Gene Fusion and Overlapping Reading Frames in the Mammalian Genes for 4E-BP3 and MASK*

Francis Poulin‡, Andrea Brueschke, and Nahum Sonenberg§

From the Department of Biochemistry and McGill Cancer Center, McGill University, Montréal, Québec H3G 1Y6, Canada

4E-BP3 is a member of the eukaryotic initiation factor (eIF) 4F-binding protein family of translational repressors. eIF4E-binding proteins (4E-BPs) inhibit translation initiation by sequestering eIF4E, the cap-binding protein, from eIF4G thus preventing ribosome recruitment to the mRNA. Previous analysis of 4E-BP3 expression uncovered an 8.5-kb mRNA variant of unknown origin. To study this splice variant, we determined the structure of the genomic locus encoding human 4E-BP3 (EIF4EBP3). EIF4EBP3 is located on human chromosome 5q31.3 and comprises three exons (A, B, and C) and two introns. Exon B contains the region of the open reading frame responsible for eIF4E binding. GenBank™ searches revealed multiple expressed sequence tags originating from the alternative splicing of exon B with unidentified upstream exons. Further studies revealed that the 8.5-kb transcript arises from the fusion of EIF4EBP3 with the mammalian homologue of Drosophila MASK (multiple ankyrin repeats, single KH domain), which is crucial for photoreceptor differentiation, cell survival, and proliferation. Surprisingly, the open reading frame of the MASK-BP3 transcript is different from that of 4E-BP3, which indicates that exon B is translated using an alternative reading frame. A gene fusion similar to that of MASK and EIF4EBP3 has been reported only once in mammals for the UEVI-Kua transcript. The use of an alternative reading frame is also very rare, having been described for two loci, INK4a/ARF and Xlais/ALEX. The simultaneous exploitation of both mechanisms underscores the flexibility of mammalian genomics and has important implications for the functional analysis of 4E-BP3 and MASK. Interestingly, both eIF4E and MASK are downstream effectors of the Ras/MEK pathway, which provides a rationale for the MASK-BP3 fusion in mammals.

The control of translation rates in eukaryotes is an important means to regulate gene expression. Translational control occurs mainly at the rate-limiting initiation step during which the 40 S ribosomal subunit is recruited to the mRNA (1, 2). Ribosomal recruitment is facilitated by the 5’-cap structure (m’GpppN, where N is any nucleotide) present on most cellular eukaryotic mRNAs (3). The cap structure is specifically recognized by eukaryotic initiation factor (eIF)3 4E, a complex comprised of three subunits: eIF4E, the cap-binding protein; eIF4A, a bi-directional RNA helicase; and eIF4G, a large modular scaffolding polypeptide that performs a bridging function between the ribosome and mRNA (1, 4). eIF4E plays a critical role in translational initiation (4), and accordingly, its activity is regulated at multiple levels (5) transcriptionally via phosphorylation and through the binding of small inhibitory proteins, the eIF4E-binding proteins (4E-BPs). The 4E-BPs specifically inhibit cap-dependent translation initiation by preventing the interaction of eIF4E with eIF4G, and thus they prevent ribosome binding to the mRNA (6). At the molecular level, the 4E-BPs mimic eIF4G and compete for eIF4E binding (6, 7). The binding of the 4E-BPs to eIF4E is regulated by their phosphorylation state. The underphosphorylated forms of 4E-BPs interact with eIF4E, whereas the hyperphosphorylated forms do not (4). Upon cell stimulation with serum, growth factors, or hormones, 4E-BPs become phosphorylated on multiple serine/threonine residues and dissociate from eIF4E to relieve translational inhibition (8, 9).

4E-BP1, 4E-BP2, and 4E-BP3 are the three known members of the mammalian 4E-BP family (10, 11). 4E-BP1 is the best characterized (6, 8–10, 12–17), and it is considered to be the prototype of the family. 4E-BP3 shares the basic structural and functional features of 4E-BP1 (11), but its regulation appears to be different (18–20). Our analysis of 4E-BP3 expression revealed that it is expressed as two different mRNAs one of which is of unknown origin (11). To understand this observation, we undertook the characterization of the human EIF4EBP3 genomic locus.

EXPERIMENTAL PROCEDURES

Northern Blotting—Human and mouse 4E-BP3 splice variants were analyzed by Northern hybridization using commercial blots (multiple tissue Northern blots MNT1 and MNTM, respectively, Clontech) according to the manufacturer’s instructions. The following DNA fragments were used as probes: mouse 4E-BP3 cDNA, exon 34 of the MAST cDNA (nt 7681–8139), exon B of the MAST-BP3/ARF cDNA (nt 7829–8265), exon A of the 4E-BP3 cDNA (GenBank™ accession number AF038869; nt 1–175), exon 0 of the MASK-BP3/ARF cDNA (nt 7626–7739), and human β-actin (Clontech).

