

The CELF Family of RNA Binding Proteins Is Implicated in Cell-Specific and Developmentally Regulated Alternative Splicing

ANDREA N. LADD, NICOLAS CHARLET-B., AND THOMAS A. COOPER*

Department of Pathology, Baylor College of Medicine, Houston, Texas 77030

Received 13 September 2000/Returned for modification 20 October 2000/Accepted 9 November 2000

Alternative splicing of cardiac troponin T (cTNT) exon 5 undergoes a developmentally regulated switch such that exon inclusion predominates in embryonic, but not adult, striated muscle. We previously described four muscle-specific splicing enhancers (MSEs) within introns flanking exon 5 in chicken cTNT that are both necessary and sufficient for exon inclusion in embryonic muscle. We also demonstrated that CUG-binding protein (CUG-BP) binds a conserved CUG motif within a human cTNT MSE and positively regulates MSE-dependent exon inclusion. Here we report that CUG-BP is one of a novel family of developmentally regulated RNA binding proteins that includes embryonically lethal abnormal vision-type RNA binding protein 3 (ETR-3). This family, which we call CELF proteins for CUG-BP- and ETR-3-like factors, specifically bound MSE-containing RNAs in vitro and activated MSE-dependent exon inclusion of cTNT minigenes in vivo. The expression of two CELF proteins is highly restricted to brain. CUG-BP, ETR-3, and CELF4 are more broadly expressed, and expression is developmentally regulated in striated muscle and brain. Changes in the level of expression and isoforms of ETR-3 in two different developmental systems correlated with regulated changes in cTNT splicing. A switch from cTNT exon skipping to inclusion tightly correlated with induction of ETR-3 protein expression during differentiation of C2C12 myoblasts. During heart development, the switch in cTNT splicing correlated with a transition in ETR-3 protein isoforms. We propose that ETR-3 is a major regulator of cTNT alternative splicing and that the CELF family plays an important regulatory role in cell-specific alternative splicing during normal development and disease.

The generation of multiple, functionally distinct protein isoforms from a single gene via alternative splicing is a common means of regulating gene expression. It has been estimated that more than one-third of human genes are alternatively spliced (21). Despite the prevalence of alternative splicing, the mechanisms by which it is regulated are not well understood. *Cis*-acting elements that mediate alternative splicing specific to different cell types have been identified in a few experimental systems (1, 3, 18, 19, 50, 66), and progress in identifying *trans*-acting factors involved in tissue-specific regulation, particularly neuron-specific splicing, has been made. KH-type splicing regulatory protein is enriched in neurons and has been isolated as a component of a complex that activates inclusion of the *c-src* neuronal N1 exon (43). The KH-type RNA binding protein Nova-1 is expressed exclusively in neurons of the central nervous system (4) and activates inclusion of exons in the glycine receptor and GABA_A receptor pre-mRNAs (27). Brain polypyrimidine tract binding protein (brPTB), a protein related to the more ubiquitous PTB but enriched in brain, binds to the same sequence recognized by PTB and antagonizes the effects of Nova-1 (48).

In addition to the use of tissue-restricted factors, cell-specific alternative splicing may also be regulated by altering the abundance or activities of constitutive pre-mRNA splicing factors such as the SR protein family and hnRNP proteins (6, 31, 34, 58). These factors participate in constitutive splicing, but their

expression is not entirely ubiquitous and undergoes tissue-specific regulation (14, 20, 30, 57, 65, 68). Regulation of alternative splice site choices by natural variation in the expression of these proteins, however, remains to be directly demonstrated.

For many genes, alternative splicing is also modulated in a developmental-stage-specific manner. Alternative splicing is determinative for cell fate decisions during development in *Drosophila melanogaster*, such as the regulation of sexual differentiation by the RNA binding protein Sex-lethal (reviewed in reference 35). Changes in the constitutive splicing machinery in nematodes also suggest a developmental role for alternative splicing regulation (54). Factors involved in the developmental regulation of alternative splicing in vertebrates, however, have not yet been reported.

Many proteins essential for striated muscle development exist in multiple isoforms generated by alternative splicing, including myogenic transcription factors, metabolic enzymes, and components of the myofibril. Skeletal-muscle-specific splicing patterns are induced during differentiation and can be induced in fibroblasts by expression of MyoD and myogenin (26, 50, 52). Alternative splicing of exon 5 of the cardiac troponin T (cTNT) gene undergoes a developmentally regulated switch such that mRNAs in embryonic striated muscle include the exon, but mRNAs in the adult do not (11). The developmental regulation of cTNT splicing is conserved in chickens, mice, rats, and humans (11, 28, 39, 60). The two cTNT isoforms generated by alternative splicing of exon 5 confer different levels of calcium sensitivity to the myofilament, thereby affecting the contractile properties of maturing muscle (16, 40).

* Corresponding author. Mailing address: Department of Pathology, Baylor College of Medicine, Houston, TX 77030. Phone: (713) 798-3141. Fax: (713) 798-5838. E-mail: tcooper@bcm.tmc.edu.

Previously, we identified four *cis*-acting elements within the introns flanking exon 5 in the chicken cTNT that are both necessary and sufficient to promote exon inclusion specifically in embryonic striated muscle (10, 46, 50). These muscle-specific splicing enhancers (MSEs) were defined in cTNT minigenes by deletions or substitutions that prevented the activation of exon 5 inclusion in embryonic skeletal muscle cultures but that had little effect on the low level of exon inclusion observed in nonmuscle cultures (10, 46, 50). One of these MSEs, MSE2, is conserved in sequence and position in chicken and human cTNT (46). MSE2 is necessary and sufficient to promote muscle-specific cTNT exon 5 inclusion in embryonic skeletal muscle cultures when present in multiple copies (10). A conserved CUG motif found within MSE2 is found adjacent to alternative exons in several genes that undergo regulated splicing in striated muscle and is required for enhancer activity, suggesting that these elements play a common role in muscle-specific splicing (10, 46).

CUG-binding protein (CUG-BP) is a highly conserved protein that was purified based on its sequence-specific binding to RNA containing CUG repeats (59). CUG-BP is proposed to play a role in the pathogenesis of the trinucleotide expansion disease, myotonic dystrophy (DM1), in which nuclear accumulation of transcripts of the myotonin protein kinase (DMPK) gene containing expanded CUG repeats is proposed to create an RNA gain-of-function mutation. We have shown that CUG-BP binds to the conserved CUG motif within cTNT MSE2 and positively regulates MSE-dependent inclusion of exon 5 (46). Embryonically lethal abnormal vision (ELAV)-type RNA binding protein 3 (ETR-3) (25), which has high sequence similarity to CUG-BP, has also recently been reported to bind to RNAs containing CUG triplet repeats (36), suggesting that it may also regulate the alternative splicing of MSE-containing transcripts.

Here we report the identification of a novel family of splicing regulators that includes CUG-BP and ETR-3, which we call CELF proteins for CUG-BP- and ETR-3-like factors. Individual CELF family members are preferentially expressed in different cell types, and at least two CELF genes express multiple isoforms via alternative splicing. We demonstrate that CELF proteins bind to MSE-containing RNAs and activate MSE-dependent exon inclusion in fibroblasts when coexpressed with cTNT minigenes. Expression of these proteins is developmentally regulated. ETR-3 protein expression in particular correlates with changes in cTNT alternative splicing during heart development in both chickens and mice and during myogenic differentiation of the mouse C2C12 myoblast cell line. We propose that ETR-3 is a major regulator of cTNT alternative splicing. Our results strongly support the hypothesis that the CELF family plays a critical role in the regulation of cell-specific alternative splicing during normal development and disease.

MATERIALS AND METHODS

Identification and analysis of CELF sequences. ETR3, CELF3, CELF4, and CELF5 were identified in searches of expressed sequence tag (EST) and high-throughput genomic sequence databases using a CUG-BP sequence query. ETR-3 was first identified in a screen of genes expressed in human fetal heart (25). Alternatively spliced forms of ETR-3 were recently found in a screen of genes induced during apoptosis in a neuroblastoma cell line and were called NAPOR-1, -2, and -3 (7); they were also found by a screen of a liver cDNA

library using a CUG-BP probe (36). A truncated version of CELF3 called CAGH4 is also found in the database (38). A potential sixth family member was identified in cosmids from chromosome 15 (accession no. AC009524) and AC009524), but corresponding ESTs could not be found.

The open reading frames for all five CELF proteins were cloned into the pcDNA3.1HisC vector (Invitrogen) in frame with the N-terminal Xpress epitope tag using 5' primers containing the initiation codon and 3' primers containing the termination codon. CUG-BP primers were GTTAGTGGATCCATGAACGGC ACCCTGGACCA and GGCCGAAGCTTTCAGTAGGGCTTGCTGT. ETR-3 primers were CGGTGAGATCTATGAACGGAGCTTTGGA and CGGTCAA GCTTTCAGTAAGGTTTGCTGTCGT. CELF3 primers were GTTAGAGGA TCCATGAAGGAGCCGGATGCCATCA and AATTACCTCGAGTCAGTA GGGCCGTTGGCATCCTTA. CELF4 primers were GTTAGAGGATCCAT GAAGGACCACGATGCCATC and AATTACCTCGAGTCAGTACGGGCG ATTGGCGTCTTTG. CELF5 primers were CATCATCGGATCCATGGCCC GCCTGACGGAGAG and CATCATCTCTAGATCACGGGTCTTTGGGCC GCT. For all CELF proteins except CELF4, the expressed open reading frames begin at the codon for the first methionine shown in Fig. 1. The open reading frame for CELF4 in the expression vector starts at the codon for the methionine at position 48, as this was the first known methionine at the time of cloning; additional ESTs that contain the putative upstream translation presented in Fig. 1 have subsequently been entered in the database. The templates for PCR included a CUG-BP cDNA clone provided by L. Timchenko (Baylor College of Medicine, Houston, Tex.) and an ETR-3 cDNA provided by C. C. Liew (University of Toronto, Toronto, Ontario, Canada). The 5' ends of CELF3 and CELF4 were not available in the EST database and were cloned by rapid amplification of cDNA ends (RACE) using the Marathon RACE kit (Clontech Laboratories, Inc.) on adult brain cDNA (Clontech). The 5' RACE products provided the sequence of initiation codons, which were then used to amplify full-length CELF3 and CELF4 from commercial cDNA from adult brain (Clontech). The 5' end of CELF5 mRNA (including the putative initiation codon) was found in cosmid R31341, and the 3' end of CELF5 mRNA (including the termination codon and 3' untranslated region [UTR]) was found in several ESTs. CELF5 was amplified using cDNA generated from human total brain RNA (Clontech) in a 20- μ l reaction mixture containing 2.5 μ g of RNA, 50 ng of oligo (dT)₁₂₋₁₈ primer (Life Technologies), First Strand buffer (Life Technologies), 0.01 M dithiothreitol [DTT], 0.5 mM deoxynucleoside triphosphate (dNTP) mixture, and 100 U of avian myeloblastosis virus reverse transcriptase (Life Technologies). The RNA and primers were heated to 65°C for 10 min and chilled on ice, and then the remaining reagents were added and the reaction mixture was incubated at 42°C for 1 h and at 70°C for 15 min and stored at -20°C. All PCRs were performed in a 50- μ l reaction volume containing *Taq* Plus Precise buffer (Stratagene, La Jolla, Calif.), 0.25 mM dNTP mixture, 2.5 μ l of dimethyl sulfoxide, 200 ng of each of the 5' and 3' primers, and 2.5 U of *Taq* Plus Precise enzyme (Stratagene) on an MJ Research PTC-100 using 30 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 2 min, followed by 1 cycle at 72°C for 10 min. PCR products were isolated from 1% agarose gels using the QIAquick gel extraction kit (Qiagen) and cloned directly into pcDNA3.1HisC. Sequences were confirmed by overlapping reads of both strands.

