Molecular Adaptation of Alanine : Glyoxylate Aminotransferase Targeting in Primates

Joanna D. Holbrook,† Graeme M. Birdsey,*†[1] Ziheng Yang,† Michael W. Bruford,‡[2] and Christopher J. Danpure*†

*MRC Laboratory for Molecular Cell Biology and †Department of Biology, University College London, London, England; and ‡Institute of Zoology, London, England

The intermediary metabolic enzyme alanine : glyoxylate aminotransferase (AGT) is targeted to different organelles (mitochondria and/or peroxisomes) in different species. Possibly under the influence of dietary selection pressure, the subcellular distribution of AGT has changed on at least eight occasions during the evolution of mammals. AGT targeting is dependent on the variable use of two alternative transcription and translation initiation sites which determine whether or not the region encoding the N-terminal mitochondrial targeting sequence is contained within the open reading frame. In the present study, we sequenced the 5′ region of the AGT gene, including both ancestral translation start sites, for 11 anthropoid primates and compared the results with data already available for two others. We show that while the more 3′ of the two translation start sites is maintained in all species, the more 5′ site has been lost in six species (five of seven catarrhines and one of six platyrrhines). In addition, the remaining two catarrhines, which have maintained the 5′ translation start site, are predicted to have lost mitochondrial targeting by a different mechanism, possibly loss of the more 5′ transcription start site. Analysis of the relative frequencies of nonsynonymous and synonymous mutations in the region encoding the extant or ancestral mitochondrial targeting sequences led us to suggest that there has been recent strong positive selection pressure to lose, or decrease the efficiency of, mitochondrial AGT targeting in several anthropoid lineages, and that the loss of mitochondrial targeting in this group of mammals is likely to have occurred on at least four, and possibly five, separate occasions.

Introduction

The localization of proteins within specific intracellular compartments is an ancient and defining feature of eukaryotic cells. Thus, almost every eukaryotic enzyme is found in a single and invariable location at which the environment is optimal for its particular catalytic function (Danpure 1995). This applies not only to enzymes in different cell types within the same species, but also to enzymes in the same cell type in different species. However, the intermediary metabolic enzyme alanine : glyoxylate aminotransferase (AGT) stands in marked contrast to this generalization, as it can be found in different subcellular locations in different species (Danpure et al. 1994). In the liver cells of some species, such as Insectivora (moles, hedgehogs, shrews), most Carnivora (cats, dogs, ferrets), Amphibia (frogs, newts), and Reptilia (terrapins), organellar AGT is mainly or entirely mitochondrial. In the liver cells of other species, such as Hominoidea (humans, gorillas, chimpanzees, orang-utans), Cercopithecidae (baboons, macaques), Cebidae (saki monkeys), Lagomorpha (rabbits), some Rodentia (guinea pigs), and some Marsupialia (wallabies, koalas), organellar AGT is entirely peroxisomal. In the liver cells of yet other species, such as Callitrichidae (marmosets, tamarins), most Rodentia (rats, mice, hamsters), at least one Carnivora (brown bear), and some Marsupialia (opossums), organellar AGT is found in both mitochondria and peroxisomes. Some species (frogs, guinea pigs) also have significant levels of AGT in the cytosol. Based on the limited number of species and phylogenetic groups so far analyzed, it has been estimated that the subcellular distribution of AGT must have changed on at least eight occasions during the evolution of mammals (Danpure et al. 1994).

Although the subcellular distribution of AGT appears to be highly variable when different species are compared, its distribution within a particular species does not normally change, and correct localization is absolutely critical for an animal's survival. This is best exemplified by the human hereditary disease primary hyperoxaluria type 1 (PH1), a lethal condition caused by a functional deficiency of AGT (Danpure and Purdue 1995). Although in most normal humans, AGT is entirely peroxisomal, in a subset of PH1 patients it is mistargeted to mitochondria (Danpure et al. 1989). Although catalytically active in the mitochondria, AGT is unable to fulfil its metabolic role of glyoxylate detoxification in this organelle.

Notwithstanding the above, in a very restricted group of rodents (e.g., rats, mice, and hamsters), but not in most other species (e.g., guinea pigs, rabbits, dogs and cats), the distribution of AGT can be modified by exogenous stimuli. For example, the normally mitochondrial and peroxisomal distribution of AGT found in the liver cells of murine rodents can be modified to become mainly mitochondrial by the administration of gluconeogenic stimuli, such as glucagon or high-protein diets (Hayashi, Sakuraba, and Noguchi 1989; Oda, Yanagisawa, and Ichiyama 1982).

There is a clear relationship between the organellar distribution of AGT and diet (Danpure et al. 1990,

[1] Present address: Centre for Clinical Pharmacology, The Rayne Institute, University College London, London, England.

[2] Present address: Cardiff School of Biosciences, Cardiff University, Cardiff, Wales.

Key words: molecular adaptation, alanine : glyoxylate aminotransferase, dietary selection pressure, protein targeting, mitochondria, peroxisomes.

Address for correspondence and reprints: Christopher J. Danpure, MRC Laboratory for Molecular Cell Biology, University College London, Gower Street, London WC1E 6BT, United Kingdom. E-mail: c.danpure@ucl.ac.uk.

Mol. Biol. Evol. 17(3):387–400. 2000

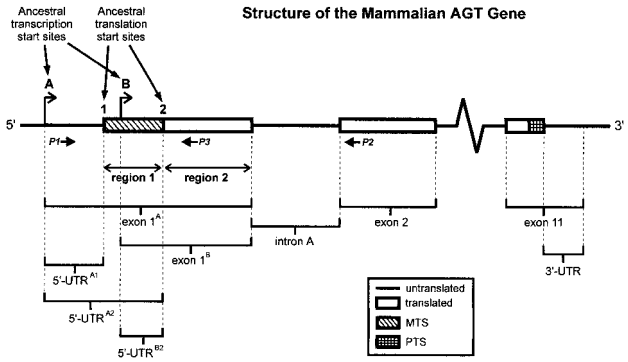


FIG. 1.—Diagrammatic structure of the archetypal mammalian AGT gene. The 5′ and 3′ ancestral translation start sites (1 and 2, respectively) are indicated, as are the 5′ and 3′ ancestral transcription start sites (A and B, respectively). Exons 1^A and 1^B are those that result from the alternative use of transcription start sites A and B, respectively. 5′-UTR^{A1}, 5′-UTR^{A2}, and 5′-UTR^{B2} are the untranslated regions that result from the alternative use of transcription start site A and translation start site 1 or 2 and from transcription start site B and translation start site 2, respectively. Region 1 refers to the part of the gene between the 5′ and 3′ translation start sites (including the 5′ site). Region 2 refers to the part of exon 1 downstream of the 3′ translation start site (including the 3′ site). P1, P2, and P3 indicate the positions to which the PCR primers map. The regions encoding the N-terminal mitochondrial targeting presequence and the C-terminal peroxisomal targeting sequence type 1 are shown hatched.

1994). Thus, there is a tendency for AGT to be peroxisomal in herbivores, mitochondrial in carnivores, and both peroxisomal and mitochondrial in omnivores. It has 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 (Danpure et al. 1990, 1994).

The single AGT gene has been cloned and functionally characterized in the human (*Homo sapiens*), the common marmoset (*Callithrix jacchus*), the rabbit (*Oryctolagus cuniculus*), the cat (*Felis catus*), the rat (*Rattus norvegicus*), and the guinea pig (*Cavia porcellus*) (Oda et al. 1987; Takada et al. 1990; Nishiyama et al. 1990; Oda, Funai, and Ichiyama 1990; Purdue, Lumb, and Danpure 1992; Lumb, Purdue, and Danpure 1994; Oatey, Lumb, and Danpure 1996; Birdsey and Danpure 1998). Such studies have shown that the archetypal mammalian AGT gene (see fig. 1) has the potential to encode an N-terminal mitochondrial targeting presequence (MTS) and a C-terminal peroxisomal targeting sequence type 1 (PTS1) (Motley et al. 1995) by the variable use of two transcription and two translation initiation sites (Danpure 1997). The final intracellular destination of AGT is dependent on the expression of the MTS rather than that of the PTS1, with the former being functionally dominant over the latter (Oatey, Lumb, and Danpure 1996). Thus, a decrease in mitochondrial targeting can automatically lead to an increase in peroxisomal targeting.

The MTS has been lost by permanent exclusion from the open reading frame on at least three temporally separated occasions in the human, rabbit, and guinea pig ancestral lines due to mutations in the more 5′ of the two potential translation initiation codons (ATG→ATA in the human, ATG→ACA in the rabbit, and ATG→GTT in the guinea pig) (Takada et al. 1990; Purdue, Lumb, and Danpure 1992; Birdsey and Danpure 1998). Both ancestral translation start sites are present in the rat, the marmoset, and the cat, and the differences in the distribution of AGT in these species reflect differences in the usage of the two ancestral transcription start sites (Oatey, Lumb, and Danpure 1996). The shift in AGT distribution in the rat following gluconeogenic stimuli results from increased expression from the more 5′ transcription start site, so that more transcripts contain the MTS in the open reading frame (Miyajima, Oda, and Ichiyama 1989).

Within Mammalia, some orders have the same distribution of AGT in all members; for example, all the species of Insectivora studied have mainly mitochondrial AGT. However, some closely related species differ in their AGT distributions. One group for which this is apparent is primates, which have provided the focus for this study. The distribution of hepatic AGT in primates appears to clearly separate the Platyrrhini from the Catarrhini. With the exception of the saki monkey (*Pithecia pithecia*), the Platyrrhini have the putatively more ancestral AGT distribution (i.e., mitochondrial and peroxisomal), also found in the Prosimii (Danpure et al. 1994). However, the Catarrhini have an exclusively peroxisomal distribution. From our previous studies on the molecular basis of AGT targeting in the human and the common marmoset, we have suggested that the loss of the 5′ translation start site in the human ancestral line probably occurred soon after the separation of the Platyrrhini and Catarrhini lineages (Danpure 1997).

In the present study, we investigated further the molecular evolution of AGT targeting in Anthropoidea. We determined the nucleotide sequence of the 5′ part of the AGT gene, including the regions containing both of the ancestral translation initiation sites, for 11 primates. We subjected these, along with the two primate sequences previously known, to a variety of analyses in an attempt to determine the nature of the mutational events, and their temporal relationships, which have led to the differences in AGT targeting within Anthropoidea.

