DNA RESEARCH 7, 339–345 (2000)

Cloning and Genomic Organization of the Mouse Gene *Slc23a1* Encoding a Vitamin C Transporter

Suzana GISPERT,[1] Amalia DUTRA,[1] Andrew LIEBERMAN,[2] Dan FRIEDLICH,[3] and Robert L. NUSSBAUM[1,*]

Genetic Disease Research Branch, National Human Genome Research Institute,[1] Neurogenetics Branch, National Institute of Neurological Disease and Stroke,[2] and Howard Hughes Medical Institute, NIH, Bethesda, Maryland 20892, USA[3]

(Received 28 August 2000)

Abstract

Vitamin C is known to exist in particularly high concentrations in brain tissue, and its free radical scavenging function is thought to represent a major antioxidative defense system. We have cloned, sequenced and analyzed the genomic structure of a mouse sodium-dependent vitamin C transporter gene, *Slc23a1* (also known as *Svct2*). The mouse *Slc23a1* cDNA is 6.4 kb long and was cloned directly from a mouse brain RNA preparation. Hybridization screening of a mouse genomic BAC library identified BAC 53L21 which contains at least the entire coding sequence of the mouse *Slc23a1* gene. Determination of the exon-intron structure of the gene revealed 17 exons ranging from 58 bp to 4407 bp extending over 50 kb of the mouse genome, with the translation start codon located in exon 3. Its 1944 nucleotide open reading frame encodes a polypeptide of 647 aa, which is highly similar to rat and human orthologs. The mouse gene was assigned to chromosome 2qG2 by fluorescence *in situ* hybridization analysis. Expression of this gene was demonstrated in a wide range of tissues, with especially high levels in brain. Neurodegenerative diseases with an established role for oxidative stress in the cytoplasm may therefore be conditions of SLC23A1 dysfunction.
Key words: gene structure; Vitamin C; transporter; oxidative stress

1. Introduction

Vitamin C (ascorbic acid or dehydroascorbate) is an important antioxidant system in living organisms. The most accepted physiological role of vitamin C is in the biosynthesis of collagen, in addition to other functions like the biosynthesis of norepinephrine and carnitine and the catabolism of tyrosine.[1] In all these functions, the role of the vitamin C is to provide electrons to keep prosthetic metal ions (Fe^{+2} and Cu^{+2}) in their reduced forms.[1] The function of vitamin C as a cofactor of dopamine β hydroxylase in the synthesis of norepinephrine explains the high levels of this vitamin in adrenal chromaffin granules. However, the high levels in brain which are only comparable with levels in the adrenal gland and are ten fold higher than in serum are still unclear. One possible explanation for the role of ascorbic acid as a protective factor against oxidative stress has been proposed: Vitamin C can scavenge singlet oxygen, hydroxy and superoxide radicals and is able to spare vitamin E, another important lipophilic antioxidant. Another explanation may be the need of brain tissue to maintain prosthetic metal ions in their reduced forms.[2]

Most terrestrial species are able to synthesize vitamin C; primates (including humans), guinea pigs and flying mammals are exceptions. These species lack a functional L-gulono-γ lactone oxidase, the last enzyme in the pathway for the synthesis of vitamin C.[3] Vitamin C is found in two different forms in the body; the reduced form, ascorbic acid, is the principal form present in blood and brain, while the oxidized form, dehydroascorbate, is present at barely detectable levels.[4]

There are two known mechanisms for the transport of vitamin C: one is mediated by the glucose transporters GLUT1 and GLUT3,[5,6] and the other by a sodium-dependent transporter. GLUT1 and GLUT3 transport only the oxidized form of the vitamin; but since this form is very scarce, this pathway of vitamin C uptake is unlikely to be important. Two sodium-dependent vitamin C transporters, SVCT1 and SVCT2, were recently discovered in rats and humans.[7–12] The expression of these two transporters in *Xenopus laevis* oocytes induces Na^+-dependent ascorbic acid uptake to occur. SVCT1 is mainly found in the epithelial system (kidney, intestine

