-transferase (GST) in the N-terminus and a Histidine-tag (His-tag) in the C-terminus (GST–MBNL1) (Fig. 4A). GST–MBNL1 was expressed in *E. coli* and purified by affinity chromatography. Gel retardation analysis suggested that GST–MBNL1 bound to a ^{32}P-labeled $CCUG_{15}$ probe (Fig. 4B, lane 3), and supershift was observed when an anti-GST antibody was added (Fig. 4B, lane 5). The extent of band shift was reduced by adding non-labeled $CCUG_{15}$ RNA (Fig. 4B, lane 4), suggesting the specificity of the binding. Figure 4C shows the effects of ion concentration. The formation of the RNA–protein complex was disrupted at higher concentrations of KCl and NaCl (Fig. 4C, lanes 9–11 and 15–17). The complex of GST–MBNL1 and $CCUG_{15}$ was subjected to RNase V1 treatment. RNase V1 specifically cleaves base-paired nucleotides. Figure 4D shows the partial protection of $CCUG_{15}$ from cleavage in the presence of GST–MBNL1 (lanes 20 and 23), further confirming the binding.

Next, we confirmed that GST–MBNL1 can bind to CUUG, CCCG and CAAG repeats, but not with CAGG and CGGG repeats (Fig. 5, left), consistent with the results of the yeast system. GST–MBNL1 also interacted with CUG repeats but not with CAG repeats, a finding not consistent with the yeast three-hybrid system. Similar results were reported in the previous study using the UV crosslink method (31). Since this discrepancy may reflect the difference in experimental systems, it is unclear whether MBNL1 actually binds to CAG repeats. Mostly consistent with the results of the yeast system, GST–MBNL1 interacted with CCUG repeats more specifically than with the other CHHG repeats, since the binding to CCUG could be detected with a much smaller amount of MBNL1 (Fig. 5, top right). We also examined the *in vitro* binding of GST–CUG–BP, but the binding to CUG and CCUG repeats was nearly undetectable at the level of GST–CUG–BP used in Figure 5 (right bottom).

We also examined the dependence of the binding between CCUG repeats and MBNL1 on the repeat length. As shown in Figure 6A, free probes of $CCUG_{27}$ and $CCUG_{35}$ disappeared at the highest dose of MBNL1, while that of $CCUG_{15}$ remained unbound at the same dose of GST–MBNL1 (lanes 4, 8 and 12). This might reflect the length-dependent formation of a particular structure that is recognized by GST–MBNL1. In this assay, the number of shifted bands represents the variety of RNA–protein complexes, mainly reflecting the number of proteins binding to a single probe. There were two bands of RNA–protein complexes with $CCUG_{15}$ (Fig. 6A, lane 4), while at least four kinds of band could be seen with $CCUG_{27}$ and $CCUG_{35}$ (Fig. 6A, lanes 6–8 and 10–12), indicating that a larger numbers of GST–MBNL1 proteins bind to CCUG probes as the RNA grows longer. These results show the length-dependent binding of MBNL1, probably accompanied by the transition of the RNA structure, and an increase in the number of binding sites for GST–MBNL1. Longer repeats were also examined by competition analysis (Fig. 6B). Consistent with Figure 6A, $CCUG_{35}$ showed more preferential binding to MBNL1 than $CCUG_{15}$, confirmed by its efficient competition (lanes 15 and 18, or 16 and 19). Moreover, $CCUG_{120}$ fully competed to the labeled probe at the lowest dose, while $CCUG_{35}$ left a bandshift at the same dose (lane 17, see also lane 20), suggesting that $CCUG_{120}$ has higher affinity to MBNL1 than $CCUG_{35}$. Consistent with the results of Figure 5, competition by CUG repeats was less evident than that by CCUG repeats (Fig. 6B, bottom). However, at the highest dose,