Isolation of Genomic Clones—Phage clones containing the mouse and human genomic regions were isolated from a lambda gypsy library using the mouse genomic DNA as a probe.

Received for publication, September 30, 2003
Published, JBC Papers in Press, October 13, 2003, DOI 10.1074/jbc.M310761200

© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.

* This work was supported by a grant from the Canadian Institutes of Health Research. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Supported by a Doctoral Award from the Canadian Institutes of Health Research. Present address: Dept. of Genome Sciences, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720.
§ Canadian Institutes of Health Research Distinguished Investigator and a Howard Hughes Medical Institute International Scholar. To whom correspondence should be addressed: Dept. of Biochemistry, McGill University, 3655 Promenade Sir William Osler, Rm. 807, Montréal, Québec H3G 1Y6, Canada. Tel.: 514-398-7274; Fax: 514-398-1287; E-mail: nahum.sonenberg@mcgill.ca.

§ Canadian Institutes of Health Research Distinguished Investigator and a Howard Hughes Medical Institute International Scholar. To whom correspondence should be addressed: Dept. of Biochemistry, McGill University, 3655 Promenade Sir William Osler, Rm. 807, Montréal, Québec H3G 1Y6, Canada. Tel.: 514-398-7274; Fax: 514-398-1287; E-mail: nahum.sonenberg@mcgill.ca.

This paper is available on line at http://www.jbc.org

Published, JBC Papers in Press, October 13, 2003, DOI 10.1074/jbc.M310761200

© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.

Received for publication, September 30, 2003
Published, JBC Papers in Press, October 13, 2003, DOI 10.1074/jbc.M310761200

© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.

Received for publication, September 30, 2003
Published, JBC Papers in Press, October 13, 2003, DOI 10.1074/jbc.M310761200

© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.

Received for publication, September 30, 2003
Published, JBC Papers in Press, October 13, 2003, DOI 10.1074/jbc.M310761200

© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.
Two Reading Frames in the Second Exon of 4E-BP3

human 4E-BP3 genes were obtained by screening genomic DNA libraries (human circulating lymphocytes, Streitagen, and mouse 129S1v, a generous gift from Tak Mak) with the human 4E-BP3 cDNA as a probe.

**MASK Cloning**—The full-length MASK cDNA was cloned by sequential 5′-RACE using a commercial kit (marathon cDNA amplification kit, Clontech) according to the manufacturer’s instructions. Four GST oligonucleotides were used: GSP1, 5′-CTC CCC AGC CCT GGT TAG CTG GCA-3′; GSP2, 5′-CTC ACT AAC ATT TTT CAC CAG-3′; GSP3, 5′-GAC TAC CAT CTG CTT CCG CTG G; and GSP4, 5′-GTC CAC ACG CAG CTA GCG TC. The four RACE products were subcloned together and with ESTs AA465291 and BE622326 to give the full-length human MASK cDNA (GenBank™ accession number AF521882) and with ESTs T82868 and BE622326 to give the full-length human MASK-4BP3™ cDNA (GenBank™ accession number AF521883). The cDNAs sequences were confirmed by sequencing.

**Quantitative RT-PCR**—A aliquots (250 ng) from the mouse total RNA master panel (BD Biosciences) or dilutions of the calibrator mouse total RNA (mouse 129SvJ, a database). These were used for quantitative single-step RT-PCR in the LightCycler System using the RNA Master SYBR Green I kit (Roche Diagnostics) as described by the manufacturer. The amplitization program consisted of an initial denaturation step at 95 °C for 30 s followed by 40 cycles of denaturation at 95 °C for 1 s, annealing at 60 °C for 3 s, and extension at 72 °C for 15 s. The amount of fluorescent product was measured at the end of every cycle with a single acquisition at 80 °C for 1 s. Duplicate determinations were made for each sample, and the gene copy number was normalized by the amount of actin. The same samples were assayed in the LightCycler using the Roche Diagnostics Quantification Software (version 1.0; Roche Diagnostics). The following transcript-specific primers were used: m4E-BP3 (ex1 forward, 5′-CTC CCC AGC ACA CCA C-3′; ex3 reverse, 5′-CTG TTC GTC ATC CCTG TAT T-3′), mMASK-BP3™ (ex33.1 forward, 5′-GGA CTT AAC AAA CCA ACA GA-3′; ex3 reverse, 5′-CTG TTC TCC ATC TTC GCAT TAT T-3′), mMASK (ex33.2 forward, 5′-GCT CCC ACT AAC ATT TTT CAC CAG-3′; ex34 reverse, 5′-CCC AGG GTC CTT CCA TCA TAG TAC-3′), and mβ-actin (actin forward, 5′-ACC AAC GGC AGC AGA CAA A-3′; actin reverse, 5′-AGC ACC ACG GGA ATC AAG CAA-3′).

