Hum Genet (2000) 106 : 627–635
Digital Object Identifier (DOI) 10.1007/s004390000309

ORIGINAL INVESTIGATION

Dirk Gründemann · Edgar Schömig

Gene structures of the human non-neuronal monoamine transporters EMT and OCT2

Received: 17 February 2000 / Accepted: 5 April 2000 / Published online: 19 May 2000

Abstract Non-neuronal monoamine transporters OCT1, OCT2, and EMT, which are all members of the amphiphilic solute facilitator family, control signal transmission by removing released transmitters, such as dopamine, noradrenaline, adrenaline, 5-hydroxytryptamine, and histamine, from the extracellular space. In the current study, we have isolated human *EMT* (gene symbol *SLC22A3*) and *OCT2* (*SLC22A2*) genes and report the gene and promoter organization. Both genes consist of 11 coding exons, with consensus GT/AG splice sites and conserved intron locations. The *EMT* gene is 77 kb, and the *OCT2* gene is 45 kb in size. For the *EMT* gene, two transcription start points were identified by inverse polymerase chain reaction based on mRNA from Caki-1 cells. The *EMT* promoter, located within a CpG island, lacks a consensus TATA box but contains a prototypical initiator element and a number of potential binding sites for ubiquitous transcription factors Sp1 and NF-1. In contrast, the *OCT2* promoter is not associated with a CpG island, contains a putative TATA box, and potential binding sites for specific transcription factors, such as HFH-8 and IK2. Since EMT and OCT2 may play important roles in catecholamine homeostasis and, as such, are candidate genes in human disease, the present results provide a basis for the analysis of genetic variation and the regulation of transcription.

h, r, and m in conjunction with a protein name designate the species as being human, rat, or mouse, respectively; human gene symbols: OCT2: *SLC22A2*, EMT: *SLC22A3*, OCTN2: *SLC22A5*, OAT2: *SLC22A7*.

D. Gründemann (✉) · E. Schömig
Department of Pharmacology, University of Heidelberg,
Im Neuenheimer Feld 366, 69120 Heidelberg, Germany
e-mail: dirk.gruendemann@urz.uni-heidelberg.de,
Tel.: +49 6221 548606, Fax: +49 6221 548607

Introduction

Since released monoamine transmitters, such as dopamine, noradrenaline, adrenaline, and 5-hydroxytryptamine, do not easily cross plasma membranes, their inactivation depends on integral membrane transport proteins. Initial functional studies with noradrenaline originally uncovered two distinct transport mechanisms, uptake$_1$ and uptake$_2$ (Iversen 1965), which, by reference to their predominant localization, were later termed neuronal and non-neuronal (extraneuronal). These transport systems differ vastly with respect to transport mechanism, drug sensitivity, substrate affinity, and substrate selectivity. Continuous scrutiny, intensified by molecular cloning about a decade ago, has generated a thorough understanding of the neuronal transporters, but much less is known about the physiological and pathophysiological significance of non-neuronal monoamine transport. Albeit, non-neuronal uptake mechanisms play a key role for terminating the actions of released catecholamines in the circulation, because pharmacological blockade dramatically increases plasma levels of adrenaline and noradrenaline in the conscious rat (Eisenhofer et al. 1996). Non-neuronal transport may also, via glia cells, contribute to the inactivation of centrally released monoamine transmitters (Russ et al. 1996).

Molecular cloning has recently revealed that not one but three transport proteins may participate in non-neuronal monoamine uptake, i.e., OCT1 (Gründemann et al. 1994; Nagel et al. 1997; Breidert et al. 1998), OCT2 (Okuda et al. 1996; Gründemann et al. 1998b; Busch et al. 1998), and EMT (Gründemann et al. 1998a; Kekuda et al. 1998). All three are members of the amphiphilic solute facilitator (ASF) family of transport proteins (Schömig et al. 1998) and accept catecholamines and neurotoxin 1-methyl-4-phenylpyridinium as substrates.

Whereas the extraneuronal monoamine transporter EMT corresponds to the classical catecholamine uptake$_2$ mechanism, the organic cation transporters OCT1 and OCT2 were first identified functionally by their affinity

for various positively charged organic solutes, i.e., xenobiotics, metabolites, and drugs. EMT is expressed in many tissues, but OCT2 has been detected solely in kidney and brain, and OCT1 is confined to liver, kidney, and intestine. Inhibitors and substrates are available to discriminate these evolutionary related carriers functionally (Russ et al. 1993; Gründemann et al. 1999). However, previous pharmacological characterizations have also revealed that, in general, EMT closely resembles OCT2, but not OCT1 (Schömig et al. 1993).

EMT and *OCT2* are candidate genes in human disease. Diminished EMT function with the resultant increase of extracellular monoamine levels could affect the sympathetic nervous system, contributing, for example, via catecholamine-stimulated glycogenolysis, to elevated blood glucose levels. Indeed, using fluorescence in situ hybridization, we have mapped the *EMTh* gene to 6q27 (Gründemann et al. 1998a), the region to which susceptibility loci for type I diabetes have been mapped (Luo et al. 1995). In close physical linkage, the genes of OCT1 and OCT2 are located at 6q26 in human (Koehler et al. 1997). The approximate arrangement of this cluster has been explored for the mouse and human genes (Verhaagh et al. 1999). OCT2 may be involved in dopamine handling and thus contribute to the control of renal sodium excretion (Gründemann et al. 1998b). As a prerequisite for the analysis of genetic variation and transcriptional regulation, it has been the aim of the current study to isolate the human genes of OCT2 and EMT and to characterize their organization.

Materials and methods

Basic molecular biology

The polymerase chain reaction (PCR), reverse transcription/PCR (RT-PCR), cloning of PCR fragments, sequencing, and preparation of total RNA and mRNA were performed as described (Gründemann et al. 1997; Schömig et al. 1998). P1 artificial chromosome (PAC) DNA was prepared by alkaline lysis. Long distance PCR (Barnes 1994) was performed with TaKaRa LA *Taq* mixture (BioWhittaker, Taufkirchen, Germany).

