

**MOLECULAR EVOLUTION OF THE
INTRACELLULAR TARGETING OF
ALANINE: GLYOXYLATE
AMINOTRANSFERASE**

A thesis presented by Joanna Dawn Holbrook for the
Degree of Doctor of Philosophy

University College London

2001

ProQuest Number: 10016050

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10016050

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code.
Microform Edition © ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The subcellular distribution of the hepatic metabolic enzyme alanine:glyoxylate aminotransferase (AGT) has changed on numerous occasions throughout the evolution of mammals. The AGT distribution in the livers of these species seems to show a relationship to diet. Thus AGT tends to be mitochondrial in carnivores, peroxisomal in herbivores and both mitochondrial and peroxisomal in omnivores. The archetypal mammalian AGT gene has the potential to encode an N-terminal mitochondrial targeting sequence (MTS) and a C-terminal peroxisomal targeting sequence type 1 (PTS1). The variable subcellular distribution of AGT results from the variable use of two alternative transcription and translation start sites which include or exclude the MTS from the open reading frame (ORF).

The functionality and evolution of the targeting sequences of AGT are extraordinary. The MTS of AGT is often excluded from the ORF and the PTS1 is protein context dependent. This study further investigates the following:

- 1. The molecular mechanism for the variable distribution of AGT in different species**
- 2. The selective pressure that acts on the targeting of AGT to cause its variable localisation**

The results presented in this thesis show the following:

The molecular explanation for the unusual mitochondrial and cytosolic distribution of amphibian AGT was determined, using the xenopus as a representative species. It is the first AGT gene studied, which does not encode a peroxisomal targeting sequence.

Xenopus AGT has the potential to encode two RNA transcripts. One of these includes the MTS and the other has no targeting information.

Two novel molecular mechanisms for a solely peroxisomal distribution of AGT were suggested for certain primate species. These were the loss of a transcription start site excluding the MTS from the RNA transcript, and accumulation of mutations in the MTS that are incompatible with its function.

The context specificity of PTS1s, in general, was explored by undertaking an extensive literature survey of C-terminal tripeptides that have some characteristics of PTS1s. This information was related to the peroxisomal targeting of AGT.

Analysis of the relative number of synonymous and nonsynonymous mutations accumulated in the region of primate AGT genes, encoding the MTS, suggested that there has been recent strong, yet episodic, positive selection pressure to lose, or diminish, mitochondrial AGT targeting in anthropoid primates. The selection pressure is possibly connected to diet.

Non-mammalian AGT homologues were identified from sequence databases. These, along with the xenopus AGT data, suggested a new hypothesis for the ancestral distribution of AGT.

Stronger evidence for the theory of diet as the selective pressure driving the evolution of AGT subcellular targeting was acquired by comparative analysis of the distribution of AGT and diet, which showed a significant correlation.

Acknowledgements

I owe very sincere thanks to Mike Lumb. This work really would not have been done without his technical guidance, patience and good humour, especially his patience.

Another person I would like to thank is Graeme Birdsey who has helped me a lot this last summer.

The most prominent people amongst the many people who have taught me about their field of expertise are Ziheng Yang, Mark Pagel and Sylvia Naga. I am very grateful to these generous people for their time and willingness to help. I am also grateful to my parents for believing me more than capable; Susanne Brink and Marie Harrisingh for understanding; and Hilary Clark for Tuesday lunchtimes.

The most influential people in my life for the last three years have been Chris, Heather and Alex. I have a lot to thank my supervisor, Chris Danpure, for. He has always been available to offer me advice and discuss my ideas. His high standards and attention to detail have taught me about being a scientist. Heather has, as always, been the one I rely on to expand my sense of proportion and she has never failed me. My last debt of gratitude is to Alex, for confidence both my own and his in me.

Contents

| | |
|---|-----------|
| Title page | |
| Abstract | 1 |
| Acknowledgements | 3 |
| Contents | 4 |
| List of figures | 9 |
| List of tables | 13 |
| Abbreviations | 14 |
| Glossary | 16 |
| 1. Introduction | 19 |
| 1.1 Compartmentalisation of eukaryotic cells | 19 |
| 1.2 Mitochondrial targeting sequences | 19 |
| 1.3 Peroxisomal targeting sequences | 23 |
| 1.4 The dual localisation of AGT protein | 26 |
| 1.5 Molecular basis of dual AGT targeting | 27 |
| 1.5.1 The context specificity of peroxisomal targeting of human AGT | 33 |
| 1.6 Metabolic function of AGT protein | 33 |
| 1.7 An adaptive explanation for variable AGT targeting | 38 |
| 1.8 Aims of this study | 39 |
| 2. Experimental and Analytical Methods | 40 |
| 2.1 Materials | 40 |
| 2.1.1 Animal tissues | 40 |
| 2.1.2 <i>Escherichia coli</i> genotypes | 40 |
| 2.2 Oligonucleotide primers | 42 |
| 2.2.1 Primate primers | 42 |
| 2.2.2 <i>Xenopus</i> primers | 42 |
| 2.3 Cloning strategies | 42 |
| 2.3.1 Primate cloning strategy | 42 |
| 2.3.2 <i>Xenopus</i> cloning strategy | 45 |
| 2.4 Sequencing strategies | 48 |
| 2.4.1 Primate sequencing strategy | 48 |
| 2.4.2 <i>Xenopus</i> sequencing strategy | 48 |
| 2.5 EMBL numbers | 49 |

| | | |
|------------|--|-----------|
| 2.6 | Experimental methods | 49 |
| 2.6.1 | Extraction of genomic DNA from animal tissue | 49 |
| 2.6.2 | Extraction of total RNA from animal tissue | 50 |
| 2.6.3 | Polymerase chain reaction | 50 |
| 2.6.4 | Dideoxy sequencing | 51 |
| 2.6.5 | Automated cycle sequencing | 51 |
| 2.6.6 | Southern blotting | 51 |
| 2.6.7 | PCR of xenopus library to clone 5' end of gene | 52 |
| 2.6.8 | Rapid amplification of cDNA ends (RACE) | 52 |
| 2.6.9 | RNase protection | 53 |
| 2.6.10 | <i>In vitro</i> transcription and translation | 53 |
| 2.6.11 | Immunoblotting | 53 |
| 2.6.12 | AGT enzyme assay | 54 |
| 2.6.13 | Expression library screening | 55 |
| 2.6.14 | Tissue culture | 56 |
| 2.6.15 | Transfection of COS cells | 56 |
| 2.6.16 | Harvesting of protein from COS cells and animal tissues | 56 |
| 2.6.17 | Immunofluorescence microscopy | 56 |
| 2.7 | Analytical methods | 57 |
| 2.7.1 | Bioinformatics | 57 |
| 2.7.2 | Maximum likelihood | 58 |
| 2.7.3 | Ancestral reconstruction | 58 |
| 2.7.4 | d_N/d_S ratios | 58 |
| 2.7.5 | Comparative analysis | 59 |
| 3. | Molecular Adaptation of AGT Targeting in Primates | 60 |
| 3.1 | Introduction | 60 |
| 3.2 | Results | 62 |
| 3.2.1 | Cloning and sequencing of the 5' region of the primate AGT genes | 62 |
| 3.2.2 | Immunoblot of human and baboons livers | 67 |
| 3.2.3 | Ancestral reconstruction of region 1 | 69 |
| 3.2.4 | The d_N/d_S ratios of regions 1* and region 2 | 69 |
| 3.2.4.1 | Region 2 | 70 |
| 3.2.4.2 | Region 1* | 72 |

