# Identification of Two Nuclear Proteins Which Bind to RNA CUG Repeats: Significance for Myotonic Dystrophy

Satyakam Bhagwati, Ashwini Ghatpande, and Betty Leung

Department of Neurology, State University of New York, HSCB, 450 Clarkson Ave, Brooklyn, New York 11203

Received September 27, 1996

Myotonic dystrophy is caused by a trinucleotide repeat expansion (CTG)n, located in the 3' untranslated region of the DM-protein kinase gene. The cellular effects of the CTG expansion and how they lead to the diverse, multi-system clinical phenotype of DM are unknown. Studies on the expression of the DM gene in affected tissue have not yielded consistent results, leading to the suggestion that alterations of DM-PK may not be the sole molecular basis for DM. We explored the hypothesis that the expanded repeat in mutant DM RNA (CUG)n binds and titrates out nuclear RNA binding proteins. Alterations in the normal function of these proteins could result in the disruption of important cellular processes. We report here the identification and magnetic bead affinity purification of two nuclear proteins, 35 and 25 kD in size, which bind to RNA (CUG)n repeats and to a varying extent with other pyrimidine rich sequences. Sequence analysis of the 35 kD protein shows that it is a novel protein. Both these proteins are widely expressed, including human brain and skeletal muscle. We speculate that these proteins may play a role in DM pre-mRNA processing or nuclear cytoplasmic trafficking of RNA. Studies into the function of these proteins should yield important insights into the complex pathogenesis of myotonic dystrophy.

Myotonic dystrophy (DM) is one of the most prevalent autosomal dominant inherited diseases in adults (incidence 1 in 8000) (1,2). It is characterized clinically by a complex multisystem involvement. Briefly these include the congenital form of DM with neonatal hypotonia, facial diplegia, respiratory distress and mental retardation and the adult form of DM with myotonia, progressive weakness and wasting of skeletal muscles, cardiac conduction defects, smooth muscle involvement, ocular cataracts, testicular atrophy and mental changes (1,2). The mutation responsible for DM has been identified as an expansion in the length of a CTG (5-40) trinucleotide repeat at chromosome 19q 13.3 (3,4,5). This triplet repeat is located in the 3' non-coding exon of a gene which encodes a putative serine-threonine kinase (Myotonin protein kinase). In normal individuals the most common alleles observed have less than 20 repeats although rarer alleles with up to 40 repeats have been found. In DM patients the number of CTG repeats is expanded beyond that seen in normals, carrying from 50 to many thousands of repeats and the size of the CTG repeat is directly correlated with the severity of the disease (1,2).

The cellular effects of the CTG expansion are unknown. Since the DM CTG expansion mutation is located in the 3' untranslated region of the DM gene, it should have no effect on the primary sequence of the protein product. Results on the studies of the effect of the repeat expansion have been conflicting. Some studies have shown decreased expression of DM RNA and protein (6-9) or unaltered transcription and translation of the DM gene in DM muscle (10) while in contrast it has also been reported that in congenital DM cases there is increased expression of the mutant allele (11). The variability in the results of these studies probably reflects the many different experimental systems and methods used to test for RNA and protein levels.

Since it is difficult to reconcile the dominant inheritance of DM with simple loss of function of the mutant allele and because of the discordant results of the expression of the DM-PK

gene in patient tissues, it has been proposed that the DM mutation may affect the function of other genes. One hypothesis proposes that this could be secondary to altered transcription of an adjacent gene due to alterations in chromatin topology related to expanded blocks of CTG repeats (12,13). However, although a large homeodomain encoding gene and another gene called DMR-NR (14) have been identified in close proximity to the CTG repeat, no abnormalities in the expression of these genes have been reported in tissue from DM patients.

In this study we explored the alternative hypothesis that mutant expanded DM RNA may act in *trans* to confer an abnormal function to the cell. The expanded CUG repeat region in mutant DM RNA may titrate out RNA binding proteins impairing essential cellular functions. In this paper we describe our search for and successful identification of two novel widely expressed nuclear proteins which bind to RNA (CUG)n repeats. Disruption of the normal function of these proteins may lead to the diverse manifestations of DM.