Southern blot analysis

Genomic DNA fragments were fractionated by gel electrophoresis in 0.5% agarose in TAE buffer (TAE buffer = 40 mM TRIS-acetic acid pH8.0, 1 mM EDTA). After downward alkaline capillary transfer (Chomczynski 1992) to Hybond-N membrane (Amersham Pharmacia, Freiburg, Germany) and UV fixation, the membrane was prehybridized for 1 h at 42°C in solution A, consisting of 6×SSPE (1× SSPE = 150 mM NaCl, 10 mM Na_2HPO_4-NaOH pH 7.4, 1 mM EDTA), 50% formamide, and 0.5% SDS, plus 5×Denhardt solution (1×Denhardt = 0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone, 0.02% Ficoll) and 0.1 mg/ml denatured herring sperm DNA. For overnight hybridization, random-primed ^{35}S-labeled double-stranded cDNA probes were used in solution A plus 5 mM dithiothreitol (DTT). Membranes were washed to a final stringency of 0.2×SSPE, 0.1% SDS at 60°C. Radioactivity was detected by radioluminography with the BAS1000 system (Fuji Photo Film, Tokyo, Japan).

Determination of gene structures

EMTh probes corresponded to positions 49–294 and 1402–1589 of the EMTh cDNA sequence (Gründemann et al. 1998a; GenBank/EBI Data Bank accession no. AJ001417). *OCT2h* probes corresponding to positions 71–549 and 1877–2155 of the OCT2h cDNA sequence (Gorboulev et al. 1997; accession no. X98333) were prepared by RT-PCR on human kidney total RNA (Clontech, Heidelberg, Germany). Probes were used to screen human PAC library no. 704 (Resource Center of the German Human Genome Project, Max-Planck-Institute for Molecular Genetics, Berlin, Germany), which consists of RPCI libraries 1, 3, 4, and 5 (Roswell Park Cancer Institute, Buffalo, New York; Ioannou et al. 1994). Positive clones were obtained from the Resource Center.

For *EMTh*, PAC clones B041200Q4, H17291Q13, and I03129Q13 were used to sequence exon 1 and its upstream region, exon 2, and exons 3–11, respectively. For *OCT2h*, clones L23786Q2 and N101138Q2 were used to sequence exon 1 and its upstream region, and exons 2–11, respectively.

Exon-containing genomic fragments were identified by restriction mapping and Southern analysis with cDNA probes and were subcloned into pUC19, pOCUS-2 (Novagen, Schwalbach, Germany), or pOX, a pOCUS-2 derivative with an extended multiple cloning site, and electroporated into *Escherichia coli* DH10B. Exons, intron/exon boundaries, and the region upstream of the translation start point were sequenced with specific or universal primers as described (Schömig et al. 1998) or by a transposon-facilitated technique (Strathmann et al. 1991) with *E. coli* strains γδ DONOR and γδ RECIPIENT (Novagen). Intron sizes were determined by restriction mapping or by long-distance PCR as indicated in Table 1. Where feasible, distances measured by PCR were confirmed by additional PCRs with primer pairs that covered two to four adjacent introns.

Determination of transcription start point

In approach I, first-strand cDNA was synthesized as described (Gründemann et al. 1997) with primer RAEMTh (5'-GGT TTC AGG CAA AAG CAT) on 1 μg mRNA from human cell line Caki-1. Following purification by phenol/chloroform extraction and ethanol precipitation, second-strand cDNA was synthesized (72°C, 10 min) with thermostable *Pwo* DNA polymerase (Peqlab, Erlangen, Germany). Products were purified on a silica spin column (QIAquick PCR purification kit, Qiagen, Hilden, Germany) and circularized by incubation with T4 DNA ligase and T4 polynucleotide kinase (Fermentas, St. Leon-Rot, Germany). Ligation products were purified by extraction, precipitated with ethanol in the presence of linear polyacrylamide, and dissolved in 8 μl TE (TE = 10 mM TRIS-HCl pH 8.0, 1 mM EDTA). Inverse PCR was performed (34 cycles: 94°C for 30 s, 58°C for 1 min, 72°C for 1 min) with 1 μl ligation product (negative control: water) and primers FXEMTh (5'-CCT CCC CTT TGC GGC AAG CAA T) and RXEMTh (5'-GCC CGC CGC CCG CA). The PCR mix (50 μl) contained 5% dimethylsulfoxide (DMSO) and 1 M formamide in addition to the standard components (Gründemann et al. 1997). After PCR, 10 μl of the reaction mix was analyzed by agarose gel electrophoresis.

In approach II, first-strand cDNA synthesized as above was purified on a silica spin column, with 50 μl water as eluent. RNA was degraded (65°C, 30 min) in 0.2 M NaOH and 1 mM EDTA (total volume: 100 μl). After neutralization with 20 μl 3 M sodium acetate pH 5.2 and 100 μl 0.2 M HCl, first-strand cDNA was precipitated with ethanol in the presence of linear polyacrylamide. 5'-Ends were phosphorylated by incubation with 25 U T4 polynucleotide kinase (T4 DNA ligase buffer; 50 μl, 37°C, 2 h). After purification by extraction and precipitation, the single-stranded cDNA (20% of the generated material) was circularized by incubation with 10 U T4 RNA ligase (Fermentas; 50 mM triethanolamine-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 10 μg/ml bovine serum albumin, 10 μM ATP; 100 μl, 20°C, 48 h). Ligation products were purified by extraction, precipitated with ethanol in the

Table 1 Organization of human *EMT* and *OCT2* genes. In the splice site nucleotide sequences, *upper-case letters* indicate exons, and *lower-case letters* indicate introns. Exon/intron boundaries were determined by comparison of sequences of genomic DNA and cDNA. The following deviations from cDNA were found with genomic DNA (with reference to published cDNA sequences; notation: cDNA nucleotide/position/genomic DNA nucleotide): EMTh: T/387/C, G/969/A; OCT2h: T/447/C, G/952/T, A/1650/G, G/2121/T, insertion of T behind position 2122, N/2159/C, insertion of A behind position 2166, N/2231/C. Database entries from the HTGS (unfinished High Throughput Genomic Sequences) division of GenBank partly cover the *EMTh* gene (AC010893, AL109933) and the *OCT2h* gene (AC022322, AC024253)