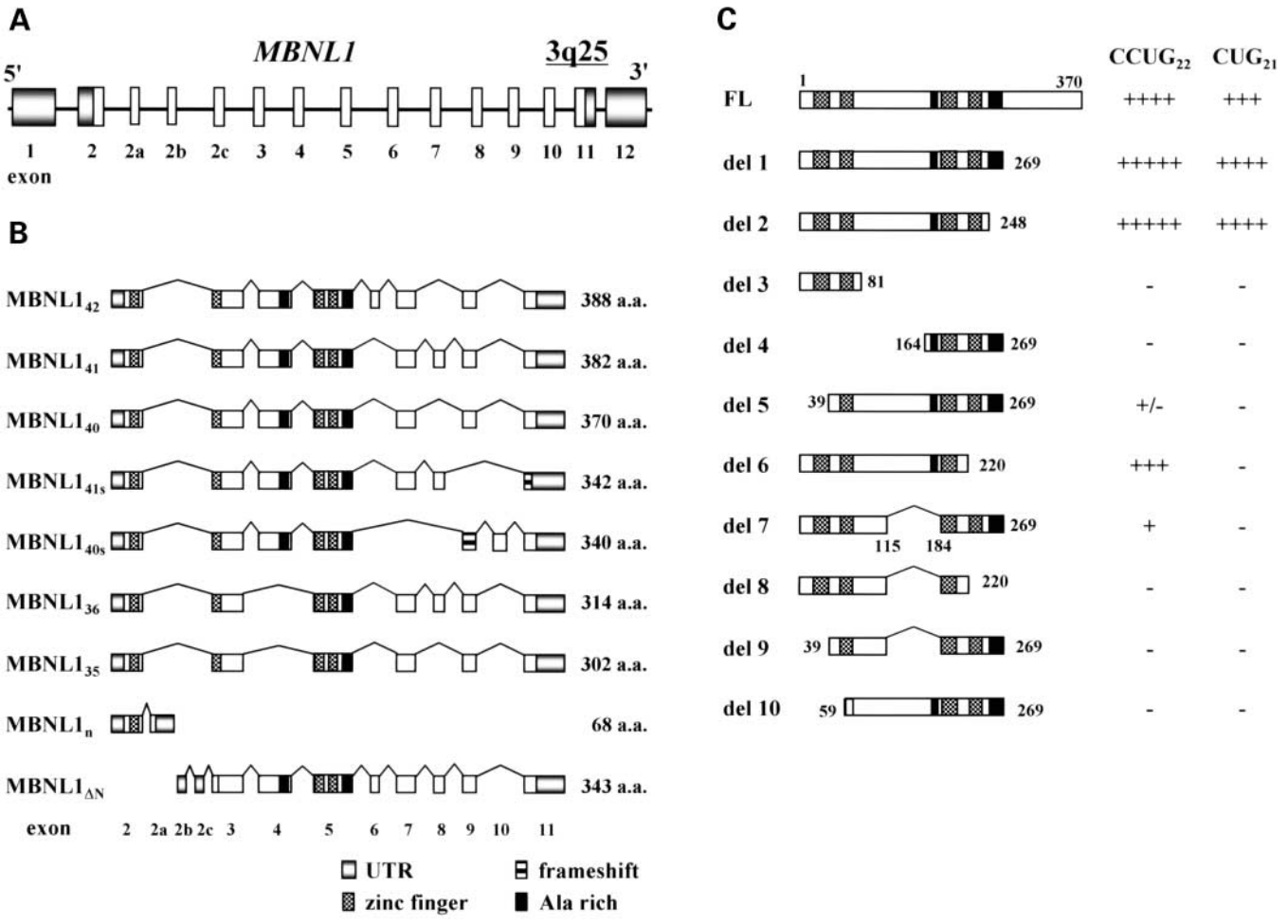

Figure 3. Splice variants and deletion analysis of MBNL1. Exons of *MBNL1* are represented as boxes. Note that the length of each box does not reflect the actual genomic structure of *MBNL1*. (**A**) Order and name of exons used in this work. (**B**) Open reading frames of each variant are indicated by open boxes. Other boxes corresponding to UTR, zinc finger motifs and alanine-rich regions are indicated in the bottom. MBNL1$_n$ and MBNL1$_{\Delta N}$ are novel isoforms. The zinc finger motif in exon3 of MBNL1$_{\Delta N}$ is disrupted since it requires a cysteine in exon 2. Striped boxes represent frameshift of some isoforms compared to the corresponding exons of other isoforms. (**C**) Full-length MBNL1$_{40}$ (FL) and its deletion mutants (del 1–10) were examined in the yeast three-hybrid system (*HIS3* assay) with CCUG$_{22}$ and CUG$_{21}$ hybrid RNAs. Numbers indicate the amino acid composition of each variant. Amino acids 116–183, deleted in del 7–del 9, correspond to the region deleted in MBNL1$_{35}$ and MBNL1$_{36}$. Results of HIS3 assay are indicated at the right with the same code as in Tables 1 and 2. Zinc finger motifs and alanine-rich regions are indicated as the same boxes in (B).

only CUG$_{130}$ showed apparent competition (lane 32), suggesting that MBNL1 prefers longer CUG repeats, which is in agreement with the previous finding using UV-crosslink method (31). Importantly, since we used the same doses (nanograms) of competitor RNAs, longer CUG/CCUG repeats should have fewer molecules than shorter repeats. Therefore, the preferential binding of MBNL1 to longer CUG/CCUG suggests that these RNAs sequester larger number of MBNL1 molecules per a single RNA than shorter ones.

The above results and those of previous studies apparently show MBNL1 to be a repeat RNA-binding protein. To examine whether MBNL1 binds to DNA, competition analysis was performed. Several unlabeled single-stranded oligo DNAs were used as cold competitors. Figure 7A shows that CCTG$_{22}$ competed with the binding of GST–MBNL1 to CUG$_{41}$ (lane 4), suggesting that MBNL1 can interact with DNA as well as RNA. Since other DNAs did not compete as efficiently as CCTG, MBNL1 might have a similar sequence specificity for DNA to that for RNA. To determine whether MBNL1 prefers RNA or DNA, the competitive efficiencies of cold CCUG$_{15}$ and CCTG$_{22}$ were compared. As shown in Figure 7B, CCUG$_{15}$ competed more efficiently than the CCTG$_{22}$ repeat, suggesting that MBNL1 favors repetitive RNA sequences for binding. Although the lengths of the CCUG and CCTG repeats used here were not the same, we can expect that CCUG$_{22}$ would be a more effective competitor than CCUG$_{15}$, because MBNL1 prefers longer CCUG repeats as suggested above. In this case, the difference in binding specificity between CCUG$_{22}$ and CCTG$_{22}$ might be more prominent than in the case of CCUG$_{15}$ and CCUG$_{22}$. The preference of MBNL1 for RNA rather than DNA may suggest its major role in RNA metabolism.

