Identification of a core promoter and a novel isoform of the human TSC1 gene transcript and structural comparison with mouse homolog

Mahmood Ali\textsuperscript{a}, Satish C. Girimaji\textsuperscript{b}, Arun Kumar\textsuperscript{a,*}

\textsuperscript{a}Department of Molecular Reproduction, Development and Genetics, Indian Institute of Science, Bangalore 560 012, India
\textsuperscript{b}Department of Psychiatry, National Institute of Mental Health and Neurosciences, Bangalore, India

Abstract

Tuberous sclerosis complex (TSC) is an autosomal dominant disorder with loci on chromosome 9q34.12 (TSC1) and chromosome 16p13.3 (TSC2). Genes for both loci have been isolated and characterized. The promoters of both genes have not been characterized so far and little is known about the regulation of these genes. This study reports the characterization of the human TSC1 promoter region for the first time. We have identified a novel alternative isoform in the 5’ untranslated region (UTR) of the TSC1 gene transcript involving exon 1. Alternative isoforms in the 5’ UTR of the mouse Tsc1 gene transcript involving exon 1 and exon 2 have also been identified. We have identified three upstream open reading frames (uORFs) in the 5’ UTR of the TSC1/Tsc1 gene. A comparative study of the 5’ UTR of TSC1/Tsc1 gene has revealed that there is a high degree of similarity not only in the sequence but also in the splicing pattern of both human and mouse TSC1 genes. We have used PCR methodology to isolate approximately 1.6 kb genomic DNA 5’ to the TSC1 cDNA. This sequence has directed a high level of expression of luciferase activity in both HeLa and HepG2 cells. Successive 5’ and 3’ deletion analysis has suggested that a ~587 bp region, from position +77 to –510 from the transcription start site (TSS), contains the promoter activity. Interestingly, this region contains no consensus TATA box or CAAT box. However, a 521-bp fragment surrounding the TSS exhibits the characteristics of a CpG island which overlaps with the promoter region. The identification of the TSC1 promoter region will help in designing a suitable strategy to identify mutations in this region in patients who do not show any mutations in the coding regions. It will also help to study the regulation of the TSC1 gene and its role in tumorigenesis.

© 2003 Elsevier B.V. All rights reserved.

Keywords: 5’ UTR; Promoter; CpG island; RACE; uORF

1. Introduction

Tuberous sclerosis complex (TSC) is an autosomal dominant disorder, which affects several organs in the human body including the brain, heart, kidneys, eyes, skin, spleen, liver and lungs. TSC is characterized by hamartomas, which rarely progress to malignancy in the affected organs. Clinical symptoms of TSC include cortical tubers in the brain, seizures, mental retardation, ungual and periungual fibromas, angiofibromas of the face, and angiomyolipomas in the kidneys (Fryer et al., 1987). TSC displays genetic heterogeneity with two known loci: TSC1 on chromosome 9q34.12 (Fryer et al., 1987) and TSC2 on chromosome 16p13.3 (Kandt, 1992). The genes for both loci have been isolated and characterized (van Slegtenhorst et al., 1997; EC16TSC Consortium, 1993).

The TSC1 gene has 23 exons and codes for a transcript of approximately 8.6 kb. The transcript contains a 3492 bases long open reading frame that codes for a 1164 amino acid protein, hamartin (van Slegtenhorst et al., 1997). The cellular role of hamartin is not clear but it is believed to regulate rho-mediated cell adhesion through interaction with ezrin (Lamb et al., 2000). In vivo, hamartin interacts with tuberin and it may regulate or modulate tuberin activity (Van

Abbreviations: LOH, loss of heterozygosity; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; TSC, tuberous sclerosis complex; TSS, transcription start site; uORF, upstream open reading frame; URR, upstream regulatory region; UTR, untranslated region.

