ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 333 (2004) 35–46
GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES
www.elsevier.com/locate/gene

The in vivo form of the murine class VI POU protein Emb is larger than that encoded by previously described transcripts☆

F. Relaix[a], S. Molinari[b], M. Lemonnier[a], B. Schäfer[c], M. Buckingham[a,*]

[a] C.N.R.S. URA 2578, Department of Developmental Biology, Pasteur Institute, 75724 Paris Cedex 15, France
[b] Dipartimento di Scienze Biomediche, Università degli studi di Modena, 41100 Modena, Italy
[c] Department of Pediatrics, University of Zurich, CH-8032 Zurich, Switzerland

Received 25 June 2003; received in revised form 28 November 2003; accepted 5 February 2004

Available online 23 April 2004

Abstract

The class VI POU domain family member known as Emb in the mouse (rat Brn5 or human mPOU/TCFβ1) is present in vivo as a protein migrating at about 80 kDa on western blots, considerably larger than that predicted (about 42 kDa) from previously cloned coding sequences. By RT-PCR and 5′ RACE strategies a full-length Emb sequence, Emb FL, is now identified. Shorter sequences encoding the –COOH terminal, and an $-NH_2$ terminal isoform, EmbN, were also isolated. Comparisons of Emb coding sequences between species, including the full-length zebra fish, POU(c), are presented, together with a compilation of the multiple transcripts produced by alternative splicing and the presence of different transcriptional start and stop sites, from the *Emb* gene.

Keywords: Class VI POU domain protein; Emb; Brn5; TCFβ1; mPOU; POU(c)

1. Introduction

The POU proteins constitute a large family of transcription factors, characterized by the presence of a classic homeobox domain and a POU domain, so-called because it was first identified in Pit1, Oct1 and Unc86. These two DNA binding motifs are separated by a linker region of variable length which confers versatility to DNA binding site recognition and also permits the protein to adopt a number of conformations leading to different regulatory possibilities. POU proteins bind as monomers or dimers, may themselves have transcriptional activation and/or repression domains and can interact with a wide range of other transcription factors and co-factors. Seven subclasses of POU domain proteins have been distinguished, mainly on the basis of the length and composition of the linker sequence (see Ryan and Rosenfeld, 1997; Phillips and Luisi, 2000). Many POU protein genes are expressed at a high level in the central nervous system, with lower level expression in other cell types; Pit1 (class I), for example, is specific to certain types of neurons and to the pituitary gland, whereas Oct1 (class II) is ubiquitously expressed. In some cases, target genes are clearly identified as for Oct1, which is required for activation of the histone 2B promoter (Hinkley and Perry, 1992). Mutant analysis has also pointed to a role for a number of POU proteins, such as Unc86 in *C. elegans*, or its Brn3 homologues in mammals, in the specification of particular types of neurons (see McEvilly and Rosenfeld, 2000).

The class VI member designated as Brn5 in rats (Andersen et al., 1993), TCFβ1 (Messier et al., 1993) or mPOU (Wey et al., 1994) in humans and Emb (Okamoto et al., 1993) or Cns-1 (Bulleit et al., 1994) in mice, has been described with a sequence of 301–303 amino acids, giving a protein with an estimated molecular weight of about 40 kDa. It is encoded by a single gene, the structure of which has been analyzed in mice (Okamoto et al., 1993) and humans (Messier et al., 1993). The murine *Emb* gene was found to have 5/6 exons with an open reading frame of 903 nucleotides, distributed over exons 2–6. The homeodomain

Abbreviations: EST, expressed sequence TAG; cDNA, DNA complementary to RNA; kDa, kilodalton(s); kb, kilobase(s).

☆ The nucleotide sequences reported in this paper have been submitted to the GenBank TM/EBI Data Bank with accession numbers AY259046 (EmbFL) and AY259047 (EmbN).

* Corresponding author. Tel.: +33-1-45-68-84-77.
E-mail address: margab@pasteur.fr (M. Buckingham).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.02.047

and 3′ part of the POU domain are encoded within exon 6 and the 5′ part of the POU domain within exon 5 (see also Messier et al., 1993). Analysis of different cDNA clones indicated that differential use of an acceptor splice site for intron 4 resulted in an additional two amino acids in exon 5, and that a truncated form of the protein (109 amino acids) was generated from an alternative, internal, ATG in intron 4 (clone C7). Further heterogeneity was observed in the 5′ non-coding sequence, due to use of alternative transcription initiation sites, resulting in the presence of an additional 5′ exon, exon 1, in cDNA clone C2 (Okamoto et al., 1993). The reported sizes of mRNA, analyzed on Northern blots—6, 5 and 2 kb—also suggest that a number of transcripts are generated from the gene (see also Andersen et al., 1993; Wey et al., 1994). Expression of *Emb*, based on Northern blots and in situ hybridization, is high in the brain and also detected in a number of tissues, including skeletal muscle and heart (Okamoto et al., 1993; Dominov and Miller, 1996). Transcripts are present in embryonic, fetal and postnatal mouse muscle (Dominov and Miller, 1996) and indeed both Emb and mPOU were cloned from muscle tissue (Wey et al., 1994; Dominov and Miller, 1996). Differentiating cells of the C2 and Sol8 muscle cell lines and NIH3T3 and C3H10T1/2 fibroblast lines also contain Emb mRNA (Dominov and Miller, 1996). In addition to the mammalian sequence encoding Brn5/mPOU/Emb, a homologue, POU(c), expressed ubiquitously in the zebra fish embryo, has been described (Johansen et al., 1993). This sequence has an open reading frame which extends further 5′, potentially generating a considerably larger protein of 578 amino acids, with an estimated molecular weight of about 63 kDa.

In this paper, we describe the identification of a number of transcripts generated from the murine *Emb* gene including such a longer sequence sharing strong similarities with the zebra fish POU(c) sequence. The open reading frame reveals, in addition to the homeo and POU domain, a third conserved domain found in bromodomain, containing chromatin remodelling proteins such as Brg-1 (Khavari et al., 1993). Western blot analysis shows that a protein migrating at 80 kDa, encoded by the longer transcript, is the major endogenous Emb form present in many mouse tissues, including brain and muscle, and in cell lines.

2. Material and methods

2.1. Cloning of the full-length cDNAS of EmbN and EmbFL

A homology screening of mouse Expressed Sequence Tag (EST) sequences was carried out using the NCBI TBLASTN search with the 5′ POU(c) protein sequence. The single mouse EST clone AA009167 was identified and used for designing the following primers to clone full-length Emb:

MPOUSTOP1: TGCGCCGGCACACGCTGACACTGAGCC
MPOUSTOP2: TACGGGATCTGAAAGACGTTCAGCTTGCTGG
MPOUSSAC2: ATTCCGCGGAGGCAACTGTGAAGTCGCTGCCCGG
MPOURSAC2: CTCCGCGGAATGGTCGGGGTCACATGACCCAGC
MPOUSAPA: TGGGGGCCCTGTTGAAGCCAGTGGACCTGC
MPOURAPA: AGGGCCCCCACTCTGGGAGCCACCAGCTGCCACAC
MPOUSSAL: AGTCGACGCCCAGCTTCCAGCCGAGGAGGAGAGC
MPOURSAL: GCGTCGACTCCCACCTCCAATACTCGGATGG

RT-PCR (using Advantage2 PCR enzyme, conditions according to the manufacturer's recommendations, Clontech) was performed using cDNA from E11.5 mouse embryos as the template. Two PCR fragments were obtained and cloned into the pGEM-T (Promega) TA cloning vector. Two rounds of 5′ RACE-PCR were performed for each pair of primers (nested PCR) to obtain full-length cDNA clones using an E11.5 mouse Marathon Ready cDNA library (Clontech) as the template. Overlapping 5′ RACE products (1.0, 1.2 and 2 kb) were cloned in the pGEM-T (Promega) TA cloning vector and sequenced by automated DNA sequencing in both directions using vector-specific primers and custom made primers. RT-PCR and RACE products were assembled by standard cloning techniques. The cloning strategy is diagrammed in Fig. 2.

2.2. Expression vectors

EmbN and EmbFL cDNAs were cloned in the pCDNA3.1+ vector (Invitrogen) for expression studies. The Flag-Tag (IBI, Kodak) empty vector was generated using annealed and phosphorylated oligonucleotides cloned into pCDNA3.1+. To generate Flag-tagged EmbN and EmbFL, PCR was utilized to allow in frame standard cloning into the empty PCDAN3.1-Flag vector, followed by sequencing.

2.3. Cell culture and transfection experiments

C2/7 is a subclone of the original C2 skeletal muscle cell line derived from C3H mice (Yaffé and Saxel, 1977). C2/7 cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing 20% fetal calf serum (FCS) at low density and, when approaching confluence, induced to differentiate with DMEM containing 2% FCS.

The 293 cell line is derived from a human embryonal kidney cell line (Graham et al., 1977) and was grown in DMEM containing 20% FCS.

FGC4 is a rat hepatoma cell line (Angrand et al., 1990); FGC4 cells were cultured in Ham's modified F12 medium supplemented with 5% FCS.