Sequence profile analyses were performed using PROSITE (24) and PSORTII (44). Nuclear localization predictions were made using both Reinhardt's NCNN and the *k*-NN nearest-neighbor prediction algorithms. The simple modular architecture research tool (SMART) (56) was used to identify proteins with a domain structure similar to that of the CELF proteins. Alignments and phylogenetic analyses were performed using DNASTAR software, which utilizes the Clustal V method (23).

Western blots. Antipeptide antibodies against ETR-3 (against a 17-amino-acid peptide [QTSATSTNANPLSTGC] conjugated to keyhole limpet hemocyanin) were generated in rabbits by Cocalico Biologicals. Antipeptide antibodies against CELF4 (against a 15-amino-acid peptide [CIHPYPAQSPSTAADP] conjugated to keyhole limpet hemocyanin) were generated in rabbits by Anaspec, Inc. The 3B1 mouse monoclonal antibody against CUG-BP was provided by M. Swanson (University of Florida, Gainesville). Tissues were dissected from embryonic-day-14, newborn, 4-day-old, and pregnant or 0- to 4-day-postpartum adult female Swiss Webster mice (Taconic Farms, Inc.) and from embryonic-day-8 to -20 and adult White Leghorn chickens (Texas A&M University) and homogenized directly in protein loading buffer (0.64 M Tris-HCl [pH 6.8], 10% glycerol, 2% sodium dodecyl sulfate [SDS], 5% β -mercaptoethanol). C2C12 murine myoblast cells were maintained at a density of 20 to 60% confluency in 100-mm-diameter tissue culture dishes in 10 ml of growth medium (low-glucose Dulbecco modified Eagle medium [DMEM] supplemented with 20% fetal bovine serum and 0.5% chick embryo extract) and cultured at 37°C in 5% CO₂. To induce differentiation, growth medium was replaced with differentiation medium (low-glucose DMEM

supplemented with 5% horse serum) when cells had reached 60% confluency. To collect total protein samples from C2C12 cultures, cells were rinsed with phosphate-buffered saline and then scraped directly into protein loading buffer and sonicated. Protein concentrations were determined using a noninterfering protein assay (Geno Technology, Inc.). Total protein samples were resolved on 10% denaturing polyacrylamide gels, with 50 to 60 μ g loaded per lane, and transferred to Immobilon-P membranes (Millipore). Membranes were incubated for 1 h in blocking solution (5% milk in phosphate-buffered saline plus 0.05% Tween-20 [PBT]) at room temperature with shaking. The primary antibody was added, the mixture was incubated for 1 h, and then the membrane was washed twice for 20 min with PBT and blocked again for 1 h. Goat anti-rabbit or goat anti-mouse horseradish peroxidase-conjugated secondary antibody (Jackson ImmunoResearch Laboratories, Inc.) was added, and the mixture was incubated for 1 h; then the membrane was washed three times for 20 min with PBT, and bands were visualized with SuperSignal chemiluminescent substrate (Pierce) on Kodak film (Eastman Kodak).

UV cross-linking. The open reading frames for CUG-BP, ETR-3, and CELF4 were cloned downstream of glutathione *S*-transferase (GST) in the pGEX-2T GST fusion expression plasmid (Novagen). BL21(DE3)pLysS competent bacteria (Stratagene) were transformed and tested for the presence of the plasmid. One colony was grown overnight at 37°C in Terrific broth (53) plus 80 μ g of Timentin/ml, diluted 1:10, and grown for an additional 2 h. IPTG (0.1 mM) was added, and cells were grown for 3 h and pelleted. The cell pellet was resuspended in 0.2 culture volumes of phosphate-buffered saline and sonicated on ice six times for 10 s each. Triton X-100 was added to a final concentration of 1%, and the total lysate was spun for 10 min at 12,000 \times g. Recombinant GST fusion proteins were purified from the supernatant according to the GST-Bind buffer kit (Novagen) and quantitated using a protein assay kit (Bio-Rad), and the quantity and purity were checked on SDS-polyacrylamide gel electrophoresis gels by Coomassie staining (53).

The templates for *in vitro* RNA synthesis were derived from the MSE-containing and MSE-lacking minigenes used for transfection (see below). The MSE-containing substrate contains 30 nucleotides of the pBluescript vector (Stratagene) downstream of the T3 promoter, 93 nucleotides of cTNT intron 4 including MSE1, a 63-nucleotide artificial exon derived from skeletal troponin I (GGTTCAACAACCATCTACGCATTGCAAGAGGCATTGGATCCGAACC AAGCAAGATGTCTGACAG), and 142 nucleotides of cTNT intron 5 including MSE2 to -4. The plasmid was linearized using *SpeI* (Roche Molecular Biochemicals). The substrate containing MSEs 2 to 4 was obtained by removing a *SexAI-BamHI* (Roche Molecular Biochemicals) fragment situated between the T3 promoter and the exon which encompassed MSE1. For the substrates containing the full-length intron and MSEs 2 to 4, the plasmids were linearized using *SpeI* (Roche Molecular Biochemicals); the MSE1 substrate was obtained by linearizing the MSE1 to 4 template with *BstBI* (Roche Molecular Biochemicals), which cuts within the exon. The linearized plasmids were transcribed with T3 RNA polymerase (Life Technologies). *In vitro* synthesis of uniformly labeled RNA was performed as described previously (9).

UV cross-linking was performed using 1 pmol of substrate (0.4×10^6 to 1.0×10^6 cpm) labeled with [α - 32 P]GTP and [α - 32 P]UTP and 500 ng of recombinant GST fusion proteins. Proteins and RNA were incubated at 30°C for 10 min with a mixture containing 30 μ g of HeLa nuclear extract (13), 2 mM magnesium acetate, 2 mM ATP, 16 mM HEPES (pH 7.9) 65 mM potassium glutamate, 0.16 mM EDTA, 0.4 mM DTT, and 16% glycerol. In some experiments, 1 μ g of tRNA (Ambion), 1 μ g of bovine serum albumin, and 0.1 μ g of heparin were substituted for HeLa nuclear extract without affecting the results. Incubation with bulk RNA and protein is necessary to decrease the nonspecific binding of the recombinant proteins. Reaction mixtures were placed on an aluminum block prechilled in ice water and irradiated 4 cm from a Philips G15T8 germicidal lamp for 8 min. Samples were digested with 0.5 μ g of each RNase A and RNase T1 (Sigma) at 37°C for 20 min. An equal volume of protein loading buffer was added, and samples were denatured at 90°C and run on a SDS-12% polyacrylamide gel. Sizes were determined using prestained markers (Bio-Rad).

Cotransfection experiments. R35C was derived from RTB33.51 (50) by filling in a *BstBI* site located 86 nucleotides upstream of the insertion site for the cTNT genomic fragment in the skeletal troponin I minigene. M2/M2TB has been described previously (10). RTBPSRAX was derived from RTBPSR (50) by removing all of the cTNT intron sequence (68 nucleotides between Asp718 and *XbaI* sites).

QT35 quail fibroblast cells were plated at a density of 1.8×10^6 cells/60-mm-diameter tissue culture dish in 3 to 5 ml of growth medium (F10 medium supplemented with 5% fetal bovine serum, 1% chick serum, 10% tryptose phosphate, and 2 mM L-glutamine) and cultured overnight at 37°C in 5% CO₂. The medium was then changed to transfection medium (low-glucose DMEM supple-

mented with 10% fetal bovine serum and 2 mM L-glutamine), and cells were transfected with 2 μ g of minigene DNA and 0 to 10 μ g of CELF expression plasmid DNA using FuGene 6 (Roche Molecular Biochemicals). Medium was replaced with growth medium at 24 h, and total RNA was harvested 48 h following transfection by the method of Chomczynski and Sacchi (8) as modified by Xie and Rothblum (63). Total protein samples were collected from parallel plates by scraping cells into 250 μ l of protein loading buffer and sonication. Fifty microliters of each sample was run on a 10% denaturing polyacrylamide gel, and Western blot analysis was performed as described above using the anti-Xpress antibody (Invitrogen) at 1:5,000. To monitor cTNT splicing during muscle differentiation, C2C12 cells were transfected with 2 μ g of R35C minigene DNA using FuGene 6 (Roche Molecular Biochemicals) at 30% confluency. Growth medium was replaced with differentiation medium the next day, and RNA was harvested every 24 h.

