3534–3542 *Nucleic Acids Research, 1999, Vol. 27, No. 17*

Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein

Susan Michalowski, Jill W. Miller[1], Carl R. Urbinati[1], Miltiadis Paliouras[1], Maurice S. Swanson[1] and Jack Griffith*

Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA and [1]Department of Molecular Genetics and Microbiology and Centers for Gene Therapy and Mammalian Genetics, University of Florida College of Medicine, Gainesville, FL 32610-0266, USA

Received May 3, 1999; Revised and Accepted July 8, 1999

ABSTRACT

Myotonic dystrophy (DM) is associated with a $(CTG)_n$ triplet repeat expansion in the 3'-untranslated region of the myotonic dystrophy protein kinase (DMPK) gene. Using electron microscopy, we visualized large RNAs containing up to 130 CUG repeats and studied the binding of purified CUG-binding protein (CUG-BP) to these RNAs. Electron microscopic examination revealed perfect double-stranded (ds)RNA segments whose lengths were that expected for duplex RNA. The RNA dominant mutation model for DM pathogenesis predicts that the expansion mutation acts at the RNA level by forming long dsRNAs that sequester certain RNA-binding proteins. To test this model, we examined the subcellular distribution and RNA-binding properties of CUG-BP. While previous studies have demonstrated that mutant DMPK transcripts accumulate in nuclear foci, the localization pattern of CUG-BP in both normal and DM cells was similar. Although CUG-BP in nuclear extracts preferentially photocrosslinked to DMPK transcripts, this binding was not proportional to $(CUG)_n$ repeat size. Moreover, CUG-BP localized to the base of the RNA hairpin and not along the stem, as visualized by electron microscopy. These results provide the first visual evidence that the DM expansion forms an RNA hairpin structure and suggest that CUG-BP is unlikely to be a sequestered factor.

INTRODUCTION

Myotonic dystrophy (DM) is the most common form of adult onset muscular dystrophy and occurs once in 8000 births (1–3). Individuals with DM show many deficits; in particular, neuromuscular involvement is prominent, with muscle weakness and wasting (1). Mapping of the DM locus in affected individuals revealed an expanded block of repeating CTG nucleotide triplets in the 3'-untranslated region (3'-UTR) of the myotonic dystrophy protein kinase (DMPK) gene (4–7). In normal individuals, this repeat block ranges in length from five to 37 repeats. Individuals with mild or no symptoms, but who have had offspring with DM, have larger repeat lengths (~50–100 repeats). The repeat block may be >1000 repeats in individuals with the severe congenital form of the disease. DM is an autosomal dominant disorder that shows anticipation from one generation to the next. Longer CTG repeat tracts are generally associated with more severe symptoms that appear earlier in life (5,8).

The mechanism by which expansion of the CTG repeat block causes DM is not known, but the multi-systemic nature of the disease suggests a complex molecular pathway (reviewed in 9). Haploinsufficiency has been proposed as a mechanism to explain the dominant nature of DM (10,11). However, DMPK knockout mice do not show the complex array of symptoms associated with the human disease (12,13). Further, no DM cases have yet been found resulting from a point mutation or deletion in DMPK coding sequences. In contrast, numerous such examples have been observed in fragile X syndrome, where individuals with mutations in the *FMR1* coding sequence exhibit a similar disease phenotype as those with CCG repeat expansions in the 5'-UTR (14). Finally, DMPK protein levels in severely affected congenital patients are generally less than half the normal levels, suggesting that the expanded CTG repeat block may adversely affect expression of both alleles (15,16).

A second model for DM pathogenesis evolved from studies on the chromatin structure of DNA containing CTG repeats and the structure of this region. Wang *et al*. (17) demonstrated that DM patient DNA with large CTG blocks had an extremely high affinity for histone octamers, resulting in the generation of hyperstable nucleosomes *in vitro*. On the basis of these and further studies (18), a chromatin structure model was proposed in which the DM expansion mutation creates hyperstable nucleosomal regions resulting in the formation of a control locus that subsequently represses adjacent genes (19). Parallel *in vivo* studies by Tapscott and colleagues (20) revealed that the region downstream of the DMPK gene acquires a nuclease-insensitive conformation when the DMPK CTG block undergoes expansion. However, this chromatin structure model as applied

*To whom correspondence should be addressed. Tel: +1 919 966 2151; Fax: +1 919 966 3015; Email: jdg@med.unc.edu

to the DMPK gene conflicts with several recent studies which indicate that transcription of the mutant allele is not severely affected by the CTG expansion (21,22). Recently, two other genes have been mapped adjacent to DMPK. The DMWD gene (formerly gene 59) lies just upstream and may share control sequences with the DMPK gene (23,24). The SIX5 gene (formerly DMAHP) maps downstream of the DMPK gene and the DMPK 3'-UTR CTG repeat overlaps the 5' control region of SIX5, suggesting functional linkage between these genes (25,26). The effect of the CTG repeat on expression of the DMWD and SIX5 genes is less clear. SIX5 expression is reported to be lower in some DM patients (27,28), while others have reported no effect on SIX5 expression in other DM patients (22). In the chromatin structure model, as the size of the expansion increases the unusual chromatin structure created by the growing array of hyperstable nucleosomes can alter gene expression at an ever increasing distance from the site of the expansion itself. Thus these apparently conflicting observations may reflect different expansion sizes.

An RNA dominant mutation model has also been proposed (3,29). There are two key features of this model. First, the expanded CUG tract in the RNA transcript folds back on itself to create a stable duplex hairpin. Second, these structures act as inappropriate protein sinks, removing certain cellular proteins from their normal functions (30,31). This depletion would prevent their association with normal target transcripts that require these proteins for pre-mRNA processing and mRNA export (reviewed in 9). Accumulating evidence suggests that single-stranded (ss)DNA composed of CTG repeats folds onto itself to form stable duplex hairpins (32–35). Since U-U base pairing in RNA can occur, long CUG repeat-containing RNA would also be predicted to form stable hairpins (36,37). In addition, a recent study using chemical modification methods has provided evidence for such hairpins in CUG RNAs containing 11–49 repeats (38). In further support of the RNA dominant mutation model, recent evidence suggests that DM cells are defective in nucleocytoplasmic export of mutant DMPK transcripts (21,22,39) and alternative pre-mRNA splicing of cardiac troponin T (cTNT) (40). A candidate for the sequestered factor, the CUG-binding protein (CUG-BP), has been identified and characterized as a protein that preferentially binds to $(CUG)_8$ oligonucleotides but does not bind to ssCTG or ssCAG DNA nor to CGG repeat RNAs (31,41). CUG-BP is localized primarily in the nucleus and exists in several different isoforms via phosphorylation (41,42). Moreover, DM cells show different CUG-BP isoforms and altered $(CUG)_8$ binding activity (31). The three RNA-binding domains (RBDs) present in CUG-BP are predicted primarily to recognize ssRNA elements (reviewed in 43,44). In addition, CUG-BP was isolated by binding to CUG_8 oligos that would not be expected to form hairpins (38). The sequestration model predicts that CUG-BP binds along the length of the CUG repeat expansion with binding proportional to the number of triplet repeats, but binding along the stem may require a dsRNA-binding protein.

