Amino Acids (2001) 21: 211–219

Amino
Acids

Analysis of the genomic organization of the human cationic amino acid transporters CAT-1, CAT-2 and CAT-4

Short Communication

R. Hammermann, G. Brunn, and **K. Racké**

Institute of Pharmacology and Toxicology, Rheinische Friedrich-Wilhelms-Universität, Bonn, Federal Republic of Germany

Accepted January 8, 2001

Summary. By screening nucleotide databases, sequences containing the complete genes of the human cationic amino acid transporters (hCATs) 1, 2 and 4 were identified. Analysis of the genomic organization revealed that hCAT-2 consists of 12 translated exons and most likely of 2 untranslated exons. The splice variants hCAT-2A and hCAT-2B use exon 7 and 6, respectively. The hCAT-2 gene structure is closely related to the structure of hCAT-1, suggesting that they belong to a common gene family. hCAT-4 consists of only 4 translated exons and 3 short introns. Exons of identical size and highly homologous to exon 3 of hCAT-4 are present in hCAT-1 and hCAT-2.

Keywords: Cationic amino acids – Human cationic amino acid transporter – Solute carrier family – Genomic organization – Exon-intron structure

Abbreviations: AC, accession number; hCAT, human cationic amino acid transporter; kB, kilobases; UTR, untranslated region.

Introduction

The amino acid L-arginine, beyond being precursor in protein synthesis, is substrate for two mediator pathways, the nitric oxide synthase and arginase – polyamine pathway, whereby the availability of L-arginine can become a limiting factor (e.g. Hey et al., 1997). In most mammalian cells L-arginine supply is carried out by uptake of extracellular L-arginine via multiple transporter systems, among these are the cationic amino acid specific transport system y^+ and several broad scope amino acid transport systems (y^+L, $b^{0,+}$, and $B^{0,+}$) (for review see Devés and Boyd, 1998). In many cells, system y^+ appears to be functionally the most important mechanism for the uptake of L-arginine. Recent studies have identified several homologous

cDNAs from different species, encoding proteins which were characterized as specific cationic amino acid transporters (CATs) with properties of system y^+ (for reviews see Closs, 1996; MacLeod and Kakuda, 1996; Devés and Boyd, 1998). The nucleic acid sequences of CAT-1 had first been described as putative murine ecotropic retrovirus receptor (Albritton et al., 1989), but short time later its properties as cationic amino acid transporter were recognized (Kim et al., 1991). So far, four additional members of this family, the splice variants CAT-2A and CAT-2B (MacLeod et al., 1990; Closs et al., 1993), CAT-3 (Hosokawa et al., 1997; Ito and Groudine, 1997) and CAT-4 (Sperandeo et al., 1998), have been identified. Initially the nomenclatur of CAT-2B was CAT-2 because it was the first of the two CAT-2 isoforms to be isolated (MacLeod and Kakuda, 1996). The nucleic acid sequences of the human transporters (hCATs) are known at present for hCAT-1 (Yoshimoto et al., 1991), hCAT-2A and hCAT-2B (Hoshide et al., 1996; Closs et al., 1997) and hCAT-4 (Sperandeo et al., 1998). hCAT-3 has been cloned (Vékony et al., 2000), but its sequence is not yet available.

So far, the genomic structure of CAT genes is only partially known for hCAT-1 (Yoshimoto et al., 1991). On the other hand, the progressing identification of the human genome (Brown, 2000) provides huge sequence data with large amounts of hidden information which will have to be uncovered. With the exception of hCAT-3 all chromosomal positions of the human cationic amino acid transporter genes are described by the "International Human Genome Project". There the nomenclature solute carrier family 7 member 1–4 and the abbreviations SLC7A1, SLC7A2, SLC7A3 and SLC7A4 for the four hCAT transporter genes is used. We tried to illuminate the genomic organization of hCAT genes by aligning known mRNA sequences of hCATs with the so far published human genomic sequences.

