Nucleic Acids Research, 1996, Vol. 24, No. 22 **4407–4414**

Identification of a $(CUG)_n$ triplet repeat RNA-binding protein and its expression in myotonic dystrophy

Lubov T. Timchenko, Jill W. Miller[1], Nikolaj A. Timchenko[2], Druery R. DeVore[1], Kshama V. Datar[1], Lijun Lin, Robert Roberts, C. Thomas Caskey[3] and Maurice S. Swanson[1,*]

Department of Medicine, Section of Cardiology, Baylor College of Medicine, Houston, TX 77030, USA, [1]Department of Molecular Genetics and Microbiology and Center for Mammalian Genetics, University of Florida College of Medicine, Gainesville, FL 32610, USA, [2]Department of Pathology, Baylor College of Medicine, Houston, TX 77030, USA and [3]Merck Research Laboratories, Sumneytown Pike, West Point, PA 19486, USA

Received August 26, 1996; Revised and Accepted October 7, 1996

DDBJ/EMBL/GenBank accession no. U63289

ABSTRACT

Myotonic dystrophy (DM) is an autosomal dominant neuromuscular disease that is associated with a $(CTG)_n$ repeat expansion in the 3′-untranslated region of the myotonin protein kinase (Mt-PK) gene. This study reports the isolation and characterization of a $(CUG)_n$ triplet repeat pre-mRNA/mRNA binding protein that may play an important role in DM pathogenesis. Two HeLa cell proteins, CUG-BP1 and CUG-BP2, have been purified based upon their ability to bind specifically to $(CUG)_8$ oligonucleotides *in vitro*. While CUG-BP1 is the major $(CUG)_8$-binding activity in normal cells, nuclear CUG-BP2 binding activity increases in DM cells. Both CUG-BP1 and CUG-BP2 have been identified as isoforms of a novel heterogeneous nuclear ribonucleoprotein (hnRNP), hNab50. The CUG-BP/hNab50 protein is localized predominantly in the nucleus and is associated with polyadenylated RNAs *in vivo*. *In vitro* RNA-binding/photocrosslinking studies demonstrate that CUG-BP/hNab50 binds to RNAs containing the Mt-PK 3′-UTR. We propose that the $(CUG)_n$ repeat region in Mt-PK mRNA is a binding site for CUG-BP/hNab50 *in vivo*, and triplet repeat expansion leads to sequestration of this hnRNP on mutant Mt-PK transcripts.

INTRODUCTION

Myotonic dystrophy (dystrophia myotonica, DM) is the most common form of adult onset muscular dystrophy (1). The clinical DM phenotype is highly variable both within and between families, and is characterized by muscle weakness and myotonia in skeletal muscle, dilated cardiomyopathy, and a number of extramuscular abnormalities including cataracts and reduced cognitive function. Myotonic dystrophy is an autosomal dominant disorder that shows genetic anticipation in which successive generations show earlier onset and increasing disease severity. The gene affected in DM has been mapped to chromosome 19q13.3, and encodes myotonin protein kinase (Mt-PK) (2–10). Nearly all affected DM individuals possess a $(CTG)_n$ triplet repeat expansion in a region of the Mt-PK gene that corresponds to the 3′-untranslated region (3′-UTR) of the mRNA. The number of $(CTG)_n$ repeats is variable ranging from 5 to 37 triplet repeats in normal cells to >700 repeats in the severe congenital form of the disease (CDM). This triplet repeat expansion also shows somatic mosaicism with repeat size variable in different tissues of affected individuals and a general increase in expansion length with age (11–13).

The effect of the $(CTG)_n$ triplet repeat expansion on expression of the Mt-PK gene was unclear until recently. Early evidence indicated both decreased expression of Mt-PK mRNA and protein in adult DM and increased expression in cells from a single CDM patient (9,14–17). However, more recent reports have consistently demonstrated that the DM expansion mutation leads to decreased expression in both adult onset DM and CDM with some CDM individuals showing almost undetectable Mt-PK mRNA and protein levels (18–20) (Timchenko *et al*., manuscript in preparation). This decreased expression has been suggested to be the result of a defect in the processing of Mt-PK pre-mRNA since the levels of mRNA are more severely affected than pre-mRNA (16,19,20). In support of the idea that loss of Mt-PK gene expression may lead to muscle disease in mammals, Mt-PK knock-out mice show a late onset myopathy in homozygotic (*Dmpk*–/–), but not heterozygotic, mutants (21). However, the degree of myopathy seen in *Dmpk*–/– mice is variable between studies, and loss of Mt-PK expression is not associated with any of the other phenotypes commonly seen in human patients including myotonia and cataracts (22). These studies support the hypothesis that the expansion mutation is exerting the dominant effect and disease is not simply the result of a loss of Mt-PK gene expression. Although the possible role of the $(CTG)_n$ expansion mutation in mediating loss of Mt-PK gene expression has remained a mystery, the corresponding $(CUG)_n$ RNA repeat might be a binding site for a pre-mRNA/mRNA-binding protein

* To whom correspondence should be addressed. Tel: +1 352 392 3082; Fax: +1 352 392 5914; Email: swanson@college.med.ufl.edu

that is required for the biogenesis, nucleocytoplasmic transport and/or translation of Mt-PK mRNA. Here we describe the isolation and characterization of a novel heterogeneous nuclear ribonucleoprotein (hnRNP) that is identical to the recently described $(CUG)_8$-binding protein (CUG-BP) (23), and demonstrate that this hnRNP binds to RNAs containing the 3′-UTR of Mt-PK mRNA *in vitro*.

MATERIALS AND METHODS

Purification of CUG-BP and bandshift analysis

Whole cell extracts were prepared as described previously (23). To prepare the cytoplasmic fraction, cells were washed twice with phosphate-buffered saline (PBS), pelleted and resuspended in buffer A (10 mM Tris–HCl, pH 7.6, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT). After a 15 min incubation on ice, cells were homogenized by passage through a 23 g needle (eight strokes), and the sample was centrifuged for 5 min at 10 000 r.p.m. (Sorvall Microspin 12S) to pellet nuclei which were used for the preparation of nuclear extracts (NE) as described previously (23). The supernatant (cytoplasmic fraction) was collected and stored at –80°C. For CUG-BP purification, HeLa cells (200 plates, 15 cm, 50% confluency) were grown in MM medium (23), and cytoplasmic proteins were fractionated by the denaturation/elution technique (24). Briefly, proteins were resolved by SDS–PAGE (12% gels) and transferred to a nitrocellulose filter. The fraction containing proteins in the 40–50 kDa range (p46) was isolated by cutting a region of the membrane near the position of the 46 kDa ovalbumin marker, and proteins were eluted in 100 µl of renaturation buffer (20 mM HEPES, pH 7.9, 50 mM KCl, 5 mM $MgCl_2$, 10% glycerol, 1% Triton X-100, 0.1 mg/ml Fraction V BSA) at 4°C. The p46 fraction, which contained 15–20 proteins as determined by SDS–PAGE and silver staining, was subjected to DEAE chromatography, and the proteins were eluted with a NaCl step gradient (0.1–1.0 M NaCl). The polypeptide composition of the DEAE-flowthrough fraction containing CUG-BP2 was analyzed by SDS–PAGE and found to contain a single 51 kDa polypeptide. The CUG-BP1 protein was eluted by 0.2–0.3 M NaCl and further purified by FPLC MonoQ chromatography using a linear salt gradient (0.02–1.0 M NaCl). The CUG-BP1 binding activity was present in fractions 18–26 with maximal activity in fraction 22 which contained a major 49 kDa polypeptide as determined by SDS–PAGE and silver staining. Each protein fraction was analyzed by bandshift analysis with the $(CUG)_8$ probe and polyacrylamide gel electrophoresis.

