ELSEVIER

Biochimica et Biophysica Acta 1353 (1997) 224–230

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Short sequence-paper

Molecular cloning of a novel alternatively spliced nuclear protein [1]

Annette M. Hartmann, Stefan Stamm *

Max-Planck Institute for Psychiatry, Am Klopferspitz 18a, 82152 Planegg Martinsried, Germany

Received 2 April 1997; revised 25 April 1997; accepted 5 May 1997

Abstract

Using the yeast two hybrid system, we isolated a rat cDNA (E3-3) coding for a new protein with no homology to any other protein in the database. E3-3 is ubiquitously expressed. Variants that most likely arise through alternative splicing encode truncated forms of the protein. Testis is the only tissue that predominantly expresses the longest protein variant. When this variant is tagged with enhanced green fluorescent protein, the protein is located in the nucleus. © 1997 Elsevier Science B.V.

Keywords: Cloning; Nuclear protein; Splicing; Yeast two hybrid system; cDNA, rat

The yeast two hybrid system [1] is a new method to clone proteins by virtue of their interaction with other proteins. With known proteins involved in pre-mRNA splicing as baits, we are using this system to find new components of the spliceosome. With SRp30c [2] as bait, we found several clones, among them rat acidic ribosomal phosphoprotein P1 (genbank accession number X15097), the rat homologue to *Bos taurus* mRNA for polyA binding protein II (Accession numbers U94858 and X89969) and the rat homologue to human SAF-b [3] (accession number L43631). Here we report another clone, E3-3, that was isolated during this screen.

Cloning and two hybrid screening were essentially performed as described [4]. Briefly, EcoRI and BamHI restriction sites were introduced into SRp30c cDNA using primers SRp30cf and SRp30cr. The resulting PCR fragment was subcloned into pGBT9 (Clontech). Using this clone as bait, approximately 150 000 colonies of an embryonic day E16 rat brain library in pADGal4 (Stratagene) were screened. From this screen 47 clones were analysed further.

The sequence of E3-3 is shown in Fig. 1. The clone isolated with the two hybrid screen contains an open reading frame of 558 nt (ORF1). It starts with an ATG that is within the consensus Kozak sequence [5]. ORF1 codes for a protein of a predicted size of 20683 dalton. The protein is slightly basic, with an isoelectric point of 7.83. The ORF is followed by a 121 nt long 5′ UTR that is terminated with a poly A signal followed by the poly A stretch. Database searches (FASTA, [6]) showed homology to rat, mouse and human ESTs coding for the same protein, but no strong homology to other known proteins. Furthermore, searches with PROSITE did not reveal any known protein patterns.

In order to determine the tissue distribution of E3-3, we performed RT-PCR from various rat tis-

Abbreviations: ORF, open reading frame; EGFP, enhanced green fluorescent protein; EST, expressed sequence tag; GAPDH, glyceraldehyde-3-phosphate dehydrogenase

* Corresponding author. Fax: +49 89 8578 3749; E-mail: stamm@assel.biochem.mpg.de

[1] GenBank accession numbers: U94858, U95160, U95161, U95162

0167-4781/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved.
PII S0167-4781(97)00090-0

