Evolution of alanine:glyoxylate aminotransferase intracellular targeting: structural and functional analysis of the guinea pig gene

Graeme M. BIRDSEY and Christopher J. DANPURE¹

MRC Laboratory for Molecular Cell Biology, University College London, Gower Street, London WC1E 6BT, U.K.

The distribution of alanine:glyoxylate aminotransferase 1 (AGT) within liver cells has changed many times during mammalian evolution. Depending on the particular species, AGT can be found in mitochondria or peroxisomes, or mitochondria and peroxisomes. In some cases significant cytosolic AGT is also present. In the livers of most rodents, AGT has what is thought to be the more 'ancestral' distribution (i.e. mitochondrial and peroxisomal). However, AGT is distributed very differently in the guinea pig, being peroxisomal and cytosolic. In this study, we have attempted to determine the molecular basis for the loss of mitochondrial AGT targeting and the apparent inefficiency of peroxisomal targeting of AGT in the guinea pig. Our results show that the former is owing to the evolutionary loss of the

more 5' of two potential transcription and translation initiation sites, resulting in the loss of the ancestral N-terminal mitochondrial targeting sequence from the open reading frame. Guinea pig AGT is targeted to peroxisomes via the peroxisomal targeting sequence type 1 (PTS1) peroxisomal import machinery, even though its C-terminal tripeptide, HRL, deviates from the standard consensus PTS1 motif. Although HRL appears to target AGT to peroxisomes less efficiently than the classical PTS1 SKL, the main reason for the low efficiency of AGT peroxisomal targeting in guinea pig cells (compared with cells from other species) lies not with guinea pig AGT but with some other, as yet undefined, part of the guinea pig peroxisomal import machinery.

INTRODUCTION

The intermediary metabolic enzyme alanine:glyoxylate aminotransferase 1 (AGT, EC 2.6.1.44) is unusual in that it is targeted to different intracellular locations in different species. In the hepatocytes of some animals, such as humans and rabbits, AGT is exclusively peroxisomal; in others, such as cats, it is mainly mitochondrial; in yet others, such as marmosets and rats, AGT is distributed more evenly between both peroxisomes and mitochondria. In some species, such as the guinea pig and frog, significant amounts of AGT are also found in the cytosol as well as in the peroxisomes (in the case of the guinea pig) or mitochondria (in the case of the frog) [1]. The subcellular distribution of AGT appears to have changed radically (i.e. it has lost mitochondrial targeting or acquired peroxisomal targeting) on at least six (probably many more) occasions during the evolution of mammals.

It has been suggested that the organellar distribution of AGT in vertebrates is under the influence of dietary selection pressure [2]. Thus, the tendency is for AGT to be peroxisomal in herbivores, mitochondrial in carnivores and both peroxisomal and mitochondrial in omnivores. The peroxisomal localization of AGT in human liver is presumably a reflection of our herbivorous ancestry rather than current diet. It has also been suggested that the dual distribution of AGT might be related to its putative dual metabolic role of gluconeogenesis (in the mitochondria) and glyoxylate detoxification (in the peroxisomes). For reasons discussed previously, it is likely that the former is the principal role of AGT in carnivores and the latter is its main role in herbivores [1,2].

The single AGT gene has been cloned and functionally characterized in the human, marmoset, rabbit, cat and rat [3–9].

Such studies have shown that the archetypal AGT gene has the potential to encode an N-terminal mitochondrial targeting sequence (MTS) and a C-terminal peroxisomal targeting sequence type 1 (PTS1), albeit of a non-standard variety [10], by the variable use of two transcription and two translation initiation sites [11]. The final intracellular destination of AGT is dependent upon the expression of the MTS rather than the PTS1, the former being functionally dominant over the latter [7]. The MTS has been lost by permanent exclusion from the open reading frame on at least two temporally separated occasions in the human and rabbit ancestral lines due to mutations in the more 5' of the two potential translation initiation codons (ATG → ATA in the human, and ATG → ACA in the rabbit) [3,4].

The ability to target AGT to the mitochondria has also been lost in the guinea pig, so that the organellar distribution is exclusively peroxisomal [1,12]. This is surprising because in other rodents, such as rats, mice, hamsters, ground squirrels and porcupines, AGT is targeted to both peroxisomes and mitochondria [1,12–16]. Loss of mitochondrial AGT targeting in the guinea pig ancestral line presumably occurred after the separation of the caviomorph and non-caviomorph rodent branches, making the event temporally distinct from those that led to the loss of mitochondrial targeting in the human and rabbit ancestral lines.

In most species, AGT targeting appears to be efficient, irrespective of which organelle it is targeted to. As a result little, if any, AGT is found in the cytosol of normal hepatocytes. The absence of cytosolic AGT is unlikely to be due to degradation of non-imported AGT, which in humans at least appears to be stable and catalytically active in the cytosol if peroxisomal import is prevented by a genetic defect in the peroxisomal import machinery [17]. However, immuno-electron microscopy studies have suggested that in guinea pig liver significant amounts of

Abbreviations used: AGT, alanine:glyoxylate aminotransferase 1; MTS, mitochondrial targeting sequence; PTS1, peroxisomal targeting sequence type 1; RACE, rapid amplification of cDNA ends; RAGE, rapid amplification of genomic ends; RT, reverse transcription (PCR).

¹ To whom correspondence should be addressed.

The nucleotide sequence described in this paper has been deposited in the EMBL database under accession number Y10727.

immunoreactive AGT are normally cytosolic, possibly implying relatively inefficient peroxisomal import compared with that in other species [1]. Why the distribution of AGT in the guinea pig liver (i.e. peroxisomes and cytosol) should be so different to that in other rodents (i.e. mitochondria and peroxisomes) is currently unknown.

The main aims of the present study were to determine the molecular basis of the loss of mitochondrial targeting and the apparent inefficiency of peroxisomal targeting of guinea pig AGT. The results presented in this paper enable us to draw the following conclusions: (1) guinea pig AGT is not targeted to the mitochondria owing to the combined effects of loss of the first ancestral translation start site (similar to that found in the human and rabbit) and loss of the first ancestral transcription start site (unlike that in the human and rabbit); and (2) the apparent inefficiency of guinea pig AGT peroxisomal import (compared with that in other species) results mainly from characteristics of the guinea pig peroxisomal import machinery rather than any intrinsic inability of guinea pig AGT to target efficiently.

EXPERIMENTAL

General molecular biology methods

Unless otherwise indicated, standard methods were used for DNA cloning, sequencing, Southern and Northern blotting, cDNA library screening, reverse transcription (RT), PCR, and RNase protection. Guinea pig liver total RNA was isolated using the QuickPrep Total RNA Extraction kit (Pharmacia Biotech). Guinea pig liver mRNA was isolated from total RNA using the PolyAtract mRNA Isolation System (Promega). Guinea pig genomic DNA was extracted from frozen liver using the Nucleon ST DNA Extraction kit (Nucleon Biosciences). Radiolabelled probes were prepared by random priming using the Megaprime DNA labelling system (Amersham Life Science). DNA dideoxy

sequencing was performed using Sequenase Version 2.0 T7 DNA polymerase (Amersham Life Science). Cycle sequencing was carried out with the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer).

cDNA library screening

To produce a homologous AGT probe for library screening, cDNA from guinea pig liver was made by RT of guinea pig liver total RNA using an oligo-dT primer. A 367 bp fragment of AGT cDNA was amplified by PCR using the degenerate primers P1 and P2 which map to highly conserved regions of the AGT gene in other mammalian species (see Table 1 for details of all the primers used in the present study and Figure 1 for a plan of the cloning strategy). This P1/P2 fragment was digested with *Pst*I to give a 75 bp fragment that was cloned into pBluescriptIIKS(+) (Stratagene) to give pCav1. From the latter clone a guinea pig specific primer (P3) was designed and used in combination with the human AGT primer P4 to PCR a 754 bp fragment which was blunt-ended and cloned into the *Eco*RV site of pBluescript to give pCav4.