Exon no.	3'-Splice acceptor site	Exon size (b)	5'-Splice donor site	Intron size (kb)
EMTh gene				
1		>456[d]	CAGCGAG gtaagggcgccccggccctt	21.4[a]
2	tgttttctttgttttctcag TTTGACC	104	CAGACAG gtaggtaccaatgtgacagc	9.6
3	ttcctgtctttctcaaaaag GTATGGC	155	GTGATTG gtaagacattcttacaccat	1.6
4	ttgtgtttgttttcctttag TGACAGA	169	ATTACTG gtaatgtggttttggttatc	1.8
5	acagcctccttctttgccag GGTGGTC	118	CTCAGAG gtaatttctttcagtatgag	26.1[a]
6	aatctgcttttcttatatag ATCACTG	98	TTGCTTG gtaagtttgacttgtgatgg	0.1
7	gtctttcaatgttgctacag GTTCACA	215	CCAGAAG gtaatcttaccccacatctg	5.2
8	caaactttctgtgtttgcag GAATAGC	109	CATTACG gtaattctaacaaacgttat	0.8
9	aattacttttcattcaacag AAATTTC	113	ATCTTTG gtaagaactcatttgctatt	3.9
10	acctttcccctattccatag GTATCCT	100	TTGGCAG gtactgtacaaaattcaatg	3.1
11	tttctcctttgtttttcag TCCACAT	>770[c]		
OCT2h gene				
1		>558[d]	CACCGAG gtaagagagtgagctcaggg	1.6[a,b]
2	cacttctccccagctggcag TTTAACC	104	CAGACAG gtaggttgaatcacctgtgg	5.9[a,b]
3	ctttgtttttcttcctgcag GTTTGGC	155	ATCCTGA gtaagaatgtttgtgcttgc	1.2
4	ttcaaattctctctctgcag TTACAGA	169	ATTACTG gtaagtccatttggaataaa	2.0
5	ccgctcaccttgtaccctag GTGCATA	115	CCTTCAG gtgagcagggccttaagtat	1.8[a,b]
6	ctttctattccttttcttag CGCCTGA	107	ACAACTG gtaaggaatatttttcactt	1.6[a,b]
7	gaactctcctctttgctcag GTTCACG	215	CCTGGTG gtaagtttcaggtgaagttg	1.2
8	actgttcttttccctcttag ATCTACA	109	TCATTAG gtgagtgcattttccttaga	0.7
9	atcagtttcctgcattctag GAATCTT	113	GTTTTCG gtaagaaatcttcacagatg	16[a]
10	ttcatctgtttggctttcag GCGTGCT	100	TGCAAAG gtaggcagttgaaagttcta	11[a]
11	atttttatttcttattgcag ACCAAGA	>553[c]		

[a]Determined by long distance PCR
[b]Confirmed by additional PCRs with primer pairs that covered two to four adjacent introns
[c]Sequencing of exon 11 was discontinued at the indicated position
[d]With reference to the 5'-end of published cDNA

presence of 10 µg tRNA, and dissolved in 5 µl TE. Inverse PCR was performed with 1 µl ligation product (negative control: single-stranded cDNA without ligation) as described above. Because of a multitude of amplicons, a nested PCR was performed with primers FAEMTh (5'-AAC ATT ACG AAA TTT CGG A) and RXEMTh (30 cycles: 94°C for 30 s, 50°C for 1 min, 72°C for 1 min, in standard buffer plus DMSO and formamide). The resulting major band at 220 bp was cloned into pUC19, and eight clones were sequenced.

Sequence analysis

Sequences were analyzed with the Wisconsin Package as implemented in HUSAR (DKFZ, Heidelberg, Germany). Pairwise sequence alignments were computed with program GAP with default parameters. Transcriptional elements were identified with program MatInspector (GBF, Braunschweig, Germany), with a core similarity of 0.80 and a matrix similarity of 0.85, based on TRANSFAC release 3.5 (Windenger et al. 1996).

Results

Isolation and organization of the human EMT gene

Screening of a human PAC library with EMTh cDNA probes identified genomic DNA containing *EMT* exons. Three PAC clones were used to analyze the organization of the entire *EMTh* gene. Sequencing of subcloned fragments revealed 11 exons. Exon and intron sizes and intron/exon boundaries are listed in Table 1. All splice donor and acceptor sites are canonical (Senapathy et al. 1990). The amino acid sequence derived from the genomic DNA is identical to the cDNA encoded sequence and thus confirms the latter, which had been obtained in part by PCR methods (Gründemann et al. 1998a). Based on the length of the published cDNA and on intron sizes, the *EMTh* gene is 77 kb long.