### **METHODS**

Whole cell, nuclear, and cytoplasmic protein extracts. These extracts were obtained from cultured human fetal myoblasts (Clonetics, CA) and human fibroblasts (Coriell Cell Repository, NJ) as described (15).

Tissue protein extracts. These extracts were prepared from human brain and skeletal muscle and the following tissues from 6 months old mice: skeletal muscle, heart, liver, kidney and lung as described (20).

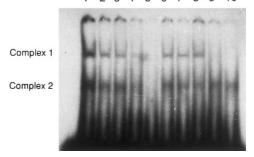
Electrophoretic mobility shift assay. The binding reactions contained 40 mM Hepes,pH 7.6, 100mM KCl, 2mM DTT, 10% glycerol and 2mM EDTA, 1 ug of E.Coli tRNA and 0.5 to 30 ug of protein extracts (protein extracts from human myoblasts, fibroblasts, tissue extracts or purified protein samples in various experiments). Samples were preincubated at room temperature for 5 min before radiolabeled DNA or RNA oligonucleotide probes (50000 cpm) were added. Specificity of binding was determined by adding excess amounts of unlabeled oligonucleotides to the binding reaction. The binding reaction was incubated at room temperature for 30 min and the reactions were electrophoresed on 0.5× TBE polyacrylamide gels, gels were dried and exposed to film.

Ultraviolet cross linking assay. 32P-labeled oligoribonucleotide probes (~50 pM) [(CUG)10, (CU)15 or (AG)15], were incubated with 30 to 50 ug of nuclear protein extract from myoblasts or fibroblasts or human brain tissue protein extracts. 2 ug of actin (Sigma) or albumin were used in control experiments. Incubations were done in 30 ul of buffer (10 mM Tris-HCl,pH 7.4, 100mM KCl, 2.5 mM MgCl2, 0.1% TritonX-100) and incubated at 25 C for 15 min. Reactions were crosslinked in a Stratagene Stratalinker 1800 UV cross linker at 3400 uW/cm2 for 45 min. Free RNAs were digested by incubation with a mixture of 20 ug of RNase A and 200 units of RNase T1 at 37 C for 1 hour. Cross linked RNA-protein complexes were separated on a 4-20% gradient Tris-glycine SDS polyacrylamide gel (Novex) by electrophoresis, dried and exposed to film.

Purification of CUG binding proteins. CUG binding proteins were isolated using Dynabeads coated with streptavidin (Dynabeads M-280 streptavidin, Dynal, NY) and a (CUG)10 probe biotinylated at the 5' end as described (16). Briefly 1200 picomoles of biotinylated (CUG)10 (obtained from Oligos,etc or BRL) was bound to 500 ul of Dynabeads streptavidin (5 mg) as described by the manufacturer (Dynal, Inc, NY). Control experiments were done with streptavidin beads alone. Whole cell protein from human myoblasts or fibroblasts (200 ug to 1mg) were used in separate experiments for binding to biotinylated CUG Dynabeads in TGED buffer ( 20mM Tris-HCl,pH 8, 1mM EDTA, 10% glycerol,1 mM DTT, 0.01% Triton X-100) with varying amounts of NaCl (50 mM, 100mM, 150mM, 200mM, 250mM). Quantitative adsorption was seen as the disappearance of specific RNA binding activity in the 'supernatant'.150mM NaCl concentration was found to be the optimal for protein binding and was used in all subsequent experiments. To eliminate non-specific binding of proteins, beads were washed once in TGED buffer with 150 mM NaCl, twice in TGED buffer containing a 10 fold excess of E.Coli tRNA and finally once with TGED buffer containing 150 mM NaCl. The bound protein was eluted with 50 ul of TGED buffer containing 1 M NaCl for 15 min on ice. Sample buffer was added to eluted proteins and samples electrophoresed on 8 % or 4-20% gradient Trisglycine gels. Proteins were stained using a silver staining kit (Novex, CA). Purifications were done several times using different protein preparations. Control experiments were done with beads carrying biotinylated CU oligonucleotides or with beads alone.