**Antibodies**—Antibody 2275 was raised in a rabbit against a synthetic peptide (CHPMHQQLSDPSTFSQ) comprising amino acids 2423–2437 of the human 4E-BP3. The peptide (2 mg) was cross-linked to keyhole limpet hemocyanin and bovine serum albumin using a commercial kit (Inject-activated immunogen conjugation kit, Pierce). The antibody was affinity-purified on a DNA fragment that is syntenic with human chromosome 5q31.3. An unrelated antibody (MAB3482) was used as a control. Antibodies 2275, 2335, and 2336 recognize both human and mouse 4E-BP3. Antibodies 2275, 2335, and 2336 were raised in rabbits against a GST-MASK fusion (Mage amino acids 2044–2523). Anti-MASK-BP3™ antibodies 2349 and 2350 were raised to a GST-MASK-BP3™ fusion (MAGE amino acids 2617). Antibodies 2275, 2335, and 2336 recognize both human and mouse 4E-BP3. Antibodies 2349 and 2350 specifically recognize MAGE-BP3™.

**Cell Culture and Transfections**—The prostate cancer cell line LNCaP (21) was purchased from the American Type Culture Collection (ATCC, CRL-1740) and maintained in RPMI 1640 medium plus 10% fetal bovine serum. Human embryonic kidney cells 293T/17 (22) were obtained from the ATCC (CRL-11268) and grown in Dulbecco’s modified Eagle’s medium containing 10% fetal bovine serum. 293T/17 cells (10-cm dish, 80% confluent) were transiently cotransfected with pcDNA3, pcDNA3-MAK, pcDNA3-MAK-MASK-BP3™, pcDNA3-MASK, or pcDNA3-MASK-BP3™ (5 μg) using LipofectAMINE-Plus (Invitrogen, 200 μl). After 6 h, the transfected cells were trypsinized, plated on a 15-cm dish, and grown for 24 h. Cell extracts were prepared by scraping cells in cold Nonidet P-40 buffer (0.5% Triton X-100, 0.5% sodium deoxycholate, and 150 mM NaCl). The supernatant was transferred to a fresh tube together with anti-HA antibody HA.11 (Babco, 5 μg) and protein G-Sepharose (Amersham Biosciences, 40 μl). The supernatant was transferred to a fresh tube together with anti-HA antibody HA.11 (Babco, 5 μg) and protein G-Sepharose (40 μl). Incubation with end-over-end rotation was carried out at 4 °C for 2.5 h. Beads were spun down and washed three times with cold radioimmune precipitation assay buffer.

**Immunoprecipitation**—For immunoprecipitation experiments, 293T/17 cells (150-mm dish) were lysed 24 h after transfection in 500 μl of cold Nonidet P-40 buffer, and debris was spun down. Total extract (1 mg) was diluted with 2 volumes of radioimmune precipitation assay buffer and precleared at 4 °C for 1 h by incubation with protein G-Sepharose (Amersham Biosciences, 40 μl). The supernatant was transferred to a fresh tube together with anti-HA antibody HA.11 (Babco, 5 μg) and protein G-Sepharose (40 μl). Incubation with end-over-end rotation was carried out at 4 °C for 2.5 h. Beads were spun down and washed three times with cold radioimmune precipitation assay buffer.

The immunoprecipitated material was eluted in Laemmli sample buffer and subjected to SDS-PAGE (6% gel).

**Immunoblotting**—Polypeptides were resolved on SDS-6% polyacrylamide gels and transferred for 16 h at 30 V onto 0.22-μm nitrocellulose membranes. The membranes were blocked for 24 h with 5% milk in phosphate-buffered saline containing 0.05% Tween 20 (PBS-T). Membranes were then incubated for 2 h with primary antibodies 2335 (1:1000), 2336 (1:1000), 2350 (1:1000), and 2275 (1:500) or a monoclonal antibody to HA (HA.11, 1:5000, Babco) all of which were diluted in PBS-T. Incubation with secondary antibody was performed with peroxidase-coupled donkey anti-rabbit immunoglobulin (1:5000 in PBS-T, Amersham Biosciences) or with peroxidase-coupled sheep anti-mouse immunoglobulin (1:5000 in PBS-T, Amersham Biosciences). Detection was performed with ECL (PerkinElmer Life Sciences).