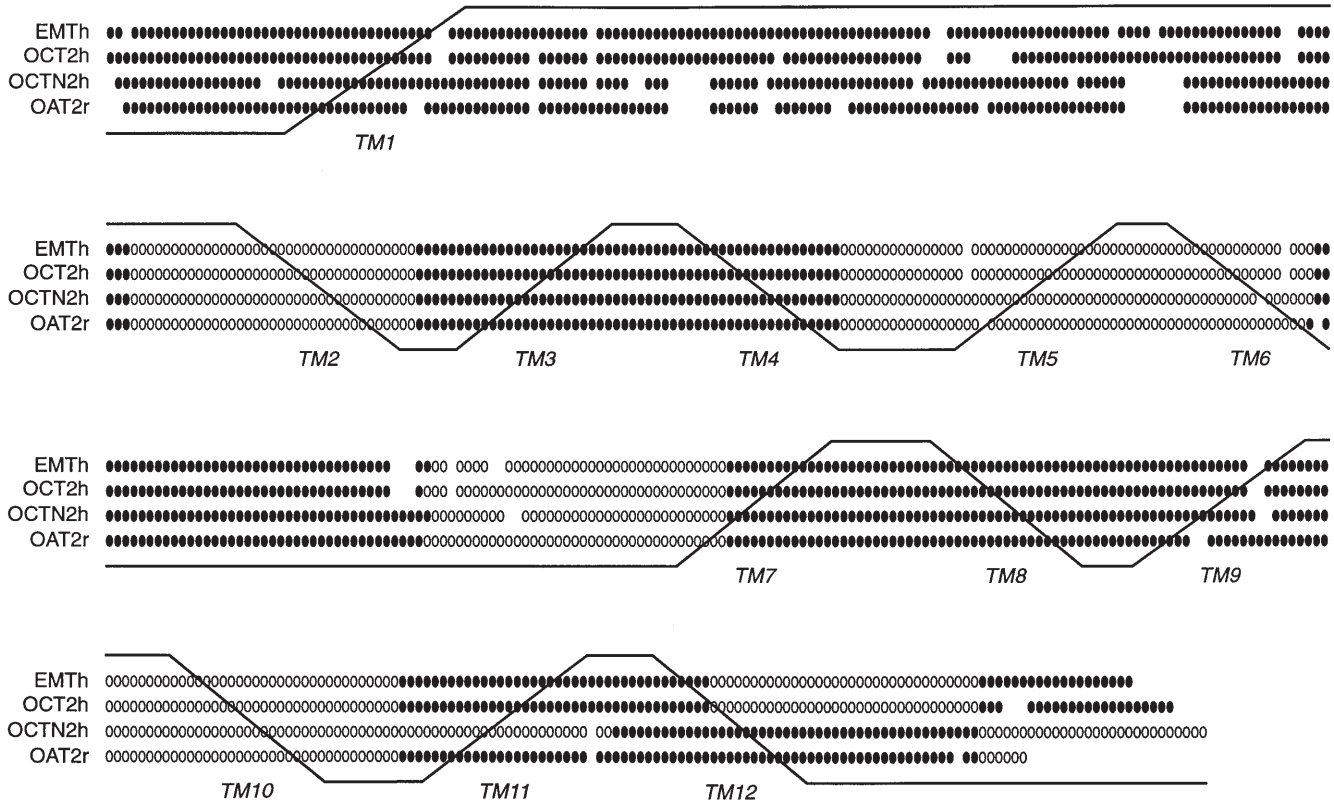

Fig. 1 Comparison of exon junctions of ASF family genes. Amino acid sequences of EMTh, OCT2h, OCTN2h, and OAT2r were aligned by program PILEUP. Each *open* or *filled ellipse* represents a single amino acid, and *spaces* indicate alignment gaps. *Runs* of *filled* or *open ellipses* mark exons. The superimposed line diagram denotes the proposed membrane topology model common to all four carriers (Schömig et al. 1998). *Horizontal line segments above* the amino acid alignment correspond to the extracellular membrane face. Transmembrane segments, denoted by *TM1–TM12*, are marked by *diagonal lines*

Isolation and organization of the human OCT2 gene

Screening of the human genomic PAC library with OCT2h cDNA fragments yielded two confirmed positive clones, which were shown by hybridization and PCR to cover fully the 2.2-kb OCT2h cDNA reported previously (Gorboulev et al. 1997). The human *OCT2* gene consists of 11 exons spread over about 45 kb (Table 1). Intron sizes were determined by restriction analysis of subclones or by PCR with the original PAC clones as templates. Again, all splice sites conform to the GT/AG rule. The amino acid sequence derived from the genomic DNA is identical to the cDNA-encoded sequence, except for position 270 (Ala versus Ser; Gorboulev et al. 1997). As detailed in the legend of Table 1, the genomic sequence also clarifies some ambiguities within the published cDNA.

Comparison of ASF family gene organization

So far, complete gene structures have been reported only for two other members of the ASF family, i.e., rat OAT2 (Simonson and Iwanij 1995) and human OCTN2 (Wu et al. 1998). Both genes consist of 10 exons. Figure 1 shows an amino acid alignment of EMTh, OCT2h, OCTN2 h, and OAT2r, with alternating symbols to delineate exons (superimposed is a topology model based on hydropathy analysis; Schömig et al. 1998). Remarkably, most of the exon boundaries are precisely conserved. Variation is confined to the vicinity of proposed transmembrane segments 11 and 12.

GC content upstream of and around exon 1 of EMTh and OCT2h

The genomic DNA sequence 4.7 kb upstream to 0.6 kb downstream of the translation start point in exon 1 was determined for *EMTh*. Likewise, for *OCT2h*, a 10-kb *Eco*RI genomic fragment containing exon 1 was subcloned, and the region 4.0 kb upstream to 0.7 kb downstream of the translation start point was completely sequenced.

Nucleotide composition analysis of this section revealed that the presumed core promoter and exon 1 of the *EMTh* gene were located within a CpG island, a region with a high GC content and clustering of CpG dinucleotides (Fig. 2). In contrast, the equivalent section of the *OCT2h* gene had an overall substantially reduced GC content and showed no prominent clustering of CpG dinucleotides. GC content peaks were confined to exon 1. Thus, the *OCT2h* gene is not associated with a CpG island.

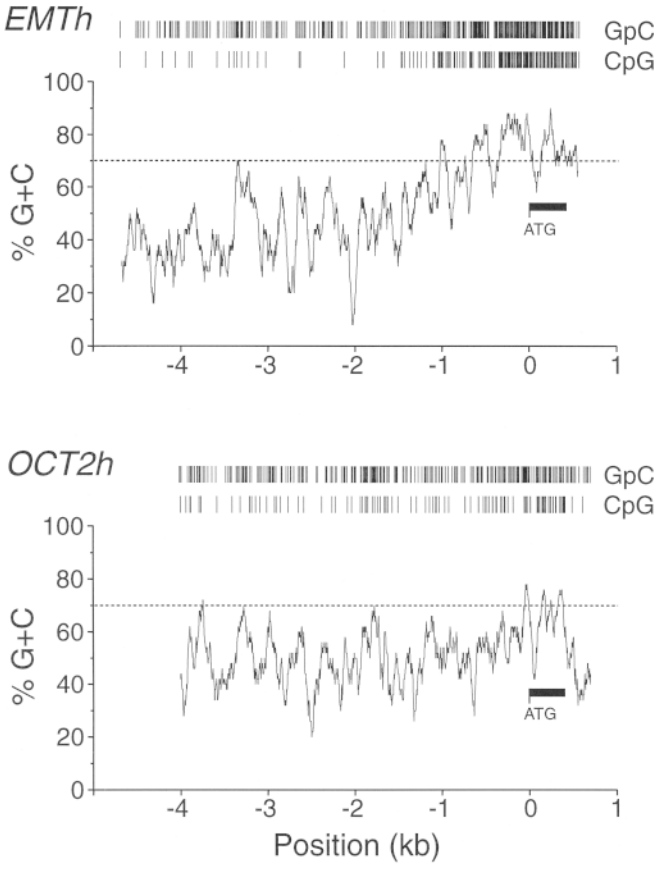

Fig. 2 GC content of human genomic DNA fragments that encompass promoter and exon 1 of *EMT* and *OCT2*. The fractional content of guanosine plus cytidine as a function of sequence position was calculated with a sliding window (width 50 bases, step size 3 bases). For both graphs, position 0 corresponds to the first base of the initiator codon. Occurrences of dinucleotides GpC and CpG are marked by *vertical dashes*. Coding regions are indicated by *solid bars* tagged with *ATG*. *Dashed horizontal lines* at 70% G+C serve to assist visual comparison. The nucleotide sequences have been submitted to GenBank/EBI Data Bank with accession nos. AJ251884 (5279 b *EMTh* fragment) and AJ251885 (4760 b *OCT2h* fragment)