Binding reactions for the bandshift assay were performed at room temperature for 30 min in a 10 µl reaction mixture containing 0.1–0.5 ng of ^{32}P-labeled DNA or RNA probe, 5–10 µg NE or 20–30 µg cytoplasmic extract, 2 µg of poly(dI–dC), 20 mM Tris–HCl, pH 7.6, 100 mM KCl, 5 mM $MgCl_2$, 5 mM DTT and 10% glycerol. When the $(CUG)_8$ and $(CGG)_8$ RNA probes were used in the binding reactions, 2 µg of total HeLa cell RNA was also added as a non-specific competitor. For single-stranded DNA probes, poly(dI–dC) was heated to 95°C for 10 min followed by incubation on ice prior to addition to the binding reaction. To determine binding specificity, a 100-fold molar excess of unlabeled $(CUG)_8$ or $(CGG)_8$ was added to the binding reaction prior to addition of the labeled probe. The ss$(CTG)_8$ DNA, ss$(CUG)_8$ RNA and ss$(CGG)_8$ RNA oligonucleotides, and the ds(CTG) DNA fragment, were synthesized and end-labeled as described (23). For supershift experiments, 0.5–3 µg of affinity-purified mAb 3B1, specific for CUG-BP/hNab50, was added to the binding reaction and incubated at room temperature for 15 min prior to addition of the ^{32}P-labeled probe. Protein–DNA and protein–RNA complexes were separated from free probe by polyacrylamide gel electrophoresis as described (23). For bandshift analysis of DM lymphoblast cytoplasmic and nuclear extracts (Fig. 2) the following cell lines were obtained from the NIGMS Human Genetic Mutant Cell Repository: DM1 (GM03986A), DM2 (GM03756), DM3 (GM03990A). Normal lymphoblasts (HH) were obtained from the Baylor College of Medicine Tissue Culture Core Facility,

Isolation of hNab50 cDNAs

The hNab50 protein was isolated using a yeast two hybrid interaction system (Clontech). The yeast strain HF7c was transformed with pNAB2.GBT9 and fusion protein expression confirmed by immunoblot analysis using the anti-Nab2p mAb 3F2 (25). Cells expressing the Gal4pBD-Nab2p fusion protein were subsequently transformed with a HeLa cell cDNA library cloned into the pGADGH plasmid. Cells were selected on SD-Leu-Trp-His plates, and clones initially tested for β-galactosidase activity using a plate assay following a protocol provided by the manufacturer. Cells that were dark blue by the plate assay were subsequently tested for β-galactosidase activity using a quantitative liquid assay. Three cDNA clones encoding the full-length CUG-BP/hNab50 protein were isolated from both human osteosarcoma and HeLa cell libraries by hybridization with the pGADGH-hNab50 clone. DNA sequences for all cDNA clones were determined and analyzed as described (25,26).

Monoclonal antibody preparation, immunoblotting and indirect cell immunofluorescence

For the preparation of anti-hNab50 polyclonal antisera, BALB/c mice were injected with an hNab50-maltose-binding protein (hNab50-MBP) fusion protein which was prepared by expression of the pMAL50.1 plasmid in *Escherichia coli* TB1 cells followed by amylose resin affinity chromatography (New England Biolabs). The pMAL50.1 plasmid was constructed by cloning a partial hNab50 cDNA clone (encoding amino acids 44–482) behind the *malE* gene. Antisera were tested by immunoblot analysis using both purified hNab50-MBP protein as well as HeLa whole cell lysates. The mAb 3B1 was prepared by fusing spleen cells from the best responding mouse with SP2/O cells, hybridoma supernatants were screened by immunoblotting and cell immunofluorescence, and positive hybridomas cloned as described (25). Affinity purified mAb 3B1 was prepared by growing hybridoma cells in Dulbecco's modified Eagle's medium (DMEM) containing 10% ultralow IgG FBS (Gibco-BRL) followed by affinity purification on Protein G–Sepharose. For immunoblot analysis of CUG-BP1 and CUG-BP2, proteins were fractionated on 12% SDS–PAGE gels and electroblotted onto nitrocellulose. The membrane was blocked with 10% milk in PBS for 1 h at room temperature, and subsequently incubated with the following monoclonal antibodies and dilutions: 3B1 (1:500); 4B10 (1:2000) (specific for the hnRNP A1 protein); 4F4 (1:2000) (specific for the hnRNP C proteins). After incubation for 1 h at room temperature, membranes were incubated for 1 h with a sheep anti-mouse secondary antibody conjugated with horseradish peroxidase and washed. All other immunoblot analyses were performed as described (25,26), and proteins were detected by