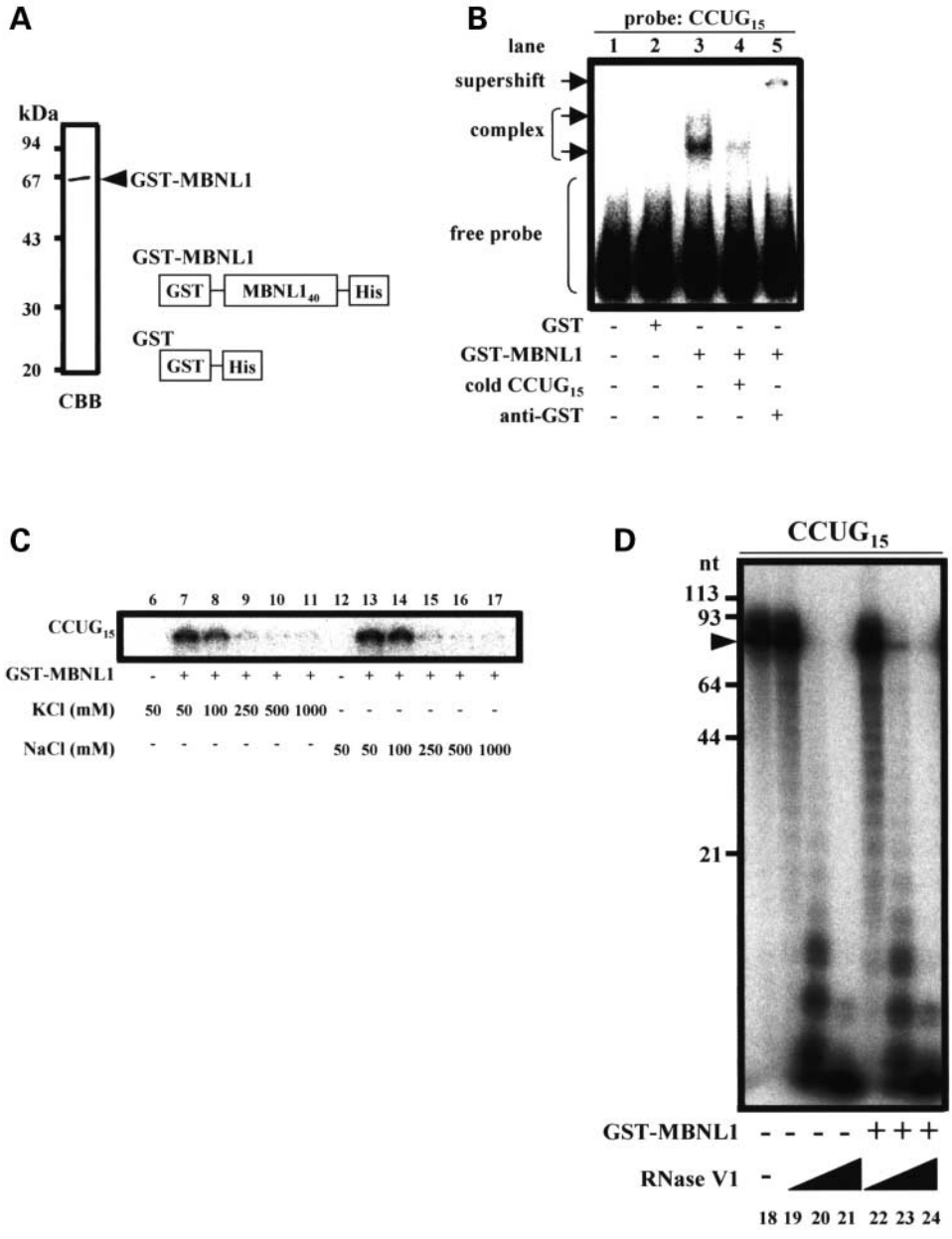

Figure 4. *In vitro* binding of recombinant GST–MBNL1 to $CCUG_{15}$. (**A**) Right, the structures of the GST–MBNL1 and GST recombinant proteins. Both proteins were also tagged with 6× histidine in the C-terminus to facilitate purification. Left, purified GST–MBNL1 stained with CBB. (**B**) Gel retardation analysis of GST–MBNL1 and ^{32}P-labeled $CCUG_{15}$ probe by non-denaturing 5% PAGE. The presence of GST–MBNL1/$CCUG_{15}$ complexes is indicated by arrowheads (lanes 3–5). The addition of non-labeled $CCUG_{15}$ RNA (400 ng) reduced the formation of the shifted band (lane 4). The addition of anti-GST antibody resulted in a supershift of the complex (lane 5). (**C**) The effect of ion concentration. RNA–protein complexes are shown in the panel. Increasing concentrations of KCl (lanes 6–11) or NaCl (lanes 12–17) were added prior to the reaction of GST–MBNL1 and $CCUG_{15}$. Higher concentrations of these salts inhibit binding (lanes 9–11 and 15–17). (**D**) Effect of RNase treatment on the RNA–protein complex. Digestion of the $CCUG_{15}$ probe was analyzed by denaturing 13% PAGE containing 8 M urea. Increasing amounts of RNase V1 (lanes 19 and 22, 0.0001 unit; lanes 20 and 23, 0.001 unit; lanes 21 and 24, 0.01 unit) were added in the presence or absence of GST–MBNL1. The arrowhead shows the original length of the $CCUG_{15}$ probe (83 nt). By comparing lane 20 with lane 23, partial protection of RNA degradation can be observed.

MBNL1 interacts with CCUG repeats in *ZNF9* with minimal lengths

Previous reports have suggested that MBNL1 and its homologs colocalize with expanded CCUG repeats in DM2 (14,26). The length of the CCUG/CCTG repeat is highly polymorphic, ranging from 11 to 26 in normal individuals and from about 75 to several thousand in patients (5,6). To address whether MBNL1 interacts with unexpanded CCUG repeats, we examined the minimal length of CCUG motifs previously reported as shown in Figure 8B (DM2-1 and DM2-2) (5). Both motifs have insertions of non-CCUG tetranucleotides, such as UCUG and GCUG. In the gel retardation assay, GST–MBNL1 interacted specifically with both DM2-1 and DM2-2 (Fig. 8C).

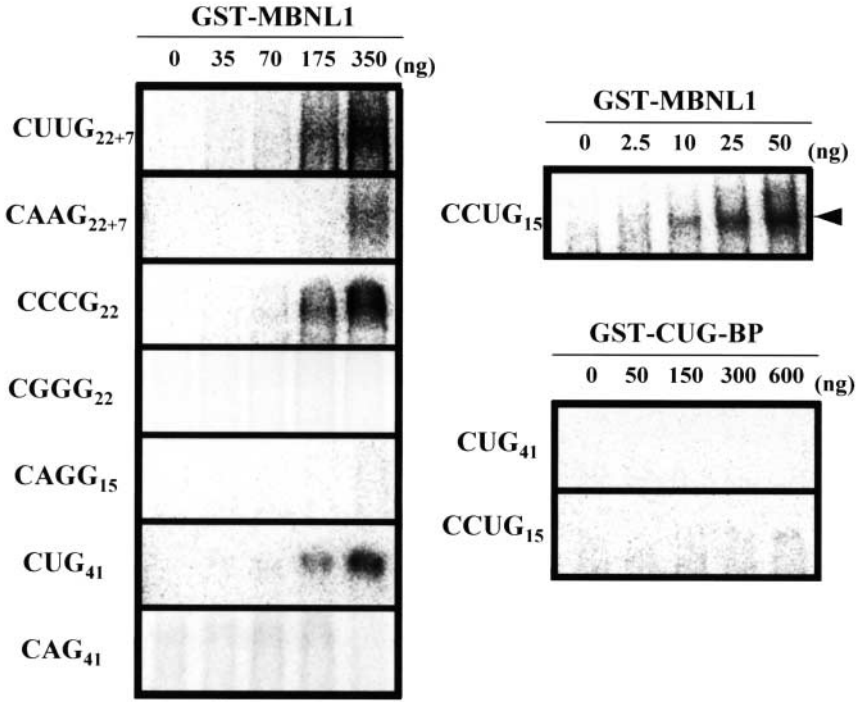

Figure 5. Dose-dependent and sequence-specific binding of GST–MBNL1. Dose-dependent binding of GST–MBNL1 to various repetitive RNA probes was analyzed by gel retardation analysis. ^{32}P-labeled probes as indicated were examined with increasing doses of GST–MBNL1 (left and right top panels). GST–CUG–BP was also analyzed for comparison (right below panel). RNA–protein complexes are shown in each panel.

From this result, we can predict two possibilities. One is that MBNL1 can bind to CCUG repeats despite the presence of the non-CCUG sequences. The other is that MBNL1 might interact with any length of CCUG repeat in human *ZNF9*, because it bound to the minimal CCUG repeat, and it might interact with longer CCUG repeats due to the length-dependence of the binding property.

DISCUSSION

We compared the specificity of three RNA-binding proteins, which were formerly suggested to interact with CUG repeat at least *in vitro.* However, two of them, CUG–BP and PKR, failed to reproduce its interaction with CUG repeats in the yeast three-hybrid system. These proteins also showed little interaction with CCUG repeats. On the other hand, MBNL1 interacted with both CUG and CCUG repeats in this system. In summary, our results indicates that double-stranded structure formed by CUG and CCUG repeats are recognized by MBNL1, but not by CUG–BP or PKR. These results for MBNL1 and CUG–BP in the yeast system are plainly consistent with the previous findings obtained by FISH (14,25,26), probably because both are *in vivo* systems. Evidence, including our present study, suggests that CUG–BP interacts with UG/GU containing single-stranded RNAs and this protein is not suitable for the direct target of the sequestration by the expanded CUG/CCUG RNAs, even though this protein is involved in the abnormal splicing in DM. There has been no report of FISH experiments on PKR or the activation of PKR *in vivo.* If our results actually reflect the *in vivo* function, PKR does not seem to be stably recruited by CUG or CCUG repeats, as in the case of CUG–BP. Yet the possibility of PKR activation by expanded RNAs cannot be ruled out, because an undetectable level of binding might cause its activation and affect cellular functions.

Toxic RNA effects and RNA-binding proteins

To date, RNA gain-of-function in DM has been suggested to have at least two types of cellular effects. Evidence has shown that abnormal splicing is involved in causing some of the symptoms of DM (18–20,41). In addition to CLC-1, IR and cTNT, the altered splicing of tau and myotubularin-related protein 1 has been reported in DM1 (42–44). Aberrant splicing of CLC-1 is also reported in DM2 (41), suggesting that the abnormal splicing event is a significant feature in the pathogenesis of both DM1 and DM2. Since the ectopic expression of the expanded repeat RNA can recapitulate the abnormal splicing patterns in cultured cells and transgenic mice (18,19,41), it is apparent that this phenotype is one of the results of RNA-gain-of-function.

