Cloning and sequencing of four new mammalian monocarboxylate transporter (MCT) homologues confirms the existence of a transporter family with an ancient past

Nigel T. PRICE[1], Vicky N. JACKSON[2] and Andrew P. HALESTRAP[3]
Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, U.K.

Measurement of monocarboxylate transport kinetics in a range of cell types has provided strong circumstantial evidence for a family of monocarboxylate transporters (MCTs). Two mammalian MCT isoforms (MCT1 and MCT2) and a chicken isoform (REMP or MCT3) have already been cloned, sequenced and expressed, and another MCT-like sequence (XPCT) has been identified. Here we report the identification of new human MCT homologues in the database of expression sequence tags and the cloning and sequencing of four new full-length MCT-like sequences from human cDNA libraries, which we have denoted MCT3, MCT4, MCT5 and MCT6. Northern blotting revealed a unique tissue distribution for the expression of mRNA for each of the seven putative MCT isoforms (MCT1–MCT6 and XPCT). All sequences were predicted to have 12 transmembrane (TM) helical domains with a large intracellular loop between TM6 and TM7. Multiple sequence alignments showed identities ranging from 20 % to 55 %, with the greatest conservation in the predicted TM regions and more variation in the C-terminal than the N-terminal region. Searching of additional sequence databases identified candidate MCT homologues from the yeast *Saccharomyces cerevisiae*, the nematode worm *Caenorhabditis elegans* and the archaebacterium *Sulfolobus solfataricus*. Together these sequences constitute a new family of transporters with some strongly conserved sequence motifs, the possible functions of which are discussed.

INTRODUCTION

In quantitative terms, lactic acid is one of the most important metabolites in the body, substantial amounts being used and/or produced by almost all mammalian cells. As such it must be rapidly transported into and out of cells. Lactic acid transport across the plasma membrane is catalysed by proton-linked monocarboxylate transporters (MCTs), which are also responsible for the transport of pyruvate and the ketone bodies acetoacetate, β-hydroxybutyrate and acetate [1]. Extensive studies in this and other laboratories of the kinetics and substrate and inhibitor specificity of monocarboxylate transport into a range of mammalian cells led us to propose that there is a family of MCTs, each member having slightly different properties related to the metabolic requirements of the tissues in which they are found [1–3]. Some cells such as cardiac myocytes seem to possess two distinct MCT isoforms within the same cell [3,4]. The proposal that several MCT isoforms exist in mammals has now been confirmed by molecular biological techniques. The first MCT was cloned from Chinese hamster ovary cells [5] and later functionally expressed in a breast tumour cell line [6]. Named MCT1, similar transporters from human, rat and mouse have been cloned and sequenced [7–9]. Western and Northern blotting has shown that MCT1 is found in the majority of tissues examined in both hamster and rat [5,6,10].

Garcia et al. [11] have cloned and sequenced a second isoform of MCT from Syrian hamster liver, a tissue from which MCT1 is absent. Named MCT2, this has approx. 60 % identity with MCT1 and has been functionally expressed in baculovirus-infected insect cells [11]. We have since cloned and sequenced MCT2 from rat and found that it is expressed in many fewer tissues in rat than in hamster [10]. During the genetic characterization of the X-inactivation centre-containing region of the X chromosome, Lafrenière et al. [12] identified a novel gene encoding a predicted 66 kDa protein whose sequence is 27 % identical with human MCT1 and, like that of MCT1, is predicted to contain 12 hydrophobic transmembrane (TM) domains characteristic of a transporter protein. The protein has an extended N-terminus containing mainly proline/glutamic residue repeats. Such 'PEST' domains are thought to be indicative of rapid or conditional proteolytic degradation [13,14]. The protein was designated XPCT (for X-linked PEST-containing transporter), and Northern blotting suggests that it is expressed in most tissues [12]. A cDNA has been identified in a chicken retinal cDNA library that encodes a developmentally regulated integral membrane protein of 50 kDa that seems to be closely related to MCT1 and MCT2. The protein was shown to be expressed in chick retinal epithelia and was initially named REMP (retinal epithelial membrane protein) [15] and subsequently functionally expressed and renamed MCT3 [16]. Hydrophobicity plots again predict 12 TM domains similar to MCT1 and MCT2, with which it shares 43 % and 45 % identity respectively [16].

Here we report the identification of four new human MCT homologues in the database of expression sequence tags (dbEST)

Abbreviations used: dbEST, database of expression sequence tags; EST, expression sequence tag; MCT, monocarboxylate transporter; REMP, retinal epithelial membrane protein; TM, transmembrane.

[1] Present address: Hannah Research Institute, Ayr KA6 5HL, Scotland, U.K.

[2] Present address: Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, Scotland, U.K.

[3] To whom correspondence should be addressed.

The nucleotide sequence results reported here will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession numbers U81800, U79745, U59299 and U79745.

and the cloning and sequencing of full-length cDNA species from human cDNA libraries. Northern blotting shows that mRNA species for these new mammalian MCT isoforms are differentially expressed in a range of human tissues.

EXPERIMENTAL

Materials

Materials were obtained from the sources described previously [9,10], with the following additions. The human placental and heart λgt11 cDNA libraries and the human multi-tissue Northern blots were from Clontech (Palo Alto, CA, U.S.A.) The human circulating-blood cDNA library was from Stratagene (La Jolla, CA, U.S.A.). The probes used for screening the cDNA libraries were generated by PCR with exact-match oligonucleotides to putative MCT sequences identified in dbEST as described below.