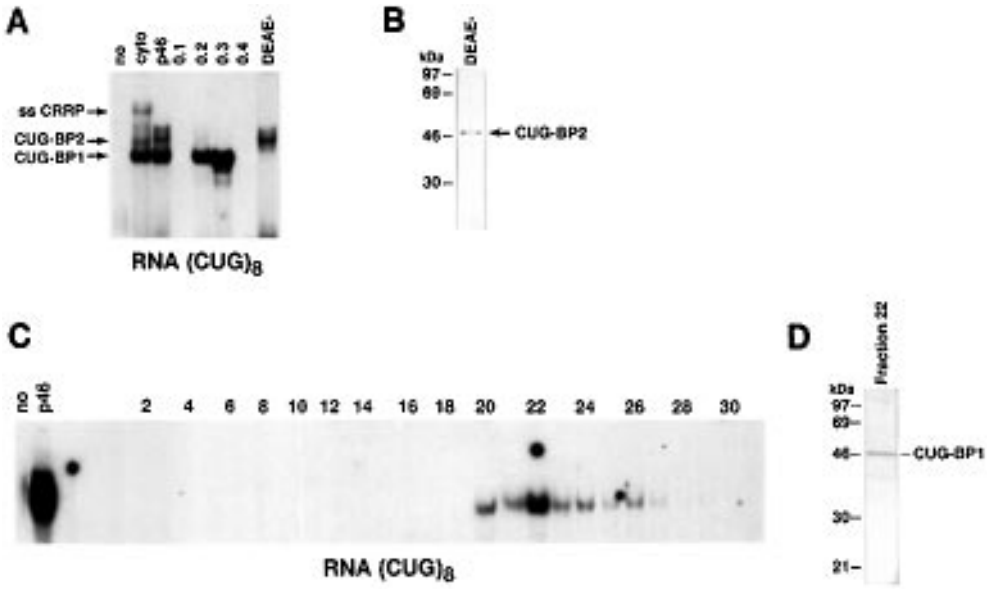

Figure 1. Purification of $(CUG)_8$ RNA-binding proteins. (**A**) The CUG-BP1 and CUG-BP2 RNA-binding activities were purified by DEAE chromatography and detected by bandshift analysis. Proteins tested by bandshift analysis with $(CUG)_8$ included no protein (no), a cytoplasmic fraction (cyto), the p46 fraction, fractions eluted from a DEAE–Sepharose column with 0.1–0.4 M NaCl (0.1, 0.2, 0.3, 0.4), and the DEAE-flowthrough (DEAE–). The single-strand CTG-repeat recognizing protein (ssCRRP) binds to both single-stranded $(CTG)_8$ DNA and $(CUG)_8$ RNA and is localized in the cytoplasm. Unbound $(CUG)_n$ probe migrates near the bottom of the gel and is not shown. (**B**) Characterization of purified CUG-BP2. The protein composition of the DEAE column flowthrough fraction, which contained purified CUG-BP2, was characterized by SDS–PAGE and silver staining. Sizes are indicated in kilodaltons (kDa). (**C**) Purification of CUG-BP1 by MonoQ chromatography and bandshift analysis. The numbers refer to chromatographic fractions from the MonoQ column. Also shown is a bandshift analysis with no protein (no) and the p46 fraction (p46). (**D**) Characterization of purified CUG-BP1. SDS–PAGE and silver staining of fraction 22 from the MonoQ column containing the purified CUG-BP1 protein.

ECL (Amersham). Indirect cellular immunofluorescence was performed essentially as described previously using a 1:500 dilution for mAbs 3B1 and 1D8 (specific for the hnRNP M proteins) (26,27) and either HeLa, Hep2, A549 or normal patient myoblast cell lines.

Photocrosslinking and immunopurification of CUG-BP/hNab50 and hnRNP complexes

For RNA–protein photocrosslinking *in vivo*, HeLa S3 cells were grown in DMEM supplemented with 10% calf serum and 1% penicillin–streptomycin to subconfluent densities. Cells were washed with ice-cold PBS and irradiated with UV light (Stratalinker, Stratagene) for 5 min in 5 ml of PBS at 4°C. Polyadenylated RNPs and immunopurified hnRNP complexes were isolated as described previously (27,28). For *in vitro* RNA binding studies, plasmids containing the 3′-UTR regions of the Mt-PK (MTPK.2, linearized with *Hin*dIII) and actin [pSP6γ-actin (29), linearized with *Bam*HI] genes were transcribed *in vitro* in the presence of [^{32}P]UTP, and purified by denaturing gel electrophoresis. The MTPK.2 plasmid was constructed by subcloning a *Bam*HI–*Hin*dIII fragment (nt 2212–2849, DDBJ/EMBL/GenBank accession no. M87312) into pSP72 (Promega). Following incubation of the ~600 nt labeled actin and Mt-PK RNAs (2.0×10^5 c.p.m.) in a 25 µl reaction mix (11 µl HeLa cell nuclear extract, 20 mM HEPES, pH 7.6, 1.3 mM $MgCl_2$, 1.5 mM ATP, 20 mM creatine phosphate) at 30°C for 10 min, 5 µg tRNA were added, samples were exposed to UV light (Stratalinker) for 5 min and RNAs were digested with 2.5 µg RNase A (30 min at 37°C). Both total and immunopurified proteins photocrosslinked to RNAs were detected by label transfer/autoradiography following SDS–PAGE. Total protein samples fractionated by SDS–PAGE corresponded to 7.5 µl of the initial 25 µl reaction. Since the hnRNP C proteins crosslink more efficiently than other hnRNPs, the amount of the crosslinked reaction volume used for immunopurification varied from 25 µl (for mAb 4F4) to 190 µl (mAb 3B1). Immunopurifications were performed at 4°C for 20 min essentially as described previously (28) except that Protein G–Sepharose was used and crosslinked samples were treated at 100°C in 1% SDS prior to dilution in PBS containing 1 mM EDTA, 1% Triton X-100, 0.5% deoxycholic acid, 0.1% SDS, 0.5% aprotinin.

RESULTS

Isolation of $(CUG)_n$ triplet repeat RNA-binding proteins

The CUG-BP activity was originally identified by characterizing HeLa cell proteins that preferentially interact with a synthetic $(CUG)_8$ RNA oligonucleotide using an electrophoretic mobility bandshift assay (23). The molecular weight of CUG-BP was estimated to be 40–50 kDa using a denaturation/elution technique (DET) (24), and $(CUG)_8$-binding activity was found predominantly in the cytoplasmic fraction. To further characterize CUG-BP, the 40–50 kDa fraction from the DET procedure (p46) was fractionated by DEAE–Sepharose chromatography using an NaCl step gradient. Each protein fraction was analyzed by the bandshift assay using a ^{32}P-labeled $(CUG)_8$ RNA oligonucleotide probe (Fig. 1A). Two different RNA–protein complexes, CUG-BP1 and CUG-BP2, were identified. The CUG-BP1 activity eluted between 0.2 and 0.3 M NaCl while CUG-BP2 was in the flowthrough fraction (Fig. 1A, DEAE–). The CUG-BP2

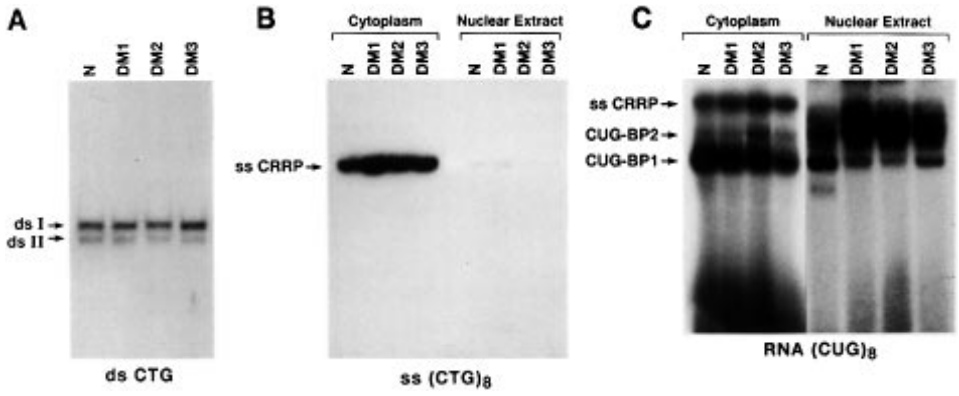

Figure 2. Alteration of nuclear $(CUG)_8$-binding activity in DM cells. (**A**) Bandshift analysis of the DNA binding activities of the dsCTG-binding proteins in normal (N) and three different DM cell (DM1, DM2, DM3) extracts. The positions of the two dsCTG bands are indicated (ds I, ds II). (**B**) Bandshift analysis of the ssCRRP binding activity using both cytoplasmic and nuclear extracts from cells described in (A). (**C**) Bandshift analysis of the $(CUG)_8$ binding activity of CUG-BP1 and CUG-BP2 using the extracts described in (B). The positions of ssCRRP, CUG-BP1 and CUG-BP2 are indicated. Unbound $(CUG)_8$ probe is shown near the bottom of the figure.