In this report, we have examined the two key features of the RNA dominant model: generation of duplex RNA hairpins and the hypothesis that CUG-BP is a sequestered factor in DM cells. Electron microscopy (EM) was used to visualize the RNA structures formed by large CUG repeats and to examine their interaction with CUG-BP. We show that RNAs composed of either CUG or CAG repeats form very stable regular duplex structures with a rise typical of duplex RNA. In contrast to the proposal that CUG-BP is sequestered by the DM triplet repeat expansion, studies of the subcellular distribution and RNA-binding properties of CUG-BP are presented to show that CUG-BP is primarily a ssRNA-binding protein that has a binding preference for CUG-rich RNA elements but not duplex CUG hairpins. The implications of these observations to the RNA dominant model for DM pathogenesis are discussed.

MATERIALS AND METHODS

Plasmid constructs and *in vitro* transcription

DMPK plasmid inserts containing six (pDMPK.8-4), 54 (pDMPK.8-16) and 90 (pDMPK.8-6) CUG repeats were generated by PCR utilizing primers flanking the CUG repeat and $(CAG)_{10}$ and $(CTG)_{10}$ primers. Three individual reactions were performed employing a subclone of the DMPK cDNA (nt 2212–2849, accession no. M87312) in pSP72 (Promega, Madison, WI). Reaction 1 contained a DMPK-specific 5' primer (5'-CTGCTGCTGCTGCTGGGGGGATCCCAGACCA-TTTCTTCTGC-3') and a SP6 3' primer. Reaction 2 contained a DMPK-specific 3' primer (5-CAGCAGCAGCAGCAGAAT-TCCCGGCTACAAGG-3') and a T7 5' primer. Reaction 3 contained only $(CAG)_{10}$ and $(CTG)_{10}$ primers. PCR reactions were performed under standard conditions. After completion of the initial reactions, a second set of reactions were performed using products from reaction 1 or 2 combined with the product from reaction 3. The products from reactions 1+3 and 2+3 were then combined and amplified for an additional 25 cycles. Products were cleaved with *Sma*I and *Hin*dIII and cloned into pSP72 (Promega) to generate plasmids DMPK/6, 54 and 90. The pDMPK.10 clone containing a deletion of the CTG repeat was created by PCR utilizing overlapping primers that lacked CTG repeats in a strategy similar to that described for plasmids DMPK/6, 54 and 90. Two reactions were performed. Reaction 1 contained a DMPK-specific 5' primer that lacked a CTG repeat (5'-CCTTGTAGCCGGGAATGGGGGGATCACAGA-CCATTC-3') and a 3' SP6 primer. Reaction 2 contained a DMPK-specific 3' primer that lacked a CTG repeat (5'-GG-TCTGTGATCCCCCCATTCCCGGCTACAAGGACC-3') and a 5' T7 primer. Products from these reactions were combined and amplified for 25 cycles. All clones were sequenced prior to use as templates for *in vitro* transcription. For EM studies, RNA substrates were transcribed from the DNA templates CTG 75 and CTG 130. The DNA templates were derived from plasmids pSH1 and pSH2 (45), respectively, by cloning into pGEM3zf+ (Stratagene, La Jolla, CA). The 58 bp of DNA 5' of the CTG repeats is identical between the DMPK and CTG plasmids. However, the DMPK plasmids contain an additional 118 bp 5' of the CTG repeat. The 145 bp of DNA 3' of the CTG repeats is 94% identical between the two plasmid types.

For photocrosslinking, RNAs were transcribed from plasmids pRPL14.1, pRPL14.2, pDMPK.8-4, pDMPK.8-16, pDMPK.8-6 and pDMPK.10 *in vitro* using T7 polymerase in the presence of [α-^{32}P]UTP and subsequently purified by denaturing gel electrophoresis as previously described (31). For EM, DNA templates were prepared for transcription by digesting either CTG 75 or CTG 130 with *Eco*RI for generation of the CAG repeat strand using SP6 RNA polymerase (New England Biolabs, Beverley, MA) or by digesting with *Hin*dIII

for generation of the CUG repeat transcript using T7 RNA polymerase (New England Biolabs). The tailed CUG 130 template contains an additional 212 nt 3' of the repeat and was prepared by digesting CTG 130 with *Pvu*II prior to transcription with T7 RNA polymerase. Transcription of the templates was carried out in 40 mM Tris–HCl (pH 7.9), 6 mM $MgCl_2$, 2 mM spermidine and 10 mM dithiothreitol supplemented with 0.5 mM each of ATP, CTP, GTP and UTP. Reactions were carried out for 2 h at 37°C for T7 RNA polymerase and at 40°C for SP6 RNA polymerase. The RNAs were purified by phenol/chloroform extraction and ethanol precipitation. The 261 bp dsRNA control was prepared as described by Wang *et al*. (45). Mung bean nuclease (MBN) was used according to the manufacturer's protocol (New England BioLabs).

To express histidine-tagged CUG-BP (hNab50) in *Escherichia coli*, full-length cDNAs encoding both proteins were cloned into pET15b (Novagen, Madison, WI). Briefly, a full-length CUG-BP cDNA was amplified by PCR with a 5' primer containing an *Nde*I site (5'-GGGCATATGAGCCAGAAGGAAGGTCCAG-3') and a 3' primer containing a *Bam*HI site (5'-CCCGGATCCAACAGCAAAACCACCA-3'). The resulting CUG-BP PCR product was digested with *Nde*I and *Bam*HI and subcloned into pET15b.

Cell immunofluorescence, photocrosslinking assay and immunopurifications

Indirect cell immunofluorescence was performed using a 1:500 dilution of the monoclonal antibody (mAb) 3B1 and a 1:10 dilution of goat anti-mouse IgG1–FITC (human adsorbed) as described previously (31). Normal human fibroblasts (GM05897B and GM03523) and DM patient fibroblasts containing 50–80 CTG repeats (GM03991), 500 CTG repeats (GM03755) and up to 2000 CTG repeats (GM03759 and GM03132) were obtained from the Coriell cell repositories (Camden, NJ). DM fibroblasts GM03755 and GM03132 have been shown to accumulate mutant DMPK transcripts in nuclear foci (39). Immunofluorescence data were collected using a CCD camera attached to a Zeiss Axiophot fluorescence microscope and images were processed using Adobe Photoshop. Photocrosslinking assays and immunopurifications were performed as described previously (31).

