Available online at www.sciencedirect.com
ScienceDirect
Gene 395 (2007) 86–97
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Functional characterization of the *TMLH* gene: promoter analysis, in situ hybridization, identification and mapping of alternative splicing variants

Jlenia Monfregola [a], Gennaro Napolitano [a], Ivan Conte [b], Armando Cevenini [a], Carmen Migliaccio [a], Michele D'Urso [a], Matilde Valeria Ursini [a,*]

[a] *Institute of Genetics and Biophysics "Adriano Buzzati-Traverso" (CNR), Via P.Castellino, 111, 80131 Naples, Italy*
[b] *Departamento de Neurobiología del Desarrollo, Instituto Cajal, CSIC, Dr. Arce 37, Madrid 28002, Spain*

Received 15 December 2006; received in revised form 9 February 2007; accepted 9 February 2007
Available online 28 February 2007
Received by A. Bernardi

Abstract

Carnitine is a molecule with well-documented pleiotropic functions whose biosynthesis involves four catalytic steps. Here, we report a detailed analysis of the expression and transcriptional control of *TMLH* gene, which codifies for the first enzyme of carnitine biosynthesis. *TMLH* maps at the extreme end of Xq28, a chromosomal region of high genomic instability. By 5′ and 3′ RACE, we identified and mapped two alternative 5′ *TMLH* first exons and seven alternative 3′-splice variants, which are spread over a genomic region of about 250 kb. While the two alternative 5′ exons have different expression profiles, all the 3′ alternative forms are ubiquitously expressed. Reporter assays revealed that the 3′-UTRs of each *TMLH* isoform might influence its own expression at post-transcriptional level. In addition, we identified a highly conserved promoter region of *TMLH*. Functional analysis of this region showed the presence of a CpG island, whose methylation-status could control the level of *TMLH* transcription. Finally, by mRNA in situ hybridization, we found that *TMLH* expression is present at E12.5 dpc in the mouse liver, lung and brain, and is then maintained in the postnatal brain with a specific neuronal pattern. Collectively, our data highlight a tight transcriptional and post-transcriptional control of *TMLH* expression.

Keywords: Carnitine biosynthesis; Xq28; 3′UTR

1. Introduction

L-Carnitine plays an indispensable role in fatty acid oxidation, but it has also additional pleiotropic functions and it is used as safe therapeutic agent for its well-tolerated properties (Virmani and Binienda, 2004; Nalecz et al., 2004). Animals obtain carnitine primarily from the diet, nevertheless most eukaryotes can synthesize it endogenously through a biochemical process that involves four catalytic steps (Vaz and Wanders, 2002). In the adult human the primary site of carnitine biosynthesis is the kidney, although also the liver, heart, muscle and brain express the enzymes of carnitine biosynthetic pathway. In addition recent reports highlighted the crucial role of carnitine production during mammalian development and suggest that carnitine biosynthesis may have an important role for the foetal metabolism (Oey et al., 2006). *TMLH* is the first and the key enzyme in carnitine biosynthesis (Swiegers et al., 2002). It is a non-heme-ferrous iron hydroxylase Fe^{++} and 2-oxoglutarate dependent, which hold a very conserved motif including an istidine at position 389 required for full enzymatic activity (Monfregola et al., 2005). Moreover TMLH is the only enzyme of carnitine biosynthesis located in the mitochondria, where it is targeted through a Mitochondrial Target Sequence (MTS), located at the N-terminal of the protein sequence (Monfregola et al., 2005).

Although a characterization of the TMLH enzyme and the identification of the catalytic domain have been produced

Abbreviations: cDNA, DNA complementary to RNA; EST, Expressed Sequence Tag; kb, kilobase; MTS, Mitochondrial Target Sequence; mRNA, messenger RNA; RACE, Rapid Amplification of cDNA Ends; RTPCR, reverse transcription-polymerase chain reaction; TMLH, Trimetil-Lysine Hydroxylase; UCSC, University of California Santa Cruz Genome Browser; UTR, Untranslated Transcription Region.

* Corresponding author. Tel./fax: +39 081 6132262.
E-mail address: ursini@igb.cnr.it (M.V. Ursini).

0378-1119/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.02.012

(Monfregola et al., 2005), the complete functional genomic characterization of *TMLH* locus is still not complete. *TMLH* is located in the distal portion of Xq28, a very intricate region in which several diseases were mapped (Kolb-Kokocinsky et al. 2006). This region has an interesting evolutionary history and specific loci within it are characterized by an instability that predisposes them to cause several diseases (Aradhya et al., 2002). Recently, the complete sequence of the X chromosome has been elucidated (Ross et al., 2005), nevertheless the complexity of the distal Xq28 makes it difficult to define the functional genomic features of many genes within this region. Indeed, this region contains various repeat sequences, including the three int22h repeats that are particularly interesting because of their sequences, copy number, location and orientation (Bagnall et al., 2005a,b). A recombination between the int22h, causing inversion and breaking of the factor VIII gene, is the major cause of Haemophilia A. Recent sequence data revealed that the int22h2 and int22h3 are contained in two regions repeated and inverted forming two ARMS of a palindrome prone to recombination. This recombination event could explain the mechanism that causes the genetic factor VIII defect in 45% of severe haemophilia A cases (Ross and Bentley, 2005; Bagnall et al., 2005a,b).