A guinea pig liver cDNA library, cloned into the *Eco*RI site of λ gt11, was a gift of Dr. Y. Suzuki, Kinki University, Japan. Using *Escherichia coli* host strain Y1090, 250 000 plaques were screened. Duplicate nylon lifts were prepared and hybridized to the guinea pig AGT probe pCav4 (see Figure 1 and the Results section). Positive plaques were prepared as plate lysates and the bacteriophage DNA extracted and analysed by Southern blotting to confirm the presence of AGT inserts which were then subcloned into the *Eco*RI site of pBluescript.

Rapid amplification of cDNA ends (RACE)

The 5' RACE reactions were performed using the 5'/3' RACE kit (Boehringer Mannheim) [18]. The kit was used according to the manufacturer's instructions. First strand cDNA synthesis was

Table 1 Description of the oligonucleotide primers used in the present study

The underlining indicates the restriction sites used in the present study – P6, *Hind*III; P7, *Bam*HI; P12, *Bam*HI; the oligo d(T)- and PCR-anchor primers, *Cl*aI (solid underline) and *Sal*I (broken underline).

Primer	Sequence	Region to which primers map*	Strand
P1†	ATGTWYCARATHATGGAYGARATHAA	477–502	(+)
P2†	CKRTGRCANARYTCNCCRAANCCRTC	844–819	(–)
P3	GATCAAGGAGCCTGGAAGCT	698–717	(+)
P4†	CGTCACGCGGTCCACATTCT	1451–1432	(–)
P5	CAGGCACTGGTACCTGTGAC	857–838	(–)
P6	CAGTGTGTAGAAAGCTCCAG	728–708	(–)
P7†	CTGGAACACGTACTGGATCCCTTCCTTGAT	527–498	(–)
P8	ACACGGGGAGCCAGGTTGGAG	430–410	(–)
P9	ACACAGCTGGTAGGAGGC	343–324	(–)
P10	ACTCCGAGCCAGCTCCTATG	73–93	(+)
P11	TGAGGCCCGGGTGGCAGGCTG	289–310	(+)
P12	AGTACAGTGGATCCCTTCTC	1639–1620	(–)
RSP‡	CAGGAAACAGCTATGAC		
T3§	ATTAACCTCACTAAAG		
Oligo d(T)-anchor	GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTT		
PCR-anchor	GACCACGCGTATCGATGTCGAC		

* Nucleotide numbering system refers to that used in the guinea pig sequence shown in Figure 5.

† Primers P1, P2, P4 and P7 were originally designed to hybridize with human AGT cDNA [4,52].

‡ RSP = pBluescript reverse sequencing primer.

§ T3 = pBluescript T3 promoter primer.

Figure 1 Guinea pig cloning strategy (refer to text for details)

The uppermost horizontal line represents the 1739 bp composite guinea pig AGT nucleotide sequence shown in Figure 5. The triplets corresponding to the first (GTT) and second (ATG) ancestral translation initiation codons are shown, as is the stop codon (TAA). The (+)-strand (P10, P11, P1 and P3) and the (—)-strand PCR primers (P9, P8, P7, P6, P2, P5, P4 and P12) are indicated. The P1/P2 PCR product was never cloned and therefore not named. pCav1 was the cloned *Pst*I fragment of this PCR product. The zigzag line at the 5' end of the library clone pCav5 represents the intron 6 sequence referred to in the text. The dotted lines at the 3' ends of the 5'-RACE (pCav20 and pCav23) and RAGE (pCav25) clones represent the sequence amplified by the first round PCR that was not amplified by the second round semi-nested PCR. The (+)-strand anchor and plasmic primers used to amplify the RACE and RAGE products, respectively, are not shown in this diagram. The RNase-protected fragment of the riboprobe pCav28 is shown as a thicker line.

carried out using guinea pig liver total RNA with cDNA synthesis primer P5 (see Table 1 for a list of all the primers used in the present study). For purification of the cDNA the High Pure PCR Product Purification kit (Boehringer Mannheim) was used according to the manufacturer's instructions. After tailing of the 3' end of the cDNA with dATP, the dA-tailed cDNA was used as template in a PCR amplification reaction using an oligo-dT anchor primer and the guinea pig AGT-specific primer P5. Five per cent of this PCR product was used as template in a second round of PCR amplifications using a pair of nested primers (the PCR-anchor primer and a guinea pig AGT specific primer P6). This gave a product of 403 bp (including primers) which was cloned into the *Sal*I and *Hind*III sites of pBluescript, to give clone pCav20 (Figure 1). A second separate 5' RACE reaction was performed. Reverse transcription was carried out using primer P6. The first round PCR amplification of the dA-tailed cDNA used the oligo-dT anchor primer and the P6 primer. A second round of PCR was then carried out on 5% of the first round product using the PCR anchor primer and the AGT-specific primer P7. This gave a product of 288 bp which was gel purified and cloned into the *Bam*HI and *Cla*I sites of pBluescript, to give pCav23 (Figure 1).

Rapid amplification of genomic ends (RAGE)

The RAGE technique used was based on that of Mizobuchi and Frohman [19]. Guinea pig liver genomic DNA (5 µg) was digested with restriction enzyme *Pst*I (Boehringer Mannheim), extracted once each with phenol/chloroform (1:1) and with chloroform and then precipitated with ethanol. pBluescript (5 µg) was digested with *Pst*I and then dephosphorylated with calf intestinal alkaline phosphatase (Promega). The plasmid DNA was extracted and precipitated as described above. The *Pst*I-digested guinea pig genomic DNA and plasmid DNA were ligated with T4 DNA ligase (Gibco-BRL) at 16 °C overnight in a reaction

volume of 100 µl. After heat inactivation of the ligase (65 °C, 15 min), 2 µl of 500 mM EDTA were added and the mixture was stored at 4 °C until use.

PCR amplification was carried out using 1 µl of the ligation mixture as template. The 50 µl PCR reaction contained 25 pmol of Reverse Sequencing Primer (RSP) and 25 pmol of the AGT-specific primer P8. The DNA was heat denatured at 95 °C for 7 min. Subsequently, *Taq* DNA polymerase (2.5 units; Gibco-BRL) was added and the DNA was allowed to anneal at 50 °C for 2 min. The DNA strands were extended at 72 °C for 40 min, followed by 30 cycles of amplification (94 °C, 40 s; 50 °C, 1 min; 72 °C, 3 min) with a final extension at 72 °C for 15 min. A band of ~450 bp was gel purified and 1% was subjected to a second round of PCR amplification with a nested plasmid T3 promoter primer and a nested AGT-specific primer (P9) under the same conditions as described above. A 5 µl aliquot of the PCR product was analysed on a 1% (w/v) agarose gel and the presence of AGT sequence was detected following Southern blotting by hybridization to a guinea-pig-specific AGT probe (pCav23). The final RAGE product (343 bp) was digested with *Pst*I and *Eco*RV and cloned into pBluescript to give pCav25 (Figure 1).

Ribonuclease protection assay

A genomic clone suitable for RNase protection was obtained by PCR of guinea pig genomic DNA using the primers P10 and P8 (see Table 1 and Figure 1). The 358 bp PCR product was blunt-end cloned into the *Eco*RV site of pBluescript to produce pCav28. Following linearization of this clone with *Hind*III, an antisense riboprobe was generated from the T7 RNA polymerase promoter in the presence of [α -³²P]UTP (600 Ci/mmol). Total RNA (10 µg) prepared from guinea pig liver was hybridized for 16 h at 45 °C with the antisense riboprobe pCav28 (10⁵ c.p.m.). Following RNase digestion, the protected fragments were separated on a 6% denaturing polyacrylamide sequencing gel.

Cells

Normal human fibroblasts (155BR) and normal guinea pig fibroblasts (JH4 clone 1) were obtained from the European Collection of Animal Cell Cultures (Porton Down, U.K.). The human PTS1(—) fibroblasts (AAL85AD) were obtained from a patient with neonatal adrenoleukodystrophy [20]. Fibroblasts were used in the AGT transfection studies (see below) because endogenous AGT expression in these cells is undetectable.

Mammalian expression plasmids

To express guinea pig AGT in tissue culture cells, pCav26 was recloned into the mammalian expression vector pcDNA3 (Invitrogen). The preparation of the clones encoding normal human AGT and a mutant human AGT construct, in which the C-terminal KKL was replaced by SKL, in the expression plasmid pHYK has been described previously [10].