Identification of potential new MCT isoforms in dbEST

We have sought to identify new putative MCTs by searching dbEST [17] with the known MCT or MCT-related protein sequences (MCT1, MCT2, REMP and XPCT). These protein sequences were compared against the expression sequence tag (EST) sequences translated in all six possible reading frames by using the TBLASTN computer program [18]. As the sequence of human MCT1 was known [7], ESTs (most of which are from human sources) encoding human MCT1 could be easily identified and were not pursued further. Sequences for putative novel MCTs mostly corresponded to sequences from the 5′ ends of clones and were used to search for overlapping sequences in dbEST that were not identified in the initial searches. In addition, in many cases it was also possible to identify additional sequences derived from 3′ untranslated regions by searching dbEST with clone identification numbers. All the overlapping sequences were aligned and assembled into contigs by using the SEQMAN package (DNAStar Lasergene software). These contigs seemed to be derived from at least four new MCT-related sequences distinct from those of MCT1 and MCT2, and thus we have named them MCT3, MCT4, MCT5 and MCT6. A single EST was also identified that seemed to encode the human isologue of hamster and rat MCT2. This sequence of 336 nt (accession number F01173) shows 77% identity with the nucleotide sequence of hamster MCT2 [11].

Generation of cDNA probes for the isolation of new MCTs

Probes for screening cDNA libraries were generated by PCR. Regions close to the 5′ end of the available sequence were chosen, to maximize the chances of obtaining full-length clones from the oligo(dT)-primed cDNA libraries. Where possible, regions containing at least two separate identical EST sequences were selected for the generation of exact-match oligonucleotides, as the original EST sequence data were not available for verification. Primer pairs were designed with the aid of the Primer Select package (DNAStar Lasergene software) to generate a distinct product of 200–500 bp for each MCT. The identity of the PCR products was confirmed by direct sequencing on an Applied Biosystems 373A sequencer with one of the original PCR oligonucleotides as a primer in each case. For each MCT the appropriate PCR primer pair was used to screen for a suitable cDNA library that contained clones with the desired cDNA, with approx. 10^5 clones per reaction. The libraries selected were human placenta (MCT4 and MCT5) and circulating blood (MCT3 and MCT6).

Screening of cDNA libraries from human placenta and circulating blood

Labelled probes for the MCTs were made by randomly priming the relevant amplified fragments of the cDNA with [α-^{32}P]dCTP and purification by gel filtration. Positive clones were isolated by screening approx. 10^6 recombinant phages and were plaque-purified by two further rounds of screening. The insert sizes of the clones were determined by PCR with primers that flanked the cloning site (λgt11 forward and reverse primers for the placental library, or T3 and T7 primers for the blood library). The MCT3 and MCT6 clones were excised from the circulating-blood λ ZAP Express library into pBK-CMV as described by the manufacturer. Sequencing the plasmid DNA with T3 and T7 primers confirmed the identity of the inserts. For the λgt11 clones (MCT4 and MCT5) the template for sequencing was prepared by pooling five separate PCR reactions (rather than a single cloned PCR product that could contain *Taq* polymerase-generated errors) with λgt11 forward and reverse primers. Sequencing the inserts with these two primers confirmed the identity of the clones. The complete sequences of the four MCTs were determined on both strands by automated sequencing with a custom primer walking strategy.

Northern blotting

Commercial human multi-tissue blots (Clontech) were probed sequentially with a portion of the coding region of relevant MCT cDNA in accordance with the manufacturer's instructions. For MCTs 3–6 the same probes as used in the library screening described above were employed. For MCT1 and XPCT suitable exact match probes (approx. 400 bp in length) were generated by PCR from human placental cDNA. For MCT2, where no complete human sequence is currently available, cDNA corresponding to the entire coding region of hamster MCT2 was used. Probes were labelled with [α-^{32}P]dCTP by random priming. After hybridization and washing as described by the manufacturer, the blot was subjected to autoradiography. Blots were then stripped of probe and washed before reprobing with the next MCT probe or finally with a β-actin probe.

RESULTS

Isolation of cDNA species encoding four novel mammalian MCT isoforms

Using the strategy outlined in the Experimental section, four new MCT-related sequences, clearly distinct from MCT1, MCT2 and XPCT, were identified in human placental or circulating-blood cDNA libraries. The MCT3 and MCT6 cDNA species isolated from the circulating-blood library both seemed to contain the full coding sequences. Both sequences contain in-frame termination codons upstream of the likely start codons. For MCT4 and MCT5, comparison of the translated assembled sequences with those of MCT1 and MCT2 suggested that neither of them was full-length at their 5′ end. To find the missing 5′ ends, PCR was used in a form of 5′ rapid amplification of cDNA ends by using λgt11 primers in combination with nested antisense primers generated against each of the two cDNA sequences, with the cDNA library as template. The resultant PCR products from the second nested PCR reaction were cloned and sequenced. For both MCT4 and MCT5, clones containing a further correct 5′ sequence were obtained. The presence of two in-frame ter-

Table 1 Known and putative MCT-related sequences from eukaryotes and prokaryotes

Abbreviations: CHO, Chinese hamster ovary; ORF, open reading frame.

Name (accession number)	Organism (source)	Notes	References
MEV (A44458)	CHO (met-18b-2 cells)	Mutant form of MCT1	[5]
MCT1 (Q03064)	CHO		[5]
MCT1	Rabbit (erythrocytes)	N-terminal protein sequence only	[40]
MCT1 (A5568)	Human (heart)	MCT1 gene structure also reported	[7]
MCT1 (X86216)	Rat (skeletal muscle)		[9]
MCT1 (U62316)	Rat (intestine)		[41]
MCT1 (X82438)	Mouse (Ehrlich Lettré cells)		[8]
MCT2 (A55626, L31957)	Syrian hamster (liver)		[11]
MCT2 (X97455)	Rat (testis)		[10]
MCT2 (U62316)	Rat (brain)		*
MCT2 (F01173)	Human (skeletal muscle)	Partial EST sequence	–
MCT3 (U81800)	Human (circulating blood)		†
MCT4 (U79745)	Human (placenta)		†
MCT5 (U59299)	Human (placenta)		†
MCT6 (U79745)	Human (circulating blood)		†
MCT7 (XPCT) (P36021)	Human (foetal brain)		[12]
REMP (MCT3?) (U15685)	Chicken (retinal epithelial cells)		[14,15]
– (Z28221) (P36032) (S38065) (YKW1)	*S. cerevisiae*	Gene YKL221W	[42]
– (Z46483) (P53918) (S55136) (YNM5)	*S. cerevisiae*	Gene YNL125c	[43]
– (Z74861)	*S. cerevisiae*	ORF YOL119c	‡
– (Z75214)	*S. cerevisiae*	ORF YOR306c	§
– (Z78545)	*C. elegans*	M03B6.2	[44]
– (U29379)	*C. elegans*	Gene K05B2.5	[44]
– (U41105)	*C. elegans*	Gene T02G5.12	[44]
– (Z70206)	*C. elegans*	Gene F46G10.6	[44]
Lactate permease (P33231)	*E. coli*	Gene lctP	[35]
Lactate permease (H64110)	*Haemophilus influenzae*		[45]
Lactate permease (D50453)	*Bacillus subtilis*		[46]
Lactate permease (Y08256)	*Sulfolobus solfataricus*	ORF c01002	[47]
– (Y08256)	*Sulfolobus solfataricus*	ORF c01003	[47]