**RESULTS**

**Cloning of Human and Mouse 4E-BP3 Gene**—In a previous analysis of 4E-BP3 mRNA expression, we observed that the human 4E-BP3 cDNA hybridized with two different messages, a mRNA of ~800 bp corresponding to the cDNA probe and an unidentified mRNA of ~8.5 kb (11). To determine whether expression of the 8.5-kb variant is conserved in the mouse, we performed a Northern analysis on multiple murine poly(A) RNA using the mouse 4E-BP3 cDNA as a probe (Fig. 1). As anticipated, the 800-bp and 8.5-kb mRNAs were both expressed in the mouse (Fig. 1, lane 7). To better understand the nature of the 8.5-kb mRNA, we isolated the human EIF4EBP3 gene and mouse Eif4ebp3 genes.

The human EIF4EBP3 gene spans 1.9 kb and comprises three exons and two introns (Fig. 2A). Data base searches indicated that the human EIF4EBP3 gene is present as a single copy in each haploid genome (data not shown), which was verified through BLAST searches of the Human Genome sequence database. This analysis also indicated that the Eif4ebp3 gene family is composed of no more than three expressed genes, Eif4ebp1, Eif4ebp2, and Eif4ebp3. The mouse Eif4ebp3 gene has a structure that is identical to the human gene, but it is 1.7 kb in length. Data base analysis established that Eif4ebp3 is located on mouse chromosome 18, on a DNA fragment that is syntenic with human chromosome 5q31. The sequence of the mouse Eif4ebp3 gene has been released in GenBank™ (accession number AC087795).

**Identification of 4E-BP3 mRNA Splice Variants**—We used the genomic structure of Eif4ebp3 to investigate potential splice variants. Exons B + C (Fig. 2A) were amplified by PCR and were used to probe a human poly(A) RNA blot (Fig. 3A). This analysis identified an 800-bp and an 8.5-kb message as was observed previously with a full-length cDNA probe (11). A second probe consisting of exon A was used on the same RNA blot (Fig. 3B). In contrast to exons B + C, the exon A probe recognized the 800-bp message but not the 8.5-kb message.

To further investigate the source of the 8.5-kb message, BLAST searches were performed on the human EST database of GenBank™. Data base probing with the sequence of exon A exclusively identified ESTs that were contained within the cDNA of 4E-BP3 (Fig. 2A, EST type 1). In contrast, two categories of ESTs could be detected when the sequence of exons B + C was used as a probe. As with exon A, the first category comprised ESTs contained in the known 4E-BP3 cDNA. The second category, however, consisted of ESTs harboring sequences identical to exons B + C plus some unidentified sequence located at the position where exon A is found in 4E-BP3 cDNA (Fig. 2A, EST type 2). One of these ESTs (GenBank™ accession number T52868) was sequenced. Oligonucleotides were designed from this sequence, and a phage clone contain-
Two Reading Frames in the Second Exon of 4E-BP3

Fig. 1. Analysis of transcript expression from the mouse Mask and Eif4ebp3 loci. A. The expression of mouse 4E-BP3 mRNA was analyzed by Northern blotting on 2 μg of poly(A) RNA purified from mouse tissues. The mouse 4E-BP3 (top panel) and human β-actin (bottom panel) cDNAs were used as probes. Lane 1, heart; lane 2, brain; lane 3, spleen; lane 4, lungs; lane 5, liver; lane 6, skeletal muscle; lane 7, kidney; lane 8, testis. B, real time quantitative RT-PCR was performed on total RNA extracted from a panel of mouse tissues. The expression levels of mouse EBP3 (black bars), hMask-BP3 (white bars), and hMask (gray bars) were quantified and normalized to the levels of β-actin present in each sample. The analysis was performed with transcript-specific primer pairs as described under “Experimental Procedures.”

The EIF4EBP3 gene was sequenced. Two novel putative exons were identified (exons 0 and 33) that are located 5.4 and 8.5 kb upstream of exon A, respectively (Fig. 2A). Data base searches using the sequence of exon 0 failed to reveal any novel splice variants. Searches with the sequence of exon 33, however, exposed a third type of EST that contains a novel putative exon (exon 34) located immediately after exon 33 (Fig. 2A, EST type 3). An EST from the third type (GenBank accession number AA465291) was sequenced. Except for exon 34 at the 3' terminus, it shared all of its upstream sequence with EST T82868. Exons 0 and 34 were amplified and used as probes for a human multiple tissues Northern blot (Fig. 3, C and D, respectively). These analyses revealed messages in which the size and pattern of expression are identical to the 8.5-kb message observed using the EIF4EBP3 exons B+C probe (Fig. 3A).