Communicated by Shoji Tsuji
* To whom correspondence should be addressed. Genetic Disease Research Branch, Building 49, Room 4A72 NIH, 49 Convent Drive, Bethesda, MD 20892-4472, USA. Tel. +1-301-402-2039, Fax. +1-301-402-2170, E-mail: rlnuss@nhgri.nih.gov
† Accession number: AY004874

and liver), while SVCT2 is more ubiquitously distributed with the highest levels in adrenal gland and brain, particularly in neurons.[8]

In this investigation, we analyzed the mouse *Slc23a1* gene with its cDNA sequence, genomic structure, chromosomal localization and expression pattern.

2. Materials and Methods

2.1. RNA preparation, reverse transcription and identification of the complete cDNA sequence

Total RNA was isolated from an adult FVB/N mouse brain, with (1 ml/50 μg tissue) TRIZOL (Life Technologies). The first strand was prepared using First-Strand cDNA Synthesis Kit (Pharmacia) and the primer *Not* I-d(T)$_{18}$ provided in the kit.

For the initial identification of mouse *Slc23a1*, we performed amplification reactions using primers based on the rat SVCT2 sequence (Accession # AF080453). The forward primer corresponds to nucleotides 310–331 (5′ ctcccggtggtgatcaatggtg 3′) and the reverse primer corresponds to nucleotides 2137–2157 (5′ aggctatactgtggcctggga 3′) of the rat sequence. The amplification reaction produced a fragment of the expected size of 1.9 kb, and this coding region cDNA was cloned in the vector pT7Blue3 with the kit Perfectly Blunt (Novagene).

The 5′ untranslated region (UTR) was determined by PCR using the mouse brain Marathon-Ready cDNA Kit (Clontech), the primer AP2 (Clontech) and a gene-specific primer (GSP) 5′ actggtcatcccccacgcac 3′. The 5′ RACE (Rapid Amplification cDNA Ends) was performed in a 20 μl final volume using 2 μl of the mouse brain Marathon-Ready cDNA template, 0.2 μM from each primer and 0.5 U Taq polymerase and buffer (Perkin Elmer).

To determine the 3′ UTR, we performed a PCR in the genomic clone BAC 53L21 with a forward primer in the 3′ end of the coding region (5′ ccatttggcatgaacattatta 3′) and a reverse primer from the very end of the human SLC23A1 EST clone KIA00238[13] (5′ acggaacagttttaaaaatacagaata 3′). For amplification, we used 3.75 U of the Expand Long Template enzyme mixture in buffer 1 (Boehringer Mannheim) and the conditions recommended by the manufacturer.

2.2. Northern blot analysis

The cloned 1.9-kb PCR product spanning the coding sequence of the mouse *Slc23a1* gene was labeled with α-^{32}P-dCTP using "Ready To Go" (Pharmacia). This probe was hybridized overnight to commercial mouse multiple tissues Northern blots (Clontech) at 42°C in 50% formamide, 5 × SSPE, 1 × Denhardt's, 1% SDS, 2% vol. sheared DNA and 50% dextran sulfate, and was subsequently washed in 2 × SSC, 0.2% SDS at 55°C.

2.3. Screening of BAC library and characterization of BAC clones

High-density filters of the mouse 129sv/ev genomic BAC library RPCI22 from Research Genetics were hybridized with the 1.9 kb coding region cDNA probe mouse *Slc23a1*. The pre- hybridization, hybridization and washes were done with Church buffer and Wash II buffer as recommended by Research Genetics. The probe was labeled with the kit Ready To Go (Pharmacia).

2.4. Shotgun library

The BAC 53L21 was completely digested with *Bam*HI or *Eco*RI and subcloned respectively into the *Bam*HI site or the *Eco*RI site of pBluescript, dephosphorylated with shrimp alkaline phosphatase (USB) 0.1 U/1.0 pmol of DNA. The library was then screened with the 1.9-kb cDNA probe.

2.5. Determination of intron-exon boundaries

Intron sequences were amplified using the mouse genomic clone BAC 53L21, primers forward and reverse across the coding region sequence, and the Long Expand Template PCR System (Boehringer Mannheim). The intron boundaries were determined by comparison with the cDNA sequence.