Materials and Methods

Species

The species investigated or discussed in this study are listed in table 1. The tissues for DNA extraction from the common chimpanzee (*Pan troglodytes, Ptr*), the common gorilla (*Gorilla gorilla, Ggo*), the anubis baboon (*Papio anubis, Pan*), the common squirrel monkey (*Saimiri sciureus, Ssc*), the golden lion tamarin (*Leontopithecus rosalia, Lro*), the white-faced saki monkey (*Pithecia pithecia, Ppi*), and the red-faced black spider monkey (*Ateles paniscus, Apa*) were kindly supplied by Andrew Cunningham, Veterinary Science Unit, Institute of Zoology, London, England. The tissue from the white-handed gibbon (*Hylobates lar, Hla*), the Diana monkey (*Cercopithecus diana, Cdi*), the Celebes ma-

Table 1
Primates Discussed in this Paper

Suborder Family *Genus species*	Common Name	Natural Diet[a]	Subcellular Distribution
Anthropoidea			
Hominoidea			
Homo sapiens (*Hsa*)	Human	Omnivore	Peroxisomal
Pan troglodytes (*Ptr*)	Common chimpanzee	Herbivore	Peroxisomal
Gorilla gorilla (*Ggo*)	Common gorilla	Herbivore	Peroxisomal
Pongo pygmaeus (*Ppy*)	Orang-utan	Herbivore-omnivore	Peroxisomal
Hylobates lar (*Hla*)	White-handed gibbon	Herbivore-omnivore	?
Cercopithecidae			
Papio anubis (*Pan*)	Anubis baboon	Omnivore	Peroxisomal
Macaca nigra (*Mni*)	Celebes macaque	Herbivore-omnivore	?
Macaca fuscata (*Mfu*)	Japanese macaque	Herbivore-omnivore	Peroxisomal
Cercopithecus diana (*Cdi*)	Diana monkey	Herbivore-omnivore	?
Callithrichidae			
Sanguinus oedipus (*Soe*)	Cotton-top tamarin	Omnivore	Mitochondrial + peroxisomal
Leontopithecus rosalia (*Lro*)	Golden lion tamarin	Carnivore-omnivore	?
Callithrix jacchus (*Cja*)	Common marmoset	Carnivore-omnivore	Mitochondrial + peroxisomal
Callithrix argentata (*Car*)	Silvery marmoset	Carnivore-omnivore	Mitochodrial + peroxisomal
Callimico goeldii (*Cgo*)	Goeldis monkey	Omnivore	?
Saimiri sciureus (*Ssc*)	Common squirrel monkey	Herbivore-omnivore	?
Cebidae			
Pithecia pithecia (*Ppi*)	White-faced saki monkey	Omnivore	Peroxisomal
Ateles paniscus (*Apa*)	Red-faced black spider monkey	Herbivore-Omnivore	?
Prosimii			
Lemuridae			
Lemur fulvis (*Lfu*)	Brown lemur	Herbivore	Mitochondrial + peroxisomal
Cheirogaleus medius (*Cme*)	Fat tailed dwarf lemur	Herbivore-omnivore	Mitochondrial + peroxisomal
Lorisidae			
Loris tardigradus (*Lta*)	Slender loris	Carnivore-omnivore	Mitochondrial + peroxisomal
Nycticebus pygmaeus (*Npy*)	Lesser slow loris	Carnivore-omnivore	Mitochondrial + peroxisomal

[a] As most primates are opportunistic omnivores, their diets will vary depending on location and availability. Therefore, only a very general categorization of dietary preferences is given.

caque (*Macaca nigra, Mni*), and the goeldis monkey (*Callimico goeldii, Cgo*) were obtained from Zoobank, the tissue bank in the Conservation Genetics Group, Institute of Zoology, London, England.

Extraction of Genomic DNA

Standard molecular biology techniques were used throughout (Sambrook, Fritsch, and Maniatis 1989). Genomic DNA was extracted from skin, muscle, and blood following incubation with proteinase K (20 mg/ml) in the presence of 10% SDS and a salt/EDTA buffer (75mM NaCl, 25 mM EDTA) at 55°C for 5 h. The samples were then subjected to phenol/chloroform purification, and the DNA was precipitated via the addition of sodium acetate and ethanol. The samples were then centrifuged at 400 × *g* for 15 min, the ethanol was removed, and the pellets were washed with 80% ethanol. After further centrifugation for 5 min at 400 × *g*, the pellets were air-dried and dissolved in water.

Polymerase Chain Reaction

The PCR primers P1 (5′-AAGCCCATCCACCAATCCTCN-3′), P2 (5′-CTGGAACACGTACTGGATCCCTTCCTTGAN-3′), and P3 (5′-AGGGGCTTGAGCAGGGCCTTG-3′) were designed to map to regions of high sequence identity. P1 (+ strand) maps to 53–32 bp upstream of the 5′ ancestral translation start (site 1 in fig. 1). P2 (− strand) maps to 12–41 bp downstream of the intron A/exon 2 boundary. P3 (− strand) maps to 32–52 bp downstream of the 3′ ancestral translation start site (site 2 in fig. 1). Primer pair P1/P2 was designed to amplify a region of the AGT gene containing all of exon 1 (including both ancestral translation start sites), intron A, and part of exon 2 (see fig. 1). Amplification of intron A was considered important to minimize the possibility of unrecognized interspecies cross contamination. In two cases in which amplification with P1/P2 was unsuccessful, amplification was achieved with primer pair P1/P3, which gave a shorter product that excluded intron A (see fig. 1).

The 5′ regions of the AGT gene were amplified by PCR with a hot start at 80°C for 3 min, followed by 35 cycles of 30 s at 94°C, 30 s at 55–65°C, and 45 s at 72°C, followed by a final extension period of 15 min at 72°C. Samples (50 μl) contained 75 μM $MgCl_2$ and 10 μM dNTPs.

Cloning and Sequencing

The PCR products were cloned into the pGEM-T Easy plasmid (Promega) and sequenced by the dideoxy method using Sequenase version 2.0 T7 DNA polymerase (Amersham Life Science). In some cases, cycle sequencing

was carried out with the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer). The sequences described in this paper have been deposited in the EMBL database under accession numbers AJ237886 (*Ptr*), AJ237887 (*Ggo*), AJ237888 (*Hla*), AJ237889 (*Pan*), AJ237890 (*Mni*), AJ237891 (*Cdi*), AJ237892 (*Lro*), AJ237893 (*Cgo*), AJ237894 (*Ssc*), AJ237895 (*Ppi*), and AJ237896 (*Apa*).

Data Analysis

Nucleotide sequences were aligned using Pileup (Devereux, Haeberli, and Smithies 1984) and CLUSTAL W (Thompson, Higgins, and Gibson 1994). Ancestral states were reconstructed by a codon-based maximum-likelihood analysis using marginal reconstruction of ancestral sequences (Yang, Kumar, and Nei 1995). The pairwise estimates of synonymous and nonsynonymous mutations per synonymous and nonsynonymous site (i.e., d_S and d_N) were calculated by the maximum-likelihood method of Goldman and Yang (1994). Estimation of d_N/d_S (ω) among lineages by maximum likelihood was performed by the method of Yang (1998) using the computer program PAML (version 2, freely available at http://abacus.gene.ucl.ac.uk/software/paml.html). Likelihood ratio tests were used to compare models, with the χ^2 distribution used as an approximation.

Results

Exon 1 is Conserved Both Upstream (Region 1) and Downstream (Region 2) of the 3′ Ancestral Translation Start Site

Using genomic DNA as the template, the 5′ region of the AGT gene was amplified by PCR in nine different anthropoid primates using primer pair P1/P2 (see fig. 1 and *Materials and Methods*). PCR failed in two cases (gorilla and golden lion tamarin), but the key regions of the AGT gene were subsequently successfully amplified in these species using primer pair P1/P3. The P1/P2 PCR products varied in size between ~680 bp and ~950 bp solely due to the presence of intron A, the size of which was calculated to vary between ~330 bp and ~600 bp.

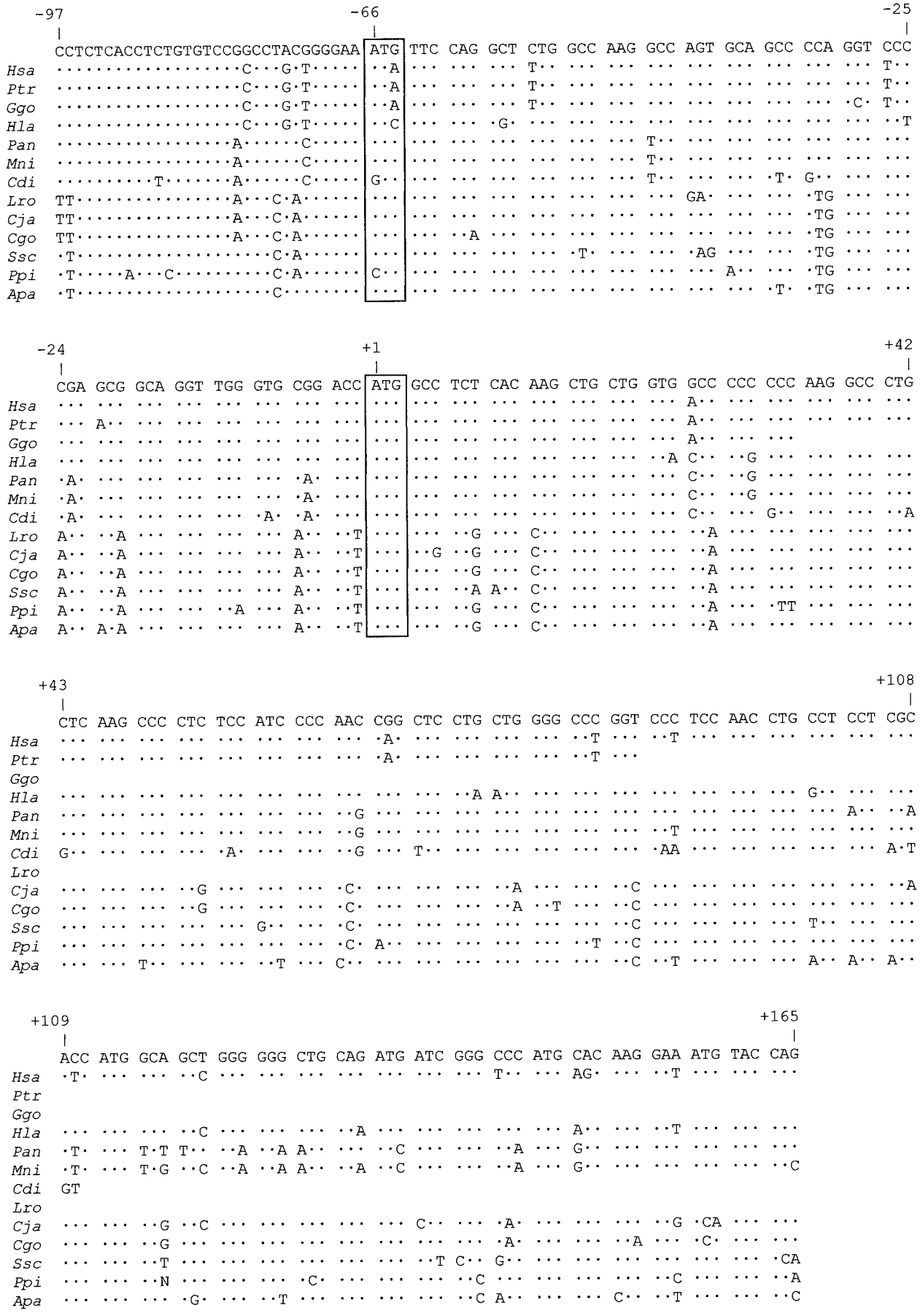
The nucleotide sequences of the P1/P2 products (excluding intron A and exon 2) and the P1/P3 PCR products are shown in figure 2, and their deduced amino acid sequences are shown in figure 3. As expected, the amplified regions were very similar to each other at both the nucleotide and the amino acid levels. Following pairwise comparisons, the nucleotide identity of region 1 (defined as nucleotides −66 to −1 and amino acids −22 to −1, with potential to encode, or act as, a MTS) varied between 78% and 100%, while the amino acid identity varied between 62% and 100%. In this region, there was a 61% nucleotide identity and a 33% amino acid identity across all 13 primates. Pairwise comparisons of region 2 (defined as nucleotides +1 to +165 and amino acids +1 to +55, where available) showed that the nucleotide identity varied between 81% and 100% and the amino acid identity varied between 67% and 100%. Excluding the missing nucleotides and residues in this region, the overall nucleotide identity was 62% and the overall amino acid identity was 48%. Overall, the levels of conservation (i.e., sequence identity) of regions 1 and 2 appear to be similar, despite the likelihood that region 1 is frequently excluded from the open reading frame (see *Introduction*). In fact, based on the absence of the 5′ ancestral start site (figs. 2 and 3), the ancestral MTS would appear to be excluded from the open reading frame in five of the seven catarrhines studied (i.e., the human, the chimpanzee, the gorilla, the gibbon, and the Diana monkey) and one out of the six platyrrhines (i.e., the saki monkey).