Methods

The CAT sequences were aligned via internet by the BLASTN (version 2.0.14) program (Altschul et al., 1997) against the human contigs of all chromosomes published by the "International Human Genome Project" (November 2000). To search also unfinished human genome sequences, alignment was additionally performed against the "htgs" database. The alignment was carried out with the following sequences: hCAT1 (accession numbers (AC): X59155, X57303 and AF078107), hCAT-2A (AC: U76368), hCAT-2B (AC: U76369 and D29990), mouse CAT-3 (AC: U70859), rat CAT-3 (AC: AB000113), and hCAT-4 (formerly named CAT-3; AC: AJ000730). Alignment between sequences of corresponding exons or regions within the exons was performed with PC-Gene™ (version 6.85).

Results and discussion

hCAT-1

The partial genomic organization of hCAT-1 had already been described by Yoshimoto et al. (1991). They have concluded from a cosmid clone, obtained

from a cosmid library derived from human lymphocytes, that the open reading frame of hCAT-1 consists of 11 exons and 10 introns. The genetic position was localized to chromosome 13q12–q14 (Albritton et al., 1992). In agreement with this, we have identified two genomic DNA sequences located to chromosome 13 (AC: AL353601 and AL163532) with the highest alignment score for hCAT-1 mRNA sequence in the "htgs" database. It was possible to identify the consecutive hCAT-1 gene sequence from exon 1–11 within the 5th unordered piece of the sequence AL353601 which consists of 6 unordered pieces (contigs). The untranslated exons (−1) and (−2) were localized within the 9th and the 3rd unordered piece of the sequence AL163532 (consists of 9 unordered pieces), respectively. Accordingly, a genomic structure of at least 13 exons (11 translated and 2 untranslated) is deduced (Fig. 1A), and a size of 21.8 kb was estimated for the gene region containing only the translated exons. The introns between the untranslated exons seem to be very large because overlapping regions of the published sequences can not be detected. All exon-intron junctions showed 5′-gt-donor and ag-3′-acceptor sites and are therefore conform to eukaryotic consensus splice junction sites (Breathnach and Chambon, 1981). Arrangement, boundaries and lengths of the exons are shown in the lower part of Fig. 1A and Table 1.

Alignment of the three published hCAT-1 cDNA sequences (X59155, X57303 and AF078107 shows only in one position a discrepancy which results in a change of deduced amino acid sequence (Table 2). The genomic sequence is identical to the cDNA sequence X59155 except for one nucleotide (C to G change of position −129; the first nucleotide of the start codon ATG being position 1) in the untranslated exon (−2).

hCAT-2

Hoshide et al. (1996) showed that the hCAT-2 gene is localized to chromosome 8p21.3–p22. In agreement with this, we have identified a genomic DNA sequence located to chromosome 8p21.3–p22 (AC: AB020863) with the highest alignment score for hCAT-2A and hCAT-2B mRNA sequences. The first nucleotide of exon (−1) starts at position 52,697 and the last nucleotide of the last exon (12) ends at position 78,957. Altogether 13 exons and 12 introns could be identified in this sequence (Fig. 1B, upper part), and a size of 21.6 kb was estimated for the gene region containing only the translated exons. As the published cDNA (AC U76368) contained additional 47 nucleotides downstream to exon (−1) which could not be identified in the genomic sequence (AC AB020863), it may be assumed that an additional untranslated exon (−2) may exist, located at least 52,7 kb downstream to exon (−1). This assumption is supported by the fact that a similar structure was identified in the gene of hCAT-1. Again, all exon-intron junctions showed 5′-gt-donor and ag-3′-acceptor sites. Arrangement, boundaries and lengths of the exons are shown in the lower part of Fig. 1B and Table 1. The splice variants hCAT-2A and hCAT-2B use exon 7 and 6, respectively, at the same position of the mRNA. The splicing occurs within a triplet (Ser 352) and the