| | |
|--|-----------|
| 3.2.5 Analysis of the d_N/d_S ratios among lineages | 76 |
| 3.2.5.1 Region 1* | 76 |
| 3.2.5.2 Region 2 | 79 |
| 3.3 Discussion | 81 |
| 3.3.1 The presence or absence of TL-1 is often, but not always correlated with AGT protein distribution | 81 |
| 3.3.2 Possible temporal relationship of the mutational events leading to the varied subcellular distribution of AGT in extant primates | 84 |
| 3.3.3 Positive selection procedure to lose mitochondrial targeting has been widespread in the evolutionary history of anthropoids | 85 |
| 3.3.4 Theoretical problems with these analyses | 86 |
| 3.3.4.1 Problem 1 (chronology) | 87 |
| 3.3.4.2 Problem 2 (threshold effect) | 88 |
| 3.3.4.3 Problem 3 (region 1 vs region 1*) | 88 |
| 3.3.4.4 Problem 4 (some substitutions are more nonsynonymous than others) | 89 |
| 3.3.4.5 Problem 5 (methodology) | 89 |
| 3.3.5 The nature of the selection pressure determining AGT targeting in primates | 90 |
| 4. A Molecular Explanation for the Unusual Distribution of AGT in Amphibian Liver | 95 |
| 4.1 Introduction | 95 |
| 4.2 Results | 97 |
| 4.2.1 Cloning the xenopus AGT gene | 97 |
| 4.2.1.1 Screening an expression library | 97 |
| 4.2.1.2 PCR and Southern blotting analysis of the xenopus expression library | 100 |
| 4.2.1.3 5' RACE of pXS | 102 |
| 4.2.1.4 RT-PCR of the xenopus AGT | 102 |
| 4.2.2 Analysis of the xenopus AGT sequence | 104 |
| 4.2.3 <i>In vitro</i> transcription and translation of the xenopus clones | 104 |
| 4.2.4 Immunoblotting of amphibian liver | 110 |
| 4.2.5 Immunoblotting of COS cells transfected with AGT clones | 110 |
| 4.2.6 AGT enzymatic activity in transfected COS cells | 110 |

| | |
|--|------------|
| 4.2.7 RNase protection | 112 |
| 4.2.8 Immunofluorescence of xenopus AGT in COS cells | 112 |
| 4.2.9 The subcellular distribution of proteins encoded by various artificial xenopus AGT constructs | 117 |
| 4.3 Discussion | 128 |
| 4.3.1 The unexpected orientation of pXS may be due to recombination | 128 |
| 4.3.2 pXall contains the entire reading frame of the xenopus AGT gene | 129 |
| 4.3.3 The molecular explanation for the mitochondrial and cytosolic distribution of AGT in xenopus liver | 131 |
| 4.3.4 The size of xenopus AGT protein is larger <i>in vivo</i> than <i>in vitro</i> | 134 |
| 4.3.5 Enzymatic activity of xenopus AGT protein is higher than that of human | 137 |
| 4.3.6 Xenopus AGT lacks the additional peroxisomal targeting information thought to be present in the human AGT protein | 138 |
| 4.4 Summary | 140 |
| 5. Evolution of eukaryotic AGT and it's targeting signals | 141 |
| 5.1 Introduction | 141 |
| 5.2 Results and discussion | 143 |
| 5.2.1 The MTS of AGT | 143 |
| 5.2.2 The PTS1 of AGT | 146 |
| 5.2.2.1 A new consensus | 152 |
| 5.2.2.2 Context dependent PTS1s | 153 |
| 5.2.2.3 Variability of PTS1s | 158 |
| 5.2.3 Recovery of homologues from the database | 160 |
| 5.2.4 Recovery of human paralogues from the database | 160 |
| 5.2.5 Classification of homologues | 160 |
| 5.2.5.1 Test 1 | 163 |
| 5.2.5.2 Test 2 | 165 |
| 5.2.6 Targeting information of eukaryotic AGT orthologues | 167 |
| 5.2.7 Alignment of AGT orthologues | 172 |
| 6. Comparative Analysis of AGT and Life Style Traits | 176 |
| 6.1 Introduction | 176 |

| | |
|---|------------|
| 6.2 Results and discussion | 177 |
| 6.2.1 Acquisition of data set | 177 |
| 6.2.2 The comparative analysis test controlling for phylogeny | 186 |
| 6.2.3 AGT distribution and diet | 187 |
| 6.2.4 AGT activity and diet | 189 |
| 6.2.5 Levels of immunoreactive AGT and AGT catalytic activity | 190 |
| 6.3 Summary | 191 |
| 7. Contributions and Future Directions | 192 |
| 7.1 Contributions of this study to our understanding of the evolution of AGT targeting | 192 |
| 7.1.1 Selection pressure | 192 |
| 7.1.2 Mechanisms of AGT targeting | 193 |
| 7.2 Directions for the future | 193 |
| Bibliography | 195 |
| Appendix | 204 |
| Reference list for Table 11 | 211 |

List of figures

| Number | Section | Figure title | Page |
|--------|---------|--|------|
| 1 | 1.2 | The mitochondrial import machinery | 22 |
| 2 | 1.3 | The peroxisomal import machinery | 25 |
| 3 | 1.4 | The subcellular distribution of hepatic AGT protein arranged on a species phylogeny | 28 |
| 4 | 1.5 | Archetypal mammalian AGT gene | 30 |
| 5 | 1.5 | Schematic diagrams of human and cat AGT genes | 31 |
| 6 | 1.5.1 | The context specificity of peroxisomal AGT targeting | 34 |
| 7 | 1.6 | Metabolic roles of AGT | 35 |
| 8 | 2.3.2 | Schematic diagram of xenopus cloning strategy | 46 |
| 9 | 2.3.2 | Cloning strategy for p <i>Hs</i> AGT-KKM | 47 |
| 10 | 3.1 | Archetypal mammalian AGT gene with region 1 and 2 indicated | 61 |
| 11 | 3.2.1 | Nucleotide sequence of the 5' region of the AGT gene in primates | 63 |
| 12 | 3.2.1 | Deduced amino acid sequence encoded by the 5' region of the AGT gene in primates | 65 |
| 13 | 3.2.1 | Pairwise percentage identities of the 5' region of primate AGT genes | 66 |
| 14 | 3.2.2 | Immunoblot of baboon and human livers | 68 |
| 15 | 3.2.3 | Ancestral reconstruction of region 1 amino acid substitutions in primate lineages | 71 |
| 16 | 3.2.4 | The d_N/d_S substitution rates of the 5' region of the AGT gene in primates | 73 |
| 17 | 3.2.4.1 | Exon specific conservation of mammalian AGT genes | 75 |
| 18 | 3.2.5 | d_N/d_S among lineages in the 5' region of the AGT gene in primates | 77 |
| 19 | 3.3.5 | AGT distribution and presence or absence of TL-1 in primates compared with diet and body mass | 92 |
| 20 | 3.3.5 | Representation of adaptive hypothesis of primate AGT evolution | 93 |
| 21 | 4.1 | Immuno electron microscopy of amphibian livers | 96 |