* Enerson, B. E., Zhdankin, O. Y. and Drewes, L. R. (1996), accession number U62316.

† This paper.

‡ Arino, J., Casamayor, A., Gamo, F. J., Gancedo, C., Lafuente, M. J., Aldea, M., Casas, C. and Herrero, E. (1996), accession number Z74861.

§ Cziepluch, C., Jauniaux, J. C., Kordes, E., Poirey, R., Pujol, A. and Tobiasch, E. (1996), accession number Z75214.

Table 2 Summary of key features of cDNA and translated protein sequences of MCT3–MCT6

The predicted number of TM domains was derived from hydropathy plots with the Kyte–Doolittle algorithm [24] with a window of nine residues and the TMPRED program [25]. Abbreviation: UTR, untranslated region.

	cDNA length (bp)			Translated protein sequence		
MCT isoform	Total	5′ UTR	3′UTR	Length	Molecular mass (kDa)	Predicted TM domains
MCT3	1982	62	522	465	49	11–12
MCT4	2529	182	883	487	54	10–12
MCT5	1720	60	141	505	55	12
MCT6	1572	165	474	523	58	10–12

mination codons before the first ATG in the MCT4 sequence suggested that the MCT4 clone was full-length. However, for MCT5 the predicted open-reading frame was not preceded by an in-frame termination codon but the putative initiation site was in a good consensus for initiation by the method of Kozak [19].

Sequence analysis of new MCTs

Full nucleotide sequences of all four new human MCT-related cDNAs are held in the EMBL database and the accession numbers are given in Table 1. In Table 2 we show the total size of each cDNA and of their 3′ and 5′ untranslated regions, and the lengths and molecular masses of the derived protein sequences. The protein sequences themselves are shown in Figure 1, where they are aligned with MCT1 (human), MCT2 (rat) and XPCT (human, subsequently referred to as MCT7). The percentage identities of the sequences with one another and their similarities (taking into account conservative substitutions) are summarized in Table 3.

We have analysed the sequences for potential glycosylation sites; the only suitable asparagine residues predicted to be in an extracellular loop (see below) and thus potential N-glycosylation sites are Asn-456 in MCT4, Asn-102 and Asn-102 in MCT5, and Asn-57 in MCT6. However, in all cases the predicted sizes of these loop regions suggest that they are too small for N-linked glycosylation to take place [20]. Thus, as in MCT1 and MCT2, it is unlikely that the new MCT isoforms are glycosylated. Indeed, for MCT1 we have confirmed this experimentally [8].

MCT3 might be the mammalian equivalent of chicken REMP (also named MCT3), with which it shares 67% identity at the protein level (76% similarity when homologous substitutions are taken into account). This is significantly greater than the 43% and 45% sequence identity that it shares with MCT1 and MCT2 respectively [16]. However, this degree of identity is lower than might be expected between species. Furthermore both the loop between TM6 and TM7 and the C-terminal cytoplasmic domain are substantially longer in chicken MCT3 than in the human isoform (by 56 and 21 residues respectively). Thus it is possible that another mammalian MCT, as yet unidentified, might exist that is more closely related to chicken MCT3.

The MCT4 cDNA contains an Alu sequence [21] within the 3′ untranslated region 135 bp downstream of the termination codon. This Alu sequence is most related to that of the Alu-Sb subfamily, which has a consensus length of 288 bp [21]. Interestingly, a similar Alu sequence from the Alu-S subfamily has been found in the 3′ untranslated region of the human anion transporter AE1 [22]. Indeed, such Alu sequences are estimated to constitute between 6 and 13% of the mass of the human genome [21]. When compared with other members of the MCT family, it is evident that the C-terminal tail of the MCT4 protein is much shorter, only eight residues compared with between 58 and 86 residues in length. Thus it is possible that the original Alu sequence insertion event interrupted the coding sequence, by introducing an insertion with an internal premature termination codon, an in-frame insertion or an insertion causing an additional frame-shift in the original C-terminal tail. In the latter case it is possible that the MCT4 protein has become truncated during evolution, so as to retain its function. The proposed 5′ untranslated region of MCT4 contains five short (2–15 codons)