2.6. Sequence and Genomic analysis

DNA sequencing was performed by Seqwright (Houston, Texas). Nucleotide sequences were compared to the full Gen Bank database using the BLAST algorithm[14] at the website http://www.ncbi.nlm.nih.gov/BLAST/. DNA sequence and amino acid sequence manipulations and primer designs were performed with the LASERGENE software package (DNASTAR Inc). Nucleic acid alignments were carried out using the MACAW (Multiple Alignment Construction and Analysis Workbench) using the segment pair overlap method.[15] Statistical significance was assessed under the null hypothesis of random alignment under a search space defined as the lengths of the actual sequences when the pairwise score exceeded 40.

Mouse Genome Database (Jackson Lab) was the source for mouse mapping data.[16]

2.7. GenBank accession numbers

The GenBank accession numbers related to these genes are: mouse SVCT2 (AB038145), rat SVCT2 (NM017316, AF080453), human SVCT2 (AF164142, AJ269478), human SLC23A1 (AF092511, NM005116), human YSPL2 (AF058319), human KIAA0238 (D87075), human SLC23A2 (NM005847), human SVCT1 (AF170911, AJ269477), human YSPL3 (AF058317), mouse *Slc23a2* (NM11397), mouse *Yspl3* (058318), rat SVCT1 (NM017315, AF080452).

Table 1. *The orthologous relationships between the two paralogous transporter genes among these three species.*

Type 1 Transporter

Species	Gene Name	NCBI Locus ID	Aliases	Map location
Human	SLC23A2	9963	SVCT1 YSPL3	5q31.2-31.3
Mouse	Slc23a2	20522	Yspl3	Chr 18, 18cM
Rat	SVCT1	50621	-	-

Type 2 Transporter

Species	Gene Name	NCBI Locus ID	Aliases	Map location
Human	SLC23A1	9962	SVCT2 YSPL2	20p13
Mouse	Slc23a1	54338	Svct2	Chr 2qG2 (this work)
Rat	SVCT2	50622	-	-

2.8. *Fluorescence* In Situ *Hybridization*

Slides with chromosome metaphase spreads (mouse spleen) were incubated for 1 hr at 37°C in 2 × SSC (0.3 M NaCl and 0.3 M sodium citrate) and then dehydrated sequentially in 70%, 80%, and 90% ethanol. Chromosomal DNA was denatured in 70% formamide, 2 × SSC, for 2 min at 72°C, followed by dehydration in serial ethanol washes of 70%, 80%, 90%, and 100%. Fluorescence *in situ* hybridization (FISH) was performed with the BAC 53L21 labeled with Spectrum Orange-dUTP (Vysis), essentially as described by Pinkel and Lichter.[17,18] On each slide, 100 ng of labeled DNA was applied. Non-unique and nonspecific DNA hybridization was blocked by pre-annealing the probes with a ten-fold excess of Mouse *Cot-1* DNA. Labeled and blocking DNAs were denatured at 75°C for 10 min and then pre-annealed at 37°C for 15 min. The hybridization mixture contained labeled DNA in 10 μl of 50% formamide, 2 × SSC, and 10% dextran sulfate at pH 7. Slides were hybridized overnight at 37°C. Post-hybridization washes were performed at 45°C as follows: 1) 50% formamide, 2 × SSC, 20 min, 2) 1 × SSC, 10 min, and 3) 0.1 × SSC, 10 min. Slides were counterstained with propidium iodide-antifade (Oncor) or DAPI, 250 ng/μl (Boehringer Mannheim) with antifade.

3. Results

Sequence comparison of multiple Genbank entries and expressed sequence tags (ESTs) from human, rat and mouse revealed that cDNAs for both of the sodium-dependent ascorbic acid transporter paralogs have been cloned from all of three species, but with conflicting and inconsistent nomenclature. For clarity, in Table 1 we list the orthologous relationships between the two paralogous transporter genes among these three species.