10T1/2 is a mouse embryonic fibroblast cell line derived from C3H mice (Reznikoff et al., 1973) and was grown in DMEM containing 10% FCS.

BC3H is a mouse skeletal muscle-like cell line derived from a brain tumor (Schubert et al., 1974) ; BC3H cells were cultured in DMEM containing 20% FCS.

Cells were grown in 100 mm diameter dishes to 50% confluence and transfected by the calcium phosphate method (see Biben et al., 1994) using 10 μg of expression vector. Precipitates were allowed to form for 30 min at room temperature. Cells were incubated with the DNA-calcium phosphate precipitates overnight, then rinsed twice with DMEM and harvested 48 h after transfection.

2.4. Nuclear and tissue extracts

Nuclear extracts were prepared using a modification of the method of Dignam et al. (1983) (Ferrari et al., 1994). Tissue extracts were prepared from embryonic day (E) 11.5 mouse embryos as described previously (Daubas et al., 2000). Nuclear and cellular extracts were aliquoted and conserved in liquid nitrogen. Protein concentration was measured using the method of Bradford (Biorad assay kit).

2.5. Western blot analysis

For the analysis of nuclear or tissue extracts, 10 μg of proteins were diluted in 1 × SDS gel loading buffer (50mM Tris−HCl pH 6.8, 100mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, 10% glycerol) and run in a 10% SDS/polyacrylamide gel. After electrophoresis, the proteins were transferred to a poly(vinylidene difluoride) membrane (Amersham). Filters were blocked by incubation with PBST (phosphate buffered saline—PBS, containing 0.1% tween-20), with added 5% low fat dry milk, for 1 h and then incubated with the primary antibodies (1:1000) for 2 h. The following antibodies were used: rabbit polyclonal anti-mPOU antibody and the corresponding pre-immune serum, mouse monoclonal FLAG antibody (SIGMA). Filters were extensively washed in PBST and then incubated with horseradish peroxidase conjugated secondary anti-rabbit or anti-mouse antibody (Amersham) (1:5000) for 1 h. After further extensive washes with PBST, antigen−antibody complexes were visualized with an enhanced chemiluminescence kit (ECL plus, Amersham).

2.6. Immunofluorescence

Thirty-six to forty-eight hours after transfection, cells were washed extensively with PBS, fixed for 20 min at room temperature with 3% paraformaldehyde in PBS, permeabilized with 0.05% Triton X-100 in PBS for 5 min, and incubated for 15 min with PBS containing 1% Bovine Serum Albumin (BSA). Cells were then incubated overnight at 4 °C with the monoclonal anti FLAG antibody (1:100), washed with PBS, incubated for 15 min with 1% BSA in PBS and incubated for a further 60 min with rhodamine conjugated goat anti-mouse IgG (1:200, Pierce). After extensive washing with PBS, cell monolayers were mounted in 10 mM Tris−HCl, pH 9, containing 60% glycerol and examined with a Zeiss Axiophot fluorescence microscope. Images were acquired using a digital camera and the SPOT32 software package (Diagnostic Instruments) and exported into Adobe Photoshop.

2.7. Immunoprecipitation

Proliferating C2/C7 cells in 100 mm dishes were transfected with 5 μg of FLAG-EMBN, FLAG-EMBFL or FLAG-EMB expression vector; 36 h after transfection, cells were washed twice with PBS and lysed in 1 ml of lysis buffer (50 mM Tris−HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100). The cell lysate was centrifuged for 10 min at 12,000 × *g* and the supernatant used for immunoprecipitation. To immunoprecipitate the FLAG fusion proteins, 20 μl of ANTI-FLAG M2 Affinity gel (SIGMA) were added to the cell lysates and the immunoprecipitation reaction was left to proceed at 4 °C for 2 h under rotation. The ANTI-FLAG M2 Affinity gel was collected and washed twice with 1 ml of lysis buffer and once with 1 ml of 10 mM Tris−HCl, pH 7.5 and 0.1% NP40. The immunoprecipitated proteins were eluted by resuspending the affinity gel in 50 μl of 2× SDS gel loading buffer and boiling for 5 min. The eluted proteins were run in a 10% SDS polyacrylamide gel and analyzed by western blot with anti-FLAG or anti-mPOU antibodies.

2.8. In vitro translation

The FLAG-EMBN, FLAG-EMBFL or FLAG-EMB fusion proteins were in vitro translated in the TnT Coupled Reticulocyte Lysate System (Promega) using plasmid vectors where the coding sequences are under the control of the bacteriophage T7 RNA polymerase promoter; 1 μg of plasmid vector was incubated with 25 μl of TnT lysate, 1 μl of 1 mM amino acid mixture minus methionine, 4 μl of ^{35}S-methionine (1000 Ci/mmol) at 10 mCi/ml, and 1 μl of T7 TnT RNA polymerase, for 1 h at 30 °C. The products of the in vitro translation were run in a 10% SDS polyacrylamide gel, which was subsequently dried under vacuum and exposed to an autoradiography film.

3. Results

3.1. The size and tissue distribution of the endogenous Emb protein

In the course of our analysis of a distal enhancer element situated 5′ to the mouse *cardiac actin* gene, we had identified a regulatory complex, present in extracts from the C2 muscle cell line, which includes Emb, a POU protein, the myogenic regulator Mef2D and the histone

deacetylase p300 (Molinari et al., 2004). However, on western blots of the muscle cell extracts, instead of a protein with the expected size of about 40 kDa (301 amino acids) (Okamoto et al., 1993; Wey et al., 1994), a protein of 80 kDa was detected. This is shown in Fig. 1. The antibody used is directed against the human mPOU protein (Wey et al., 1994) corresponding to the −COOH terminal part of the Emb sequence. It is specific for this POU domain class and does not interact with Oct1 for example (Molinari et al., 2004). In Fig. 1A, western blots of extracts from different mouse cell lines are shown. As we had observed previously, in dividing myoblasts and differentiated myotubes of the C2 muscle cell line, a major 80 kDa form is seen. This is also the case in muscle cells of the BC3H cell line and in C3H10T1/2 embryonic fibroblasts. In hepatocytes of the FGC4 line, two slightly larger bands are seen which are not evident with the pre-immune serum. In western blots with tissue extracts (Fig. 1B), a major band at 80 kDa is again seen with preparations from limb, tail, thorax and head of 11.5 day mouse embryos. Thorax (and limb) are enriched in skeletal muscle, while the brain (mesencephalon and diencephalon) is a major component of the head extract. With extracts from adult cerebellum there is a major 80 kDa protein. In adult liver and heart, two slightly larger proteins are seen, as in the case of the liver cell extracts. A larger protein is also just detectable in adult skeletal muscles. Other minor forms are also seen. In the cerebellum, the band at about 42 kDa would also appear to be above background. These observations suggest alternative splicing of the *Emb* gene in vivo.

3.2. Identification of further Emb coding sequences

The principal Emb form of about 80 kDa is much larger than the size predicted by the Emb/Brn5 coding sequences