Protein sequencing. Purified proteins were separated by electrophoresis on 8% Tris glycine SDS gels and transferred to a PVDF membrane (Applied Biosystems) using 25mM Tris. 192 mM glycine, 20% methanol, pH 8.3 transfer

RNA (CUG)<sub>10</sub> Competition: - - - - - 5x 10x 25x 125x SS (CTG)<sub>10</sub> Competition: - 5x 10x 25x 125x - - - - - 1 2 3 4 5 6 7 8 9 10



**FIG. 1.** Autoradiogram of a mobility shift assay with a 30 mer ss (CTG)10 probe and human brain protein extract. Lanes 1 and 6 have no competition. Lanes 2 to 5 represent competition with increasing amounts of unlabeled ss DNA probe (CTG)10, (original magnification  $5 \times$ ,  $10 \times$ ,  $25 \times$ ,  $125 \times$ ), and lanes 7 to 10 represent competition with increasing amounts of unlabeled RNA (CUG)10 probe (original magnification  $5 \times$ ,  $10 \times$ ,  $25 \times$ ,  $125 \times$ ).

buffer. After transfer, proteins were stained with 0.1 5 Ponceau S in 1 % acetic acid for 5 min and destained in 5% acetic acid followed by several distilled water washes. Strips of membrane carrying the protein band of interest were sent to the Protein/DNA Technology Center, Rockfeller University for sequencing.

#### RESULTS

# Identification of RNA (CUG Repeat ) Binding Proteins

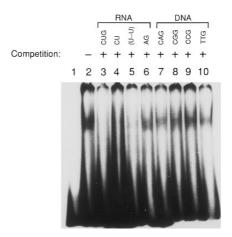
Proteins initially identified as single stranded DNA binding proteins have sometimes been found to also bind RNA (17). We decided therefore to test the relative affinity of protein binding to single stranded DNA (CTG)n or RNA (CUG)n probes on mobility shift assays. Human brain protein extract was incubated with radiolabeled single stranded (CTG)10 probe and competition with unlabeled (CTG)10 or (CUG)10 was done at 5,10, 25 or 125 fold excess (Fig 1). Two protein-CTG complexes were seen. Complex 1 was completely inhibited by unlabeled CUG probe (lane 9) and also by unlabeled CTG probe (lane 4) indicating dual binding to CTG and CUG repeats. Complex 2 was only inhibited by ss CTG and not by CUG and therefore identifies a single stranded CTG binding protein.

Next the presence of a (CUG)n RNA binding protein was confirmed using an end labeled oligoribonucleotide probe (CUG)10 in mobility shift assays with protein extracts from cultured normal human fibroblasts (Fig 2). Sequence specificity of the CUG binding protein was assessed by competition with different unlabeled oligonucleotide probes in 300 fold molar excess. The CUG protein complex could be completely inhibited by homologous (CUG) competition, and also interestingly by vast excess of other pyrimidine rich RNA oligonucleotides such as (CU)n and UU—UU, but only minimally by the purine oligoribonucleotide (AG)15 or by the single stranded DNA competition (CAG,CGG,CCG or TTG) (all 30 mer).

### Tissue Distribution and Cellular Localization of CUG Binding Proteins

In order to determine the tissue distribution of the CUG binding protein, we performed mobility shift assays on protein extracts from various mouse and human tissues. The CUG-protein complex was detected in human brain and skeletal muscle and mouse cardiac and skeletal muscle, liver and kidney but not in mouse lung (data not shown).

Next we sought to determine the cellular localization of this CUG binding protein in nuclear, cytoplasmic, and whole cell protein fractions from myoblasts by mobility shift assays. Results



in Fig 3 show that the CUG protein complex is formed using nuclear extracts but not using cytoplasmic protein extracts. The purity of the nuclear and cytoplasmic protein preparations used for this assay was confirmed by immunoblot analysis with an antibody to PCNA (proliferating cell nuclear antigen)(Sigma) which is expressed in the nucleus (data not shown).