Construction of the DM lymphoblast cDNA library and isolation of RPL14 clones

Polyadenylated RNAs were isolated from human DM1 lymphoblasts (GM03986A) and purified by oligo(dT) chromatography. A cDNA library was prepared using a commercially available cDNA synthesis kit and λ packaging extracts (Stratagene). Libraries were screened by hybridization for cDNAs containing $(CTG)_n$ repeats using a $(CAG)_{10}$ RNA probe. Filters were hybridized overnight at 50°C in 50% deionized formamide, 6× SSC, 1% SDS, 0.1% Tween-20, 100 μg/ml tRNA and 5×10^5 c.p.m./ml of labeled transcript. The filters were washed twice in 1× SSC, 0.1% SDS at room temperature for 30 min and then twice in 0.1× SSC, 0.1% SDS at 65°C for 30 min. Positive plaques were purified and the corresponding plasmids isolated and characterized by DNA sequencing. All of the cDNA inserts encoded a protein highly related to the rat ribosomal 60S subunit protein L14 (46) and therefore the corresponding gene was named RPL14. Two different types of RPL14 cDNAs were isolated that differed only in a polymorphic CUG repeat region (10 or 22 repeats) which encoded 10 or 22 alanine residues near the C-terminus of the protein. Database searches revealed that the human RPL14 gene had been previously characterized as CAG-isl 7, a cDNA clone isolated during a search for diabetes genes from a human islet cell cDNA library (47).

Purification of recombinant CUG-BP

To purify recombinant histidine-tagged CUG-BP (hNab50), bacterial strains carrying pHIS-Nab50 were grown to an OD_{600} = 0.6 in 500 ml TB (1.2% bactotryptone, 2.4% yeast extract, 0.4% glycerol, 0.1 M KPO_4 buffer, pH 7.15) supplemented with 0.5 mg/ml carbenicillin. Cells were collected by centrifugation at 1600 *g* for 10 min, resuspended in 500 ml TB supplemented with 0.5 mg/ml carbenicillin and 1 mM isopropyl β-D-thiogalactoside (Sigma) and subsequently grown at 30ºC with vigorous shaking for 2 h. Cells were then harvested by centrifugation at 7200 *g* for 10 min, resuspended in 20 ml of ice-cold binding buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris–HCl, pH 7.9). Resuspended cells were lysed by a single pass through a French press (1250 lb/in^2). Lysate was clarified by centrifugation at 17 000 *g* for 20 min and the supernatant filtered through a 0.45 μM filter. Recombinant protein was purified using a His-Bind affinity column (Novagen) according to the manufacturer's recommendation except that bound protein was eluted with 100 mM EDTA, 500 mM NaCl, 10 mM $NaPO_4$, pH 6.5). Eluted protein was dialyzed against 2 l of dialysis buffer (20 mM Tris–HCl, pH 7.6, 100 mM NaCl, 1 mM dithiothreitol, 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride).

Preparation of RNA and protein–RNA complexes for EM

The buffer used to prepare protein–RNA complexes consisted of 20 mM HEPES (pH 7.4) and 100 mM KCl. Reactions included 100 ng of RNA template and 0–500 ng of CUG-BP as indicated in the text and figure legends. Incubations were carried out for 10 min at 30°C and the samples fixed with glutaraldehyde (0.6%) at room temperature for 5 min. The samples were filtered through columns of Bio-Gel A-5m (Bio-Rad) equilibrated with 10 mM Tris–HCl (pH 7.6), 0.1 mM EDTA. The filtered samples were mixed with a buffer containing spermidine, adsorbed to glow charged thin carbon foils, dehydrated through a water/ethanol series and rotary shadow cast with tungsten as described (48). RNA samples were prepared for EM by mixing the RNA (0.5 ng/μl) with a buffer containing spermidine, adsorbed to glow charged thin carbon foils and prepared as described above. Samples were visualized in a Philips 400 instrument. Micrographs for publication were scanned from negatives using a Nikon LS 4500 multiformat film scanner and the contrast optimized and panels arranged using Adobe Photoshop. Morphometry measurements were done using a Summagraphics digitizer coupled to a Macintosh computer programmed with software developed by J.D.G.

RESULTS

Visualization of CUG and CAG repeat-containing RNAs

Previous studies (38) have concluded that RNA containing tandem CUG repeats form dsRNA hairpins; however, long CUG repeat-containing RNAs have not been visualized. RNA

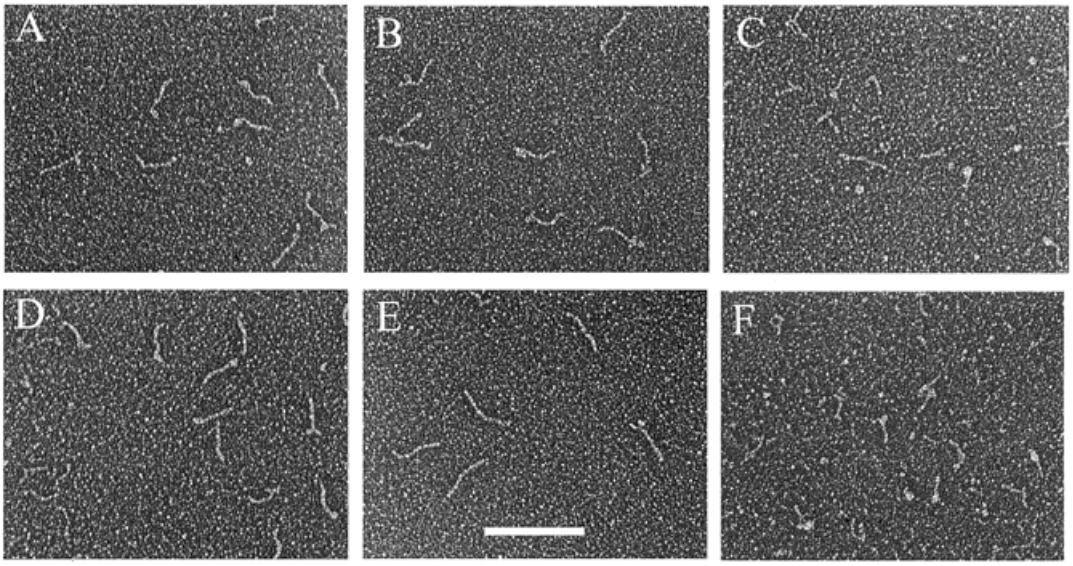

Figure 1. CUG repeat-containing RNA forms hairpins. RNAs were produced by transcribing from various plasmids using either T7 or Sp6 RNA polymerase. Electron micrographs (shown in reverse contrast) of RNAs prepared for EM by direct adsorption to carbon foils, washing, air drying and rotary shadow casting with tungsten (see Materials and Methods). (**A**) CUG 130 contains 130 repeats (390 nt) and 203 nt of flanking sequence; (**B**) CUG 130T contains 130 repeats and an additional 212 nt to CUG 130 3' of the repeat; (**C**) CUG 75 contains 75 repeats (225 nt) and 203 nt of flanking sequence; (**D**) CAG 130 contains 130 CAG repeats and is the antisense transcript of CUG 130; (**E**) dsRNA control (261 bp); (**F**) DMPK 90 contains 90 repeats (270 nt) and 321 nt of flanking sequence. The bar represents 500 bp.