Primers for the 5′ and 3′ end of the coding region of the rat SVCT2 gene were chosen from regions of similarity between human and rat, and used in PCR to isolate one 1.9-kb fragment from mouse brain cDNA. This fragment was cloned and sequenced. 5′ RACE in mouse brain cDNA identified an additional 300-bp fragment containing the rest of the coding sequence of *Slc23a1* plus 202 bp of the 5′ UTR. When assembled, these sequences contained the complete coding region, encoding 647 amino acids of the mouse *Slc23a1* gene.

We screened the RPCI22 mouse BAC library with the 1.9-kb cDNA fragment as a probe in order to isolate *Slc23a1* genomic DNA. Eleven positive clones were identified by filter hybridization. The presence of the *Slc23a1* gene was confirmed in three clones (53L21, 77J14 and 125H10) by Southern blot hybridization, using specific probes in the 5′ and 3′ ends of the coding region. One of them, 53L21, has an estimated insert size of 145 kb and was chosen for further studies. PCR from 53L21 allowed the isolation of a 4172-bp fragment of 3′ UTR transcribed within a single exon. Assembly of the 3′ UTR with the rest of the sequence generated 6346 bp of cDNA representing a nearly full-length mouse *Slc23a1* cDNA (GenBank Accession # AY004874). At the level of nucleic acid sequence, a comparison of the mouse, human, and rat orthologues of *Slc23a1* revealed very strong similarity in the coding region (Fig. 2). The 5′ UTR of the mouse gene also showed strong similarity to the rat 5′ UTR sequence (AF080453) but not to the human SLC23A1 sequence AF164142.

In contrast, the 3′ UTR of mouse and human (D87075) contained numerous segments of strong similarity. The relative similarity of protein sequence among the rat, human and mouse sodium-dependent vitamin C transporters is presented in Fig. 3 as a dendrogram of protein-coding sequences performed using Megalign (DNASTAR Inc).

To determine exon-intron organization, we constructed

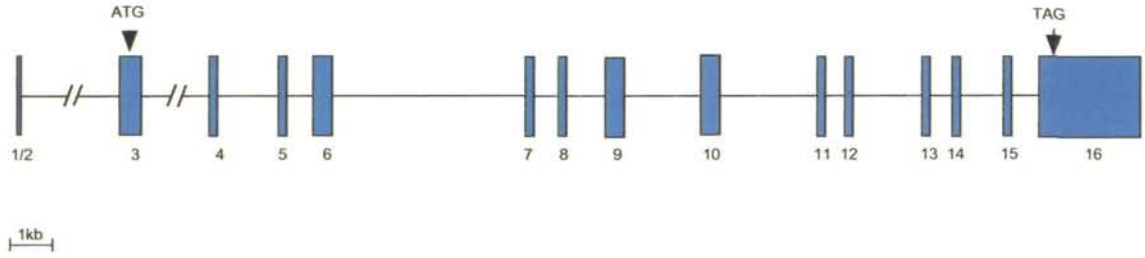

Figure 1. Schematic representation of the relative exon-intron size of the mouse *Slc23a1* gene. Shaded boxes represent exons.