| | | | |
|----|---------|---|------------|
| 22 | 4.2.1.1 | Cloning vector of xenopus cDNA library | 99 |
| 23 | 4.2.1.2 | Southern blot of PCR products from xenopus library PCR | 101 |
| 24 | 4.2.1.3 | Products of 5'RACE reaction | 103 |
| 25 | 4.2.2 | Nucleotide and predicted amino acid sequence of xenopus AGT | 105 |
| 26 | 4.2.2 | Alignment of xenopus AGT peptide sequence and human AGT peptide sequence | 107 |
| 27 | 4.2.3 | Characterisation of xenopus clones | 108 |
| 28 | 4.2.7 | RNase protection of the xenopus AGT gene | 114 |
| 29 | 4.2.8 | AGT distribution in COS cells following transfection with <i>pHsAGT</i> | 115 |
| 30 | 4.2.8 | AGT distribution in COS cells following transfection with <i>pHsAGT</i> ⁺ | 116 |
| 31 | 4.2.8 | AGT distribution in COS cells following transfection with pXL (1) | 119 |
| 32 | 4.2.8 | AGT distribution in COS cells following transfection with pXL (2) | 120 |
| 33 | 4.2.8 | AGT distribution in COS cells following transfection with pXS (1) | 121 |
| 34 | 4.2.8 | AGT distribution in COS cells following transfection with pXS (2) | 122 |
| 35 | 4.2.9 | GFP distribution in COS cells following transfection with pXmtsGFP | 123 |
| 36 | 4.2.9 | AGT distribution in COS cells following transfection with <i>pHsAGT-KKM</i> | 124 |
| 37 | 4.2.9 | AGT distribution in COS cells following transfection with pXS-SKL | 125 |
| 38 | 4.2.9 | AGT distribution in COS cells following transfection with pXS-KKL | 126 |
| 39 | 4.2.9 | AGT distribution in COS cells following transfection with pXL-KKL | 127 |
| 40 | 4.3.3 | Schematic diagram of the xenopus AGT gene | 135 |
| 41 | 5.2.1 | Helical analysis of N-terminal amino acids of known AGT orthologues | 144 |

| | | | |
|----|---------|--|------------|
| 42 | 5.2.1 | The position specific constraints on AGT MTS evolution in mammals | 145 |
| 43 | 5.2.5.1 | Clustering relationships of AGT homologues | 164 |
| 44 | 5.2.5.2 | Clustering relationships of the putative AGT orthologues | 166 |
| 45 | 5.2.6 | Helical wheel analysis of the N-terminal amino acids of putative <i>C. elegans</i> AGT | 168 |
| 46 | 5.2.6 | Theoretical evolution of AGT targeting | 171 |
| 47 | 5.2.7 | Multiple alignment of consensus orthologue sequences | 173 |
| 48 | 6.2.1 | The relationships between diet and AGT activity or subcellular distribution | 180 |
| 49 | 6.2.1 | The relationships between body size and AGT activity or subcellular distribution | 181 |
| 50 | 6.2.1 | The relationships between length of gestation and AGT activity or subcellular distribution | 182 |
| 51 | 6.2.1 | The relationships between the level of AGT immunoreactive protein AGT activity or subcellular distribution | 183 |
| 52 | 6.2.1 | The relationship between AGT activity and AGT subcellular distribution | 184 |
| 53 | 6.2.2 | Results of comparative analysis of mammalian species traits | 188 |
| A1 | 3.2.4 | The d_N/d_S substitution rates of region 1 of the AGT gene in primates (with TL-1 included in the analysis) | 205 |
| A2 | 3.2.5 | d_N/d_S among lineages in the 5' region of the AGT gene in primates with TL-1 included in the analysis | 207 |
| A3 | 3.3.4.5 | The d_N/d_S substitution rates of the 5' region of the AGT gene in primates, as estimated by the method of Li (1993) | 208 |
| A4 | 3.3.4.5 | The d_N/d_S substitution rates of the 5' region of the AGT gene in primates, as estimated by the method of Nei and Gojobori (1986) | 209 |
| A5 | 4.2.8 | Colour images of the dual labelled immunofluorescence experiments | 210 |
| A6 | 6.2.2 | Results of comparative analysis of mammalian species | |

| | | | |
|----|-------|---|------------|
| | | traits using a phylogeny generated from a taxonomy (figure A6) | 220 |
| A7 | 6.2.2 | Taxonomy of mammalian species | 221 |
| A8 | 6.2.2 | Phylogeny of mammalian species based on mitochondrial DNA | 222 |

List of tables

| Number | Section | Table title | Page |
|--------|---------|--|------|
| 1 | 1.5 | Previously cloned AGT genes and the subcellular distributions of their gene products | 32 |
| 2 | 2.1.1 | Species discussed in this study | 41 |
| 3 | 2.2 | Oligonucleotide primers | 43 |
| 4 | 3.2.4 | Mean values for d_N and d_S in pairwise comparisons | 74 |
| 5 | 3.2.5 | Log-likelihood values and maximum likelihood estimates of parameters under different models | 48 |
| 6 | 3.2.5 | Tests of evolutionary hypotheses | 80 |
| 7 | 4.2.1.1 | BLAST searching the library screen positives | 98 |
| 8 | 4.2.6 | AGT activities of transfected COS cells | 111 |
| 9 | 4.3.3 | Subcellular distribution of clones in COS cells, as elucidated by immunofluorescent microscopy | 132 |
| 10 | 4.3.4 | Estimated sizes of xenopus AGT proteins | 133 |
| 11 | 5.2.2 | C-terminal tripeptides with at least some properties of PTS1s | 148 |
| 12 | 5.2.2 | Frequency of certain residues at positions with PTS1s | 154 |
| 13 | 5.2.2.1 | C-terminal tripeptides in functional human, <i>S. cerevisiae</i> and trypanosome systems | 155 |
| 14 | 5.2.2.2 | The variability of PTS1s found on mammalian peroxisomal enzymes | 159 |
| 15 | 5.2.3 | AGT homologues recovered from BLAST searches | 161 |
| 16 | 5.2.4 | AGT paralogues and their percentage identity to human AGT | 162 |
| 17 | 5.2.6 | The targeting information of C-terminal tripeptides of AGT orthologues | 170 |
| 18 | 6.2.1 | AGT and life style traits examined by comparative analysis | 178 |
| A1 | 3.3.4.3 | Mean values for d_N and d_S in pairwise comparisons | 206 |
| A2 | 5.2.5.1 | Homologue test 2 | 217 |
| A3 | 5.2.5.2 | Homologue test 2 | 217 |
| A4 | 6.2.2 | Data for comparative analysis | 218 |

