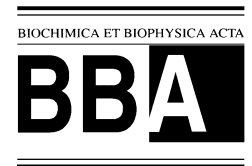




ELSEVIER

Biochimica et Biophysica Acta 1447 (1999) 107–112



www.elsevier.com/locate/bba

Short sequence-paper

Alternative processing of the human and mouse Raly genes¹

Irina Khrebtukova², Alexander Kuklin³, Richard P. Woychik², Edward J. Michaud^{*}

Life Sciences Division, Oak Ridge National Laboratory, P.O. Box 2009, Oak Ridge, TN 37831-8077, USA

Received 20 May 1999; accepted 22 July 1999

Abstract

A human homolog (RALY) of the mouse Raly gene was isolated and sequenced, and shown to encode a novel protein isoform containing a 16 amino acid in-frame insert in the variable region of the protein. Analysis of the corresponding region of the mouse Raly gene demonstrated that this novel protein isoform is also present in the mouse. Comparative analysis of RALY cDNA and EST sequences suggests the presence of additional alternatively processed RALY transcripts. As in the mouse, the human RALY gene is widely expressed as a 1.7-kb transcript. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: RALY; Alternative splicing; p542; hnRNP C

The mouse Raly (hnRNP associated with lethal yellow) gene, which was previously isolated and characterized in our laboratory [1] and by others [2], is closely linked to the agouti gene on chromosome 2. A spontaneous deletion of 170 kb, which includes the entire coding region of the Raly gene, causes the recessive embryonic lethality observed in the mouse mutation, lethal yellow [3]. Raly is a ubiquitously expressed gene of unknown function, which encodes a new member of the heterogeneous nuclear ribonucleoprotein (hnRNP) gene family, and is similar in sequence to human hnRNP C [1,2,4]. We were interested in the human homolog of Raly because it could play a similar critical role in human embryonic development. Here we describe the cloning of a human

homolog of Raly, different from previously published forms of the mouse and human Raly proteins, and also describe alternative processing of Raly in the human and mouse.

Using a cross-species hybridization approach, we isolated and sequenced RALY by screening a human testis cDNA library (Clontech) with a ³²P-labeled 5'-non-coding fragment of the mouse Raly cDNA, corresponding to nucleotides 87–169 [1]. The cDNA sequence of human RALY (hRALY, GenBank Accession number AF148457) is 1589 bp long and contains a 306-amino acid open reading frame (ORF) with putative 5'- and 3'-untranslated regions of 298 and 370 bp, respectively. Sequence alignment of this ORF to the predicted mouse Raly (mRaly) protein (PIR|A47318, [1]) revealed that the human protein has 89% overall identity and 96% similarity, but contains an additional 16 residues inserted in-frame after the first 109 amino acids (see below). While this work was in progress, the cDNA sequence of the p542 gene, isolated from a human B-lymphoblastoid cell line, was reported [5]. p542 was first recognized as an autoantigen in individuals infected with the Epstein–

* Corresponding author. Fax: +1-423-574-1271;

E-mail: michaudej@bio.ornl.gov

¹ GenBank accession numbers: AF148457 and AF148458.

² Present address: Parke-Davis Laboratory for Molecular Genetics, 1501 Harbor Bay Parkway, Alameda, CA 94502, USA.

³ Present address: Transgenomic, Inc., 2032 Concourse Drive, San Jose, CA 94041, USA.

Barr virus [6], and later identified as a human homolog of mRaly due to the great similarity between the two genes [5]. Interestingly, this form of human RALY (p542) also lacks the additional 16 amino acids found in the hRALY isoform reported here. Fig. 1B shows a Clustal W sequence alignment comparison of the human and mouse Raly proteins, and the human and mouse hnRNP C1 and C2 proteins. To avoid confusion, the isoform of human RALY cloned by us is designated as hRALY, whereas the previously cloned isoform is designated as p542. In addition to the 16-amino acid insert in hRALY (double underlined in Fig. 1B), there are three amino acid differences between hRALY and p542 (marked by colons under the sequence in Fig. 1B). Compared to hRALY, there is one extra residue in p542 (a serine at position 215 in p542) in the glycine-rich motif, which is the most divergent region between the human and mouse Raly proteins (single underline in Fig. 1B). There are also two amino acid differences between hRALY and p542 (positions 215 and 216 in hRALY); curiously, these residues are identical between hRALY and mRaly, but different in p542. These amino acid changes between hRALY and p542 may represent polymorphic differences in the human RALY gene and protein, or different tissue sources from which the cDNA clones were isolated.