To better characterize the 8.5-kb message, 5'RACE was performed on poly(A) RNA isolated from LNCaP cells. Four RACE products were successively generated, and they were spliced together, along with ESTs obtained from the IMAGE consortium. Two cDNAs were constructed, the cDNA for splice variant 3 contains exons 33+34 at its 3' end (Fig. 2A), whereas the cDNA for splice variant 2 contains exons 33+0+B+C (Fig. 2A). The full-length cDNA for splice variant 3 (GenBank accession number AF521882) is 8139-bp long and consists of a 5'-UTR of 60 bp, an ORF of 7829 bp, and a 3'-UTR of 450-bp. The 5'-UTR of the cDNA is most probably close to full-length as the size of the mRNA observed in Fig. 3D (8.5 kb) is close to the size of the cDNA (8139 bp), taking into consideration the presence of a poly(A) tail on the mRNA. Moreover, there is an in-frame stop codon located 45 bp upstream from the initiator ATG. The full-length cDNA for splice variant 2 (GenBank accession number AF521883) is 8265-bp long and consists of a 5'-UTR of 60 bp that is identical to that of splice variant 3, an ORF of 7854 bp, and a 3'-UTR of 351 bp.

Splice Variant 3 Encodes the Human Homologue of Drosophila MASK—The ORF of splice variant 3 encodes a predicted 2542-amino acid protein with a molecular weight of 269,487. The polypeptide contains two identifiable structural motifs, two blocks of ankyrin repeats (10 and 15 repeats) and a KH domain. Although it is shorter, this putative polypeptide is highly similar to the Drosophila melanogaster protein MASK (multiple ankyrin repeats, single KH domain 23). The conservation is especially striking in the regions containing the ankyrin repeats (Fig. 4, A and B) and the KH domain (Fig. 4C). Because of this homology we have named the protein encoded by splice variant 3 human MASK (hMASK).

As previously observed (23), data base searches of GenBank define a family of hMASK-related transcripts expressed in several organisms from Caenorhabditis elegans to humans. As expected, the most highly conserved parts of MASK are the ankyrin repeats and the KH domain (Fig. 4). Analysis of the human genome also identified a gene that appears to have arisen from a duplication of MASK. This gene is located on chromosome 4q13.3, and it expresses a transcript that encodes a protein named GTAR (gene trap ankyrin repeat, GenBank accession number NM_032217). The human GTAR displays 71% overall homology to hMASK with higher identity in the ankyrin repeats and KH domain regions (Fig. 4).

Splice Variant 2 Results from the Fusion of MASK and EIF4EBP3—The gene for hMASK is located on chromosome 5q31.3. It is flanked upstream by the SLCA4A9 gene and downstream by the EIF4EBP3 gene. The transcript that we isolated as splice variant 2 (Fig. 2A) results from readthrough transcription of MASK followed by alternative splicing (Fig. 2B). Because splice variant 2 results from the fusion of MASK exons with EIF4EBP3 exons (Fig. 2, A and B), we have named the encoded protein human MASK-BP3 (hMASK-BP3). The hMASK-BP3 ORF encodes a predicted 2617-amino-acid protein with a molecular weight of 277,206. The first 2542 amino acids result from the translation of exons 0, B, and C. Intriguingly, computer analysis predicts that exons B and C of the MASK-BP3 transcript are translated in an in-frame stop codon located 45 bp upstream from the initiator ATG. The full-length cDNA for splice variant 2 (GenBank accession number AF521883) is 8265-bp long and consists of a 5'-UTR of 60 bp that is identical to that of splice variant 3, an ORF of 7854 bp, and a 3'-UTR of 351 bp.

Because of this homology we have named the protein encoded by splice variant 3 human MASK (hMASK).