Ancestral Reconstruction of Region 1 Suggests that the MTS of AGT Could Have Been Lost on at Least Four Different Occasions in Anthropoids

We attempted to determine the number of events (and their temporal relationships) that are likely to have led to the abolition or diminution of AGT mitochondrial targeting by subjecting region 1 to a codon-based maximum-likelihood analysis using reconstruction of ancestral sequences (Yang, Kumar, and Nei 1995). The results are shown in figure 4. Numerous mutations have occurred that would be expected to abolish mitochondrial AGT targeting. For example, our analysis suggests that the 5′ ancestral translation start site has been lost in branches 5, 6, 10, and 22 and a stop codon has been generated in branch 22. Other substitutions have occurred that would be likely to interfere with mitochondrial AGT targeting, if not abolish it. For example, as MTSs require basic residues but are usually deficient in acidic residues (von Heijne 1986), the loss of two basic amino acids in branch 8 and the acquisition of two acidic amino acids in branches 10 and 19 would be expected to markedly decrease the effectiveness of region 1 as an MTS. In several branches (e.g., 8, 10, and 22), numerous substitutions have occurred, and in these cases it is clearly not possible to determine the order of events.

→

FIG. 2.—Nucleotide sequences of the 5′ region of the AGT gene in various primates. The first line is the consensus sequence generated by majority vote and has no evolutionary implications. In the aligned sequences below, nucleotides identical to the consensus sequence are indicated by periods. The codons equivalent to the two ancestral translation start sites are boxed. The numbering starts from the first base of the more 3′ ancestral translation start site (+1), so that the first base of the more 5′ translation start site is numbered −66. Thus, the whole sequence runs from −97 to +165. For four species (*Ptr, Ggo, Cdi,* and *Lro*) only incomplete sequence was available. The species abbreviations are defined in table 1. Note that the 5′ translation start site has been lost in the human (*Hsa*), the chimpanzee (*Ptr*), the gorilla (*Ggo*), the gibbon (*Hla*), the Diana monkey (*Cdi*), and the saki monkey (*Ppi*).


-97
-66
-25
CCTCTCACCTCTGTGTCCGGCCTACGGGGAAATGTTCCAGGCTCTGGCCAAGGCCAGTGCAGCCCCAGGTCCC
Hsa ·····C···G·T·······A··· ··· ··· T·· ··· ··· ··· ··· ··· ··· ··· T··
Ptr ·····C···G·T·······A··· ··· ··· T·· ··· ··· ··· ··· ··· ··· ··· T··
Ggo ·····C···G·T·······A··· ··· ··· T·· ··· ··· ··· ··· ··· ··· ·C· T··
Hla ·····C···G·T·······C··· ··· ·G· ··· ··· ··· ··· ··· ··· ··· ··· ··T
Pan ·····A·····C········ ··· ··· ··· ··· ··· ··· T·· ··· ··· ··· ··· ··· ···
Mni ·····A·····C········ ··· ··· ··· ··· ··· ··· T·· ··· ··· ··· ··· ··· ···
Cdi ··········T·····A·····C·····G·· ··· ··· ··· ··· ··· ··· T·· ··· ··· ·T· G·· ··· ···
Lro TT·····A···C·A········ ··· ··· ··· ··· ··· ··· ··· GA· ··· ··· ·TG ··· ···
Cja TT·····A···C·A········ ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ·TG ··· ···
Cgo TT·····A···C·A········ ··· ··A ··· ··· ··· ··· ··· ··· ··· ··· ·TG ··· ···
Ssc ·T·····C·A········ ··· ··· ··· ··· ·T· ··· ··· ·AG ··· ··· ·TG ··· ···
Ppi ·T·····A···C·····C·A·····C·· ··· ··· ··· ··· ··· ··· ··· ··· A·· ··· ·TG ··· ···
Apa ·T·····C········ ··· ··· ··· ··· ··· ··· ··· ··· ··· ·T· ·TG ··· ···
-24
+1
+42
CGA GCG GCA GGT TGG GTG CGG ACC ATG GCC TCT CAC AAG CTG CTG GTG GCC CCC CCC AAG GCC CTG
Hsa ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· A·· ··· ··· ··· ··· ···
Ptr ··· A·· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· A·· ··· ··· ··· ··· ···
Ggo ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· A·· ··· ···
Hla ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··A C·· ··G ··· ··· ··· ···
Pan ·A· ··· ··· ··· ··· ··· ··· ·A· ··· ··· ··· ··· ··· ··· ··· ··· C·· ··G ··· ··· ··· ···
Mni ·A· ··· ··· ··· ··· ··· ··· ·A· ··· ··· ··· ··· ··· ··· ··· ··· C·· ··G ··· ··· ··· ···
Cdi ·A· ··· ··· ··· ··· ·A· ·A· ··· ··· ··· ··· ··· ··· ··· ··· ··· C·· ··· G·· ··· ··· ··A
Lro A·· ··A ··· ··· ··· ··· A·· ··T ··· ··· ··G ··· C·· ··· ··· ··· ··A ··· ··· ··· ··· ···
Cja A·· ··A ··· ··· ··· ··· A·· ··T ··· ··G ··G ··· C·· ··· ··· ··· ··A ··· ··· ··· ··· ···
Cgo A·· ··A ··· ··· ··· ··· A·· ··T ··· ··· ··G ··· C·· ··· ··· ··· ··A ··· ··· ··· ··· ···
Ssc A·· ··A ··· ··· ··· ··· A·· ··T ··· ··· ··A A·· C·· ··· ··· ··· ··A ··· ··· ··· ··· ···
Ppi A·· ··A ··· ··· ··A ··· A·· ··T ··· ··· ··G ··· C·· ··· ··· ··· ··A ··· ·TT ··· ··· ···
Apa A·· A·A ··· ··· ··· ··· A·· ··T ··· ··· ··G ··· C·· ··· ··· ··· ··A ··· ··· ··· ··· ···
+43
+108
CTC AAG CCC CTC TCC ATC CCC AAC CGG CTC CTG CTG GGG CCC GGT CCC TCC AAC CTG CCT CCT CGC
Hsa ··· ··· ··· ··· ··· ··· ··· ··· ·A· ··· ··· ··· ··· ··T ··· ··T ··· ··· ··· ··· ··· ···
Ptr ··· ··· ··· ··· ··· ··· ··· ··· ·A· ··· ··· ··· ··· ··T ···
Ggo
Hla ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··A A·· ··· ··· ··· ··· ··· ··· ··· G·· ··· ···
Pan ··· ··· ··· ··· ··· ··· ··· ··G ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· ··· A·· ··A
Mni ··· ··· ··· ··· ··· ··· ··· ··G ··· ··· ··· ··· ··· ··· ··· ··· ··T ··· ··· ··· ··· ··· ···
Cdi G·· ··· ··· ··· ·A· ··· ··· ··G ··· T·· ··· ··· ··· ··· ··· ·AA ··· ··· ··· ··· ··· A·T
Lro
Cja ··· ··· ··· ··G ··· ··· ··· ·C· ··· ··· ··· ··A ··· ··· ··C ··· ··· ··· ··· ··· ··· ··A
Cgo ··· ··· ··· ··G ··· ··· ··· ·C· ··· ··· ··· ··A ··T ··· ··C ··· ··· ··· ··· ··· ··· ···
Ssc ··· ··· ··· ··· ··· G·· ··· ·C· ··· ··· ··· ··· ··· ··· ··C ··· ··· ··· ··· T·· ··· ···
Ppi ··· ··· ··· ··· ··· ··· ··· ·C· A·· ··· ··· ··· ··· ··T ··C ··· ··· ··· ··· ··· ··· ···
Apa ··· ··· T·· ··· ··· ··T ··· C·· ··· ··· ··· ··· ··· ··· ··C ··T ··· ··· ··· A·· A·· A··
+109
+165
ACC ATG GCA GCT GGG GGG CTG CAG ATG ATC GGG CCC ATG CAC AAG GAA ATG TAC CAG
Hsa ·T· ··· ··· ··C ··· ··· ··· ··· ··· ··· ··· T·· ··· AG· ··· ··T ··· ··· ···
Ptr
Ggo
Hla ··· ··· ··· ··C ··· ··· ··· ··A ··· ··· ··· ··· ··· A·· ··· ··T ··· ··· ···
Pan ·T· ··· T·T T·· ··A ··A A·· ··· ··C ··· ··· ··A ··· G·· ··· ··· ··· ··· ···
Mni ·T· ··· T·G ··C ··A ··A A·· ··A ··C ··· ··· ··A ··· G·· ··· ··· ··· ··· ··C
Cdi GT
Lro
Cja ··· ··· ··G ··C ··· ··· ··· ··· ··· C·· ··· ·A· ··· ··· ··· ··G ·CA ··· ···
Cgo ··· ··· ··G ··· ··· ··· ··· ··· ··· ··· ··· ·A· ··· ··· ··A ··· ·C· ··· ···
Ssc ··· ··· ··T ··· ··· ··· ··· ··· ··· ··T C·· G·· ··· ··· ··· ··· ··· ··· ·CA
Ppi ··· ··· ··N ··· ··· ··· ·C· ··· ··· ··· ··C ··· ··· ··· ··· ··C ··· ··· ··A
Apa ··· ··· ··· ·G· ··· ··T ··· ··· ··· ··· ··C A·· ··· ··· C·· ··T ··· ··· ··C