Until recently, hnRNP C-related proteins were traditionally divided into two distinct domains; an N-terminal RNA-binding domain (RBD) and a C-terminal auxiliary domain [7,8]. The RBD (90–100 amino acids) is a well-characterized domain, evolutionarily conserved from yeast to man, and found in many hnRNPs and other RNA-binding proteins [7]. It assembles into four-stranded antiparallel β -sheets and two α -helices [9], and was shown to bind poly U tracts *in vitro* [10]. Most of the similarity between the Raly and hnRNP C proteins lies within this region ($\sim 80\%$ identity). The structure and function of the auxiliary domain (about 200 amino acids) are not well known. Four distinct regions within the auxiliary domain of hnRNP C were revealed recently [11,12] (Fig. 1A). Following the RBD, there is a region called variable due to the high degree of variability in the length of this region. Next, there is a region consisting mostly of basic residues (basic region), and a leucine zipper motif that is characterized by the presence of hydrophobic residues at every

seventh position and also at the fourth position. The basic region and the leucine zipper in hnRNP C are similar in organization to the basic leucine zipper motif of DNA-binding proteins and have been suggested to function as a novel RNA-binding motif [12]. The C-terminal region of the hnRNP C proteins is rich in acidic residues (acidic region). Although this region does not show a great deal of sequence identity between the Raly and hnRNP C proteins, suggesting a different function for these two groups of proteins, the Raly proteins do have numerous acidic residues in this region, many of which are identical between the Raly and hnRNP C proteins. There is a glycine–serine-rich stretch within the C-terminal regions of the Raly proteins (33 residues in mRaly, and 27 and 28 residues in hRALY and p542, respectively; single underline in Fig. 1B), which is absent in the hnRNP C proteins. The 28 amino acid region of p542 was identified as a cross-reactive epitope in autoimmune diseases [6].

As can be seen in Fig. 1B, the 16-amino acid in-frame insert that is alternatively processed in the hRALY protein (double underlined) is located within the variable region, at the same location where the hnRNP C1 and C2 proteins differ by the 13-amino acid insert in C2. Therefore, not only do RALY and hnRNP C have considerable sequence similarities, but they are also processed in a similar manner. The insertion of small peptides in hnRNP proteins by alternative splicing might be a general mechanism for diversification among hnRNPs. For example, two other members of the hnRNP family, the A2 and B1 proteins, also differ by only a 12-amino acid insert in B1 [4].

Because hRALY and p542 appear to be two isoforms of the same protein, we were interested in determining if the mRaly protein is also alternatively processed, with a 16-amino acid insert in the same region of the protein. To this end, we performed reverse transcriptase-polymerase chain reaction (RT-PCR) analysis [1] across this region of the mouse mRNA, using RNA from several different tissues. Two primers were used (Fig. 1C) that should amplify a 244-bp fragment from the previously published mRaly isoform (GB|L17076, [1]). If the mRaly gene also contains a 16-amino acid insert, we would expect to identify PCR fragments of 244 and 292 bp in length. As shown in Fig. 1C, in addition to a 244-