* Corresponding author. Tel.: +91-80-293-2998; fax: +91-80-360-0999, 80-360-0683.
E-mail address: karun@mrdg.iisc.ernet.in (A. Kumar).
the vector (Promega, USA) and 12 clones were sequenced using The amplification product was cloned in the pGEM-T region (UTR) of the gene. Binding sites, AP1 and AP2. For 5' GAAGGCGCTGTG-3' Ready first round of PCR, the adaptor ligated HeLa Marathon-polymerase chain reaction (PCR) were performed. In the gene specific primer, HR1 (5'-CTTCAGTTTCCAGCTGTG- CAACCTG-3') located in exon 3 of the TSC1 cDNA. A nested PCR was performed on the product from the first round of PCR using AP2 and the second gene specific primer, HR2 (5'-CTTCAGTTTCCAGCTGTG- CAACCTG-3') located in exon 2 of the TSC1 cDNA. The amplification product was cloned in the pGEM-T vector (Promega, USA) and 12 clones were sequenced using the fmol* DNA cycle sequencing system (Promega).

2.2. In silico analysis of the 5' upstream regulatory region and 5' untranslated region

In order to identify putative transcription factor binding sites, an analysis of the 5'-upstream sequence of the TSC1 gene was performed in silico by using the MatInspector Professional program (Quandt et al., 1995). The CpG island analysis was performed using CpG plot/CpG report (Rice et al., 2000) of the European Molecular Biology Open Software Suite (EMBOSS; http://www.ebi.ac.uk/emboss/cpgplot/). The function of the program ‘cpgplot’ is to plot CpG rich areas, and ‘cpgreport’ to report all CpG rich regions. In silico analysis of the 5' UTR for identifying known functional elements was done by using a sequence analysis tool known as UTRscan (Pesole and Liuni, 1999; http://bighost.area.ba.cnr.it/BIG/UTRScan/). UTRscan looks for UTR functional elements by searching through user submitted query sequences for the patterns defined in the UTRsite collection (http://bighost.area.ba.cnr.it/srs6/).

2.3. Comparison of human and mouse 5' upstream sequences

To compare the human and mouse TSC1 upstream regulatory regions (URRs), we obtained mouse Tsc1 genomic sequences (GenBank accession no. 036482) from the GenBank database. Multiple alignment of approximately 600 bp of the sequence upstream to the transcription start site region was carried out using the multiple sequence alignment program Clustal 1.8 at the BCM search launcher (http://searchlauncher bcm.tmc.edu/). Shading of the multiple alignment was carried out using the program BOX-SHADE 3.21 (EMBnet server; http://www.ch.embnet.org/software/BOX_form.html).

2.4. Construction of reporter plasmids

In order to generate deletion constructs, PCR fragments were amplified using primers containing KpnI and HindIII restriction enzyme sites in forward and reverse primers, respectively. Primer sequences were derived using DNA sequences of a genomic clone available in the public database (GenBank accession no. NT_035014) which contained the 5' end of the TSC1 gene. PCR was performed in a 25 µl reaction mixture containing 50–100 ng genomic DNA, 50 ng of each primer, 200 µM of each dNTP and 1 unit Dynazyme DNA polymerase with proofreading activity (Finnzyme, Finland) in a standard PCR buffer supplied by the manufacturer. After an initial denaturation step at 95 °C for 5 min, the reaction was carried out for 30 cycles at 94 °C for 45 s, 67 to 69 °C for 45 s, and 72 °C for 1–2 min with a final extension of 5 min at 72 °C in a Minicycler™ (MJ Research, USA). The fragments were purified by using a gel extraction kit (Concert™ Rapid Gel Extraction System, Life Technologies, USA) and cloned in the T/A cloning vector pGEM-T (Promega). Positive clones were double digested.
with KpnI and HindIII (Gibco-BRL, USA) in React™ 1 buffer and the inserts released were subcloned directionally upstream of the fire fly luciferase reporter gene of the pGL3-basic vector (Promega). Transfection quality DNA was isolated using the Wizard® plus Midipreps DNA purification system (Promega).