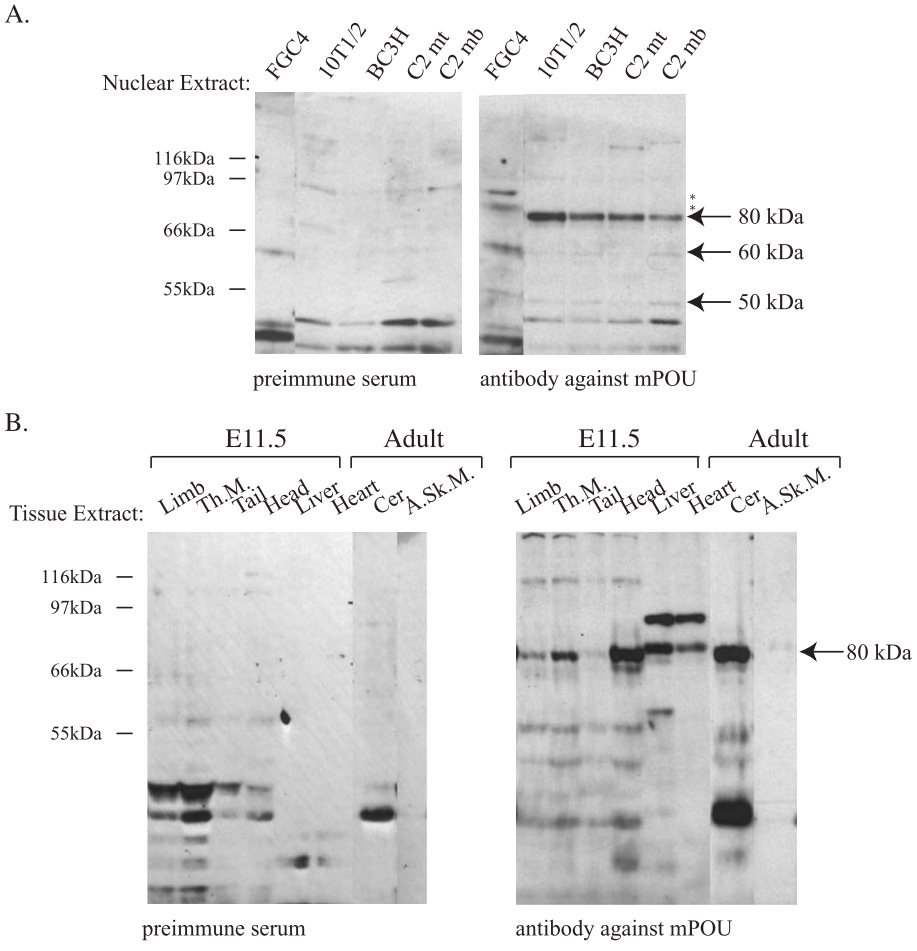


Fig. 1. Detection of Emb isoforms in different cell lines and tissues, by western blot analysis, using an antibody to the human homologue, mPOU. (A) The mPOU antibody reveals a band with an apparent molecular weight of 80 kDa present in nuclear extracts from the mouse embryonic fibroblast line, 10T1/2, the BC3H muscle cell line, and dividing myoblasts (mb) and differentiated myotubes (mt) of the C2 mouse skeletal muscle cell line. In the FGC4 rat liver cell line, two slightly larger bands (>80 kDa, asterisks) are detected which are not seen with the pre-immune serum. Additional minor bands are also detected in the different cell lines. (B) The mPOU antibody reveals a major band with an apparent molecular weight of 80 kDa in tissue preparations of limb, thorax (Th. M.), which is enriched in skeletal muscle, and head (mesencephalon and diencephalon) from 11.5-day-old mouse embryos (E11.5). This is also a major form in adult cerebellum (Cer.), where a second band, with a molecular weight of about 42 kDa, also appears to be enriched relative to an equivalent band in the pre-immune serum. In adult liver and heart, two larger bands (>80 kDa, asterisks) are detected. The smaller of the two is detectable in adult skeletal muscle (A. Sk. M.). Other minor bands are detected with the antibody and not the pre-immune serum in some tissues.

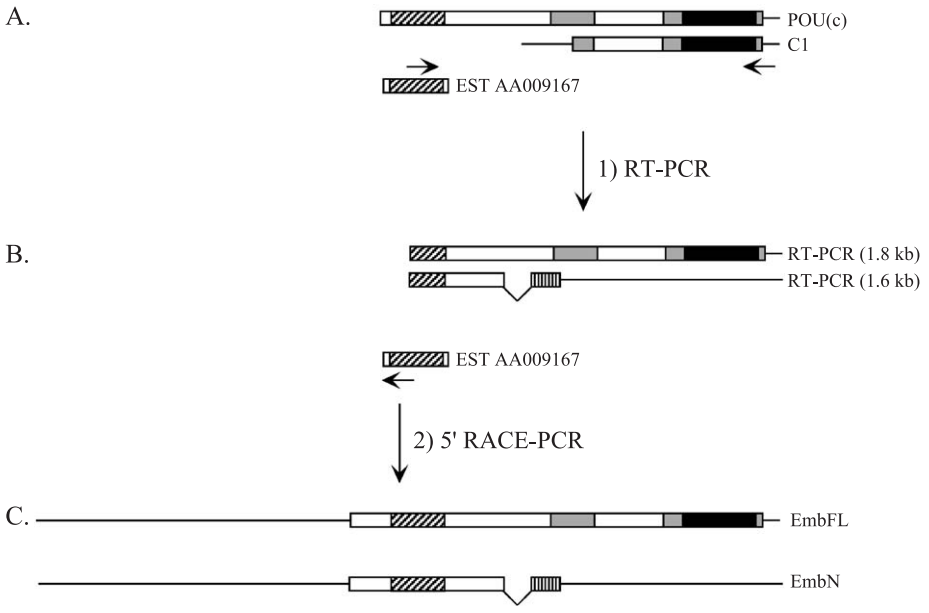
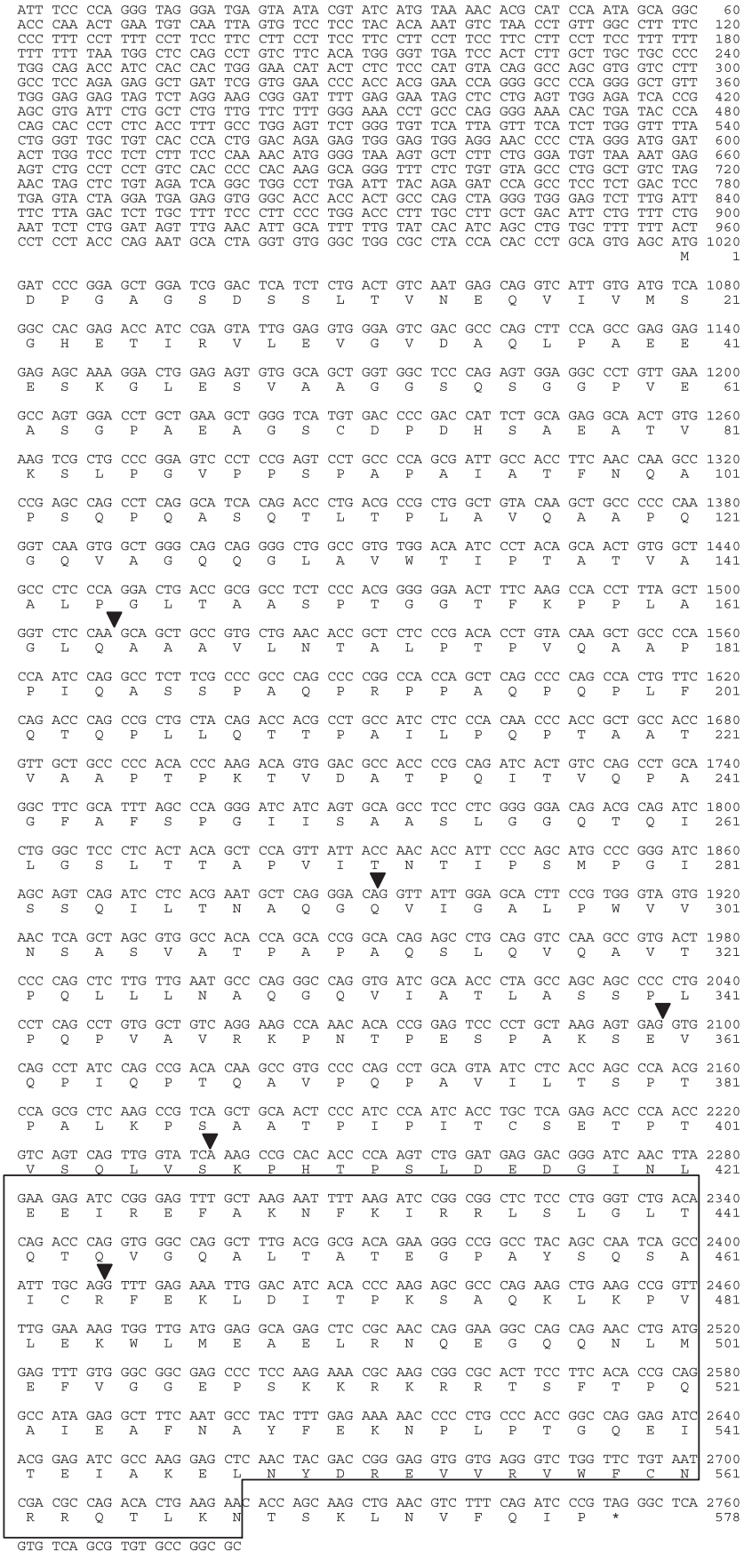


Fig. 2. Cloning strategy to obtain EmbFL and EmbN splice variants. (A) A TBLASTn search using the 5′ region of POU(c) revealed a single mouse EST clone (accession number AA009167) suggesting that Emb (CI, Okamato et al., 1993) may have an incomplete 5′ sequence. (B) RT-PCR on cDNA to RNA from whole E11.5 mouse embryos gave two major bands of 1.6 and 1.8 kb which were cloned, sequenced and revealed two splice variants of the coding sequence (sequences in Fig. 3A and B). (C) 5′ RACE-PCR using a Matchmaker E11.5 mouse embryo cDNA library (Clontech) was performed to obtain the full-length 5′ UTR (2 kb). The full-length cDNA clones were generated by ligating the RT-PCR product to the 5′ RACE products using the silent restriction sites introduced in the PCR primers. The POU domain is boxed in black, the 5′ acidic domain is indicated as diagonal stripes, other conserved sequences are in grey. The divergent sequence in EmbN is indicated as vertical stripes.