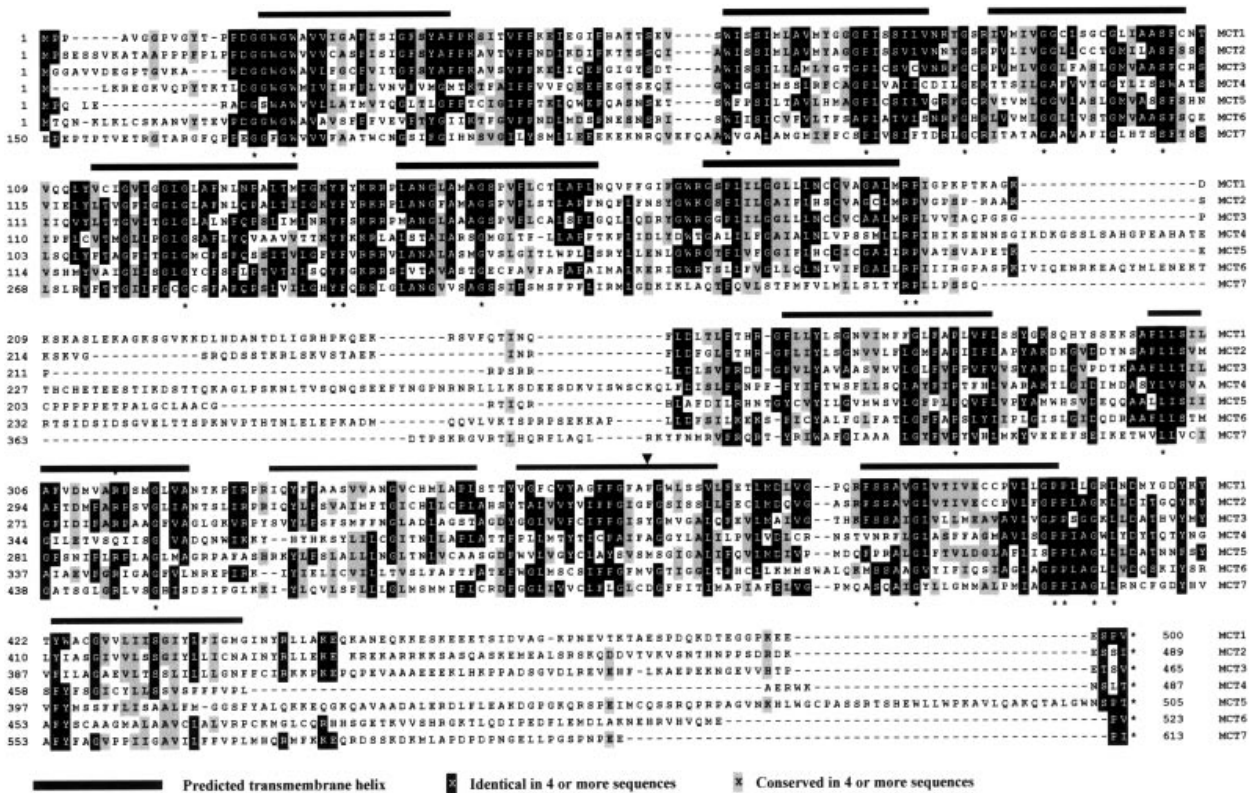

Figure 1 Amino acid sequences of the new mammalian MCT isoforms

The sequences for MCT1, MCT2 and MCT7 (XPCT) are taken from the sources shown in Table 1; the sequences for the new MCTs (3–6) are reported in the present paper. All sequences shown are from human sources except for MCT2, which is the rat sequence [10]. Sequences were aligned with the ClustalW algorithm by using MEGALIGN software (DNAStar Lasergene). Residues identical in four or more sequences are highlighted as white on black, with those identical in all sequences being marked with an asterisk; residues conserved in at least four sequences are highlighted black on grey. Mutation of Phe-360 (▼) in TM10 of Chinese hamster ovary MCT1 converts the protein into a mevalonate transporter (pMev) [5,6]. It should be noted that for MCT7 (XPCT) the first 150 amino acid residues (including the PEST domain) are not shown. The predicted TM segments are indicated by bars.

overlapping open reading frames. Such minicistrons are also present in the upstream region of human Na^+/H^+ exchanger NHE-1 and are known to be inhibitory to translational efficiency [23]. Thus it seems possible that MCT4 might be subject to translational regulation.

Predicted membrane topology

Hydropathy plots with the Kyte–Doolittle algorithm [24] with a window of nine residues and the TMPRED program [25] have been used to predict the number of TM α-helical domains; results are summarized in Table 2. In all cases between 10 and 12 TM domains are predicted, as for MCT1 and MCT2 [5,11]. When all sequences are taken together it seems most probable that there are 12 TM domains with the N-terminus and C-terminus located within the cytoplasm. For MCT1 we have tested this topological prediction experimentally by using proteolytic digestion and labelling techniques, and the results fully support the prediction [26,27]. The sequence alignment (Figure 1) clearly shows that it is in the putative TM regions and the shorter loop regions between them that the greatest sequence conservation is observed. In contrast the hydrophilic regions of the sequences (the N-terminal residues preceding TM1, the loop region between TMs 6 and 7, and the C-terminal residues after TM12) show little conservation. Indeed the size of the loop region between TM6 and TM7 varies substantially from 105 and 93 residues in MCT4 and MCT6, through 67, 49, 47 and 40 residues in MCT1, MCT2, MCT5 and MCT7, to only 29 residues in MCT3.

Northern blots to show the distribution of MCT isoforms in human tissues

The expression in a range of human tissues of mRNA encoding the seven MCT isoforms was investigated by Northern blotting, with actin being used to give an indication of mRNA loading. Results are shown in Figure 2. With the exception of MCT2, mRNA of the expected size for each isoform was observed, although there are clear differences in the tissue distribution of each. These results confirm that the newly cloned MCT isoforms are likely to be real entities, although the extent of their expression at the protein level must await the development of appropriate antibodies. The apparent lack of MCT2 mRNA in any tissue was confirmed on two separate blots and could not be attributed to the use of an inappropriate probe because it has been shown to detect MCT2 mRNA in Northern blots of mRNA extracted from a variety of tissues from rat, mouse and hamster [10]. We have reported previously that there are considerable species differences in the expression of MCT2 [10], and it seems that the expression of MCT2 in human tissues might be very limited. Indeed, after searching dbEST we detected only one putative MCT2 sequence fragment of 336 nt (accession number F01173)

Table 3 Similarities and identities between the different mammalian MCT isoforms

Percentage identities and similarities were calculated with the Wisconsin Package Version 9.0 (Genetics Computer Group, Madison, WI, U.S.A.), which uses the alignment method of Needleman and Wunsch [48].