As previously observed (23), data base searches of GenBank define a family of hMASK-related transcripts expressed in several organisms from Caenorhabditis elegans to humans. As expected, the most highly conserved parts of MASK are the ankyrin repeats and the KH domain (Fig. 4). Analysis of the human genome also identified a gene that appears to have arisen from a duplication of MASK. This gene is located on chromosome 4q13.3, and it expresses a transcript that encodes a protein named GTAR (gene trap ankyrin repeat, GenBank accession number NM_032217). The human GTAR displays 71% overall homology to hMASK with higher identity in the ankyrin repeats and KH domain regions (Fig. 4).
4E-BP3, MASK, and MASK-BP3 Are Differentially Expressed in Mouse Tissues—The distribution of m4E-BP3, mMASK, and mMASK-BP3 in multiple mouse tissues was analyzed by real time quantitative RT-PCR (Fig. 1B). The relative abundance of the three mRNAs in various tissues was observed to differ substantially. m4E-BP3 is very strongly expressed in the liver with a 10-fold higher expression than in the kidney (Fig. 1B, black bars). Other tissues express variable, and significantly lower, levels of the m4E-BP3 mRNA. This is in contrast to the mMASK mRNA in which expression is more evenly distributed across tissues with slightly higher levels detected in the brain and the eye (Fig. 1B, gray bars). The mMASK-BP3 mRNA expression is also more uniformly distributed in different tissues with higher expression detected in the kidney and testis (Fig. 1B, white bars).

MASK and MASK-BP3-ARF Are Both Expressed in LNCaP Cells—To determine whether MASK and MASK-BP3-ARF proteins are expressed in cells, we generated antibodies with various specificities toward the two proteins (see “Experimental Procedures”). Because they were raised against polypeptides present in both MASK and MASK-BP3-ARF, antibodies 2275, 2335, and 2336 should recognize both proteins. On the other hand, antibodies 2349 and 2350 should be specific for MASK-BP3-ARF as they were raised against a fusion protein (GST-MASK-BP3-ARF) designed from the MASK-BP3-ARF-specific exons (0 +B+C).

To verify that the antibodies displayed the correct specificity, we transiently transfected 293T/17 cells with HA-tagged MASK or HA-tagged MASK-BP3-ARF (Fig. 5A). Overexpression of the two proteins was verified using an antibody directed against the HA epitope (Fig. 5A, top panel). As expected, HA-MASK and HA-MASK-BP3-ARF were both detected by antibodies 2275 and 2335 (Fig. 5A, second and third panels). Antibody 2335 could also detect the endogenous MASK present in 293T/17 cells (Fig. 5A, third panel, lane 1). Contrary to antibodies 2275 and 2335, the MASK-BP3-ARF-specific antibodies 2350 and 2349 could detect only the overexpressed HA-MASK-BP3-ARF protein (Fig. 5A, bottom panel, and data not shown). To further assess the specificity of antibodies 2335 and 2350, untagged MASK and MASK-BP3-ARF were overexpressed in 293T/17 cells (Fig. 5B). As anticipated, antibody 2335 detected both MASK and MASK-BP3-ARF (Fig. 5B, top panel), whereas antibody 2350 detected only MASK-BP3-ARF (Fig. 5B, bottom panel). In agreement with their predicted molecular weight, MASK and MASK-BP3-ARF migrated as ~270-kDa polypeptides on SDS-polyacrylamide gels (Fig. 5B).

Expression of the endogenous MASK and MASK-BP3-ARF in LNCaP cells was established by immunoblot analysis (Fig. 5C). LNCaP cells were chosen because they express the mRNAs for both MASK and MASK-BP3-ARF (data not shown). The five antibodies that were generated against MASK or MASK-BP3-ARF recognized polypeptides of almost identical size (Fig.
BP3ARF expression detected by the exon 0 probe. The expression of transcripts from the EIF4EBP3 locus was analyzed by Northern blotting on 2 μg of poly(A)+ RNA purified from human tissues. DNA probes derived from the exons illustrated in Fig. 2 were used for the detection of specific transcripts. A, EIF4EBP3 and MASK-BP3 translation initiation codon, the amount of mRNA in each lane was quantified. B, EIF4EBP3 expression detected by the exon A probe. C, MASK-BP3 translation initiation codon, the amount of mRNA in each lane was quantified. D, EIF4EBP3 expression detected by the exon 0 probe. E, the amount of mRNA in each lane was controlled using a human β-actin probe. Lane 1, heart; lane 2, brain; lane 3, placenta; lane 4, lungs; lane 5, liver; lane 6, skeletal muscle; lane 7, kidney; lane 8, pancreas.