Abbreviations

| | |
|---------|--|
| AGT | Alanine glyoxylate aminotransferase |
| CAT | Chloramphenicol acetyl transferase |
| d_N | Nonsynonymous substitution rate per nonsynonymous site |
| d_S | Synonymous substitution rate per synonymous site |
| ECL | Enhanced chemiluminescence |
| GFP | Green fluorescent protein |
| HIS tag | 6 Histidine residue epitope tag |
| Hsp | Heat shock protein |
| LR | Likelihood ratio statistic |
| MGE | Mitochondrial ATP/ADP exchange factor |
| MPP | Mitochondrial processing peptidase |
| MSF | Mitochondrial import stimulation factor |
| MTS | Mitochondrial targeting sequence |
| My | Million years |
| ORF | Open reading frame |
| PAGE | Polyacrylamide gel electrophoresis |
| PBS | Phosphate buffered saline |
| PEX | Peroxin genes |
| PH1 | Primary hyperoxaluria type 1 |
| PTS1 | Peroxisomal targeting sequence type 1 |
| PTS2 | Peroxisomal targeting sequence type 2 |
| RACE | Rapid amplification of cDNA ends |

| | |
|----------------|------------------------------------|
| SD | Standard deviation |
| SDS | Sodium dodecyl sulphate |
| TIM | Translocator of the inner membrane |
| T _M | Melting temperature |
| TOM | Translocator of the outer membrane |
| TPR | Tetratricopeptide repeat |
| UTR | Untranslated region |

Glossary

| | |
|------------------------|--|
| Anthropoids | Primates including the catarrhines (old world monkeys and apes) and the platyrrhines (new world monkeys) but excluding the prosimians. |
| Callitrichidae | New world monkeys such as the marmoset. |
| Carnivore | Meat eating; for the purposes of this study including eggs and insects |
| Catarrhines | Old world monkeys and apes. |
| Cebidae | New world monkeys such as the saki monkey. |
| Cercopithecidae | Old world monkeys such as the baboon. |
| COS | SV40 transformed green monkey kidney cells. |
| Folivorous | Leaf eating. |
| Frugivorous | Fruit eating. |
| Herbivore | Plant eating; for the purposes of this study including sap and resin. |
| Homidae | Old world apes. |
| Insectivorous | Insect eating. |
| Orthologue | Evolutionarily related gene or protein found in different species but with a similar function, e.g. human and rabbit AGT are orthologous. |
| Paralogue | Evolutionarily related gene or protein with different function, e.g. human AGT and human AAT are paralogous as are human AGT and yeast AAT. |
| Phylogenetic inference | The tendency for closely related species to have similar traits due to the short time in which they have been diverging from a common ancestor . |
| pHsAGT | Human AGT in pcDNA3. |
| pHsAGT ⁺ | Human AGT with TL-1 reinstated to include the ancestral MTS in the ORF, in pcDNA3. |
| Platyrrhines | New world monkeys. |

| | |
|------------|--|
| pRACE-L | The long RACE product in pGEM-T Easy. |
| pRACE-S | The short RACE product in pGEM-T Easy. |
| pXL | Long xenopus AGT clone in pcDNA3, including both translation start sites and the MTS. |
| pXL-5' | The 5' end of XL, from 5' of the <i>EcoRV</i> site, including both translation start sites, in pcDNA3. |
| pXL-KKL | XL with the sequence encoding the C-terminal tripeptide changed to code for KKL, in pcDNA3. |
| pXmtsGFP | The MTS of xenopus AGT, including TL-1 and up to but not including TL-2, fused to GFP, in pcDNA3. |
| pXS | Short xenopus AGT clone in pcDNA3, including only one translation start site; identical to library clone 4. |
| pXS-3' | The 3' end of XS, from 3' of the <i>XbaI</i> site, including the C-terminal tripeptide and polyA tail, in pcDNA3. |
| pXS-KKL | XS with the sequence encoding the C-terminal tripeptide changed to code for KKL in pcDNA3. |
| pXS-SKL | XS with the sequence encoding the C-terminal tripeptide changed to code for SKL in pcDNA3. |
| pXSΔ3' | XS without its 3' end, 3' of the <i>XbaI</i> site, in pBluescript. |
| Region 1 | The region of the primate AGT genes, between the translation start sites, including TL-1 and up to but not including TL-2. |
| Region 1* | As region 1 but not including TL-1. |
| Region 2 | Exon 1 of the primate AGT genes excluding the region 5' of TL-2. |
| TA cloning | The cloning of PCR products into the pGEM-T Easy vector using the overhanging A nucleotides at the 5' and 3' ends. |
| TL-1 | The more 5' translation start site, translation start site 1 (figure 4). |

| | |
|----------|---|
| TL-2 | The more 3' translation start site, translation start site 2 (figure 4). |
| TS-A | The more 5' transcription start site, transcription start site A (figure 4). |
| TS-B | The more 3' transcription start site, transcription start site B (figure B). |
| Xall | The PCR product including nearly all of the xenopus AGT cDNA sequence known. It has never been possible to clone it.. |
| ω | d_N/d_S ratio. |

1. Introduction

1.1 Compartmentalisation of eukaryotic cells

An ancient and defining feature of eukaryotes is the sequestration of proteins into intracellular compartments. Organisation of the interior of a cell in such a manner allows the enzyme environment to be optimised, metabolic pathways to be more tightly regulated, substrates to be concentrated and toxins to be kept away from vulnerable cellular components.

Some intracellular compartments are termed organelles. Organelles contain subsets of cellular protein, bounded by their membranes. Most of the proteins within organelles are encoded by nuclear DNA and synthesised in the cytosol. So organellar proteins must be selected from thousands of cellular proteins and directed down discrete trafficking pathways to their ultimate subcellular location. Trafficking pathways start with the interaction between the targeting sequences within polypeptides and their cognate receptors. Proteins that lack targeting signals remain in the cytosol, by default.

1.2 Mitochondrial targeting sequences

Proteins destined for the mitochondrial matrix, and many inner membrane and intermembrane space proteins, are synthesised as precursor peptides with mitochondrial targeting sequences (MTS) at their amino termini^{1;2}. MTSs are defined as amino acid sequences that are both necessary and sufficient to direct proteins to mitochondria. More often than not MTSs are cleaved after import³. Following import into the

mitochondrial matrix, part of the N-terminus (often between 10-50 amino acids) including at least part of the MTS is cleaved. While having no significant primary sequence identity, MTSs of different proteins carry out the same functions. Therefore MTS function must rely on attributes other than primary sequence⁴. The lengths of MTSs vary but are usually in the range of 20 - 40 amino acid residues long. The typical MTS, forms an amphiphilic α -helical conformation with positively charged amino acid residues on one side of the helix and hydrophobic amino acid residues on the other side⁴. It seems that amphiphilicity is essential for mitochondrial targeting but helical structure is not⁵. However increasing the helicity of an MTS has been shown to improve its efficiency⁶. It has been shown that, in at least one case, the helical structure of an MTS may not be required for targeting the precursor protein to the mitochondria at all, but may instead be important for maintaining the peptide in an import competent conformation⁷. It follows that, success at maintaining an unfolded conformation would improve efficiency of mitochondrial targeting indirectly, as only unfolded, or loosely folded, monomeric polypeptides are capable of import into mitochondria⁸.

MTSs participate in many stages of mitochondrial import¹(figure 1-page 22). MTSs can bind the cytosolic mitochondrial stimulation factor (MSF) which, along with cytosolic chaperones like heat shock protein (Hsp) 70⁹, keep the mitochondrial protein in an unfolded conformation by binding tightly to individual precursors. MSF specifically recognises the MTS and directly targets precursors to receptors on the mitochondrial surface. MSF also actively disperses aggregates of mitochondrial precursors^{10;11}. At the cytosolic face of the mitochondrial outer membrane, the MSF-MTS complex is bound by the Tom37-Tom70 receptor complex and with ATP hydrolysis the MSF disassociates and the MTS is transferred to Tom20. Binding to Tom20 is dependent on the MTS¹. MSF is not needed for the delivery of all mitochondrial proteins to the

mitochondrial outer membrane. Most MTS containing proteins bind Tom20/22 directly.