molecules 429 or 593 bases in length containing 75 or 130 CUG trinucleotide repeats, respectively, were generated by *in vitro* transcription. Complementary RNAs of 424 and 588 bases containing 75 and 130 CAG repeats were produced in parallel (Materials and Methods). The repeat tracts in these RNAs are flanked by non-repeat RNA sequences. The purified RNAs were examined by EM following direct adsorption to thin carbon foils, dehydration and rotary shadow casting with tungsten. Examination of the CUG and CAG repeat RNAs revealed fields of molecules consisting of stiff rod-like segments with a ball of unstructured material at one end (Fig. 1A–D). Scoring such fields, >99% of the molecules present had this appearance, in contrast to fully ball-like particles. The rod-like segment was very similar in thickness and appearance to fully duplex RNAs (Fig. 1E) and the ball-like appendage was typical of unstructured RNA segments prepared by these EM procedures (49). The length of the rod-like segment was greater for the RNAs containing 130 (Fig. 1A and D), in contrast to 75 (Fig. 1C), repeats and their precise measurements are described below. Greater than 90% of the rods were straight with no sharp kinks that would have resulted from one or more triplet repeats bulging out from the duplex segment. This suggested that the rod segments resulted from the RNA repeat segments folding into very long perfect duplex hairpins with a loop at one end and a mass of unstructured RNA at the other. The duplex segments are highly stable, since the hairpin structures reformed rapidly during the EM mounting procedures when the RNAs were boiled and quickly cooled, even under low salt conditions.

To confirm that the ball-like appendages were the result of the non-repeat RNA present at each end of the repeat blocks, transcripts were made from *Pvu*II-digested CTG 130 plasmids using T7 RNA polymerase. These transcripts (CTG 130T) have an additional 212 nt present at the 3'-end of the molecule (Fig. 1B). As shown in Figure 1 (compare A and B), the ball-like structure at the base of the RNA rods increased in size with inclusion of the additional 212 nt. This tail of RNA frequently appeared α-shaped, but also as a ball or T-shaped appendage. RNAs were also transcribed from several plasmids containing the DMPK 3'-UTR together with different sized CUG repeat blocks. Examination of the RNA transcript containing 90 repeats revealed a duplex hairpin protruding from the non-repeat-containing RNA that appeared as a large ball (Fig. 1F). Examination of DMPK transcripts containing either no or six CUG repeats revealed ball-like structures with no evidence of hairpin formation (data not shown).

To measure the length of the hairpin segments, a dsRNA molecule 261 bp in length was prepared by *in vitro* transcription to generate complementary RNAs followed by strand annealing (45). This fully duplex RNA (Fig. 1E) and the CUG- and CAG-containing RNAs were prepared for EM. The lengths of the duplex segments for 100 molecules were measured and the results plotted as a percentage of the length of the dsRNA control (Fig. 2). The lengths of the hairpins were compared to the predicted lengths based on 2.6 Å/base pair of duplex RNA and the assumption that the repeats folded exactly in half (Table 1). These measurements demonstrated that the rod-like duplex segments are precisely the length expected for a fully duplex segment of 2.6 Å/base pair. These measurements and the lack of a significant number of kinks along their length argue that these segments consist of highly ordered duplex RNA.

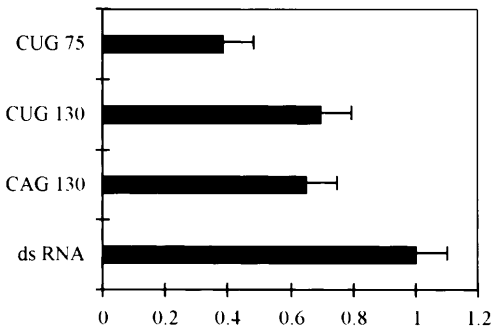

Figure 2. The length of the CUG repeat-containing RNA is consistent with the formation of hairpins. Histogram of RNA lengths plotting (n = 100) RNAs from the experiment described in Figure 1 as a percentage of the dsRNA control molecule.

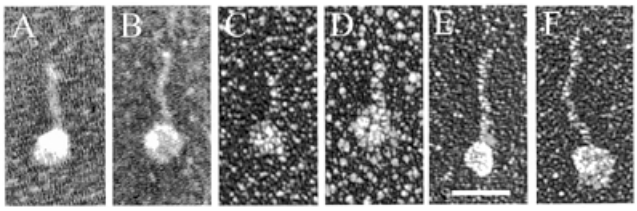

Figure 3. CUG-BP binds to the base of CUG repeat-containing hairpins and not along the stem. The protein was incubated with RNA at a 5:1 mass ratio as described in Materials and Methods. Electron micrographs (shown in reverse contrast, prepared as in Fig. 1) of complexes of CUG-BP with CUG 130 (**A** and **B**), DMPK 90 (**C** and **D**) and dsRNA (**E** and **F**). The bar represents 125 bp.

Table 1. Comparison of the predicted and experimental lengths of the RNAs from Figure 1

RNA	Predicted	Actual
dsRNA	100	100
CAG 130	72.4	65.1
CUG 130	72.4	69.6
CUG 75	41.8	38.4

The RNA hairpin is consistent with dsRNA of 2.6 Å/base pair of duplex RNA. The predicted lengths were calculated assuming the entire CUG repeat forms a hairpin, such that the predicted length of the dsRNA hairpin is equal to half the number of bases in the CUG repeat.

CUG-BP binds the base of $(CUG)_n$ hairpins

To examine the binding of CUG-BP, a known CUG-binding protein, to the RNAs characterized above, the protein was incubated with the different RNAs and prepared for EM. For these incubations, the protein:RNA ratio was varied from as low as 1 protein monomer/25 bp to as high as 1 monomer/3 bp (based on a 50 kDa molecular weight of CUG-BP) and EM was used to determine optimal levels at which there was a high frequency of protein binding without a great excess of free protein. Examples of the RNA–protein complexes are shown in Figure 3. Irrespective of the RNA used, CUG-BP bound to one end of the RNA, in contrast to binding along the length of the duplex RNA segment. To determine whether CUG-BP binds to the base of the hairpin or to the single-stranded loop at the apex of the hairpin, CUG-BP was incubated with CUG 130T RNA. In this molecule, the RNA tail is distinct (Fig. 1B). In these CUG-BP–RNA complexes, the protein was consistently observed at the base of the hairpin and obscured the RNA tail (data not shown). Binding was not observed along the duplex stem nor was protein seen at the tip of the hairpin, which would have generated a dumbbell-shaped molecule.