2.5. Cell culture, transfection and reporter gene analysis

Both HeLa cells and HepG2 cells were cultured in Dulbecco’s minimal essential medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin and 100 μg/ml streptomycin at 37 °C. A total of 2.5 × 10^4 HeLa or HepG2 cells/well were seeded into a 24-well culture plate 1 day prior to transfection. One hundred nanograms of each human TSC1-luciferase reporter gene construct was transfected to cells using the Effectene™ reagent (Qiagen, USA). The pRL-TK vector (Promega) harboring the Renilla luciferase gene was co-transfected as an internal control to normalize for transfection efficiency. An individual construct was transfected in triplicate for each transfection experiment and each experiment was repeated independently at least three times. After 36 h, cells were washed with phosphate buffered saline (PBS) and lysed with 100 μl passive lysis buffer (Promega). Luciferase activity was assayed using the Dual-Luciferase® Reporter Assay System (Promega). The intensity of chemiluminescence in the supernatant was measured using a luminometer (TD-20/20, Turner Designs, USA).

3. Results and discussion

3.1. Mapping of the transcription start site

We have made an attempt to identify the transcription start site (TSS) of the TSC1 gene using 5’ RACE (Campbell et al., 2002; Song and Goodman, 2000). The first round of PCR amplification using AP1 and HR1 primers did not show a distinct amplification product except for a faint streak (data not shown). However, the nested PCR using the PCR amplification product with AP1 and HR1 primers as a template and AP2 and HR2 primers showed several bands in the size range of 130 to 160 bp, which formed a broad band on agarose gel (data not shown). This band was gel purified and cloned in the pGEM-T easy vector (Promega). DNA sequencing of 12 clones revealed inserts with two different 5’ ends which were, respectively, 4 and 19 bases shorter than the published 5’ UTR end of the TSC1 cDNA (Fig. 1) (van Slegtenhorst et al., 1997). A TSC1 EST clone containing a 234 bp 5’ UTR was retrospectively found in the human EST database (GenBank accession no. BI463204); the clone contains a more upstream transcription start site (Fig. 1). This EST clone (BI463204) was obtained from a cDNA library constructed using the cap-trapper method (Carninci et al., 1997) and hence was expected to have a full-length 5’ UTR. Thus we have designated the 5’-most end of the cDNA obtained by the cap-trapper method as the transcription start site of the TSC1 gene. Multiple transcription initiation sites are known to occur in eukaryotic genes. Our finding of two TSSs in HeLa cells supports this observation (Fig. 1). Sp1 binding to GC boxes in TATA-less GC rich promoter is critical for transcription initiation, which is often directed from multiple sites (Sargsyan et al., 2002).