Notably, abnormal splicing is accompanied by the up-regulation of CUG–BP in DM1 tissues and model cells (19,20). Moreover, the overexpression of CUG–BP alone can induce abnormal splicing patterns without the expression of expanded repeats in normal cells (19). Importantly, these indicate that aberrant splicing is induced by the increased activity of CUG–BP, but not by the loss-of-function of this protein. From this point of view, the absence of interaction between CUG–BP and CUG/CCUG repeats shown above and in the FISH analyses is compatible with the role of CUG–BP in

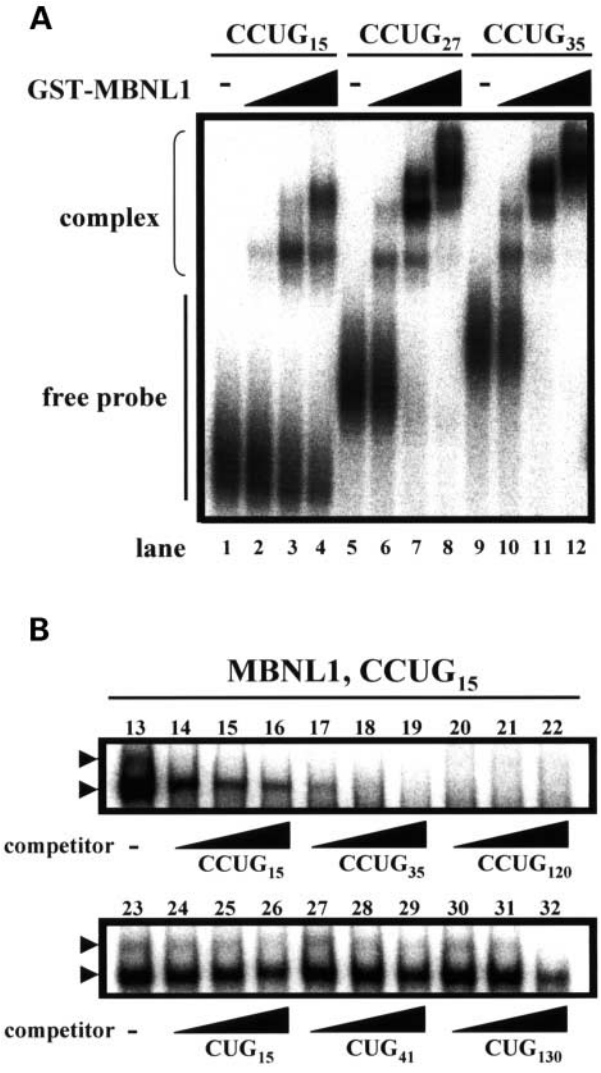

Figure 6. Length-dependent binding of GST–MBNL1 to CCUG repeats. (**A**) $CCUG_{15}$, $CCUG_{27}$ and $CCUG_{35}$ probes were examined with increasing doses of GST–MBNL1 in the gel retardation assay. GST–MBNL1 was added to the reaction at amounts of 0 ng (lanes 1, 5 and 9), 10 ng (lanes 2, 6 and 10), 50 ng (lanes 3, 7 and 11), or 250 ng (lanes 4, 8 and 12). Free $CCUG_{15}$ probe remains even at the highest dose of GST–MBNL1 (lane 4), while those of $CCUG_{27}$ and $CCUG_{35}$ are almost undetectable (lanes 8 and 12). (**B**) All lanes contain GST–MBNL1 and ^{32}P-labeled $CCUG_{15}$ probe. Increasing doses of $CCUG_{15}$ (lanes 14–16), $CCUG_{35}$ (lanes 17–19), $CCUG_{120}$ (lanes 20–22), CUG_{15} (lanes 24–26), CUG_{41} (lanes 27–29) and CUG_{130} (lanes 30–32) were added as cold competitors in advance of the probe addition. The amount of RNA was 100 ng (lanes 14, 17, 20, 24, 27 and 30), 300 ng (lanes 15, 18, 21, 25, 28 and 31) or 600 ng (lanes 16, 19, 22, 26, 29 and 32). The competition by $CCUG_{120}$ was more efficient than that by $CCUG_{15}$ or $CCUG_{35}$ (lanes 14, 17 and 20). CUG_{130} also competed best among CUG repeats (lane 32).

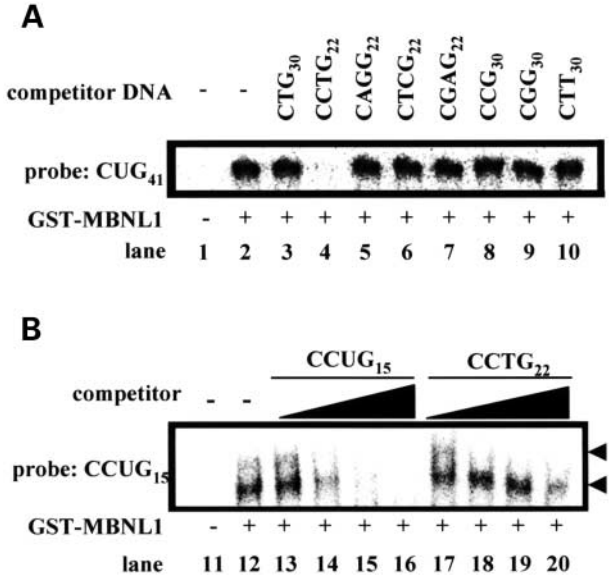

Figure 7. *In vitro* binding of GST–MBNL1 to CCTG DNA repeat. (**A**) Unlabeled oligo DNAs (300 ng each) were added as competitors into the reaction of GST–MBNL1 and CUG_{41} probe. The complex of CUG_{41} and GST–MBNL1 is indicated as a band in the panels. Competition by $CCTG_{22}$ was observed (lane 4). (**B**) Comparison of unlabeled CCUG RNA repeats and CCTG DNA repeats in the inhibition of the binding of GST–MBNL1 to the $CCUG_{15}$ probe. Increasing doses of $CCUG_{15}$ or $CCTG_{22}$ were added at amounts of 50 ng (lanes 13 and 17), 300 ng (lanes 14 and 18), 1000 ng (lanes 15 and 19), 3000 ng (lanes 16 and 20).

the abnormal splicing, because if CUG–BP interacts with these expanded repeats, the binding of CUG–BP to its target RNAs should be reduced, not leading to the aberrant splicing. Rather, since the affinity of CUG–BP to double-stranded CUG and CCUG repeats is low, the presence of expanded repeats may not interrupt the binding of CUG–BP to its target RNAs of splicing. Even if CUG–BP is sequestered by these repeats, the up-regulation of CUG–BP does not appear to be explained directly by the sequestration. Therefore, it seems reasonable to postulate an indirect mechanism as leading to the up-regulation of CUG–BP.