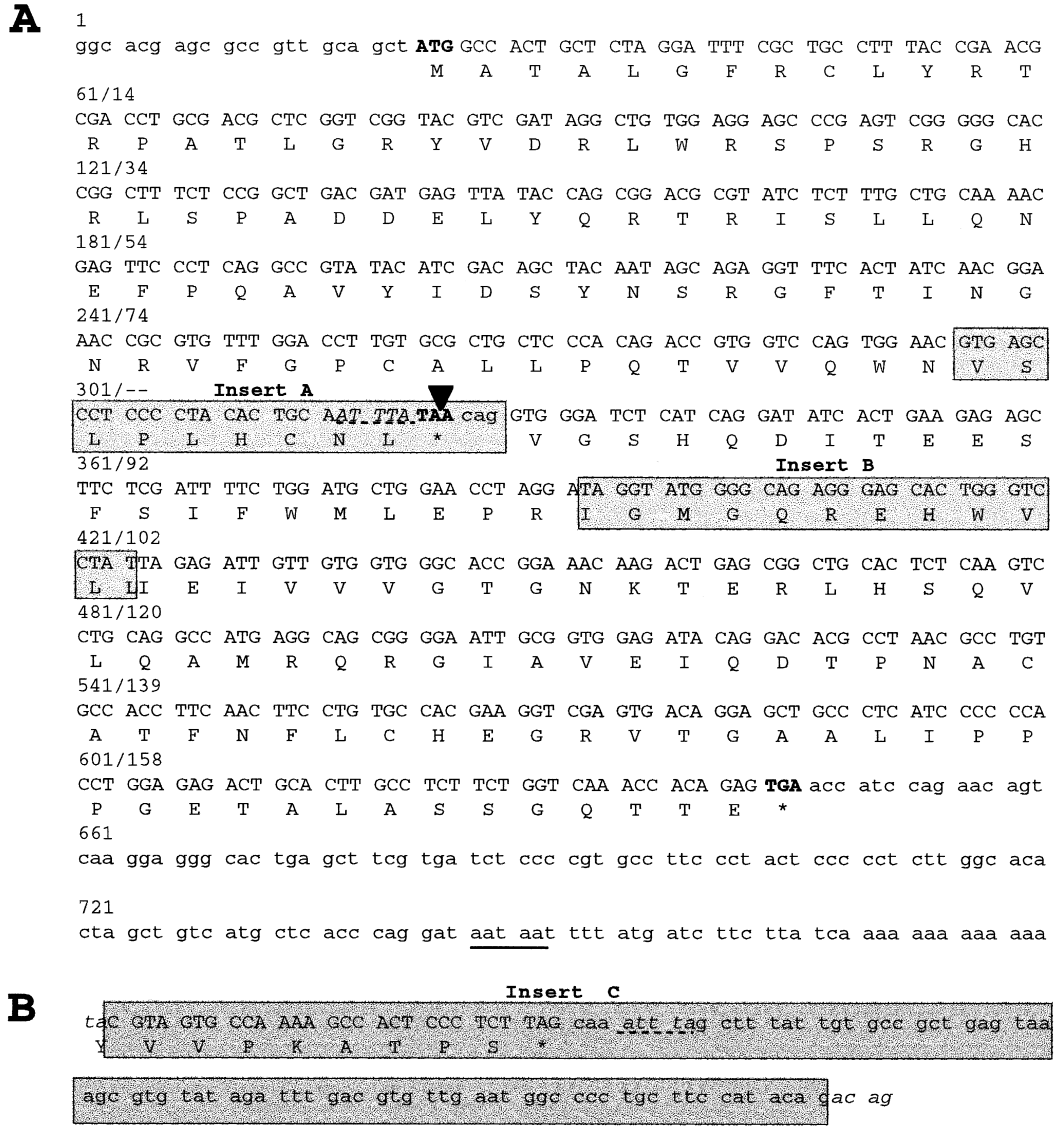