binding activity was not detected previously since it was less abundant and masked by the predominant CUG-BP1 activity (23). To characterize the specificity of CUG-BP1 and CUG-BP2, a bandshift competition analysis was performed. Addition of a 100-fold excess of unlabeled $(CUG)_8$, but not a $(CGG)_8$ RNA oligonucleotide, abolished the binding of both CUG-BP1 and CUG-BP2 to the labeled $(CUG)_8$ probe (data not shown). The polypeptide composition of the DEAE-flowthrough fraction containing CUG-BP2 was analyzed by SDS–PAGE and found to contain predominantly a single 51 kDa polypeptide (Fig. 1B). A similar analysis of the CUG-BP1 fraction eluted from DEAE–Sepharose with 0.3 M NaCl indicated that it was composed of multiple proteins, and therefore it was further fractionated by MonoQ FPLC chromatography (Fig. 1C). The CUG-BP1 binding activity was present in fractions 18–26 with maximal activity in fraction 22 which contained a major 49 kDa polypeptide (Fig. 1D). Bandshift analysis using the $(CUG)_8$ RNA probe demonstrated that these highly purified CUG-BP1 and CUG-BP2 proteins form RNA–protein complexes with similar electrophoretic mobility shift patterns as the proteins present in the p46 fraction (Fig. 3).

RNA-binding activity of CUG-BP1 and CUG-BP2 is altered in DM cells

Numerous reports have documented a correlation between the size of the $(CTG)_n$ repeat expansion and expression of the Mt-PK gene at both the mRNA and protein levels (9,14–20). The isolation of $(CUG)_8$ RNA-binding proteins permitted a direct test of the hypothesis that these proteins might be involved in the regulation of Mt-PK gene expression and DM pathogenesis (23). We first analyzed the RNA-binding activities of CUG-BP1 and CUG-BP2 by bandshift analysis using nuclear and cytoplasmic extracts derived from lymphoblast cell lines obtained from both normal and DM patients. A previous study identified a single-stranded (ss)CTG-repeat recognizing protein (ssCRRP) that was localized in the cytoplasmic fraction (23). We therefore also assayed binding actitivites for ssCTG and double-stranded (ds)CTG repeats. The dsCTG binding activity was similar in both normal and three different DM lymphoblast lines (Fig. 2A). In

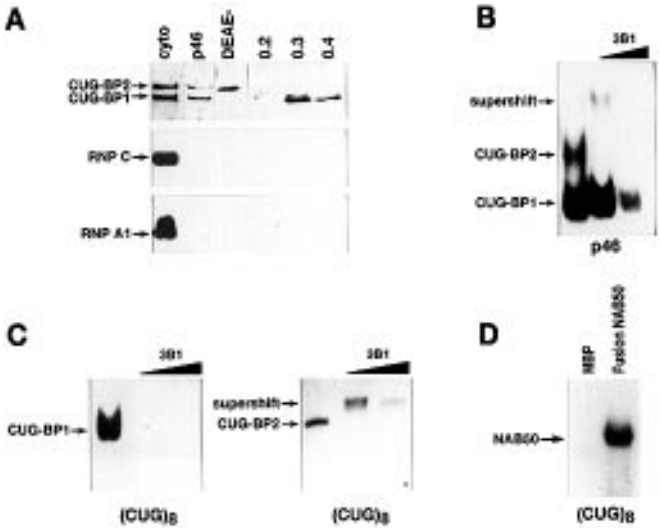

Figure 3. CUG-BP1 and CUG-BP2 are identical to the hNab50 protein. (**A**) Immunoblot analysis of proteins from HeLa cytoplasmic extracts (cyto), the p46 fraction (p46) and the flowthrough of various salt (0.2, 0.3, 0.4) fractions from the DEAE–Sepharose column described in Figure 1A. The hNab50 proteins were detected with mAb 3B1, the hnRNP C proteins (C1 and C2 comigrate under these conditions) with mAb 4F4, and the hnRNP A1 protein with mAb 4B10. (**B**) A monoclonal antibody to hNab50 supershifts/neutralizes both the CUG-BP1 and CUG-BP2 complexes present in the p46 fraction. Two different concentrations of mAb 3B1 (1 and 3 µg) were used. (**C**) Purified CUG-BP1 and CUG-BP2 are supershifted/neutralized by the mAb 3B1. (**D**) Recombinant hNab50 (Fusion NAB50), but not MBP alone, bandshifts the $(CUG)_8$ probe.

agreement with previous results using HeLa cell extracts, ssCRRP was localized in the cytoplasmic fraction and showed only a slight increase in binding to ss$(CTG)_8$ in DM cells (Fig. 2B). In contrast, CUG-BP1 and CUG-BP2 were distributed in both the cytoplasmic and nuclear fractions (Fig. 2C). In normal lymphoblasts, the majority of the $(CUG)_8$ binding activity of CUG-BP1 was in the cytoplasmic fraction, but significant activity was also present in the nucleus while CUG-BP2 was predominantly nuclear. Although the $(CUG)_8$ binding activity of CUG-BP1

declined in both the cytoplasmic and nuclear fractions in DM lymphoblasts, there was a consistent ~2-fold increase of binding activity associated with CUG-BP2 in nuclear extracts from DM cells. This increase in nuclear CUG-BP2 activity was confirmed using four additional lymphoblast, as well as two myoblast, cell lines (data not shown). These results demonstrated that the CUG-BP proteins were distributed in both nuclear and cytoplasmic fractions, and that nuclear CUG-BP1 and CUG-BP2 activities were altered in DM cells.

Identification of hNab50 as the CUG repeat RNA-binding protein

Although CUG-BP1 and CUG-BP2 specifically bound CUG repeats *in vitro*, no evidence existed that these proteins were associated with mRNAs *in vivo*. Proteins that directly bind to pre-mRNAs and mRNAs in the nucleus are heterogeneous nuclear ribonucleoproteins (hnRNPs) while cytoplasmic mRNA binding proteins are mRNPs (30,31). However, some hnRNPs shuttle between the nucleus and cytoplasm, and subcellular fractionation invariably results in the presence of hnRNPs in the cytoplasmic fraction (32). Since the CUG-BP proteins were present in both nuclear and cytoplasmic fractions, we determined whether CUG-BP1 or CUG-BP2 were previously identified nuclear RNA-binding proteins by immunoblot analysis using monoclonal antibodies against several human hnRNPs. Remarkably, a monoclonal antibody (mAb) to the hNab50 protein, mAb 3B1, reacted against both CUG-BP1 and CUG-BP2 in cytoplasmic and partially purified p46 fractions (Fig. 3A). Fractionation of the two different CUG-BP activities by DEAE–Sepharose chromatography and subsequent immunoblot analysis indicated that both CUG-BP1 and CUG-BP2 were specifically recognized by mAb 3B1. In contrast, monoclonal antibodies against two other abundant hnRNPs, the hnRNP A1 and C proteins, detected the corresponding proteins only in the cytoplasmic fraction (Fig. 3A).