We have analyzed the structure of the gene encoding the translational inhibitor 4E-BP3. Comparison of EIF4EBP3 with the previously described EIF4EBP1 and EIF4EBP2 genes (24) reveals that they possess a very similar structure, although their sizes are different, 1.9 kb for EIF4EBP3, 19.7 kb for EIF4EBP2, and 29.8 kb for EIF4EBP1. The three genes comprise three exons and two introns, with the first intron being longer than the second intron. The position of the first intron is absolutely conserved in human and mouse EIF4EBP genes. The position of the second intron is similar in the three genes, but the position of the splice donor and acceptor site of the flanking exons is variable; the second exon of EIF4EBP2 is 6 nt longer at the 3' end when compared with EIF4EBP1 and EIF4EBP3, whereas the third exon of EIF4EBP3 is 3 nt shorter at the 5' end when compared with EIF4EBP1 and EIF4EBP2. The first intron is more ancestral as it is the only one present in the 4E-BP gene from D. melanogaster, where it occupies the same position that is observed in the human EIF4EBP genes. Because Drosophila possesses a single 4E-BP gene, and the three functional human genes are located on different chromosomes, all members of the mammalian EIF4EBP family most likely originated from the duplication of an original gene. The quantitative analysis of m4E-BP3, mMASK-BP3, and mMASK mRNA in mouse tissues revealed that their respective expression levels and patterns differ considerably (Fig. 1B). This is surprising in light of the assumption that mMASK and mMASK-BP3 should be transcribed from a common promoter. Thus it is tempting to speculate that the expression of the mRNA for mMASK-BP3 is regulated at the level of splicing, which would logically explain the differences that we observe between mMASK and mMASK-BP3. Such a regulation would have to be tissue-specific, because each mRNA is preferentially expressed in a different subset of tissues (Fig. 1B). An intriguing observation about MASK, MASK-BP3, and 4E-BP3 is that their respective expression patterns are very different in mouse and human (compare Fig. 1 with Fig. 3). For example, h4E-BP3 has a broad expression domain (Fig. 3B), whereas m4E-BP3 is almost exclusively expressed in the liver and the kidney (Fig. 1). Moreover, and in contrast to mouse tissues, the three human transcripts display very similar expression patterns (Fig. 3, B–D). This implies that the two loci are still undergoing change in mammalian lineages and could be subjected to differential selective pressure. Moreover it indicates that studies of MASK, MASK-BP3, and 4E-BP3 function in the mouse will have to be interpreted carefully before being extrapolated to humans. Dissection of the promoter regions of both human and mouse 4E-BP3 and MASK will be required to get a better understanding of the molecular basis for the differences in expression.

The structure of EIF4EBP3 allowed us to investigate the origin of an alternative transcript that we had observed previously (11). We isolated the corresponding cDNA and observed that it is derived from the fusion of EIF4EBP3 with a gene located upstream, the human homologue of D. melanogaster MASK (23). In the fly, MASK is an effector of receptor tyrosine-kinase signaling, and it functions either downstream of Ras/ MAPK or on a parallel pathway. Moreover, the Drosophila MASK is critical for photoreceptor differentiation, cell survival, and proliferation (23). We cloned the human MASK cDNA and used it to determine the structure of the MASK gene. We thus established that the fused MASK-BP3 transcript arises from readthrough transcription of MASK. This is the simplest explanation for our observations, given that MASK-BP3 incorporates an intermediate exon that is located between MASK and EIF4EBP3 (Fig. 2, A and B). To bypass premature termination codons, the MASK-BP3 chimeric transcript employs alternative splicing to remove the exons containing the MASK stop codon and the 4E-BP3 translation initiation codon.

Analysis of the D. melanogaster genome revealed that 4E-BP...
FIG. 4. Alignment of hMASK homologous proteins. Regional alignments of hMASK conserved motifs with homologous regions from similar proteins. Black shading denotes identical residues, and gray shading denotes similar residues. The sequences of human MASK, Drosophila MASK, human GTAR, and mouse GTAR have been published in GenBank™ (accession numbers AF521882, AF425651, NM_032217, and AY026253, respectively). Macaque protein sequence is predicted from a cDNA clone (GenBank™ accession number AB049837). The sequences of mouse MASK, Anopheles gambiae EAA08897, and C. elegans R11A8.7 are derived from computational analyses of their respective genome sequences. A, alignment of the amino-terminal stretch of 15 ankyrin repeats. B, alignment of the carboxyl-terminal stretch of 10 ankyrin repeats. C, alignment of the KH domain.
and MASK are located on different chromosomes and, consequently, are different loci. In the human genome, however, these two loci are adjacent by several kilobases on chromosome 5q31, and a portion of RNA transcripts from the two genes is fused into a single mRNA. Searches of the GenBank™ EST data base revealed that such a fusion of MASK and EIF4EBP3 is observed only in mammals (Homo sapiens, Mus musculus, Rattus norvegicus, Sus scrofa, Bos taurus) and is therefore a relatively recent innovation. The gene fusion event may be the result of a weak termination signal for the entire second exon of EIF4EBP3 and the fusion of proteins often indicates that they are functionally interacting (26–28). Although we do not have any evidence that MASK and EIF4EBP3 are fused together, both genes express separate transcripts and are therefore likely to have kept their original separate functions. This could facilitate the appearance of a novel function following gene fusion. One additional possibility is that the flexibility of the two loci in mammals is enhanced by the fact that both genes possess paralogs that can continue to provide the original functionality; EIF4EBP3 expresses a protein that is structurally related to 4E-BP1 and 4E-BP2 (11), whereas MASK encodes a protein that is highly similar to GTAR (Fig. 4).