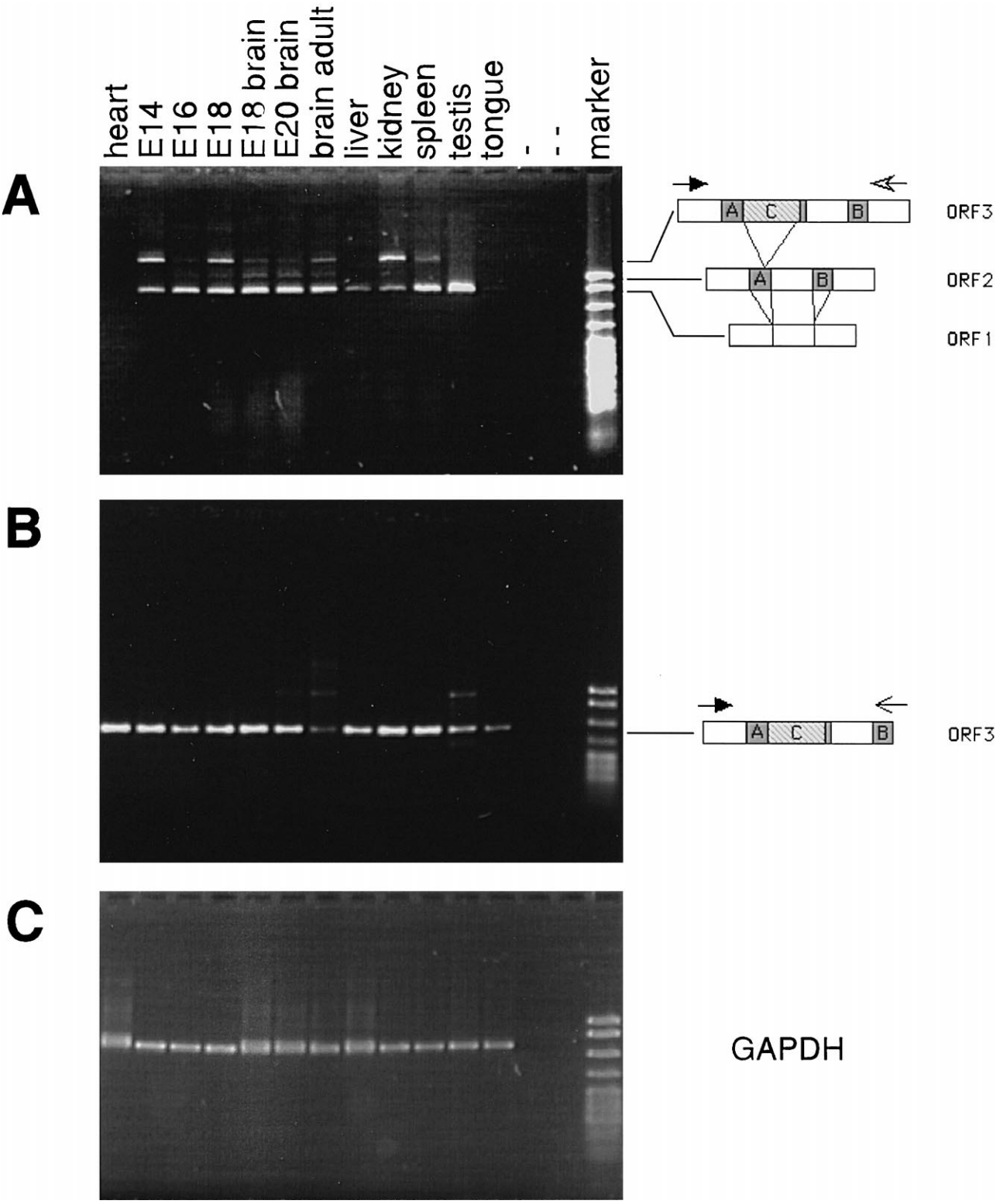