MTSs are thought to mediate transport across the outer membrane by sequential binding to acidic receptor sites. This is called the acid chain hypothesis¹². The mitochondrial targeting signal also contributes to intra-mitochondrial sorting, precursors destined for the outer membrane stop at the Tom20 receptor whilst precursors destined for the mitochondrial inner membrane and matrix, will be passed from the Tom20 receptor to the internal Tim23 receptor¹². The interaction between Tim23 and the MTS is the initial event of the protein translocation across the inner membrane¹, Tim17 and Tim23 are integral membrane proteins and are believed to form the protein translocation core of the Tim complex, whereas a peripheral membrane protein, Tim44, is loosely associated with Tim23 and also binds mitochondrial hsp70 in the matrix. The membrane bound mitochondrial hsp70 functions as an ATP-driven import motor that pulls the preproteins into the matrix^{13;14}.

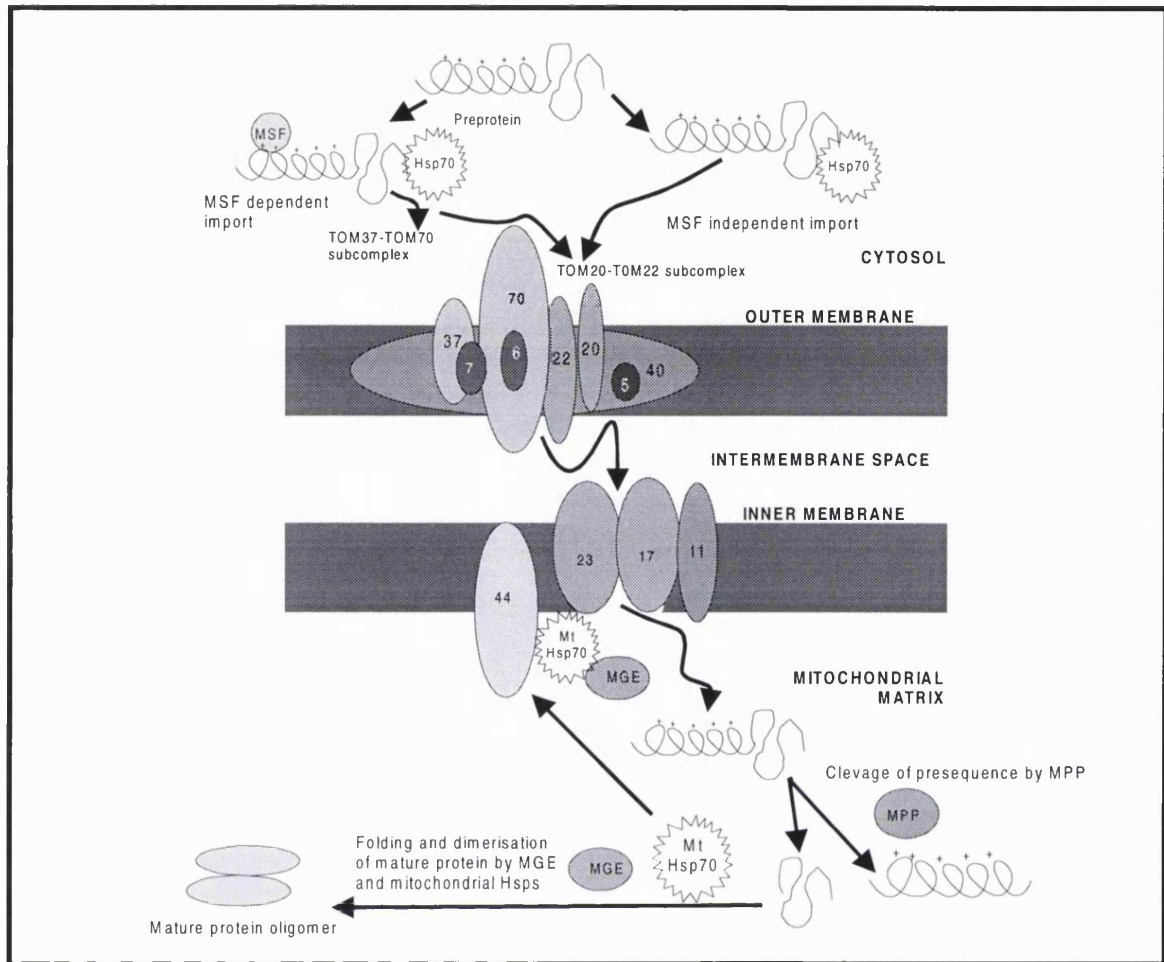


Figure 1 - The mitochondrial import machinery

The mitochondrial import machinery is made up of the TOM complex and the TIM complex. Proteins with a MTS are bound by cytosolic chaperones. Most MTS containing preproteins are recruited to the TOM20-TOM22 complex directly. Other preproteins bind MSF and are recruited to the TOM37-TOM70 complex and then bind to the TOM20-TOM22 complex. From TOM20-TOM22 the proteins are passed to the TIM complex. After import through the TIM complex the preprotein is in the mitochondrial matrix. In the matrix the presequence is cleaved by MPP and the mature peptide folding is aided by mitochondrial chaperones and co-chaperones, including mitochondrial Hsps and the ATP/ADP exchange factor, MGE. Adapted from Ryan and Jensen (1995)¹⁵.

As soon as proteins are imported into the matrix, the matrix processing peptidase (MPP) catalyses the removal of their presequences. MPP seems to be specific to mitochondrial proteins; it does not show non-specific protease activity¹. The majority of presequences are cleaved in one step. The requirements for recognition of the presequence as substrate for MPP are arginine at the -2 or -3 position relative to the mature amino terminus, and another basic (arginine or lysine) residue upstream of the -2/-3 arginine. There is also a preference for hydrophobic or especially aromatic residues at the +1 position relative to the mature amino terminus¹⁶. Some higher order structure seems to be important for cleavage, specifically a helix-linker-helix structure, in which glycine and /or proline residues serve as a α -helix breaking linker¹⁷.

A subset of proteins have leader sequences that are removed in two sequential steps, the first step being catalysed by MPP and the second by intermediate specific matrix protease. This subset share an amino acid motif of arginine at position -10, a hydrophobic residue at position -8, and serine, threonine or glycine at position -5 relative to the mature amino terminus¹⁸. The mature polypeptides then fold and oligomerise, if necessary, with the help of mitochondrial chaperones and co-chaperones, such as mitochondrial Hsp70 and MGE.

