Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 85 (2005) 493–502

GENOMICS

www.elsevier.com/locate/ygeno

Cloning of the murine ER71 gene (*Etsrp71*) and initial characterization of its promoter☆

Luciano De Haro, Ralf Janknecht*

Department of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905, USA

Received 1 September 2004; accepted 16 December 2004
Available online 29 January 2005

Abstract

The ER71 protein belongs to the ETS transcription factor family and is testis-specifically expressed in adult mice. Here we describe the cloning of the respective *Etsrp71* gene and promoter. The murine *Etsrp71* gene is relatively compact, spanning 3 kb, and is arranged into seven exons and six introns, the majority of which are highly conserved in rat and human. Its promoter is devoid of a TATA box and transcription starts at multiple sites. Furthermore, two ER71 isoforms exist that differ by 22 N-terminal amino acids, but show no difference in DNA binding or transactivation. Close to the transcription initiation sites, we identified a binding site for the transcription factor Sp1. Mutation of this binding site severely diminished the ability of Sp1 to activate the *Etsrp71* promoter. The findings reported here may provide avenues for further research elucidating the regulation of *Etsrp71* gene activity during embryogenesis and in adult testes.

Keywords: Alternative splicing; ER71; ETS protein; Etsrp71; ETV2; Promoter; Sp1; Transcription

The ETS family of transcription factors is characterized by a highly conserved region of ~85 amino acids, the ETS domain, that mediates DNA binding to target sequences containing a 5′-GGA(A/T)-3′ core motif [1–3]. Over 20 different *ETS* genes have been found in mammals, and their gene products often share extensive homology outside the ETS domain. However, one ETS protein that shares no further homology with other ETS proteins aside from its DNA binding domain is ER71 (ETS-related 71).

Originally, the respective murine *Etsrp71* (Ets-related protein 71) cDNA (GenBank Accession No. NM_007959) was cloned from an 8.5-day mouse embryo cDNA library with degenerate oligonucleotides corresponding to two highly conserved regions within the ETS domain [4]. Further analysis revealed that—in addition to 8.5- and 10.5-day embryos—*Etsrp71* mRNA was expressed in adult mice solely in the testes. The size of the testis *Etsrp71* transcript is 1.7 kb and thus exceeds the reported size of the *Etsrp71* cDNA of 1046 bp, indicating the need to unravel missing cDNA sequences.

Characterization of the murine ER71 protein demonstrated that it interacts with DNA sequences of the consensus 5′-(G/C)(G/C)(C/A)**GGA***(***A/T***)*(G/A)(T/C)C-3′ [4]. Specifically, ER71 was shown to bind to the *MMP1* (matrix metalloproteinase-1) gene promoter and regulate its transcriptional activity [5]. As matrix metalloproteinases are important for the extensive tissue remodeling during spermatogenesis [6–9], *Etsrp71* expression in the adult testis may contribute to this process. Conceivably, the extensive tissue remodeling in embryos, in particular in developing testes [10,11], may also be supported through ER71-dependent activation of *MMP1* transcription.

Analysis of the murine ER71 molecule has revealed that it is a constitutively nuclear protein, most likely due to a bipartite nuclear localization signal within its ETS domain. Furthermore, ER71 DNA binding is strongly enhanced by its 21 C-terminal amino acids, whereas N-terminal amino acids slightly suppress DNA binding. In addition, the

☆ Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. AY805222.

* Corresponding author. Fax: +1 507 284 1767.
E-mail address: janknecht.ralf@mayo.edu (R. Janknecht).

0888-7543/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2004.12.003

N-terminus of ER71 contains an extremely potent transactivation domain, which can stimulate transcription >80-fold when fused to a heterologous DNA binding domain [5]. As such, ER71 is a typical modular transcription factor whose DNA binding and transactivation activities might be regulated in vivo through posttranslational modifications and interactions with other transcription factors.

In this report, we have determined the structure of the murine *Etsrp71* gene, indicating that two isoforms of the ER71 protein can exist. In addition, we characterized the TATA-box-devoid *Etsrp71* promoter and provide evidence that it is regulated by the transcription factor Sp1.

Results and discussion

Cloning of the murine Etsrp71 gene and its 5′- and 3′-flanking sequences

To clone genomic DNA corresponding to the published murine *Etsrp71* cDNA, we screened a murine embryonic stem cell genomic phage library with a cDNA fragment corresponding to the first 292 amino acids of the reported [4] ER71 protein sequence. Positive genomic clones were further checked for the presence of DNA sequences encoding both the N- and the C-terminus of the ER71 protein by Southern blotting and restriction analyses. Finally, we sequenced 3463 bp of one genomic clone (Fig. 1) that encompassed all *Etsrp71* cDNAs known previously.