The results described above suggested that different isoforms of hNab50 might be responsible for both the CUG-BP1 and CUG-BP2 $(CUG)_8$ RNA-binding activities. To test this possibility, we investigated the effect of affinity-purified mAb 3B1 on $(CUG)_8$ binding *in vitro* by bandshift/supershift analysis. The p46 fraction containing both CUG-BP1 and CUG-BP2 was incubated with the $(CUG)_8$ probe in the presence of increasing amounts of mAb 3B1. Addition of low amounts of mAb 3B1 (1 µg) to the binding reaction resulted in the complete disappearance of CUG-BP2 and the appearance of a new supershifted band (Fig. 3B). Higher antibody concentrations (>3 µg) were required to neutralize formation of the CUG-BP1 complex. Bandshift/supershift experiments with purified CUG-BP1 and CUG-BP2 proteins yielded identical results (Fig. 3C). To determine if recombinant hNab50 could bind specifically to the $(CUG)_8$ probe, a hNab50 cDNA clone (see Materials and Methods) was fused in-frame with the *E.coli* maltose-binding protein (MBP) *malE* gene to form pMAL50.1. The pMAL50.1 fusion protein was synthesized in *E.coli*, purified to homogeneity by affinity chromatography, and assayed for RNA binding by bandshift analysis. The MBP-hNab50 protein bound to the $(CUG)_8$ probe and formed a shifted complex (Fig. 3D). This interaction was specific since addition of a 100-fold excess of unlabeled $(CUG)_8$, but not $(CGG)_8$, abolished binding (data not shown). No detectable binding of the MBP alone was observed (Fig. 3D). We conclude that CUG-BP1 and CUG-BP2 are isoforms of hNab50.

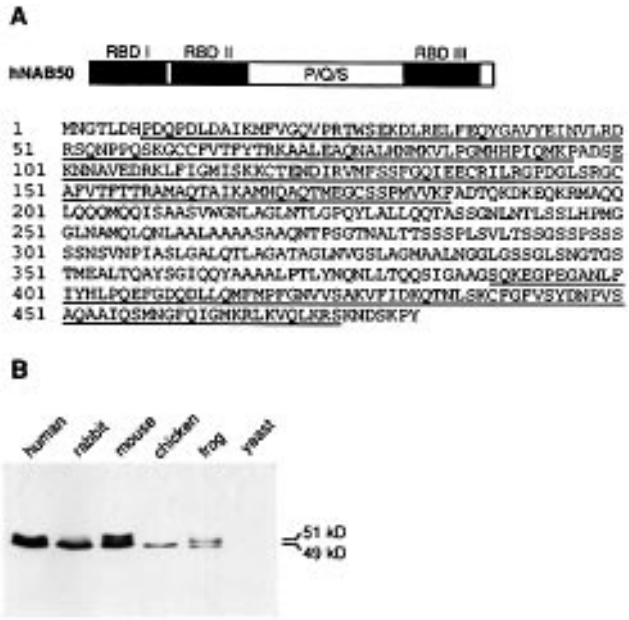

Figure 4. Characterization of hNab50. (**A**) Structure and deduced amino acid sequence of CUG-BP/hNab50. The three RBDs are indicated by black boxes or underlined in the sequence. The region between RBD II and RBD III, which is rich in proline, glutamine and serine (P/Q/S), is indicated. (**B**) Immunoblot analysis using mAb 3B1 against hNab50. Total cellular proteins were isolated from either human (HeLa), rabbit (RK13), mouse (3T3), chicken (CEF), frog (XL1) or budding yeast (BJ926) cells grown in culture.

CUG-BP/hNab50 is a novel human pre-mRNA/mRNA-binding protein

The hNab50 protein was identified during studies on the function of a yeast hnRNP, Nab2p (25). To elucidate conserved pathways in which Nab2p might play an important role, we sought to identify human proteins that interact with yeast Nab2p *in vivo* using the two hybrid interaction system. Several human cDNAs were isolated from a HeLa cell cDNA-activation domain library. A full-length cDNA clone encoding the hNab50 protein was isolated from a human osteosarcoma cDNA library, and the deduced amino acid sequence indicated that hNab50 is a basic (pI = 8.75) ~52 kDa protein (Fig. 4A). The hNab50 protein is related to a family of RNA-binding proteins which possess three RNA-binding domains (RBDs) and are differentially expressed in the vertebrate nervous system (33,34). Two proteins were detectable in HeLa cells by immunoblot analysis using an anti-hNab50 monoclonal antibody. The sizes of these two proteins, 49 and 51 kDa, were identical to CUG-BP1 and CUG-BP2, respectively. Proteins immunologically related to hNab50 were also present in a variety of vertebrate cells from human to frog, but were not present in *Saccharomyces cerevisiae* (Fig. 4B).

Two criteria were used to establish that hNab50 was an authentic human hnRNP (25,26). First, hNab50 was shown to be directly bound to poly(A)$^+$ RNA *in vivo* using a photocrosslinking assay (Fig. 5A). Second, hNab50 was localized predominantly in the nucleus by immunofluorescence microscopy (Fig. 6). Although