The extent of minigene exon inclusion was determined by reverse transcription-PCR (RT-PCR) analysis. cDNA was generated as described above. PCR was performed in a 40- μ l volume containing *Taq* MgCl₂-free buffer (Promega), 1.75 mM MgCl₂, 0.2 mM dNTPs, 200 ng of each oligonucleotide (CATTCACC ACATGGTGTGC and AGGTGCTGCCGCCGGGGTGGCTG), 1.5 ng of γ - 32 P-labeled 5' primer, and 2.5 U of *Taq* (Promega) for 18 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min, followed by 1 cycle at 72°C for 5 min. The 5' primer was labeled with γ - 32 P by incubation with a mixture containing kinase buffer (70 mM Tris [pH 7.5], 1 mM MgCl₂, 0.5 mM DTT), 0.7 U of polynucleotide kinase (USB Corporation), enough oligonucleotide for 1.5 ng per PCR, and equal molar amounts of [γ - 32 P]ATP in a 10- μ l volume at 37°C for 30 min. PCR products were resolved on 5% nondenaturing polyacrylamide gels, and bands were quantitated using a PhosphorImager (Molecular Dynamics). The extent of endogenous cTNT exon inclusion in mouse heart was determined by RT-PCR analysis using primers AGCCGAGAGCATGTCTGACGCCGAGGA GGTGGT (in exon 3) and CAGCATCTTTGGCTTCATCAGGACCAACT (in exon 6) and 30 cycles of amplification. Primer extension of chicken heart RNA was performed using a 5'-end-labeled oligo nucleotide complementary to cTNT exon 6 as described previously (64), and the ratios of primer extension products were quantitated by a PhosphorImager.

Nucleotide sequence accession numbers. Accession numbers for the sequences of CELF3, CELF4, and CELF5 are AF329264, AF329265, and AF329266, respectively.

RESULTS

Identification of the CELF protein family. A screen of human EST databases was performed to identify CUG-BP paralogs based on sequence similarity. Four proteins closely related to CUG-BP were found, including ELAV-type RNA binding protein 3 (ETR-3) and three novel proteins we call CELF3, CELF4, and CELF5 (Fig. 1B). A potential sixth family member was identified in cosmid sequences from chromosome 15 (see Materials and Methods), but no corresponding ESTs were found and this family member was not studied further. The three novel protein sequences presented are derived from open reading frames amplified by PCR from human brain cDNA. All five proteins have the same domain structure (Fig. 1A); three RNA binding domains composed of the RNP-containing RNA recognition motif (RRM) (reviewed in reference 5) and a domain of unknown function separating RRM2 and RRM3. We have termed the region separating RRM2 and RRM3 the divergent domain, because sequence similarities within this domain split the CELF family into two subgroups of closely related proteins (Fig. 1C), the first containing CUG-BP and ETR-3 (78% identical) and the second containing CELF3, CELF4, and CELF5 (CELF3 and CELF5 have 60.8 and 63.8% identity with CELF4, respectively). BLAST searches reveal that the divergent domains are unique. No known protein-protein, protein-RNA, or protein-DNA interaction motifs, targeting signals, or predicted secondary structure were identified within the divergent domain. Variable regions most likely due

A.



B.

CUG-BP -----MNGTLDHPDQP-----DLDAIKMFVGGVPRTWSEKDLRELFQYGAIVEINVLDRDSQ
 ETR-3 -----MNGALDHS DQP-----DPDAIKMFVGGVPRSWSEKELKELFEPYGAIVEINVLDRDSQ
 CELF3 -----MKEPDAIKLFGVQIPRHLEEKDLKPIFEQFGRIFELTVIKDK--
 CELF4 MYIKMATLANGQADNASLSTNGLGSSPGSAGHMNGLSHSPGNPSTIPMKDHD AIKLF IQIPRNLEKDLKPLFEFEGKIYELTVLKDR--
 CELF5 M----ARL TESEARRQ Q--QQLLQPRPSPVGS-SGPEPPGGQPDG--MKDLDAIKLFGVQIPRHDDKDLKPLFEQFGRIYELTVLKDP--
RNP2

RRM1
RRM2

CUG-BP NPPQSKGCCFVTFYTRKAAL EAQNALHNMKVLP GMHHP IQMKPADSE-----KNAVEDRKLFI GMI SKKCTENDIRVMFSSFGQIEE
 ETR-3 NPPQSKGCCFVTFYTRKAAL EAQNALHNIK TLP GMHHP IQMKPADSE-----KNAVEDRKLFI GMI SKKCTENDIRVMFSSFGQIEE
 CELF3 YTGLHKGCAFLTYCARD SALKAQ SALSALHEQKTLP GMNRPI QVKPADSESRG-----EDRKLFGMLGKQQTDEDVRKMFEPFGTIDE
 CELF4 FTGMHKGCAFLTYCERESALKAQSALHEQKTLP GMNRPI QVKPADSESRGSSCLRQPPSQDRKLFGMLNKQSEDDVRRLEAFGNIEE
 CELF5 YTG MHK GCAFLTYCARD SAIKAQTALHEQKTLP GMARPI QVKPADSESRGGR-----DRKLFGMLNKQSEEDVLRLEFPFGVIDE
RNP1
RNP2

CUG-BP CRILRGP DGLSRGCAFVTFSTRAMAQTAIKAMHQ AQTMEGCSSPMVVKFADTKDKEQKRMAQQLOQQMQQISAASVWGNLAGLNTLGPQY
 ETR-3 CRILRGP DGLSRGCAFVTFSTRAMAQNAIKAMHQ SQTMEGCSSPIVVKFADTKDKEQRRLQQQLAQMQQLNTAT-WGNLTGLGLLTPQY
 CELF3 CTVLRGPDGTSKGC AFVKFQTHAEAQA AINTLHSSRTLP GASSSLVVKFADTEKERGLRRM-QQVATQLGMFSP-IA----LQFGAYSAYT
 CELF4 CTILRGP DGNSKGC AFVKYSSHAEAQA AINALHGSQTMPGASSSLVVKFADTKERTMRRM-QQMAQMGMFNP-MA----IPFGAYGAYA
 CELF5 CTVLRGPDGSSKGC AFVKFSSHTEAQA AIALHGSQTMPGASSSLVVKFADTKERTLRRM-QQMVGQLGILTPSLT----LPFSPYSAYA
RNP1

CUG-BP LALLQQTASSGNLNTLSS-LHPMGGLNAMQLQNLALALAAAA-SAAQNTPSGTNA--LTTSSSPLSVLTSS--GSSPSSSSNSVNP IASLG
 ETR-3 LALLQQATSSSNLGA FSG-IQQMAGMNALQLQNLATLAAAA-AAAQTSATSTNANPLSTSSALGALTS PVAASTPNSTAGAAMNSLTSLG
 CELF3 QALMQQQAALVA--AHSAYLSPMATMAAVQM QHMAAINANGLIATPI TSSSGTSTPPAIAATPVSAIPAALGVNGYSPVPTQP----TGQP
 CELF4 QALMQQQAALMASVAQGGYLNPMAAF AAAQM QMAALNMNGLAAAPMTPTSGGSTPPGITAPAVSIPSPIGVNGFTGLPPQA----NGQP
 CELF5 QALMQQQTTVLSTS--GSYLSPGVAFSPCHIQQ IGAVSLNGLPATPVAPASGLHSPPLLGTAVPGLVAPIT-NGFAGVVFPF----GGHP

CUG-BP ALQTLAGATAGLN-----VGS LAGMAALNGGLGS-SGLSNGTGSTMEAL TQAYSIGIQYAAAALPTLYNQNL TQQSIGAAGSQKEGPE
 ETR-3 TLQGLAGATVGLNNINALAVAQMLSGMAALNGGLGA-TGLTNGTAGTMDAL TQAYSIGIQYAAAALPTLYSQSLLQQQS--AAGSQKEGPE
 CELF3 APDALYPN--GVHPYPAQSPVAPVDPLQQAYAGMQHYT--AAYP-AAYSLVAPAFPQ----PPALVAQQPPPPPPQQQQQQQQQQQREGPD
 CELF4 AA EAVFAN--GIHPYPAQSP TAA-DPLQQAYAGVQQYAGPAAYP-AAYGQISQAFPQ----PPPMPQQ-----QREGPE
 CELF5 AL ETVYAN--GLVPYPAQSP TVA-ETLHPAFSGVQYQT--AMYPTAAITPIAHSVQ P----PPPLL-----QQQQREGPE

RRM3

CUG-BP GANLF IYHLPQEF GDQDL LQMFMPF GNVVSAKVFIDKQTNLSKCFGFVSYDNPVSAQA AIQSMNGFQIGMKRLKVQLKR SKNDSKPY
 ETR-3 GANLF IYHLPQEF GDQHLQMFMPF GNVISAKVFIDKQTNLSKCFGFVSYDNPVSAQA AIQAMNGFQIGMKRLKVQLKR SKNDSKPY
 CELF3 GCNIF IYHLPQEF TDS EILQMFVFPFGHVISA KVFVDRATNQS KCFGFVSYDNPASQA AIQAMNGFQIGMKRLKVQLKR PKDANRPY
 CELF4 GANLF IYHLPQEF GDAELMQMFLPFG-----FVSYDNPASQA TAIQAMNGFQIGMKRLKVQLKR PKDANRPY
 CELF5 GCNLF IYHLPQEF GDTELTQMF LPPFGNI ISSKVFMDRATNQS KCFGFVSYDNPASQA AIQAMNGFQIGMKRLKVQLKR PKD---P
RNP2
RNP1

C.

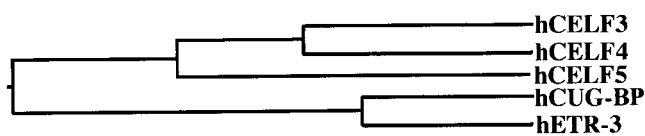


FIG. 1. The CELF family members are closely related. (A) All of the CELF proteins possess the same domain structure: three RRM s and a divergent domain of unknown function between RRM2 and RRM3. Nomenclature is designed to consider CUG-BP and ETR-3 as CELF1 and CELF2, respectively. (B) Sequences of five human CELF proteins derived from PCR-amplified cDNAs (see Materials and Methods). Blue, conserved nonpolar amino acids; red, conserved uncharged polar residues; green, conserved positively charged residues; gold, conserved negatively charged residues; black, nonconserved residues. Known alternatively spliced regions are shaded. Italic's, a region of allelic variation in CELF3 due to a variable number of CAG repeats. (C) Phylogenetic relationship of human CELF proteins.

to alternative splicing were identified in both ETR-3 and CELF4 (Fig. 1B). CELF3 contains a region of allelic variation in which a variable number of CAG repeats encode a stretch of glutamines of unfixed length (Fig. 1B). All of the CELF proteins contain multiple potential protein kinase C and casein kinase II phosphorylation sites. All are predicted to have predominantly nuclear localization, and CELF3, CELF4, and CELF5 each possess a consensus nuclear localization signal sequence near the C terminus. CUG-BP is known to be distributed within both the nucleus and cytoplasm and is detectable by Western blot analysis and gel shift assays in two isoforms that differ in phosphorylation state and intracellular distribution (49, 59). Among vertebrate proteins, the CELF family is most closely related to the Hu family of RNA binding proteins, homologs of *Drosophila* nuclear protein ELAV that

have been implicated in the regulation of mRNA stability and translation (reviewed in reference 32).