Percentage similarity

Percentage identity

	MCT1	MCT2	MCT3	MCT4	MCT5	MCT6	MCT7
MCT1	-	65.4	52.6	37.4	41.1	39.4	35.6
MCT2	56.9	-	53.3	38.5	42.4	40.6	35.4
MCT3	45.0	44.3	-	37.3	47.4	42.0	37.4
MCT4	25.3	29.9	27.1	-	37.8	34.1	36.6
MCT5	30.2	31.1	38.2	27.2	-	38.9	35.4
MCT6	29.6	30.7	33.2	27.0	29.4	-	33.5
MCT7	25.22	25.0	26.5	24.7	25.3	21.0	-

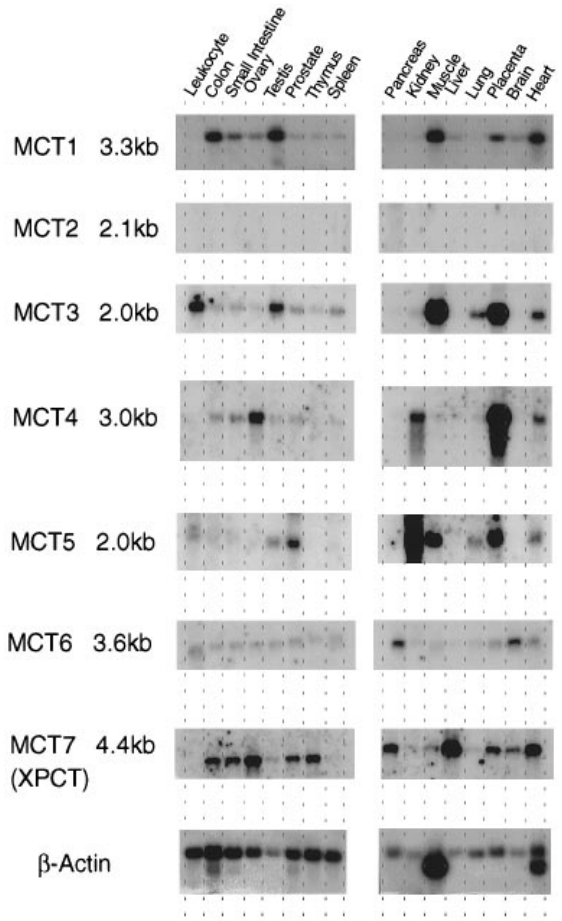

Figure 2 Northern blots showing the tissue distribution of MCT isoforms

The two human multi-tissue Northern blots (Clontech) were probed in accordance with manufacturer's recommendations as described in the Experimental section. The sizes of the MCT mRNA species were estimated by comparison with the position of migration of RNA standards already present on the blot.

that showed 77% identity with hamster MCT2 over the 336 nt, whereas many ESTs derived from a wide range of cDNA libraries were detected for all the other isoforms.

DISCUSSION

A family of mammalian MCTs

The results we present in this paper demonstrate that in human tissues there are at least seven MCT-related mRNAs expressed. It is highly probable that each sequence represents a different MCT isoform because computer analysis of protein sequences within the known transporter families/superfamilies has shown that proteins showing a high degree of sequence similarity throughout their lengths usually serve the same function [28]. However, care needs to be exercised with less closely related members of the family: in several cases, transporters that mediate import or export of structurally unrelated compounds show some degree of relatedness in sequence to one another [29]. Thus it is possible that some of the less related members of the MCT family might actually be involved in the transport of unrelated molecules sharing a similar transport mechanism. Ultimately, expression and kinetic characterization or other unequivocal biochemical identification is the only way to confirm that a sequence does indeed encode an MCT. So far only mammalian MCT1 and MCT2 and chicken MCT3 have been expressed in an active form [6,11,16] and it is clearly important to find a heterologous expression system in which to characterize fully the kinetics and the substrate and inhibitor specificities of all the new isoforms. This is not a trivial task because heterologous expression of membrane proteins in an active form and in the right cellular compartment often proves to be difficult. With MCT expression there is an added level of complexity, because all mammalian cell lines that we have investigated have endogenous monocarboxylate transport activity, including the human breast carcinoma cell line used by Kim-Garcia et al. [6] to express MCT1 activity. Furthermore the measurement by others [6,11,16] of a modest increase in [^{14}C]pyruvate accumulation in cells expressing the desired MCT isoform might in part reflect altered pyruvate metabolism rather than transport [1].

The Northern blots presented in Figure 2 indicate that each MCT isoform is likely to have a unique pattern of expression, and that several isoforms might be expressed in a single tissue. This is already known to be true in heart, from functional studies [3,4], and in the liver through the use of specific antibodies [10]. It is most likely that individual isoforms possess different monocarboxylate specificities or are reserved primarily for either import or export of monocarboxylates. For example, the high level of expression of MCT3 in skeletal muscle strongly suggests that this is the isoform that is the major lactate transporter in this tissue, a role that the low expression of MCT1 and MCT2 cannot fulfil [30]. Another possibility is that the different isoforms have different preferences for the co-transported cation, e.g. Na^+ as opposed to H^+. Thus it is known that the Na^+/glucose transporter can act as a H^+/glucose transporter with an affinity for protons 1000-fold greater than for Na^+, although at the prevailing ion concentration it is the Na^+-driven process that predominates [31]. MCT5 would be a suitable candidate for such a Na^+-linked transporter because its mRNA expression is particularly high in kidney, which is known to possess a very active Na^+-linked MCT [32]. It is possible that there are more MCT isoforms to be discovered, but their apparent absence from dbEST makes it unlikely that they will be major isoforms. It might be significant that at present only seven members of the GLUT family of glucose transporters are known [33].