3.2. Identification of a novel alternatively spliced form

Cloning and sequencing of the 5’ RACE product from HeLa cells cDNA led to the serendipitous discovery of a novel alternative isoform in the 5’ UTR of the TSC1 gene involving exon 1 (isoform-2) (Fig. 2A). Several clones show a deletion of 29 bases near the end of exon 1. The deletion occurred precisely from a GT dinucleotide at position 63 from the TSS (Fig. 2B). It has been proposed that the 5’ donor splice site of the splice variants differs from the consensus splice site MAG|GTRAGT (where M is A or C and R is A or G; G of the GT dinucleotide is numbered as +1) at the +4 and +5 positions (Stamm et al., 2000). To test this possibility, we have aligned the 5’ donor splice site involved in the alternative splicing of isoform-2 with the consensus sequence (Stamm et al., 2000). Our analysis shows that this splice site differs at the +5 (A instead of G) and −3 (G instead of A or C) positions. A database search using the 5’ UTR of the human TSC1 as query against the human EST database (http://www.ncbi.nlm.nih.gov/dbEST/) has revealed five clones derived from different cell types such as neuroblastoma cells (GenBank accession no. AL528790), pre-B cells (GenBank accession nos. BE244888 and BE245145) and medulla (GenBank accession nos. BI827800 and BI826974) which also show alternative splicing including a deletion of 29 bases near the end of exon 1 similar to the RACE clones we have found in HeLa cells. This suggests that alternative splicing occurs in vivo and is not an experimental artifact. The presence of this isoform in different tissues indicates that it is not specific to HeLa cells. Moreover, a similar alternatively spliced isoform-2 was also found in the mouse EST database (http://www.ncbi.nlm.nih.gov/dbEST/) (GenBank accession nos. BB617455, BB654305, 622283). In addition to this novel isoform, the human TSC1 gene has also been shown to undergo alternative splicing of exon 2 (isoform-3) (van Slegtenhorst et al., 1997) (Fig. 2A). We have found a mouse EST clone (GenBank accession no. BI855559) in the mouse EST database that also show alternative splicing of exon 2. Moreover, there was a high degree of conservation (85% identity) in the sequence of the 5’ UTRs of human and mouse TSC1/Tsc1 genes (Fig. 2B). We believe that a high degree of conservation in the 5’ UTR sequence and the similar types of spliced isoforms are of functional significance.
Fig. 1. Analysis of the 5' upstream sequences. The nucleotide sequence representing the 5' flanking region and the first three exons of the \( TSC1 \) gene. Transcription start site (TSS) is represented by an upward arrow. The published end of the \( TSC1 \) cDNA is represented by an arrowhead below the sequence. The ends of the 5' UTR obtained during the RACE analysis is indicated by downward arrows above the sequence. The consensus sequences for the putative transcription factor binding sites and other sites are boxed and indicated above the sequence. The amino acid sequence starting from the ATG codon (marked by a rightward arrow) is shown below the sequence. The TSS is numbered as +1 and the rest of the sequence is numbered relative to it. Nucleotide numbers are indicated on the right side of the sequence.
Alternative splicing events that exclusively involve the 5′ UTRs of mammalian mRNAs are rather uncommon. Alterations to this region of the mRNA are known to have the potential to introduce post-transcriptional regulatory elements (Sonenberg, 1994). Several mechanisms of translational control by the 5′ UTR have been proposed. These include, but are not limited to, the primary structure of the mRNA affecting translational initiation (e.g. upstream AUGs affecting ribosome scanning), the secondary or tertiary structures modulating translational machinery (e.g. stem-loops, pseudoknots), and RNA binding proteins (e.g. IRP), among others. In many of these examples, translational regulation mediated by the 5′ UTR plays a key role in tissue and developmental stage specific expression of the gene involved. The long 5′ UTR of TSC1 may also harbor an internal ribosome entry site (IRES) which could be responsible for its translation. A database search to identify the functional elements in the UTR was carried out using the computer program UTRscan which looks for UTR functional elements by searching through user-submitted sequence data for patterns defined in the UTRsite collection (Pesole and Liuni, 1999). We could not detect any consensus pattern matching with the collection at the UTRsite. Interestingly, we have detected three non-overlapping upstream open reading frames (uORFs) in the 5′ UTR of TSC1 (Fig. 2B). The first uORF (uORF-1) is 11 codons long and is present in exon 1. It occurs precisely from nucleotide positions 29 to 61 from the TSS (Fig. 2B), just one base pair before the GT dinucleotide that is involved in the alternative splicing of isoform-2 (see Section A).
3.2). The second uORF (uORF-2) is 22 codon long. It starts in exon 2 from nucleotide position 97 and ends in exon 3 at nucleotide position 162 from the TSS. This uORF-2 is not present in isoform-3 (Fig. 2B). The third uORF (uORF-3) is eight codons long and is present in exon 3 (Fig. 2B). It occurs from nucleotide positions 171 to 194 from the TSS. All but the uORF-3 are in good Kozak’s context. All the three uORFs are also present in the 5′ UTR of the mouse Tsc1 gene. The lengths of the three uORFs are very well conserved in both mouse and human. Multiple sequence alignment of the amino acid sequence encoded by uORF-1, uORF-2 and uORF-3 showed 100%, 85% and 57% identity, respectively, between the human and mouse (data not shown). The significance of such a high identity between these short putative polypeptides is not known.