(42 kDa). This surprising result led us to re-investigate transcripts produced by the *Emb* gene. Previous results on the expression of Emb/mPOU/Brn5 had been based on transcript analysis using Emb probes. Since a transcript with a longer open reading frame, POU(c), had been described in zebra fish (Johansen et al., 1993), we looked in the mouse EST database for sequences homologous to the POU(c) sequence, lying 5′ to the region which overlaps with the previously described cDNA clones, C1 and C2 for mouse Emb (Okamoto et al., 1993). A single mouse EST, EST AAOO9167, was identified at the time and used to generate primers, together with 3′ Emb sequences, which were used for 5′ RACE PCR and RT-PCR with cDNAs from embryonic day (E) 11.5 mouse embryos, as indicated in Fig. 2. Assembly of the RT-PCR and 5′ RACE PCR products resulted in two sequences, EmbFL and EmbN. The latter corresponds to the $-NH_2$ part of EmbFL, with an additional divergent 3′ domain due to alternative splicing. EmbN has an open reading frame of 180 amino acids and EmbFL of 578 amino acids. The sequence of EmbFL is presented in Fig. 3A, and the divergent sequence of the truncated form, EmbN, in Fig. 3B. In the course of this study, a large number of new ESTs have appeared in the databases, confirming that EmbFL encodes a longer form of Emb which is the most abundant transcript generated by the *Emb* gene. However, shorter transcripts such as EmbN or the other previously described Emb isoforms, although detectable by RT-PCR experiments (data not shown), appear to be rare transcripts, not found in the EST databases. In order to correlate what is seen on western blots with the Emb coding sequences, we transfected Emb expression vectors into cells and identified the resultant protein also by western blotting. Since EmbN lacks the 3′ part of Emb which codes for the −COOH terminal POU domain containing region, recognised by the mPOU antibody, we added a FLAG tag to the NH_2 termini of EmbN, EmbFL and Emb. Transfection of these sequences into 293 or FGC4 cells, followed by western blotting of cell extracts, using the anti-mPOU antibody, shows that EmbFL encodes a protein which migrates at about 80 kDa, in the FLAG tag fusion configuration, slightly faster than the endogenous protein (Fig. 4A), which in 293 cells, used for their high transfection efficiency, also migrates at 80 kDa. Immunoprecipitation with the FLAG antibody, followed by western blot analysis with this or the mPOU antibody, confirms the result for EmbFL and shows that EmbN and Emb sequences encode proteins which migrate at about 32 and 42 kDa, respectively (Fig. 4B). In vitro translation of EmbN and Emb also confirmed this result. In the case of EmbFL it was difficult to detect full-length (80 kDa) ^{35}S labelled protein (Fig. 4C). Other observations (not shown) also suggest that this isoform is susceptible to proteolytic attack following overexpression in cells.

Immunofluorescent detection of tagged EmbFL or EmbN, after transfection into C2 muscle cells, showed that the protein was present in both the nucleus and cytoplasm of myoblasts and myotubes. In contrast, Emb was mainly nuclear in these experiments. In some of the transfection

A.


ATT TCC CCA GGG TAG GGA TGA GTA ATA CGT ATC ATG TAA AAC ACG CAT CCA ATA GCA GGC 60
ACC CAA ACT GAA TGT CAA TTA GTG TCC TCC TAC ACA AAT GTC TAA CCT GTT GGC CTT TTC 120
CCC TTT CCT TTT CCT TCC TTC CTT CCT TCC TTC CTT CCT TCC TTC CTT CCT TCC TTT TTT 180
TTT TTT TAA TGG CTC CAG CCT GTC TTC ACA TGG GGT TGA TCC ACT CTT GCT TGC TGC CCC 240
TGG CAG ACC ATC CAC CAC TGG GAA CAT ACT CTC TCC CAT GTA CAG GCC AGC GTG GTC CTT 300
GCC TCC AGA GAG GCT GAT TCG GTG GAA CCC ACC ACG GAA CCA GGG GCC CCA GGG GCT GTT 360
TGG GAG GAG TAG TCT AGG AAG CGG GAT TTT GAG GAA TAG CTC CTG AGT TGG AGA TCA CCG 420
AGC GTG ATT CTG GCT CTG TTG TTC TTT GGG AAA CCT GCC CAG GGG AAA CAC TGA TAC CCA 480
CAG CAC CCT CTC ACC TTT GCC TGG AGT TCT GGG TGT TCA TTA GTT TCA TCT TGG GTT TTA 540
CTG GGT TGC TGT CAC CCA CTG GAC AGA GAG TGG GAG TGG AGG AAC CCC CTA GGG ATG GAT 600
ACT TGG TCC TCT CTT TCC CAA AAC ATG GGG TAA AGT GCT CTT CTG GGA TGT TAA AAT GAG 660
AGT CTG CCT CCT GTC CAC CCC CAC AAG GCA GGG TTT CTC TGT GTA GCC CTG GCT GTC TAG 720
AAC TAG CTC TGT AGA TCA GGC TGG CCT TGA ATT TAC AGA GAT CCA GCC TCC TCT GAC TCC 780
TGA GTA CTA GGA TGA GAG GTG GGC ACC ACC ACT GCC CAG CTA GGG TGG GAG TCT TTG ATT 840
TTC TTA GAC TCT TGC TTT TCC CTT CCC TGG ACC CTT TGC CTT GCT GAC ATT CTG TTT CTG 900
AAT TCT CTG GAT AGT TTG AAC ATT GCA TTT TTG TAT CAC ATC AGC CTG TGC TTT TTT ACT 960
CCT CCT ACC CAG AAT GCA CTA GGT GTG GGC TGG CGC CTA CCA CAC CCT GCA GTG AGC ATG 1020
M 1
GAT CCC GGA GCT GGA TCG GAC TCA TCT CTG ACT GTC AAT GAG CAG GTC ATT GTG ATG TCA 1080
D P G A G S D S S L T V N E Q V I V M S 21
GGC CAC GAG ACC ATC CGA GTA TTG GAG GTG GGA GTC GAC GCC CAG CTT CCA GCC GAG GAG 1140
G H E T I R V L E V G V D A Q L P A E E 41
GAG AGC AAA GGA CTG GAG AGT GTG GCA GCT GGT GGC TCC CAG AGT GGA GGC CCT GTT GAA 1200
E S K G L E S V A A G G S Q S G G P V E 61
GCC AGT GGA CCT GCT GAA GCT GGG TCA TGT GAC CCC GAC CAT TCT GCA GAG GCA ACT GTG 1260
A S G P A E A G S C D P D H S A E A T V 81
AAG TCG CTG CCC GGA GTC CCT CCG AGT CCT GCC CCA GCG ATT GCC ACC TTC AAC CAA GCC 1320
K S L P G V P P S P A P A I A T F N Q A 101
CCG AGC CAG CCT CAG GCA TCA CAG ACC CTG ACG CCG CTG GCT GTA CAA GCT GCC CCC CAA 1380
P S Q P Q A S Q T L T P L A V Q A A P Q 121
GGT CAA GTG GCT GGG CAG CAG GGG CTG GCC GTG TGG ACA ATC CCT ACA GCA ACT GTG GCT 1440
G Q V A G Q Q G L A V W T I P T A T V A 141
GCC CTC CCA GGA CTG ACC GCG GCC TCT CCC ACG GGG GGA ACT TTC AAG CCA CCT TTA GCT 1500
A L P G L T A A S P T G G T F K P P L A 161
GGT CTC CAA GCA GCT GCC GTG CTG AAC ACC GCT CTC CCG ACA CCT GTA CAA GCT GCC CCA 1560
G L Q A A A V L N T A L P T P V Q A A P 181
CCA ATC CAG GCC TCT TCG CCC GCC CAG CCC CGG CCA CCA GCT CAG CCC CAG CCA CTG TTC 1620
P I Q A S S P A Q P R P P A Q P Q P L F 201
CAG ACC CAG CCG CTG CTA CAG ACC ACG CCT GCC ATC CTC CCA CAA CCC ACC GCT GCC ACC 1680
Q T Q P L L Q T T P A I L P Q P T A A T 221
GTT GCT GCC CCC ACA CCC AAG ACA GTG GAC GCC ACC CCG CAG ATC ACT GTC CAG CCT GCA 1740
V A A P T P K T V D A T P Q I T V Q P A 241
GGC TTC GCA TTT AGC CCA GGG ATC ATC AGT GCA GCC TCC CTC GGG GGA CAG ACG CAG ATC 1800
G F A F S P G I I S A A S L G G Q T Q I 261
CTG GGC TCC CTC ACT ACA GCT CCA GTT ATT ACC AAC ACC ATT CCC AGC ATG CCC GGG ATC 1860
L G S L T T A P V I T N T I P S M P G I 281
AGC AGT CAG ATC CTC ACG AAT GCT CAG GGA CAG GTT ATT GGA GCA CTT CCG TGG GTA GTG 1920
S S Q I L T N A Q G Q V I G A L P W V V 301
AAC TCA GCT AGC GTG GCC ACA CCA GCA CCG GCA CAG AGC CTG CAG GTC CAA GCC GTG ACT 1980
N S A S V A T P A P A Q S L Q V Q A V T 321
CCC CAG CTC TTG TTG AAT GCC CAG GGC CAG GTG ATC GCA ACC CTA GCC AGC AGC CCC CTG 2040
P Q L L L N A Q G Q V I A T L A S S P L 341
CCT CAG CCT GTG GCT GTC AGG AAG CCA AAC ACA CCG GAG TCC CCT GCT AAG AGT GAG GTG 2100
P Q P V A V R K P N T P E S P A K S E V 361
CAG CCT ATC CAG CCG ACA CAA GCC GTG CCC CAG CCT GCA GTA ATC CTC ACC AGC CCA ACG 2160
Q P I Q P T Q A V P Q P A V I L T S P T 381
CCA GCG CTC AAG CCG TCA GCT GCA ACT CCC ATC CCA ATC ACC TGC TCA GAG ACC CCA ACC 2220
P A L K P S A A T P I P I T C S E T P T 401
GTC AGT CAG TTG GTA TCA AAG CCG CAC ACC CCA AGT CTG GAT GAG GAC GGG ATC AAC TTA 2280
V S Q L V S K P H T P S L D E D G I N L 421
GAA GAG ATC CGG GAG TTT GCT AAG AAT TTT AAG ATC CGG CGG CTC TCC CTG GGT CTG ACA 2340
E E I R E F A K N F K I R R L S L G L T 441
CAG ACC CAG GTG GGC CAG GCT TTG ACG GCG ACA GAA GGG CCG GCC TAC AGC CAA TCA GCC 2400
Q T Q V G Q A L T A T E G P A Y S Q S A 461
ATT TGC AGG TTT GAG AAA TTG GAC ATC ACA CCC AAG AGC GCC CAG AAG CTG AAG CCG GTT 2460
I C R F E K L D I T P K S A Q K L K P V 481
TTG GAA AAG TGG TTG ATG GAG GCA GAG CTC CGC AAC CAG GAA GGC CAG CAG AAC CTG ATG 2520
L E K W L M E A E L R N Q E G Q Q N L M 501
GAG TTT GTG GGC GGC GAG CCC TCC AAG AAA CGC AAG CGG CGC ACT TCC TTC ACA CCG CAG 2580
E F V G G E P S K K R K R R T S F T P Q 521
GCC ATA GAG GCT TTC AAT GCC TAC TTT GAG AAA AAC CCC CTG CCC ACC GGC CAG GAG ATC 2640
A I E A F N A Y F E K N P L P T G Q E I 541
ACG GAG ATC GCC AAG GAG CTC AAC TAC GAC CGG GAG GTG GTG AGG GTC TGG TTC TGT AAT 2700
T E I A K E L N Y D R E V V R V W F C N 561
CGA CGC CAG ACA CTG AAG AAC ACC AGC AAG CTG AAC GTC TTT CAG ATC CCG TAG GGC TCA 2760
R R Q T L K N T S K L N V F Q I P * 578
GTG TCA GCG TGT GCC GGC GC