Intranuclear microinjection

Cells were grown in a 1:1 mixture of Dulbecco's modified Eagle's medium and Ham's F10 supplemented with 10% foetal calf serum (Gibco-BRL) at 37 °C under 5% CO₂. Twenty-four hours before micro-injection the cells were seeded at 50% confluence onto 13 mm diameter glass microscope coverslips. Expression plasmids were injected into the nuclei of at least 200 cells at a concentration of 200 µg/ml in reverse phosphate-buffered saline (4 mM Na₂HPO₄, 1 mM KH₂PO₄, 140 mM KCl, pH 7.3), using an Eppendorf microinjector. Following injection the cells were

incubated at 37 °C for 18–24 h and processed for indirect immunofluorescence microscopy [7].

Immunofluorescence microscopy

The cells were washed in PBS (0.8 % NaCl, 0.02 % KCl, 0.11 % Na_2HPO_4 , 0.02 % KH_2PO_4 , pH 7.2) and fixed in freshly prepared 3 % (w/v) paraformaldehyde for 15 min at room temperature, followed by permeabilization with 1 % Triton X-100 for 15 min at room temperature [7]. The cells were then processed for double-label indirect immunofluorescence using various combinations of the following polyclonal antibodies: rabbit anti-human AGT, rabbit anti-human catalase, rabbit anti-SKL peptide [21] and guinea pig anti-human AGT. Exogenous AGT, endogenous catalase and endogenous SKL-containing polypeptides (the latter two acting as peroxisomal markers) were visualized with fluorescein isothiocyanate-conjugated goat anti-(rabbit IgG) (Sigma) or biotinylated goat anti-(guinea pig IgG) (Vector Laboratories) followed by avidin-labelled Texas Red (Vector Laboratories). All incubations were performed at room temperature for 15 min in PBS containing 3 % BSA. The cells were washed extensively in PBS after each incubation. The cover slips were mounted on to glass slides using Mowiol (Harlow Chemical Co. Ltd.) containing diazabicyclo[2.2.2]octane (DABCO, Sigma). The fluorescent staining pattern was visualized in a Bio-Rad MRC1000 confocal laser-scanning microscope and the images were recorded on Kodak TMAX100 ASA B/W film.

Subcellular fractionation

Guinea pig liver (4 g) was minced and then homogenized in a Potter–Elvehjem apparatus to give a 20 % (w/v) homogenate in ice-cold 250 mM sucrose (containing 1 mM EDTA, 5 mM Mops, pH 7.4). The nuclei and cell debris were removed by centrifugation at 600 *g* for 10 min (H2000B rotor, Sorvall RC3C+ centrifuge). A sample (4 ml) of the postnuclear supernatant was layered on to a sucrose gradient (1.05–1.30 g/ml) and fractionated by isopycnic centrifugation at 72 400 *g* for 60 min at 4 °C (Sorvall TV860 rotor, Beckman L70 Ultracentrifuge). Of the remaining postnuclear supernatant, 10 ml were centrifuged at 31 000 *g* for 30 min at 4 °C (JA-25.50 rotor, Beckman Avanti J-25 centrifuge). The pellet was resuspended in 20 ml of ice-cold 250 mM sucrose (containing 1 mM EDTA, 5 mM Mops, pH 7.4) and 4 ml layered onto a Nycodenz gradient (1.03–1.24 g/ml) and centrifuged at 82 300 *g* for 70 min at 4 °C (Sorvall TV860 rotor, Beckman L70 Ultracentrifuge). Sixteen fractions of ~ 2 ml were collected from the bottom of both sucrose and Nycodenz gradients and were assayed for a variety of marker enzymes.

Enzyme assays and immunoblotting

The catalytic activities of AGT (EC 2.6.1.44) [22], D-amino acid oxidase (DAO, EC 1.4.3.3) [22], catalase (CAT, EC 1.11.1.60) [22], glutamate dehydrogenase (GDH, EC 1.4.1.2) [23], urate oxidase (UO, EC 1.7.3.3) [24] and lactate dehydrogenase (LDH, EC 1.1.2.3) [25] were assayed as described previously. Immunoreactive AGT and catalase in the gradient fractions were determined by SDS/PAGE on 12 % gels followed by immunoblotting using polyclonal rabbit anti-(human AGT IgG) and catalase antisera [1].

RESULTS

AGT in guinea pig liver is both peroxisomal and cytosolic

To quantify the subcellular distribution of AGT catalytic activity and AGT immunoreactive protein, especially the proportion in

Figure 2 Subcellular fractionation of guinea pig liver on an isopycnic sucrose gradient

(A) and (B) Distribution of enzyme activities (Δ , lactate dehydrogenase; $\square$, glutamate dehydrogenase; $\circ$, D-amino acid oxidase; $\blacktriangle$, catalase; $\blacksquare$, uricase; $\bullet$, AGT). (C) and (D) Distribution of immunoreactivities (C, AGT; D, catalase). Ordinates = enzyme activity in each fraction as a percentage of that recovered from the total gradient (A and B only). Abscissa = fraction number (1 = bottom of gradient).

the cytosol, guinea pig liver was fractionated by isopycnic sucrose (Figure 2) and Nycodenz (Figure 3) density gradient centrifugation. The results showed that both AGT catalytic activity and immunoreactivity are distributed between the organelle and cytosolic compartments. Although it was not possible to clearly separate guinea pig liver peroxisomes and mitochondria on sucrose gradients, previous immuno-electron microscopic studies have shown that organellar AGT is entirely peroxisomal [1]. In the sucrose gradient, about 39 % of AGT catalytic activity was located in the top three fractions, which also contained 94 % of the cytosolic marker enzyme lactate dehydrogenase. Notwithstanding the possibility of differential enzyme leakage from otherwise intact peroxisomes, this soluble AGT was unlikely to have been released due to bulk rupture of the peroxisomal membrane, as the activities of two other peroxisomal enzymes, uricase and D-amino acid oxidase, at the top of the gradient were low (1 and 9 %, respectively) (Figure 2). It has been known for some time that catalase is largely cytosolic in guinea pig liver [26–28], a finding that was confirmed in the present study by the observation that 89 % of the catalase was located in the top three fractions (Figure 2). The distributions of AGT catalytic activity and immunoreactivity were similar, indicating that most of the cytosolic AGT activity measured was unlikely to be due to

Figure 3 Subcellular fractionation of guinea pig liver on an isopycnic Nycodenz gradient

(A) Distribution of enzyme activities (□, glutamate dehydrogenase; ▲, catalase; ●, AGT). (B) and (C) Distribution of immunoreactivities (B, AGT; C, catalase). Ordinate = enzyme activity in each fraction as a percentage of that recovered from the total gradient (A). Abscissa = fraction number (1 = bottom of gradient).

interference from other aminotransferases, such as glutamate:glyoxylate/alanine:2-oxoglutarate aminotransferase [29]. A much better separation between peroxisomes and mitochondria was achieved in the Nycodenz gradient (Figure 3). The distribution of AGT catalytic activity and immunoreactivity was bimodal and paralleled the distribution of the peroxisomal marker catalase. As less than 2% of the cytoplasmic lactate dehydrogenase activity contaminated the organellar pellet loaded onto the Nycodenz gradient, it is unlikely that the low density AGT and catalase at the top of the gradient result from cytosolic contamination. It is more likely that the bimodal distribution is a reflection of the physical heterogeneity of the peroxisomal population or peroxisomal damage during the re-suspension of the organelle pellet. Although the presence of some AGT in the mitochondria cannot be absolutely excluded from these results, it is clear that the great majority of the organellar AGT is peroxisomal. This conclusion is supported not only by previous immuno-electron microscopic data but also by subsequent molecular and immunofluorescence studies (see below). As suggested from previous studies, the proportion of AGT that is cytosolic in guinea pig liver is considerably greater than that found previously in other species [1]. For example, in human liver less than 10% of the AGT catalytic activity and almost none of the AGT immunoreactivity were found in the 'cytosolic' fractions of sucrose gradients [30,31], a finding that agrees with the near absence of cytosolic immunoreactive AGT as determined by immuno-electron microscopy [32]. The small proportion of AGT catalytic activity in the 'cytosol' of human liver can be attributed to cross-over from another enzyme, glutamate:glyoxylate aminotransferase [29]. Although a previous study [12] has also demonstrated similar amounts (32%) of AGT in the 'soluble fraction' of guinea pig liver, the distribution of immunoreactive AGT was not determined. Also using the same technique, the same authors showed that 30–35% of AGT in human liver was also in the

Figure 4 Southern blotting of guinea pig DNA

Guinea pig genomic DNA was digested with *EcoRI* (E), *HindIII* (H) and *XbaI* (X), electrophoresed on an agarose gel and probed with pCav4. The positions of the size markers are shown in kb.

soluble fraction [33], a result that is clearly at variance with other studies on the distribution of immunoreactive AGT in the human.