CELF proteins are widely expressed in adult tissues. To examine the expression of CELF family members, commercial RNA dot blots containing samples from a variety of human tissues were probed with 3'-UTR probes. CELF3 and CELF5 mRNAs were restricted to brain, whereas CUG-BP, ETR-3, and CELF4 displayed broader patterns of expression (data not shown). Consistent with these results, CELF3 and CELF5 ESTs were identified only in brain-derived libraries.

To examine the distribution of the three CELF family members that are not restricted to brain, Western blot analyses were performed on total protein samples extracted from a variety of adult mouse tissues (Fig. 2). A mouse monoclonal antibody against CUG-BP and rabbit polyclonal antipeptide antibodies against unique epitopes in ETR-3 and CELF4 were used. Each antipeptide antibody failed to recognize other CELF proteins that were transiently expressed in QT35 cells or as recombinant proteins (data not shown). Every tissue examined expressed one or more of the three CELF proteins, with skeletal muscle, heart, and brain showing the highest levels of CELF proteins overall. Each CELF family member was expressed in multiple tissues with a distinct pattern of tissue distribution. The multiple bands observed for ETR-3 and CELF4 suggest the presence of multiple isoforms, with different cell types showing different combinations of isoforms. At least some of these isoforms are likely to be generated by alternative splicing (Fig. 1B and data not shown). ETR-3 and CELF4 were particularly abundant in striated muscle and brain, where multiple isoforms were observed for each. In skeletal muscle, the pattern of these isoforms differed depending on the source of the tissue, perhaps due to differences in fiber type composition.

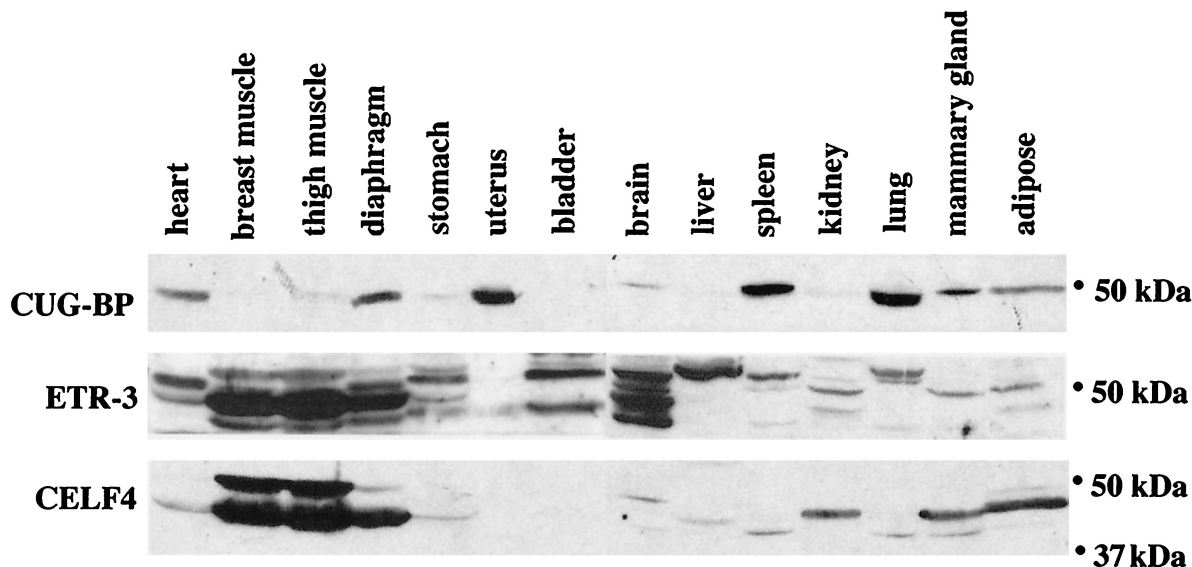


FIG. 2. CELF proteins are widely expressed. Western blot analysis was performed on total protein samples extracted from adult, pregnant or postpartum female Swiss Webster mice. Fifty micrograms of total protein was loaded in each lane. ETR-3 and CELF4 polyclonal antipeptide antibodies failed to recognize other CELF proteins as transiently expressed or recombinant proteins. An abundant 30-kDa protein was also detected in brain samples on CELF4 blots (data not shown). The relative intensity of this band was reduced somewhat by the addition of a protease inhibitor cocktail to the homogenization buffer, suggesting abundant CELF4 protein expression and cleavage or degradation in brain.

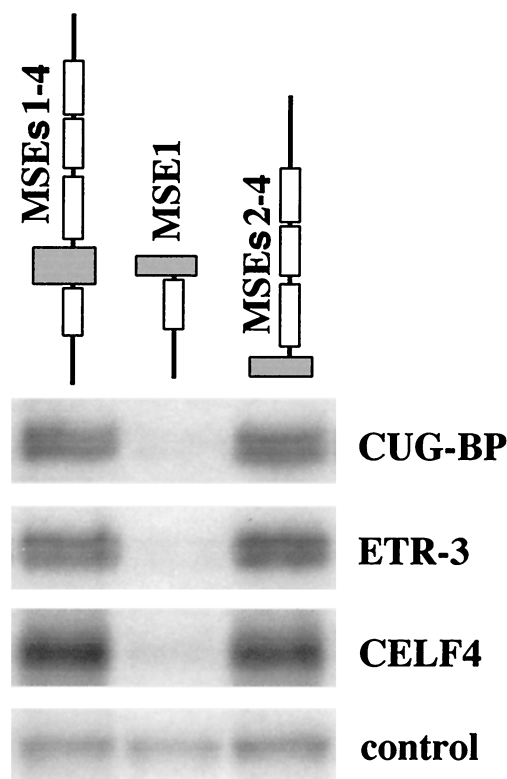


FIG. 3. CELF proteins bind specifically to a downstream MSE-containing region. UV cross-linking experiments were performed using purified recombinant GST-CELF fusion proteins and uniformly ^{32}P -labeled RNA (G and U) containing partial or full-length segments of the cTNT-regulated region. Proteins and RNA were preincubated in HeLa nuclear extract to increase the specificity of binding. Similar results were obtained when tRNA, bovine serum albumin, and heparin were used instead of HeLa nuclear extract (not shown). Each lane contains equal molar amounts of RNA. An approximately 100-kDa protein that bound nonspecifically to all three RNA substrates in HeLa cell nuclear extracts is also shown.

CELF family members bind to cTNT MSEs in vitro and positively regulate MSE-dependent splicing in vivo. We have previously shown that CUG-BP binds to a human MSE2-like sequence flanking the human cTNT alternative exon 5 and promotes MSE-dependent exon inclusion (46). To determine whether CELF proteins differ in their ability to bind MSEs, we tested the binding of CUG-BP, ETR-3, and CELF4, the three CELF proteins expressed in striated muscle, to chicken cTNT MSEs. Sequence-specific binding of purified recombinant GST fusion proteins was tested by UV cross-linking using uniformly labeled RNAs containing MSEs 1 to 4, MSE1, or MSEs 2 to 4 from the chicken cTNT gene. All three CELF proteins bound to MSEs 1 to 4 and to MSEs 2 to 4 but not to an upstream fragment that contains MSE1 alone (Fig. 3), demonstrating that the binding of CELF proteins is restricted to the downstream intron. This binding is not due to the GST portion of the fusion proteins, as CELF fusion proteins made with other epitope tags demonstrate the same sequence-specific binding (data not shown). In experiments performed in the presence of HeLa nuclear extracts, a nonspecific 100-kDa RNA binding protein demonstrated equivalent levels of binding to all sub-

strates, indicating that all three RNAs were intact and competent for binding. Therefore, the CELF proteins bind to a 142-nucleotide intronic region of the chicken cTNT gene that is sufficient for activation of exon inclusion in striated muscle (10, 50). Analysis using scanning mutations has identified multiple CELF protein binding sites within this region (unpublished data).

To determine whether different CELF family members can activate MSE-dependent exon inclusion, expression vectors for all five CELF proteins were cotransfected with the R35C minigene that contains a heterologous alternative exon flanked by MSEs 1 to 4 into QT35 quail fibroblast cells (Fig. 4A). This exon is appropriately regulated in embryonic skeletal muscle cultures and is predominantly skipped in QT35 cells, as in other nonmuscle cells (10, 50). Western blots probed with antibodies against the N-terminal epitope tag demonstrated that all of the expression vectors express proteins of the expected sizes at comparable levels and that increasing amounts of expression plasmid resulted in increasing amounts of expressed proteins (data not shown). Cotransfection with each of the five CELF proteins enhanced the level of exon inclusion in a dose-responsive manner (Fig. 4B). To determine whether enhanced exon inclusion by CELF proteins is MSE dependent, 300 ng of each CELF expression plasmid was cotransfected with a minigene containing the same alternative exon flanked by human β -globin intron 1 sequences which lack MSEs (Fig. 4C). The extent of exon inclusion in the presence of 300 ng of CUG-BP, CELF5, CELF4, or CELF3 expression plasmid did not differ from the level observed for the minigene alone (Fig. 4D). Exon inclusion was elevated by the addition of 300 ng of ETR-3 expression plasmid but remained well below levels observed with the MSE-containing minigene. Therefore, the CELF family members promote the MSE-dependent exon in vivo on RNA substrates to which they bind in a sequence-specific manner in vitro.

Differential ability of CELF proteins to promote exon inclusion via MSE2 alone. We previously reported that multiple copies of a single, 39-nucleotide MSE, MSE2, is sufficient to drive robust muscle-specific inclusion of a heterologous exon (10). To determine whether the CELF proteins can promote exon inclusion via this element alone, CUG-BP, ETR-3, CELF3, CELF4, and CELF5 were cotransfected in QT35 fibroblasts with the M2/M2TB minigene, which contains a 59-nucleotide alternative exon flanked on both sides by three copies of MSE2 (Fig. 5A). Both CELF3 and CELF4 promoted exon inclusion (Fig. 5B). CELF5 also promoted exon inclusion, but to a much lesser extent. In contrast, neither CUG-BP nor ETR-3 affected M2/M2TB exon inclusion (Fig. 5B), even when up to 10 μg of expression plasmid was cotransfected with the minigene (data not shown). This difference in ability to promote exon inclusion via MSE2 alone indicates a functional division of the CELF protein family into two subgroups, consistent with the division suggested by the sequence analysis discussed above.