B.

```
ATT TCC CCA GGG TAG GGA TGA GTA ATA CGT ATC ATG TAA AAC ACG CAT CCA ATA GCA GGC   60
ACC CAA ACT GAA TGT CAA TTA GTG TCC TCC TAC ACA AAT GTC TAA CCT GTT GGC CTT TTC  120
CCC TTT CCT TTT CCT TCC TTC CTT CCT TCC TTC CTT CCT TCC TTC CTT CCT TCC TTT TTT  180
TTT TTT TAA TGG CTC CAG CCT GTC TTC ACA TGG GGT TGA TCC ACT CTT GCT TGC TGC CCC  240
TGG CAG ACC ATC CAC CAC TGG GAA CAT ACT CTC TCC CAT GTA CAG GCC AGC GTG GTC CTT  300
GCC TCC AGA GAG GCT GAT TCG GTG GAA CCC ACC ACG GAA CCA GGG GCC CCA GGG GCT GTT  360
TGG GAG GAG TAG TCT AGG AAG CGG GAT TTT GAG GAA TAG CTC CTG AGT TGG AGA TCA CCG  420
AGC GTG ATT CTG GCT CTG TTG TTC TTT GGG AAA CCT GCC CAG GGG AAA CAC TGA TAC CCA  480
CAG CAC CCT CTC ACC TTT GCC TGG AGT TCT GGG TGT TCA TTA GTT TCA TCT TGG GTT TTA  540
CTG GGT TGC TGT CAC CCA CTG GAC AGA GAG TGG GAG TGG AGG AAC CCC CTA GGG ATG GAT  600
ACT TGG TCC TCT CTT TCC CAA AAC ATG GGG TAA AGT GCT CTT CTG GGA TGT TAA AAT GAG  660
AGT CTG CCT CCT GTC CAC CCC CAC AAG GCA GGG TTT CTC TGT GTA GCC CTG GCT GTC TAG  720
AAC TAG CTC TGT AGA TCA GGC TGG CCT TGA ATT TAC AGA GAT CCA GCC TCC TCT GAC TCC  780
TGA GTA CTA GGA TGA GAG GTG GGC ACC ACC ACT GCC CAG CTA GGG TGG GAG TCT TTG ATT  840
TTC TTA GAC TCT TGC TTT TCC CTT CCC TGG ACC CTT TGC CTT GCT GAC ATT CTG TTT CTG  900
AAT TCT CTG GAT AGT TTG AAC ATT GCA TTT TTG TAT CAC ATC AGC CTG TGC TTT TTT ACT  960
CCT CCT ACC CAG AAT GCA CTA GGT GTG GGC TGG CGC CTA CCA CAC CCT GCA GTG AGC ATG 1020
                                                                          M     1

GAT CCC GGA GCT GGA TCG GAC TCA TCT CTG ACT GTC AAT GAG CAG GTC ATT GTG ATG TCA 1080
 D   P   G   A   G   S   D   S   S   L   T   V   N   E   Q   V   I   V   M   S    21

GGC CAC GAG ACC ATC CGA GTA TTG GAG GTG GGA GTC GAC GCC CAG CTT CCA GCC GAG GAG 1140
 G   H   E   T   I   R   V   L   E   V   G   V   D   A   Q   L   P   A   E   E    41

GAG AGC AAA GGA CTG GAG AGT GTG GCA GCT GGT GGC TCC CAG AGT GGA GGC CCT GTT GAA 1200
 E   S   K   G   L   E   S   V   A   A   G   G   S   Q   S   G   G   P   V   E    61

GCC AGT GGA CCT GCT GAA GCT GGG TCA TGT GAC CCC GAC CAT TCT GCA GAG GCA ACT GTG 1260
 A   S   G   P   A   E   A   G   S   C   D   P   D   H   S   A   E   A   T   V    81

AAG TCG CTG CCC GGA GTC CCT CCG AGT CCT GCC CCA GCG ATT GCC ACC TTC AAC CAA GCC 1320
 K   S   L   P   G   V   P   P   S   P   A   P   A   I   A   T   F   N   Q   A   101

CCG AGC CAG CCT CAG GCA TCA CAG ACC CTG ACG CCG CTG GCT GTA CAA GCT GCC CCC CAA 1380
 P   S   Q   P   Q   A   S   Q   T   L   T   P   L   A   V   Q   A   A   P   Q   121

GGT CAA GTG GCT GGG CAG CAG GGG CTG GCC GTG TGG ACA ATC CCT ACA GCA ACT GTG GCT 1440
 G   Q   V   A   G   Q   Q   G   L   A   V   W   T   I   P   T   A   T   V   A   141

GCC CTC CCA GGA CTG ACC GCG GCC TCT CCC ACG GGG GGA ACT TTC AAG CCA CCT TTA GCT 1500
 A   L   P   G   L   T   A   A   S   P   T   G   G   T   F   K   P   P   L   A   161

GGT CTC CAA GAT CAG TGC AGC CTC CCT CGG GGG ACA GAC GCA GAT CCT GGG CTC CCT CAC 1560
 G   L   Q   D   Q   C   S   L   P   R   G   T   D   A   D   P   G   L   P   H   181

TAC AGC TCC AGT TAT TAC CAA CAC CAT TCC CAG CAT GCC CGG GAT CAG CAG TCA GAT CCT 1620
 Y   S   S   S   Y   Y   Q   H   H   S   Q   H   A   R   D   Q   Q   S   D   P   201

CAC GAA TGC TCA GGG ACA GGT TAT TGG AGC ACT TCC GTG GGT AGT GAA CTC AGC TAG CGT 1680
 H   E   C   S   G   T   G   Y   W   S   T   S   V   G   S   E   L   S   *       239

GGC CAC ACC AGC ACC GGC ACA GAG CCT GCA GGT CCA AGC CGT GAC TCC CCA GCT CTT GTT 1740
GAA TGC CCA GGG CCA GGT GAT CGC AAC CCT AGC CAG CAG CCC CCT GCC TCA GCC TGT GGC 1800
TGT CAG GAA GCC AAA CAC ACC GGA GTC CCC TGC TAA GAG TGA GGT GCA GCC TAT CCA GCC 1860
GAC ACA AGC CGT GCC CCA GCC TGC AGT AAT CCT CAC CAG CCC AAC GCC AGC GCT CAA GCC 1920
GTC AGC TGC AAC TCC CAT CCC AAT CAC CTG CTC AGA GAC CCC AAC CGT CAG TCA GTT GGT 1980
ATC AAA GCC GCA CAC CCC AAG TCT GGA TGA GGA CGG GAT CAA CTT AGA AGA GAT CCG GGA 2040
GTT TGC TAA GAA TTT TAA GAT CCG GCG GCT CTC CCT GGG TCT GAC ACA GAC CCA GGT GGG 2100
CCA GGC TTT GAC GGC GAC AGA AGG GCC GGC CTA CAG CCA ATC AGC CAT TTG CAG GTT TGA 2160
GAA ATT GGA CAT CAC ACC CAA GAG CGC CCA GAA GCT GAA GCC GGT TTT GGA AAA GTG GTT 2220
GAT GGA GGC AGA GCT CCG CAA CCA GGA AGG CCA GCA GAA CCT GAT GGA GTT TGT GGG CGG 2280
CGA GCC CTC CAA GAA ACG CAA GCG GCG CAC TTC CTT CAC ACC GCA GGC CAT AGA GGC TTT 2340
CAA TGC CTA CTT TGA GAA AAA CCC CCT GCC CAC CGG CCA GGA GAT CAC GGA GAT CGC CAA 2400
GGA GCT CAA CTA CGA CCG GGA GGT GGT GAG GGT CTG GTT CTG TAA TCG ACG CCA GAC ACT 2460
GAA GAA CAC CAG CAA GCT GAA CGT CTT TCA GAT CCC GTA GGG CTC AGT GTC AGC GTG TGC 2520
CGG CGC                                                                      2526
```