Guinea pig AGT is encoded by a single gene

Southern blotting of *EcoRI*-, *HindIII*- or *XbaI*-digested guinea pig genomic DNA, using pCav4 as a probe, yielded single hybridizing bands varying between ~12 and ~18 kb (Figure 4). This is compatible with the presence of only a single AGT gene, as has been found in all the other species examined to date. Therefore, both the peroxisomal and cytosolic isoforms of AGT (comprising ~61 and ~39% of the total cellular AGT respectively) must be derived from the same gene.

Cloning of the guinea pig AGT gene

To determine the molecular basis for the unusual distribution of guinea pig AGT, cDNA and genomic clones were isolated, sequenced and functionally analysed. The homologous AGT cDNA probe pCav4 was used to screen a guinea pig liver cDNA library in λ gt11 (see the Experimental section). A number of clones were isolated, the longest being 1105 bp which was cloned into the *EcoRI* site of pBluescript to give pCav5. By analogy with the nucleotide sequences of human AGT, pCav5 was shown to consist of part of intron 6, all of exons 7–11 and the 3'-UTR, including the poly(A)⁺ signal and poly(A)⁺ tail. None of the other clones isolated extended more 5' than the original PCR-generated clone pCav1 (Figure 1).

To acquire more 5' sequence, the strategy of 5'-RACE was adopted (see the Experimental section). The first clone generated (pCav20) using the semi-nested (–)-primers P5 and P6 did not include either of the two ancestral potential translation start sites (see the Introduction). Therefore, further 5'-RACE was attempted using more proximal (–)-strand primers (P6 and P7). The resulting product (pCav23) extended the known 5' sequence to include the second, but not the first, of the two ancestral translation start sites (Figure 1). Further attempts at 5'-RACE failed to yield any more 5' sequence.

Because it was important to obtain sequence that included that equivalent to the first ancestral translation initiation site and because it was possible that this sequence was not represented in

```

1 GTGAGTTTTCAGACCCCTGACTAAGCAGGATCCATTGGCTGGGTCTCTGGCGCTATGG 60
61 GCACCGATCTGCACCTCCGAGCCAGCTCTTATGAGTCCCTGCCACAGGGTTTCAGCTC 120
121 CCAGGCTGCCGGTCTTGACAGCAGGACAAATGGTCTAAAAGTTTCAATACAGTGGCAG 180
181 AGGCATGAGCATGGGGCCAGCAGCAGACATGATCAGCCTCCGCCCTGTGATTTCGATA 240
241 GGGCCACCTGGAAGGTTTCCGGGGCACTGGCCAGGTCAGTCTGCCCTGAGGCCCGGG 300
      ↑           ↑           ↑
301 TGGCAGGCTGGGCACAGACCATGGCTCTTACCAGCTGTGTGTGGCCCCGCCAGAGGCC 360
      M A S Y Q L C V A P P E A L

361 TGGCTCAAGCCGCTCTCCATCCCAACGGCTCCTGCTGGGGCTGGCCCCCTCAACCTGG 420
      L K P L S I F N R L L L G P G P S N L A

421 CTCCCGCTGTCCAGCGGGCTGGCGGGCTGCAGATGATCGGGCAGATGCACAAGGAGACAT 480
      P R V Q A A G G L Q M I G H M H K E T F

481 TCCAGGTCATGGACAGATCAAGAAGGGCATCCAGTACGTCTTCCAGACAGGAACCCG 540
      Q V M D E I G A H V L H A M I K E P G S Y Y

541 TCACGCTGGCCATCAGCGGCTTGGACACTGCGCCCTAGAGGTGGCCCTGTTCAATCTCC 600
      T L A I S G S G H C A L E V A L F N L L

601 TGGAGCCAGGGGACTCTCTCTGGTGGGGCCAAAGCGGAGTGGGGCAGAGCGGGCTCAG 660
      E P G D S F L V G A N G E W G R R A A E

661 AGATTGGGAGCGGATCGGAGCCCATGTGCACGCCATGATCAAGGAGCCTGGAAGCTACT 720
      I G E R I G A H V L H A M I K E P G S Y Y

721 ACACACTGCAGGAGGTGGAAGAGGGCTTGGCCAGCACAAGCGGTGCTGCTTTCTGA 780
      T L Q E V E E G L A Q H K P V L L F L T

781 CCCAGGGGAGTCTCTCCACCGGGGTGCTGCAGCCCTTGGATGGCTATGGGAGCTGTGTC 840
      Q G E S S T G Y L D G E Y L G E L C H

841 ACAGGTACCACTGCCTGCTGCTGGTGGATGCTGTGGCATCCTGGCGGGGTCCCCATCT 900
      R Y Q C L L L V D A V A S L G G V P I F

901 TCATGACCATCAAGGCATTTGATGCCCTGTATTCTGGCTCCCAAGAGTCTCAACGGCC 960
      M D H Q G I D A L Y S G S Q K V L N G P

961 CCCCCGGCACCTCTCTCTCTCCTTCAGCGACAGGGCCAAAGCACAAGGTCTACTCTCGCA 1020
      P G T S L L R A K S D R A L K H K V Y S R K

1021 AGACGAAGCCCTTCTCTTCTCATGGACATTACATTGCTGGCCAACCTTGGGGCTGTG 1080
      T K P F S F L M D I T L L A N L W G C D

1081 ACGACAAGCCAGGATGTACCATCACCACGCCCATCATTTGGCTTGTTCACCTGAGGG 1140
      D K P R M Y H H T T P I I G L F T L R E

1141 AGAGCCTGGCACTCATCTCGGAGCAGGGCCTGGAGAGTAGCTGGCGGAGCAGCCTGAGG 1200
      S L A L I S E Q G L E S S W R Q H R E A

1201 CCATGAGCACCTGCAACAGCCGTGTCAGGGCTGGCGCTGCAGCTCTTCTGTCAGGAGCC 1260
      M E H L H S R L Q G L G L Q L F V R D P

1261 CGGCAGCGCTTACCTACAGTACACACCGTGGCTGTGCGCGCTGGCTACAACCTGAGGG 1320
      A S R L P T V T T V A V P A G Y A N W R D

1321 ACATCGTGACTACGTATGAGTCTCTTTCAGCATTGAGATCTCGGTGGGCTTGGACCTT 1380
      I V D Y V M D S F S I E I S G G L A G P S

1381 CCCTGGGGAAGGTGCTGCGGATCGGCTGATGGCTGTAATGCCAGCGGGAGAGCTGG 1440
      L G K V L R I G L M G C N A S R E N V D

1441 ACCGCTGGCCAGGCCCTGAAGGAGGCCCTGCAGCGCTGCCCGCTCACAGGCTGTAAT 1500
      R L A Q A L K E A L Q R C P R H R L *

1501 ACGGTACACACCCACCCCTACCTGGGCTGGGCTGCAAGGGATCTCAGGCTGTGTGT 1560
1561 CACCAGGCTGAGGGCAGGAGAGCTGTGCTGATGCTGCCCCCTCTGCGGAGCTGCTAG 1620
1621 AGAAGGATTCAGTGTACTTCTTCATGGGATCAGCCAGCCTGGCTTCAAGGAACATCA 1680
1681 ATAAAGTGCTGGCCAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1739

```

Figure 5 The guinea pig AGT nucleotide sequence and predicted amino acid sequence

The sequence shown is a composite of that obtained from cDNA (generated by RT-PCR, library screening and RACE) and genomic (generated by RAGE) clones. The 5' extremities of the cDNA/mRNA, as suggested by RACE and RNase protection, are indicated by ↑ and ↑, respectively. The putative poly(A) signal is indicated by single underlining. The first and second ancestral translation start sites are indicated by double underlining (note that the first site is lost in the guinea pig). This sequence has been deposited in the EMBL database under accession number Y10727.

the pool of guinea pig liver AGT transcripts, we isolated a guinea pig AGT genomic clone using RAGE (see the Experimental section). The resulting clone (pCav25) generated by the semi-nested (—)-strand primers P8 and P9 contained sequence equivalent to both of the ancestral potential translation initiation sites (Figure 1).