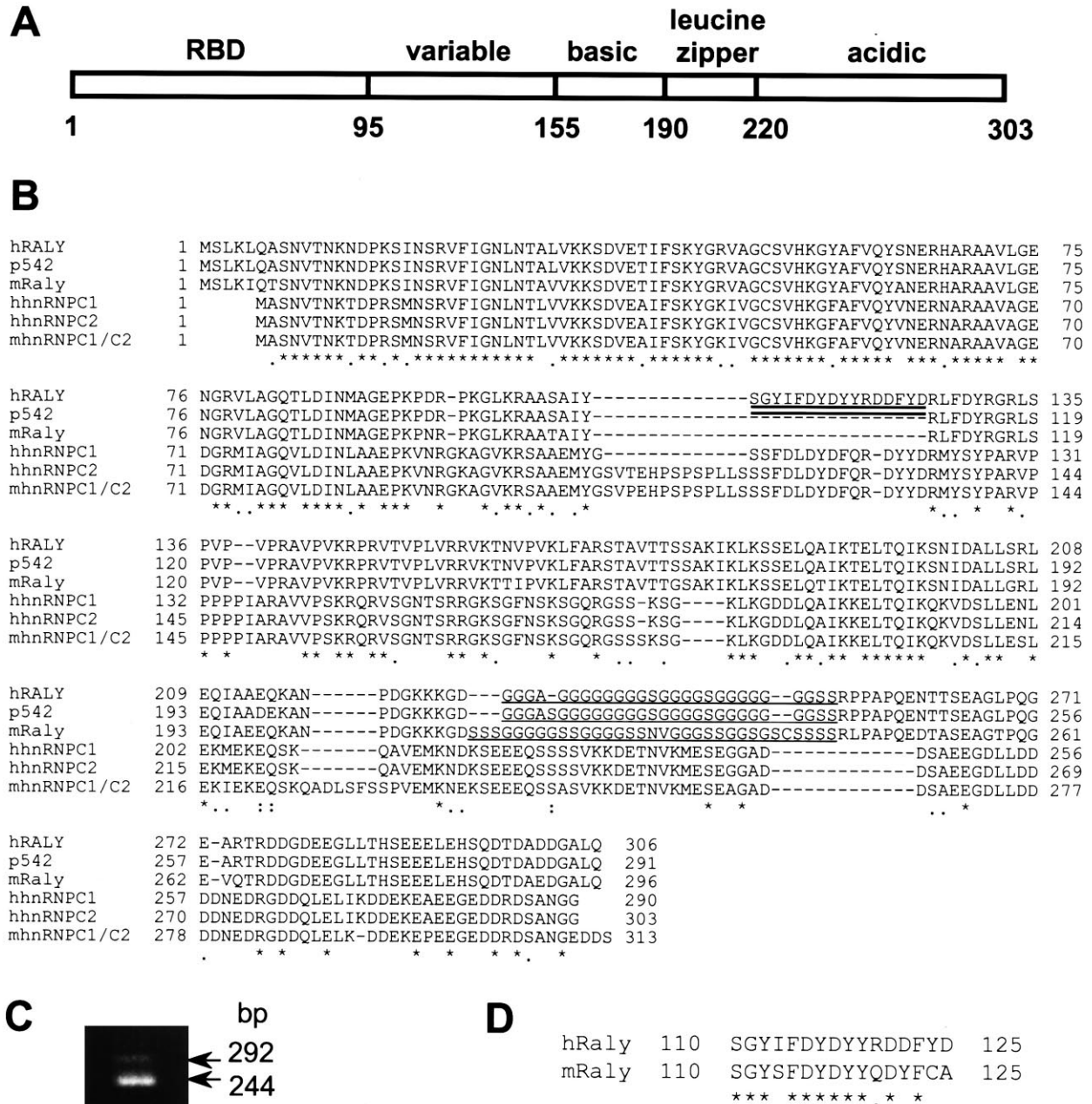


Fig. 1. Isoforms of the human and mouse Raly and hnRNP C proteins are derived by peptide insertion in the variable region. (A) Domain structure of hnRNP C-related proteins [11,12]. Residues at the beginning and end of each region correspond to hnRNP C2, as previously described [11]. (B) Clustal W sequence alignment of human RALY (hRALY in this report, AF148457; and p542, GB|AAC28898 [5]), mouse Raly (mRaly, PIR|A47318 [1]), human hnRNP C2 and C1 (hhnRNPC2, SP|P07910 and hhnRNPC1, PIR|A26885 [4]), and mouse hnRNP C1/C2 (mhnRNPC1/C2, GB|AAD03717, D.J. Williamson, J. DeGregori, H.E. Ruley, direct submission to GenBank). Residues identical in all sequences are marked by asterisks under the sequence. Residues reflecting conservative changes are indicated by dots under the sequence. Gaps are represented by dashes between the residues. The 16-amino acid in-frame insert that is alternatively processed in the hRALY protein is double underlined. The three amino acid differences between hRALY and p542 are marked by colons under the sequence. The glycine-serine-rich region within the C-terminal regions of the Raly proteins is single underlined. (C) RT-PCR analysis of mouse testis RNA for the presence of the alternate Raly transcript with an in-frame insert in the variable region. Fragment sizes are shown in basepairs (bp) to the right of the ethidium bromide-stained gel. Primers used: 5'-TGTCAGTATGCCAATGAGC-3' (forward), corresponding to nucleotides 438–457; and 5'-GCGAACCAAAGGGACTGTAA-3' (reverse), nucleotides complementary to 662–681 (GB|L17076, [1]). (D) Predicted amino acid sequence of the mRaly 16 amino acid insert compared to the corresponding insert in hRALY.