Table 1. Comparison of exon lengths and homologies of corresponding exons of the human cationic amino acid transporters hCAT-1, hCAT-2, and hCAT-4. Homologies between CAT-1 and CAT-2, respectively, to CAT-4 are only shown from the region with highest homology and from the exons with identical size. The results of other alignments depend largely on software parameters and are therefore not convenient

Gene						Homology								
CAT-1		CAT-2		CAT-4		CAT-1 to CAT-2			CAT-1 to CAT-4			CAT-2 to CAT-4		
Ex	[bp]	Ex	[bp]	Ex	[bp]	Ex	[%]	Ex	Ex	[%]	Ex	Ex	[%]	Ex
−2	≥33	−2	?	−1	?									
−1	100	−1	126	1	1,025	−1	54.0	−1						
1	384	1	398			1	72.9	1						
2	159	2	156			2	69.2	2						
3	175	3	166			3	69.9	3						
4	122	4	134			4	75.4	4						
5	223	5	223			5	73.5	5						
6	140	6 (B)	140	2	641	6	80.0	6 (B)						
		7 (A)	137			6	73.0	7 (A)						
7	103	8	103			7	83.5	8	7	67.0	2	8	69.9	2
8	218	9	206			8	66.5	9						
9	167	10	167			9	69.5	10						
10	109	11	109	3	109	10	69.7	11	10	70.6	3	11	64.2	3
11	225	12	213	4	512	11	66.7	12						

Table 2. Documentation of sequence differences between published cDNA and genomic sequences of hCAT-1 and hCAT-2 which lead to changes in deduced protein sequences. Compared are the cDNA sequences X59155, X57303, AF078107 (hCAT-1), U76368 (hCAT-2A), D29990 (hCAT-2B) and the genomic sequence AB020863 (hCAT-2). The amino acids (three-letter symbols) which are deduced predominatly from the corresponding nucleic acid sequences are highlighted in bold letters

Nucleotides	Amino acids position	name	Nucleotides	Amino acids position	name	Nucleotides	Amino acids position	name
hCAT-1								
X59155 (CAT-1 cDNA)			X57303 (CAT-1 cDNA)			AF078107 (CAT-1 cDNA)		
G	23	**Arg**	C	23	Pro	G	23	**Arg**
hCAT-2								
U76368 (CAT-2A cDNA)			D29990 (CAT-2B cDNA)			AB020863 (CAT-2A/B genomic DNA)		
T	134	**Trp**	G	134	Gly	T	134/134	**Trp**
T	364	**Ile**	A	364	Asn	T	364/364	**Ile**
AC	519	**Thr**	TA	520	Tyr	AC	519/520	**Thr**
A	530	Thr	G	531	**Ala**	G	530/531	**Ala**
A	546	**Gln**	A	547	**Gln**	T	546/547	Leu
A	551	**Gln**	G	552	Arg	A	551/552	**Gln**
G	565	**Ala**	A	566	Thr	G	565/566	**Ala**
GC	567	**Ser**	CG	568	Thr	GC	567/568	**Ser**

first five deduced amino acids are identical between the two variants. Thus, the splicing starts earlier than expected from differences in the amino acid sequences (Closs et al., 1997).