To check that the various cDNA clones (i.e. pCav1, pCav4, pCav5, pCav20 and pCav23) could actually be derived from the same transcript, guinea pig liver total RNA was amplified by

```

-22 1 38
gpig: MASYQLCVAPPEALLKPLSIPNRLLLGPGSNLAPRVQ
rat: MFRMLAKASVTLSRAASVWRN.G.H.L.P...S...K...L
rab: ...R.L...R...CT.H...P...L
cat: MFRALARASATLGPQVAGWART...T.C...L...R...L
marm: MFQALAKASALGPRAAGWVRT...H.L...K...T...P...TM
hum: ...HK.L.T.K...Q...P...IM

39 98
gpig: AAGGLQMIHMHKETFQVMEIKKIYQVQTRNPLTLAISGSGHCALEVALFNLEPGD
rat: ...S.R...Q...M...I...Q...V...V...M...T...
rab: ...E.MY...Q...A...V...T...
cat: V...K...M...I...D...Q...K...A...I...
marm: ...L...Y...I...E...V...A...I...V...
hum: ...S.S.DMY.I...E...V...A...V...

99 158
gpig: SFLVANGEWGRRAAEIGERIGAHVHMIKEPGSYITLQVEEGLAQHKPVLLFTQGES
rat: ...T.I...I...A...R...Q...K...EH...H...
rab: A...I...Q...V...R...P...D...H...C...H...
cat: P...V...I...Q...D...R...P...D...NH...L...A...
marm: ...V...I...Q...D...L...R...P...T...D...GH...H...
hum: ...I...Q...VD...R...P...T...D...GH...H...

159 218
gpig: STGVLPDLGDELCHRYQCLLLVDAVASLGGVPIFMDHQDIALYSGSKVLNPPPGTS
rat: ...F...S...S...S...S...S...S...S...S...S...
rab: ...F...K...S...S...A...Y...Q...V...A...A...
cat: ...S...N...S...S...C...T...Y...Q...V...S...
marm: ...S...F...K...S...A...L...Y...Q...I...A...
hum: ...F...K...S...S...T...LY...R...I...A...A...

219 278
gpig: LLSFSDRAKHKVSRTKPFSLMDITLLANLWGGDDKPRMYHTTPIGLFTLRESIAL
rat: ...I...N.K...S...V...YT...Y.SK...EG.T.VI...L.V.S.YC...
rab: ...I...K...S.I.A...Y...VQ...I...G...V...I.A...
cat: ...I...K...N.I.T...V...YL.MKW...I...G...I...VVS.YS...
marm: ...T...N.I...S...YL.VKY...G...VVS.YS...G...
hum: ...I...K...K.M...YL...KW...F...Q...I.V.S.YS...

279 338
gpig: ISEQGLESSWRQREAMEHLRLGLGLQFVRDPASRLPTVTTVAVPAGYNWRDIVDY
rat: ...L.N.R...TA...KC.RE...KF...K.EI...I...T...S...
rab: LV...K...QR...VAQ...YR...E...K...L...I...S.R...S...
cat: ...A...N...VTAY...G...K...L...D...N...
marm: L...N...K...AAY...G...A...R...K...L...T...D...S...
hum: ...A...N...AAY...G...A...K...L...D...S...

339 392
gpig: VMDSFSIEISGGLGSLGKVLRIGLMGCNASRENVDRLAQLKEALQRCPRHRL
rat: ...L.H.N...ED...L.Y...TT...A.V.E...R...H...KNK...
rab: ...HH.G...T...AD...L...T...T...R...H.AQSQ...
cat: ...H.D...T...M...L...T...VI...Q...S.NK...
marm: LIER.G...T...T...T...T...LVTE...R...H...KKK...
hum: ...I.H.D...M...T...L...T...VTE...RA...H...KKK...

```

Figure 6 Comparison of the predicted guinea pig AGT amino acid sequence with those of other species

The sequence of guinea pig AGT (gpig) is compared with that of rat, rabbit (rab), cat, marmoset (marm) and human (hum) AGTs. Residues identical to the guinea pig sequence are indicated by dots. The numbering starts at the initiating Met in the guinea pig protein, so that the MTS, when present, is represented by amino acids -22 to -1. The 58 amino acids underlined indicate the highly conserved regions referred to in the text. The asterisk indicates the putative pyridoxal phosphate binding site at K209.

RT-PCR using the guinea pig specific primers P11, which mapped close to the 5' extremity of the most 5' cDNA clone (pCav23), and P12, which mapped to the putative 3'-UTR. A product of expected size (1338 bp) was obtained, which was gel purified, blunt-ended, digested with *Bam*HI and then cloned into *Eco*RV/*Bam*HI-digested pBluescript to give pCav26 (Figure 1).

The guinea pig AGT gene has lost the ancestral 5' translation initiation site

Figure 5 shows the composite guinea pig AGT nucleotide sequence constructed from the various overlapping genomic and cDNA clones. The fifth ATG triplet (nucleotides 321–323) is predicted to be the translation initiation codon, the four upstream ATGs all being followed closely by stop codons. In addition, the nucleotide sequence around the fifth ATG, but not the others, fits the consensus sequence for optimum translation efficiency [34,35]. The ability of translation to initiate at this site was

	Human	Marmoset	Rabbit	Rat	Cat	Guinea pig
Human		89	80	80	83	78
Marmoset	90		80	78	82	77
Rabbit	62	71		78	80	79
Rat	67	67	52		77	76
Cat	62	71	43	57		79
Guinea pig	57	67	52	48	67	

amino acids -21 to -1

amino acids 1 to 392

Figure 7 Percentage amino acid sequence identity between AGTs of different species

The percentage identity is based on the sequence alignment in Figure 6. Amino acids 1 to 392 contain the predicted polypeptide sequence following translation initiation from the second ancestral site (i.e. the total polypeptide in the guinea pig, rabbit and human, the peroxisomal form of AGT in the rat and marmoset and the mature, processed, form of mitochondrial AGT in the rat, marmoset and cat). Amino acids -21 to -1 contain the leader sequence of pre-mitochondrial AGT in the rat, marmoset and cat (ignoring the N-terminal Met), and the equivalent theoretical sequence in the guinea pig, rabbit and human if it had been contained within the open reading frame. The overall identity for all six species is 14% (for residues -21 to -1) and 61% (for residues 1 to 392).

Figure 8 Northern blotting of guinea pig RNA

Total guinea pig RNA was electrophoresed on an agarose gel and probed initially with the cDNA clone pCav4, then stripped and re-probed with a clone derived from the intron 6 portion of the library clone pCav5. The positions of the size markers are shown in kb. Northern blotting of guinea pig polyA⁺ RNA (probed with pCav4) gave an identical result to that found with total RNA (results not shown).

confirmed by *in vitro* transcription/translation of pCav26, which yielded a product of expected size (~43 kDa) (results not shown). Thus, the open reading frame encodes a polypeptide of 392 amino acids, a size that is the same as that found for peroxisomal AGT in all other species. The amino acid sequence shows considerable similarity (76–79% identical) with that of AGT from other species (Figures 6 and 7). Surprisingly, the sequence is least like that of its nearest relative, the rat. An internal region of ~58 amino acids, previously noted to be highly conserved [4,5], is also well conserved in guinea pig AGT, as is the putative pyridoxal phosphate binding site [9] (see Figure 6). The C-

Figure 9 RNase protection

Total RNA from guinea pig liver was hybridized with the riboprobe pCav28. The RNase-protected fragments were run out on a denaturing agarose gel. The control lanes show the absence of any protected fragments when total RNA was replaced by tRNA, and the probe-sized fragment of 434 bp when the probe was run undigested on the gel.

terminal tripeptide of guinea pig AGT is HRL, which like all other AGTs studied so far (KKL in human and marmoset, NKL in rat and cat, and SQL in rabbit) [3–5,9] has only a two out of three match with the consensus PTS1 motif [36,37].