CELF protein expression is dynamic during striated muscle and brain development. The expression of CUG-BP, ETR-3, and CELF4 in striated muscle and brain and the ability of these proteins to regulate the splicing of cTNT, a transcript that undergoes developmental stage-specific alternative splicing, prompted us to investigate whether the expression of these

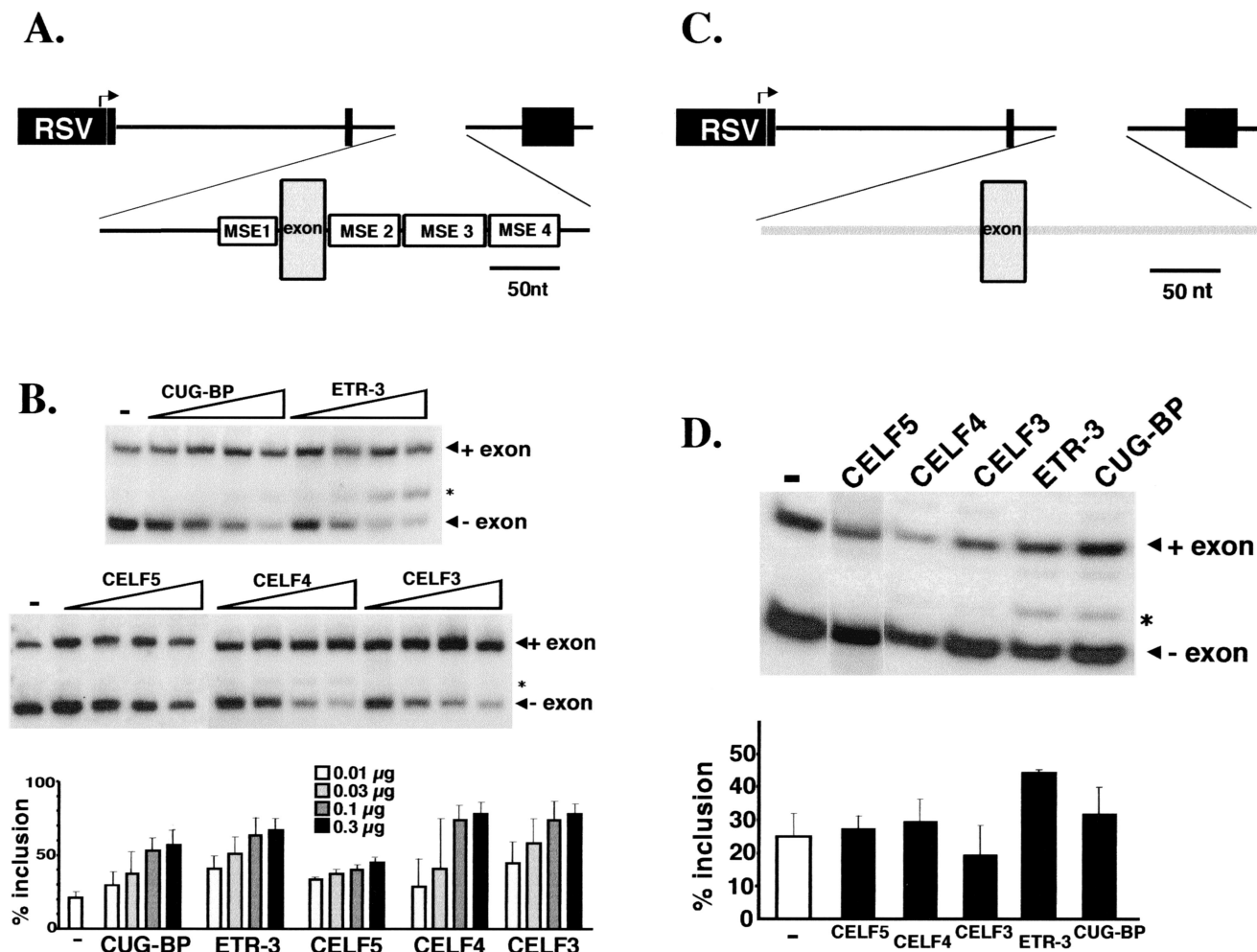


FIG. 4. CELF proteins promote MSE-dependent exon inclusion. (A) The MSE-containing R35C minigene contains an alternative exon flanked upstream by the last 99 nucleotides of cTNT intron 4, which includes MSE1, and downstream by the first 142 nucleotides of cTNT intron 5, which includes MSEs 2 to 4. This intron-exon-intron cassette is inserted between exons 2 and 4 of the constitutively spliced chicken skeletal troponin I gene. RSV, Rous sarcoma virus. (B) Quail QT35 fibroblasts were cotransfected with 2 μg of R35C and 0 (-), 10, 30, 100, or 300 ng of CUG-BP, ETR-3, CELF5, CELF4, or CELF3 expression plasmid. Increasing concentrations of expression vector showed increasing amounts of protein expression (data not shown). RNA was harvested after 48 h, and RT-PCR analysis was performed to determine the extent of exon inclusion. (C) The MSE-lacking substrate RTBPSRAX contains the same alternative exon as R35C flanked upstream by the last 78 nucleotides of human β-globin intron 1 and downstream by the first 96 nucleotides of the same intron. (D) QT35 cells were cotransfected with 2 μg of RTBPSRAX alone (-) and with 300 ng of CELF5, CELF4, CELF3, ETR-3, or CUG-BP expression plasmid. Asterisk, cryptic splice observed in some PCRs in which the exon is included but the 30-nucleotide exon upstream of the alternative exon is skipped. The appearance of this cryptic splice is inconsistent and represents at most 15% of spliced mRNAs. It is unknown why this cryptic splice is enhanced most efficiently by ETR-3, but enhancement does not appear to be MSE dependent as it is observed for both MSE-containing and MSE-lacking minigenes. Loading was adjusted so that approximately equal counts were loaded in each lane.

CELF proteins is developmentally regulated in muscle and brain. To examine the developmental profiles of the three widely expressed CELF proteins, Western blot analyses were performed on total protein samples collected from mice at different stages of embryonic and postnatal life using the antibodies against CUG-BP, ETR-3, and CELF4 (Fig. 6). All three family members displayed changes during development. The abundance of CUG-BP decreased significantly throughout skeletal muscle development to low levels in adult thigh muscle. CUG-BP remained constant during early stages of brain development and was reduced to a very low but detectable level in the adult brain. ETR-3 increased in abundance in

skeletal muscle and brain and underwent a transition from high-molecular-mass isoforms (primarily a band at approximately 52 kDa) in embryonic limb to predominantly lower-molecular-mass isoforms (approximately 42 and 50 kDa) in adult thigh. CELF4 levels increased in adult skeletal muscle without any apparent change in isoform distribution, while a low-molecular-mass isoform (approximately 42 kDa) is lost in adult brain.

Changes in ETR-3 protein correlate with developmental changes in cTNT splicing in cardiac and skeletal muscle. To test whether changes in CELF proteins occur concomitantly with changes in cTNT splicing during development, Western

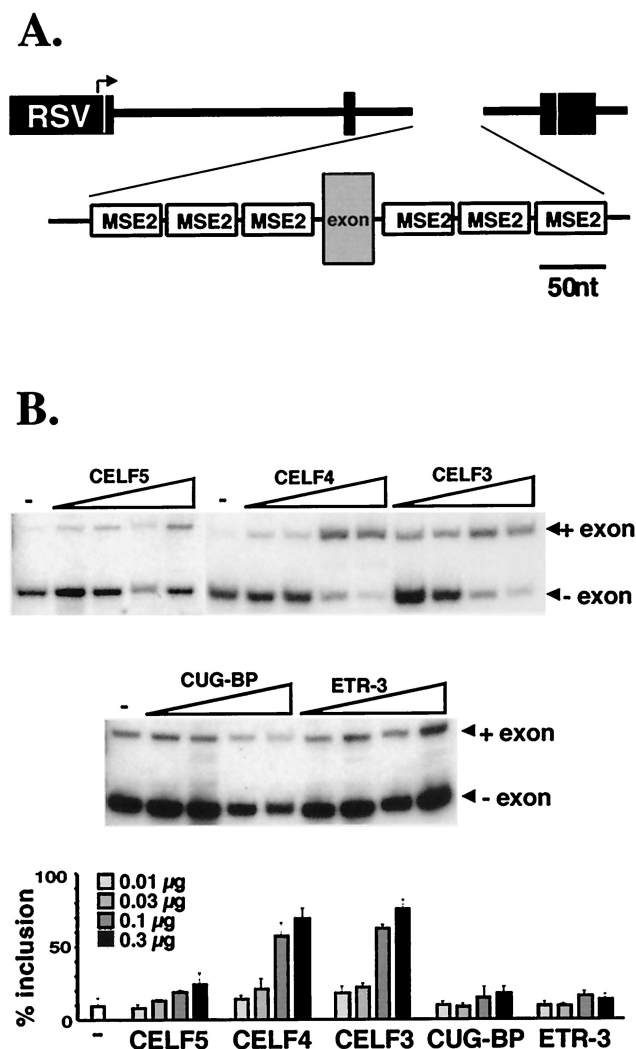
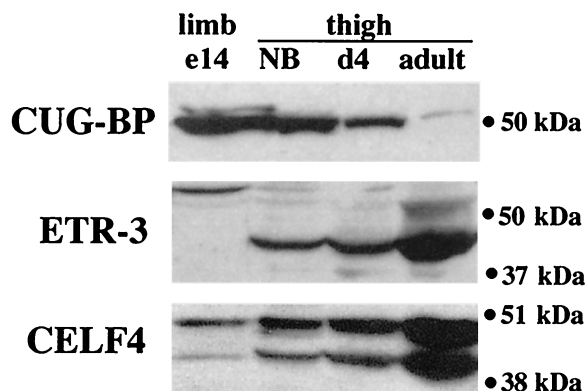


