Gene 533 (2014) 488–493

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Cloning and characterization of the promoter regions from the parent and paralogous creatine transporter genes

Joseph D.T. Ndika [a,b], Vera Lusink [a], Claudine Beaubrun [a], Warsha Kanhai [a], Cristina Martinez-Munoz [a], Cornelis Jakobs [a], Gajja S. Salomons [a,b,c,*]

[a] Department of Clinical Chemistry, Metabolic Unit, VU University Medical Center, Amsterdam, The Netherlands
[b] Neuroscience Campus, VU University Medical Center, Amsterdam, The Netherlands
[c] Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Accepted 2 October 2013
Available online 18 October 2013

Keywords:
Creatine transport
Transcriptional regulation
Promoter
Pseudogene

ABSTRACT

Interconversion between phosphocreatine and creatine, catalyzed by creatine kinase is crucial in the supply of ATP to tissues with high energy demand. Creatine's importance has been established by its use as an ergogenic aid in sport, as well as the development of intellectual disability in patients with congenital creatine deficiency. Creatine biosynthesis is complemented by dietary creatine uptake. Intracellular transport of creatine is carried out by a creatine transporter protein (CT1/CRT/CRTR) encoded by the *SLC6A8* gene. Most tissues express this gene, with highest levels detected in skeletal muscle and kidney. There are lower levels of the gene detected in colon, brain, heart, testis and prostate. The mechanism(s) by which this regulation occurs is still poorly understood. A duplicated unprocessed pseudogene of *SLC6A8–SLC6A10P* has been mapped to chromosome 16p11.2 (contains the entire SLC6A8 gene, plus 2293 bp of 5′flanking sequence and its entire 3′UTR). Expression of SLC6A10P has so far only been shown in human testis and brain. It is still unclear as to what is the function of SLC6A10P. In a patient with autism, a chromosomal breakpoint that intersects the 5′flanking region of SLC6A10P was identified; suggesting that SLC6A10P is a non-coding RNA involved in autism. Our aim was to investigate the presence of cis-acting factor(s) that regulate expression of the creatine transporter, as well as to determine if these factors are functionally conserved upstream of the creatine transporter pseudogene.
Via gene-specific PCR, cloning and functional luciferase assays we identified a 1104 bp sequence proximal to the mRNA start site of the *SLC6A8* gene with promoter activity in five cell types. The corresponding 5′flanking sequence (1050 bp) on the pseudogene also had promoter activity in all 5 cell lines. Surprisingly the pseudogene promoter was stronger than that of its parent gene in 4 of the cell lines tested. To the best of our knowledge, this is the first experimental evidence of a pseudogene with stronger promoter activity than its parental gene.

1. Introduction

1.1. Creatine pathway

Creatine is a nitrogenous organic acid whose intracellular pool is maintained by both endogenous synthesis and nutritional uptake. The creatine transporter gene, also known as *SLC6A8* (GeneID 6535) or CT1, is located on the human chromosome Xq28 and consists of 13 exons (Gregor et al., 1995; Sandoval et al., 1995). Mutations in the *SLC6A8* gene lead to congenital creatine deficiency (Salomons et al., 2001). The importance of creatine transport is highlighted by in vitro studies in cerebellar granule cells which show that creatine transport seems to be more efficient in building up total intracellular creatine than de novo synthesis (Carducci et al., 2012). The two enzymes responsible for endogenous creatine synthesis are AGAT (from arginine and glycine) and GAMT (via methylation of guanidinoacetate produced by AGAT). Genetic deficiency of either creatine synthesis enzymes results in depletion of brain creatine, similar to SLC6A8 deficiency (OMIM: 300036). Key clinical features of creatine deficiency are intellectual disability, autism like behavior, movement disorders and epilepsy (Mercimek-Mahmutoglu et al., 2009). Creatine supplementation therapies, especially when initiated in the newborns, have been successful in preventing disease onset and in treating several cases of creatine biosynthesis deficiency (Bianchi et al., 2000, 2007;

Abbreviations: ATP, adenosine triphosphate; bp, base pair; kb, kilo base pair; cDNA, DNA complementary to RNA; GAPDH, Glyceraldehyde 3-phosphate dehydrogenase; gDNA, genomic DNA; EGFP, enhanced green fluorescent protein; SV40, Simian virus 40; chr, chromosome; MEFs, mouse embryo fibroblasts; HEK293, human embryonic kidney cell line; 3T3 Swiss, mouse fibroblast cell line; SK-N-SH, human neuroblastoma cell line; mRNA, messenger RNA; miRNA, microRNA; ORF, open reading frame; PCR, polymerase chain reaction; RTPCR, reverse transcriptase-polymerase chain reaction; SLC6A8, solute carrier family 6, member 8; SLC6A10P, solute carrier family 6, member 10 pseudogene; UTR, untranslated region.