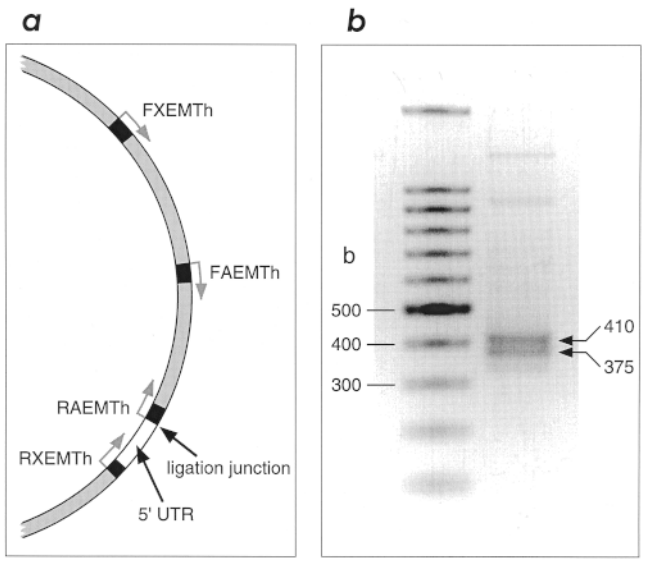

Fig. 3a, b Determination of *EMTh* transcription start point by inverse PCR. **a** Illustration of cDNA ring template prior to inverse PCR with binding sites and orientations of primers. Given a full-length copy of the mRNA, then the outmost base of the 5' untranslated region at the ligation junction corresponds to the transcription start point. **b** Inverted photographic image of ethidium-stained agarose gel with DNA size marker (*left lane*) and products from inverse PCR (*right lane*)

Determination of the transcription start point of the EMTh gene

In order to assign a core promoter, it is essential to determine the transcription start point (tsp). Preferably, experimental results should be corroborated by different methods. For the *EMTh* tsp, two approaches were successfully executed, both based on RT with an EMTh-specific primer of mRNA from Caki-1 cells, followed by ring-ligation and inverse PCR (Fig. 3a). The Caki-1 cell line produces relatively high levels of EMTh mRNA (Gründemann et al. 1998a).

In the first procedure, thermostable *Pwo* DNA polymerase was used at 72°C to generate double-stranded cDNA (see Methods for details). Inverse PCR yielded two major bands (Fig. 3b), estimated by regression analysis at 410±4 bp and 375±4 bp. This suggested that there were at least two tsps for the *EMTh* gene. The longer band was cloned, and seven clones were sequenced. The consensus 5'-terminus was 5'-AAAGAGAGGC, with small length variation among clones. The shorter band was confirmed to be authentic by secondary PCR with nested primers (data not shown).

In the second approach, single-stranded cDNA, after elimination of hybrid RNA, was phosphorylated at the 5'-end, circularized by T4 RNA ligase, and used in the inverse PCR as above. In this case, a nested PCR was necessary to produce a single major band that incorporated both primers (data not shown). Cloning of this band and sequencing of eight clones uniformly revealed the same 5'-terminus as above, with minor, probably PCR-acquired sequence differences.

Examination of promoters

Figure 4a shows, based on the results of tsp determination, the *EMTh* core promoter. The second tsp, about 35 bp downstream from the first, is probably contained in an initiator (Inr) element (also known as the cap signal) that perfectly matches the consensus (Bucher 1990). Canonical TATA box and CCAAT sequence elements are absent. However, a motif related to the TATA box recognition matrix (Bucher 1990) is present at the first tsp, at the usual distance upstream of the second tsp. There are several potential binding sites for ubiquitous transcription factors: seven GC boxes, putative sites for Sp1, are located within 220 bp upstream of the second tsp, and six of these are arranged in pairs. Moreover, there are four sites