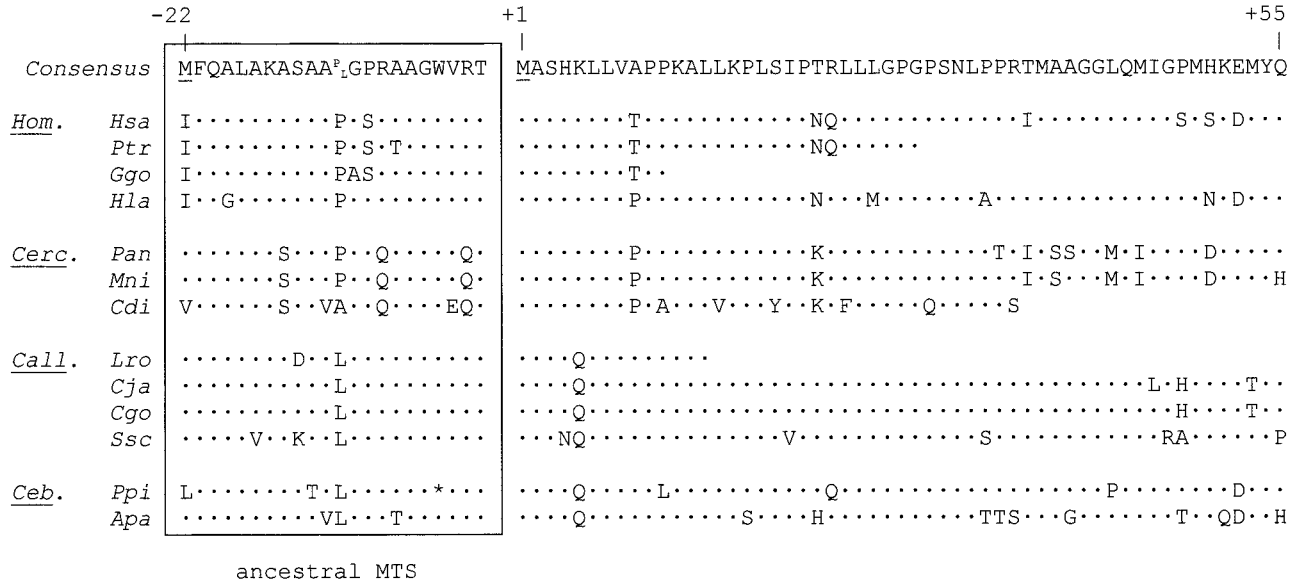


FIG. 3.—Deduced amino acid sequences encoded by the 5′ region of the AGT gene in various primates. The consensus sequence was generated by majority vote and has no evolutionary implications. In the aligned sequences below, residues identical to the consensus sequence are indicated by periods. The amino acids encoded by the two ancestral translation start sites are underlined in the consensus sequence. The numbering starts from the initiating methionine of the more 3′ translation start site (+1), so that the residue nominally encoded by the more 5′ translation start site (not always methionine) is numbered −22. The deduced amino acid sequence from residues −22 to −1 assumes the presence of this region in the open reading frame, even though in many cases it is clearly not (see text for details). The species abbreviations are defined in table 1. Note that the more N-terminal initiating methionine encoded by the more 5′ translation start site has been lost in the human (*Hsa*), the chimpanzee (*Ptr*), the gorilla (*Ggo*), the gibbon (*Hla*), the Diana monkey (*Cdi*), and the saki monkey (*Ppi*).

The d_N/d_S Ratios Indicate that Region 1 and Region 2 Have Been Subjected to Different Selection Pressures

Because ancestral reconstruction of the region 1 sequence indicated that the mitochondrial targeting of AGT could have been lost on at least four occasions (fig. 4), we attempted to determine whether this was likely to be an adaptive response to positive selection pressure for change. Therefore, using a maximum-likelihood method (Goldman and Yang 1994), we determined the relative number of nonsynonymous differences per nonsynonymous site (d_N) compared with the number of synonymous differences per synonymous site (d_S) for each pairwise species comparison. The d_N/d_S ratio (ω) is frequently taken as a measure of the selection pressure having acted on genes or parts of genes (Messier and Stewart 1997). Thus, $d_N/d_S < 1$ suggests that there has been pressure for conservation, while $d_N/d_S > 1$ suggests that there has been pressure for change, and $d_N/d_S = 1$ suggests that there has been no pressure (i.e., the sequence is neutral).

The values for d_N plotted against those for d_S in region 1 (which would be expected to be under variable selection pressure because it can be included or excluded from the open reading frame) and region 2 (which would be expected to be under continuous functional constraint because it is always included in the open reading frame) are shown in figure 5. The 5′ ancestral translation start codon was excluded from the d_N and d_S analysis of region 1 (designated as region 1*) because of its unique position in determining whether the remainder of the region is included within the open reading frame or not. In addition, any change to this codon from the ancestral condition would exclude this site from the open reading frame, and therefore the terms synonymous and nonsynonymous would have no meaning in this context.

Several findings emerge from this analysis. When region 2 is compared between all 13 species (i.e., 78 individual comparisons), $d_N/d_S < 1$ in 71 cases, and $d_N/d_S > 1$ in only two cases. In the remaining five cases, both d_N and $d_S = 0$. The large preponderance of comparisons in which $d_N/d_S < 1$ is strongly suggestive of functional constraint, as expected for a protein-coding sequence. In this respect, region 2 appears to be fairly representative of the whole coding region. For example, we calculated that in the human, marmoset, rat, rabbit, cat, and guinea pig AGT genes, the average value for d_N/d_S in region 2 is very similar to the average value for all 11 exons (i.e., 0.261 ± 0.099, compared with 0.253 ± 0.032). In our primate analysis, the mean values for d_S and d_N in region 2 were 0.434 ± 0.302 and 0.086 ± 0.044, respectively, giving an overall d_N/d_S ratio of 0.200 (table 2).

When d_N/d_S is determined for region 1*, a rather different picture emerges, depending on whether distant or close relatives are compared (fig. 5 and table 2). Thus, when the Catarrhini are compared with the Platyrrhini, $d_N/d_S < 1$ in 30 of 42 cases, and $d_N/d_S > 1$ in 12 of 42 cases, 6 of which involve comparisons with the same species (i.e., the Diana monkey). The presence of ratios both less than and greater than 1 might be indicative of variable functional constraint and positive selection pressure on a sequence that is predicted to be sometimes within the open reading frame and sometimes not. When

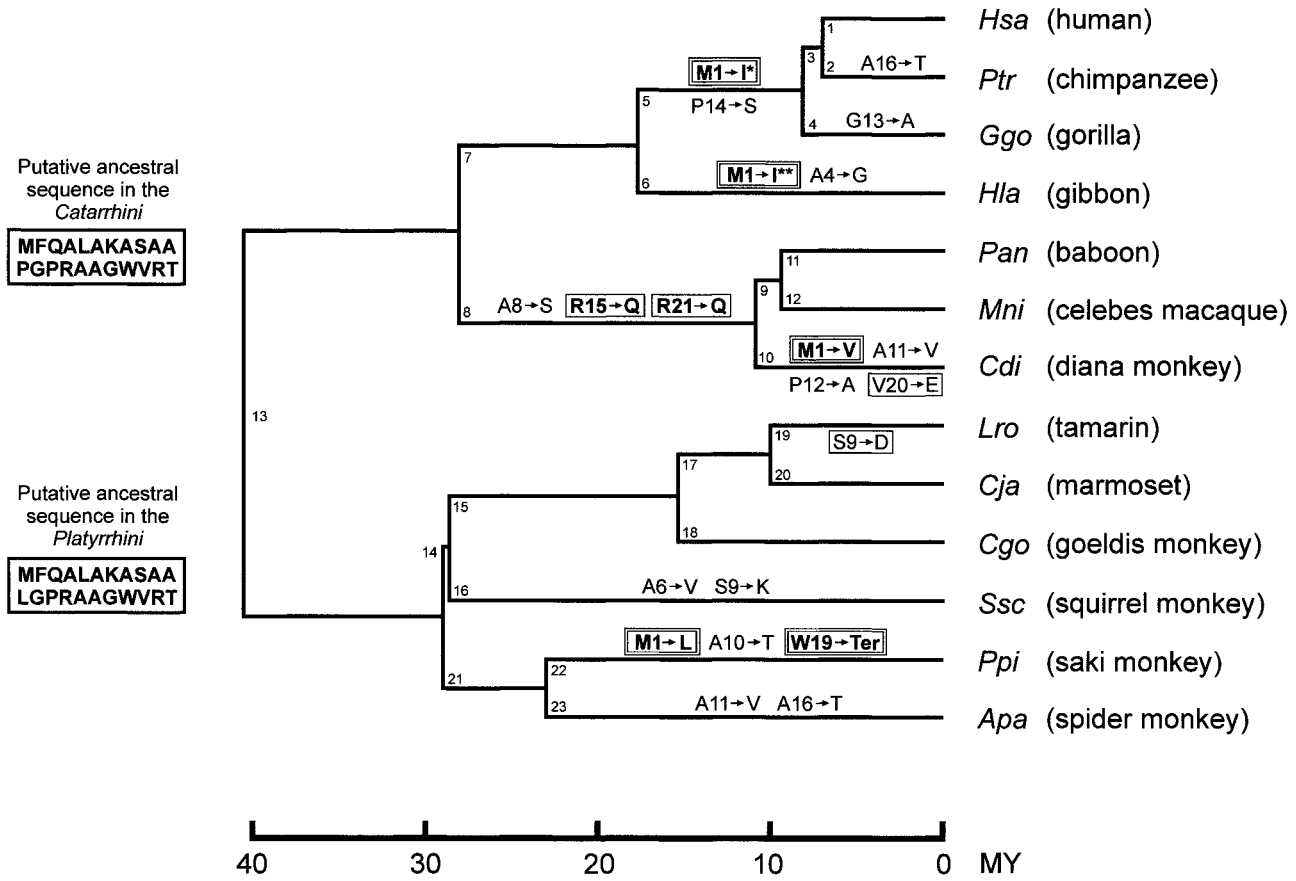


FIG. 4.—Ancestral reconstruction of AGT region 1 amino acid substitutions in primate lineages. The reconstructions are based on a maximum-likelihood method using the computer program PAML (see *Materials and Methods*). The phylogenetic tree and timescale are derived from Purvis (1995). The branches are arbitrarily labeled 1–23. Following the convention used elsewhere in this paper (e.g., fig. 3), the amino acid substitutions shown assume that region 1 remains within the open reading frame, even though those following loss of the 5′ translation start site (i.e., M1→I/V/L substitutions) clearly do not. Mutations that will abolish mitochondrial targeting are contained within a double box, and those that are likely to diminish, if not abolish, mitochondrial targeting are contained within a single box (see text). On the basis of this analysis, mitochondrial AGT targeting could have been abolished in four different lineages (branches 5, 6, 10, and 22) and diminished in three lineages (branches 8, 10, and 19). The loss of the ancestral 5′ translation start site in branch 5 (*) is a separate mutational event from its loss in branch 6 (**). Not shown on this tree is the putative loss of the 5′ transcription start site suggested for the baboon/macaque (*Pan*/*Mni*) ancestral line (see text). The only difference between the reconstructed Catarrhini and Platyrrhini ancestral sequences is residue 12 (P in the former and L in the latter). As the residue at this position is L in the rat and the cat (the only other species possessing mitochondrial AGT in which the gene has been sequenced), it is likely that the ancestral anthropoid sequence also had L at this position.

the Catarrhini are compared with each other, $d_N/d_S > 1$ in 19 of 21 cases, and $d_N/d_S < 1$ in only 1 of 21 cases. In the one remaining case, $d_N = d_S = 0$. When the Platyrrhini are compared with each other, $d_N/d_S > 1$ in 13 of 15 cases, and $d_N/d_S < 1$ in only 2 of 15 cases. The high preponderance of short-range comparisons in which $d_N/d_S > 1$ (i.e., 32 of 36 comparisons) is strongly suggestive of positive selection pressure for change acting on region 1*.