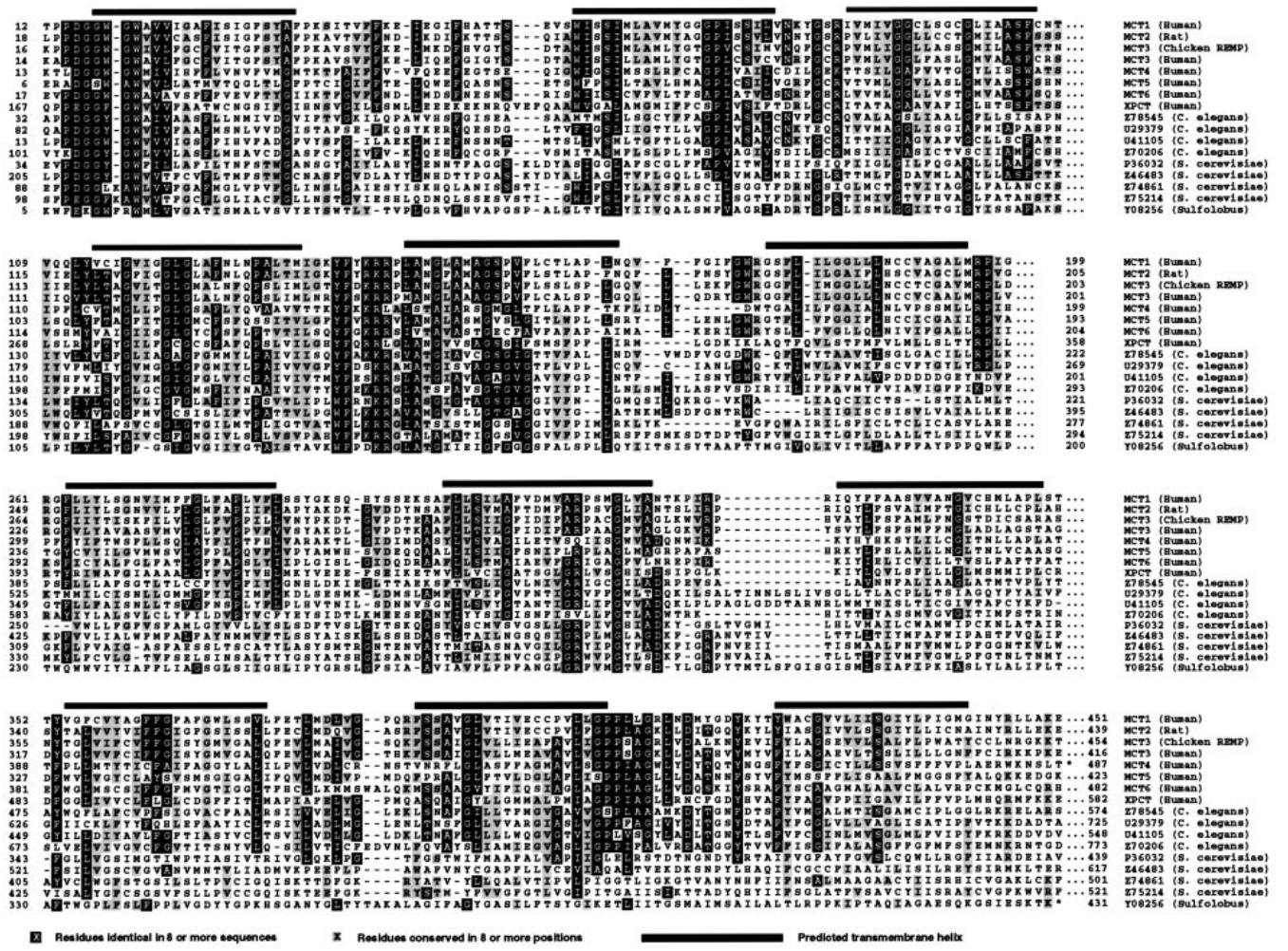

Figure 3 Alignment of TM segments of the wider family of MCT-related sequences

Sequences were aligned with the ClustalW algorithm by using MEGALIGN software (DNAStar Lasergene). Residues identical in eight or more of the sequences are highlighted as white on black; those conserved in at least eight sequences are highlighted as black on grey. The predicted TM segments of MCT1 are indicated as black bars.

The MCT family has a long evolutionary history: MCT homologues from lower eukaryotes and prokaryotes

Searching the protein sequence databases reveals ORFs for four potential transporters from *Saccharomyces cerevisiae* and four from *Caenorhabditis elegans* (see Table 1), although none of these sequences has been heterologously expressed to confirm their MCT activity. As the yeast genome-sequencing project has now been completed [34] we know that four represents the full set of putative MCTs for this species. In a unicellular organism such as *S. cerevisiae*, different MCT isoforms cannot reflect kinetic adaptation for tissue-specific requirements and are therefore more likely to possess different substrate specificities. The sequence of a putative MCT from the archaebacterium *Sulfolobus solfataricus* (see Table 1) also shows similarity to the eukaryotic MCT family with 36% similarity and 25% identity with human MCT1 over the first 200 residues. Interestingly, this sequence is directly adjacent in the genome to a sequence encoding another putative membrane protein with homology (32% identity, 58% similarity) to the *Escherichia coli* lactate permease (lctP) [35]. However, lctP shows only 12% similarity to the eukaryotic MCT family member sequences and thus it is unlikely that the mammalian plasma membrane MCTs evolved from the prokaryotic lactate permease.