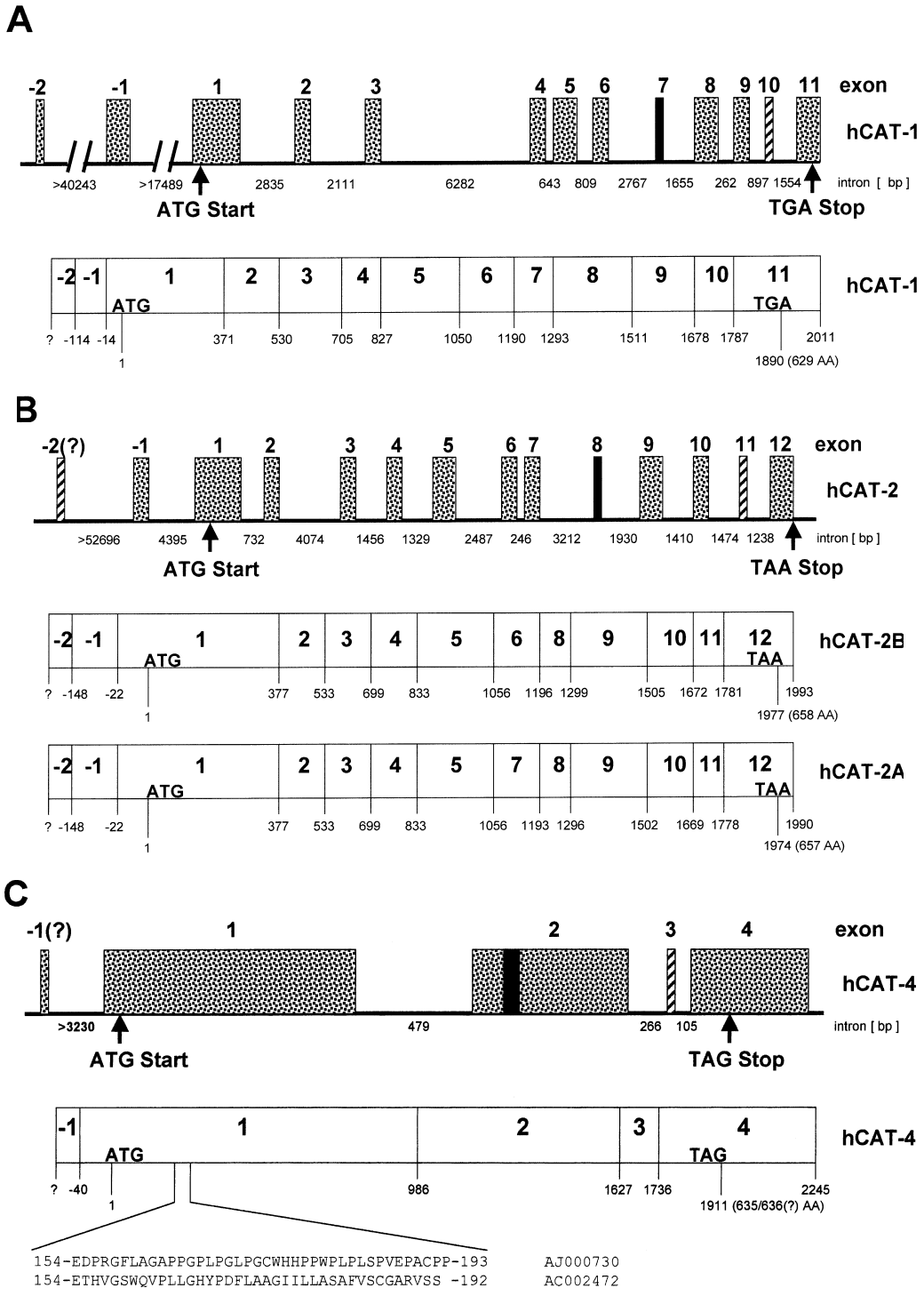
Furthermore, the genomic sequence allows the clarification of the unidentified position (−34) of the mRNA sequence (AC U76368): N should be replaced by GC. Of the few differences between the published mRNAs and the genomic sequences only those resulting in changes of the deduced amino acids might be mentioned and are summarized in Table 2. Whether these differences reflect sequencing errors or genetic variability within the population which might even have functional significance remains to be evaluated. Noteworthy, most of the differences are located at the carboxy terminus of the protein.

hCAT-3

Although hCAT-3 has already been cloned, its nucleotide sequence is not yet available, but its deduced amino acid sequence has been given as 83.4% and 81.6% identical to rat and mouse CAT-3, respectively (Vékony et al., 2000). Using the mouse or rat CAT-3 nucleotide sequence a homologous human genomic sequence could not be identified in the databases.