uORFs are known to regulate translation in eukaryotes (Geballe and Sachs, 2000). There are several mechanisms by which this can be achieved. uORFs can encode peptides that control reinitiation at the downstream ORF [e.g. the GCN4 system] or stall ribosomes and block scanning [e.g. the arginine attenuator peptide] (Geballe and Sachs, 2000). They could have a passive effect and influence reinitiation efficiency as a consequence of the distance between their termination codon and the initiation codon of the downstream ORF (Geballe and Sachs, 2000). There have also been reports for additional roles for uORFs such as stimulation of translation and effects on mRNA stability (Geballe and Sachs, 2000). The long 5′ UTR of TSC1 may also harbor an internal ribosome entry site (IRES). Therefore, the 5′ UTR of the TSC1 gene seems to be unique in the sense that it offers a very good system for the post-transcriptional regulation of this gene.

3.3. Analysis of the 5′ flanking sequence of the TSC1 gene

The 5′ flanking sequences of the TSC1 gene were analyzed for the presence of transcription factor binding sites. Analysis of an approximately 1.6 kb 5′ flanking sequence (−1474/+86) of the TSC1 gene using the MatInspector Professional Program (Quandt et al., 1995) revealed consensus binding sites for many transcription factors (Fig. 1). No consensus TATA or CAAT boxes are found in the sequence analyzed. Transcription factor Sp1 binding sites are found at four positions (Fig. 1). Sp1 transcription factor binding sites (known as GC boxes) are commonly found in the promoters of viral and cellular housekeeping genes. In addition, binding sites for several other transcription factors, namely GATA, E2F, c-Ets, CdxA, HSF2, SRY, Ik2 and USF are also found in the 5′ flanking region (Fig. 1). Binding of Sp1 to GC boxes of TATA-less promoters has been shown to be critical for transcriptional initiation of numerous constitutively active genes (Sargsyan et al., 2002). Moreover, E box-specific binding proteins, such as USF1 and USF2 belonging to the basic helix-loop-helix family, have also been shown to be important for the basal activity of a number of TATA-less and Inr-less promoters (Sargsyan et al., 2002). Furthermore, it has been demonstrated that the cooperation of Sp1 and USF factors is crucial for the transcriptional activation of the human transcobalamin II TATA-less promoter (Li and Seetharam, 1998). Additionally, E2F and Sp1 cooperation has been shown to be essential for the full promoter activity of several TATA-less genes (Huang et al., 2000). Taken together, these data suggest that transcription factors such as Sp1, USF and possibly E2F could function as key regulators for the basal activation of the TSC1 gene.

3.4. Function delineation of the TSC1 core promoter

To identify regions of promoter activity, HeLa and HepG2 cells were transfected with reporter plasmid constructs expressing luciferase controlled by a series of putative TSC1 promoter fragments. Eight different PCR fragments with length ranging from 250 to 1600 bp were cloned upstream of the luciferase reporter gene. Transiently transfected cells were lysed and assayed for luciferase activity. The four constructs (L1.6, L0.65, L0.58 and L1.3) exhibited significantly higher levels of reporter activity than the negative control (pGL3-basic) (Fig. 3). This suggests that the four constructs harbor elements that are essential for promoter function. The construct L0.58 shows maximum reporter gene activity whereas the activity of the larger constructs L1.6 (−1526/+77), L0.65 (−693/−55) and L1.3 (−1526/−242) is much lower. This suggests that a negative element lies between −510 to −1526 from the TSS. Deletion of 242 bp upstream of the TSS (L1.3) does not result in a significant reduction of luciferase activity when compared to the construct carrying the largest fragment (L1.6). By contrast, deletion of 536 bp upstream of the 5′ UTR (L1.0) as well as other constructs harboring larger deletions (L0.8, L0.5 and L0.25) lowers promoter activity to a level similar to that of the negative control (Fig. 3). These findings demonstrate that deletion of 242 bp of the 5′ URR reduces the promoter activity. The region from +77 to −242 seems to be necessary for optimal promoter function but is not essential for basal TSC1 expression. Therefore, the region spanning −242 to −510 harbors regulatory elements indispensable for basal transcription and presumably represents the core promoter region. We have found comparable results in both HeLa and HepG2 cells (Fig. 3B).