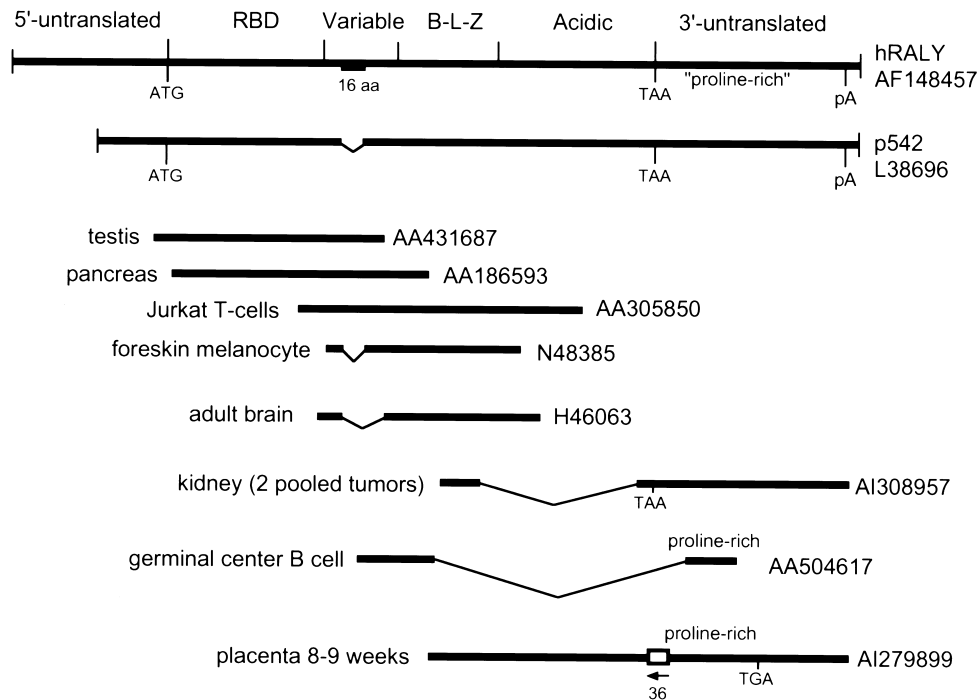


Fig. 2. Putative alternative processing of hRALY transcripts based on comparative analysis of known RALY cDNAs and selected RALY EST sequences. hRALY and p542 are full-length cDNAs encoding two human RALY isoforms, with and without a 16-amino acid insert in the variable region of the protein, respectively. Bold lines represent sequences corresponding to the hRALY cDNA sequence and thin lines represent putative splicing. The tissue origin of each EST is shown on the left and the accession numbers of the full-length RALY cDNAs (hRALY and p542) and ESTs are shown on the right. Regions of the hRALY transcript that encode putative protein domains are indicated above hRALY. The basic and leucine zipper regions are marked as B-L-Z. 'Proline-rich' designates a region in the 3'-untranslated portion of hRALY that is present in ORFs encoded by the germinal center B-cell and placenta ESTs. The translation initiation codon (ATG) and the stop codon (TAA) within the cDNA sequences are indicated, as are the 36-bp inversion and position of the putative TGA stop codon in the placenta EST.

bp fragment, a second larger fragment was observed in mouse testis RNA. The same results were also obtained for mouse brain and lung RNA (not shown). Sequencing confirmed that the larger fragment had an additional 48 nucleotides (GenBank accession number AF148458) that encode an additional 16 amino acids in mRaly, exactly in the same place as in the hRALY protein. Fig. 1D shows the predicted amino acid sequence of the mouse insert in comparison to the corresponding insert in hRALY. Therefore, in addition to the previously described alternative splicing of 83 bp of sequence in the 5'-untranslated region of the mRaly transcript [1,2], another form of alternative processing of Raly is demonstrated here for both the human and mouse genes, involving an in-frame inclusion of 16-amino acids in the variable region of the protein.