hCAT-4

hCAT-4 had been localized to chromosome 22q11.2 (Sperandeo et al., 1998). In agreement to this, we have again identified two genomic sequences located


A
-2 -1 1 2 3 4 5 6 7 8 9 10 11 exon
hCAT-1
>40243 >17489 2835 2111 6282 643 809 2767 1655 262 897 1554 intron [bp]
ATG Start
TGA Stop
-2 -1 1 2 3 4 5 6 7 8 9 10 11
hCAT-1
ATG
TGA
? -114 -14 371 530 705 827 1050 1190 1293 1511 1678 1787 2011
1
1890 (629 AA)
B
-2(?) -1 1 2 3 4 5 6 7 8 9 10 11 12 exon
hCAT-2
>52696 4395 732 4074 1456 1329 2487 246 3212 1930 1410 1474 1238 intron [bp]
ATG Start
TAA Stop
-2 -1 1 2 3 4 5 6 8 9 10 11 12
hCAT-2B
ATG
TAA
? -148 -22 377 533 699 833 1056 1196 1299 1505 1672 1781 1993
1
1977 (658 AA)
-2 -1 1 2 3 4 5 7 8 9 10 11 12
hCAT-2A
ATG
TAA
? -148 -22 377 533 699 833 1056 1193 1296 1502 1669 1778 1990
1
1974 (657 AA)
C
-1(?) 1 2 3 4 exon
hCAT-4
>3230 479 266 105 intron [bp]
ATG Start
TAG Stop
-1 1 2 3 4
hCAT-4
ATG
TAG
? -40 986 1627 1736 2245
1
1911 (635/636(?) AA)
154-EDPRGFLAGAPPGPLPGLPGCWHHPPWPLPLSPVEPACPP-193 AJ000730
154-ETHVGSWQVPLLGHYPDFLAAGIILLASAFVSCGARVSS -192 AC002472

to chromosome 22q11.2–q13.2 (AC: AC002472 and AB002059) with the highest alignment score for hCAT-4 mRNA sequence. The first nucleotide of exon 1 starts at position 71,086 and the last nucleotide of the last exon (4) ends at position 74,217. This results in a gene structure of only 4 translated exons and 3 short introns (Fig. 1C, upper part), and a size of 3.1 kb was estimated for the gene region containing only the translated exons. As the published cDNA (AC AJ000730) contained additional 38 nucleotides downstream to exon 1 which could not be identified in the genomic sequence (AC AB002059), it may be assumed again that an additional untranslated exon (−1) may exist, located at least 3.2 kb downstream to exon 1. Again, all exon-intron junctions showed 5′-gt-donor and ag-3′-acceptor sites. Arrangement, boundaries and lengths of the exons are shown in the lower part of Fig. 1C and Table 1.

Surprisingly, the published cDNA sequence (AJ000730) and the genomic sequences (AC002472 and AB002059) reveal sequence differences in three positions which lead to a frame shift in deduced hCAT-4 protein sequence over a stretch of 40 amino acids (Fig. 1C). On principle, genomic sequences published by the "International Human Genome Project" appear to have a higher level of verification than cDNA sequences characterized by a single group. Furthermore, the unusual amino acid sequence deduced from the cDNA sequence (14 prolines within this region of 40 amino acids) questioned the cDNA sequence (Fig. 1C). Nevertheless, further studies of hCAT-4 sequence and future alignments to other species are needed to solve this discrepancy.