The other type of cellular effect by the expanded repeat is a defect in myogenesis accompanied by the altered expression of myogenic markers. Mahadevan and colleagues have suggested the abnormalities in differentiation of their cell culture models that expresses expanded CUG repeats in the 3′-UTR of *DMPK* (45). In the same cells, MyoD has been suggested to be down-regulated in the post-transcriptional level (46). In another report, an impaired cell cycle withdrawal due to a reduced expression of p21 in DM1 cells was also suggested (24). Interestingly, the reduction of MyoD protein was also suggested in the report (24). These properties may be involved in the muscle development abnormalities in congenital DM1 and/or defects in muscle regeneration, possibly leading to muscle wasting in adult patients (1). Thus, myogenic regulation is also a noteworthy feature of DM. However, it is probable that the abnormalities in splicing and myogenic differentiation may share relevance to each other rather than being two separable events.

Since the expression of MBNL1 is induced during myoblast differentiation (31), this protein might be involved in myogenic differentiation as in the case of muscleblind in fly. Interestingly, the expression of MBNL3 has recently been shown to be reduced during differentiation (47), in contrast to MBNL1. Furthermore, MBNL3 has an inhibitory role in myogenesis, in which the overexpression of MBNL3 suppresses the progress of differentiation (47). One speculation is that MBNL1 is a positive regulator of muscle differentiation and competes with MBNL3 for RNA substrates. FISH analysis has shown that not only MBNL1 but also MBNL2 and MBNL3 co-localize with

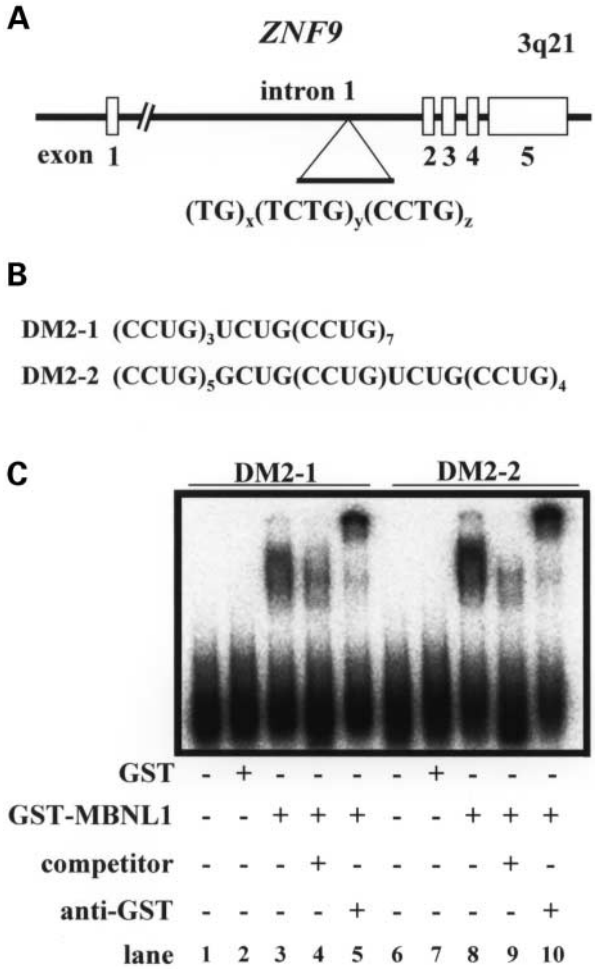

Figure 8. GST–MBNL1 bound to minimal CCUG repeat motifs in *ZNF9*. (**A**) Gene structure of *ZNF9*. Exons are indicated as open boxes. A repetitive sequence containing variable lengths of TG, TCTG and CCTG repeat elements is located in intron 1. (**B**) Two of the shortest CCUG elements of *ZNF9* ever reported were chosen as probes for the experiment below. Both contains non-CCUG tetranucleotide(s). (**C**) *In vitro* binding of GST–MBNL1 to both DM2-1 and DM2-2 RNAs. GST or GST–MBNL1 was incubated with ^{32}P-labeled probes. RNA–protein complexes are observed in the gel. The addition of unlabeled DM2-1 or DM2-2 reduced the shifted band (lanes 4 and 9), while the addition of anti-GST antibody induced supershifts of the complex (lanes 5 and 10).

CUG and CCUG repeats (26), suggesting that these three proteins have similar RNA-binding properties. Indeed, our preliminary results in the yeast system have suggested that MBNL2 also interacts with CCUG and CCCG repeats (unpublished data). However, it is also conceivable that, even though the binding specificities of MBNL proteins overlap partially, they may have different target RNAs. Since the myogenic defect is an important issue in DM, the myogenic roles of MBNL proteins must be clarified by further experiments.

RNA-binding properties of MBNL1

Although the length of CUG and CCUG repeats used in this study are smaller than those found in DM1 or DM2, competition analysis demonstrates the length-dependent interaction of MBNL1 with up to CUG_{130} or $CCUG_{120}$ *in vitro*, whose lengths are longer than the minimum lengths of pathogenesis. Although we cannot exclude the possibility that longer expanded CUG/CCUG repeats form other RNA structures and have other binding properties, the tendency of length-dependent binding may allow the prediction that MBNL1 binds to even longer expansion than the lengths used in this experiment, which is supported by the FISH results.

Previously, long CUG repeats have been shown to form a hairpin structure comprising a long stem and a single loop rather than with multiple stem-loops (34). In addition, detailed biochemical analyses in a recent report have shown that all CNG repeats and CCUG repeats form stable hairpin structures *in vitro*, and these RNAs contain heterogeneous conformers (folding variants) in the population due to alternative alignment of the hairpin structure (48). Notably, $CNG_{16\text{-}17}$ and $CCUG_{14}$ are sufficient lengths to form stable hairpins (48), suggesting that the $CCUG_{15}$ probe used in this study may be double-stranded. In the gel retardation assays, some portion of $CCUG_{15}$ probe remained unbound even in the overdose of MBNL1, and such population decreased when longer CCUG repeats were analyzed. This might indicate that some of $CCUG_{15}$ conformers do not bind to MBNL1, and as the length gets longer, the relative amount of such conformers decreases due to the stabilization of hairpin structure. It is not known what kind of conformers MBNL1 binds to or not. It is also unclear whether unbound conformers contain single-stranded RNAs. Next, the tendency that G-rich sequences such as CGG and CGGG are not preferred by MBNL1 seems to reflect the sequence selectivity of MBNL1 rather than its structure selectivity, because at least CGG repeats form a hairpin structure similar to those by CHG repeats (48). However, the authors also pointed out the peculiarity of CCG repeats in the formation of conformers and loop structure of the hairpins, may be due to the stability of CCG repeats highest among CNG repeats. These issues will be clarified by further biochemical and structural analyses of MBNL1.

According to the deletion analysis, some isoforms of MBNL1 did not bind to CCUG repeats. Nevertheless, they might bind to some RNAs other than CCUG, since all these isoforms have at least one zinc finger motif. Indeed, a peptide containing only the first CCCH motif of TTP (tristetraprolin), an RNA-binding protein with two CCCH motifs involved in the deadenylation of tumor necrosis factor-α mRNA (49), has been shown to bind to a UUAUU RNA, suggesting that each CCCH motif has the ability to bind to RNA by itself (50). The existence of multiple MBNL1 isoforms implies that there is some functional antagonism or feedback regulation among MBNL1 isoforms through the interaction with common RNAs or proteins.