Fig. 1. Sequence of E3-3. (A) Sequence of E3-3 ORF1 and ORF2. ORF1 refers to sequence without, ORF2 with the shaded boxes. Start codon and stop codons are shown in bold, the polyadenylation signal is underlined. The boxed sequence represents the alternatively spliced inserts in ORF2. The triangle (▼) in insert A indicates the insertion point of insert C. The deduced amino acid sequence is shown below the coding region and numbered 1 to 171. Amino acids encoded by the alternative inserts (gray shading) are omitted from this numbering, because of the stop codon introduced by the first alternative exon. The poly A addition signal is underlined. RNA destabilisation motifs are shown in *italics* and are marked with dashes. Coding nucleotides are indicated with upper case. (B) Sequence of insert C (boxed and gray). The flanking nucleotides shown in Figure A are indicated in *italics*, RNA destabilisation motifs are shown in *italics* and are marked with dashes.

sues. RT-PCR was performed using 1.5 μg of total rat RNA, 2 mM DTT, 0.2 mM dNTP, 1 × Superscript RT-buffer, 3 U of RNAse Inhibitor, 60 U of H^- Superscript MMLV (GIBCO) and 40 pmol of E3-3-r1 antisense primer. For the PCR, 28 cycles (94°C, 1′; 60 → 50°C 1′; 72°C, 2′) were used. The result of the RT-PCR using the primer E3-3-f2 and E3-3-r1 is shown in Fig. 2A. E3-3 expression was observed in all tissues. Muscle tissue like heart and tongue showed only weak expression, that was clearly detectable when more cycles were used in the PCR (data not shown). As shown in Fig. 2A, we observed three


heart
E14
E16
E18
E18 brain
E20 brain
brain adult
liver
kidney
spleen
testis
tongue
-
--
marker
A
A
C
B
ORF3
A
B
ORF2
ORF1
B
A
C
B
ORF3
C
GAPDH

PCR fragments. One had the expected size of 551 nt and the others were 617 and 723 nt long. To determine their difference, these PCR products were subcloned into pCRII (Invitrogen). The 551 nt fragment was the expected PCR amplicon derived from ORF1. The 617 nt fragment contained two inserts, named A and B (Fig. 1A, Fig. 2A). Insert A contains an in frame stop codon. Therefore, usage of insert A results in an ORF2 that encodes a slightly acidic protein of 14.5 kD with an isoelectric point of 6.04. The 723 nt amplicon had another insertion, termed C (Fig. 1B, Fig. 2A). Interestingly, insert C is placed into the TAA stop codon of insert A. However, insert C contains an in frame stop codon after 30 nucleotides, again resulting in a truncated protein (ORF3) of 12 213 dalton ($pI = 7.04$).

Skipping of insert A and usage of insert B would result in an open reading frame containing eleven additional amino acids from insert B. We therefore determined whether there is an additional splice variant containing only insert B. We performed RT-PCR using an insert B specific antisense primer (E3-3r5) and the upstream primer E3-3f2 with RNA from a variety of tissues. We obtained a 380 nt long PCR fragment corresponding to ORF3 in all tissues. However, we were not able to detect a PCR fragment of 234 nt in length that would correspond to E3-3 without insert A and C.

Using the flanking primers E3-3f2 and E3-3r1, PCR products with inserts were detectable in almost all tissues, with liver and testis being the exceptions. As the expression of ORF1 in testis is higher than in liver, we conclude that the protein product of ORF1 is most abundant in testis.

To determine the quality of the RNA, we performed RT-PCR using GAPDH primer. The presence of GAPDH amplicons in all tissues demonstrated the integrity of the RNA (Fig. 2C).

The variants of E3-3 are most likely created by alternative splicing, a common mechanism to create protein isoforms [7,8]. One of the biological functions of alternatively spliced exons is to create truncated proteins by introducing frameshifts or stop codons [9,10]. In addition to introducing stop codons, both exon A and C add AUUUA motifs to the RNA which are known to be responsible for mRNA destabilisation [11]. Therefore, inserts A and B will most likely destabilize the RNA and truncate the protein encoded by ORF1.

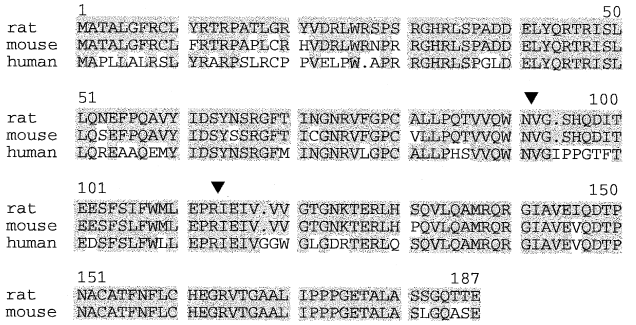


Fig. 3. Sequence comparison of rat, mouse and human E3-3. The rat sequence was determined in this paper. The human and mouse sequences are the translational products from the ESTs given in the text. Triangles indicate the insertion points of insert A and B. Amino acids that are identical to the rat sequence are gray. Only 150 amino acids of the human sequence are shown, because then the sequence of the ESTs is unreliable.