FIG. 5. CELF3, CELF4, and to a lesser extent CELF5, but not CUG-BP or ETR-3, can promote exon inclusion via MSE2 alone. (A) The M2/M2TB minigene contains a 52-nucleotide exon flanked upstream and downstream by three copies of the 39-nucleotide cTNT MSE2 from intron 5. RSV, Rous sarcoma virus. (B) QT35 cells were cotransfected with 2 μ g of the M2/M2TB minigene and 0 (–), 10, 30, 100, or 300 ng of CELF5, CELF4, CELF3, CUG-BP, or ETR-3 expression plasmid. RT-PCR analysis was performed on RNA harvested 48 h posttransfection. Loading was adjusted so that approximately equal counts were loaded in each lane.

blot analyses were performed on total protein extracts from heart at different stages of development using antibodies against CUG-BP, ETR-3, and CELF4. Because the developmental switch in cTNT splicing is conserved among species, whole-tissue extracts were collected from both chickens and mice at various stages of embryonic and postnatal life. Primer extension or RT-PCR analysis was performed on parallel tissue samples to determine the extent of cTNT exon 5 inclusion over the same time course. In both chickens and mice, a dramatic transition from high- to low-molecular-weight ETR-3 isoforms concomitant with a switch from cTNT exon inclusion to exon skipping was observed (Fig. 7). Preliminary experiments indicate that this transition is not due to changes in

A. Skeletal Muscle



B. Brain

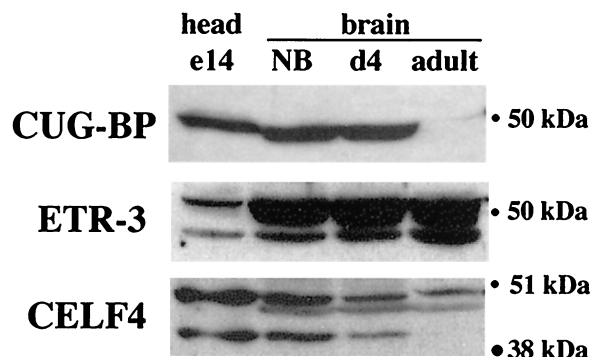


FIG. 6. Developmental profile of CELF proteins in skeletal muscle and brain. Western blot analysis was performed on total protein samples extracted from embryonic-day-14 (e14), newborn (NB), postnatal-day-4 (d4), and adult pregnant or 0- to 4-day-postpartum female Swiss Webster mice. Fifty micrograms of total protein was loaded in each lane. In brain samples, high levels of a 30-kDa protein are observed on CELF4 blots (data not shown), suggesting that a high degree of tissue-specific cleavage or degradation of CELF4 may be occurring in brain as described for Fig. 2.

phosphorylation, and we have cloned several alternatively spliced forms of ETR-3 from mouse heart tissue (data not shown). CUG-BP protein levels are somewhat reduced in the adult heart in both chickens and mice but do not change concomitantly with changes in cTNT splicing (data not shown). CELF4 is also slightly reduced in the adult chicken heart but does not undergo any apparent changes in the avian embryo or in the developing or adult mouse heart (data not shown). Therefore, the conserved developmental switch in cTNT splicing occurs concomitantly with a conserved switch in ETR-3 isoforms.

Muscle-specific splicing is induced as part of the myogenic program, since it can be induced in fibroblasts by expression of the myogenic regulators MyoD and myogenin (50). To monitor cTNT splicing during muscle differentiation, we transfected the R35C minigene containing MSEs 1 to 4 into C2C12 mouse

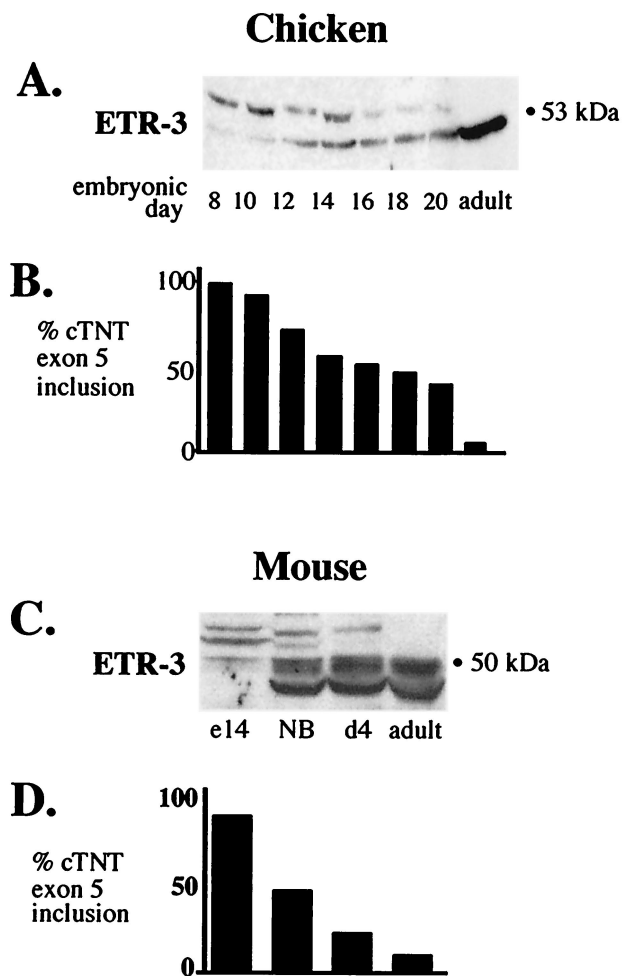


FIG. 7. An ETR-3 isoform switch that correlates with cTNT splicing is conserved in avian and mammalian heart development. (A) ETR-3 Western blot of total protein samples extracted from embryonic and adult chicken hearts. Sixty micrograms of total proteins was loaded in each lane. (B) Results from primer extension of endogenous cTNT mRNA from the same time course using a ³²P-labeled oligonucleotide complementary to exon 6. (C) ETR-3 Western blot of total protein samples from embryonic-day-14 (e14), newborn (NB), postnatal-day-4 (d4), and adult mouse hearts. Fifty micrograms of total protein was loaded per lane. (D) RT-PCR of endogenous cTNT was performed over the same time course and quantitated by PhosphorImager.

myoblast cells induced to differentiate and monitored the extent of exon inclusion over time. An exogenous minigene was used to examine splicing during myoblast differentiation because endogenous cTNT mRNA cannot be detected until differentiation is well under way. As shown in Fig. 8B, cTNT exon inclusion is strongly induced during C2C12 differentiation. This effect is MSE dependent as alternative exons flanked by introns that lack MSEs maintain a constant level of exon inclusion (data not shown). Western blot analysis on total protein extracts from differentiating C2C12 cultures demonstrates a dramatic increase in ETR-3 protein expression during the same time course (Fig. 8A). CUG-BP and CELF4 protein levels were invariant during differentiation (41; data not shown).

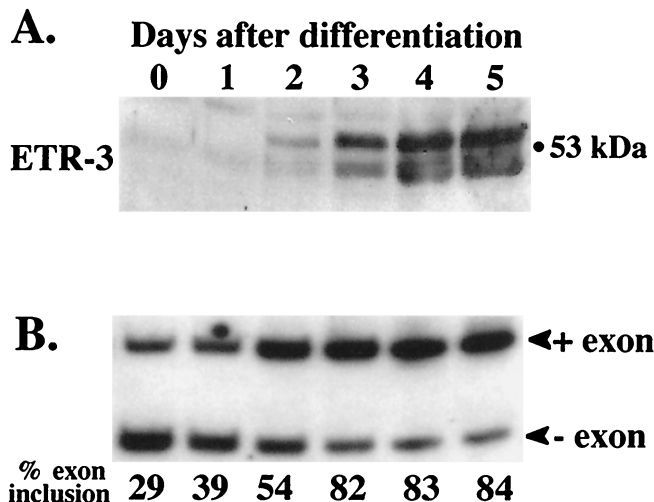


FIG. 8. Induction of ETR-3 protein expression and cTNT exon inclusion occur simultaneously during skeletal muscle differentiation. (A) ETR-3 Western blot of total protein extracts from C2C12 murine myoblast cells induced to differentiate. Fifty micrograms of total protein was loaded in each lane. (B) RT-PCR analysis of cTNT minigene RNA from similar cultures transfected with R35C harvested over the same time course. Loading was adjusted so that approximately equal counts were loaded in each lane.

DISCUSSION

A novel family of splicing regulators. In this study, we provide evidence that members of the CELF family of RNA binding proteins are regulators of cell-specific alternative splicing. This family includes CUG-BP, a protein that has previously been reported to bind to a conserved intronic splicing element and regulate alternative splicing in a sequence-dependent manner (46). By screening human EST databases, we identified four paralogs of CUG-BP including ETR-3, originally identified in a fetal heart library (25), and three novel factors we called CELF3, CELF4, and CELF5. Following submission of the manuscript of this report, another group reported the identification of these proteins and called them Bruno-like (BRUNOL) proteins (17). CUG-BP, ETR-3, and CELF4 all bind to RNA substrates containing chicken cTNT MSEs, and all five CELF proteins promote MSE-dependent exon inclusion in chicken cTNT minigene pre-mRNAs in quail fibroblast cells. The ability of the human CELF proteins to regulate MSE-dependent alternative splicing of avian cTNT pre-mRNAs, the conservation in sequence and position of MSE2 in chicken and human cTNT (46), and the high degree of conservation in the splicing pattern of cTNT transcripts in birds and mammals (11, 28, 39, 60) all indicate that MSE-dependent regulation of alternative splicing by the CELF family is a conserved mechanism of regulating gene expression during development in vertebrates.

Several muscle-specific genes whose pre-mRNAs undergo regulated alternative splicing contain potential MSE sequences (10), suggesting that the CELF family may be involved in the regulation of the alternative splicing of a subset of genes in muscle. Although the total abundance of the three widely expressed CELF proteins was particularly high in striated muscle tissues, every tissue examined by Western blot analysis

contains at least one CELF family member. Dot blots indicate that CELF3 and CELF5 are restricted to brain. Together, these results suggest that each CELF protein may play a distinct role in alternative splicing in different tissues or act on different target pre-mRNAs within tissues where more than one family member is present.