Analysis of EM fields of CUG-BP–RNA complexes was carried out in which molecules were scored either as being free of protein, with protein at the base of the hairpin or bound in some other fashion, including along the stem or at the top of the hairpin (Table 2). The fraction of CUG repeat RNAs with CUG-BP bound at an end was 60–70% and this percentage was insensitive to the number of CUG repeats. The binding of CUG-BP to DMPK transcripts containing either no or six CUG repeats was examined by EM. Since these transcripts do not contain hairpins, their ball-like structure is covered by the bound CUG-BP. Therefore, EM could not be used to determine accurate scores of free versus bound RNA. However, in reactions containing CUG-BP and DMPK0, the low amount of free RNA (30%) and the high amount of protein (70%, including both RNA-bound and free protein) suggests that CUG-BP binds to DMPK transcript even in the absence of CUG repeats (see also Fig. 6B). The extent of CUG-BP binding to CAG repeat RNA was similar (67 versus 70%) with binding solely at the base of the hairpin. Binding was also observed when the control dsRNA was used with CUG-BP bound to one but not both ends of 45% of these RNAs (Table 2). The dsRNA control molecule contains 9 nt of ssRNA at one end and 13 bases at the other. The binding of CUG-BP to one end of the dsRNA control suggests that this protein recognizes either only ssRNA or the junction between ssRNA and dsRNA with some sequence specificity. To further examine the binding properties of CUG-BP, both the dsRNA control and CUG 130 RNAs were treated with MBN to remove the single-stranded tails. The RNAs were purified, incubated with CUG-BP and prepared for EM. Following this treatment, the binding of CUG-BP to the MBN-treated CUG 130 RNA was reduced >4-fold (70 to 17%). Binding to the dsRNA control molecule was also reduced >6-fold (45 to 7%), suggesting that CUG-BP had a strong preference for ssRNA. The CUG repeat hairpins may still contain short single-stranded regions even following MBN treatment, since a single-stranded tail could be recreated due to slippage of the RNA structure (38). Therefore, the binding of CUG-BP to pure ds$(CUG)_n$ RNA may be <17%. The preference for one end of the dsRNA control is either due to nucleotide sequence specificity or to a difference in the length of the single-stranded regions

(9 versus 13 nt). In summary, recombinant CUG-BP bound to the base of the RNA hairpin generated by large CUG repeats and some ssRNA was required.

Table 2. Summary of the binding of CUG-BP to various RNAs as shown in Figure 3

RNA	Bound at 1 end (%)	Free (%)	Bound other (%)
CUG 130	69.4	30.6	0.0
CUG 75	70.3	29.7	0.0
DMPK 54	63.7	36.3	0.0
DMPK 90	71.9	28.1	0.0
CAG 130	66.5	33.0	0.5
dsRNA	45.1	51.9	3.0
dsRNA/MBN	7.3	90.7	2.0
CUG 130/MBN	17.2	82.3	0.0

Molecules were scored as either being free of protein, having protein bound at the base of the hairpin or other binding, which includes binding along the stem, binding to the tip or binding by more than one protein.

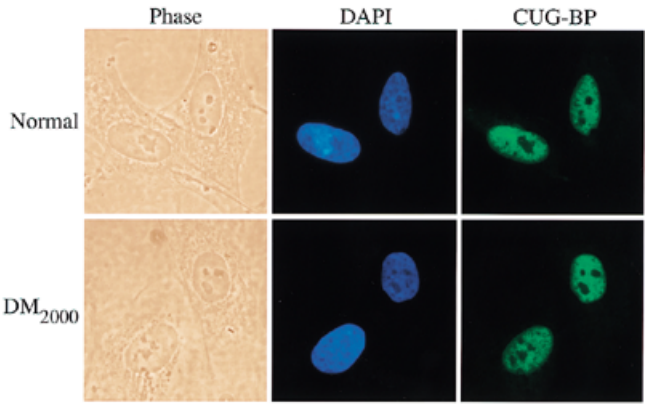

Figure 4. CUG-BP does not accumulate in intranuclear foci in DM cells. The intracellular distribution of CUG-BP in normal and DM fibroblasts was examined by indirect cell immunofluorescence using the mAb 3B1 (CUG-BP). Locations of the cells are shown using phase contrast (Phase) and chromosomal DNA by DAPI staining.

Intracellular distribution of CUG-BP is not altered in DM cells

Previous observations using fluorescence *in situ* hybridization (FISH) analysis have shown that transcripts from the DMPK mutant allele accumulate in the nucleus of DM cells in discrete foci (21,39). The sequestration hypothesis predicts that the large CUG repeats present in mutant DMPK transcripts bind and sequester CUG expansion-binding proteins (3,30,31). If CUG-BP is sequestered by the CUG repeat expansion *in vivo*, the localization of this protein may be altered in DM cells. To test this possibility, we characterized the intracellular location of CUG-BP in normal versus DM fibroblasts. In agreement with a previous report, CUG-BP was localized predominantly in the nucleus in normal fibroblasts (Fig. 4) (31). However, a significant change in this subcellular distribution pattern was not apparent in DM fibroblasts (Fig. 4). Furthermore, when compared to normal fibroblasts, no discernable difference in CUG-BP localization was observed in DM fibroblasts previously shown to accumulate mutant DMPK transcripts in nuclear foci (39; not shown). More detailed analysis using digital imaging/deconvolution microscopy failed to detect any nuclear foci enriched in CUG-BP in these DM cells (data not shown). Since a potential problem with the interpretation of this CUG-BP localization study is impaired antibody accessibility or reactivity in DM cells, we examined the RNA-binding properties of CUG-BP in more detail.

CUG-BP preferentially recognizes DMPK RNAs but binding is not proportional to CUG repeat size

The previous characterization of CUG-BP was based on its binding to ssRNA [$(CUG)_8$] oligonucleotides (31). Since a recent report has suggested that larger CUG repeats form stable dsRNA hairpin structures (38), we determined whether CUG-BP was also a dsRNA-binding protein. As described in Materials and Methods, we identified several ribosomal protein mRNAs that also contain polymorphic $(CUG)_n$ repeats. These RPL14 RNAs possess either 10 or 22 CUG repeats in the coding region. Size-matched RNAs (~650 nt) were transcribed in the presence of [^{32}P]UTP from either the DMPK 3'-UTR or the RPL14 cDNAs (Fig. 5A). These labeled transcripts were purified by denaturing acrylamide gel electrophoresis, incubated in HeLa cell nuclear extracts, irradiated with UV light to form covalent RNA–protein crosslinks, treated with RNase A and the labeled proteins visualized by SDS–PAGE and autoradiography. In agreement with previous observations (31), CUG-BP photocrosslinked to DMPK 3'-UTR RNAs (Fig. 5B). In contrast, CUG-BP did not crosslink to RPL14 transcripts with either 10 or 22 CUG repeats, while the hnRNP C proteins cross-linked to all three transcripts. Therefore, CUG-BP in nuclear extracts did not indiscriminately bind to any RNA with CUG repeats but instead preferentially recognized RNA sequence elements in the DMPK transcript. Examination of the DMPK RNA sequence used in these crosslinking experiments revealed the presence of additional CUG-containing regions in which the CUG trinucleotides were non-contiguous (Fig. 6A). We therefore tested the binding properties of CUG-BP to DMPK RNAs that varied only in the size of the CUG repeat expansion. Crosslinking of CUG-BP to DMPK RNAs without contiguous CUG repeats at the DM mutation site was detectable at a low level while RNA containing six repeats was more efficient (~2-fold, Fig. 6B). However, varying the number of CUG repeats from six to 90 only resulted in a negligible difference in CUG-BP crosslinking. This small increase was not due to limiting CUG-BP concentration in the nuclear extract, since increasing the concentration of nuclear extract or decreasing the transcript concentration had only a small effect on the relative levels of CUG-BP crosslinking (data not shown). In summary, crosslinking of CUG-BP in nuclear extracts to DMPK RNAs was affected by the presence of CUG repeats at the DM mutation site, but the extent of binding was not proportional to CUG repeat size. In conjunction with the CUG-BP cellular localization