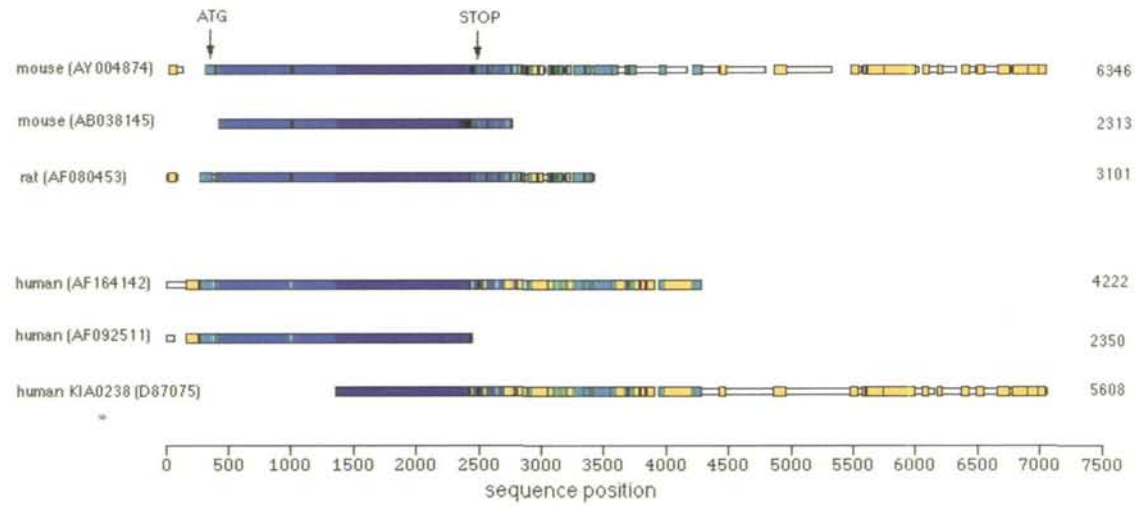

Figure 2. Schematic diagram of MACAW alignment of nucleic acid sequence of mouse *Slc23a1* cDNA generated in this report as compared with other mouse, rat and human sequences (shown with their accession numbers). Regions of significant sequence similarity are shown as shaded rectangles. Similarity between two sequences is shown as shading in light yellow, with the shading of the rectangles becoming increasingly dark with greater number of sequences that share a region of similarity.

a library of short genomic DNA fragments from BAC 53L21 into plasmid pBlueScript. The clones containing *Slc23a1* were isolated with the 1.9-kb cDNA clone as a probe. First, we sequenced the different clones with T3 and T7 primers to identify any exon sequence present within 500–600 bp of the end of the clone. In a second step, the BAC was sequenced directly with forward and reverse primers randomly distributed across the 1.9-kb cDNA fragment. The precise location of exon/intron boundaries was determined by aligning the cDNA sequence and the sequences of the genomic fragments. We delineated 15 coding exons, beginning with exon3. During preparation of this manuscript, "the first draft" sequence of the sequence of human BAC RP11 27C24 (Accession # AC068582) became available. Comparison with human cDNA for SLC23A1 (Accession # AF092511) revealed 7 exons; the first human coding exon was exon 3. We adopted this numbering system for the mouse gene as well, although we have not confirmed that the human and mouse genes have the same 5′ exon/intron structure since it is unlikely that we have the complete 5′ UTR of the mouse gene. Exons ranged in size from 58 bp (exon 1) to 4.4 kb (exon 16) (Table 2). All the splice junctions conformed to the "GT-AG" consensus sequence.[19] We assume that the site of translation initiation is the ATG at position 209, in exon 3, since it is the first one compatible with a Kozak consensus sequence and sequence homology between mouse and human cDNA sequence becomes very high at this point. Distances between exons were estimated by electrophoresis of PCR products from the BAC 53L21 with primers from adjacent exons (Table 2).

To determine the pattern of expression of the *Slc23a1* gene in mouse tissue, the 1.9-kb cDNA fragment was used as a probe on a mouse multiple tissue Northern blot. An approximately 7-kb transcript was identified, with expression in all the tissues represented except in skeletal muscle where no expression was detectable and high expression levels were observed in brain (Fig. 4).

We mapped the mouse *Slc23a1* gene using BAC 53L21

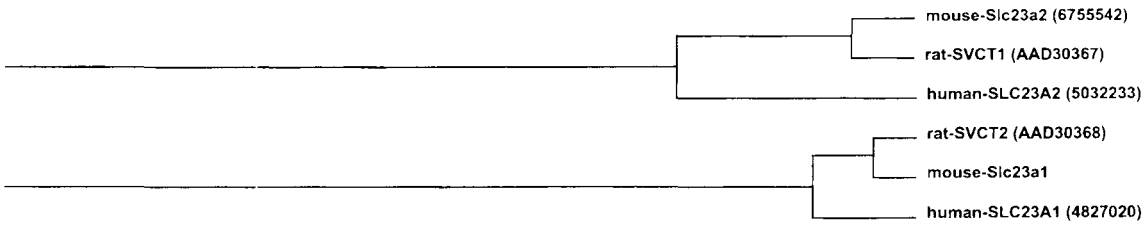

Figure 3. Phylogenetic tree of the sodium dependent vitamin C transporters SVCT1 and SLC23A1 in human, rat and mouse (based on amino acid sequence). The numbers in parenthesis are the accession numbers in GenBank.