The CELF proteins differ in their abilities to activate splicing by MSE2 alone. We have previously shown that CUG-BP binds to an MSE2-like sequence downstream of human cTNT exon 5 in correlation with regulation *in vivo* (46). CUG-BP (and ETR-3) failed, however, to activate inclusion of an exon flanked by multiple copies of MSE2 from the chicken cTNT gene. It should be noted that, in the previous study, MSE2 was in the context of the natural introns, which presumably contain additional regulatory elements. Although the binding of CUG-BP to the human MSE2-like sequence is clearly required for regulation, the results presented in Fig. 5 indicate that MSE2 binding is not sufficient. Additional elements flanking MSE2, different spacing between elements, or elements within human MSE2 not found in chicken MSE2 are required for CUG-BP function.

CELF proteins as regulators of alternative splicing during normal development and disease. The shift in alternative splicing of cTNT exon 5 from exon inclusion in the embryo to exon skipping in adult striated muscle is highly conserved in birds and mammals (11, 28, 39, 60). Western blot analysis demonstrates that in both chicken and mouse heart, ETR-3 undergoes a developmental transition from high- to low-molecular-weight isoforms concomitant with the switch to cTNT exon 5 skipping. CUG-BP and CELF4 remain largely invariant over this same time course, strongly suggesting that ETR-3 may be the primary regulator of cTNT alternative splicing *in vivo*. Preliminary evidence suggests that the embryonic and adult ETR-3 isoforms observed in heart are generated by alternative splicing, and at least some of the larger ETR-3 isoforms expressed in embryonic mouse heart are also present in differentiated C2C12 cells (A. N. Ladd and T. A. Cooper, unpublished observations). Different splice forms of ETR-3 may differ in their abilities to recognize or bind cTNT RNA, their abilities to promote exon 5 inclusion, or their nuclear and cytoplasmic localizations. We have cloned several different ETR-3 splice variants and are currently addressing this issue. It is possible that a different splicing factor is responsible for regulating both the switch in cTNT splicing and the transition in ETR-3 isoforms during heart development. ETR-3 mRNAs contain potential binding sites for CUG-BP and ETR-3 (36), which raises the possibility of CELF protein autoregulation or cross talk in the developing heart, where CUG-BP, ETR-3, and CELF4 are all highly expressed. There is also a striking correlation, however, between the level of ETR-3 protein (but not CUG-BP or CELF4) and the extent of cTNT exon 5 inclusion during differentiation of myoblasts into myotubes. Preliminary studies of variants cloned from mouse heart indicate that high-molecular-weight isoforms promote cTNT exon inclusion (A. N. Ladd and T. A. Cooper, unpublished observations). Both high- and low-molecular-weight isoforms of ETR-3 protein are induced concomitantly with the increase in cTNT exon inclusion during C2C12 differentiation. Therefore, in two independent systems (heart development and skeletal muscle differentiation), changes in cTNT alternative splicing directly

correlate with changes in the expression of high-molecular-weight ETR-3 protein isoforms, ETR-3 shows sequence-specific binding to a conserved intronic element demonstrated to be required for regulated splicing of cTNT exon 5, and coexpression of ETR-3 with cTNT minigenes promotes element-dependent exon inclusion in fibroblasts. Together, these results strongly suggest that ETR-3 plays a major role in cTNT alternative splicing regulation.

cTNT is likely to be just one of multiple targets of CELF protein regulation. Although neither CUG-BP nor CELF4 expression correlated with changes in cTNT splicing in the heart, both underwent developmental changes in skeletal muscle. The abundance of CUG-BP decreased significantly during high muscle development, while CELF4 levels increased dramatically during this same period. CUG-BP, ETR-3, and CELF4 also displayed changes in expression during brain development. This is consistent with their involvement in the developmental-stage-specific regulation of other target pre-mRNAs. CUG-BP has been suggested to play a role in neuron-specific splicing of an exon in the γ_2 of GABA_A transcripts; GABA_A gene expression is known to undergo changes during brain development (67). The identification of other CELF family target genes will be an important goal for future studies.

CELF family members are likely to play a critical role during development. ETR-1, a putative homolog of CUG-BP in *Caenorhabditis elegans*, is essential for muscle differentiation (42). CUG-BP has also been implicated in the pathogenesis of myotonic dystrophy (59), a neuromuscular disease caused by an expansion of an unstable region of CTG repeats in the 3' UTR of the DMPK gene (2, 15, 37). Congenital patients have numerous developmental problems, including arrested or delayed muscle maturation (55), the absence or reduction of some muscle fiber types, and mental retardation (22). Overexpression of DMPK transcripts containing expanded CUG repeats suppresses myogenic differentiation in culture (51, 61), suggesting a direct link between CUG expansion and the developmental defects observed in congenital patients.

The importance of CELF family function is probably not diminished in adulthood. DM1 is the most common form of adult onset muscular dystrophy and is associated with manifestations in several different tissues. The most affected are striated muscle and brain, the tissues that demonstrate the highest levels of CELF protein expression. It has been proposed that the expanded CUG repeats in the DMPK transcripts of affected individuals create a gain-of-function mutation that affects the processing of other mRNAs by disrupting RNA processing proteins (62). Consistent with this model, DMPK is actively transcribed and both expanded DMPK transcripts (12) and the hypophosphorylated isoform of CUG-BP (59) accumulate in the nuclei of DM1 cells. The splicing of cTNT is also disrupted in striated-muscle cultures from DM1 patients and in normal cells expressing transcripts with large CUG repeats (46), consistent with the altered function of CELF proteins. If CELF family function is disrupted in DM1, the missplicing of a battery of target pre-mRNAs including cTNT could explain many of the different manifestations of the disease in affected individuals.

Other potential functions of the CELF proteins. Phylogenetic analysis indicates that the CELF family is most closely related to the Hu family, which regulates mRNA stability and

translation (reviewed in reference 32). Embryo deadenylation element binding protein, a homolog of CUG-BP identified in *Xenopus laevis*, has been shown to regulate mRNA deadenylation (45). Bruno, a *Drosophila* protein similar to CUG-BP, regulates translation of *oskar* mRNAs not localized to the posterior pole of the oocyte (33). Based on the similarities of the mammalian CELF proteins to these factors, it is likely that CELF protein function is not limited to regulation of alternative splicing. Other splicing regulatory factors have been implicated in transport or processing of mRNAs within the cytoplasm, including PTB (29), hnRNP A1 (47), and several SR proteins (6). CUG-BP is known to be localized in both the nucleus and cytoplasm (49), consistent with its having roles in both nuclear and cytoplasmic RNA processing events such as splicing, translation, regulation of mRNA stability, and mRNA shuttling. Indeed, the loss of cytoplasmic CUG-BP in DM1 cells may contribute as much to the pathogenesis of DM1 through changes in translation or mRNA stability as the changes in alternative splicing caused by its accumulation in the nucleus.

ACKNOWLEDGMENTS

We thank Lubov Timchenko, C. C. Liew, and Maurice Swanson for kind gifts of reagents or clones, Claire Lo, Gopal Singh, and Wade Haaland for their technical assistance, and Sue Berget, Rajesh Savkur, Chris Smith, and Maurice Swanson for helpful discussions on the manuscript.

This work was supported by grants to T.A.C. from the Muscular Dystrophy Association and NIH (AR45653). A.N.L. was supported by a postdoctoral NRSA fellowship from the NIH.