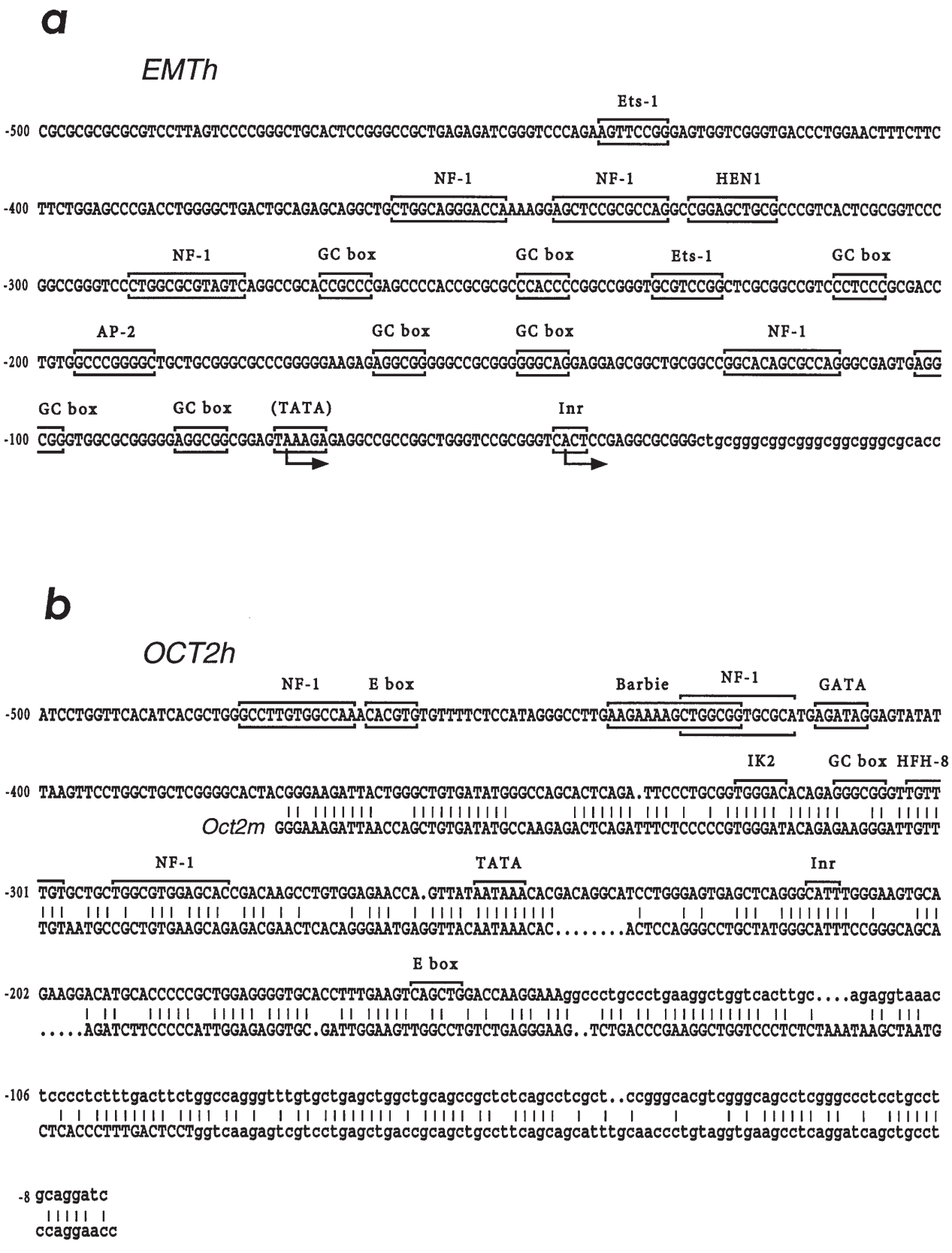

Fig. 4 Transcriptional elements of the *EMTh* promoter (**a**) and of the *OCT2h* promoter (**b**). For both genes, 500 bases genomic DNA sequence immediately upstream of the start codon are listed. The published 5' untranslated cDNA sequence is given in *lowercase letters*. Promoter elements and potential binding sites are indicated by *labeled square brackets* (*GC box* binding site for Sp1, *Barbie box* cis-element for barbiturate-induced expression). For the *EMTh* promoter, *arrows* indicate experimentally defined transcription start points. For the *OCT2h* promoter, an alignment is shown to all of the *Oct2m* promoter published so far (Mooslehner and Allen 1999). *Vertical dashes* Bases conserved among human and mouse, *dots* alignment gaps

for NF-1. In contrast, there are only single sites for specific transcription factors AP-2 and HEN1, and two sites for Ets-1.

At present, no original human cell line is known that expresses OCT2. Thus, no attempt was made here to determine the tsp. However, inspection of the presumed *OCT2h* promoter region (Fig. 4b) revealed a TATA box that closely matches the consensus, at position –250 relative to the translation start codon. At a center-to-center distance of 36 bp (Bucher 1990) downstream of the TATA box, a proper transcriptional initiator element is present. The presence of both elements in specific spacing is compatible with a tsp slightly upstream of the indicated Inr element (Javahery et al. 1994). Similar to the *EMTh* promoter, three sites for NF-1 are present. However, by marked contrast, only a single GC box is available in the *OCT2h* promoter. No sites for Ets-1, AP-2, or HEN-1 are discernible, but potential Barbie, GATA, and E boxes are present.

If we assume that important promoter elements are conserved among mammalian species, then sequence alignment and comparison of *OCT2* promoters from human and mouse may improve the identification of potential transcription factor binding sites. The latter sequence (Fig. 4b) has been recently published (Mooslehner and Allen 1999). There are marked differences between both promoters. Notably, the TATA box, the CCAAT sequence, and the IL6 site assigned to the mouse promoter have no match at equivalent positions in the human promoter, and the GC box and downstream E box in the human promoter are absent from the mouse promoter. However, there are also many conserved stretches (10 b long, 0 or 1 bases different), such as the IK2 and HFH-8 sites and the TATA box plus Inr element. Additional homologous regions do not fit with any of the known recognition matrices in the transcription factor database.

Discussion

The current study establishes the gene and promoter structures of the human non-neuronal monoamine transporters EMT and OCT2. To date, the genomic organization has been reported for only two ASF family members: first, OAT2r (originally termed NLT), an organic anion transporter (Simonson et al. 1994; Simonson and Iwanij 1995; Sekine et al. 1998) that is predominantly expressed in liver, and second, OCTN2h, a sodium-dependent carnitine transporter that is expressed in many tissues (Wu et al. 1998; Tamai et al. 1998). Gene sizes differ profoundly among *OAT2r* (6 kb), *OCTN2h* (27 kb), *OCT2h* (46 kb), and *EMTh* (77 kb), as do substrate specificities. Nevertheless, the locations of exon/intron borders are virtually identical (Fig. 1). This implies that evolutionary pressure has conserved exon junctions within this family of transporters and possibly reflects some common basic protein structure. Conversely, the variation of intron number and position around proposed transmembrane segments 11 and 12 may indicate fundamental differences in transporter structure in this region.

It has been reported, for other transporter families, that single transmembrane segments are encoded by separate exons (Lam et al. 1993). This is not generally correct for the ASF family, because exons 3, 4, and 7 encompass two or more transmembrane segments. It is however true that exon disruptions, with the exception of junctions at TM7 and TM12, lie within hydrophilic protein domains (Lam et al. 1993).

One distinctive feature of the EMTh cDNA is the very high GC content at its 5' terminus. With a peak value of 90% (window size 50 bases), the EMTh cDNA is among the top rating cDNAs known so far, e.g., human basic fibroblast growth factor (94%), human insulin-like growth factor II receptor (94%), and human huntingtin (92%). An explanation comes from nucleotide composition analysis of the genomic DNA (Fig. 1). The promoter and exon 1 of the *EMTh* gene are part of a CpG island, a region with relatively high GC content and clustering of CpG dinucleotides. CpG dinucleotides are generally found in vertebrate genomes at a much lower frequency than expected on statistical grounds (Gardiner-Garden and Frommer 1987). In CpG islands, however, they are found in large numbers. A CpG island has also been reported for the gene of EMTm (Verhaagh et al. 1999). CpG islands are commonly but not exclusively found at promoters of house-keeping proteins (Gardiner-Garden and Frommer 1987) and so the presence of a CpG island is not sufficient evidence to conclude that a protein has a maintenance function. The purpose of CpG islands is not entirely clear, but methylation of CpG island DNA evidently determines gene activity (Cross and Bird 1995).