# UV Crosslinking of RNA-Protein Complexes

In an attempt to identify the specific proteins which bind to CUG repeats, human brain protein extracts were incubated in separate reactions with 30 bases long (CUG)10, (CU)15 or (AG)15 labeled oligoribonucleotides and RNA-protein complexes were crosslinked by ultraviolet radiation and complexes analyzed by SDS-polyacrylamide gel electrophoresis. Only two proteins 35 kD and 25kD in size were consistently identified from all sources of protein extracts as binding specifically to CUG repeats but not to the AG or CU oligonucleotides (Fig

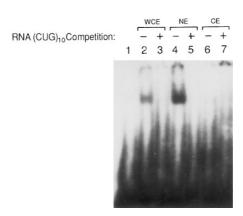


FIG. 3. Autoradiogram of mobility shift assay with a 30 mer RNA(CUG)10 probe and fibroblast whole cell, nuclear and cytoplasmic protein extracts. Probe alone (lane 1), whole cell extract (WCE) without (lane 2) and with unlabeled (CUG)10 competition (lane 3), nuclear protein extract (NE) without (lane 4) and with unlabeled (CUG)10 competition (lane 5), and cytoplasmic extract (CE) without (lane 6) and with (CUG)10 competition (lane 7).

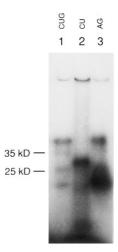


FIG. 4. UV light crosslinking assay. P32 labeled oligoribonucleotides were incubated with human brain protein extract and cross linked with UV light. Cross linked products were resolved by SDS-PAGE and autoradiography. Identity of each RNA molecule is indicated at the top of each lane and the 35 and 25 kD CUG binding proteins are indicated. Similar results were obtained with nuclear extracts from myoblasts and fibroblasts.

4). Similar results were obtained with nuclear protein extracts from myoblasts and fibroblasts. No bands were seen in control experiments done with actin and albumin.

# Magnetic RNA Affinity Purification of CUG and CTG Binding Proteins

In order to purify the CUG binding proteins we used streptavidin magnetic beads (Dynal, NY) coupled to biotinylated (CUG)10. Nuclear protein extracts from human fetal myoblasts and fibroblasts were used in separate purifications. Two proteins, 35 kD and 25 kD in size were consistently isolated from nuclear protein extracts from cultured human myoblasts and fibroblasts (Fig 5). RNA(CU)n carrying beads revealed many non-specific protein bands. No non-specific protein binding was seen using beads alone.

We attempted to separate these two proteins after electrophoresis and elution from SDS-polyacrylamide gels in a renaturation buffer. However mobility shift assays done with these eluted proteins did not retain binding activity presumably because of inefficient renaturation. We therefore analyzed the magnetic bead purified protein fraction containing both the 35 kD

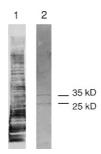


FIG. 5. Magnetic RNA affinity purification of CUG binding proteins from human myoblast nuclear proteins. Purified proteins were resolved by SDS-PAGE and silver stained for visualization. Unpurified protein extract (lane 1), purified protein (lane 2). The 35 and 25 kD proteins are indicated. Similar results were obtained with human fibroblasts.

and the 25 kD protein and confirmed binding to CUG repeats by mobility shift assays. Similar assays showed that these proteins also bound to CU and UU-UU and ssCTG repeat oligonucleotides but showed no affinity for GAC, AAC, AAU,GGG,GGA,GAA,CCC, AGA ribooligonucleotides (not shown). This shows that both proteins select a specific subset of pyrimidine rich RNA sequences for binding. Although we have identified two proteins using magnetic beads, only one CUG-protein complex was seen on mobility shift assays. This may indicate that both proteins participate in the same RNA-protein complex or that the two separate complexes cannot be separately resolved by gel electrophoresis.

### PROTEIN SEQUENCING

Partial N terminal amino acid sequencing of the 35 kD CUG binding protein has revealed a sequence of 12 amino acids, which does not have homology to any known protein.

### DISCUSSION

The mechanism by which the DM expansion mutation leads to the diverse clinical phenotype of DM is unknown. Knock out mice bred to homozygosity with no DM-PK expression show minimal changes in muscle and have none of the features of DM (18,19). Transgenic mice overexpressing DM-PK develop hypertrophic cardiomyopathy but do not develop any features of DM either (19). Also, no consistent abnormality in the level of expression of DM-PK in patient tissues has been described (6-11). These results would indicate that differences in DM-PK expression alone may not be crucial in disease pathogenesis and altered levels of DM-PK may not be the sole molecular basis for DM. It has been proposed that expanded CUG repeats in mutant DM RNA may titrate out RNA binding proteins and disturb their physiological functions.