Analysis of the d_N/d_S Ratios Among Lineages Indicates Positive Pressure Acting on Region 1 in Several Recent Anthropoid Branches but Not on the Ancestral Branch Linking Catarrhines with Platyrrhines

In order to determine the evolutionary basis of the differences in d_N/d_S ratios obtained by pairwise comparisons (see above), we used the likelihood ratio test to calculate d_N/d_S ratios among the different lineages (fig. 6) (Yang 1998). The data were fitted to various models of sequence evolution, and the models were compared by the likelihood ratio test using the χ^2 approximation (tables 3 and 4).

Again, the results of the analyses of regions 1* and 2 were very different from each other. When a model that assumes an independent ω ratio ($\omega = d_N/d_S$) for each branch in the phylogeny (the ''free-ratios'' model) is applied to region 1*, estimates of the ω ratios are greater than 1 in the majority (8 out of 12) of the branches in the phylogeny with nonzero estimated lengths (i.e., branches 2, 4, 8, 10, 16, 19, 22, and 23) (fig. 6*B*). Eleven of the branches in region 1* have no substitutions at all and therefore have zero length and cannot be included in our analyses (fig. 6). Of the four branches in region 1* in which ω is less than 1 (i.e., 5, 6, 13, and 18), two of the more recent ones (i.e., 5 and 6) have ratios approaching 1 (i.e., 0.78), whereas the long ancestral branch 13 joining the Catarrhini and the Platyrrhini has a ratio of only 0.13.

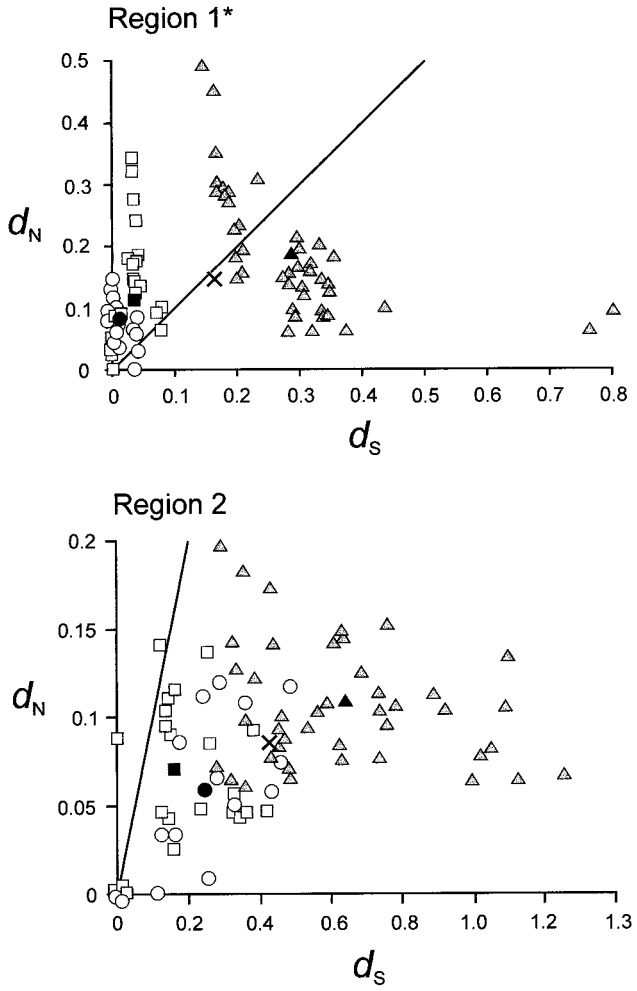


FIG. 5.—Comparison of the synonymous and nonsynonymous substitution rates in the 5′ region of the AGT gene in primates. The synonymous (d_S) and nonsynonymous (d_N) substitution rates were calculated using the maximum-likelihood method of Goldman and Yang (1994) for each of the 78 individual interspecies pairwise nucleotide comparisons. There are 21 Catarrhini-Catarrhini comparisons (open squares), 15 Platyrrhini-Platyrrhini comparisons (open circles), and 42 Catarrhini-Platyrrhini comparisons (shaded triangles). The upper and lower panels show the plots of d_N against d_S for region 1 (* excluding the 5′ ancestral translation start codon) and region 2, respectively (see fig. 1). The black symbols show the means for each group of comparisons, and the cross indicates the overall mean (the mean values for d_N and d_S are given in table 2). The diagonal line indicates $d_N = d_S$. In region 1*, d_N was greater than d_S in 90% of the Catarrhini-Catarrhini comparisons and in 87% of the Platyrrhini-Platyrrhini comparisons, but in only 29% of the Catarrhini-Platyrrhini comparisons (half of which were comparisons against the Diana monkey (*Cdi*). The equivalent values for region 2 were 9.5%, 0%, and 0%, respectively.

A model that assumes the same ω ratio for all branches (the ''one-ratio'' model) gave ω = 2.00. To test whether this ω ratio is significantly greater than 1, we compared it with a one-ratio model in which ω was fixed at 1. When the likelihood ratio statistic ($2\Delta\ell$) was compared with the χ^2 distribution, the difference between the models was not significant ($P = 0.110$) (see tables 3 and 4).

In order to test our hypothesis (based on our subjective assessment of the data to date) that positive selection has been driving only the recent evolution of region 1*, we fitted a ''two-ratio'' model, which assumes that the ancestral branch joining the Catarrhini with the Platyrrhini (i.e., branch 13) has an ω ratio (ω_0) that is different from that of all the other (more recent) branches (ω_1). The two-ratio model fits the data of region 1* significantly better than the one-ratio model ($P = 0.0014$), and it follows that ω ratios are significantly different between the old (branch 13) and recent (branches 2, 4–6, 8, 10, 16, 18, 19, 22, and 23) lineages. Furthermore, we found that ω_1 for these recent lineages was significantly different from (greater than) 1 ($P = 0.0038$) (see table 4).

In contrast, region 2 shows a typical pattern of functional constraint. When the free-ratios model is applied to this region, estimates of ω ratios for all branches are less than 1, except for two branches whose estimated lengths are very small (fig. 6*C*). The short branches contain less information about the ω ratio. Seven branches in region 2 had no substitutions and therefore had zero length and could not be included in the analysis. The model assuming one ω ratio for all lineages gives the maximum-likelihood estimate ω = 0.258. This is significantly different from (less than) 1, with $P < 0.001$.

Discussion

The Presence or Absence of the 5′ Ancestral Translation Start Site is Often, but Not Always, Correlated with AGT Protein Distribution in Primates

Previous evidence suggested that the intracellular compartmentation of AGT is dependent on the presence or absence of the 5′ ancestral translation start site (i.e., the codon at nucleotide positions −66 to −64) (see *Introduction*). In its absence, all translation would be expected to start at the 3′ site (nucleotide positions +1 to +3). Although one or other of the ancestral transcription start sites (see fig. 1) can also be lost, this has usually been considered a neutral secondary event (Danpure 1997). For example, our previous studies have shown that the presence of ATG at the 5′ translation start site in the common marmoset and its mutation to ATA in the human ancestral line is entirely responsible for the presence of both mitochondrial and peroxisomal AGT in the former, yet only peroxisomal AGT in the latter (Takada et al. 1990; Purdue, Lumb, and Danpure 1992).

In addition to the absence of the 5′ ancestral translation start site in humans (Takada et al. 1990), we have shown in the present study that it is also absent in the following catarrhines—the common chimpanzee, the common gorilla, the white-handed gibbon, and the Diana monkey, as well as the white-faced saki monkey, which is a platyrrhine (see figs. 2 and 3). Of these, the human, the chimpanzee, the gorilla and the saki monkey are known to have a peroxisomal distribution of AGT (Danpure et al. 1994; unpublished observations). The subcellular distribution of AGT in the gibbon and the Diana monkey is not known. In addition to the presence of the 5′ ancestral translation start site in the common marmoset (Purdue, Lumb, and Danpure 1992), we have shown in this study that it is also present in the black spider monkey, the golden lion tamarin, the goeldis

Table 2
Mean Values for d_N and d_S in Pairwise Comparisons

	N	Mean d_N (±SD)	Mean d_S (±SD)	Ratio of Mean d_N/ Mean d_S[a]
Region 1[b]				
Catarrhini vs. Catarrhini	21	0.141 ± 0.093	0.031 ± 0.046	4.54
Platyrrhini vs. Platyrrhini	15	0.075 ± 0.039	0.012 ± 0.019	6.25
Catarrhini vs. Platyrrhini	42	0.185 ± 0.101	0.290 ± 0.133	0.64
Overall	78	0.152 ± 0.099	0.167 ± 0.166	0.91
Region 2				
Catarrhini vs. Catarrhini	21	0.067 ± 0.043	0.174 ± 0.124	0.39
Platyrrhini vs. Platyrrhini	15	0.058 ± 0.045	0.247 ± 0.153	0.23
Catarrhini vs. Platyrrhini	42	0.106 ± 0.035	0.632 ± 0.261	0.17
Overall	78	0.086 ± 0.044	0.434 ± 0.302	0.20

[a] The means of the d_N/d_S ratios cannot be calculated because some individual comparisons give $d_S = 0$.
[b] Excluding the 5′ ancestral translation start site.

monkey, and the common squirrel monkey (all platyrrhines), as well as the anubis baboon and the Celebes macaque (both catarrhines) (see figs. 2 and 3). AGT is known to be both mitochondrial and peroxisomal in the marmoset and yet peroxisomal in the baboon. The distribution in the other species is unknown, but as all Callitrichidae studied to date have both mitochondrial and peroxisomal AGT, it is likely that the golden lion tamarin, the goeldis monkey, and the common squirrel monkey do as well. In addition, it is likely that Celebes macaque has peroxisomal AGT, as its close relative the Japanese macaque does also (Takada and Noguchi 1982). Unfortunately, there are no clues to the subcellular distribution of AGT in the black spider monkey.