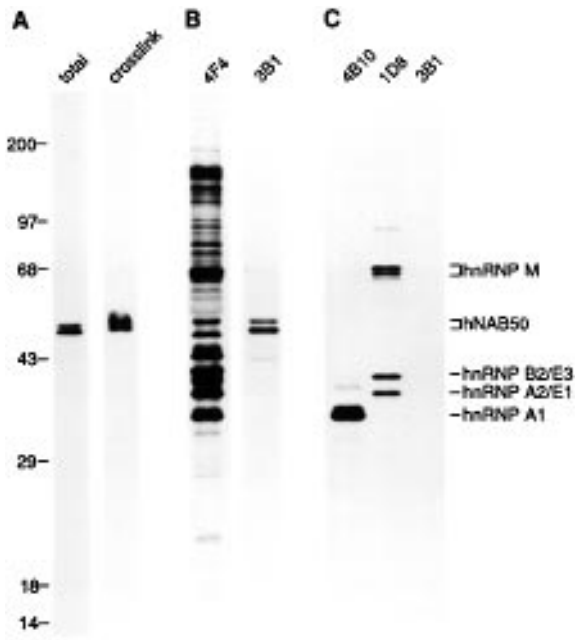

Figure 5. The CUG-BP/hNab50 protein is associated with poly(A)$^+$ RNA *in vivo* but fails to co-immunopurify with hnRNP complexes. (**A**) hNab50 is associated with poly(A)$^+$ RNA *in vivo*. Total HeLa cell proteins (total) or proteins photocrosslinked to poly(A)$^+$ RNAs *in vivo* (crosslink) were immunoblotted with mAb 3B1. The decrease in relative mobility of hNab50 in the crosslink lane is due to crosslinked nucleotides which remain following nuclease digestion. Sizes are indicated in kilodaltons. (**B**) A monoclonal antibody against CUG-BP/hNab50 fails to immunopurify the hnRNP complex. HeLa cells were labeled with [^{35}S]methionine and hnRNP complexes immunopurified using mAb 4F4 against the hnRNP C proteins (4F4). Parallel immunopurifications were performed using mAb 3B1 against CUG-BP/hNab50 (3B1). (**C**) CUG-BP/hNab50 is not a major component of the immunopurified hnRNP complex. HnRNP complexes were isolated as described in (B) using mAb 4F4. RNA–protein complexes were then dissociated by 1% SDS and boiling followed by dilution to 0.1% SDS. Monoclonal antibodies were then used to immunopurify hnRNP A1 (4B10), hnRNP M (1D8) and CUG-BP/hNab50 (3B1). Under these conditions, mAb 1D8 also immunopurifies proteins that co-migrate with the hnRNP A2/E1 and B2/E3 proteins.

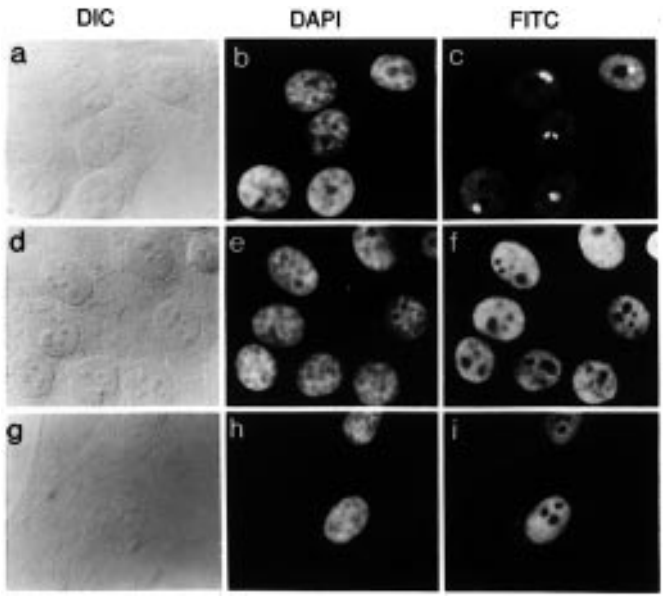

Figure 6. CUG-BP/hNab50 is a nuclear RNA-binding protein. The CUG-BP/hNab50 and hnRNP M proteins were localized within cells by indirect cellular immunofluorescence using either the 3B1 (c, i) or 1D8 (f) mAbs, respectively. Both HeLa cells (**a**–**f**) and normal patient myoblasts (**g**–**i**) are shown. The positions of the cells are shown by differential interference contrast (DIC) microscopy (a, d, g), and the chromosomal DNA by DAPI staining (b, e, h). The HeLa cell in the upper right corner of (c) shows a higher nucleoplasmic signal since it is in a different focal plane than the other four cells.

hNab50 was primarily nuclear in distribution, it also appeared to accumulate in a peri-nucleolar region in HeLa cells. This subnuclear localization pattern is similar to that previously described for the hnRNP I/PTB protein and several Y Ro RNP-associated RNAs within a peri-nucleolar compartment (35,36). This pattern is different from the majority of previously characterized hnRNPs, such as the hnRNP M proteins (28), which are distributed throughout the nucleoplasm (Fig. 6d–f). In contrast with HeLa cells, hNab50 was distributed uniformly throughout the nucleoplasm in both normal patient myoblasts (Fig. 6g–i) and fibroblasts (data not shown). This striking difference in the subcellular distribution of CUG-BP/hNab50 has also been detected in several other cell lines including Hep2 cells, which show the peri-nucleolar localization of CUG-BP/hNab50, and A549 cells, which do not (data not shown). Therefore, CUG-BP/hNab50 is localized primarily in the nucleus, but its intranuclear distribution is highly variable between different cell types.

Many hnRNPs have been isolated based upon their co-immunopurification with the hnRNP complex (30,31). This complex is composed of >20 major heterogeneous nuclear RNA (hnRNA)-binding proteins and requires intact hnRNA for stability. To ascertain if hNab50 was located within the hnRNP complex, the anti-hnRNP C mAb 4F4 was used to immunopurifiy hnRNP complexes from [^{35}S]methionine labeled HeLa cell nucleoplasm. As previously described (30), numerous proteins co-immunopurify with the hnRNP C proteins under these conditions (Fig. 5B). However, when the anti-hNab50 mAb 3B1 was used for immunopurification, only hNab50 was detected. This result suggested that either hNab50 was not a major component of the hnRNP complex or that mAb 3B1 was unable to recognize hNab50 within the complex. To distinguish between these two possibilites, hnRNP complexes were first immunopurified with mAb 4F4, the complexes subsequently dissociated by heating to 100°C in 1% SDS followed by immunopurification of the hnRNP A1 protein with the mAb 4B10, the hnRNP M proteins with mAb 1D8, and the hNab50 protein with mAb 3B1 under conditions which minimize non-specific protein–protein interactions (Fig. 5C). Although the hnRNP A1 and M proteins were efficiently immunopurified under these conditions, CUG-BP/hNab50 was not. These results demonstrated that hNab50 was not a major component of the immunopurified hnRNP complex, and suggested the possibility that this hnRNP might be associated with a subpopulation of mRNAs.

CUG-BP/hNab50 binds to RNAs containing the Mt-PK 3′-UTR

The results described above demonstrated that CUG-BP/hNab50 bound to $(CUG)_n$ repeats *in vitro* and was a pre-mRNA/mRNA-binding protein *in vivo*. To test if this hnRNP was able to bind to

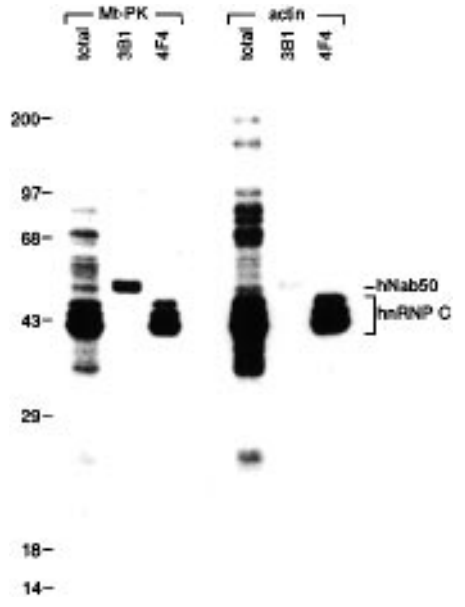

Figure 7. The CUG-BP/hNab50 protein in HeLa cell nuclear extracts preferentially photocrosslinks to RNAs containing the 3′-UTR of Mt-PK. Labeled RNAs containing Mt-PK and actin 3′-UTR sequences were synthesized *in vitro*, incubated in HeLa cell nuclear extracts, and photocrosslinked with UV light. Total or immunopurified proteins were then fractionated by SDS–PAGE, and proteins that bound were detected by label transfer/autoradiography. Immunopurifications were performed using either an anti-hnRNP C mAb (4F4) or an anti-hNab50 mAb (3B1). Sizes are indicated in kilodaltons.