Fig. 3. Predicted nucleotide and protein sequence of EmbFL and EmbN. (A) Sequence of EmbFL. Nucleotides and amino acids are numbered on the right. The POU domain is boxed. About 1 kb of 5′ UTR sequence is shown. Known splice sites are indicated by black arrowheads. The entire sequence is reported in GenBank submission AY259046. (B) Sequence of EmbN. Nucleotides and amino acids are numbered on the right. Known splice sites are indicated by black arrowheads. The divergent 3′ coding sequence, generated by a frame shift of this splice variant, is underlined. The entire sequence is reported in GenBank submission AY259047.

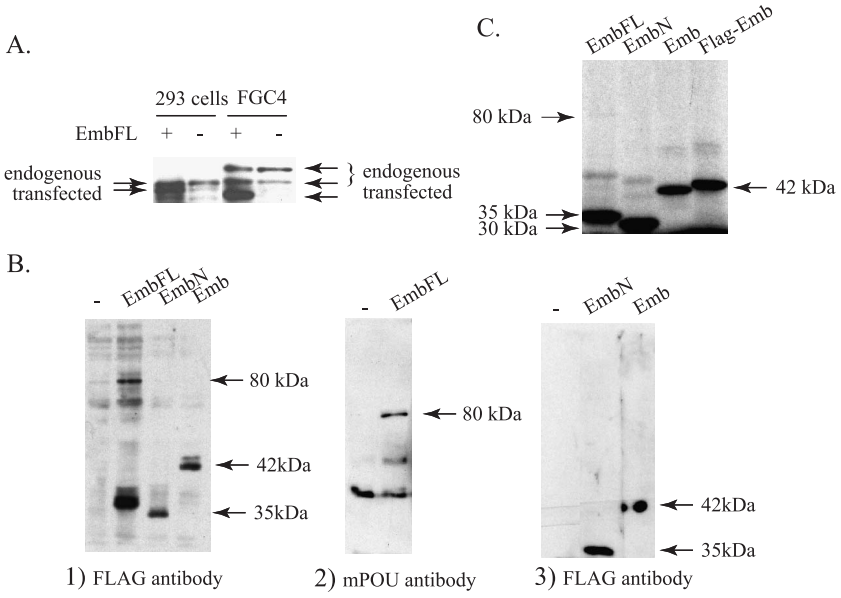


Fig. 4. Analysis of the proteins encoded by Emb, EmbFL and EmbN sequences. (A) Western blot analysis, using the antibody to the human Emb homologue, mPOU, with nuclear extracts from the human kidney cell line 293, and from the rat liver cell line, FGC4, transfected with the full-length EmbFL expression vector (+) or the empty vector (−). In each case the transfected protein, in the FLAG tagged configuration, migrates slightly faster than the endogenous protein. (B) Immunoprecipitation of FLAG-fusion proteins. The empty expression vector (−), full-length Emb (EmbFL), $-NH_2$ terminal form (Emb N) and −COOH terminal form (Emb) were transfected into myoblasts of the C2 muscle cell line. In (1), the products were revealed with a FLAG antibody after western blotting of the myoblast extracts. In (2) and (3), the extracts were first immunoprecipitated with the FLAG antibody and revealed with mPOU antibody (2) or FLAG antibody (3). EmbFL gives a band with an apparent molecular weight of 80 kDa and other smaller bands, EmbN gives a band with an apparent molecular weight of 30 kDa and Emb of 42 kDa. (C) Autoradiograph of ^{35}S in vitro translated EmbFL, EmbN, Emb and Emb with a FLAG tag, after separation by gel electrophoresis. A band migrating at 80 kDa is just detectable on in vitro translation of Emb FL. As expected, Emb N gives a band migrating at 32 kDa and Emb at about 42 kDa.

experiments, overexpression of EmbN tended to lead to cell death (data not shown).

3.3. Emb sequence comparisons between mouse and other species

In Fig. 5, the comparison between EmbFL, EmbN and the zebra fish sequence POU(c) (Johansen et al., 1993) is shown. EmbFL probably corresponds to a full-length sequence, based on that of POU (c) and also on the size of the protein encoded in transfection experiments which is similar to the major endogenous form in most cell types. The DNA binding domains (610–710) in the −COOH terminal part are highly conserved. An acidic domain at the $-NH_2$ terminal is also conserved between zebra fish and mouse sequences. This conserved region is also found in bromodomain containing proteins, such as Brg-1 (Khavari et al., 1993), implicated in chromatin remodelling (Fig. 5B). A human coding sequence, RDPF-1 (Zhou et al., 1996), also a member of the class VI family, while clearly not a homologue of mouse EmbFL, shows striking similarities, with conservation of the DNA binding POU domains and of some other regions of the sequence, also conserved between mammalian and zebra fish sequences.

A schematic representation of the isoforms of Emb is shown in Fig. 6A. The longest mammalian coding sequence previously described for this protein corresponds to the mouse C1/C2 Emb (Okamoto et al., 1993), rat Brn5 (Andersen et al., 1993) or human TCFβ1 (Messier et al., 1993).

Nucleotide alignments to the available human genomic sequence indicates that the Emb isoforms are encoded by at least eight exons as indicated in Fig. 6B.

4. Discussion

Visualization of the Emb protein by western blotting shows that it is the large 80 kDa form, referred to as EmbFL (full length) which accumulates in brain and other tissues where mammalian Emb/Brn5/mPOU mRNA had previously been detected. It is therefore a longer transcript, similar to that described as POU(c) in zebra fish (Johansen et al., 1993) which is the major sequence expressed from the *Emb* gene. As shown here, this extends for a further 821 bp of open reading frame 5′ to the previously identified ATG. This mammalian sequence has two conserved domains in the NH_2 terminus, the more 5′ of which is an acidic domain found in bromodomain containing proteins such as Brg1 (Khavari et al., 1993).

Comparison with the human genome sequence indicates that the coding sequence of EmbFL is composed of at least eight exons; the numbers of exons in the mouse 5′ UTR sequence, which is diverged from the human, is unknown.