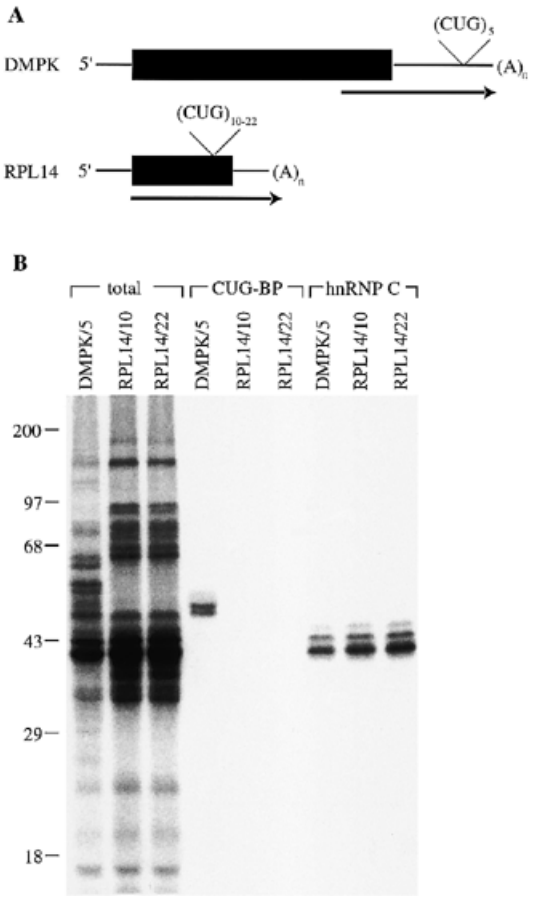

Figure 5. CUG-BP/hNab50 specifically photocrosslinks to DMPK mRNA and not to RPL14. (**A**) Structure of RPL14 mRNAs that contain either 10 or 22 CUG repeats in the coding region. The CUG repeat region [$(CUG)_{10,22}$] and the polyadenylate tail [$(A)_n$] are indicated. The arrows indicate the regions transcribed *in vitro*. (**B**) Label transfer/SDS–polyacrylamide gel analysis of HeLa cell nuclear extract proteins that photocrosslink to DMPK RNA with five CUG repeats (DMPK/5), RPL14 RNA with 10 CUG repeats (RPL14/10) and RPL14 RNA with 22 CUG repeats (RPL14/22). HeLa cell nuclear extracts were incubated with ^{32}P-labeled RNAs, photocrosslinked with UV light, treated with RNase A and either directly fractionated by SDS–PAGE (total) or following immunopurification with mAb 3B1 (CUG-BP) or mAb 4F4 (hnRNP C). Sizes are indicated in kDa.

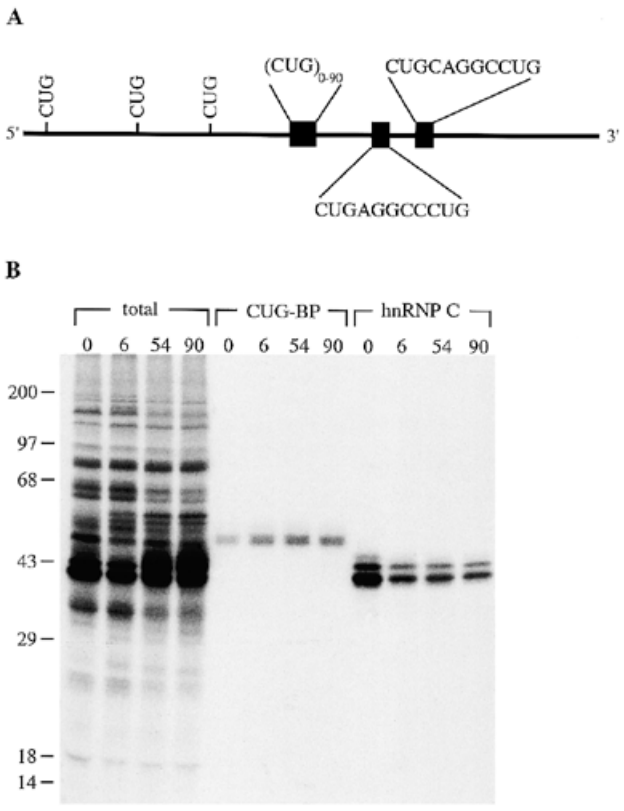

Figure 6. Binding and photocrosslinking of CUG-BP to DMPK RNAs is not proportional to $(CUG)_n$ repeat size. (**A**) CUG-rich elements in the 3'-UTR of DMPK mRNA. (**B**) Label transfer/SDS–PAGE of either total HeLa cell nuclear extract proteins (total), immunopurified CUG-BP (CUG-BP) or immunopurified hnRNP C proteins (hnRNP C). The RNAs contained either 0, 6, 54 or 90 $(CUG)_n$ repeats.

studies, which failed to show nuclear accumulation of CUG-BP in DM cells, this *in vitro* crosslinking analysis supported the idea that CUG-BP does not recognize large CUG repeat expansions.

DISCUSSION

In this study, we have tested several features of the RNA-binding protein sequestration hypothesis. First, we have shown by EM that RNAs transcribed from CTG/CAG-containing templates form hairpins. These hairpins are regular structures that do not appear to contain kinks or bubbles and their lengths are consistent with an A-form dsRNA helix. Second, since mutant transcripts accumulate in intranuclear foci in DM cells, we have studied the subcellular localization of CUG-BP. This protein does not appear to accumulate in disease-associated foci since the distribution of CUG-BP in normal and DM cells is very similar. Third, the transcript-binding properties of CUG-BP in HeLa cell nuclear extracts have been studied using photocrosslinking. This analysis confirmed the binding preference of CUG-BP for DMPK transcripts previously reported. However, CUG-BP binding was not significantly responsive to the size of the CUG repeat block. Purified CUG-BP was observed to bind to the base of the CUG-containing RNA hairpin and not along the stem and, in agreement with the photocrosslinking studies, the length of the CUG repeat does not significantly affect its binding. Cumulatively, these results demonstrate that large CUG repeats form RNA hairpins that are not recognized by CUG-BP and predict that if sequestered factors exist they may be dsRNA-binding proteins.