Unlike AGT in the marmoset, rat and cat, but like AGT in the human and rabbit, guinea pig AGT does not contain the ancestral MTS within the open reading frame. This is owing to the loss of the first ancestral translation start site, the equivalent codon (nts 255–257) being GTT instead of ATG.

The guinea pig AGT gene has lost the ancestral 5' transcription initiation site

Northern blotting of RNA isolated from guinea pig liver, using pCav4 as a probe, demonstrated the presence of a single major transcript of ~1.6 kb (Figure 8). A faint hybridizing band of ~2.5 kb was detected, estimated by phosphorimaging analysis to comprise ~5% of the total. Interestingly, this band, but not the 1.6 kb band, also hybridized to an intron 6 probe (derived from the cDNA library clone pCav5), and therefore might represent a minor splicing variant. The functional significance of this minor transcript is not clear as an equivalent polypeptide has never been encountered. It is possible, therefore, that pCav5 might represent a partial cDNA clone of the 2.5 kb transcript prematurely truncated following RT stalling during the creation of the library.

Results from the 5'-RACE (see above) indicated that cDNA could be amplified with a 5' extremity that mapped to nucleotide 279, between the two ancestral translation start sites (refer to Figure 1). To determine whether any transcripts initiated upstream of the more 5' of these sites, RNase protection analysis was carried out (Figure 9). One major protected fragment of

Figure 10 Immunofluorescence analysis of the distribution of guinea pig AGT in transfected normal and mutant human fibroblasts

Normal human fibroblasts (**A–D**) and mutant human PTS1(—) fibroblasts (**E, F**) were micro-injected with cDNA encoding guinea pig AGT. Cells were double-labelled for AGT (**A, C, E**) and catalase (**B, D, F**). To see better the co-localization of the AGT and catalase in the normal cells, parts of **A** and **B** are magnified by a linear factor of ~ 2.1 to give **C** and **D**, respectively. Bars = $10\ \mu\text{m}$ (panels **A, B, E** and **F** are magnified to the same extent, as are panels **C** and **D**).

~ 133 bp was identified which corresponded to a 5' extremity mapping to nucleotide 297 (refer to Figure 1). This is 18 bases further downstream of the most 5' nt in the RACE clone pCav23, but still between the two ancestral translation initiation sites. Two minor fragments of ~ 100 bp and ~ 139 bp were also detected, corresponding to 5' termini which map to nucleotides 330 and 291. No protected fragments were identified larger than 175 bp, which would have been expected if any transcripts initiated upstream of the more 5' ancestral translation start site. The combined data from these experiments suggest that the vast majority of transcripts initiate within an 18 bp region that maps between the two ancestral translation initiation codons, suggesting that the more 5' of the two ancestral transcription start sites has been lost in addition to the more 5' ancestral translation

start site. Either of these events on their own could explain the exclusion of the ancestral MTS from the open reading frame and the failure of AGT to be targeted to the mitochondria in the guinea pig [7].

Guinea pig AGT is targeted efficiently to peroxisomes in human cells via the PTS1 import machinery

Expression of guinea pig AGT in normal human fibroblasts following intranuclear microinjection of cDNA (pCav26 recloned into pcDNA3) showed that it was localized in punctate structures that co-localized with catalase (i.e. peroxisomes) [21] (Figures 10A–10D). Partly owing to the variable levels of expression of the transfected cDNA and partly owing to variations between

Table 2 AGT distribution in populations of transfected fibroblasts

AGT: GP, normal guinea pig AGT (ending in HRL); Hum, normal human AGT (ending in KKL); Hum-SKL, human AGT in which the C-terminal KKL is replaced by SKL. Cells: GP, guinea pig fibroblasts; Hum, normal human fibroblasts. Distribution (%): percentage of cells expressing AGT in which the distribution, as judged subjectively by immunofluorescence, is entirely peroxisomal (P), entirely cytosolic (C) or both peroxisomal and cytosolic (P + C). *N* = Total number of cells counted, with the number of different experiments in brackets.

AGT	Cells	Distribution (%)			<i>N</i>
		P	P + C	C	
GP (HRL)	GP	19	78	3	293 (4)
	Hum	48	52	0	122 (3)
Hum (KKL)	GP	1	96	3	102 (2)
	Hum	53	47	0	43 (1)
Hum-SKL	GP	70	30	0	144 (3)
	Hum	88	12	0	99 (2)

different cells with respect to numbers of peroxisomes etc., the degree of diffuse (cytosolic) labelling was variable. Analysis of 122 cells in three different experiments showed that in about half of the cells the distribution was entirely punctate (peroxisomal) and in about half it was both punctate and diffuse (peroxisomal and cytosolic) (Table 2). The distribution of guinea pig AGT in human cells was very similar to the distribution of normal human AGT (Table 2).

Expression of guinea pig AGT in human fibroblasts obtained from a patient with neonatal adrenoleukodystrophy, in which there had previously been shown to be a selective deficiency in the import of PTS1-containing peroxisomal proteins due to a mutation in the PTS1 import receptor Pex5p [20,38], gave a distribution that was entirely diffuse (i.e. cytosolic) (Figures 10E and 10F). This indicates that guinea pig AGT is imported into peroxisomes via the PTS1 import machinery and, as a corollary to this, that its C-terminal HRL is likely to be a PTS1, albeit of a non-standard variety, as suggested for AGT in other species [10].

The inefficient peroxisomal targeting of AGT in the guinea pig is a characteristic of guinea pig cells rather than guinea pig AGT

Expression of guinea pig AGT in guinea pig fibroblasts was rather different from its expression in human fibroblasts. Although the AGT was also clearly targeted to peroxisomes in the guinea pig cells, significantly more remained diffusely distributed in the cytosol (Figures 11A and 11B). Analysis of 293 transfected cells in four different experiments showed that a much greater proportion (78%) localized AGT in both peroxisomes and cytosol than was the case in the human cells (Table 2). This apparently less efficient peroxisomal targeting in guinea pig cells was not restricted to guinea pig AGT, as the number of cells with partly diffuse labelling was even greater (96%) when the cells were transfected with human AGT (Figures 11C and 11D and Table 2). Although no specific analysis was carried out, there was no obvious difference between the number of peroxisomes in the guinea pig and normal human fibroblasts used.

The efficiency of peroxisomal targeting of human AGT in guinea pig and human cells is improved by the addition of the consensus PTS1 SKL

To determine whether the apparent lower efficiency of AGT (both guinea pig and human) targeting in the guinea pig cells

(compared with the human cells) was related to the presence of a non-standard PTS1, guinea pig and human cells were transfected with human AGT in which the C-terminal KKL was replaced by SKL (Figures 11E and 11F). The efficiency with which this construct was imported into peroxisomes of both guinea pig and human cells was markedly increased compared with that found for normal human AGT (Table 2). However, even with the more efficiently targeted human AGT-SKL construct, the proportion of cells with exclusively peroxisomal labelling was greater with human cells than guinea pig cells, although the difference was somewhat less noticeable than with the naturally occurring AGTs (Table 2).

DISCUSSION

Frequent evolutionary loss of ancestral transcription and/or translation start sites accounts for the variable intracellular compartmentalization of AGT in mammals

The unparalleled variability in the intracellular targeting of AGT during the evolution of mammals has been the subject of investigation, both at the phenomenological and mechanistic levels (see the Introduction). Previous studies on the human, marmoset, rabbit, cat and rat have shown that the intracellular distribution of AGT in these species is dependent upon the presence or absence of an N-terminal MTS in the open reading frame. The unusual structure of the archetypal AGT gene with two in-frame potential translation start sites 66 bp apart, and two potential transcription start sites, one upstream of both translation start sites and one in between, allows alternative transcription or translation initiation at these sites to determine the intracellular distribution of AGT [11]. A surprising feature of this process is how frequently mutations appear to have occurred during the evolution of mammals, which have led to the loss of one or more of these sites.