1.3 Peroxisomal targeting sequences

There are at least two types of peroxisomal import pathways for peroxisomal matrix proteins; these correspond to two targeting sequences¹⁹⁻²². Peroxisomal targeting sequence type 1 (PTS1) is a C-terminal tripeptide which has a consensus sequence (S/A/C)-(K/R/H)-(L/M)^{23,24}. Proteins with a PTS1 bind to the PTS1 receptor Pex5^{25,26} and are imported across the peroxisomal membrane by the peroxisomal import

machinery (figure 2). Peroxisomal targeting sequence type 2 (PTS2) is an N-terminal, or almost N-terminal, sequence [(X)_n R L (X)₅ H/Q L]^{27;28}. PTS2s interact with the PTS2 receptor Pex7 and are imported across the peroxisomal membrane in a pathway that converges with the PTS1 pathway via Pex14²⁹ (figure 2). Pex5 is involved in PTS2 import as well³⁰ and the different import pathways are mediated by different domains Pex5³¹. Once in the peroxisomal matrix PTS2 peptides are cleaved by an unknown factor²⁷.

Pex5 includes tetratricopeptide repeat (TPR) domains³² which have been shown to be essential for PTS1 binding³³. TPRs are degenerate 34-amino acid repeated motifs that are widespread in evolution³⁴. TPRs have a high probability of forming amphipathic α -helices and seem to mediate protein-protein interactions³⁵. As well as being essential for Pex5-PTS1 binding, TPR domains are involved in TOM20-MTS binding³⁶. A model for the binding of Pex5p and the canonical PTS1 tripeptide SKL has been proposed in which the tripeptide appears to sit in the groove formed by two adjacent TPRs³⁷.

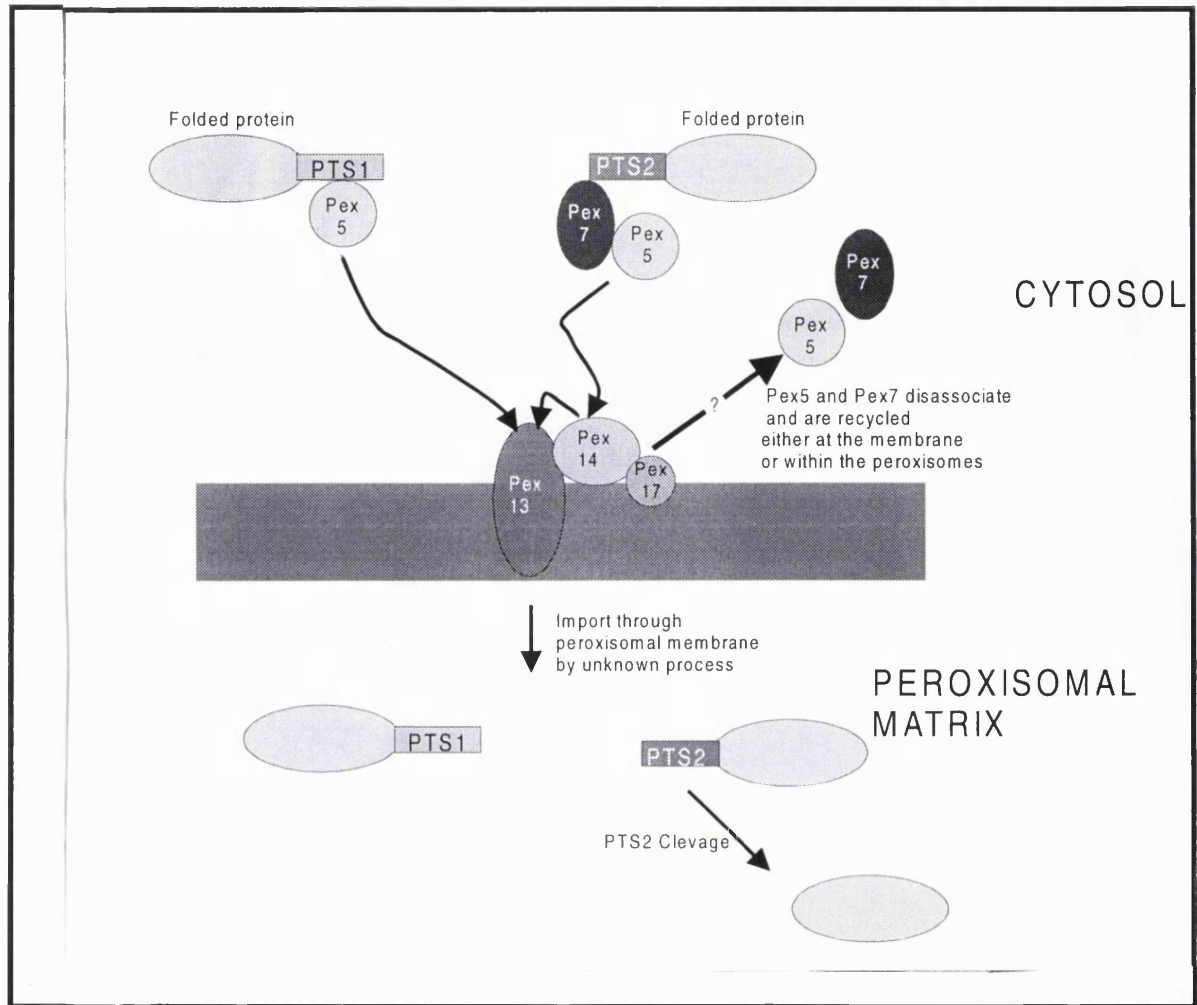


Figure 2 - The peroxisomal import machinery

The peroxisomal import channel is made up of Peroxin proteins. Proteins containing a PTS1 are bound in the cytosol by Pex5 which is then recruited by Pex13; whilst proteins containing a PTS2 are bound by Pex7. PTS2 import also involves Pex5. Pex7 takes the PTS2 containing protein to Pex14, it is passed to Pex13 and the PTS1 and PTS2 pathways converge. The method of translocation across the peroxisomal membrane is unknown. Once in the peroxisomal matrix, the PTS2 signal is cleaved by an unknown factor

In contrast to mitochondrial import, proteins can be imported into the peroxisomes fully folded³⁸. It has been shown that in the case of thiolase and chloramphenicol acetyl transferase (CAT). Subunits lacking PTS1s can be “piggybacked” into peroxisomes by oligomerising with subunits possessing PTS1s³⁸⁻⁴⁰. It is unknown how the passage through the peroxisomal membrane is achieved. There seems to be a requirement for ATP but not for a membrane potential⁴¹.

1.4 The dual localisation of AGT protein

The intracellular location of individual proteins tends to be highly conserved from early on in eukaryotic evolutionary history. As a result of natural selection for optimised function, the vast majority of enzymes, have a single and unvarying location irrespective of cell type or species, for examples see Danpure (1995)⁴².

Alanine glyoxylate aminotransferase (AGT) is a conspicuous exception. The liver-specific, intermediary metabolic enzyme AGT is found in a variety of different subcellular locations depending on the vertebrate species investigated⁴³ (figure 3 - page 28). There are mammalian species in which hepatic AGT is peroxisomal (e.g. human, macaque, gorilla, orang utan, saki monkey, koala, wallaby and rabbit) and in which it is mainly mitochondrial (e.g. domestic cat, dog, ferret, shrew, mole, and hedgehog). AGT can also be located in both organelles in the livers of some mammals (e.g. common marmoset, golden lion tamarin, hamster, rat and mouse). In a few cases, significant amounts of AGT have been found in the cytosol, for example in the guinea pig, in which AGT is peroxisomal and cytosolic and in the common frog in which AGT is mitochondrial and cytosolic⁴³. The distribution of AGT protein has changed on

numerous occasions during the evolution of mammals, both between and within orders (figure 3).