We have identified, purified and partially characterized two nuclear proteins that bind to (CUG)n RNA. Characterization of the binding spectrum of these proteins showed that they bind to a specific subset of pyrimidine rich RNAs and ss CTG but not to a wide range of other purine rich RNAs. These proteins are 35 kD and 25 kD in size and are widely expressed in cultured human myoblasts and fibroblasts, and in mouse and human tissues including brain and skeletal muscle. Partial amino acid sequence analysis of the 35 kD protein has not revealed homology to any known protein. Two previous reports have described proteins that bind to specific subsets of DNA triplet repeat sequences (AGC, GGC, AGT) but not to others (GAC, AAC).(20,21). Only one previous report (15) has described in Hela cell protein extracts a (CUG)n RNA binding protein (40 to 50 kD) and a cytoplasmic protein which binds to both ss CTG and CUG, whose molecular weight could not be precisely determined (27). The 25 kD and 35kD CUG and ss CTG binding proteins we have detected are different proteins as they are detected in nuclear protein extracts. It is possible that many different nuclear or cytoplasmic proteins bind to the mutant DM CTG or CUG repeats and alter cellular functions in complex ways.

What could the role of these CUG binding proteins be in the pathogenesis of myotonic dystrophy. Recent studies(22-24)have found evidence for impaired processing of pre-mRNA in DM tissues. One study (22) found that although the levels of pre-mRNA products of mutant and wild type allele are normal the level of processed mRNA product of the DM allele in patient skeletal muscle was reduced relative to the normal allele. In another study (23) using fluorochrome conjugated (CTG)n and (CAG)n probes, bright foci of DM-PK transcripts were seen in the nuclei of DM patient fibroblasts, but not in normals, suggesting that they may represent aberrant processing or nucleo-cytoplasmic transport of DM RNA. Another study (24) found dramatic decreases in poly (A)+ RNA products of both normal and DM alleles with relatively unchanged total RNA in DM tissues, again suggesting impaired processing of both normal and mutant DM pre-mRNAs.

Some heterogenous nuclear riboproteins, (hnRNPs), which play important roles in RNA splicing preferentially interact with polypyrimidine rich sequences located adjacent to the 3' splice site (25). hnRNPs are known to bind to RNAs with a range of affinities (25,26). Binding sites that conform to a specific motif are bound most tightly and there is a commensurate loss of affinity for binding sites as they diverge from this sequence. We hypothesize that the presence of large CUG repeat sequences in mutant DM RNA within the nuclei of cells may lead to the aberrant binding of hnRNPs which normally have a preference and higher affinity for specific polypyrimidine rich sequences. This could lead to titration and decreased availability of that protein to bind to its physiological substrate. It could be conceptualized therefore in keeping with previous speculation (15,24) that alterations in the level of proteins important in pre-mRNA splicing could lead to profound effects on the metabolism of different development and tissue specific RNAs which share similar regulatory elements. Alternatively, CUG binding proteins may play a role in the polyadenylation of nascent transcripts or in RNA transport from nucleus to cytoplasm. Finally as these proteins are nuclear in location and also bind to ss CTG, they may play a role in the transcription of DM-PK or other adjacent genes.

Cloning and sequencing the genes encoding these two proteins and generating specific antibodies to study the expression of these proteins in normal and DM tissues should prove very useful in unraveling the complex pathogenesis of DM.

#### **ACKNOWLEDGMENT**

This work was supported by an NIH (NINDS) research grant to S.B.