3.5. Identification of a CpG island

Analysis of the approximately 1.6 kb sequence flanking the TSS revealed the presence of a CpG island of approximately 521 bp (Fig. 4). The 521 bp CpG island spans bases +34 to −487 relative to the TSS and includes the region exhibiting promoter activity (see Section 3.4). This region of TSC1 has a 66.6% C + G content and exhibits an observed/expected value of 0.95, which easily meets the criteria for a CpG island (Larsen et al., 1992). Hypermethylation of CpG islands has been correlated with
reduced gene expression. Future studies of the methylation status of the CpG island in the core promoter of the TSC1/Tsc1 may help resolve mechanisms of tumor development as suggested in other tumor suppressor genes (Jones, 1996).

### 3.6. Comparison of human and mouse upstream regulatory regions (URRs)

To compare the human and the mouse TSC1 upstream regulatory regions and to identify potentially conserved regulatory regions, we have carried out multiple sequence analysis of nearly 600 bp region flanking the TSS of both human and mouse (Fig. 5). The two orthologous URRs reveal an overall sequence identity of 55% within 600 bp upstream of the TSS. Regions of higher conservation are confined to specific motifs (Fig. 5). The conserved regions coincided with putative transcription factor binding sites identified in silico using the MatInspector Professional program. Sp1 sites in the URRs are well conserved in both human and mouse. In addition, the transcription factor binding sites for E2F, CdxA, Ik-2, GATA, c-Ets, HSF2, USF and SRY are also conserved in both human and mouse. These conserved sequences may be important for expression and/or regulation of the TSC1/Tsc1 gene. Future expression studies using mutant constructs for these transcription factor-binding sites may uncover their importance.

Comparative analysis of the genomic organization of the mouse Tsc1 and human TSC1 genes carried out by Cheadle et al. (2000b) has revealed a conserved region of around 500 bp upstream of the TSS. This region was compared with the matrices representing consensus transcription factor binding sites from the TRANSFAC database (it is a database of transcription factors, their genomic binding sites and DNA-binding profiles; http://transfac.gbf.de/TRANSFAC/) with the MatInspector program (Cheadle et al., 2000b). Conserved putative transcription factor binding sites for Sp1, USF and TCF11 were identified (Cheadle et al., 2000b). The results of comparison of mouse and human URRs carried out by Cheadle et al. (2000b) are in agreement with the observations we made in the present study (Fig. 1). However, there is some difference between the results obtained by Cheadle et al. (2000b) using MatInspector and those found in the present study using MatInspector Professional, possibly due to the fact that the MatInspector Professional program is an advanced version. Cheadle et al. (2000b) have identified one site each for SP1, TCF11 and USF transcription factors which were conserved in human and mouse, whereas we have identified a total of 15 sites: four for SP1, one for USF, two for E2F, three for GATA, one for c-Ets, one for CdxA, one for HSF2, one for SRY and one for Ik2.

The absence of TATA-like sequences, the presence of Sp1 binding sites and more importantly the presence of a CpG
island in the regulatory regions as found in the TSC1 gene are the characteristic features of the promoters of various housekeeping genes and ubiquitously expressed tumor suppressor genes (Kuzmin et al., 1995 and references therein). The ubiquitous expression of the tumor suppressor gene TSC1 is consistent with the structural characteristics of its 5' flanking region.