1.5 Molecular basis of dual AGT targeting

The archetypal mammalian AGT gene (figure 4 - page 30) has the potential to encode an N-terminal MTS and a C-terminal PTS1⁴⁴. 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 latter⁴⁴. The MTS is included or excluded from the open reading frame (ORF) due to variable use of two transcription start sites (TS-A and TS-B), and two in-frame translation start sites (TL-1 and TL-2). In what is assumed to be the more primitive condition, all four sites are present⁴⁵. In this condition, presently found in extant mammalian species such as the rat⁴⁶ and marmoset⁴⁷, two proteins of different sizes are synthesised. The longer form which contains the MTS is targeted to the mitochondria whilst the shorter form is targeted to the peroxisomes⁴⁷. However if one or more of these sites are absent the subcellular distribution of AGT protein is altered (figure 5 and table 1). For example, the MTS has been lost by permanent exclusion from the ORF on at least four temporally separated occasions in human⁴⁸ (figure 5 and table 1), rabbit⁴⁷ and guinea pig⁴⁹ ancestral lines due to mutations in TL-1. This has resulted in a peroxisomal distribution of AGT in these species⁴⁵. Transcription start sites B (TS-B) has been lost in the cat ancestral line leading to a mainly mitochondrial distribution of AGT in cat liver (figure 5 and table 1). Therefore, theoretically, the subcellular distribution of AGT

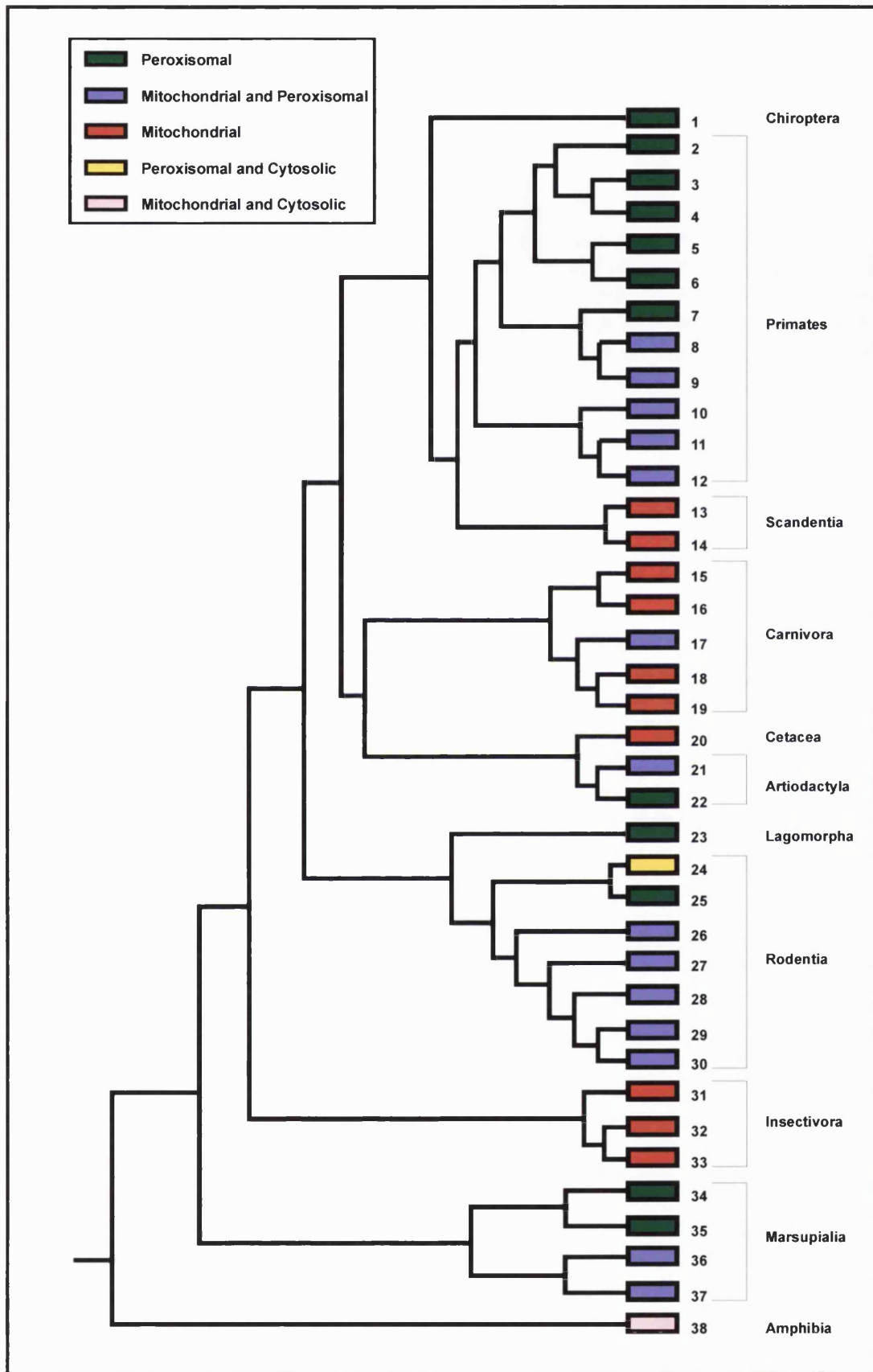


Figure 3 - Subcellular distribution of hepatic AGT in species arranged on a phylogeny

Figure 3 - Subcellular distribution of hepatic AGT in species arranged on a phylogeny

Boxes indicate species, 1 = seba's fruit bat, 2 = orang-utan, 3 = gorilla, 4 = human, 5 = baboon, 6 = macaque, 7 = saki monkey, 8 = marmoset, 9 = tamarin, 10 = lemur, 11 = flemur, 12 = loris, 13 = treeshrew, 14 = dwarf treeshrew, 15 = ocelot, 16 = cat, 17 = bear, 18 = ferret, 19 = dog, 20 = porpoise, 21 = pig, 22 = sheep, 23 = rabbit, 24 = guinea pig, 25 = agouti, 26 = porcupine, 27 = squirrel, 28 = hamster, 29 = rat, 30 = mouse, 31 = mole, 32 = hedgehog, 33 = shrew, 34 = wallaby, 35 = koala, 36 = opossum, 37 = dunnart, 38 = frog. Data from Danpure and co-workers (1994)⁴³ and unpublished results, Danpure *et al.* Phylogeny is adapted from mitochondrial DNA studies⁵⁰⁻⁵⁸. Boxes are coloured due to the subcellular distribution of AGT in the livers of the species, as key. For taxonomic species names see table 2 - page 41.