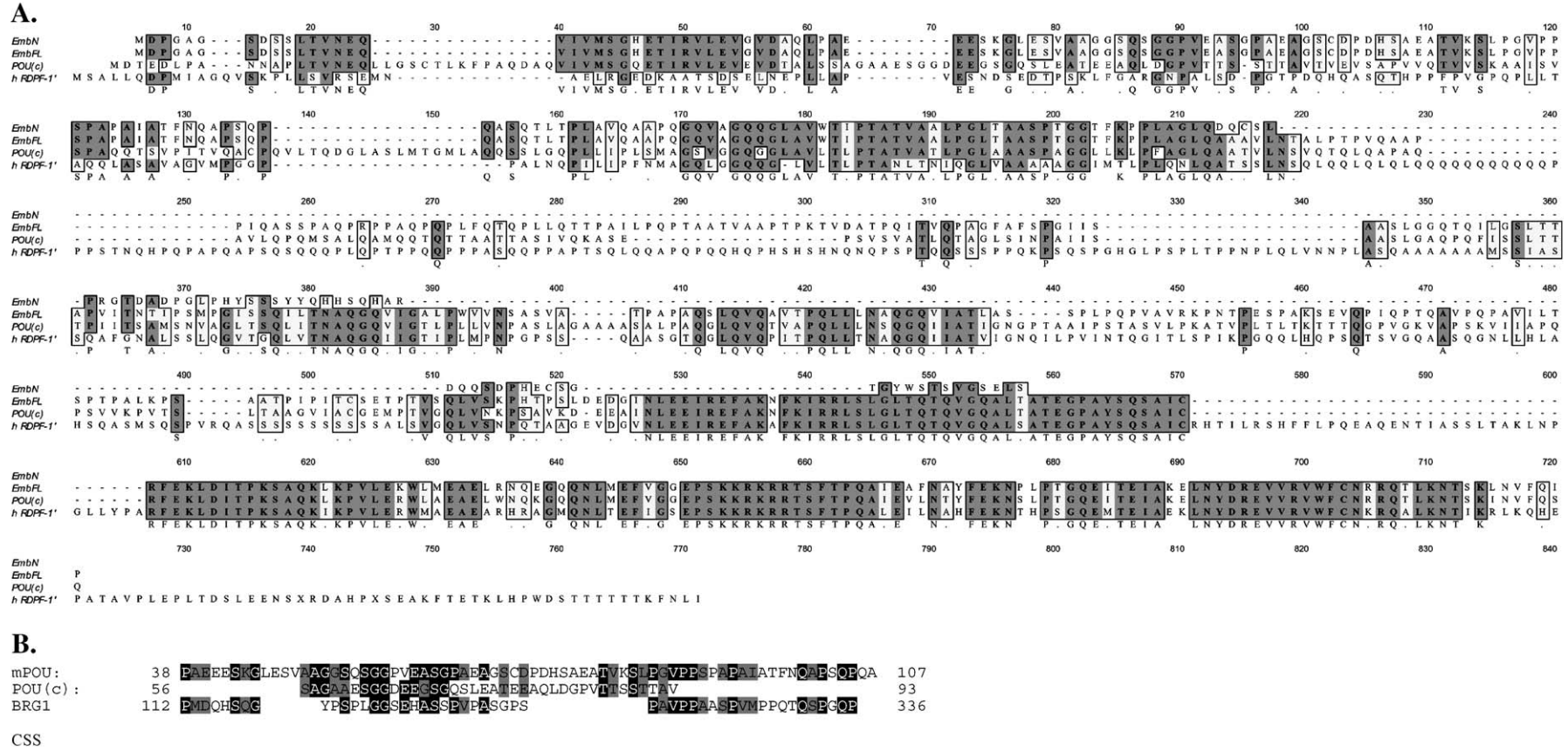


Fig. 5. Comparison of the predicted amino acid sequences of EmbN (mouse), EmbFL (mouse), zebra fish POU(c) (Johansen et al., 1993) and human RDPF-1 (Zhou et al., 1996), another POU VI subfamily protein. (A) Amino acid sequences are aligned by the CLUSTAL W programme (Mac Vector, IBI). Identical amino acids are highlighted in dark grey and similar residues in lighter grey. Dashes represent gaps to bring the sequences into alignment. Conserved residues between the four protein sequences are represented below the sequence alignments. (B) Alignment of the 5′ conserved mPOU sequence with BRG-1. Identical amino acids are highlighted in black and similar residues in grey. Dashes represent gaps to bring the sequences into alignment.

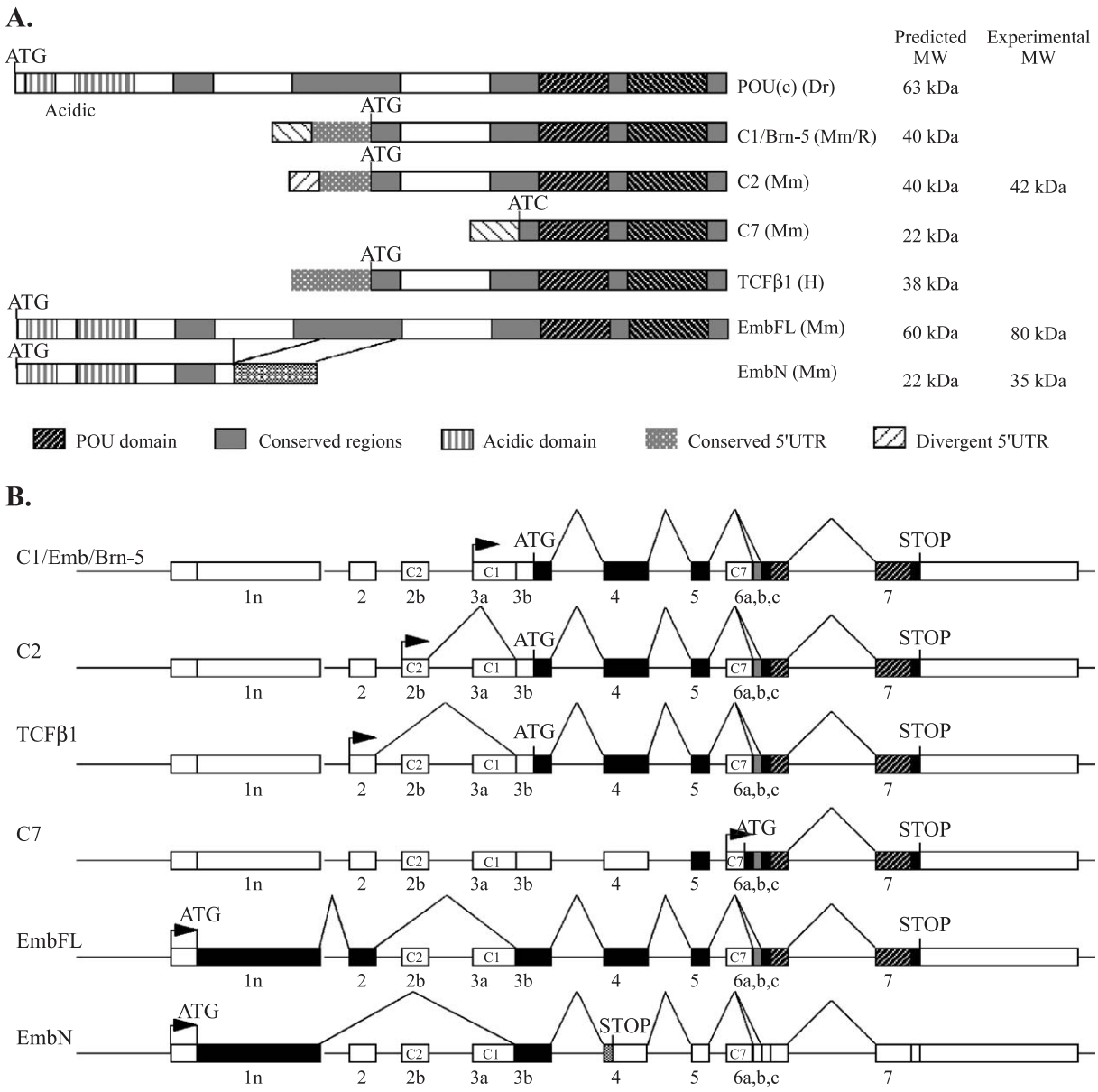


Fig. 6. Protein structure and genomic organisation of all the Emb isoforms identified from different species. (A) Protein domains of Emb isoforms identified as POU(c) in zebra fish (Dr) (Johansen et al., 1993), mouse(Mm) Emb/C1 (Okamoto et al., 1993) or rat(R) Brn5 (Andersen et al., 1993), mouse C2 and C7 (Okamoto et al., 1993), human (H) TCFβ1 (Messier et al., 1993), mouse EmbFL and EmbN, described in this paper. The translation start site methionine (ATG) is indicated, as well as the predicted and experimental (when known) molecular weight for each isoform on the right. Conserved protein domains are boxed as indicated. (B) Genomic organisation of the *Emb* gene based on the mouse Emb/C1, C2 and C7 sequences (Okamoto et al., 1993), the rat Brn5 sequence (Andersen et al., 1993), the human TCFβ1 sequence (Messier et al., 1993) and the mouse data (EmbFL; EmbN) reported in this paper. The non-coding sequences are indicated as white boxes, coding sequences as black, and the POU domain as grey. The first methionine (ATG) and the stop codon (STOP) are indicated. The number of 5′ exons (shown as 1n) is unknown since the mouse Emb genomic sequence is not currently available and the 5′ UTRs are very divergent between mouse and human. Schema not to scale.

As a result of multiple transcription start sites and differential splicing, several other transcripts are generated from the gene including the previously described Emb and a shorter –COOH version (C7), also including the POU and homeodomains (Okamoto et al., 1993). These sequences will bind to DNA, and may act as negative competitors with EmbFL. The NH_2 terminal sequence (EmbN) which lacks these DNA binding motifs may also regulate EmbFL function. The possibility of further transcripts, possibly generated from more 5′ genomic sequence, arises because of the presence of larger (>80 kDa) forms of EmbFL in extracts from liver and heart. However, it is also possible that these reflect secondary modifications which affect EmbFL migration (e.g. phosphorylation). The 80 kDa protein already migrates with a higher molecular weight than the 60 kDa predicted from the coding sequence (see Fig. 6A). This is frequently observed in acidic proteins, and may also suggest that the protein is modified. Northern blots had revealed a

number of mRNAs hybridizing to the Emb/Brn5/mPOU sequence, with a 6 kb sequence predominating in most murine tissues, including brain and striated muscle (Okamoto et al., 1993). The open reading frame of EmbFL requires 1.734 kb of coding sequence and this therefore implies extensive 3′ or 5′ non-coding regions if the 6 kbp corresponds to the processed mRNA; one of the mouse cDNA clones (C3) isolated by Okamoto et al. (1993), had a 3′ non-coding sequence of >4 kb and the longer 5′ RACE product reported here, has 1.8 kb of 5′ UTR.

The genes of other POU family members also give rise to more than one coding sequence. For example, differential splicing leads to Oct 2 isoforms with different –COOH termini (Wirth et al., 1991), with and without a transcriptional activation domain (Lillycrop and Latchman, 1992).