With the notable exception of two of the Cercopithecidae, the anubis baboon and the Celebes macaque,

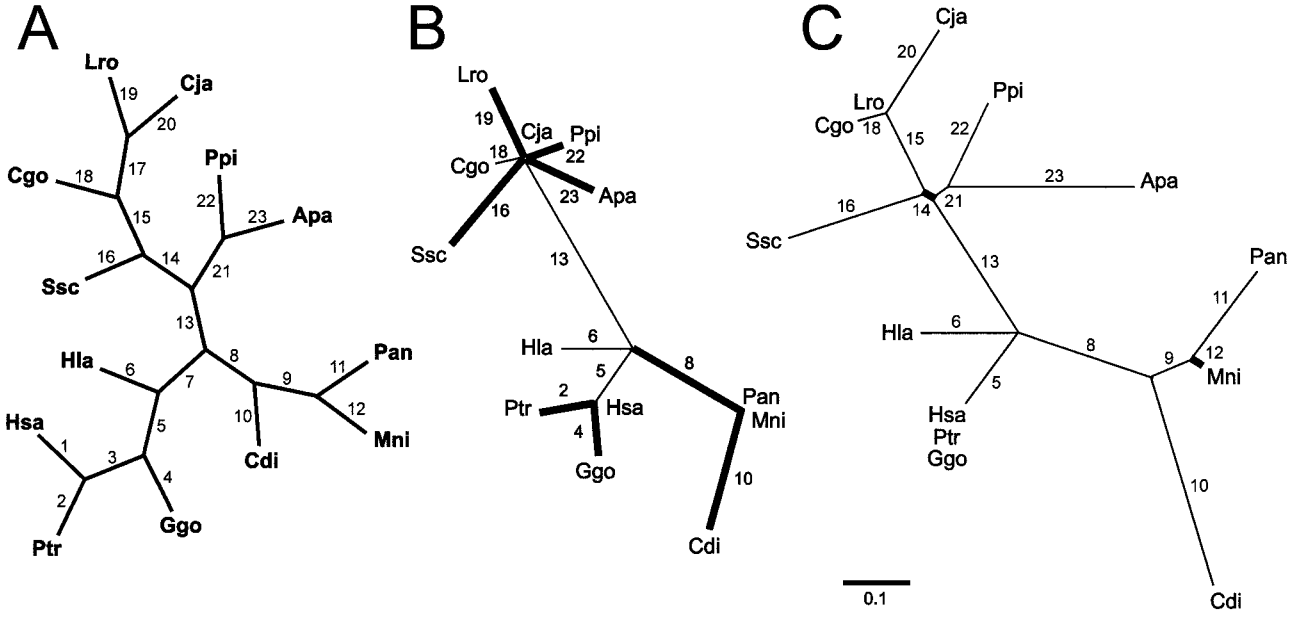


FIG. 6.—Nonsynonymous/synonymous substitution ratios (d_N/d_S) among lineages in the 5′ region of the AGT gene in primates. Three trees are shown. The phylogenetic tree A is based on that of Purvis (1995) (see fig. 4) but with all branches (numbered 1–23 as in fig. 4) drawn equal length. The derived trees B and C were generated using the computer program PAML (see *Materials and Methods*) and drawn using the program TREEVIEW (Page 1996). The branch lengths were calculated using the one-ratio model and the d_N/d_S ratios calculated using the free-ratios model (Yang 1998). The bar represents the expected number of nucleotide substitutions per codon estimated using the one-ratio model (Yang 1998). The apparent loss of some branches results from the fact that in some cases $d_N = d_S = 0$, and therefore the branch length is zero. The roots of these trees (not shown) would join onto the ancestral branch 13 linking the Platyrrhini and Catarrhini. Results of the analysis of region 1* (region 1 excluding the 5′ ancestral translation start site) are shown in tree B, and those for region 2 are shown in tree C. In trees B and C, thick lines denote branches where $d_N > d_S$, whereas thin lines denote branches where $d_N < d_S$. Species abbreviations (defined in table 1) are attached to terminal nodes. The calculated d_N and d_S values (to three decimal places) along various lineages for region 1* (excluding the 5′ ancestral start codon) (tree B) are as follows: 1 = 0 (d_N), 0 (d_S); 2 = 0.028, 0; 3 = 0, 0; 4 = 0.029, 0; 5 = 0.029, 0.037; 6 = 0.029, 0.037; 7 = 0, 0; 8 = 0.092, 0.001; 9 = 0, 0; 10 = 0.096, 0.001; 11 = 0, 0; 12 = 0, 0; 13 = 0.029, 0.227; 14 = 0, 0; 15 = 0, 0; 16 = 0.086, 0.001; 17 = 0, 0; 18 = 0, 0.037; 19 = 0.057, 0.001; 20 = 0, 0; 21 = 0, 0; 22 = 0.028, 0; 23 = 0.056, 0.001. The d_N and d_S values along lineages for region 2 (tree C) are as follows: 1 = 0 (d_N), 0 (d_S); 2 = 0, 0; 3 = 0, 0; 4 = 0, 0; 5 = 0.038, 0.055; 6 = 0.019, 0.141; 7 = 0, 0; 8 = 0.057, 0.105; 9 = 0, 0.091; 10 = 0.096, 0.116; 11 = 0.017, 0.174; 12 = 0.008, 0; 13 = 0.033, 0.217; 14 = 0.009, 0; 15 = 0.015, 0.129; 16 = 0.05, 0.121; 17 = 0, 0; 18 = 0, 0.063; 19 = 0, 0; 20 = 0.008, 0.178; 21 = 0, 0.038; 22 = 0.024, 0.114; 23 = 0.085, 0.091.

Table 3
Log-Likelihood Values and Maximum-Likelihood Estimates of Parameters Under Different Models

Analysis	ℓ	ω	κ
Region 1			
(1) One-ratio model with ω estimated	−219.13	2.00	5.35
(2) One-ratio model with $\omega = 1$ fixed	−220.41	[1]	5.11
(3) Two-ratio model with ω_0 and ω_1 estimated	−214.05	$\omega_0 = 0.125$ $\omega_1 = 4.77$	5.36
(4) Two-ratio model with $\omega_1 = 1$ fixed	−218.24	$\omega_0 = 0.125$ [$\omega_1 = 1$]	5.36
Region 2			
(5) One-ratio model with ω estimated	−719.14	0.258	2.42
(6) One-ratio model with $\omega = 1$ fixed	−737.37	[1]	2.17

NOTE.—ℓ = log likelihood value; $\omega = d_N/d_S$ (ω_0 is for ancestral branch 13, ω_1 is for all other branches; see fig. 6); κ = transition/transversion rate ratio. Region 1 excludes the 5′ ancestral start codon.

the sequences at the 5′ ancestral translation start sites in the primate AGT genes studied in this paper are compatible with the known or likely subcellular distributions of AGT. Thus, when the triplet ATG is present at this site, AGT is both mitochondrial and peroxisomal. However, when any other sequence (e.g., ATA, ATC, GTG, or CTG in this study) is present at this site, AGT is only peroxisomal. Thus, in most of the primates studied, loss of mitochondrial AGT targeting is due to loss of the 5′ translation start site and hence the exclusion of the region encoding the MTS from the open reading frame, as advocated previously (Danpure 1997). Clearly, this cannot be the case for the baboon and the macaque, which must have lost mitochondrial AGT targeting by a different mechanism.

Mitochondrial AGT Has Been Lost in the Baboon and the Macaque, Probably by Loss of the 5′ Ancestral Transcription Start Site

There are at least two possible mechanisms by which the baboon and the macaque could have lost the ability to target AGT to the mitochondria without loss of the 5′ translation start site. They could have lost the 5′ transcription start site (site A in fig. 1), or they could have accumulated nonsynonymous mutations in the MTS that prevent it from functioning as such. Loss of the 5′ transcription start site, which is upstream of both translation start sites, would result in the exclusion of region 1 (encoding the MTS) from the open reading frame, as does the loss of the 5′ translation start site. The 5′ transcription start site has been lost in the guinea pig, which has also lost the ability to target AGT to the mitochondria, but this could be a secondary event to the putative earlier loss of the 5′ translation start site (Birdsey and Danpure 1998).

Table 4
Test of Evolutionary Hypothesis

Hypothesis Being Tested (H_0)	Models Compared (see table 3)	Test Statistic ($2\Delta\ell$)	P
Region 1			
(a) $\omega = 1$	(1) and (2)	2.56	0.110
(b) $\omega_0 = \omega_1$	(1) and (3)	10.16	0.0014
(c) $\omega_1 = 1$	(3) and (4)	8.38	0.0038
Region 2			
(d) $\omega = 1$	(5) and (6)	36.46	<0.001

NOTE.—Hypotheses being tested: (a) whether the d_N/d_S ratio in region 1 averaged over all lineages is different from 1; (b) whether the d_N/d_S ratio in region 1 for the ancient branch (branch 13 in fig. 6) is different from the ratio for all other lineages; (c) whether the d_N/d_S ratios in region 1 in the more recent lineages are different from 1; and (d) whether the d_N/d_S ratio in region 2 averaged over all lineages is different from 1. Region 1 excludes the 5′ ancestral start codon.

On the assumption that region 1 is contained within the open reading frame in the baboon and the macaque, the deduced amino acid sequence shows a number of differences from those expected for an efficient MTS, which are usually positively charged amphiphilic α-helices (von Heijne 1986). For example, Arg(−2) and Arg(−8) have both been replaced by Gln (fig. 3). This would not only decrease substantially the net positive charge of the sequence, but the loss of Arg(−2) would also be predicted to interfere with presequence cleavage (Schneider et al. 1998). In addition, the presence of three juxtaposed helix breakers, Pro(−11), Gly(−10), and Pro(−9), would call into question the ability of this region to fold into an α-helix.

Although both of the mechanisms proposed for the loss of mitochondrial targeting in the baboon and the macaque are possible, we consider that the former is most likely (i.e., the loss of the 5′ transcription start site), because our previous studies (Danpure et al. 1990) and recent unpublished observations show that AGT in the baboon liver is similar in size to human AGT (i.e., it lacks the 22-aa mitochondrial leader sequence). Although it is conceivable that baboon AGT could be initially targeted to mitochondria, have its leader MTS (encoded by region 1) cleaved by the mitochondrial processing peptidase, and then be released back into the cytosol for subsequent import into peroxisomes (by a mechanism like that proposed for *Saccharomyces cerevisiae* fumarase [Stein et al. 1994]), it seems improbable to us, at least. If our conclusions are correct, then this is the first example of the loss of mitochondrial AGT targeting caused by loss of the 5′ transcription start site (site A in fig. 1).

Possible Temporal Relationship of the Mutational Events Leading to the Varied Subcellular Distribution of AGT in Extant Primates

So far, there is evidence for only one platyrrhine having lost mitochondrial AGT targeting (i.e., the saki

monkey), and this is clearly independent of the loss of mitochondrial AGT targeting in the Catarrhini. In addition, the saki monkey is the only platyrrhine in which the 5′ ancestral translation start site has been lost. As region 1 in the saki monkey also contains a stop codon, it is not possible to determine whether the loss of the 5′ translation start was the primary event and the acquisition of stop codon a neutral secondary event or vice versa. Nevertheless, both of these changes probably occurred after the separation of the saki monkey and the spider monkey ancestral lines about 23 MYA (see fig. 4).