Target RNA motifs of MBNL1

Molecular targets of MBNL1 are important for understanding the effect of its sequestration by the expanded repeats and could be involved in the pathogenesis of DM. From our results, targets of MBNL1 may be CHHG or CHG repeat-containing RNAs. Strikingly, CCUG repeats showed much higher affinity than CUG and other CHG/CHHG repeats assayed in the present study. A simple interpretation of the results is that the major target RNAs of MBNL1 may have a CCUG repeat

sequence. However, the interactions with CUG repeats or other minor sequences are also important because these interactions might be interfered easily by the presence of expanded pathogenic repeats. At least, the fact that expanded CUG repeats can sequester MBNL1 *in vivo* suggests the physiological interaction between MBNL1 and CUG repeats, although the length of CUG repeats in physiological target genes might be much shorter than the mutant RNA in DM1.

We have also shown that MBNL1 is able to interact with CCUG repeats containing insertions (DM2-1 and DM2-2). These insertions are seen in normal alleles of *ZNF9* and might be related to the stability of the CCTG tract in *ZNF9* (51). The insertions may alter the predicted secondary structure of CCUG repeats, but both DM2-1 and DM2-2 probes still have affinity to MBNL1. Currently, we do not know whether these insertions increase or decrease the affinity of MBNL1 to RNA repeats. Some of the target RNAs of MBNL1 might have such insertions in repetitive motifs. Importantly, length-dependent binding of MBNL1 to CUG/CCUG repeats suggests MBNL1's preference for stem and bulge structure rather than stem–loop structure, if a long and uninterrupted CUG or CCUG repeat actually folds as a long and stable hairpin structure. Mixture of CHHG and CHG motifs is also possible interactors of MBNL1. Indeed, two mixed sequences, $(CUUG)(CCUG)_4(CCCG)(CCUG)(CCAG)$ and $(CCUG)_7(CCCG)(CCUG)_5$ interacted with GST–MBNL1 in gel retardation assays (unpublished data), suggesting that MBNL1 binds to CHHG-mixed sequences and that an RNA containing eight tetranucleotides has sufficient length for binding to MBNL1. Thus, complex variations in repetitive sequences can be predicted as the binding motifs of MBNL1.

We searched the human and mouse genome databases for CHHG-containing genes and found many such genes. Since MBNL1 is a conserved protein, we postulated that CHHG motifs in these genes are also conserved between human and mouse. However, we have not still found conserved motifs between humans and mice in these genes, suggesting that physiological target motifs of MBNL1 may be complex configurations and/or that MBNL1 might have broader RNA-binding specificities. The CCUG tract in *ZNF9*, a potential target of MBNL1 as shown above, is conserved among human, chimpanzee and gorilla, but it is not found in mouse or rat, suggesting that CCUG motifs were acquired in the process of evolution (51). If *ZNF9* is an actual target of MBNL1, it is possible that MBNL1 might be involved in the metabolism of *ZNF9* pre-mRNA in the former three species. Interestingly, more than 500 bp up- or downstream of the CCUG tract is highly conserved among these three species. In addition, the UG repeat franking with the CCUG repeat (Fig. 8A) is conserved among all five above species (51). However, we do not know whether these regions have any physiological role or not. As we have examined CCUG tract in *ZNF9* only *in vitro*, the physiological significance of the interaction between MBNL1 and the CCUG tract should be examined *in vivo.*

Notably, a knockout mouse model of MBNL1 showed myotonia, cataracts and abnormal muscle histology in a recent report (52), strongly supporting the hypothesis that a loss of function of this protein by the expanded repeats leads to pathogenesis of DM. These mice also exhibit abnormalities in splicing of mouse CLC-1 (clc-1), and cardiac and fast skeletal muscle troponin T (tnnt2 and tnnt3). Although the molecular pathway causing these abnormalities is still unclear, MBNL1 may have some roles in the regulation of alternative splicing. From our results, it is possible that MBNL1 acts with CHHG/CHG repeat-like sequences in the process of splicing. To date, abnormal splicing of CLC-1 has been suggested to be caused by the altered expression of CUG–BP. It is mysterious whether both or either of these two proteins are involved in the splicing of CLC-1. There might be some direct or indirect crosstalk between CUG–BP and MBNL proteins. Our results presented here will be a help for further understanding of MBNL1 as well as the pathogenesis of DM.

MATERIALS AND METHODS

Vectors

For protein expression in yeast, $MBNL1_{40}$ and its deletion mutants, the structures of which are indicated in Figure 3, were amplified by polymerase chain reaction (PCR) so as to have *Bam*HI and *Sal*I sites in the 5′ and 3′ regions, respectively, of the open reading frame, and subcloned into pGAD424 (Clontech). pGAD424, containing CUG–BP, was the same as previously described (28). The p20 fragment of PKR was amplified using primers 5′-CATGGCTGGTGATCTTTCAGC AGGTTTCGG-3′ and 5′-ACGCGTCGACAGTAGCAAAAG AACCAGAGG-3′, and inserted into the *EcoR*I and *Sal*I sites of pGAD424. For bacterial expression, pET32b (Novagen) was adopted by replacing thioredoxin-tag and S-tag with GST. The resulting plasmid was cut by *Sac*I in the polylinker region, blunted with T_4 DNA polymerase, and self-ligated to adjust the reading frame of the His-tag. The newborn vector, designated pET-GX, is able to express a fusion protein with GST in the N-terminus and a His-tag in the C-terminus. Fragments of MBNL1 and CUG–BP were inserted into the polylinker sites of pET-GX. Oligo DNAs of repetitive sequences were purchased from Proligo. To make longer repetitive sequences, oligo DNAs were amplified and extended by the non-template PCR method (53), and inserted into the *Hinc*II site of either pBluescript SK+ (Stratagen) or pUC118 (Takara). After sequencing, repetitive fragments were cut off with appropriate restriction enzymes, blunted with T_4 DNA polymerase and inserted into the *Sma*I site of pIIIA/MS2-2 vector (38). For *in vitro* transcription, we used pBluescriptII SK+ vector, which was cut by *Sac*I and *Kpn*I, and ligated with linker nucleotides consisting of 5′-CGTCGACGAGCT-3′ and 5′-CGTCGACGG TAC-3′, both of which have a *Hinc*II site. The resulting vector was named pBSDM. Some of the repetitive DNA fragments above were inserted into the *Hinc*II site of pBSDM. DM2-1 and DM2-2 fragments (their sequences are shown in Fig. 8B) were also purchased and inserted into the *Hinc*II site of pBSDM. All constructs were confirmed by sequencing.