A search of GenBank for RALY expressed sequence tag (EST) sequences that span the variable

region confirmed the presence of both isoforms in human ESTs. For example, clones from human testis (GB|AA431687), pancreas (GB|AA186593) and Jurkat T-cell libraries (GB|AA305850) were found to contain the 16-amino acid insert, whereas a clone from a foreskin melanocyte cDNA library (GB|N48385) contained the RALY isoform without the insert (Fig. 2). Interestingly, there was one RALY EST (GB|H46063) from a human adult brain library that splices out 87 bp of DNA, resulting in the removal of the alternative 16 amino acids, as well as the next 13 residues in the deduced protein sequence of hRALY (Fig. 2). It is not clear from one clone whether or not this EST represents a legitimate transcript. However, it is worth noting that, in addition to this clone, there are other unique RALY EST sequences, at least suggesting the possibility of additional alternatively processed RALY protein isoforms (Fig. 2). For example, a kidney tumor EST

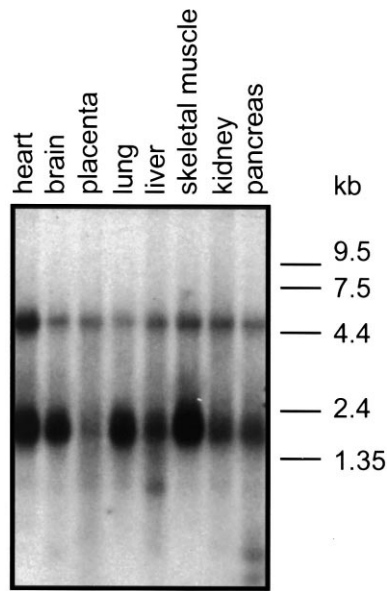


Fig. 3. Expression analysis of the hRALY gene. A ^{32}P -labeled full-length hRALY cDNA clone was hybridized to a poly A⁺ RNA Northern blot (2 μg per lane, Clontech) of various human tissues (indicated above the blot). Positions of RNA size markers, in kilobases (kb), are indicated on the right.

(GB|AI308957) has sequence corresponding to nucleotide 885 of hRALY cDNA spliced to nucleotide 1187, resulting in the removal of coding sequence for half of the leucine zipper and most of the C-terminal acidic regions of the protein. In another clone (GB|AA504617) from a germinal center B-cell library, hRALY nucleotide 794 is spliced to nucleotide 1282, resulting in the removal of coding sequence for half of the basic domain, all of the leucine zipper region and the acidic C-terminus, including the stop codon (nucleotides 1217–1219). Skipping the stop codon results in the continuation of the ORF and an additional 29 amino acids (the EST sequence does not contain any in-frame stop codon) that are rich in proline residues. Similarly, an EST clone (GB|AI279899) from a placenta library contains a 36-bp inversion around the stop codon (nucleotides 1207–1242 of hRALY cDNA), resulting in the removal of the stop codon and the addition of 66 amino acids that are rich in proline residues (as in the B-cell clone above; Fig. 2), before reaching a new in-frame stop codon. Again, additional experimentation would be needed to determine the validity of these apparently new transcripts and their functional significance. Proline-rich peptides have been shown to

be involved in protein–protein interactions, especially with Src homology 3 (SH3) and WW domains present in a number of signaling and regulatory proteins [13,14].

As previously shown for the mRaly gene [1,2], hRALY is widely expressed as an ~ 1.7 -kb transcript. Fig. 3 shows a poly A⁺ RNA Northern blot of various human tissues (Clontech) hybridized with the ^{32}P -labeled full-length RALY cDNA clone. A transcript size of ~ 1.7 kb is consistent with the 1589-bp cDNA clone, plus a poly A tail of approximately 100 bases. The two RALY transcripts differ in size by only 48 bases (that encode the 16 amino acid insert in the protein) and are therefore seen to co-migrate on the Northern blot. An additional hybridizing fragment, slightly larger than 4.4 kb in size and co-migrating with 28S RNA, may represent cross-hybridization to 28S, partially processed RALY message, or another, as yet unidentified, RALY transcript. The p542 gene was reported to be expressed as a single 4.4-kb transcript in three B-cell lines, one T-cell line, and in HeLa (epithelial) cells [5]. The authors did not report hybridization of the p542 probe to a 1.7-kb transcript in these cell lines. If these cell lines do not express the prominent 1.7-kb RALY message, it would indicate that this RALY message is not necessary for cell viability.