Concluding remarks

Strikingly, the genomic organization of the genes of hCAT-1 and hCAT-2 is almost identical. All exons have similar (7 exons) or even equal length (5 exons) (Table 1), but different sizes of introns between corresponding exons of hCAT-1 and hCAT-2. This indicates that "important information" (exons) resulting in deduced protein sequences is much more conserved than "unimportant information" (introns) during evolution. Nevertheless, the gene sizes from the regions containing only the translated exons of hCAT-1 and hCAT-2 are almost identical. Additionally, both genomic structures contain

Fig. 1. Genomic organization of the human cationic amino acid transporter hCAT-1 **(A)**, hCAT-2 **(B)** and hCAT-4 **(C)** and their relationship to the spliced mRNAs. In the upper panel of each part, the genomic structure of the genes is illustrated schematically as speckled bars (exons) and open regions (introns) with the exact intron length in basepairs given under the line. Arrows indicate start and stop codons. Hatched bars represent an exon which has an corresponding exon of exactly 109 basepairs in each of the three hCAT genes. The black bars show the region of the highest homology between all three genes. In the lower panel of each part the exon boundaries of the mature mRNAs are shown in open boxes. The entries of the exons are given with respect to the ATG start codon and the numbers of deduced amino acids (AA) are presented in parenthesis. Part C reveals additionally the alignment of a deduced region of hCAT-4 protein sequence (AA 154–193) which is different between the two published nucleic acid sequences (Accession numbers of mRNA sequence [AJ000730] and of genomic sequence [AC002472])

an untranslated exon (−1) of similar size and probably an additional exon (−2) may exist with a relatively large distance downstream to exon (−1) (Fig. 1). The high sequence homology between corresponding exons of hCAT-1 and hCAT-2 is a further argument that these two genes are evolutionary closely related. In contrast, the genomic organization and size of the hCAT-4 gene is largely different, indicating a greater evolutionary distance to the two other CAT genes. This genomic distance seems to correlate to functional differences, as hCAT-4 appears to require additionally expressed factors to function as transporter for L-arginine (Wolf et al., 2000).

The homology analysis of the three hCAT genes (Table 1) revealed several interesting points: 1) The highest homology is observed between exon 7 of hCAT-1 and exon 8 of hCAT-2 and both exons show high homology to a sequence segment within exon 2 of hCAT-4 (marked in Fig. 1 as black bars, Table 1). The high conservation of this region suggests that it may encode an important functional or regulatory domain. 2) Exon 3 of hCAT-4 has corresponding exons of identical length (109 basepairs, Table 1) in hCAT-1 and hCAT-2 (given as hatched bars in Fig. 1). 3) Exon 6 of hCAT-1 has a much higher homology to exon 6 of hCAT-2 (= hCAT-2B) than to exon 7 of hCAT-2 (= hCAT-2A). This correlates with the functional properties of CAT-1 and CAT-2B being high affinity CATs and CAT-2A being a low affinity CAT.

The present analysis reveal that the genes of hCAT-1, hCAT-2 and hCAT-4 contain large 5′-untranslated sequences with at least 1–2 untranslated exons. Interestingly, the mouse CAT-2 gene shows the presence of diverse 5′ untranslated sequences in mCAT-2 transcripts which are dispersed over 18 kb and which were obviously expressed in a cell specific manner (Finley et al., 1995). Therefore, it appears possible that the hCAT genes may contain additional, so far unidentified, untranslated exons which may be used alternatively. Altogether, the complex structure of CAT genes suggest that very precise mechanisms may control their expression. The illumination of these processes will be a very interesting, but difficult task in the future.

References

Albritton LM, Tseng L, Scadden D, Cunningham JM (1989) A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. Cell 57: 659–666

Albritton LM, Bowcock AM, Eddy RL, Morton CC, Tseng L, Farrer LA, Cavalli-Sfroza LL, Shows TB, Cunningham JM (1992) The human cationic amino acid transporter (ATRC1): physical and genetic mapping to 13q12–q14. Genomics 12: 430–434

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402

Breathnach R, Chambon P (1981) Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem 50: 349–383

Brown K (2000) The human genome business today. Sci Am 7: 40–45

Closs EI (1996) CATs, a family of three distinct mammalian cationic amino acid transporters. Amino Acids 11: 193–208