Determination of transcription start and polyadenylation sites

The previously reported *Etsrp71* cDNA had a poly(A) tail starting at the corresponding position +2938 of our annotated genomic sequence (see Fig. 1). This is preceded by an AATAAA hexanucleotide sequence (+2916 to +2921) as would be expected for a polyadenylation site. To assess if this is the only polyadenylation site, we performed 3′-RACE (rapid amplification of cDNA ends) on mouse testis mRNA. Our analyses showed that only 1 3′-RACE product was visible on an agarose gel (Fig. 2A). Indeed, we could confirm exactly the same polyadenylation site as reported before in 9 of 11 3′-RACE products sequenced. However, we found an additional polyadenylation site 10 bp downstream of the previous one at position +2948 in the remaining 2 3′-RACE products sequenced.

Similarly, we performed 5′-RACE to determine the transcription initiation site(s). Surprisingly, we obtained 2 major discernable 5′-RACE products that differed in size by ~0.2 kb (Fig. 2B); please note that the smaller 5′-RACE product was much less abundant than the larger one. Sequence analysis of 16 5′-RACE products revealed that various transcription start sites were utilized (Fig. 2B), and we assigned the one farthest upstream of the *Etsrp71* coding region the coordinate +1, which extends the previously reported *Etsrp71* cDNA by 442 bp. In agreement with transcription starting over a range of 90 bp, no TATA box is present that could direct transcription to initiate at a specific site ~30 bp downstream of its location. In addition, as a reflection of the presence of the minor smaller 5′-RACE product (Fig. 2B), sequence analysis of the 5′-RACE products also revealed the existence of infrequent alternative splicing: 219 bp (position +152 to +370, underlined in Fig. 1) within exon 1 can be excluded from the *Etsrp71* mRNA.

Northern blotting demonstrated that the *Etsrp71* transcript in testis has a size of ~1.7 kb ([4] and Fig. 2C). Joining all exons gives a maximal *Etsrp71* cDNA length of 1503 bp, which matches very well the 1.7-kb mRNA size considering that the average size of a poly(A) tail is 0.1–0.2 kb. Consistent with a rare occurrence of alternative splicing within exon 1 in vivo, a longer exposure of the Northern blot revealed an additional minor *Etsrp71* transcript of ~1.5 kb in testis (Fig. 2C, right). Furthermore, our Northern blot analyses confirm the testis-specific expression of *Etsrp71* in adult mice, as no expression was observable in brain, heart, kidney, liver, lung, muscle, skin, small intestine, spleen, stomach, and thymus (Fig. 2C).

To study further the expression of *Etsrp71* in testis, we performed RT-PCR on a murine Sertoli (15P-1) and a Leydig (MLTC-1) tumor cell line. As a control, we utilized the mouse mammary epithelial cell line HC11, which, in contrast to normal or cancerous breast, expresses high levels of *Etsrp71* [12]. As shown in Fig. 2D, 15P-1 cells also express high levels of *Etsrp71,* whereas MLTC-1 cells show a much reduced level of *Etsrp71* expression, especially taking into account that apparently more MLTC-1 RNA was utilized for the RT-PCR as indicated by the stronger signal for the housekeeping gene *Gapd* (glyceraldehyde-3-phosphate dehydrogenase). Thus, it is likely that *Etsrp71* is highly expressed in Sertoli but not Leydig cells of the testis, but further research is needed to validate this.

Two ER71 isoforms are encoded by Etsrp71

Analysis of the derived *Etsrp71* cDNA revealed that the previously reported 336-amino-acid-long ER71 protein may have an N-terminal extension of 22 further amino acids (see Fig. 3A for a sketch of ER71), mostly encoded by the previously unrecognized exon 1. The ATG encoding methionine 1, giving rise to a 358-amino-acid-long ER71 protein, is preceded 108 nt upstream by an in-frame stop codon (see shaded TGA in Fig. 1), or if the rare alternative splicing occurs, 144 nt upstream (see shaded TAA in Fig. 1). Neither methionine 1 (ATGAA**ATG**G) nor methionine 23 (CATC-C**ATG**G) is encoded within a perfect Kozak consensus translation start site, (C/T)(C/T)(G/A)(C/T)(C/T)**ATG**G, suggesting that translation might start at both of these methionines. If so, two ER71 isoforms, $ER71_{1–358}$ and $ER71_{23–358}$, whose calculated molecular masses of 39.6 and 37.2 kDa, respectively, differ by 2.4 kDa, should exist.