#### REFERENCES

- 1. Ashley, C. T., and Warren, S. T. (1995) Annu. Rev. Genet. 29, 703-727.
- 2. Wieringa, B. (1994) Hum. Mol. Genet. 3, 1-7.
- 3. Mahadevan, M. S., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C. E., Narang, M., Barcelo, J., and O'Hoy, K. (1992) *Science* 255, 1253–1255.
- Fu, Y. H., Pizzuti, A., Fenwick, R. G., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., de Jong, P., Wieringa, B., Korneluk, R., Perryman, M. B., Epstein, H. F., and Caskey, C. T. (1992) Science 255, 1256–1258.
- Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburtani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., Sohn, R., Zemelman, B., Snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P. S., Shaw, D. J., and Housman, D. E. (1992) Cell 68, 799–808.
- 6. Carango, P., Noble, J. E., Marks, J. H., and Funange, V. L. (1992) Genomics 18, 340-348.
- 7. Hoffmann-Radvanyi, H., et al. (1993) Hum. Mol. Genet. 2, 1263-1266.
- 8. Novelli, G., et al. (1993) Biochem. Mol. Biol. Int. 29, 291-297.
- 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., and Caskey, C. T. (1993) Science 260, 235–238.
- 10. Bhagwati, S., Ghatpande, A., and Leung, B. (1996) Biochim. Biophys. Acta, (in press).
- 11. Sabourin, L. A., Mahadevan, M. S., Narang, M., Lee, D. S. C., Surh, L. C., and Korneluk, R. G. (1993) *Nat. Genet.* 4, 233–238.
- 12. Wang, Y. H., Amirhaeri, S., Kang, S., Wells, R. D., and Griffith, J. D. (1994) Science 265, 669-671.
- 13. Otten, A. D., and Tapscott, S. J. (1995) Proc. Natl. Acad. Sci. USA 92, 5465-5469.
- Boucher, C. A., King, S. K., Carey, N., Krahe, R., Winchester, C. L., Rahman, S., Creavin, T., Meghji, P., Bailey, M. E. S., Chartier, F. L., Brown, S. D., Siciliano, M. J., and Johnson, K. J. (1995) *Hum. Mol. Genet.* 4, 1919–1925.
- 15. Timchenko, L. T., Timchenko, N. A., Caskey, C. T., and Roberts, R. (1996) Hum. Mol. Genet. 5, 115-121.
- 16. Gabrielsen, O. S., and Huet, J. (1993) Methods in Enzymology 218, 508-525.
- 17. Ishikawa, F., Matunis, M. J., Dreyfuss, G., and Cech, T. R. (1993) Mol. Cell. Biol. 13, 4301-4310.
- Jansen, G., Groenen, P. T. A., Bachner, D., Jap, P. H. K., Coerwinkle, M., Oerlemans, F., Broek, W., Gohlsch, B., Pette, D., Plomp, J. J., Molenaar, P. C., Nederhoff, M. G. J., Echteld, C. J. A., Dekker, M., Berns, A., Hameister, H., and Wieringa, B. (1996) *Nat. Genet.* 13, 316–323.
- 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., Gordon, R. J., and Housman, D. (1996) *Nat. Genet.* 13, 325–335.

- 20. Yanagisawa, H. Y., Li, Y., Wang, H., and Kohwi, Y. (1995) Nucl. Acid Res. 23, 2654-2660.
- 21. Richards, R. I., Holman, K., Yu, S., and Sutherland, G. R. (1993) Hum. Mol. Genet. 2, 1429-1435.
- Krahe, R., Ashizawa, T., Abbruzzese, C., Roeder, E., Carango, P., Giacanelli, M., Funange, V. L., and Siciliano, M. J. (1995) Hum. Mol. Genet. 28, 1–14.
- Taneja, K. L., McCurrach, M., Schalling, M., Housman, D., and Singer, R. H. (1995) J. Cell Biol. 128, 995– 1002.
- 24. Wang, J., Pegoraro, E., Menegazzo, E., Gennarelli, M., Hoop, R. C., Angelini, C., and Hoffman, E. P. (1995) *Hum. Mol. Genet.* **4,** 599–606.
- 25. Dreyfuss, G., Matunis, M., Serafin, P., and Burd, C. G. (1993) Annu. Rev. Biochem. 62, 289-321.
- 26. Singh, R., Valcarcel, V., and Green, M. R. (1995) Science 268, 1173-1176.
- 27. Roscigno, R. F., Weiner, M., and Garcia-Blanco, M. A. (1993) J. Biol. Chem. 25, 11,222-11,229.