Closs EI, Albritton LM, Kim JW, Cunningham JM (1993) Identification of a low affinity, high capacity transporter of cationic amino acids in mouse liver. J Biol Chem 268: 7538–7544

Closs EI, Gräf P, Habermeier A, Cunningham JM, Förstermann U (1997) Human cationic amino acid transporters hCAT-1, hCAT-2A, and hCAT-2B: three related carriers with distinct transport properties. Biochem 36: 6462–6468

Devés R, Boyd CAR (1998) Transporters for cationic amino acids in animal cells: discovery, structure, and function. Physiol Rev 78: 487–545

Finley KD, Kakuda DK, Barrieux A, Kleeman J, Huynh PD, MacLeod CL (1995) A mammalian arginine/lysine transporter uses multiple promoters. Proc Natl Acad Sci USA 92: 9378–9382

Hey C, Boucher JL, Vadon-Le Goff S, Ketterer G, Wessler I, Racké K (1997) Inhibition of arginase in rat and rabbit alveolar macrophages by N^{ω}-hydroxy-D,L-indospicine, effects on L-arginine utilization by nitric oxide synthase. Br J Pharmacol 121: 395–400

Hoshide R, Ikeda Y, Karashima S, Matsuura T, Komaki S, Kishino T, Niikawa N, Endo F, Matsuda I (1996) Molecular cloning, tissue distribution, and chromosomal localization of human cationic amino acid transporter 2 (hCAT2). Genomics 38: 174–178

Hosokawa H, Sawamura T, Kobayashi S, Ninomiya H, Miwa S, Masaki T (1997) Cloning and characterization of a brain-specific cationic amino acid transporter. J Biol Chem 272: 8717–8722

Ito K, Groudine M (1997) A new member of the cationic amino acid transporter family is preferentially expressed in adult mouse brain. J Biol Chem 272: 26780–26786

Kim JW, Closs EI, Albritton LM, Cunningham JM (1991) Transport of cationic amino acids by the mouse ecotropic retrovirus receptor. Nature 352: 725–728

MacLeod CL, Kakuda DK (1996) Regulation of CAT: cationic amino acid transporter gene expression. Amino Acids 11: 171–191

MacLeod CL, Finley K, Kakuda D, Kozak CA, Wilkinson MF (1990) Activated T cells express a novel gene on chromosome 8 that is closely related to the murine ecotropic retroviral receptor. Mol Cell Biol 10: 3663–3674

Sperandeo MP, Borsani G, Incerti B, Zollo M, Rossi E, Zuffardi O, Castaldo P, Taglialatela M, Andria G, Sebastio G (1998) The gene encoding a cationic amino acid transporter (SLC7A4) maps to the region deleted in the velocardiofacial syndrome. Genomics 49: 230–236

Vékony N, Wolf S, Boissel JP, Closs EI (2000) Human cationic amino acid transporter-3 (hCAT-3): Cloning of a cDNA comprising the complete coding region and analysis of expression in different human tissues. Naunyn-Schmiedeberg's Arch Pharmacol 361 [Suppl]: R5

Wolf S, Vékony N, Closs EI (2000) The expression of the human cationic amino acid transporter-4 (hCAT-4) in the plasma membrane does not correlate with its transport activity. Naunyn-Schmiedeberg's Arch Pharmacol 361 [Suppl]: R8

Yoshimoto T, Yoshimoto E, Meruelo D (1991) Molecular cloning and characterization of a novel human gene homologous to the murine ecotropic retroviral receptor. Virology 185: 10–17

Authors' address: Dr. Rainer Hammermann, Institut für Pharmakologie und Toxikologie, Universität Bonn, Reuterstraße 2B, D-53113 Bonn, Federal Republic of Germany, Fax 0228/73-5404, E-mail: r.hammermann@uni-bonn.de

Received September 8, 2000