* Corresponding author at: Department of Clinical Chemistry, Metabolic Unit, VU University Medical Center, Amsterdam, The Netherlands. Tel.: +31 204442880; fax: +31 204440305.
E-mail address: g.salomons@vumc.nl (G.S. Salomons).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.10.008

Mercimek-Mahmutoglu et al., 2006; Ndika et al., 2012; Schulze and Battini, 2007). No treatment options yet exist for creatine transporter deficient patients. Thus unraveling aspects of creatine transporter regulation remain crucial; both in terms of understanding the pathophysiology of creatine biosynthesis deficiencies — to improve existing treatment regimens, and also to explore possible avenues for therapy vis-a-vis SLC6A8 deficiency.

1.2. Intellectual disability

Intellectual disability (now the preferred term for mental retardation) is the most common developmental disorder with a worldwide prevalence of 1.37% (Maulik and Mascarenhas, 2011). For the majority of cases of inherited intellectual disability, the genetic cause has not yet been elucidated. X-linked intellectual disability is estimated to account for 5%–12% of all cases of intellectual disability (Herbst and Miller, 1980). The frequency of *SLC6A8* mutations in an XLMR population of 288 patients was as high as 2.1% (Rosenberg et al., 2004). In another study, a 1% prevalence of mutations in *SLC6A8* was found in boys with intellectual disability of unknown etiology (Clark et al., 2006). So far patients are being diagnosed either via metabolic workup (i.e. creatine and guanidinoacetic acid measurements in urine and plasma) via cranial MRS or genetic testing. In the latter, so far only the coding exons and the neighboring splice sites of the *SLC6A8* gene are being analyzed. The unknown promoter region has not yet been included in these analyses. The functional relevance of such (promoter) regions is highlighted by Dunham et al., based on their observation that disease-relevant single nucleotide polymorphisms (SNPs) are enriched within non-coding functional elements — a majority of which reside within or in the vicinity of regions that are outside of protein-coding genes (Dunham et al, 2012).