Several ESTs in the database code for proteins similar to E3-3. Using mouse (genbank accession numbers AA000183, AA108485, AA153006, W34172, AA139298, AA068840) and human ESTs (genbank accession numbers R84301, H42278, T55917, T80723, D81472, D81230, N33784, H16144 and R17947) we were able to compare rat, mouse and human sequences. This comparison is shown in Fig. 3. Only ESTs that are similar to ORF1 are present in the database. With one exception, all ESTs apparently used the same start codon as E3-3 or their sequence started further downstream, indicating that the clone we isolated is full length. An EST (accession number AA023151) from mouse placenta had an ATG that started 24 amino acids in frame upstream from the ATG in the E3-3 clone. As this sequence diverges from the other mouse ESTs, it might also be

Fig. 2. Tissue distribution of E3-3 and its splice variants. (A) RT-PCR using E3-3f2 and E3-3r1 primer to amplify cDNA from the tissues indicated. Cloning and sequencing of the three PCR products revealed the structure of the PCR products that is schematically shown on the right side. ‘‘−’’ indicates absence of reverse transcriptase, ‘‘− −’’ indicates absence of template. (B) RT-PCR using E3-3f2 and an insert B specific primer E3-3r5. ORF3 is the main product. The additional bands seen in testis and brain are PCR artifacts due to mispriming. (C) RT-PCR using GAPDH specific primers GAPDHf and GAPDHr to demonstrate the integrity of the RNA used.

encoded by an alternative exon. The sequence comparison demonstrated that rat, mouse and human sequences are highly related. Protein sequence identity is 71% between rat and human and 93% between rat and mouse.

In order to gain insight into the function of the protein, we determined the intracellular localisation of the protein encoded by ORF1. We cloned ORF1 into pEGFP-C2 (Clonetech) and transfected A293 cells with this construct (pEGFPC2-E3-3). Confocal