Human RALY (p542) has been mapped by the International RH Mapping Consortium to chromosome 20, within the reference interval 39.9–50.2 cM (<http://www.ncbi.nlm.nih.gov/genemap/loc.cgi?ID=30442>), close to the human AGOUTI gene (reference interval 50.2–54.9 cM; <http://www.ncbi.nlm.nih.gov/genemap/loc.cgi?ID=30442>), which is consistent with the large region of syntenic homology between distal mouse chromosome 2 and human chromosome 20q.

The work presented here has demonstrated that both the human and mouse Raly genes are alternatively processed to generate at least two protein isoforms that are similar in both overall sequence and processing compared to the hnRNP C proteins. As expected, the hRALY gene is widely expressed as a 1.7-kb message. In general, the hnRNP proteins appear to be modular in nature, with multiple functional domains. Recent data on the hnRNP C proteins are beginning to elucidate the potential functional roles of these various domains [12,15].

The generation of targeted mutations within specific functional domains of the mRaly gene, and other hnRNP genes, is one approach that may provide significant insight into the *in vivo* function of these domains.

We thank Jeff Schrick for providing the human cDNA library membranes that were screened with the mRaly probe. The work reported here was supported by the Office of Health and Environmental Research, United States Department of Energy under contract DE-AC05-96OR22464 with Lockheed Martin Energy Research Corporation.

References

- [1] E.J. Michaud, S.J. Bultman, L.J. Stubbs, R.P. Woychik, *Genes Dev.* 7 (1993) 1203–1213.
- [2] D.M. Duhl, M.E. Stevens, H. Vrieling, P.J. Saxon, M.W. Miller, C.J. Epstein, G.S. Barsh, *Development* 120 (1994) 1695–1708.
- [3] E.J. Michaud, S.J. Bultman, M.L. Klebig, M.J. van Vugt, L.J. Stubbs, L.B. Russell, R.P. Woychik, *Proc. Natl. Acad. Sci. USA* 91 (1994) 2562–2566.
- [4] C.G. Burd, M.S. Swanson, M. Gorlach, G. Dreyfuss, *Proc. Natl. Acad. Sci. USA* 86 (1989) 9788–9792.
- [5] G.H. Rhodes, J.R. Valbracht, M.D. Nguyen, J.H. Vaughan, *J. Autoimmun.* 10 (1997) 447–454.
- [6] J.H. Vaughan, J.R. Valbracht, M.D. Nguyen, H.H. Handley, R.S. Smith, K. Patrick, G.H. Rhodes, *J. Clin. Invest.* 95 (1995) 1306–1315.
- [7] C.G. Burd, G. Dreyfuss, *Science* 265 (1994) 615–621.
- [8] G. Biamonti, S. Riva, *FEBS Lett.* 340 (1994) 1–8.
- [9] M. Wittekind, M. Gorlach, M. Friedrichs, G. Dreyfuss, L. Mueller, *Biochemistry* 31 (1992) 6254–6265.
- [10] M. Gorlach, M. Wittekind, R.A. Beckman, L. Mueller, G. Dreyfuss, *EMBO J.* 11 (1992) 3289–3295.
- [11] W. Jiang, X. Guo, V.P. Bhavanandan, *Biochim. Biophys. Acta* 1399 (1998) 229–233.
- [12] J.G. McAfee, L. Shahied-Milam, S.R. Soltaninassab, W.M. LeSturgeon, *RNA* 2 (1996) 1139–1152.
- [13] M. Sudol, H.I. Chen, C. Bougeret, A. Einbond, P. Bork, *FEBS Lett.* 369 (1995) 67–71.
- [14] G.B. Cohen, R. Ren, D. Baltimore, *Cell* 80 (1995) 237–248.
- [15] L. Shahied-Milam, S.R. Soltaninassab, G.V. Iyer, W.M. LeSturgeon, *J. Biol. Chem.* 273 (1998) 21359–21367.