Table 2. Exon/intron organization of the mouse *Slc23a1* gene.

exon	exon length (bp)	acceptor site	donor site	intron length (kb)
1/2	58 5′ UTR	CGGCG	TGCAG	n.d.
3	255	tgc**ag**GTGAT	TCCCG**gt**aag	> 4
4	98	ttc**ag**GTGGT	AGAAG**gt**acc	~ 3
5	116	tcc**ag**AGCTC	TGCAG**gt**gag	~ 1.5
6	157	agc**ag**CACTA	TGCAG**gt**gag	~ 10
7	88	ttt**ag**GTTAC	CACAG**gt**aat	~ 1.3
8	64	tca**ag**AGATT	AAGAG**gt**aaa	~ 2
9	181	ttc**ag**ATCCA	ATGCT**gt**aag	~ 4.5
10	121	ctc**ag**GACGA	TTCCT**gt**gag	~ 5
11	157	tcc**ag**ATAAT	CCCAT**gt**aag	~ 0.5
12	147	ctc**ag**TTCAG	AACAG**gt**ctg	~ 1.4
13	105	ttt**ag**GGGTA	CAAAG**gt**ata	~ 3.5
14	129	cac**ag**GTCGG	CTTTG**gt**aag	~ 1.6
15	137	tcc**ag**GAATG	TACAG**gt**agg	~ 2.5
16	95	tcc**ag**GAATA	CCCAG**gt**atg	~ 1.6
17	235 coding + 4172 3′ UTR	agc**ag**GTACC	TATAG AATAAA polyadenylation	

n.d: not determined

as a probe for FISH to mouse chromosomes. We localized the *Slc23a1* gene to the murine chromosome 2qG2 (Fig. 5). This band, located at approximately 70–75 cM on MMU2, contains many genes whose human orthologs map to 20p13; the mouse *Slc23a1* gene represents further conservation of this syntenic region.

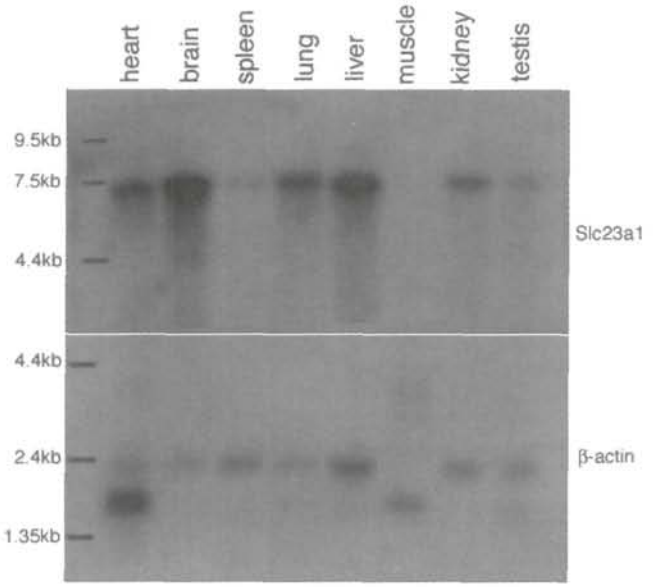

Figure 4. Tissue distribution of mouse *Slc23a1* expression. Northern blot analysis of mRNA from a Clontech multiple tissue blot was performed using a radiolabeled 1.9-kb cDNA probe. The blot was also probed with β-actin cDNA as a control for the amount of mRNA loaded in each line.