the 3′-UTR of Mt-PK mRNA, we employed an *in vitro* photocrosslinking assay. Labeled RNAs containing the 3′-UTRs of the Mt-PK and actin genes were prepared by *in vitro* transcription, incubated in HeLa cell nuclear extracts, and proteins were crosslinked to the RNAs by exposure to UV light. Following RNase digestion, proteins crosslinked to RNAs were fractionated by SDS–PAGE and detected by autoradiography. The majority of proteins in HeLa cell nuclear extracts, including the hnRNP C proteins, crosslinked more efficiently to the actin 3′-UTR than to the Mt-PK RNA (Fig. 7). In contrast, the CUG-BP/hNab50 protein crosslinked preferentially to the Mt-PK RNA. These results demonstrate that CUG-BP/hNab50 binds to RNAs containing the 3′-UTR of Mt-PK, and suggest the possibility that this hnRNP may possess transcript-specific binding properties.

DISCUSSION

Myotonic dystrophy is one of several diseases in humans which are associated with the expansion of a trinucleotide repeat (37,38). These trinucleotide repeat expansions occur in various regions of the affected gene, and generally result in either loss of the correct gene product or a dominant gain-of-function phenotype in which the structure of the gene product is altered. Three models have been proposed to explain the puzzling observation that the $(CTG)_n$ triplet repeat expansion in the 3′-UTR of the Mt-PK gene results in an autosomal dominant and variable phenotype. First, the precise level of the Mt-PK protein may be critical to normal cellular function and repeat expansion in one of the Mt-PK alleles may lead to haploinsufficiency and disease (9). Second, the mutant DM allele may alter chromatin structure by changing nucleosome positioning and affect the expression of both Mt-PK and other linked genes (39–41). Third, the $(CTG)_n$ repeat expansion may be a dominant gain-of-function mutation either exerted in trans at the RNA level or this repeat may be a binding site for a specific nuclear RNA-binding protein (19,23). Recent studies have provided evidence which support the proposal that Mt-PK gene expression may be affected at the post-transcriptional level in DM cells. Hoffman and co-workers have shown that $(CTG)_n$ repeat expansion results in the reduction of poly(A)$^+$ mRNAs from both the normal and DM mutant alleles even though there is only a minor effect on the transcription and accumulation of Mt-PK pre-mRNAs (19). Another study using Mt-PK mRNA-specific *in situ* hybridization analysis has shown that transcripts from the DM mutant allele accumulate within intranuclear foci although both wild-type and mutant Mt-PK mRNAs are detectable in the cytoplasm (42).

In this study, we identify a novel human hnRNP as a candidate for the first triplet repeat eukaryotic RNA-binding protein to be characterized. We present three lines of evidence that hNab50 is responsible for $(CUG)_n$ triplet repeat RNA-binding activity: (i) polyclonal and monoclonal antibodies against hNab50 specifically recognize purified CUG-BP1 and CUG-BP2; (ii) anti-hNab50 antibodies supershift/neutralize CUG-BP activity; (iii) recombinant hNab50 and CUG-BP isolated from human cells have identical and specific $(CUG)_8$ RNA-binding activities. In addition, the CUG-BP/hNab50 protein binds to RNAs containing the 3′-UTR of Mt-PK mRNA *in vitro*, and nuclear extracts from DM cells show alterations in CUG-BP activities compared with normal cells. We conclude that CUG-BP/hNab50 is a $(CUG)_n$ triplet repeat RNA-binding protein in human cells, and propose that the Mt-PK $(CUG)_n$ triplet repeat is a binding site for CUG-BP/hNab50 *in vivo*.

How might a $(CUG)_n$ RNA-binding protein be involved in the regulation of gene expression and DM pathogenesis? In prokaryotes, a triplet repeat RNA-binding protein has been previously characterized that regulates both transcription and translation. The *trp* RNA-binding attenuation protein (TRAP) of *Bacillus subtilis* binds specifically to an RNA secondary structure, the antiterminator region, in the nascent *trp* operon leader transcript. Binding of TRAP, which is dependent on the presence of 11 G/UAG triplet repeats within the leader transcript, results in the disruption of the antiterminator and transcriptional termination upstream of the *trp* structural genes (43–45). TRAP also appears to play a role in translation by binding to a G/UAG-rich region that overlaps the ribosome binding site in *trpG* transcripts. In eukaryotes, hnRNPs bind to RNA polymerase II transcripts following transcripitional initiation (30,31). The association of these abundant nuclear pre-mRNA/mRNA-binding proteins to nascent transcripts is believed to play an important role in facilitating the formation of pre-mRNA structures amenable to subsequent pre-mRNA processing events. Recent work has also demonstrated that hnRNPs are not exclusively nuclear proteins, and therefore they could potentially function in the nucleocytoplasmic export, translation and turnover of mRNAs (30–32). Expansion of the Mt-PK 3′-UTR triplet repeat in DM cells would lead to a large increase in the number of potential binding sites for a $(CUG)_n$ triplet repeat mRNA-binding protein. If CUG-BP/hNab50 is important for the biogenesis and/or turnover of both Mt-PK as well as other mRNAs, then a reduction in the availability of this hnRNP might affect the processing and/or turnover of these mRNAs. This RNA processing/turnover defect

could manifest itself by affecting different RNAs in different tissues, thus accounting for the highly variable DM phenotype seen in humans.

ACKNOWLEDGEMENTS

We thank L. Green and L. Matthews for assistance with antibody preparation, M. Sardana for help with the purification of CUG-BP, G. Dreyfuss for providing antibodies to hnRNP A1 and C. S. Willingham for photography, and members of our laboratories for comments on the manuscript. This work was supported by grants from the NIH (GM46272), the NHLBI Specialized Centers of Research (HL54313-01), the NIH Training Center in Molecular Cardiology (T32-HL07706), the AHA Burgher Foundation Center for Molecular Biology (86-2216) and an AHA Medical Student Research Fellowship to J.W.M.