Yeast three-hybrid system

The transformation of yeast cells and reporter gene assays were performed as previously described (28,38). In the *HIS3* assay, yeast strain *L40-coat* was transformed with pGAD and pIIIA/MS2-2 vectors coding RNA-binding proteins and RNA, respectively, and

selected on plates lacking leucine and uracil. Yeast transformants were picked up and spotted onto selection plates lacking in leucine, uracil, and histidine, with or without 0.1, 0.5 or 1 mM 3-AT. The plates were incubated at 30°C for about 1 week and the viability of the yeast transformants was analyzed. We classified the binding activity as (+++++), (++++), (+++) and (++), when yeast growth was observed on the plates containing 1, 0.5, 0.1 and 0 mM 3-AT, respectively; (+) yeast grew in the absence of 3-AT after more than 1 week, (−) no growth of yeast transformants was observed even after prolonged incubation. Positive and negative controls were included to confirm the validity of this system using combinations of iron response element (IRE) and iron regulatory protein (IRP), or pIIIA/MS2-2 and pGAD424 empty vectors, respectively. According to the classification above, yeast transformants of IRE and IRP were (++++), while those of pIIIA/MS2-2 and pGAD424 were (−). The liquid β-gal assay was performed as described previously (28,38).

Bacterial expression and purification of recombinant proteins

pET-GX vector containing $MBNL1_{40}$ or CUG–BP was transformed into BL21(DE3) strain. An overnight culture of transformant in LB medium was diluted and shaken at 37°C for 1.5 h (or until the OD reached 0.3–0.4), then 0.1 mM IPTG was added. During induction by IPTG, the culture was shaken at 27°C for 1.5 h. Bacterial cells were collected by centrifugation, washed with PBS and suspended in 0.05 culture volumes of sonication buffer containing 50 mM sodium phosphate (pH 7.5), 150 mM NaCl, 1 mM PMSF, and 1/1000 volume of protease inhibitor cocktail (Sigma). Lysozyme was added at a final concentration of 1 mg/ml, then incubated for 30 min on ice. The reaction was sonicated on ice nine times for 10 s each. Triton-X was added to a final concentration of 1% and the lysate was centrifuged. The supernatant of the lysate was subjected to affinity purification using Glutathione Sepharose 4B (Amersham Biosciences) and Talon Metal Affinity Resin (Clontech), according to the manufacturer's protocol. The quantity and purity were checked on SDS–polyacrylamide electrophoresis gels by staining with Coomassie brilliant blue (CBB). The identity of GST–MBNL1 was verified by peptide mass finger printing with a mass spectrometry (AXIMA-CFR, SHIMADZU) following trypsinization of GST–MBNL1.

In vitro binding assay

To synthesize RNAs, pBSDM with repetitive sequences were cut by *Pvu*II and either *Kpn*I or *Sac*I, depending on the direction of the transcription, and blunted with T_4 DNA polymerase. Restriction fragments containing repetitive sequences were purified by agarose gel electrophoresis. *In vitro* transcription was performed using mMESSAGE mMACHINE (Ambion) according to the manufacturer's directions. When labeling RNAs, 1 µl of [α-^{32}P]-CTP (~800 Ci/mmol) was added to the reaction. Unincorporated nucleotides were discarded using NucAway Spin Columns (Ambion). The gel mobility shift assay was performed as described (34). In brief, a purified protein and RNA were mixed in reaction buffer containing 50 mM Tris–HCl (pH 7.6), 100 mM NaCl, 5% glycerol, 1 mg/ml yeast tRNA, 1 mg/ml bovine serum albumin and 1 mM DTT, and incubated for 10 min at room temperature. Typically, $\sim 5 \times 10^5$ cpm RNA was added to the reaction. When necessary, non-labeled RNA competitors, anti-GST antibody (Santa Cruz) or RNaseV1 (Ambion), were added as indicated in each figure legend. The reaction was resolved by native PAGE using 4–5% polyacrylamide gels and 0.5× Tris borate buffer. In the case of RNase treatment as shown in Figure 4D, denaturing 13% polyacrylamide gels containing 8 M urea were used. The gels were dried after fixation with 10% methanol and 10% acetic acid, autoradiographed, and visualized by a BAS-2500 imaging analyzer (Fujifilm).

ACKNOWLEDGEMENTS

We would like to thank Dr Marvin Wickens of the University of Wisconsin at Madison for kindly providing us with the yeast three-hybrid system. We also thank Dr S. Suo for valuable discussions and encouragement. Y.K. is supported by a JSPS Research Fellowship for Young Scientists. This work is supported by research grants from the Ministry of Health, Labor and Welfare, Japan, and 'Ground-based Research Announcement for Space Utilization' promoted by the Japan Space Forum.

REFERENCES

1. Harper, P.S. (2001) *Myotonic Dystrophy*, 3rd edn. WB Saunders, London.
2. 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.
3. Fu, Y.H., Friedman, D.L., Richards, S., Pearlman, J.A., Gibbs, R.A., Pizzuti, A., Ashizawa, T., Perryman, M.B., Scarlato, G., Fenwick, R.G. *et al.* (1993) Decreased expression of myotonin–protein kinase messenger RNA and protein in adult form of myotonic dystrophy. *Science*, **260**, 235–238.
4. 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.
5. 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.
6. Day, J.W., Ricker, K., Jacobsen, J.F., Rasmussen, L.J., Dick, K.A., Kress, W., Schneider, C., Koch, M.C., Beilman, G.J., Harrison, A.R. *et al.* (2003) Myotonic dystrophy type 2: molecular, diagnostic and clinical spectrum. *Neurology*, **60**, 657–664.
7. Reddy, S., Smith, D.B., Rich, M.M., Leferovich, J.M., Reilly, P., Davis, B.M., Tran, K., Rayburn, H., Bronson, R., Cros, D. *et al.* (1996) Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy. *Nat. Genet.*, **13**, 325–335.
8. Jansen, G., Groenen, P.J., Bachner, D., Jap, P.H., Coerwinkel, M., Oerlemans, F., van den Broek, W., Gohlsch, B., Pette, D., Plomp, J.J. *et al.* (1996) Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. *Nat. Genet.*, **13**, 316–324.
9. Berul, C.I., Maguire, C.T., Aronovitz, M.J., Greenwood, J., Miller, C., Gehrmann, J., Housman, D., Mendelsohn, M.E. and Reddy, S. (1999) DMPK dosage alterations result in atrioventricular conduction abnormalities in a mouse myotonic dystrophy model. *J. Clin. Invest.*, **103**, R1–7.
10. Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M. and Thornton, C.A. (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science*, **289**, 1769–1773.
11. 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.