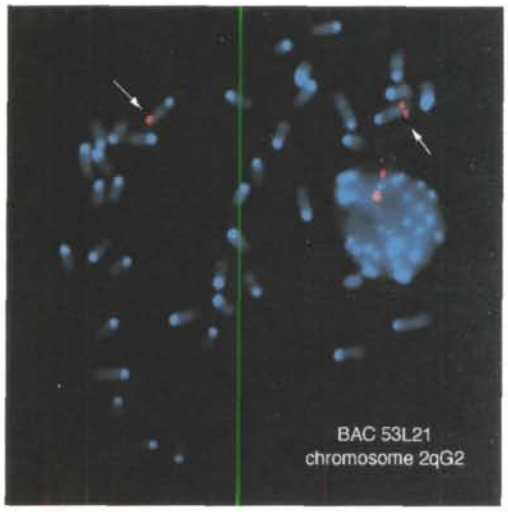

Figure 5. Chromosomal mapping of the *Slc23a1* gene by fluorescence *in situ* hybridization (FISH) to normal metaphases from mouse spleen. The probe was the mouse BAC 53L21.

4. Discussion

We describe here the sequence, genomic structure and expression of the mouse sodium-dependent vitamin C transporter gene type 2. This gene shows high homology with the orthologous gene in human and rat with 94.9% and 98.7% similarity at the protein level.

We assume that the initiation codon is the ATG at position 202, since this sequence atgATGg resembles best the Kozak consensus sequence with an A in position -3 and a G in position $+4$.[20] This assumption contrasts with the interpretation of the rat sequence by Tsukaguchi et al. who assumed the initiation codon to be at position 367 in spite of this ATG being neither the best (no purine in position -3) nor the first possible initiation site. Our assumption also contrasts with the interpretation of the human sequence, where the initiation codon is thought to be the ATG immediately before the one in position 202,[7,9–11] where no purine in position -3 is present or a G in position $+4$, resulting in a very weak Kozak consensus sequence. Definitive identification of the start codon awaits amino acid sequence analysis of the protein.

The length of the complete cDNA sequence of 6324 bp without the polyA tail corresponds well to the unique band of approximately 7 kb observed on Northern blots of all tissues analyzed, with exception of skeletal muscle where no transcript was observed.

The approximately 6.3 kb are distributed in 208 bp for the 5′ UTR, 1944 for the coding region, and a long 3′ UTR of 4172 bp. The comparison between the human and the mouse gene showed only 34.8% similarity in the 5′ UTR region of both genes. The comparison between the human EST clone KIA0238 and the mouse 3′ UTR shows regions of high homology interrupted by gaps with low or no homology, for an average of 37%. Similarity between the mouse and the human 5′ and 3′ UTRs are both low as compared to the average interspecies conservation of 70% observed by Makalowski and Boguski[21] (Fig. 5).

Another interesting feature was the expression pattern. In mouse, rat and human, one approx. 7-kb transcript was observed in most of the tissues. The expression in brain is consistently high in all three species: mouse, rat and human. Subtle differences in tissue distribution of expression where observed, with liver showing high levels of expression in humans and the mouse, but not in the rat. The lung, in contrast, showed no expression in humans but good expression in the mouse and rat. The presence of high concentrations of ascorbic acid in brain and the suggested role of this vitamin as a radical scavenger in the organism make SLC23A1 a relevant candidate gene to be investigated in human neurodegenerative disorders where oxidative stress appears to play an important role.

Acknowledgements: This work was supported by the Division of Intramural Research, NHGRI and NINDS.