In the present report we provide an analysis of the regulatory regions of the *TMLH* locus, by mapping its transcription initiation, termination sites and promoter region. Using parallel approaches consisting of comparative evolutionary studies, EST mapping and computational analyses of the human sequence of the X chromosome now available, we reveal a complex structure and evolution of the *TMLH* locus, which confirms the dynamic nature of this genomic region. By 5′ and 3′ RACE we identified two 5′ and seven 3′ variants, which all arise from alternative splicing. Some of these forms include sequences located at a great distance from the bulk of the gene, in a region of recombination between the two Int22h sequences. Moreover this structure is only present in human *TMLH*, suggesting that it is the product of a recent evolutionary event.

We analysed the expression and the functional significances of alternative *TMLH* 3′-isoforms in an in vitro reporter system. We also identified the regulatory region responsible for the *TMLH* transcriptional regulation, characterizing a CpG island and defining that its methylation-status plays a role in *TMLH* transcriptional control.

In addition, we describe a detailed study of the expression of the *TMLH* gene using in situ hybridization in the mouse, showing that it is abundant in the liver, lung and brain at E12.5 dpc developmental stage and that maintains a cell-specific neuronal pattern in the mouse brain at stage P7.

2. Materials and methods

2.1. Cell cultures

HEK293 (Human Embryonic Kidney) cell lines were grown in DMEM (Dulbecco's modified Eagle medium) supplemented with 10% foetal bovine serum and routine antibiotics in a humidified atmosphere containing 5% CO_2 and 95% air at 37 °C.

2.2. Gene expression analysis by RT-PCR

Human total RNAs from different tissues (BD Clontech, Inc. Palo Alto, CA) were used as a source of cDNA. Reverse transcription was done with the Superscript III enzyme (Invitrogen, corp. Carlsbad, CA) according to the manufacturer's protocol. The *TMLH* 5′-UTR variants were detected using a unique reverse primer in exon 3 (ex3r, TTAGAGTTGTAGCACGATGC) and a specific forward for each first exon: (1aF, AATTCTGCCCACTCTGCGGA; 1bF, TTGTGGCTGGTCTGAGGTT). The 3′UTR variants were detected with a unique forward primer in exon 6 (CTACCCATGGAATAAAGAGC) and a specific reverse primer for each 3′UTR (*TMLHc,* CTCGCGTTTGAATATGATATG; *TMLHd,* AAGACATCTTTTGTTTATTGAAC; *TMLHe,* AAGTGTTTAATACATCATCTCTT). For cDNA testing we used *β-actin* specific primers (Salvemini et al., 1999). To detect the *TMLHg* and *TMLHf* forms we used a forward primer annealing in the first exon (1aF, AATTCTGCCCACTCTGCGGA) and a reverse primer annealing in the fourth exon (4aR, AGTCCCTCGTTGGTTTCTA). The full *TMLH* expression was detected using primers annealing in the exon 4 (4bF, CTATGAAGGGCAGAAACAA) and in the exon 6 (6aR, ACATTCTCCAACATCTTCAA). All amplified fragments were verified by sequencing.

2.3. Bioinformatic analysis

Analysis of the nucleotide sequences was performed using annotations available at UCSC (University of California, Santa Cruz Genome Browser). The genomic region containing the *TMLH* putative promoter was analysed using the software m-Vista: http://www-gsd.lbl.gov/vista/VistaInput that allowed us to align two chosen sequences and to determine the percentage of identity between them. We determined conserved segments using as criteria 80% identity for 70 bp length. Promoters were predicted with Promoter Scan: http://bimas.dcrt.nih.gov/molbio/proscan/ and PromoterInspector soft-ware; predicted transcription factor binding sites were scanned by the Genomatix suite http://www.genomatix.de/. The CpG island and relative plot were generated as previously described (Ambrosio et al., 2002).

2.4. Construction of the expression-plasmid, transient transfection and Luciferase assay

All promoter fragments of *TMLH* were amplified by PCR from human genomic DNA using the primers listed in Table 1. Amplified fragments were double digested with XhoI/HindIII and cloned in pGL3-Basic vector (Promega, Madison WI) upstream of the *Photinus pyralis* luciferase coding sequence, which contains an acceptor splice site at its 5′ end (Crispi et al., 2004).

The *TMLH* 3′ UTRs were amplified by PCR from the Human Adult Brain (BD Clontech, Inc. Palo Alto, CA) cDNA using the primers listed in Table 1. The amplified products were digested with *Fse1/Sal1* and cloned in pTP1-Luc vector downstream of the Luciferase coding sequence.