The situation for the Catarrhini appears to be much more complicated. Although, as is so far known, the distribution of AGT in the Catarrhini is uniform (i.e., peroxisomal), our molecular analysis and ancestral reconstruction suggests the possibility of four different reasons, namely mutation of the 5′ ancestral translation start site to ATA in the human, the chimpanzee and the gorilla, to ATC in the gibbon, and to GTG in the Diana monkey, and the loss of the 5′ transcription start site in the baboon and the macaque. The difficulty is in determining which events in the evolutionary history of the Catarrhini had functional consequences (i.e., led to the loss of mitochondrial targeting) and which were neutral secondary consequences of an earlier event that led to the loss of mitochondrial targeting (see fig. 4).

This is especially the case for the Diana monkey, the more recent ancestral branches (i.e., 8 and 10) of which contain six nonsynonymous mutations, as well as the loss of the 5′ translation start site (fig. 4). Superimposed on this already complex situation is the predicted loss of the 5′ transcription start site in the baboon/macaque ancestral lineage, which presumably could have occurred in branches 8 or 9 (see fig. 4). Because the cell biological effects of the various amino acid substitutions found in the ancestral MTS are not currently known with certainty, and because their temporal relationships can never be known with certainty, the precipitating event (i.e., that which actually causes loss of mitochondrial AGT targeting) can only be surmised.

Notwithstanding these difficulties, it is likely that mitochondrial AGT targeting has been lost on at least four or five occasions in anthropoid evolution.

Positive Selection Pressure to Lose Mitochondrial Targeting Has Been Widespread in the Recent Evolutionary History of Anthropoids

One of the difficulties in determining the sequence of molecular events that have led to the variable distribution of AGT in primates is an inability to understand the nature of, and to quantify, the selection pressures that might have led to the changes. Ratios of d_N/d_S greater than 1 are often taken as evidence of positive selection pressure (i.e., pressure for change) acting on a sequence of interest (Messier and Stewart 1997). All our analyses point to region 1 (i.e., the MTS or the region encoding the MTS) behaving very differently from region 2 (see figs. 4–6).

When region 1 is included in the open reading frame, it has a function (i.e., directs the targeting of AGT to the mitochondria). It might, therefore, be expected to be evolutionarily constrained, preventing it from acquiring features that interfere with its function. Equally, when this region is excluded from the open reading frame, it is presumed to have no function and therefore would not be so constrained. In the first case, d_N would be expected to be less than d_S. In the second case, it would be expected that there would be no difference in the values of d_N and d_S. On the other hand, if there was selection pressure to interfere with the function of region 1 while it was still in the open reading frame (i.e., to diminish its effectiveness as an MTS), then at stages in evolutionary history it might be expected that d_N would be greater than d_S.

There are many more pairwise comparisons in which d_N/d_S is greater than 1 in region 1* than there are in region 2 (fig. 5). In addition, $d_N/d_S > 1$ is much more common when region 1* is compared between closely related species than when it is compared between distantly related species. This might suggest that the pressures on the MTS might be different in ancient anthropoid evolutionary history compared with more recent evolutionary history. In other words, ancient pressure to conserve mitochondrial AGT targeting might have led to more recent pressure to lose it. If this is generally true, then the Catarrhini would appear to have been more ''successful'' than the Platyrrhini at succumbing to that pressure.

Estimation of d_N/d_S ratios among lineages gives a slightly different, but not incompatible, picture (fig. 6). In region 1*, 8 of the more recent branches have $d_N/d_S > 1$, 3 branches have $d_N/d_S < 1$ (although only slightly so in two), and 11 branches have no substitutions at all. Significantly, the ancestral branch joining the Catarrhini to the Platyrrhini had a d_N/d_S ratio very much lower than one. Statistical analysis confirmed the ancient/recent divide, although clearly not all recent branches show evidence of positive selection pressure. Thus, the adaptive evolution of the region encoding the ancestral MTS of AGT would appear to be episodic, as has been found before with primate lysozyme (Messier and Stewart 1997), but with the adaptive changes being concentrated in more recent lineages.

There are two major difficulties with such analyses, one unique to the MTS of AGT and one more generally applicable to the comparison of short, closely related sequences. First, pressure to lose or diminish mitochondrial AGT targeting could result in two consequences at the sequence level—(1) loss of the 5′ ancestral translation start site, and (2) changes in the amino acid sequence which adversely affect the ability of the ancestral MTS to provide adequate topogenic information. The latter is likely to be an incremental stochastic process and could result in high d_N/d_S ratios. This would be typical of the situation in almost all other proteins studied in which positive selection pressure has been shown to occur. However, the former is an instantaneous switch and would not be reflected in the d_N/d_S ratios. Therefore, high d_N/d_S ratios resulting from positive selection pressure can only occur while the MTS and the 5′ translation start site still exist. Once the translation start site has

been lost, any changes to the region (except perhaps those that might reinstate the MTS) would be expected to be neutral.

Molecular analysis of the anthropoid lineages clearly shows that there is a relationship between the branches with high d_N/d_S ratios (fig. 6) and those in which amino acid replacements are predicted to have occurred (fig. 4). The exceptions are branches 5 and 6, but even in these cases, $d_N/d_S > 1$ if the 5′ initiating methionine is included in the analysis (data not shown—see rationale for excluding this codon from the d_N/d_S analyses in the *Results* section). Therefore, it is possible that in branches where the 5′ translation start site has been predicted to have been lost (i.e., branches 5, 6, 10, and 22; see fig. 4), the high d_N/d_S ratios reflect positive selection pressure to diminish the effectiveness of the MTS before it was lost completely. Why some branches downstream of others in which the 5′ translation start site has been lost should still have $d_N/d_S > 1$ (i.e., branches 2 and 4) is unclear (but see the threshold effect below).

The second difficulty with analyses such as these is a general problem associated with the comparison of small sequences (only 63 nucleotides in the case of region 1*) in closely related species in which only relatively few mutational events have occurred. This is especially noticeable when region 1* is compared within the Platyrrhini, where d_S in most comparisons is zero (fig. 5). This situation is not found in region 2. Although marked differences in d_S across individual genes have been found before (Endo, Ikeo, and Gojobori 1996; Alvarez-Valin, Jabbari, and Bernardi 1998), the reasons remain unclear. As the total number of nonsynonymous sites is usually greater than the number of synonymous sites, a threshold effect could artifactually elevate the d_N/d_S ratio in very closely related short sequences. Although this could conceivably contribute to the high d_N/d_S ratio when the Platyrrhini are compared with each other, it is less likely to contribute significantly to the high d_N/d_S ratio when the Catarrhini are compared with each other.

The Nature of the Selection Pressure Determining AGT Targeting in Primates

Although molecular adaptation is predicted to be a consequence of Darwinian natural selection, there are surprisingly few examples in the literature of it actually having occurred (Golding and Dean 1998). Based on $d_N/d_S > 1$, only relatively few genes, or parts of genes, show any evidence of having been the subject of positive selection pressure (Endo, Ikeo, and Gojobori 1996). Some of the best known examples of genes, or parts of genes, that do show such pressure include primate lysozymes (Messier and Stewart 1997), the antigen recognition sites of MHC genes (Hughes, Ota, and Nei 1990), the surface antigens of parasites and viruses (Endo, Ikeo, and Gojobori 1996), and abalone fertilization proteins (Vacquier, Swanson, and Lee 1997). Even though sequence comparisons might provide strong evidence of positive selection pressure, it is not always easy to identify what that selection pressure actually is.

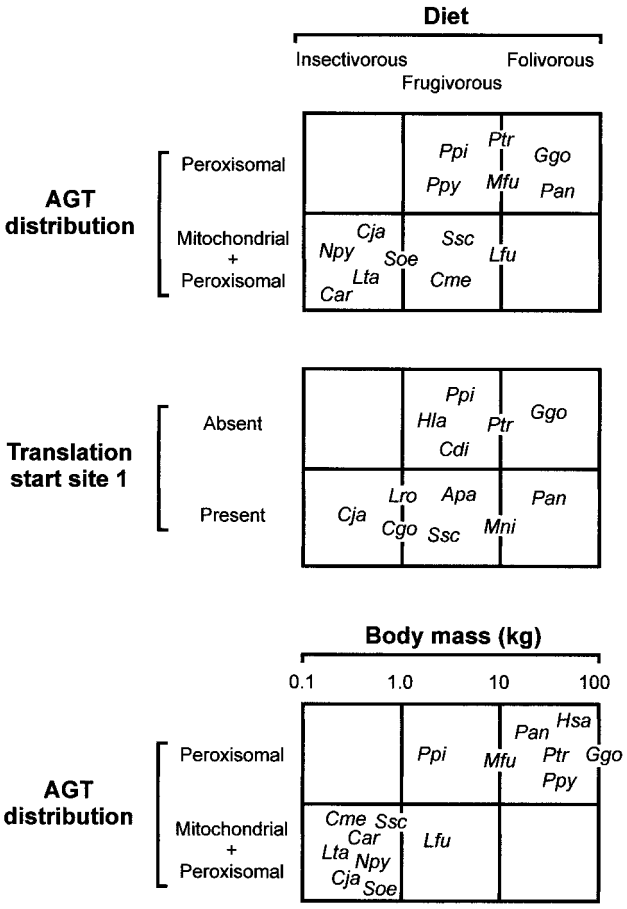


FIG. 7.—AGT distribution and presence or absence of the 5′ ancestral translation start site in primates compared with diet and body mass. Extant diets of primates are divided into three categories: mainly insectivorous, mainly frugivorous, and mainly folivorous (based on information provided in Fleagle [1988] and Nowak [1991]). Subcellular distribution is divided into peroxisomal and mitochondrial + peroxisomal. Body masses were taken from Smith and Jungers (1997). Species abbreviations are defined in table 1. Note that insectivorous primates tend to have both mitochondrial and peroxisomal AGT and have maintained the 5′ ancestral translation start site. However, frugivorous and folivorous primates can have either both mitochondrial and peroxisomal AGT or only peroxisomal AGT. In addition, they have either maintained or lost the 5′ translation start site. The human (*Hsa*) has been omitted from the dietary analysis. Large primates tend to have peroxisomal AGT, whereas the smaller primates (mainly Platyrrhini and Prosimii) tend to have both mitochondrial and peroxisomal AGT. Note that the saki monkey (*Ppi*) is the largest platyrrhine studied and is the only one with a peroxisomal AGT distribution.

There are a number of metabolic reasons for suggesting that diet is the best candidate for the pressure to lose, or decrease the efficiency of, mitochondrial AGT targeting in Mammalia as a whole (Danpure et al. 1994), but whether dietary selection pressure is also responsible for the predicted frequent loss of mitochondrial targeting in Anthropoidea (especially the Catarrhini) is much less clear. Our previous studies on mammals, which included carnivores (including insectivores), herbivores (including frugivores and folivores), omnivores, and almost all combinations, suggested that the optimal distribution of AGT was mitochondrial in carnivores, peroxisomal in herbivores, and both mitochondrial and peroxisomal in omnivores. Unfortunately, the dietary range in Primates,

most of which are opportunistic omnivores, is much smaller (see fig. 7) (Kay 1984). Nevertheless, there appears to be a weak relationship between diet and AGT distribution (or the presence or absence of the 5′ ancestral translation start site) in Primates. For example, insectivorous primates tend to have both mitochondrial and peroxisomal AGT, whereas frugivorous or folivorous primates can have AGT either in both organelles or only in peroxisomes. However, it is not clear whether there is any rational metabolic basis for this dietary relationship in Primates as there is for Mammalia as a whole (Danpure et al. 1994).