References

1. Padh, H. 1991, Vitamin C: Newer insights into its biochemical functions, *Nutr. Rev.*, **49**, 65–70.
2. Rice, M. E. 2000, Ascorbate regulation and its neuroprotective role in the brain, *Trends Neurosci.*, **23**, 209–216.
3. Englard, S. and Seifter, S. 1986, The biochemical functions of ascorbic acid, *Annu. Rev. Nutr.*, **6**, 365–406.
4. Dhariwal, K., Hartzell, W., and Levine, M. 1991, Ascorbic acid and dehydroascorbic acid measurements in human plasma and serum, *Am. J. Clin. Nutr.*, **54**, 712–716.
5. Agus, D. B., Gambhir, S. S., Pardridge, W. M., Spielholz, C., Baselga, J., Vera, J. C., and Golde, D. W. 1997, Vitamin C crosses the blood-brain barrier in the oxidized form through the glucose transporters, *J. Clin. Invest.*, **100**, 2842–2848.
6. Rumsey, S. C., Kwon, O., Xu, G. W., Burant, C. F., Simpson, I., and Levine, M. 1997, Glucose transporter isoforms GLUT1 and GLUT3 transport dehydroascorbic acid, *J. Biol. Chem.*, **272**, 18982–18989.
7. Hogue, D. L. and Ling, V. 1999, A human nucleobase transporter-like cDNA (SLC23A1): member of a transporter family conserved from bacteria to mammals, *Genomics*, **59**, 18–23.
8. Tsukaguchi, H., Tokui, T., Mackenzie, B., Berger, U. V., Chen, X. Z., Wang, Y., Brubaker, R. F., and Hediger, M. A. 1999, A family of mammalian Na^+-dependent L-ascorbic acid transporters, *Nature*, **399**, 70–75.
9. Daruwala, R., Song, J., Koh, W. S., Rumsey, S. C., and Levine, M. 1999, Cloning and functional characterization of the human sodium-dependent vitamin C transporters hSVCT1 and hSVCT2, *FEBS Lett.*, **460**, 480–484.
10. Rajan, D. P., Huang, W., Dutta, B., Devoe, L. D., Leibach, F. H., Ganapathy, V., and Prasad, P. D. 1999, Human placental sodium-dependent vitamin C transporter (SVCT2): molecular cloning and transport function, *Biochem. Biophys. Res. Commun.*, **262**, 762–768.
11. Faaland, C. A., Race, J. E., Ricken, G., Warner, F. J., Williams, W. J., and Holtzman, E. J. 1998, Molecular characterization of two novel transporters from human and mouse kidney and from LLC-PK1 cells reveals a novel conserved family that is homologous to bacterial and Aspergillus nucleobase transporters, *Biochim. Biophys. Acta*, **1442**, 353–360.
12. Wang, Y., Mackenzie, B., Tsukaguchi, H., Weremowicz, S., Morton, C. C., and Hediger, M. A. 2000, Human vitamin C (L-ascorbic acid) transporter SVCT1, *Biochem. Biophys. Res. Commun.*, **267**, 488–494.
13. Nagase, T., Seki, N., Ishikawa, K., Ohira, M., Kawarabayasi, Y., Ohara, O., Tanaka, A., Kotani, H., Miyajima, N., and Nomura, N. 1996, Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain, *DNA Res.*, **3**, 321–329, 341–354.
14. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, **25**, 3389–3402.
15. Schuler, G. D., Altschul, S. F., and Lipman, D. J. 1991, A Workbench for Multiple Alignment Construction and Analysis, *Genetics*, **9**, 180–190.
16. Blake, J. A., Eppig, J. T., Richardson, J. E., and Davisson, M. T. 2000, The Mouse Genome Database (MGD): expanding genetic and genomic resources for the laboratory mouse. The Mouse Genome Database Group, *Nucleic Acids Res.*, **28**, 108–111.
17. Pinkel, D., Straume, T., and Gray, J. W. 1986, Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization, *Proc. Natl. Acad. Sci. USA*, **83**, 2934–2938.
18. Lichter, P., Cremer, T., Borden, J., Manuelidis, L., and Ward, D. C. 1988, Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries, *Hum. Genet.*, **80**, 224–234.
19. Shapiro, M. B. and Senapathy, P. 1987, RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression, *Nucleic Acids Res.*, **15**, 7155–7174.
20. Kozak, M. 1996, Interpreting cDNA sequences: some insights from studies on translation, *Mamm. Genome*, **7**, 563–574.
21. Makalowski, W. and Boguski, M. S. 1998, Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences, *Proc. Natl. Acad. Sci. USA*, **95**, 9407–9412.