One peculiarity of MASK-BP3ARF is that it is translated in a reading frame that differs from 4E-BP3. Therefore, two different proteins are encoded by the same exon of EIF4EBP3, and MASK-BP3ARF does not have any similarity with 4E-BP3 at the level of its primary structure. We want to draw attention to the fact that the entire second exon of EIF4EBP3 is translated in an alternative reading frame such that no premature termination codon is observed in MASK-BP3 that could result in nonsense-mediated decay. As for the functionality of the additional sequence observed in MASK-BP3, it cannot be directly derived from the amino acid sequence, because the sequence encoded by the alternative reading frame does not share similarity with any known proteins. It has been shown that the fusion of proteins often indicates that they are functionally interacting (26–28). Although we do not have any evidence that MASK and 4E-BP3 interact, it remains possible that the two proteins participate in the same biochemical pathway. Indeed, the mammalian MASK may function downstream of Ras/MAPK (23), and the function of eIF4E is regulated by the Ras/MAPK-signaling pathway (29). It will therefore be interesting to investigate whether MASK function is in any way
modified by the additional exons present in MASK-BP3\(^{T6R}\) and if this impinges on eIF4E regulation.

There is only one report of a gene fusion similar to that of MASK and EIF4EBP3, the UVE1-Kua chimera (30). The use of overlapping reading frames in mammalian genes is also very rare, having been described only for the INK4a/ARF (31) and the X\(\alpha\)/ALEX loci (31, 32). We have presented here the first instance where both gene fusion and utilization of an alternative reading frame occur in the same cellular transcript. Thus, we have uncovered a complex mechanism that can be used in mammalian genomes to evolve new functions from existing genes.

Acknowledgments—We thank Colin Lister for excellent technical assistance, Jarred Chicoine, David Lubell, and Sandra Brouillard for help with MASK cloning, Hiroaki Imataka for pcDNA3-HA, and Mathieu Miron for critical reading of the manuscript.

REFERENCES

1. Poulin, F., and Sonenberg, N. (2003) in Translation Mechanisms (Lapointe, J., and Brakier-Gingras, L., eds) pp. 280–297, Landes Bioscience, Austin, TX
2. Mathews, M. B., Sonenberg, N., and Hershey, J. W. B. (2000) in Translational Control of Gene Expression (Sonenberg, N., Hershey, J. W. B., and Mathews, M. B., eds) pp. 1–31, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
3. Shatkin, A. J. (1985) Cell 40, 223–224
4. Gingras, A.-C., Raught, B., and Sonenberg, N. (1999) Annu. Rev. Biochem. 68, 913–963
5. Raught, B., and Gingras, A.-C. (1999) Int. J. Biochem. Cell Biol. 31, 43–57
6. Haghghati, A., Mader, S., Pause, A., and Sonenberg, N. (1995) EMBO J. 14, 5701–5709
7. Marrotrigiano, J., Gingras, A.-C., Sonenberg, N., and Burley, S. K. (1999) Mol. Cell 3, 707–716
8. Gingras, A.-C., Raught, B., Gyg, P., Niedzwiecka, A., Miron, M., Burley, S. K., Polakiewicz, R. D., Wyslouch-Cieszyńska, A., Aebbersold, R., and Sonenberg, N. (2001) Genes Dev. 15, 2852–2864
9. Gingras, A.-C., Gyg, P., Raught, B., Polakiewicz, R. D., Abraham, R. T., Hekstra, M. F., Aebbersold, R., and Sonenberg, N. (1999) Genes Dev. 13, 1422–1437
10. Pause, A., Belsham, G. J., Gingras, A.-C., Donzé, O., Lin, T. A., Lawrence, J. C., Jr., and Sonenberg, N. (1994) Nature 371, 762–767