Comparison of all known MCT-related sequences and the significance of conserved regions

An alignment of all 17 MCT sequences described above is shown in Figure 3. All these sequences are consistent with a topology of 12 TM segments, with the hydrophilic regions of the sequences (the N-terminal residues preceding TM1, the loop region between TMs 6 and 7, and the C-terminal residues after TM12) showing little conservation. The alignment shown in Figure 3 excludes these regions. Such divergent hydrophilic regions are a common feature of 12-TM transporter family sequences [28], and it is unlikely that these regions are directly involved in transport. Rather they might be critical for other aspects of function such as substrate specificity or regulation of transport activity. A cladogram to show the possible evolutionary history of the MCT family derived from the data of Figure 3 is shown in Figure 4. This suggests that mammalian MCT1, MCT2, MCT3, MCT5 and MCT6 form a subgrouping that includes chicken MCT3, whereas MCT4 is more closely related to the four *C. elegans*

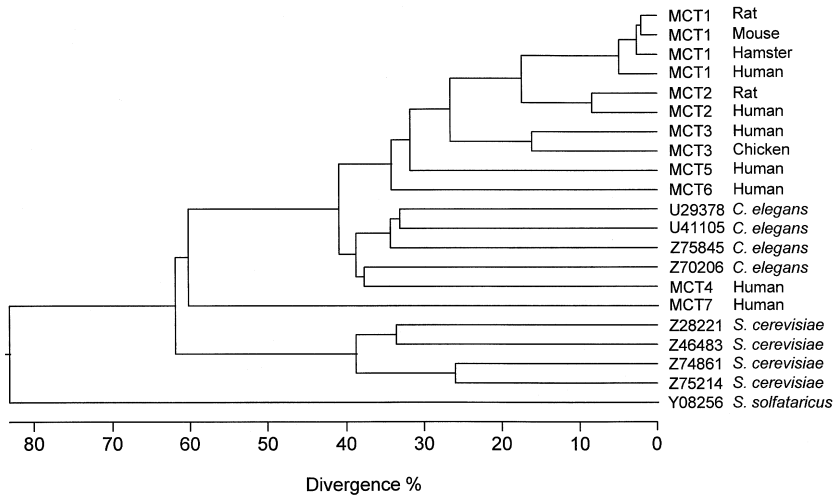

Figure 4 Cladogram representing the proposed evolutionary history of the MCT family

The sequences shown in Figure 3, together with those for hamster [5], mouse [8] and rat [9] MCT1 and hamster MCT2 [11], were aligned with the ClustalW algorithm by using MEGALIGN software (DNAStar Lasergene), which calculated the relatedness of the sequences to one another and displayed the results as the cladogram shown.

MCTs. MCT7 seems to be the result of an earlier evolutionary divergence, as would the four putative MCTs found in *S. cerevisiae*, whereas the single MCT found in *Sulfolobulus solfataricus* seems to be a more distant relative.

It is noteworthy that the highly conserved motifs are all predicted to be in regions that include part of a loop and the start or end of a TM helix. Glycine residues comprise a large proportion of these conserved residues, which presumably reflects an important structural role. They are likely to be involved in the formation of turns between TM segments, the packing of helices and the provision of flexibility for conformational changes [28]. Similarly, the conserved proline and hydrophobic residues are likely to have structural significance. The conserved charged and hydrophilic residues might have a catalytic role [28] as discussed below.

In TMs 1–6 there are four residues that are absolutely conserved across all 17 sequences (Figure 3), and many more positions at which only conservative substitutions occur. Two highly conserved regions stand out in particular: the sequence with consensus P[**D**/**E**]G[**G**/**S**]W[G/A]**W**V[V/I]V, which traverses the lead into TM1, and the consensus [**Y**/**W**]FXK[**R**/**K**]-[**R**/**L**]XLAX[**G/A/S**]XAXA**G**, which leads into TM5 (residues shown in bold are conserved in all of the sequences, residues in square parentheses indicate alternative amino acids, and residues in normal type are the consensus amino acid at that position). The C-terminal half of the molecule (TMs 7–12) shows a slightly lower level of conservation with only a single invariant residue. TM 11 and the loop regions on each side are most highly conserved in this region, with the motif LXGPPXXGXLXD on the proposed exofacial boundary of TM 11 standing out. In general, N-terminal domains of different members of transporter families are more strongly conserved than their C-terminal domains, and thus the MCT family fits this pattern [28]. The two halves of the molecule (TM helices 1–6 and 7–12) have been proposed to have different functional roles, the N-terminal domains being more important for energy (e.g. H^+ or Na^+) coupling, membrane insertion and/or correct structure maintenance, whereas the C-terminal domains are probably more important for substrate specificity determination [28]. This phenomenon is supported by inhibitor and photo-affinity labelling studies on sugar transporters [33] and by the characteristics of chimaeras constructed with regions from different glucose transporter isoforms [36]. Thus it might be significant that in the MCT family the only conserved aspartic/glutamic residue, a possible candidate for proton translocation, is at the start of TM1 (Asp-16 of MCT1). There is also evidence that the C-terminal half of MCT1 is involved in substrate specificity. Thus the conversion of Phe-360 in TM 10 of Chinese hamster MCT1 (indicated by the arrow in Figure 1) to Cys changes MCT1 from a lactate/pyruvate transporter to a mevalonate transporter [5,6]. Furthermore the binding site of MCT1 for 4,4′-di-isothiocyanostilbene-2,2′-disulphonate, which binds at or near the substrate binding site [37], is in the C-terminal half of the transporter [26].

A positively charged group that binds the CO_2^- anion is a feature that is likely to be present in all MCTs. In erythrocyte MCT1 arginine probably fulfils this role because the arginine-specific reagent phenylglyoxal inhibits MCT1-mediated lactate transport [1,38]. This is also true in lactate dehydrogenase [39]. An arginine residue in TM8 (Arg-313 of human MCT1) is conserved in all the putative MCTs from higher eukaryotes except MCT4, and also in all but one of the *S. cerevisiae* and *C. elegans* sequences (see Figure 3). It is also in the C-terminal half of the molecule thought to be involved in substrate binding. The only other Arg/Lys residue that is totally conserved in all of these sequences is present between TMs 4 and 5. This residue is within the relatively highly conserved region [**Y**/**W**]FXK[**R**/**K**] [**R**/**L**]XLAX[**G/A/S**]XAXA**G** mentioned above.

This work was supported by project grants from the British Heart Foundation, the Wellcome Trust and a University postgraduate scholarship (to V.N.J.).