12. 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.
13. 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.
14. Mankodi, A., Urbinati, C.R., Yuan, Q.P., Moxley, R.T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M.S. and Thornton, C.A. (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.*, **10**, 2165–2170.
15. Timchenko, L.T., Miller, J.W., Timchenko, N.A., DeVore, D.R., Datar, K.V., Lin, L., Roberts, R., Caskey, C.T. and Swanson, M.S. (1996) Protein Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucl. Acids Res.*, **24**, 4407–4414.
16. Timchenko, L.T., Timchenko, N.A., Caskey, C.T. and Roberts, R. (1996) Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy. *Hum. Mol. Genet.*, **5**, 115–121.
17. Ladd, A.N., Charlet, N. and Cooper, T.A. (2001) The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol. Cell. Biol.*, **21**, 1285–1296.
18. 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.
19. 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.
20. 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.
21. Timchenko, N.A., Welm, A.L., Lu, X. and Timchenko, L.T. (1999) CUG repeat binding protein (CUGBP1) interacts with the 5′ region of C/EBPbeta mRNA and regulates translation of C/EBPbeta isoforms. *Nucl. Acids Res.*, **27**, 4517–4525.
22. Welm, A.L., Mackey, S.L., Timchenko, L.T., Darlington, G.J. and Timchenko, N.A. (2000) Translational induction of liver-enriched transcriptional inhibitory protein during acute phase response leads to repression of CCAAT/enhancer binding protein alpha mRNA. *J. Biol. Chem.*, **275**, 27406–27413.
23. Paillard, L., Legagneux, V. and Osborne, H.B. (2003) A functional deadenylation assay identifies human CUG-BP as a deadenylation factor. *Biol. Cell*, **95**, 107–113.
24. Timchenko, N.A., Iakova, P., Cai, Z.J., Smith, J.R. and Timchenko, L.T. (2001) Molecular basis for impaired muscle differentiation in myotonic dystrophy. *Mol. Cell. Biol.*, **21**, 6927–6938.
25. Fardaei, M., Larkin, K., Brook, J.D. and Hamshere, M.G. (2001) *In vivo* co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucl. Acids Res.*, **29**, 2766–2771.
26. Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshere, M.G., Harper, P.S. and Brook, J.D. (2002) Three proteins, MBNL, MBLL and MBXL, co-localize *in vivo* with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.*, **11**, 805–814.
27. Michalowski, S., Miller, J.W., Urbinati, C.R., Paliouras, M., Swanson, M.S. and Griffith, J. (1999) Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. *Nucl. Acids Res.*, **27**, 3534–3542.
28. Takahashi, N., Sasagawa, N., Suzuki, K. and Ishiura, S. (2000) The CUG-binding protein binds specifically to UG dinucleotide repeats in a yeast three-hybrid system. *Biochem. Biophys. Res. Commun.*, **277**, 518–523.
29. Suzuki, H., Jin, Y., Otani, H., Yasuda, K. and Inoue, K. (2002) Regulation of alternative splicing of alpha-actinin transcript by Bruno-like proteins. *Genes Cells*, **7**, 133–141.
30. Timchenko, L.T., Iakova, P., Welm, A.L., Cai, Z.J. and Timchenko, N.A. (2002) Calreticulin interacts with C/EBPalpha and C/EBPbeta mRNAs and represses translation of C/EBP proteins. *Mol. Cell. Biol.*, **22**, 7242–7257.
31. Miller, J.W., Urbinati, C.R., Teng-Umnuay, 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.
32. Begemann, G., Paricio, N., Artero, R., Kiss, I., Perez-Alonso, M. and Mlodzik, M. (1997) *muscleblind*, a gene required for photoreceptor differentiation in *Drosophila*, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. *Development*, **124**, 4321–4331.
33. Artero, R., Prokop, A., Paricio, N., Begemann, G., Pueyo, I., Mlodzik, M., Perez-Alonso, M. and Baylies, M.K. (1998) The *muscleblind* gene participates in the organization of Z-bands and epidermal attachments of *Drosophila* muscles and is regulated by *Dmef2*. *Dev. Biol.*, **195**, 131–143.
34. Tian, B., White, R.J., Xia, T., Welle, S., Turner, D.H., Mathews, M.B. and Thornton, C.A. (2000) Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA*, **6**, 79–87.
35. Williams, B.R. (1999) PKR; a sentinel kinase for cellular stress. *Oncogene*, **18**, 6112–6120.
36. Bevilacqua, P.C. and Cech, T.R. (1996) Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR. *Biochemistry*, **35**, 9983–9994.
37. Tian, B. and Mathews, M.B. (2001) Functional characterization of and cooperation between the double-stranded RNA-binding motifs of the protein kinase PKR. *J. Biol. Chem.*, **276**, 9936–9944.
38. SenGupta, D.J., Zhang, B., Kraemer, B., Pochart, P., Fields, S. and Wickens, M. (1996) A three-hybrid system to detect RNA-protein interactions *in vivo*. *Proc. Natl Acad. Sci. USA*, **93**, 8496–8501.
39. Bernstein, D.S., Buter, N., Stumpf, C. and Wickens, M. (2002) Analyzing mRNA-protein complexes using a yeast three-hybrid system. *Methods*, **26**, 123–141.
40. Peel, A.L., Rao, R.V., Cottrell, B.A., Hayden, M.R., Ellerby, L.M. and Bredesen, D.E. (2001) Double-stranded RNA-dependent protein kinase, PKR, binds preferentially to Huntington's disease (HD) transcripts and is activated in HD tissue. *Hum. Mol. Genet.*, **10**, 1531–1538.
41. 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.
42. Sergeant, N., Sablonniere, B., Schraen-Maschke, S., Ghestem, A., Maurage, C.A., Wattez, A., Vermersch, P. and Delacourte, A. (2001) Dysregulation of human brain microtubule-associated tau mRNA maturation in myotonic dystrophy type 1. *Hum. Mol. Genet.*, **10**, 2143–2155.
43. Seznec, H., Agbulut, O., Sergeant, N., Savouret, C., Ghestem, A., Tabti, N., Willer, J.C., Ourth, L., Duros, C., Brisson, E. *et al.* (2001) Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities. *Hum. Mol. Genet.*, **10**, 2717–2726.
44. Buj-Bello, A., Furling, D., Tronchere, H., Laporte, J., Lerouge, T., Butler-Browne, G.S. and Mandel, J.L. (2002) Muscle-specific alternative splicing of myotubularin-related 1 gene is impaired in DM1 muscle cells. *Hum. Mol. Genet.*, **11**, 2297–2307.
45. 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.
46. Amack, J.D., Reagan, S.R. and Mahadevan, M.S. (2002) Mutant DMPK 3′-UTR transcripts disrupt C2C12 myogenic differentiation by compromising MyoD. *J. Cell Biol.*, **159**, 419–429.
47. Squillace, R.M., Chenault, D.M. and Wang, E.H. (2002) Inhibition of muscle differentiation by the novel muscleblind-related protein CHCR. *Dev. Biol.*, **250**, 218–230.
48. Sobczak, K., de Mezer, M., Michlewski, G., Krol, J. and Krzyzosiak, W.J. (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucl. Acids Res.*, **31**, 5469–5482.
49. Carballo, E., Lai, W.S. and Blackshear, P.J. (1998) Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin. *Science*, **281**, 1001–1005.
50. Michel, S.L., Guerrerio, A.L. and Berg, J.M. (2003) Selective RNA binding by a single CCCH zinc-binding domain from Nup475 (Tristetraprolin). *Biochemistry*, **42**, 4626–4630.
51. Liquori, C.L., Ikeda, Y., Weatherspoon, M., Ricker, K., Schoser, B.G., Dalton, J.C., Day, J.W. and Ranum, L.P. (2003) Myotonic dystrophy type 2: human founder haplotype and evolutionary conservation of the repeat tract. *Am. J. Hum. Genet.*, **73**, 849–862.
52. 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) A muscleblind knockout model for myotonic dystrophy. *Science*, **302**, 1978–1980.
53. Takahashi, N., Sasagawa, N., Suzuki, K. and Ishiura, S. (1999) Synthesis of long trinucleotide repeats *in vitro*. *Neurosci. Lett.*, **262**, 45–48.