Our suggestion (see above) that there has been relatively recent pressure to lose or diminish the function of the MTS of AGT in at least some anthropoid lineages is compatible with various hypotheses which suggest that the diets of primates, especially Anthropoidea, have indeed changed over the past 30–40 Myr or so. Extant members of the suborder are believed to be much larger than ancestral members, and this increase in body size is thought to accompany a shift in diet from one that is more insectivorous to one that is more folivorous (Kay and Simons 1980; Kay 1984; Fleagle and Kay 1985). Changes in the dentition of the Cercopithecidae also indicate relatively recent changes in diet (Fleagle 1988). In addition, Catarrhines tend to be larger than Platyrrhines (Smith and Jungers 1997) (see fig. 7), a difference that also parallels a dietary shift from being mainly insectivorous and frugivorous to being mainly frugivorous and folivorous (Kay 1984; Harvey, Martin, and Clutton-Brock 1987) and the observation that the Catarrhines have been rather more 'successful' at losing mitochondrial AGT than have the Platyrrhines.

Thus, although the nature of the evolutionary selection pressure leading to changes in AGT distribution in Anthropoidea is less easily identified than it is within Mammalia as a whole, it is clear that the variable compartmentalization of AGT is a unique and remarkable example of molecular adaptation as a consequence of positive (probably dietary) selection pressure.

Acknowledgments

We would like to thank Leslie Aiello, Department of Anthropology, University College London, for helpful discussions about the evolution of primate anatomy and lifestyles, Caro-Beth Stewart, Department of Biological Sciences, State University of New York at Albany, for helpful comments concerning an earlier draft of our paper, and Andrew Cunningham, Veterinary Science Unit, Institute of Zoology, University of London, for providing some of the samples. We acknowledge the financial support of the U.K. Medical Research Council (J.D.H., G.M.B., and C.J.D.) and the U.K. Biotechnology and Biological Sciences Research Council (Z.Y.).

LITERATURE CITED

ALVAREZ-VALIN, F., K. JABBARI, and G. BERNARDI. 1998. Synonymous and nonsynonymous substitutions in mammalian genes: intragenic correlations. J. Mol. Evol. **46**:37–44.

BIRDSEY, G. M., and C. J. DANPURE. 1998. Evolution of alanine : glyoxylate aminotransferase intracellular targeting. Structural and functional analysis of the guinea pig gene. Biochem. J. **331**:49–60.

DANPURE, C. J. 1995. How can the products of a single gene be localized to more than one intracellular compartment? Trends Cell Biol. **5**:230–238.

———. 1997. Variable peroxisomal and mitochondrial targeting of alanine : glyoxylate aminotransferase in mammalian evolution and disease. Bioessays **19**:317–326.

DANPURE, C. J., P. J. COOPER, P. J. WISE, and P. R. JENNINGS. 1989. An enzyme trafficking defect in two patients with primary hyperoxaluria type 1: peroxisomal alanine/glyoxylate aminotransferase rerouted to mitochondria. J. Cell Biol. **108**:1345–1352.

DANPURE, C. J., P. FRYER, P. R. JENNINGS, J. ALLSOP, S. GRIFFITHS, and A. CUNNINGHAM. 1994. Evolution of alanine : glyoxylate aminotransferase 1 peroxisomal and mitochondrial targeting. A survey of its subcellular distribution in the livers of various representatives of the classes Mammalia, Aves and Amphibia. Eur. J. Cell Biol. **64**:295–313.

DANPURE, C. J., K. M. GUTTRIDGE, P. FRYER, P. R. JENNINGS, J. ALLSOP, and P. E. PURDUE. 1990. Subcellular distribution of hepatic alanine : glyoxylate aminotransferase in various mammalian species. J. Cell Sci. **97**:669–678.

DANPURE, C. J., and P. E. PURDUE. 1995. Primary hyperoxaluria. Pp. 2385–2424 *in* C. R. SCRIVER, A. L. BEAUDET, W. S. SLY, and D. VALLE, eds. The metabolic and molecular bases of inherited disease. McGraw-Hill, New York.

DEVEREUX, J., P. HAEBERLI, and O. SMITHIES. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.

ENDO, T., K. IKEO, and T. GOJOBORI. 1996. Large-scale search for genes on which positive selection may operate. Mol. Biol. Evol. **13**:685–690.

FLEAGLE, J. G. 1988. Primate adaptation and evolution. Academic Press, San Diego.

FLEAGLE, J. G., and R. F. KAY. 1985. The paleobiology of catarrhines. Pp. 23–36 *in* E. DELSON, ed. Ancestors: the hard evidence. Alan R. Liss, New York.

GOLDING, G. B., and A. M. DEAN. 1998. The structural basis of molecular adaptation. Mol. Biol. Evol. **15**:355–369.

GOLDMAN, N., and Z. YANG. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**:725–736.

HARVEY, P. H., R. D. MARTIN, and T. H. CLUTTON-BROCK. 1987. Life histories in comparative perspective. Pp. 181–196 *in* B. B. SMUTS, D. L. CHENEY, R. M. SEYFARTH, R. W. WRANGHAM, and T. T. STRUHSAKER, eds. Primate societies. University of Chicago Press, Chicago.

HAYASHI, S., H. SAKURABA, and T. NOGUCHI. 1989. Response of hepatic alanine : glyoxylate aminotransferase 1 to hormone differs among mammalia. Biochem. Biophys. Res. Commun. **165**:372–376.

HUGHES, A. L., T. OTA, and M. NEI. 1990. Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol. Biol. Evol. **7**:515–524.

KAY, R. F. 1984. On the use of anatomical features to infer foraging behaviours in extinct primates. Pp. 21–53 *in* P. S. RODMAN and J. G. CANT, eds. Adaptations for foraging in nonhuman primates. Columbia University Press, New York.

KAY, R. F., and E. L. SIMONS. 1980. The ecology of oligocene African Anthropoidea. Int. J. Primatol. **1**:21–37.

LUMB, M. J., P. E. PURDUE, and C. J. DANPURE. 1994. Molecular evolution of alanine/glyoxylate aminotransferase 1 in-

tracellular targeting: analysis of the feline gene. Eur. J. Biochem. **221**:53–62.

MESSIER, W., and C. B. STEWART. 1997. Episodic adaptive evolution of primate lysozymes. Nature **385**:151–154.

MIYAJIMA, H., T. ODA, and A. ICHIYAMA. 1989. Induction of mitochondrial serine : pyruvate aminotransferase of rat liver by glucagon and insulin through different mechanisms. J. Biochem. Tokyo **105**:500–504.

MOTLEY, A., M. J. LUMB, P. B. OATEY, P. R. JENNINGS, P. A. DE ZOYSA, R. J. WANDERS, H. F. TABAK, and C. J. DANPURE. 1995. Mammalian alanine : glyoxylate aminotransferase 1 is imported into peroxisomes via the PTS1 translocation pathway. Increased degeneracy and context specificity of the mammalian PTS1 motif and implications for the peroxisome-to-mitochondrion mistargeting of AGT in primary hyperoxaluria type 1. J. Cell Biol. **131**:95–109.

NISHIYAMA, K., G. BERSTEIN, T. ODA, and A. ICHIYAMA. 1990. Cloning and nucleotide sequence of cDNA encoding human liver serine-pyruvate aminotransferase. Eur. J. Biochem. **194**:9–18.

NOWAK, R. M. 1991. Walker's mammals of the world. John Hopkins University Press, Baltimore.

OATEY, P. B., M. J. LUMB, and C. J. DANPURE. 1996. Molecular basis of the variable mitochondrial and peroxisomal localisation of alanine : glyoxylate aminotransferase. Eur. J. Biochem. **241**:374–385.

ODA, T., T. FUNAI, and A. ICHIYAMA. 1990. Generation from a single gene of two mRNAs that encode the mitochondrial and peroxisomal serine : pyruvate aminotransferase of rat liver. J. Biol. Chem. **265**:7513–7519.

ODA, T., H. MIYAJIMA, Y. SUZUKI, and A. ICHIYAMA. 1987. Nucleotide sequence of the cDNA encoding the precursor for mitochondrial serine : pyruvate aminotransferase of rat liver. Eur. J. Biochem. **168**:537–542.

ODA, T., M. YANAGISAWA, and A. ICHIYAMA. 1982. Induction of serine : pyruvate aminotransferase in rat liver organelles by glucagon and a high-protein diet. J. Biochem. (Tokyo) **91**:219–232.

PAGE, R. D. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12**:357–358.

PURDUE, P. E., M. J. LUMB, and C. J. DANPURE. 1992. Molecular evolution of alanine : glyoxylate aminotransferase 1 intracellular targeting. Analysis of the marmoset and rabbit genes. Eur. J. Biochem. **207**:757–766.

PURVIS, A. 1995. A composite estimate of primate phylogeny. Philos. Trans. R. Soc. Lond. B Biol. Sci. **348**:405–421.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York.

SCHNEIDER, G., S. SJOLING, E. WALLIN, P. WREDE, E. GLASER, and G. VON HEIJNE. 1998. Feature-extraction from endopeptidase cleavage sites in mitochondrial targeting peptides. Proteins **30**:49–60.

SMITH, R. J., and W. L. JUNGERS. 1997. Body mass in comparative primatology. J. Hum. Evol. **32**:523–559.

STEIN, I., Y. PELEG, S. EVEN RAM, and O. PINES. 1994. The single translation product of the FUM1 gene (fumarase) is processed in mitochondria before being distributed between the cytosol and mitochondria in Saccharomyces cerevisiae. Mol. Cell. Biol. **14**:4770–4778.

TAKADA, Y., N. KANEKO, H. ESUMI, P. E. PURDUE, and C. J. DANPURE. 1990. Human peroxisomal L-alanine : glyoxylate aminotransferase. Evolutionary loss of a mitochondrial targeting signal by point mutation of the initiation codon. Biochem. J. **268**:517–520.

TAKADA, Y., and T. NOGUCHI. 1982. Subcellular distribution, and physical and immunological properties of hepatic alanine: glyoxylate aminotransferase isoenzymes in different mammalian species. Comp. Biochem. Physiol. B **72**:597–604.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

VACQUIER, V. D., W. J. SWANSON, and Y. H. LEE. 1997. Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? J. Mol. Evol. **44**(Suppl. 1):S15–S22.

VON HEIJNE, G. 1986. Mitochondrial targeting sequences may form amphiphilic helices. EMBO J. **5**:1335–1342.

YANG, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.

YANG, Z., S. KUMAR, and M. NEI. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics **141**:1641–1650.

SIMON EASTEAL, reviewing editor

Accepted November 8, 1999