Upon examination of CUG-containing RNA by EM, we observed rod-like structures consistent with the formation of hairpins. Although CUG repeat-containing RNAs have yet to be shown to form hairpins *in vivo*, a double-stranded conformation is likely based on the calculated free energy (38). In addition,

efforts to examine denatured RNA by EM were unsuccessful because the hairpins reformed during the mounting procedures even under low salt conditions, suggesting that these are stable structures that are likely to be formed *in vivo*. It has been previously shown that a single base bulge in a double-stranded helix creates a bend of ~30° (50). If a single CUG triplet bulged out from the hairpin, we would expect to see a 90° bend in the duplex. The lack of bends in the RNAs examined suggests that these hairpins are paired along the entire length of the triplet repeat. The formation of long stable hairpins supports the dominant RNA mutation and protein sequestration hypotheses in that these unusual structures in the nucleus could be recognized by a particular set of proteins. From the EM data we can say conclusively that CUG-BP has a significantly reduced affinity for purely dsCUG RNA and that the binding requires a ssRNA tail. CUG-BP bound to only one end of the dsRNA control and thus we can conclude that the binding has some sequence specificity and does not simply recognize the presence of a ds/ssRNA junction. This protein binds to the base of dsCAG RNA hairpins that contain seven CUGs in the single-stranded flanking region. Since the CUG-BP protein was originally isolated because of its ability to bind to tandem CUGs present in a presumably single-stranded CUG_8 oligo, the presence of CUGs could account for the binding of CUG-BP to dsCAG RNA. We do not know why the protein binds to the single-stranded tails of the dsRNA control. The binding specificity of CUG-BP needs to be examined further.

The RNA dominant mutation model for DM pathogenesis suggests that this disease is caused by a gain-of-function at the RNA level (3,18,30). One possibility envisioned by this model is that the unusual RNA structure of the DM $(CUG)_n$ expansion interferes with the normal binding properties of a CUG-binding protein (30,31,42). If the binding of this protein is proportional to the number of CUG repeats, then large expansions would result in the accumulation of this protein on mutant allele transcripts. Aberrant binding might result in the depletion of this expansion-binding protein from other transcripts which contain CUG-rich elements essential for normal pre-mRNA processing and/or mRNA nucleocytoplasmic export. A candidate for this sequestered factor, CUG-BP, has been previously characterized and disease-associated alterations in $(CUG)_8$-binding activity documented (30,31). Large CUG repeats form stable hairpin structures (38; this report) and the sequestration hypothesis predicts that a protein or set of proteins bind to these dsRNA hairpins. However, the type of RNA-binding domain present in CUG-BP is predicted to primarily recognize ssRNA elements (reviewed in 43,44) and we have shown that CUG-BP does not bind to purely dsCUG RNA. Furthermore, a recent study provided evidence that DM cells show enhanced CUG-BP activity in the nucleus compared to normal cells, although the sequestration hypothesis predicts disease-associated loss of RNA-binding activity (40). Cumulatively, this data suggests that CUG-BP is unlikely to be the candidate factor proposed by the protein sequestration hypothesis.

Does CUG-BP play a role in DM disease pathogenesis or the biogenesis/nucleocytoplasmic export of DMPK or other mRNAs? CUG-BP is a member of the hnRNP family of RNA-binding proteins (31). The majority of hnRNPs contain RBDs that preferentially recognize ssRNA sequence elements implicated in a variety of post-transcriptional regulatory processes (reviewed in 44). For example, recent work has implicated several hnRNPs in the splicing of pre-mRNAs transcribed from muscle and neural genes (40,51). Alternative splicing of cTNT pre-mRNA exon 5 is regulated by an RNA enhancer located within the downstream intron (40). This enhancer is composed of several dispersed CUG repeats and transient overexpression of CUG-BP *in vivo* induces exon 5 inclusion. Both DM cells, as well as normal human skeletal muscle cells expressing DM minigenes with CUG repeat expansions, show constitutive increases in exon 5 inclusion, suggesting enhanced CUG-BP splicing activity. In neural cells, CUG-BP may also bind to a CUG-rich intronic enhancer to promote inclusion of clathrin light chain B exon EN and the *N*-methyl-D-aspartate receptor NR1 subunit exon 5 (51). These observations support the idea that CUG-BP plays an important regulatory role in the splicing of pre-mRNAs that possess CUG-rich RNA enhancer elements. Although we failed to detect a disease-associated difference in the overall distribution of CUG-BP in fibroblasts or myoblasts, hypophosphorylated CUG-BP isoforms are differentially localized in DM cells (42). Moreover, DMPK has been shown to directly phosphorylate CUG-BP *in vitro* (42). Therefore, loss of DMPK expression in DM cells may lead to accumulation of hypophosphorylated CUG-BP in the nucleus which subsequently alters splicing for a variety of pre-mRNAs that possess intronic or exonic CUG-rich RNA regulatory elements. Since CUG-BP does not appear to be a sequestered factor, we propose that the nuclear accumulation of hypophosphorylated CUG-BP occurs as a result of altered DMPK expression. According to this proposal, altered CUG-BP activity is not the primary cause, but an effect, of DM pathogenesis. While CUG-BP may not be a sequestered factor *in vivo*, it may play an important role in DMPK gene expression at the post-transcriptional level. The existence of the CUG direct repeat, as well as adjacent CUG-rich RNA elements, in the DMPK transcript 3'-UTR suggests that this hnRNP may function in the 3'-end processing, translation or turnover of DMPK transcripts. Future work will be focused on these potential CUG-BP functions and on the characterization of dsRNA-binding proteins that preferentially recognize $(CUG)_n$ triplet repeat expansions.

ACKNOWLEDGEMENTS

We thank M. Stenberg for recombinant CUG-BP protein preparation, P. Non for graphics assistance and members of our laboratories for comments on the manuscript. S.M. and C.R.U. were supported by NIH Training grants T32 CA09156-24 and T32 AI07110-18, respectively. This work was also supported by grants from the NIH (GM31819 and CA70343) to J.G. and an Established Investigator Award from the American Heart Association to M.S.S.

REFERENCES

1. Harper,P.S. (1989) In *Myotonic Dystrophy*, 2nd Edn. W.B. Saunders, London, UK.
2. Wieringa,B. (1994) *Hum. Mol. Genet.*, **3**, 1–7.
3. Timchenko,L., Monckton,D. and Caskey,C.T. (1995) *Semin. Cell Biol.*, **6**, 13–19.
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*, **68**, 799–808.