Interestingly, although the genes of EMTh and OCT2h have probably evolved from a common ancestor, *OCT2h* is not associated with a CpG island. Here, there is only moderate CpG clustering, GC content is substantially lower, and peaks are confined to exon 1.

The high GC content poses a serious obstacle for EMTh DNA amplification by PCR (Gründemann et al. 1998a) and may also impede propagation in bacteria. Another consequence could be the failure in this study to determine the *EMTh* transcription start point by primer extension or RNA primer ligation to decapped mRNA (not shown; Schaefer 1995). Moreover, it is presently unclear whether experiments for *EMTh* tsp determination (Fig. 3) may have been biased toward GC-depleted enzyme-accessible mRNA 5'-ends, i.e., ends with a low potential of forming hairpins, and so results must be viewed with caution.

A comparison of promoters of *EMTh* and *OCT2h* reveals fundamental differences (Fig. 4). The *EMTh* promoter is a prototype of a TATA-less promoter. In such a promoter, RNA polymerase II may be positioned by a strong initiator element, as seems to be the case for the second tsp. A large number of potential binding sites for ubiquitous transcription factors Sp1 (Kadonaga et al. 1986) and NF-1 (Sumner et al. 1996) should provide for substantial basal transcription of the *EMTh* gene. This is in accordance with the observed broad tissue-distribution of EMT mRNA (Gründemann et al. 1998a; Kekuda et al.

1998; Verhaagh et al. 1999). However, EMT expression levels clearly differ among tissues and also vary during embryogenesis. Thus, tissue-specific transcription factors are likely to modulate transcription initiation. Potential sites for known specific factors are rare in the *EMTh* promoter. Among these, AP-2 (Williams and Tijan 1991) and HEN1 (Brown and Baer 1994), both of which are involved in the regulation of neural development and differentiation, may exert temporal and spatial control on the level of transcription.

In the *OCT2h* promoter, the start point of transcription may be determined by a combination of TATA box and Inr element. In contrast to the *EMTh* promoter, the *OCT2h* core promoter hardly contains binding sites for ubiquitous transcription factors, such as Sp1 and NF-1. Rather, it holds many possible regulatory sites for specific factors, such as E boxes, GATA, and IK2.

A salient feature of OCT2 is its very narrow tissue-distribution. Based on results from RT-PCR and *in situ* hybridization, transcription of the *OCT2* gene is limited to S3 segments of renal proximal tubules (Gründemann et al. 1998b) and, much more weakly, to certain brain areas (Gründemann et al. 1997; Busch et al. 1998; Mooslehner and Allen 1999). What factor(s) then limit expression of OCT2 to the observed locations? One interesting candidate is indicated by a potential binding site for HFH-8, a site that is conserved between mouse and human *OCT2* promoters. HFH-8 is a member of the forkhead/winged helix class of regulators. Members of this multigene family are involved in the control of tissue-specific gene transcription. The expression of human HFH-3, for instance, is limited to the renal distal tubule (Overdier et al. 1997).

Obviously, detailed promoter activity studies are needed on *EMTh* and *OCT2h* promoter elements to verify the efficacy of putative transcription-factor-binding sites found in this work and to understand the cell-specific and tissue-specific expression of both transporters. This may involve the search for transcription factors that bind to conserved promoter elements (Fig. 4b). Furthermore, some attention should be directed to the long 5' untranslated region of OCT2 mRNA, a region that provides ample space for regulating the initiation of translation. In contrast, the 5' untranslated regions of EMTh mRNA are quite short, which is conceivably yet another consequence of the high GC content.

In summary, we have uncovered the gene and promoter structures of the human non-neuronal monoamine transporters EMT and OCT2. Since both carriers may play important roles in catecholamine homeostasis and, as such, are candidate genes in human disease, the present results provide a basis for the analysis of genetic variation and transcriptional regulation. Moreover, it is now feasible to search for intragenic repetitive sequences that may serve as polymorphic markers.

Acknowledgements We thank Katarzyna Baran, Anke Ripperger, and Barbara Wallenwein for skillful technical assistance, P. Kioschis for screening the PAC library, and G. Rappold for providing RNA. This work was supported by Deutsche Forschungsgemeinschaft (SFB 601/A4 and FOR 302/1–2/B2).

References

Barnes WM (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. Proc Natl Acad Sci USA 91:2216–2220

Breidert T, Spitzenberger F, Gründemann D, Schömig E (1998) Catecholamine transport by the organic cation transporter type 1 (OCT1). Br J Pharmacol 125:218–224

Brown L, Baer R (1994) HEN1 encodes a 20-kilodalton phosphoprotein that binds an extended e box motif as a homodimer. Mol Cell Biol 14:1245–1255

Bucher P (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol 212:563–578

Busch AE, Karbach U, Miska D, Gorboulev V, Akhoundova A, Volk C, Arndt P, Ulzheimer JC, Sonders MS, Baumann C, Waldegger S, Lang F, Koepsell H (1998) Human neurons express the polyspecific cation transporter hOCT2, which translocates monoamine neurotransmitters, amantadine, and memantine. Mol Pharmacol 54:342–352

Chomczynski P (1992) One-hour downward alkaline capillary transfer for blotting of DNA and RNA. Anal Biochem 201:134–139

Cross SH, Bird AP (1995) CpG islands and genes. Curr Opin Genet Dev 5:309–314

Eisenhofer G, McCarty R, Pacak K, Russ H, Schömig E (1996) Disprocynium 24, a novel inhibitor of the extraneuronal monoamine transporter, has potent effects on the inactivation of circulating noradrenaline and adrenaline in conscious rat. Naunyn-Schmiedebergs Arch Pharmacol 354:287–294

Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196:261–282

Gorboulev V, Ulzheimer JC, Akhoundova A, Ulzheimer-Teuber I, Karbach U, Quester S, Baumann C, Lang F, Busch AE, Koepsell H (1997) Cloning and characterization of two human polyspecific organic cation transporters. DNA Cell Biol 16: 871–881