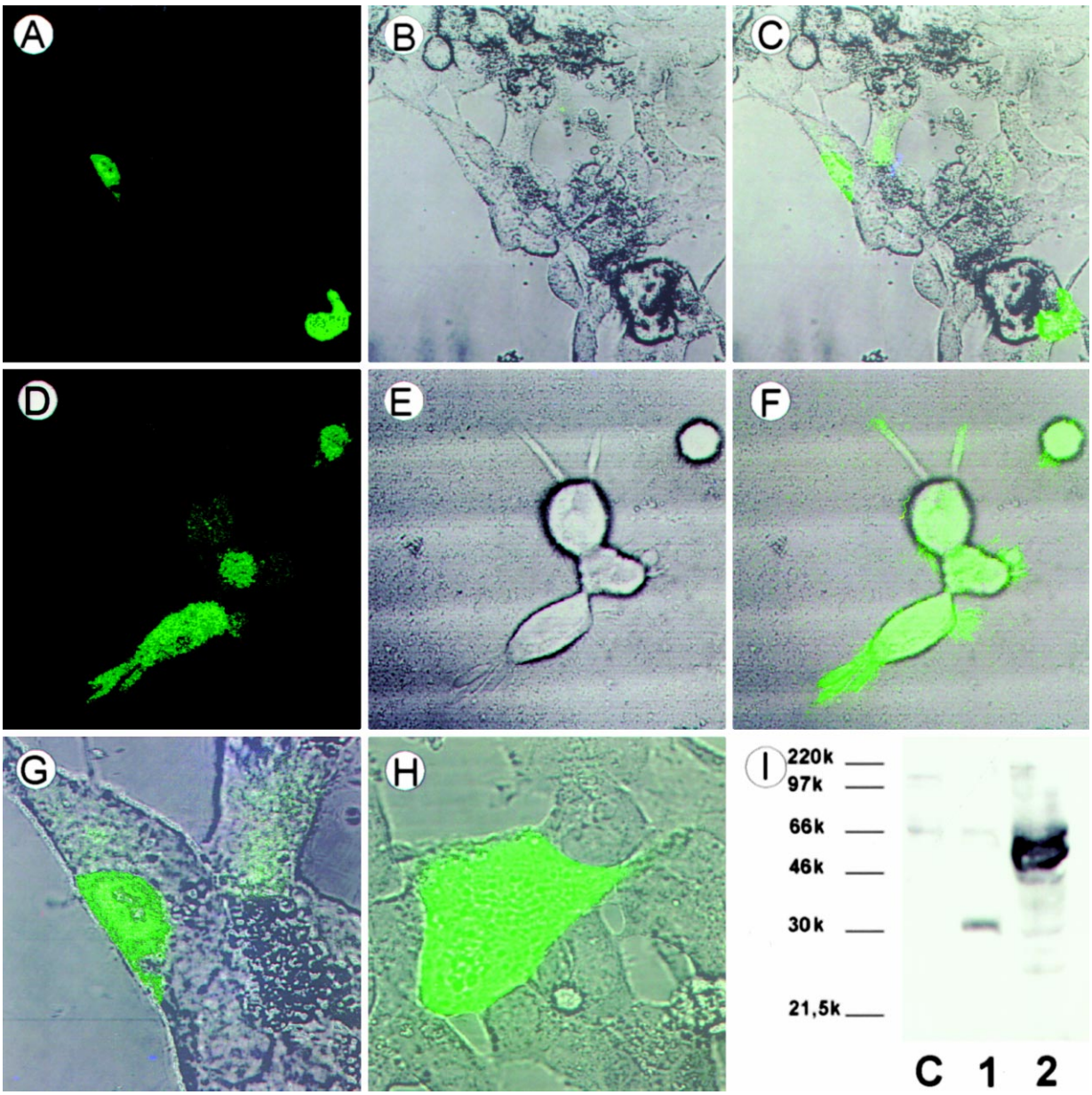


Fig. 4. Intracellular localisation of E3-3. A293 cells were transfected with EGFP tagged E3-3 and analysed by confocal microscopy (Leica). Magnification for panels A–F is 63 ×. (A) EGFP-E3-3 staining of a group of cells. Staining is predominant in the nucleus. (B) Same field in phase contrast. (C) Overlay of (A) and (B). (D) EGFP staining of a group of cells. Staining is uniformly distributed. (E) Phase contrast of the same group of cells. (F) Overlay of (D) and (E). (G) Higher magnification of the cell in panel (C). (H) Higher magnification of an EGFP stained cell. (I) Western blot analysis of mock transfected cells (C), pEGFP-C2 transfected cells (1) and pEGFPC2-E3-3 (2) transfected cells. An anti-GFP antibody was used for detection.

microscopy analysis demonstrated that E3-3 tagged with EGFP is localized uniformly in the nucleus. We did not see a fluorescence pattern that revealed any subnuclear structures. In contrast, EGFP alone is evenly distributed throughout the whole cell (Fig. 4). The integrity of the transfected protein was further analysed by Western blot, using an EGFP specific antibody (Clontech). Using this antibody, we observed a band of 46.6 kD in pEGFPC2-E3-3 transfected cells and a band of 26 kD in pEGFP-C2 transfected cells, which corresponded to the expected sizes. These data suggest that E3-3 is a nuclear protein with a uniform distribution in the nucleus. As we cloned E3-3 by its interaction with SRp30c, we tested it against other splicing components in the yeast two hybrid system. None of the proteins tested [4] (SRp20, SF2/ASF, SC35, SRp40, SRp55, SRp75, U2AF65 and SmN) interacted with E3-3. In addition, retesting of E3-3 with SRp30c resulted in poor growing yeast. We therefore tested E3-3 for self-activation. Transformation of E3-3 into HF7c and plating on double dropout plates lacking leucine and histidine revealed that the protein did not selfactivate. We therefore think that E3-3 might interact with SRp30c, but due to the weakness of this interaction and the lack of binding to other splicing proteins tested, we cannot rule out that this interaction is a two-hybrid screening artifact.