REFERENCES

1 Poole, R. C. and Halestrap, A. P. (1993) Am. J. Physiol. **264**, C761–C782
2 Jackson, V. N. and Halestrap, A. P. (1996) J. Biol. Chem. **271**, 861–868
3 Wang, X.-M., Levi, A. J. and Halestrap, A. P. (1994) Am. J. Physiol. **267**, H1759–H1769
4 Wang, X.-M., Levi, A. J. and Halestrap, A. P. (1996) Am. J. Physiol. **270**, H476–H484

5 Kim, C. M., Goldstein, J. L. and Brown, M. S. (1992) J. Biol. Chem. **267**, 23113–23121
6 Kim-Garcia, C., Goldstein, J. L., Pathak, R. K., Anderson, R. G. W. and Brown, M. S. (1994) Cell **76**, 865–873
7 Garcia, C. K., Li, X., Luna, J. and Francke, U. (1994) Genomics **23**, 500–503
8 Carpenter, L., Poole, R. C. and Halestrap, A. P. (1996) Biochim. Biophys. Acta **1279**, 157–163
9 Jackson, V. N., Price, N. T. and Halestrap, A. P. (1995) Biochim. Biophys. Acta **1238**, 193–196
10 Jackson, V. N., Price, N. T., Carpenter, L. and Halestrap, A. P. (1997) Biochem. J. **324**, 447–453
11 Garcia, C. K., Brown, M. S., Pathak, R. K. and Goldstein, J. L. (1995) J. Biol. Chem. **270**, 1843–1849
12 Lafrenière, R. G., Carrel, L. and Willard, H. F. (1994) Hum. Mol. Genet. **3**, 1133–1139
13 Rechsteiner, M. and Rogers, S. W. (1996) Trends Biochem. Sci. **21**, 267–271
14 Rogers, S., Wells, R. and Rechsteiner, M. (1986) Science **234**, 364–368
15 Philp, N., Chu, P., Pan, T., Zhang, R. Z., Chu, M., Stark, K., Boettiger, D., Yoon, H. and Kieber-Emmons, T. (1995) Exp. Cell Res. **219**, 64–73
16 Yoon, H., Fanelli, A., Grollman, E. F. and Philp, N. J. (1997) Biochem. Biophys. Res. Commun. **234**, 90–94
17 Boguski, M. S., Tolstoshev, C. M. and Bassett, Jr., D. E. (1994) Science **265**, 1993–1994
18 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
19 Kozak, M. (1989) Mol. Cell. Biol. **9**, 5073–5080
20 Landolt-Marticorena, C. and Reithmeier, R. A. F. (1994) Biochem. J. **302**, 253–260
21 Schmid, C. W. (1996) Prog. Nucleic Acid Res. **53**, 283–319
22 Tanner, M. J. A. (1993) Semin. Hematol. **30**, 34–57
23 Fliegel, L. and Dyck, J. R. B. (1995) Cardiovasc. Res. **29**, 155–159
24 Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. **157**, 105–132
25 Hofmann, K. and Stoffel, W. (1993) Hoppe-Seyler's Z. Biol. Chem. **347**, 166
26 Poole, R. C., Sansom, C. E. and Halestrap, A. P. (1996) Biochem. J. **320**, 817–824
27 Poole, R. C. and Halestrap, A. P. (1997) J. Biol. Chem. **272**, 14624–14628
28 Saier, M. H. (1994) Microbiol. Rev. **58**, 71–93
29 Griffith, J. K., Baker, M. E., Rouch, D. A., Page, M. G. P., Skurray, R. A., Paulsen, I. T., Chater, K. F., Baldwin, S. A. and Henderson, P. J. F. (1992) Curr. Opin. Cell. Biol. **4**, 684–695
30 Juel, C. (1997) Physiol. Rev. **77**, 321–358
31 Hirayama, B. A., Loo, D. D. F. and Wright, E. M. (1994) J. Biol. Chem. **269**, 21407–21410
32 Wright, E. M. (1985) Annu. Rev. Physiol. **47**, 127–141
33 Gould, G. W. and Holman, G. D. (1993) Biochem. J. **295**, 329–341
34 Clayton, R. A., White, O., Ketchum, K. A. and Venter, C. J. (1997) Nature (London) **387**, 459–462
35 Dong, J. M., Taylor, J. S., Latour, D. J., Iuchi, S. and Lin, E. C. (1993) J. Bacteriol. **175**, 6671–6678
36 Arbuckle, M. I., Kane, S., Porter, L. M., Seatter, M. J. and Gould, G. W. (1996) Biochemistry **35**, 16519–16527
37 Poole, R. C. and Halestrap, A. P. (1992) Biochem. J. **283**, 855–862
38 Carpenter, L. and Halestrap, A. P. (1994) Biochem. J. **304**, 751–760
39 Hart, K. W., Clarke, A. R., Wigley, D. B., Waldman, A. D. B., Chia, N., Barstow, D. A., Atkinson, T., Jones, J. B. and Holbrook, J. J. (1987) Biochim. Biophys. Acta **914**, 294–298
40 Poole, R. C. and Halestrap, A. P. (1994) Biochem. J. **303**, 755–759
41 Takanaga, H., Tamai, I., Inaba, S., Sai, Y., Higashida, H., Yamamoto, H. and Tsuji, A. (1995) Biochem. Biophys. Res. Commun. **217**, 370–377
42 Alexandraki, D. and Tzermia, M. (1994) Yeast **10**, S81–S91
43 Mallet, L., Busseraeu, F. and Jacquet, M. (1995) Yeast **11**, 1195–1209
44 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J. et al. (1994) Nature (London) **368**, 32–38
45 Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. et al. (1995) Science **269**, 496–512
46 Yamane, K., Kumano, M. and Kurita, K. (1996) Microbiology **142**, 3047–3056
47 Sensen, C. W., Klenk, H. P., Singh, R. K., Allard, G., Chan, C. C. Y., Liu, Q. Y., Penny, S. L., Young, F., Schenk, M. E., Gaasterland, T. et al. (1996) Mol. Microbiol. **22**, 175–191
48 Needleman, S. B. and Wunsch, E. (1970) J. Mol. Biol. **48**, 443–453

Received 20 June 1997/21 August 1997; accepted 18 September 1997