There are several familial cases linked to the TSC1 locus that do not show any mutations in the coding region (see Section 1) (Cheadle et al., 2000a). It is hypothesized that mutations in these families could be in the promoter region. Mutations affecting the promoter regions of several genes have been described in many disorders. For example, mutations in the promoter region of the β-globin gene result in β-thalassemia, which is due to low abundance of β-globin protein. Many mutations in the β-globin gene promoter result in the decrease or total loss of transcription (McDonagh and Nienhuis, 1993). Similarly, mutations affecting the promoter have been described in the retinoblastoma tumor suppressor gene Rb1 (Cowell and Cragg, 1996). The identification of the TSC1 promoter in this study will help in designing a suitable strategy to identify mutations in this region in patients who do not show mutations in the coding sequences. Methylation of a CpG

Fig. 4. Identification of a CpG island surrounding the transcriptional start site. (A) Diagram of the region of the human TSC1 gene subjected to CpG island analysis. The region showing the core promoter activity is shown as a hatched box. (B) Diagram showing occurrence of observed/expected ratio of CpG dinucleotides. (C) Analysis of the G + C content of this region using a window size of 100. (D) Location of a CpG island identified by CpGplot with a threshold set at 0.6 and a minimal length of 200. All panels are spatially aligned with each other.
island has been correlated with reduced gene expression. Hypermethylation in the \( Rb1 \) gene was found to be associated with a unilateral, sporadic form of retinoblastoma (Ohtan-Fujita et al., 1997). By analogy, studies of the methylation status of the CpG island in the core promoter of the \( TSC1/Tsc1 \) may resolve mechanisms of tumor development in TSC patients.

As stated above, the tuberous sclerosis complex is due to mutations in two genes, \( TSC1 \) and \( TSC2 \). Little is known about the regulation of these genes and their potential regulatory elements. During the present study, we have identified a novel isoform of the \( TSC1 \) gene transcript involving the 5' UTR and discussed its possible role in the \( TSC1 \) gene regulation. We have identified uORFs in the conserved 5' UTR of the \( TSC1 \) gene and discussed their potential roles in the post-transcriptional regulation of the \( TSC1 \) gene. We have also identified the promoter of the \( TSC1 \) gene using a luciferase reporter system. As stated above, it is hypothesized that some of the mutations lie in the promoter of the \( TSC1 \) gene. Based on the hypothesis that mutations in non-coding regions and/or aberrant methylation pattern in regulatory regions of the \( TSC1 \) gene are responsible for some cases of TSC, the identification of the promoter will facilitate in designing PCR primers that will help us to look for mutations in this region. We have identified several

---

Fig. 5. Comparison of the human and mouse upstream regulatory regions (URRs) and results of the in silico transcription factors binding sites analysis. The nucleotide positions are numbered with respect to TSS, which is shown by a downward arrow above the sequence. The transcription factor binding sites are boxed and labeled accordingly. The solid black shading represents the sequence homology and stippled grey shading represents a mismatch.
putative transcription factor-binding sites and discussed their potential role in the transcription of the TSC1 gene. These putative transcription factor-binding sites can now be tested for their role in TSC1 gene regulation and the disease process.

4. Conclusion

A novel alternative splicing in the 5′ UTR of the TSC1 gene has been found. We have also identified three uORFs in the 5′ UTR of the TSC1/Tsc1 gene. Putative binding sites for several known transcription factors, namely Sp1, E2F, CdxA, GATA, c-Ets, HSF2, Ik2, USF and SRY are found in the upstream region. In this study, we have reported the isolation and characterization of 1.6 kb of the 5′ flanking region of TSC1. Maximal promoter activity is present in a 587-bp region +77 to –510 bp with respect to the TSS in the TSC1 upstream region. A CpG island of 521 bp is found in a region which overlaps with the promoter region identified in this study.

Acknowledgements

This work was financially supported by a research grant from the Department of Biotechnology, Government of India, New Delhi to AK and SCG, and a CSIR JRF to MA. We thank Dr. P. Kondaiah and Mrs. Urvashi Bahadur for their help in tissue culture and suggestions. We also thank Prof. V. Nanjundiah, Dr. S. Das and Dr. U. Varshney for their comments on the manuscript. We also thank the anonymous reviewers for their valuable suggestions to improve the manuscript.

References


Pesole, G., Liuni, S., 1999. Internet resources for the functional analysis of 5′ and 3′ untranslated regions of eukaryotic mRNA. TIG 15, 378.