5. Buxton,J., Shelbourne,P., Davies,J., Jones,C., Tongeren,T., Aslandis,C., John,P., Jansen,G., Anvert,M., Riley,B., Williamson,R. and Johnson,K. (1992) *Nature*, **355**, 547–548.
6. Fu,Y.-H., Pizzuti,A., Fenwick,R.G.Jr, King,J., Rajanarayan,S., Dunne,P.W., Dubel,J., Nasser,G.A., Ashizawa,T., de Jong,P. *et al*. (1992) *Science*, **255**, 1256–1258.
7. Mahadeven,M., Tsilfidis,C., Sabourin,L., Shutler,G., Amemiya,C., Jansen,G., Neville,C., Narang,M., Barcelo,J., O'Hoy,K. *et al*. (1992) *Science*, **255**, 1253–1255.
8. Harley,H., Rundle,S., MacMillan,J., Myring,J., Brook,J., Crow,S., Reardon,W., Fenton,I., Shaw,D. and Harper,P. (1993) *Am. J. Hum. Genet.*, **52**, 1164–1174.
9. Korade-Mirnics,Z., Babitzke,P. and Hoffman,E. (1998) *Nucleic Acids Res.*, **26**, 1363–1368.
10. Fu,Y.-H., Friedman,D.L., Richards,S., Pearlman,J.A., Gibbs,R.A., Pizzuti,A., Ashizawa,T., Perryman,H.B., Scarleto,G., Fenwick,R.G. and Caskey,C.T. (1993) *Science*, **260**, 235–238.
11. Carango,P., Nobler,J.E., Marks,H.G. and Funanage,V.L. (1993) *Genomics*, **18**, 340–348.
12. Reddy,S., Smith,D., Rich,M., Leferovich,J., Reilly,P., Davis,B., Tran,K., Rayborn,H., Bronson,R., Cros,D., Balice-Gordon,R. and Housman,D. (1996) *Nature Genet.*, **13**, 325–335.
13. Jansen,G., Groenen,P.J.T.A., Bächner,D., Jap,P.H.K., Coerwinkel,M., Oerlemans,F., van den Broek,W., Gohlsch,B., Pette,D., Plomp,J.J. *et al*. (1996) *Nature Genet.*, **13**, 316–324.
14. Lugenbeel,K.A., Peier,A.M., Carson,N.L., Chudley,A.E. and Nelson,D.L. (1995) *Nature Genet.*, **10**, 483–485.
15. Hofmann-Radvanyi,H., Lavedan,C., Rabes,S., Savoy,D., Duros,C., Johnson,K. and Junien,C. (1993) *Hum. Mol. Genet.*, **2**, 1263–1266.
16. Novelli,G., Gennarelli,M., Zelano,G., Pizzuti,A., Fattorini,C., Caskey,C. and Dallapiccola,B. (1993) *Biochem. Mol. Biol. Int.*, **29**, 291–297.
17. Wang,Y.H., Amirhaeri,S., Kang,S., Well,R.D. and Griffith,J.D. (1994) *Science*, **265**, 669–671.
18. Wang,Y.-H. and Griffith,J. (1995) *Genomics*, **25**, 570–573.
19. Wolffe,A.P. (1999) *Nucleic Acids Res.*, **27**, 711–720.
20. Otten,A. and Tapscott,S. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 5465–5469.
21. Davis,B., McCurrach,M., Taneja,K., Singer,R. and Housman,D. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 7388–7393.
22. Hamshere,M.G., Newman,E.E., Alwazzan,M., Athwal,B.S. and Brook,J.D. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 7394–7399.
23. Shaw,D.S., McCurrach,M., Rundle,S.A., Harley,H.G., Crow,S.R., Sohn,R., Thirion,J.-P., Hamshere,M.G., Buckler,A.J., Harper,P.S. *et al*. (1993) *Genomics*, **18**, 673–679.
24. Jansen,G., Mahadeven,M., Amemiya,C., Wormskamp,N., Segers,B., Hendricks,W., O'Hoy,K., Bairel,S., Sabourin,L., Lennon,G. *et al*. (1992) *Nature Genet.*, **1**, 261–266.
25. Harris,S., Moncrieff,C. and Johnson,K. (1996) *Hum. Mol. Genet.*, **5**, 1417–1423.
26. Johnson,K.J., Boucher,C.A., King,S.K., Winchester,C.L., Bailey,M.E., Hamilton,G.M. and Carey,N. (1996) *Biochem. Soc. Trans.*, **24**, 510–513.
27. Klessert,T., Otten,A., Bird,T. and Tapscott,S. (1997) *Nature Genet.*, **16**, 402–406.
28. Thorton,C., Wyner,J., Simmons,Z., McClain,C. and Moxley,R. (1997) *Nature Genet.*, **16**, 407–409.
29. Wang,J., Pegoraro,E., Menegazzo,E., Gennarelli,M., Hoop,R.C., Angelini,C. and Hoffman,E.P. (1995) *Hum. Mol. Genet.*, **4**, 599–606.
30. Caskey,C.T., Swanson,M.S. and Timchenko,L.T. (1996) *Cold Spring Harbor Symp. Quant. Biol.*, **61**, 607–614.
31. 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) *Nucleic Acids Res.*, **24**, 4407–4414.
32. Gacy,A.M., Goellner,G., Juranic,N., Macura,S. and McMurray,C.T. (1995) *Cell*, **81**, 533–540.
33. Petraska,J., Arnheim,N. and Goodman,M.F. (1996) *Nucleic Acids Res.*, **24**, 1992–1998.
34. Zheng,M., Huang,X., Smith,G.K., Yang,X. and Gao,X. (1996) *J. Mol. Biol.*, **264**, 323–336.
35. Wells,R. (1996) *J. Biol. Chem.*, **271**, 2875–2878.
36. Baeyans,K.J., DeBondt,H.L. and Holbrook,S.R. (1995) *Nature Struct. Biol.*, **2**, 56–62.
37. Koch,K.S. and Leffert,H.L. (1998) *J. Theor. Biol.*, **192**, 505–514.
38. Napierala,M. and Krzyzosiak,W.J. (1997) *J. Biol. Chem.*, **272**, 31079–31085.
39. Taneja,J.L., McCurrach,M., Schalling,M., Housman,D. and Singer,R.H. (1995) *J. Cell Biol.*, **128**, 995–1002.
40. Philips,A.V., Timchenko,L.T. and Cooper,T.A. (1998) *Science*, **280**, 737–741.
41. Timchenko,L.T., Timchenko,N.A., Caskey,C.T. and Roberts,R. (1996) *Hum. Mol. Genet.*, **5**, 115–121.
42. Roberts,R., Timchenko,N.A., Miller,J.W., Reddy,S., Caskey,C.T., Swanson,M.S. and Timchenko,L.T. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 13221–13226.
43. Burd,C.G. and Dreyfuss,G. (1994) *Science*, **265**, 615–621.
44. Krecic,A.M. and Swanson,M.S. (1999) *Curr. Opin. Cell Biol.*, **11**, 363–371.
45. Wang,Y.-H., Howard,M.T. and Griffith,J.D. (1991) *Biochemistry*, **30**, 5443–5449.
46. Chan,Y.L., Olvera,J. and Wool,I.G. (1996) *Biochem. Biophys. Res. Commun.*, **222**, 427–431.
47. Aoki,M., Koranyi.L., Riggs,A.C., Wasson,J., Chiu,K.C., Vaxillaire,M., Froguel,P., Gough,S., Liu,L., Donis-Keller,H. and Permutt,M.A. (1996) *Diabetes*, **45**, 157–164.
48. Griffith,J.D. and Christensen,G. (1978) *Annu. Rev. Biophys. Bioeng.*, **7**, 19–35.
49. Wang,Y.-H., Murphy,F., Cech,T. and Griffith,J.D. (1994) *J. Mol. Biol.* **236**, 64–71.
50. Wang,Y.-H., Barker,P., Sancar,A. and Griffith,J. (1992) *J. Biol. Chem.*, **267**, 4911–4915.
51. Zhang,L., Liu,W. and Grabowski,P.J. (1999) *RNA*, **5**, 117–130.