REFERENCES

- Black, D. L. 1992. Activation of c-src neuron-specific splicing by an unusual RNA element in vivo and in vitro. *Cell* **69**:795–807.
- Brook, J. D., M. E. McCurrach, H. G. Harley, A. J. Buckler, D. Church, H. Aburatani, K. Hunter, V. P. Stanton, J. P. Thirion, T. Hudson, R. Sohn, B. Zeman, R. G. Snell, S. A. Rundle, S. Crow, J. Davies, P. Shelbourne, J. Buxton, C. Jones, V. Juvonen, K. Johnson, P. S. Harper, D. J. Shaw, and D. E. Housman. 1992. Molecular basis of myotonic dystrophy—expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**:799–808.
- Buckanovich, R. J., and R. B. Darnell. 1997. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol. Cell. Biol.* **17**:3194–3201.
- Buckanovich, R. J., J. B. Posner, and R. B. Darnell. 1993. Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron* **11**:657–672.
- Burd, C. G., and G. Dreyfuss. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**:615–621.
- Caceres, J. F., G. R. Sreaton, and A. R. Krainer. 1998. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev.* **12**:55–66.
- Choi, D. K., T. Ito, Y. Mitsui, and Y. Sakaki. 1998. Fluorescent differential display analysis of gene expression in apoptotic neuroblastoma cells. *Gene* **223**:21–31.
- Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**:156–159.
- Cooper, T. A. 1992. In vitro splicing of cardiac troponin T precursors. Exon mutations disrupt splicing of the upstream intron. *J. Biol. Chem.* **267**:5330–5338.
- Cooper, T. A. 1998. Muscle-specific splicing of a heterologous exon mediated by a single muscle-specific splicing enhancer from the cardiac troponin T gene. *Mol. Cell. Biol.* **18**:4519–4525.
- Cooper, T. A., and C. P. Ordahl. 1985. A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternate splicing. *J. Biol. Chem.* **260**:11140–11148.
- Davis, B. M., M. E. McCurrach, K. L. Taneja, R. H. Singer, and D. E. Housman. 1997. Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl. Acad. Sci. USA* **94**:7388–7393.
- Dignam, J. D., R. M. Lebovitz, and R. G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**:1475–1489.
- Fu, X. D. 1995. The superfamily of arginine/serine-rich splicing factors. *RNA* **1**:663–680.
- Fu, Y. H., A. Pizzuti, R. G. Fenwick, J. King, S. Rajnarayan, P. W. Dunne, J. Dubel, G. A. Nasser, T. Ashizawa, P. Dejong, B. Wieringa, R. Korneluk, M. B. Perryman, H. F. Epstein, and C. T. Caskey. 1992. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**:1256–1258.
- Godt, R. E., R. T. H. Fogaca, I. K. Silva, and T. M. Nosek. 1993. Contraction of developing avian heart muscle. *Comp. Biochem. Physiol. [A]* **105**:213–218.
- Good, P. J., Q. Chen, S. J. Warner, and D. Herring. 2000. A family of human RNA-binding proteins related to the *Drosophila* Bruno translational regulator. *J. Biol. Chem.* **275**:28583–28592.
- Gooding, C., G. C. Roberts, G. Moreau, B. Nadal-Ginard, and C. W. Smith. 1994. Smooth muscle-specific switching of alpha-tropomyosin mutually exclusive exon selection by specific inhibition of the strong default exon. *EMBO J.* **13**:3861–3872.
- Guo, W., G. J. Mulligan, S. Wormsley, and D. M. Helfman. 1991. Alternative splicing of beta-tropomyosin pre-mRNA: cis-acting elements and cellular factors that block the use of a skeletal muscle exon in nonmuscle cells. *Genes Dev.* **5**:2096–2107.
- Hanamura, A., J. F. Caceres, A. Mayeda, B. R. Franza, Jr., and A. R. Krainer. 1998. Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA* **4**:430–444.
- Hanke, J., D. Brett, I. Zastrow, A. Aydin, S. Delbruck, G. Lehmann, F. Luft, J. Reich, and P. Bork. 1999. Alternative splicing of human genes: more the rule than the exception? *Trends Genet.* **15**:389–390.
- Harper, P. S. 1998. Myotonic dystrophy as a trinucleotide repeat disorder—a clinical perspective, p. 115–130. *In* R. D. Wells and S. T. Warren (ed.), *Genetic instabilities and hereditary neurological diseases*. Academic Press, Boston, Mass.
- Higgins, D. G., and P. M. Sharp. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *Comput. Appl. Biosci.* **5**:151–153.
- Hofmann, K., P. Bucher, L. Falquet, and A. Bairoch. 1999. The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27**:215–219.
- Hwang, D. M., W. S. Hwang, and C. C. Liew. 1994. Single pass sequencing of a unidirectional human fetal heart cDNA library to discover novel genes of the cardiovascular system. *J. Mol. Cell Cardiol.* **26**:1329–1333.
- Ichida, M., H. Endo, U. Ikeda, C. Matsuda, E. Ueno, K. Shimada, and Y. Kagawa. 1998. MyoD is indispensable for muscle-specific alternative splicing in mouse mitochondrial ATP synthase gamma-subunit pre-mRNA. *J. Biol. Chem.* **273**:8492–8501.
- Jensen, K. B., B. K. Dredge, G. Stefani, R. Zhong, R. J. Buckanovich, H. J. Okano, Y. Y. Yang, and R. B. Darnell. 2000. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25**:359–371.
- Jin, J. P., and J. J. Lin. 1989. Isolation and characterization of cDNA clones encoding embryonic and adult isoforms of rat cardiac troponin T. *J. Biol. Chem.* **264**:14471–14477.
- Kaminski, A., S. L. Hunt, J. G. Patton, and R. J. Jackson. 1995. Direct evidence that polypyrimidine tract binding protein (PTB) is essential for internal initiation of translation of encephalomyocarditis virus RNA. *RNA* **1**:924–938.
- Kamma, H., D. S. Portman, and G. Dreyfuss. 1995. Cell type-specific expression of hnRNP proteins. *Exp. Cell Res.* **221**:187–196.
- Kanopka, A., O. Muhlemann, S. Petersen-Mahrt, C. Estmer, C. Ohrmalm, and G. Akusjarvi. 1998. Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature* **393**:185–187.
- Keene, J. D. 1999. Why is Hu where? Shuttling of early-response-gene messenger RNA subsets. *Proc. Natl. Acad. Sci. USA* **96**:5–7.
- Kim-Ha, J., K. Kerr, and P. M. Macdonald. 1995. Translational regulation of *oskar* mRNA by bruno, an ovarian RNA-binding protein, is essential. *Cell* **81**:403–412.
- Lin, C. H., and J. G. Patton. 1995. Regulation of alternative 3' splice site selection by constitutive splicing factors. *RNA* **1**:234–245.
- Lopez, A. J. 1998. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* **32**:279–305.
- Lu, X., N. A. Timchenko, and L. T. Timchenko. 1999. Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy. *Hum. Mol. Genet.* **8**:53–60.
- Mahadevan, M., C. Tsilfidis, L. Sabourin, G. Shuttler, C. Amemiya, G. Jansen, C. Neville, M. Narang, J. Barcelo, K. Ohoy, S. Leblond, J. Earle MacDonald, P. J. Dejong, B. Wieringa, and R. G. Korneluk. 1992. Myotonic dystrophy mutation—an unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**:1253–1255.
- Margolis, R. L., M. R. Abraham, S. B. Gatchell, S. H. Li, A. S. Kidwai, T. S. Breschel, O. C. Stine, C. Callahan, M. G. McInnis, and C. A. Ross. 1997. cDNAs with long CAG trinucleotide repeats from human brain. *Hum. Genet.* **100**:114–122.
- McAuliffe, J. J. 1994. Delineation of the cardiac troponin T expression pattern during murine development. *J. Cell. Biochem.* **18D**(Suppl.):W347.

40. McAuliffe, J. J., L. Z. Gao, and R. J. Solaro. 1990. Changes in myofibrillar activation and troponin C Ca²⁺ binding associated with troponin T isoform switching in developing rabbit heart. *Circ. Res.* **66**:1204–1216.
41. Miller, J. W., C. R. Urbinati, P. Teng-Umuay, M. G. Stenberg, B. J. Byrne, C. A. Thornton, and M. S. Swanson. 2000. Recruitment of human muscle-blind proteins to (CUG)_n expansions associated with myotonic dystrophy. *EMBO J.* **19**:4439–4448.
42. Milne, C. A., and J. Hodgkin. 1999. ETR-1, a homologue of a protein linked to myotonic dystrophy, is essential for muscle development in *Caenorhabditis elegans*. *Curr. Biol.* **9**:1243–1246.
43. Min, H., C. W. Turck, J. M. Nikolic, and D. L. Black. 1997. A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes Dev.* **11**:1023–1036.
44. Nakai, K., and M. Kanehisa. 1992. A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* **14**:897–911.
45. Paillard, L., F. Omilli, V. Legagneux, T. Bassez, D. Maniey, and H. B. Osborne. 1998. EDEN and EDEN-BP, a *cis* element and an associated factor that mediate sequence-specific mRNA deadenylation in *Xenopus* embryos. *EMBO J.* **17**:278–287.
46. Phillips, A. V., L. T. Timchenko, and T. A. Cooper. 1998. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280**:737–741.
47. Piñol-Roma, S., and G. Dreyfuss. 1992. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* **355**:730–732.
48. Polydorides, A. D., H. J. Okano, Y. Y. Yang, G. Stefani, and R. B. Darnell. 2000. A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. *Proc. Natl. Acad. Sci. USA* **97**:6350–6355.
49. Roberts, R., N. A. Timchenko, J. W. Miller, S. Reddy, C. T. Caskey, M. S. Swanson, and L. T. Timchenko. 1997. Altered phosphorylation and intracellular distribution of a (CUG)_n triplet repeat RNA-binding protein in patients with myotonic dystrophy and in myotonin protein kinase knockout mice. *Proc. Natl. Acad. Sci. USA* **94**:13221–13226.
50. Ryan, K. J., and T. A. Cooper. 1996. Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle. *Mol. Cell. Biol.* **16**:4014–4023.
51. Sabourin, L. A., K. Tamai, M. A. Narang, and R. G. Korneluk. 1997. Over-expression of 3'-untranslated region of the myotonic dystrophy kinase cDNA inhibits myoblast differentiation in vitro. *J. Biol. Chem.* **272**:29626–29635.
52. Saitoh, O., E. N. Olson, and M. Periasamy. 1990. Muscle-specific RNA processing continues in the absence of myogenin expression. *J. Biol. Chem.* **265**:19381–19384.
53. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
54. Sanford, J. R., and J. P. Bruzik. 1999. Developmental regulation of SR protein phosphorylation and activity. *Genes Dev.* **13**:1513–1518.
55. Sarnat, H. B., and S. W. Silbert. 1976. Maturation arrest of fetal muscle in neonatal myotonic dystrophy. A pathologic study of four cases. *Arch. Neurol.* **33**:466–474.
56. Schultz, J., F. Milpetz, P. Bork, and C. P. Ponting. 1998. SMART, a simple modular architecture research tool: identification of signalling domains. *Proc. Natl. Acad. Sci. USA* **95**:5857–5864.
57. Screaton, G. R., J. F. Caceres, A. Mayeda, M. V. Bell, M. Plebanski, D. G. Jackson, J. I. Bell, and A. R. Krainer. 1995. Identification and characterization of three members of the human SR family of pre-mRNA splicing factors. *EMBO J.* **14**:4336–4349.
58. Singh, R., J. Valcarcel, and M. R. Green. 1995. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**:1173–1176.
59. Timchenko, L. T., J. W. Miller, N. A. Timchenko, D. R. Devore, K. V. Datar, L. J. Lin, R. Roberts, C. T. Caskey, and M. S. Swanson. 1996. Identification of a (CUG)_n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res.* **24**:4407–4414.
60. Townsend, P. J., P. J. R. Barton, M. H. Yacoub, and H. Farza. 1995. Molecular cloning of human cardiac troponin T isoforms: expression in developing and failing heart. *J. Mol. Cell Card* **27**:2223–2236.
61. Usuki, F., S. Ishiura, N. Saitoh, N. Sasagawa, H. Sorimachi, H. Kuzume, K. Maruyama, T. Terao, and K. Suzuki. 1997. Expanded CTG repeats in myotonin protein kinase suppresses myogenic differentiation. *Neuroreport* **8**:3749–3753.
62. Wang, J., E. Pegoraro, E. Menegazzo, M. Gennarelli, R. C. Hoop, C. Angelini, and E. P. Hoffman. 1995. Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation. *Hum. Mol. Genet.* **4**:599–606.
63. Xie, W. Q., and L. I. Rothblum. 1991. Rapid, small-scale RNA isolation from tissue culture cells. *BioTechniques* **11**:324–327.
64. Xu, R., J. Teng, and T. A. Cooper. 1993. The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Mol. Cell. Biol.* **13**:3660–3674.
65. Zahler, A. M., K. M. Neugebauer, W. S. Lane, and M. B. Roth. 1993. Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science* **260**:219–222.
66. Zhang, L., M. Askiya, T. G. Sherman, and P. J. Grabowski. 1996. Essential nucleotides direct neuron-specific splicing of gamma-2 pre-mRNA. *RNA* **2**:682–698.
67. Zhang, L., W. Liu, and P. J. Grabowski. 1999. Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB. *RNA* **5**:117–130.
68. Zhang, W.-J., and J. Y. Wu. 1996. Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. *Mol. Cell. Biol.* **16**:5400–5408.