1.3. Duplicated paralogues of SLC6A8 on chromosome 16

There is a poor understanding of why some genes (paralogous genes) are amplified in our genome during evolution and whether they have a function. Until recently it was believed that paralogous genes were "faults" of nature and had no function. However, nowadays several paralogue genes are known to be expressed and even functional (Pei et al., 2012). A duplicated paralogue of *SLC6A8–SLC6A10* (alias CT2) with a transversion of the T in its initial termination codon to a G, extending its open reading frame by 50 amino acids, was mapped to chromosome 16p11.2 (Iyer et al., 1996; Xu et al., 1997). However in chromosome 16 the predicted amino acid sequence was found to harbor a nonsense mutation in "exon 4" (compared to CT1), indicating that a creatine transporter protein cannot be translated from the SLC6A10 mRNA and is most likely a pseudogene (Eichler et al., 1996; Iyer et al., 1996). This established the basis for a change in nomenclature from *SLC6A10* to *SLC6A10P* (Gene ID: 386757). Further clarification on the nature of the SLC6A8 duplication on chromosome 16 was provided by Höglund and colleagues (Höglund et al., 2005). By searching the May 2004 assembly of the human genome on the UCSC genome browser, they found out that there are two adjacent pseudogenes of SLC6A8 on chromosome 16: one at 32797531–32799840 on the reverse strand and the other at 33690486–33692794 on the forward strand. Both loci share a 95.8% sequence similarity with SLC6A8. All published instances of SLC6A10P referred to the pseudogene on the reverse strand, now denoted as SLC6A10pA (Gene ID: 386757), while that on forward strand, SLC6A10pB is listed in Entrez Gene as a predicted gene (Gene ID: 653562). However, with a sequence similarity of 99.6%, it is very likely that some of the cDNA identified as SLC6A10pA also included transcripts from SLC6A10pB. SLC6A10P transcripts (most likely *SLC6A10pA* and *SLC6A10pB* messages) have been reported in testis (Iyer et al., 1996) and brain (Bayou et al., 2008). Moreover, according to data from the Gene Expression Atlas — a database of publicly available gene expression data obtained from functional genomics experiments by the European Bioinformatics Institute (EMBL-EBI), differential expression of SLC6A10P (*SLC6A10pA* and *SLC6A10pB*) has been seen for some tissues (kidney, skeletal muscle, spinal cord, etc) and cell lines (HT1080, Hela, A498, etc) (http://www.ebi.ac.uk/gxa/gene/ENSG00000214617). This differential and tissue-specific expression pattern suggests that there is/are functional role(s) associated with the *SLC6A10P* pseudogenes. Supportive of this hypothesis, a translocation breakpoint on chromosome 16p11.2 was mapped to disrupt the 5′flanking sequence of SLC6A10pA in a patient presenting with autism (Bayou et al., 2008). Approximately 2.3 kb of the 5′flanking sequence of *SLC6A10pA* and *SLC6A10pB* shares 95% homology to the same region on their parent gene, and 99.9% homology with one another. Other than their differential expression across tissues and experimental conditions, nothing else is known of the regulation and even function of the *SLC6A10P* pseudogenes. As a first step towards understanding the physiological relevance of these pseudogenes we investigated the presence of a functional promoter and its activation across different cell types. Sequence analysis of the cloned promoter reveals it to be the 5′flanking sequence of the pseudogene on the reverse strand (SLC6A10pA). For simplicity we will refer to both SLC6A10pA and SLC6A10pB as SLC6A10P, except otherwise indicated.

2. Methodology

2.1. Sequence alignment and analysis

Using the genome browser feature on UCSC Genome Browser (February 2009 version), we obtained a sequence of approximately 3 kb (2868 bp) upstream of the start of the 5′UTR of SLC6A8. Next we performed a BLAT alignment (Kent et al., 2002) of this region (including the entire 5′UTR [279 bp] and 33 bp into the *SLC6A8* open reading frame [ORF]). Restriction site analysis of the selected potential upstream regulatory region (URR) — 3180 bp in total, was done in Vector NTI™ (Invitrogen, Carlsbad, CA, USA).

2.2. Isolation of 5′flanking sequences

DNA isolation was carried out using a genomic DNA (gDNA) isolation kit (Promega). Via PCR on isolated gDNA, 5′flanking sequences of *SLC6A8* were amplified from whole blood of a healthy donor (male) while 5′flanking sequences of *SLC6A10* were amplified from the fibroblast cells of an anonymized individual (male) with a complete deletion of the SLC6A8 locus. Primers were designed to amplify and clone 5′flanking sequences of −3146/+34 base pairs (bp) relative to the first methionine of *SLC6A8* and −1016/+34 bp relative to the same methionine on *SLC6A10pA*/*SLC6A10pB*. The primers used to obtain the 5′flanking sequence of SLC6A8 were gene-specific while those used to obtain 5′flanking sequences of the pseudogenes could also amplify SLC6A8; however this is avoided by carrying out the PCR on gDNA of the individual with a deletion of the SLC6A8 locus. PCR primers were designed using Vector NTI™ and synthesized by IDTdna (Leuven, Belgium).

All PCRs were carried out using a high fidelity polymerase (KAPA HiFi™ HotStart — KapaBiosystems, MA, USA) as specified by the manufacturer. A gradient of different annealing temperatures (Ta) was used to determine the most optimum Ta for each primer pair. In order to enable cloning of PCR products into the reporter vector, forward and reverse primers were designed to contain restriction site sequences at their 5′ends. Potential 5′flanking regulatory regions as well as PCR primers are depicted in Table 1. Truncated 5′flanking sequences of 2345 bp (−2311/+34) and 1104 bp (−1070/+34) were derived by making use of native *XhoI* and *SacI* restriction sites within the −3146/+34 — *SLC6A8* fragment and a created *HindIII* restriction site at its 3′end.