Gründemann D, Gorboulev V, Gambaryan S, Veyhl M, Koepsell H (1994) Drug excretion mediated by a new prototype of polyspecific transporter. Nature 372:549–552

Gründemann D, Babin-Ebell J, Martel F, Örding N, Schmidt A, Schömig E (1997) Primary structure and functional expression of the apical organic cation transporter from kidney epithelial LLC-PK1 cells. J Biol Chem 272:10408–10413

Gründemann D, Schechinger B, Rappold GA, Schömig E (1998a) Molecular identification of the corticosterone-sensitive extraneuronal monoamine transporter. Nat Neurosci 1:349–351

Gründemann D, Köster S, Kiefer N, Breidert T, Engelhardt M, Spitzenberger F, Obermüller N, Schömig E (1998b) Transport of monoamine transmitters by the organic cation transporter type 2 (OCT2). J Biol Chem 273:30915–30920

Gründemann D, Liebich G, Kiefer N, Köster S, Schömig E (1999) Selective substrates for non-neuronal monoamine transporters. Mol Pharmacol 56:1–10

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, Jong PJ de (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nat Genet 6:84–89

Iversen LL (1965) The uptake of catechol amines at high perfusion concentrations in the rat isolated heart: a novel catechol amine uptake process. Br J Pharmacol 25:18–33

Javahery R, Khachi A, Lo K, Zenzie-Gregory B, Smale ST (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol Cell Biol 14:116–127

Kadonaga JT, Jones KA, Tijan R (1986) Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem Sci 11:20–23

Kekuda R, Prasad PD, Wu X, Wang H, Fei Y-J, Leibach FH, Ganapathy V (1998) Cloning and functional characterization of a potential-sensitive, polyspecific organic cation transporter (OCT3) most abundantly expressed in placenta. J Biol Chem 273:15971–15979

Koehler MR, Wissinger B, Gorboulev V, Koepsell H, Schmid M (1997) The two human organic cation transporter genes SLC22A1 and SLC22A2 are located on chromosome 6q26. Cytogenet Cell Genet 79:198–200

Lam DM-K, Fei J, Zhang X-Y, Tam ACW, Zhu L-H, Huang F, King SC, Guo L-H (1993) Molecular cloning and structure of the human (GABATHG) GABA transporter gene. Mol Brain Res 19:227–232

Luo DF, Bui MM, Muir A, Maclaren NK, Thomson G, She JX (1995) Affected-sib-pair mapping of a novel susceptibility gene to insulin-dependent diabetes mellitus (IDDM 8) on chromosome 6q25-q27. Am J Hum Genet 57:911–919

Mooslehner KA, Allen ND (1999) Cloning of the mouse organic cation transporter 2 gene, Slc22a2, from an enhancer-trap transgene integration locus. Mamm Genome 10:218–224

Nagel G, Volk C, Friedrich T, Ulzheimer JC, Bamberg E, Koepsell H (1997) A reevaluation of substrate specificity of the rat cation transporter rOCT1. J Biol Chem 272:31953–31956

Okuda M, Saito H, Urakami Y, Takano M, Inui K (1996) cDNA cloning and functional expression of a novel rat kidney organic cation transporter, OCT2. Biochem Biophys Res Commun 224:500–507

Overdier DG, Ye H, Peterson RS, Clevidence DE, Costa RH (1997) The winged helix transcriptional activator HFH-3 is expressed in the distal tubules of embryonic and adult mouse kidney. J Biol Chem 272:13725–13730

Russ H, Engel W, Schömig E (1993) Isocyanines and pseudoisocyanines as a novel class of potent noradrenaline transport inhibitors: synthesis, detection, and biological activity. J Med Chem 36:4208–4213

Russ H, Staudt K, Martel F, Gliese M, Schömig E (1996) The extraneuronal transporter for monoamine transmitters exists in cells derived from human central nervous system glia. Eur J Neurosci 8:1256–1264

Schaefer BC (1995) Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. Anal Biochem 227:255–273

Schömig E, Babin-Ebell J, Russ H (1993) 1,1'-Diethyl-2,2'-cyanine (decynium 22) potently inhibits the renal transport of organic cations. Naunyn Schmiedebergs Arch Pharmacol 347: 379–383

Schömig E, Spitzenberger F, Engelhardt M, Martel F, Örding N, Gründemann D (1998) Molecular cloning and characterization of two novel transport proteins from rat kidney. FEBS Lett 425:79–86

Sekine T, Cha SH, Tsuda M, Apiwattanakul N, Nakajima N, Kanai Y, Endou H (1998) Identification of multispecific organic anion transporter 2 expressed predominantly in the liver. FEBS Lett 429:179–182

Senapathy P, Shapiro MB, Harris NL (1990) Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol 183:252–278

Simonson GD, Iwanij V (1995) Genomic organization and promoter sequence of a gene encoding a rat liver-specific type-I transport protein. Gene 154:243–247

Simonson GD, Vincent AC, Roberg KJ, Huang Y, Iwanij V (1994) Molecular cloning and characterization of a novel liver-specific transport protein. J Cell Sci 107:1065–1072

Strathmann M, Hamilton BA, Mayeda CA, Simon MI, Meyerowitz EM, Palazzolo MJ (1991) Transposon-facilitated DNA sequencing. Proc Natl Acad Sci USA 88:1247–1250

Sumner C, Shinohara T, Durham L, Traub R, Major EO, Amemiya K (1996) Expression of multiple classes of the nuclear factor-1 family in the developing human brain: differential expression of two classes of NF-1 genes. J Neurovirol 2:87–100

Tamai I, Ohashi R, Nezu J, Yabuuchi H, Oku A, Shimane M, Sai Y, Tsuji A (1998) Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2. J Biol Chem 273:20378–20382

Verhaagh S, Schweifer N, Barlow DP, Zwart R (1999) Cloning of the mouse and human solute carrier 22a3 (Slc22a3/SLC22A3) identifies a conserved cluster of three organic cation transporters on mouse chromosome 17 and human 6q26-q27. Genomics 55:209–218

Williams T, Tijan R (1991) Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. Genes Dev 5:670–682

Windenger E, Dietze P, Karas H, Knüppel R (1996) TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res 24:238–241

Wu X, Prasad PD, Leibach FH, Ganapathy V (1998) cDNA sequence, transport function, and genomic organization of human OCTN2, a new member of the organic cation transporter family. Biochem Biophys Res Commun 246:589–595