Based mainly on morphological criteria, the guinea pig has been traditionally classified as a rodent. However, this classification has been challenged on molecular grounds [39,40]. Following the suggestion by Graur et al. [39] that the family Caviomorpha (which includes the guinea pig) should be removed from the order Rodentia and elevated to ordinal status in its own right, the phylogenetic relationship between guinea pigs and other rodents (rats, mice, etc.) has been hotly disputed. Different laboratories, using the same or similar molecular analyses, have come out both for and against the classification of guinea pigs as rodents. Although a consensus has yet to be reached, strong arguments have been advanced recently for upholding the traditional classification of guinea pigs as rodents [41,42].

Nevertheless, a variety of sequence and functional analyses have shown there to be significant differences between many guinea pig and rat proteins, including AGT. Even in the absence of any sequence data, the differences between guinea pig and rat AGT were drawn to attention by Noguchi et al. [43]. Thus immunoreactivity, substrate specificity, tissue distribution and, most remarkably, subcellular distribution of AGT are very different in the two species.

On the assumption that Caviomorpha are correctly classified within Rodentia, the ability to target AGT to the mitochondria must have been lost in the guinea pig ancestral line after its divergence from the rat ancestral line. Logic demands that this loss must have been temporally distinct from the similar events that also occurred in the rabbit and human ancestral lines. In the present state of our knowledge, it is not possible to determine whether the initial mutational event leading to loss of mitochondrial targeting in the guinea pig was the loss of the more 5' transcription start site or the more 5' translation start site. Both

Figure 11 Immunofluorescence analysis of the distribution of guinea pig AGT and normal and mutant human AGT constructs in transfected guinea pig fibroblasts

Guinea pig fibroblasts were micro-injected with cDNA encoding normal guinea pig AGT (**A, B**), normal human AGT (**C, D**) and mutant human AGT in which the C-terminal KKL was replaced by SKL (**E, F**). Cells were double-labelled for AGT (**A, C, E**) and SKL-containing polypeptides (**B, D**) or catalase (**F**). The inserts show parts of the main photographs magnified by a linear factor of ~ 2.1 , to show better the co-localization of AGT with the peroxisomal markers. This is especially necessary for **A** and **C** where the significant cytosolic immunofluorescence masks the rather weak peroxisomal labelling at the low magnifications. Bar = $10\ \mu\text{m}$ (all the main panels are magnified to the same extent, as are all the inserts).

would have resulted in the exclusion of the ancestral MTS from the open reading frame. In the human and rabbit ancestral lines, it is probable that the losses of the 5' translation start site was the key event, the losses of the more 3' transcription start sites being secondary events. Nevertheless it is remarkable that mutations to the same (translation initiation) codon that have the potential to lead to major alterations in the behaviour and functions of the resulting gene product have occurred on at least three separate occasions during the evolution of mammals.

Although the absence of transcription initiation from the more 5' site would not be expected to have any significance with respect to the subcellular distribution of guinea pig AGT, owing to the loss of the ancestral 5' translation start and consequential

permanent exclusion of the MTS from the open reading frame, it could be of importance with respect to the control of AGT expression, especially the influence of dietary fluctuations. For example, AGT expression in rat liver is variously regulated by gluconeogenic stimuli (glucagon or high protein diet), insulin and glucocorticosteroids [44,45]. Glucagon and insulin induction of AGT results from increased expression from the more 5', not the more 3', transcription start site [44–46]. The resulting 'long' transcript contains the MTS in the open reading frame so that all the induced AGT is localized to the mitochondria. The rat AGT gene contains numerous upstream regulatory elements, including a cAMP response element and a negative glucocorticoid response element [47]. Why these *cis* elements should control transcription

initiation from the more 5' site and not the more 3' site is not understood. Whether the guinea pig AGT gene possesses such upstream elements is currently unknown, but unlike all other rodents, such as rats, mice, hamsters and chipmunks, AGT expression in guinea pig liver is not induced by glucagon [48].

Inefficient targeting of some peroxisomal proteins in the guinea pig might indicate discrimination of the guinea pig peroxisomal import machinery against non-standard PTS1s

All AGTs so far studied possess C-terminal PTS1s that have only a two out of three match with the minimal consensus PTS1 tripeptide motif of S/A/C-K/R/H-L/M [36,37]. For example, the C-terminus of AGT is KKL in humans and marmosets [3,4], SQL in rabbits [4], NKL in cats and rats [5,9] and HRL in guinea pigs (present study). These non-standard PTS1s are necessary for the correct targeting of AGT but are insufficient to target reporter proteins, such as firefly luciferase, chloramphenicol acetyltransferase and green fluorescent protein, to peroxisomes [10] (P. B. Oatey and C. J. Danpure, unpublished work). The apparent context specificity of these non-standard PTS1s is contrasted by their greater degeneracy. For example, human AGT can be targeted to peroxisomes by the non-consensus tripeptides KKL, SQL, NKL and SSL, as well as the consensus tripeptide SKL [10]. The C-terminal tripeptide of guinea pig AGT (i.e. HRL), which has not been found before in any other peroxisomal protein, extends this degeneracy even further.

Various attempts have been made to measure, albeit subjectively, the efficiency of peroxisomal targeting sequences in transfected cells (for example see [37]). However, the cellular and molecular determinants of peroxisomal import efficiency have remained elusive. Assessment of the intrinsic efficiency of peroxisomal protein import in whole cells is hampered because the observable phenomenon (i.e. the relative distribution between the peroxisomes and cytosol) is the net result of several processes, including the possible degradation of non-imported protein in the cytosol. However, there is no evidence to suggest that the peroxisomal proteins described in the present study are unstable in the cytosol. For example, the total cell activities of many peroxisomal enzymes, including catalase, D-amino acid oxidase and AGT, are not reduced when peroxisomal import is inhibited due to genetic defects in the peroxisomal import machinery [17].

Although the molecular basis of the inefficient peroxisomal targeting of guinea pig AGT in guinea pig cells has yet to be fully elucidated, the data presented in this paper suggest that it might be due to the combined effects of an intrinsically weak (non-standard) PTS1 and a peroxisomal import machinery that cannot handle efficiently such a PTS1. This, albeit tentative, conclusion is drawn from the following observations: (1) significantly more AGT is cytosolic in guinea pig liver than in human liver, (2) all AGT constructs target less efficiently to peroxisomes in transfected guinea pig fibroblasts than in transfected human fibroblasts, (3) the peroxisomal targeting efficiency of guinea pig AGT is no worse than that of normal human AGT in either cell type and (4) the targeting efficiency of human AGT increases significantly when its non-standard C-terminal tripeptide KKL is replaced by the standard tripeptide SKL, the increase being more obvious in guinea pig cells than in human cells.

Although AGT is only the second guinea pig PTS1-containing peroxisomal protein to have its sequence determined (the other being a bifunctional protein which ends in SKL [49]), it is conceivable that the putative lower efficiency of peroxisomal protein targeting in the guinea pig is at least partly restricted to proteins possessing non-standard PTS1s. In a variety of other mammalian species catalase has been shown to possess a highly

conserved non-standard C-terminal tripeptide (i.e. ANL in both humans and rats), and both D-amino acid oxidase and uricase have been shown to possess equally highly conserved standard PTS1s (i.e. SHL in human, mouse, rabbit and pig D-amino acid oxidase, and SRL in baboon, mouse, pig, rabbit and rat uricase) (see [50] for a list of peroxisomal protein C-termini). Endogenous catalase is very inefficiently targeted to peroxisomes in guinea pig liver cells (even more so than AGT) (present study and [26–28]), whereas D-amino acid oxidase and uricase appear to target efficiently (present study). Whether the targeting efficiency of a protein like catalase would improve if its C-terminal tripeptide was replaced by SKL remains to be seen.

The working definition of a 'non-standard' PTS1 is more a functional one than a molecular one. It is a C-terminal tripeptide based on, but not conforming to, the SKL motif that is necessary, but not sufficient, for peroxisomal import. Non-standard PTS1s appear to be both context specific and degenerate. But the molecular basis for either of these characteristics is unknown. Presumably, such non-standard PTS1 proteins must contain extra targeting information, which might simply be an extra amino acid adjacent to the C-terminal tripeptide, as shown for catalase [51], or more complex sequences/structures next to or remote from the C-terminus, as suggested for AGT [10]. This extra targeting information could directly or indirectly affect the binding of the C-terminal component of the PTS1 to the tetratricopeptide repeat domain of the import receptor Pex5p or some other downstream component of the peroxisomal import pathway.