The function of E3-3 remains unclear, mainly because the clone does not contain any known structural motif and bears no homology to previously described proteins. From our analysis we can conclude that the E3-3 gene is expressed in all tissues. The gene product is alternatively spliced, creating a 14.5 kD acidic ($pI = 6.04$), a 12.2 kD neutral ($pI = 7.04$) and a 20.6 kD basic ($pI = 7.8$) nuclear protein. The alternative spliced exons creating the 12.2 kD and 14.5 kD isoforms introduce destabilisation signals into the RNA. Testis is the exception of these splicing events, as this tissue predominantly skips the alternative exons probably resulting in the synthesis of the 20.6 kD isoform as the main protein. This indicates that E3-3 might play a role in cell nuclei in testis.

List of primers used

E3-3f1	CCGGCTGACGATGAGTTATACC
E3-3f2	CCAGCGGACGCGTATCTCTTT
E3-3f3	GATAGAGATTGTTGTGGTGG
E3-3r1	ATAAGAAGATCATAAATTATTA
E3-3r2	TGGCACAGGAAGTTGAAGGTG
E3-3r3	CCAAACACGCGGTTTCCGTTG
E3-3r5	CCAGTGCTCCCTCTGCCCCA
E3/3-HAf1	AAAGCTTACCATGTACCCATAC-GATGTTCCAGATTACGCTGAA-TTCGGCACGAGCGCCGT
E3/3-HAr1	AGACCAGTTTGGTGTCTCACTT-GGGGATCCCCTCGAGGA
GAPDHf	ACCACAGTCCATGCCATCAC
GAPDHr	TCCACCACCCTGTTGCTGTA
SRp30cf	CGCGAATTCTCGGGCTGGGCGG-ACGAGCGC
SRp30cr	GAGGGATCCTCAGTAGGGCCTG-AAAGGAGAG

We would like to thank Claudia Cap for sequencing, J. Chalcroft and R. Schorner for help with artwork. The work is supported by the Max-Planck-Society.

References

[1] C.-T. Chien, P.L. Bartel, R. Sternglanz, S. Field, The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest, Proc. Natl. Acad. Sci. USA 88 (1991) 9578–9582.

[2] G.R. Screaton, J.F. Caceres, A. Mayeda, M.V. Bell, M. Plebanski, D.G. Jackson, J.I. Bell, A.R. Krainer, Identification and characterization of three members of the human SR family of pre-mRNA splicing factors, EMBO J. 14 (1995) 4336–4349.

[3] A. Renz, F.O. Fackelmayer, Purification and molecular cloning of the scaffold attachment factor B (SAF-B), a novel human nuclear protein that specifically binds to S/MAR-DNA, Nucl. Res. Acids 24 (1996) 843–849.

[4] B. Beil, G. Screaton, S. Stamm, Molecular cloning of htra2-beta-1 and htra2-beta-2, two human homologues of tra-2 generated by alternative splicing, DNA 16 (1997) 679–690.

[5] M. Kozak, Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs, Nucl. Acids Res. 12 (1984) 857–872.

[6] W.R. Pearson, D.J. Lipman, Improved tools for biological sequence analysis, Proc. Natl. Acad. Sci. USA 85 (1988) 2444–2448.

[7] A. Krämer, The structure and function of proteins involved in mammalian pre-mRNA splicing, Annu. Rev. Biochem. 65 (1996) 367–409.

[8] M. McKeown, Alternative mRNA splicing, Annu. Rev. Cell Biol. 8 (1992) 133–155.

[9] P.M. Bingham, T.-B. Chou, I. Mims, Z. Zachar, On/off regulation of gene expression at the level of splicing, Trends Genet. 4 (1988) 134–138.
[10] S. Stamm, M.Q. Zhang, T.G. Marr, D.M. Helfman, A sequence compilation and comparison of exons that are alternatively spliced in neurons, Nucl. Acids Res. 22 (1994) 1515–1526.
[11] J.E.G. McCarthy, H. Kollmus, Cytoplasmatic mRNA–protein interaction in eucaryotic gene expression, Trends Biochem. Sci. 20 (1995) 191–197.