In the cell and tissue systems which we have examined, the longer form of the Emb protein (EmbFL) predominates. This is confirmed in the EST databases, where EmbFL is clearly an abundant transcript, whereas EmbN and the other Emb isoforms previously described are not found as ESTs. The truncated forms, Emb and EmbN, could potentially have a dominant negative effect. In the overexpression experiments described here, no striking change was observed in the transfected cells. This may reflect the inadequacies of this artificial approach to the biological function of POU proteins (see Wu et al., 2001). The possible negative effect of the truncated EmbN sequence on muscle cell survival is intriguing in the context of the effects of the longer ($-NH_2$ terminal) isoform of the class IV POU domain protein, Brn3a, on the activation of the cell survival gene *BclII*. Ectopic expression of Brn5, the truncated –COOH form of EmbFL, has been shown to inhibit neuronal cell proliferation (Cui and Bulleit, 1997). This shorter form of Emb has also been variously described as a weak transcriptional activator (Andersen et al., 1993; Messier et al., 1993) or as exerting some degree of transcriptional repression (Wey et al., 1994), possibly through competition for DNA binding with other POU domain proteins or indeed with EmbFL. The in vivo significance of these observations is not clear since the EST data suggest that the shorter forms of Emb are rare transcripts. However, a shorter form may accumulate, as well as EmbFL, in the cerebellum (Fig. 1B). The fact that this form of Emb, when transfected into cells, appears to be concentrated in the nucleus may reflect a heightened efficiency for DNA binding to POU protein targets, compared to EmbFL or EmbN which, when overexpressed, are not localised exclusively in the nucleus. Overexpression may also lead to disregulated cellular distribution following interaction of EmbFL with specific partners.

We have shown (Molinari et al., 2004) that EmbFL binds to a regulatory sequence in the *cardiac actin* enhancer, in the form of a complex with the myogenic factor Mef2D and the histone transacetylase p300. Mutations, which interfere with Emb binding, reduce enhancer activity in differentiating muscle cells and the enhancer is necessary for expression of the *cardiac actin* gene in skeletal muscle in vivo (Biben et al., 1996). Furthermore, the sequence to which the Emb complex binds may be involved in controlling transcriptional accessibility of the locus. In the context of a transcriptional role for Emb, the homeo and POU domains are clearly of importance in DNA binding and, potentially, in providing an interface for interaction with other transcription factors or co-factors (see Phillips and Luisi, 2000). The conserved regions in the $-NH_2$ terminal domain of EmbFL are also of potential interest in this context, notably the acidic sequence which is similar to that in bromodomain containing proteins such as Brg1 (Khavari et al., 1993), involved in chromatin remodelling.

Acknowledgements

This work was supported by the Pasteur Institute and the C.N.R.S. (URA 2578), with grants from the A.F.M., the M.R.T. (ACI programme) and the E.C. 5th framework programme (Bio4-CT95-0228). S. Molinari had an E.C. Training and Mobility Research fellowship (ERBFMBICT95-0240) and an A.F.M. fellowship. F. Relaix is supported by Inserm.

References

Andersen, B., Schonemann, M.D., Pearse II, R.V., Jenne, K., Sugarman, J., Rosenfeld, M.G., 1993. Brn-5 is a divergent POU domain factor highly expressed in layer IV of the neocortex. J. Biol. Chem. 268, 23390–23398.

Angrand, P.O., Kallenbach, S., Weiss, M.C., Rousset, J.P., 1990. An exogenous Albumin promoter can become silent in dedifferentiated hepatoma variants as well as intertypic hybrids. Cell Growth Differ. 1, 519–526.

Biben, C., Kirschbaum, B.J., Garner, L., Buckingham, M., 1994. Novel muscle-specific enhancer sequences upstream of the cardiac actin gene. Mol. Cell. Biol. 14, 3504–3513.

Biben, C., Hadchouel, J., Tajbakhsh, S., Buckingham, M., 1996. Developmental and tissue-specific regulation of the murine cardiac actin gene in vivo depends on distinct skeletal and cardiac muscle-specific enhancer elements in addition to the proximal promoter. Dev. Biol. 173, 200–212.

Bulleit, R.F., Cui, H., Wang, J., Liu, X., 1994. NMDA receptor activation in differentiating cerebellar cell cultures regulates the expression of a new POU gene, Cns-1. J. Neurosci. 14, 1584–1595.

Cui, H., Bulleit, R.F., 1997. Expression of the POU transcription factor Brn-5 inhibits proliferation of NG18-15 cells. B.B.R.C. 236, 693–696.

Daubas, P., Tajbakhsh, S., Hadchouel, J., Primig, M., Buckingham, M., 2000. Myf5 is a novel early axonal marker in the mouse brain and is subjected to post-transcriptional regulation in neurons. Development 27, 319–331.

Dignam, J.D., Lebovitz, R.M., Roeder, R.G., 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. 11, 1475–1489.

Dominov, J.A., Miller, J.B., 1996. POU homeodomain genes and myogenesis. Dev. Genet. 19, 108–118.

Ferrari, S., Battini, R., Molinari, S., 1994. Specific binding to vitamin D response elements of chicken intestinal DNA-binding activity is not related to the vitamin D receptor. Mol. Endocrinol. 8, 173–181.

Graham, F.L., Smiley, J., Russell, W.C., Nairn, R., 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J. Gen. Virol. 36, 59–72.

Hinkley, C., Perry, M., 1992. Histone H2B gene-transcription during Xenopus early development requires functional co-operation between proteins bound to the CCAAT and octamer motifs. Mol. Cell. Biol. 12, 4400–4411.

Johansen, T., Moens, U., Holm, T., Fjose, A., Krauss, S., 1993. Zebrafish pou(c): a divergent POU family gene ubiquitously expressed during embryogenesis. Nucleic Acids Res. 21, 475–483.

Khavari, P.A., Peterson, C.L., Tamkun, J.W., Mendel, D.B., Crabtree, G.R., 1993. BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. Nature 366, 170–174.

Lillycrop, K.A., Latchman, D.S., 1992. Alternative splicing of the Oct-2 transcription factor RNA is differentially regulated in neuronal cells and B cells and results in protein isoforms with opposite effects on the activity of octamer/TAATGARAT-containing promoters. J. Biol. Chem. 267, 24960–24965.

McEvilly, R.J., Rosenfeld, M.G., 2000. The role of POU domain proteins in the regulation of mammalian, pituitary and nervous system development. Prog. Nucleic Acid Res. Mol. Biol. 63, 223–255.

Messier, H., Brickner, H., Gaikwad, J., Fotedar, A., 1993. A novel POU domain protein which binds to the T-cell receptor β enhancer. Mol. Cell. Biol. 13, 5450–5460.

Molinari, S., Relaix, F., Lemonnier, M., Kirschbaum, B., Schäfer, B., Buckingham, M., 2004. A novel complex regulates cardiac actin gene expression through interaction of Emb, a class VI POU domain/bromodomain protein, MEF2D and the histone transacetylase, p300. Mol. Cell. Biol. 24, 2944–2957.

Okamoto, K., Wakamiya, M., Noji, S., Koyama, E., Taniguchi, S., Takemura, R., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Muramatsu, M., Hamada, H., 1993. A novel class of murine POU gene predominantly expressed in central nervous system. J. Biol. Chem. 268, 7449–7457.

Phillips, K., Luisi, B., 2000. The virtuoso of versatility: POU proteins that flex to fit. J. Mol. Biol. 302, 1023–1039.

Reznikoff, C.A., Brankow, D.W., Heidelberger, C., 1973. Establishment and characterization of a cloned line of C3H mouse embryo cells sensitive to postconfluence inhibition of division. Cancer Res. 33, 3231–3238.

Ryan, A.K., Rosenfeld, M.G., 1997. POU domain family values: flexibility, partnership and developmental codes. Genes Dev. 11, 1207–1225.

Schubert, D., Heinemann, S., Carlisle, W., Tarikas, H., Kimes, B., Patrick, J., Steinbach, J.H., Culp, W., Brandt, B.L., 1974. Characterization of a unique muscle cell line. J. Cell Biol. 61, 398–413.

Wey, E., Lyons, G.E., Schäfer, B.W., 1994. A human POU domain gene, mPOU, is expressed in developing brain and specific adult tissues. Eur. J. Biochem. 220, 753–762.

Wirth, T., Priess, A., Annweiler, A., Zwilling, S., Oeler, B., 1991. Multiple Oct-2 isoforms are generated by alternative splicing. Nucleic Acids Res. 19, 43–51.

Wu, R., Jurek, M., Sundarababu, S., Weinstein, D.E., 2001. The POU gene Brn-5 is induced by neuregulin and is restricted to myelinating Schwann cells. Mol. Cell. Neurosci. 17, 683–695.

Yaffé, D., Saxel, O., 1977. Serial passaging and differentiation of myogenic cells isolated from dystrophic muscle. Nature 270, 725–777.

Zhou, H., Yoshioka, T., Nathans, J., 1996. Retina-derived POU-domain factor-1: A complex POU-domain gene implicated in the development of retinal ganglion and amacrine cells. J. Neurosci. 16, 2261–2274.