Non-standard PTS1s, such as those described in this paper, are being identified with increasing frequency in peroxisomal proteins. Molecular analysis of a peroxisomal import system, such as that in the guinea pig cells, that appears to discriminate to a much greater extent than that in human cells between standard and non-standard PTS1 proteins might provide insights into the perplexing characteristics, such as lower targeting efficiency, context specificity and degeneracy, possibly generally associated with non-standard sequences.

We thank Dr. Y. Suzuki (Kinki University, Japan) for the guinea pig liver cDNA library, and Dr. S. J. Gould (Johns Hopkins University, Baltimore, U.S.A.) for the anti-SKL antibody. We acknowledge the technical contribution of P. Cummings to the isolation of the library clone pCav5. The work described in this paper was supported by the Medical Research Council.

REFERENCES

- 1 Danpure, C. J., Fryer, P., Jennings, P. R., Allsop, J., Griffiths, S. and Cunningham, A. (1994) *Eur. J. Cell Biol.* **64**, 295–313
- 2 Danpure, C. J., Guttridge, K. M., Fryer, P., Jennings, P. R., Allsop, J. and Purdue, P. E. (1990) *J. Cell Sci.* **97**, 669–678
- 3 Takada, Y., Kaneko, N., Esumi, H., Purdue, P. E. and Danpure, C. J. (1990) *Biochem. J.* **268**, 517–520
- 4 Purdue, P. E., Lumb, M. J. and Danpure, C. J. (1992) *Eur. J. Biochem.* **207**, 757–766
- 5 Lumb, M. J., Purdue, P. E. and Danpure, C. J. (1994) *Eur. J. Biochem.* **221**, 53–62
- 6 Nishiyama, K., Berstein, G., Oda, T. and Ichiyama, A. (1990) *Eur. J. Biochem.* **194**, 9–18
- 7 Oatey, P. B., Lumb, M. J. and Danpure, C. J. (1996) *Eur. J. Biochem.* **241**, 374–385
- 8 Oda, T., Funai, T. and Ichiyama, A. (1990) *J. Biol. Chem.* **265**, 7513–7519
- 9 Oda, T., Miyajima, H., Suzuki, Y. and Ichiyama, A. (1987) *Eur. J. Biochem.* **168**, 537–542
- 10 Motley, A., Lumb, M. J., Oatey, P. B., Jennings, P. R., De Zoysa, P. A., Wanders, R. J., Tabak, H. F. and Danpure, C. J. (1995) *J. Cell Biol.* **131**, 95–109
- 11 Danpure, C. J. (1997) *BioEssays* **19**, 317–326
- 12 Takada, Y. and Noguchi, T. (1982) *Comp. Biochem. Physiol. B* **72**, 597–604
- 13 Noguchi, T., Takada, Y. and Oota, Y. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 919–927
- 14 Noguchi, T., Minatogawa, Y., Takada, Y., Okuno, E. and Kido, R. (1978) *Biochem. J.* **170**, 173–175

- 15 Yokota, S. and Oda, T. (1984) *Histochemistry* **80**, 591–595
- 16 Takada, Y. and Noguchi, T. (1982) *Biochem. Biophys. Res. Commun.* **108**, 153–157
- 17 Danpure, C. J., Fryer, P., Griffiths, S., Guttridge, K. M., Jennings, P. R., Allsop, J., Moser, A. B., Naidu, S., Moser, H. W., MacCollin, M. and DeVivo, D. C. (1994) *J. Inher. Metab. Dis.* **17**, 27–40
- 18 Frohman, M. A., Dush, M. K. and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998–9002
- 19 Mizobuchi, M. and Frohman, L. A. (1993) *BioTechniques* **15**, 214–216
- 20 Motley, A., Hettema, E., Distel, B. and Tabak, H. (1994) *J. Cell Biol.* **125**, 755–767
- 21 Gould, S. J., Krisans, S., Keller, G. A. and Subramani, S. (1990) *J. Cell Biol.* **110**, 27–34
- 22 Danpure, C. J. and Jennings, P. R. (1988) *Clin. Sci.* **75**, 315–322
- 23 Ellis, G. and Goldberg, D. M. (1972) *Clin. Chem.* **18**, 523–527
- 24 Crane, D., Holmes, R. and Masters, C. J. (1982) *Biochim. Biophys. Acta* **712**, 57–64
- 25 Plummer, D. T. and Wilkinson, J. H. (1963) *Biochem. J.* **87**, 423–429
- 26 Roels, F., de Coster, W. and Goldfischer, S. (1977) *J. Histochem. Cytochem.* **25**, 157–160
- 27 Yamamoto, K., Volkl, A., Hashimoto, T. and Fahimi, H. D. (1988) *Eur. J. Cell Biol.* **46**, 129–135
- 28 Bulitta, C., Ganea, C., Fahimi, H. D. and Volkl, A. (1996) *Biochim. Biophys. Acta* **1293**, 55–62
- 29 Thompson, J. S. and Richardson, K. E. (1966) *Arch. Biochem. Biophys.* **117**, 599–603
- 30 Danpure, C. J. and Jennings, P. R. (1986) *FEBS Lett.* **201**, 20–24
- 31 Wise, P. J., Danpure, C. J. and Jennings, P. R. (1987) *FEBS Lett.* **222**, 17–20
- 32 Cooper, P. J., Danpure, C. J., Wise, P. J. and Guttridge, K. M. (1988) *J. Histochem. Cytochem.* **36**, 1285–1294
- 33 Noguchi, T. and Takada, Y. (1978) *J. Biol. Chem.* **253**, 7598–7600
- 34 Kozak, M. (1989) *J. Cell Biol.* **108**, 229–241
- 35 Kozak, M. (1991) *J. Biol. Chem.* **266**, 19867–19870
- 36 Gould, S. J., Keller, G. A., Hosken, N., Wilkinson, J. and Subramani, S. (1989) *J. Cell Biol.* **108**, 1657–1664
- 37 Swinkels, B. W., Gould, S. J. and Subramani, S. (1992) *FEBS Lett.* **305**, 133–136
- 38 Dodd, G., Braverman, N., Wong, C., Moser, A., Moser, H. W., Watkins, P., Valle, D. and Gould, S. J. (1995) *Nat. Genet.* **9**, 115–125
- 39 Graur, D., Hide, W. A. and Li, W. (1991) *Nature (London)* **351**, 649–652
- 40 D'Erchia, A. M., Gissi, C., Pesole, G., Saccone, C. and Arnason, U. (1996) *Nature (London)* **381**, 597–600
- 41 Cao, Y., Okada, N. and Hasegawa, M. (1997) *Mol. Biol. Evol.* **14**, 461–464
- 42 Allard, M. W., McNiff, B. E. and Miyamoto, M. M. (1996) *Mol. Phylog. Evol.* **5**, 78–88
- 43 Noguchi, T., Fujiwara, S., Hayashi, S. and Sakuraba, H. (1994) *Comp. Biochem. Physiol. B* **107**, 179–182
- 44 Oda, T., Yanagisawa, M. and Ichiyama, A. (1982) *J. Biochem. (Tokyo)* **91**, 219–232
- 45 Miyajima, H., Oda, T. and Ichiyama, A. (1989) *J. Biochem. (Tokyo)* **105**, 500–504
- 46 Uchida, C., Funai, T., Oda, T., Ohbayashi, K. and Ichiyama, A. (1994) *J. Biol. Chem.* **269**, 8849–8856
- 47 Oda, T., Nishiyama, K. and Ichiyama, A. (1993) *Genomics* **17**, 59–65
- 48 Hayashi, S., Sakuraba, H. and Noguchi, T. (1989) *Biochem. Biophys. Res. Commun.* **165**, 372–376
- 49 Caira, F., Cherkaoui Malki, M., Hoefler, G. and Latruffe, N. (1996) *FEBS Lett.* **378**, 57–60
- 50 de Hoop, M. J. and Ab, G. (1992) *Biochem. J.* **286**, 657–669
- 51 Purdue, P. E. and Lazarow, P. B. (1996) *J. Cell Biol.* **134**, 849–862
- 52 Purdue, P. E., Lumb, M. J., Fox, M., Griffo, G., Hamon Benais, C., Povey, S. and Danpure, C. J. (1991) *Genomics* **10**, 34–42