REFERENCES

1 Harper, P.S. (1989) *Myotonic Dystrophy.* Saunders, London/Phildelphia, 2nd Ed.
2 Harley, H.G., Brook, J.D., Rundle, S.A., Crow, S., Reardon, W., Buckler, A.J., Harper, P.S., Housman, D.E. and Shaw, D.J. (1992) *Nature* **355**, 545–546.
3 Buxton, J., Shelbourne, P., Davies, J., Jones, C., Van Tongeren, T., Aslanidis, C., de Jong, P., Jansen, G., Anvret, M., Riley, B. *et al.* (1992) *Nature* **355**, 547–548.
4 Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Mahadevan, M., Tsilfidis, C., Chen, C., Alleman, J., Wormskamp, N.G., Vooijs, M. *et al.* (1992) *Nature* **355**, 548–551.
5 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.
6 Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barceló, J., O'Hoy, K. *et al.* (1992) *Science* **255**, 1253–1255.
7 Jansen, G., Mahadevan, M., Amemiya, C., Wormskamp, N., Segers, B., Hendriks, W., O'Hoy, K., Baird, S., Sabourin, L., Lennon, G. *et al.* (1992) *Nature Genet.* **1**, 261–266.
8 Mahadevan, M.S., Amemiya, C., Jansen, G., Sabourin, L., Baird, S., Neville, C.E., Wormskamp, N., Segers, B., Batzer, M., Lamerdin, J. *et al.* (1993) *Hum. Mol. Gen.* **2**, 299–304.
9 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.. Jr *et al.* (1993) *Science* **260**, 235–238.
10 Shaw, D.J., 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.
11 Anvret, M., Åhlberg, G., Grandell, U., Hedberg, B., Johnson, K. and Edström, L. (1993) *Hum. Mol. Gen.* **2**, 1397–1400.
12 Jaspert, A., Fahsold, R., Grehl, H. and Claus, D. (1995) *J. Neurol.* **242**, 99–104.
13 Wong, L.-J. C., Ashizawa, T., Monckton, D.G., Caskey, C.T. and Richards, C.S. (1995) *Am. J. Hum. Genet.* **56**, 114–122.
14 Sabourin, L.A., Mahadevan, M.S., Narang, M., Lee, D.S.C., Surh, L.C. and Korneluk, R.G. (1993) *Nature Genet.* **4**, 233–238.
15 Hofmann-Radvanyi, H., Lavedan, C., Rabès, J.P., Savoy, D., Duros, C., Johnson, K. and Junien, C. (1993) *Hum. Mol. Genet.* **2**, 1263–1266.
16 Carango, P., Noble, J.E., Marks, H.G. and Funanage, V.L. (1993) *Genomics* **18**, 340–348.
17 Hofmann-Radvanyi, H. and Junien, C. (1993) *Neuromuscul. Disord.* **3**, 497–501.
18 Koga, R., Nakao, Y., Kurano, Y., Tsukahara, T., Nakamura, A., Ishiura, S., Nonaka, I. and Arahata, K. (1994) *Biochem. Biophys. Res. Commun.* **202**, 577–585.
19 Wang, J., Pegoraro, E., Menegazzo, E., Gennarelli, M., Hoop, R.C., Angelini, C. and Hoffman, E.P. (1995) *Hum. Mol. Genet.* **4**, 599–606.
20 Krahe, R., Ashizawa, T., Abbruzzese, C., Roeder, E., Carango, P., Giacanelli, M., Funanage, V.L. and Siciliano, M.J. (1995) *Genomics* **28**, 1–14.
21 Reddy, S., Smith, D.B.J., Rich, M.M., Leferovich, J.M., Reilly, P., Davis, B.M., Tran, K., Rayburn, H., Bronson, R., Cros, D. *et al.* (1996) *Nature Genet.*, **13**, 325–335.
22 Jansen, G., Groenen, P.J., Bächner, D., Jap, P.H., Coerwinkel, M., Oerlemans, F., van den Broek, W., Gohlsch, B., Pette, D., Plomp, J.J. *et al.* (1996) *Nature Genet.* **13**, 316–324.
23 Timchenko, L.T., Timchenko, N.A., Caskey, C.T. and Roberts, R. (1996) *Hum. Mol. Genet.* **5**, 115–121.
24 O'Brien, R.M., Lucas, P.C., Yamasaki, T., Noisin, E.L. and Granner, D.K. (1994) *J. Biol. Chem.* **269**, 30419–30428.
25 Anderson, J.T., Wilson, S.M., Datar, K.V. and Swanson, M.S. (1993) *Mol. Cell. Biol.* **13**, 2730–2741.
26 Wilson, S.M., Datar, K.V., Paddy, M.R., Swedlow, J.R. and Swanson, M.S. (1994) *J. Cell Biol.* **127**, 1173–1184.
27 Piñol-Roma, S., Swanson, M.S., Gall, J.G. and Dreyfuss, G. (1989) *J. Cell Biol.* **109**, 2575–2587.
28 Datar, K.V., Dreyfuss, G. and Swanson, M.S. (1993) *Nucleic Acids Res.* **21,** 439–446.
29 Enoch, T., Zinn, K. and Maniatis, T. (1986) *Mol. Cell. Biol.* **6**, 801–810.
30 Dreyfuss, G., Matunis, M.J., Piñol-Roma, S. and Burd, C.G. (1993) *Ann. Rev. Biochem.* **62**, 289–321.
31 Swanson, M.S. (1995) In Lamond, A.I. (ed.) *Pre-mRNA Processing.* Springer-Verlag, New York, pp. 17–33.
32 Piñol-Roma, S. and Dreyfuss, G. (1992) *Nature* **355**, 730–732.
33 Richter, K., Grunz, H., and Dawid, I.B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8086–8090.
34 Kato, K. (1992) *Trends Neurosci.* **15**, 319–323.
35 Ghetti, A., Piñol-Roma, S., Michael, W.M., Morandi, C. and Dreyfuss, G. (1992) *Nucleic Acids Res.* **20**, 3671–3678.
36 Matera, A.G., Frey, M.R., Margelot, K. and Wolin, S.L. (1995) *J. Cell Biol.* **129**, 1181–1193.
37 Ashley, C.T., Jr and Warren, S.T. (1995) *Annu. Rev. Genet.* **29**, 703–728.
38 Paulson, H.L. and Fischbeck, K.H. (1996) *Ann. Rev. Neurosci.* **19**, 79–107.
39 Wang, Y.-H., Amirhaeri, S., Kang, S., Wells, R.D. and Griffith, J.D. (1994) *Science* **265**, 669–671.
40 Wang, Y.-H. and Griffith, J. (1995) *Genomics* **25**, 570–573.
41 Jansen, G., Bächner, D., Coerwinkel, M., Wormskamp, N., Hameister, H. and Wieringa, B. (1995) *Hum. Mol. Genet.* **4**, 843–852.
42 Taneja, K.L., McCurrach, M., Schalling, M., Housman, D. and Singer, R.H. (1995) *J. Cell Biol.* **128**, 995–1002.
43 Babitzke, P., Stults, J.T., Shire, S.J. and Yanofsky, C. (1994) *J. Biol. Chem.* **269**, 16597–16604.
44 Antson, A.A., Otridge, J., Brzozowski, A.M., Dodson, E.J., Dodson, G.G., Wilson, K.S., Smith, T.M., Yang, M., Kurecki, T. and Gollnick, P. (1995) *Nature* **374**, 693–700.
45 Babitzke, P., Bear, D.G. and Yanofsky, C. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7916–7920.