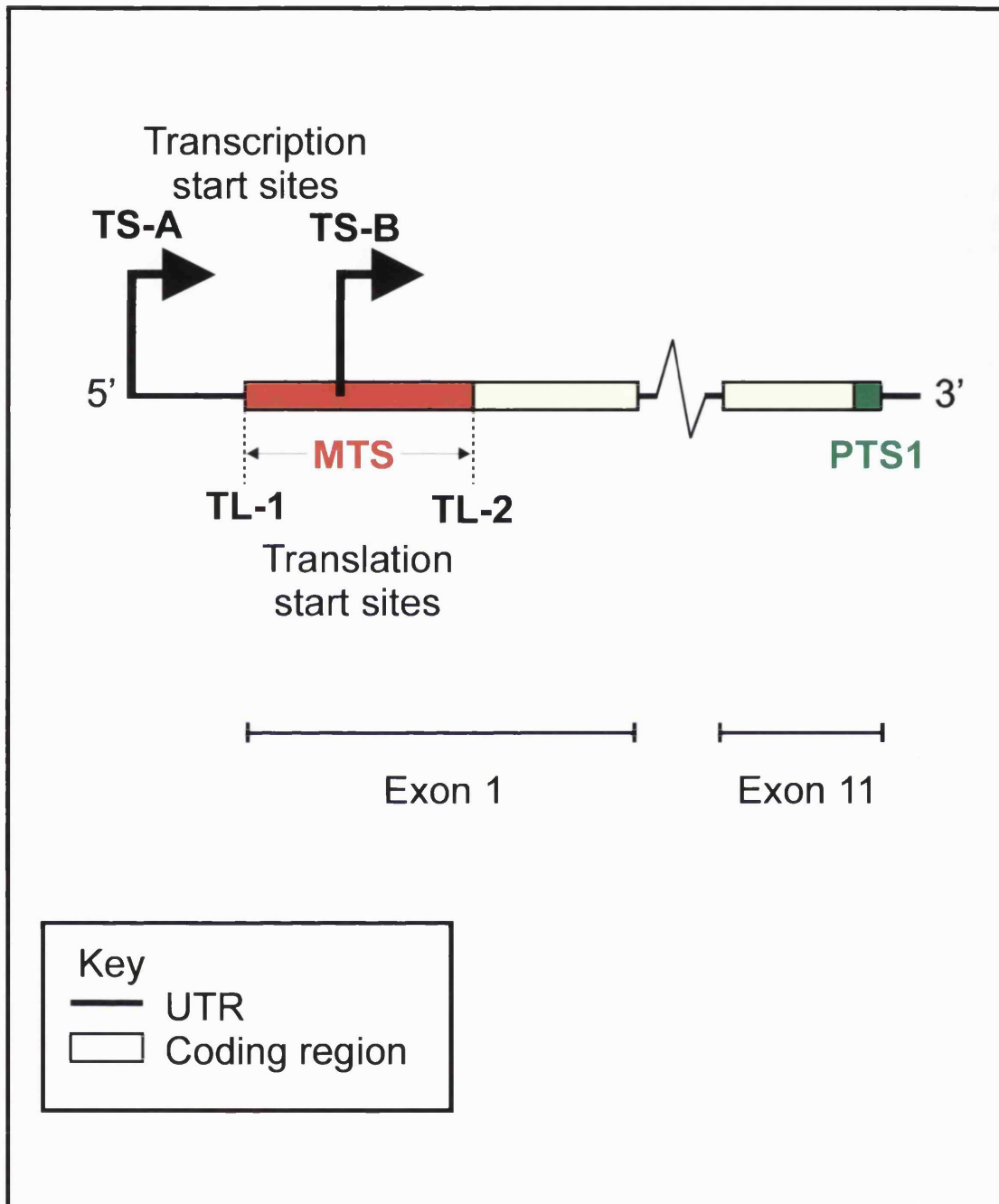
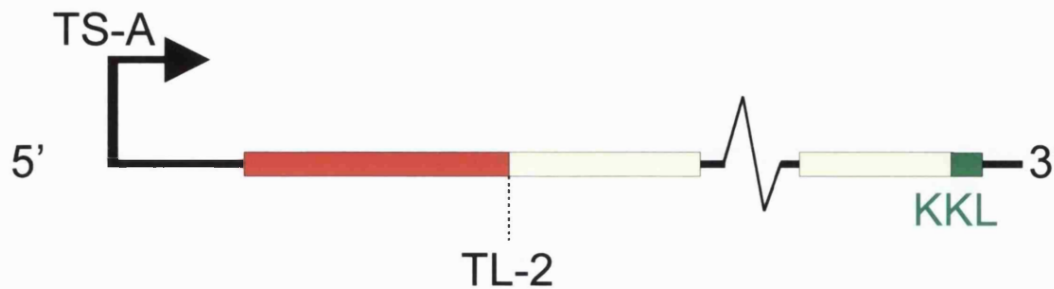


Figure 4 - Archetypal mammalian AGT gene

All mammalian AGT genes that have been cloned encode a PTS1 and some encode a MTS. Some include two transcription and two translation start sites, some include only a subset of the four archetypal sites. The presence or absence of these sites determines the subcellular distribution of AGT protein in the liver of the species.

Human AGT gene encoding a peroxisomal product



Cat AGT gene encoding a mainly mitochondrial product

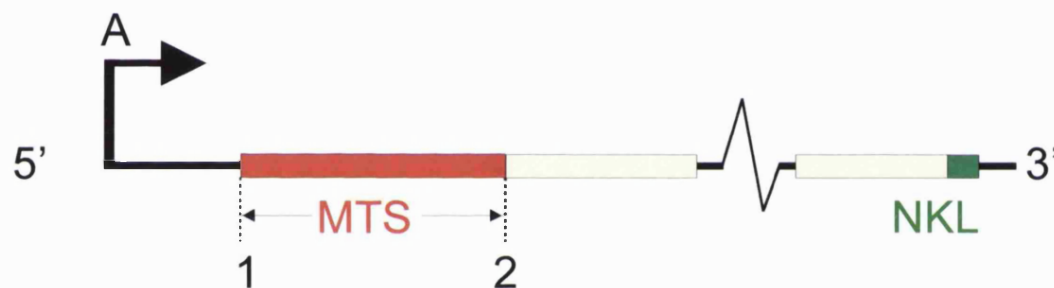


Figure 5 - Schematic diagrams of human and cat AGT genes

Human AGT is translated from TL-2, as TL-1 is absent⁴⁸. Therefore the MTS is excluded from the ORF and the protein includes only a PTS1 and so is targeted to the peroxisomes. TS-B may also be absent in the human AGT gene but this would have no functional consequences due to the loss of TL-1. Cat AGT is transcribed from TS-A as TS-B is absent⁵⁹. Therefore, the majority of translation initiates from TL-1. AGT protein translated from TL-1 includes a MTS and is mitochondrial. Some leaky ribosome scanning causes about 5% of translation to initiate at site TL-2 and this minor transcript encodes a peroxisomal protein. Therefore Cat AGT is mainly mitochondrial with 5 % peroxisomal⁵⁹.

Table 1 - Previously characterised AGT genes and the subcellular distributions of their gene products

| Species | TS-A | TL-1 | TS-B | TL-2 | Subcellular Distribution |
|------------|------|------|------|------|--------------------------|
| Human | + | - | - | + | P |
| Marmoset | + | + | + | + | M+P |
| Cat | + | + | - | + | M |
| Rabbit | + | - | (+) | + | P |
| Rat | + | + | + | + | M+P |
| Guinea pig | - | - | + | + | P+C |

The table includes all the mammalian AGT genes that have been analysed to date. The presence and absence of functional transcription and translation start sites result in AGT gene products being differentially located: P = peroxisomal, M= mainly mitochondrial and C = cytosolic. The human AGT gene has lost both TL-1 and TS-B, but loss of TS-B is thought to be a secondary loss due to the lack of functional constraint on the site since the loss of TL-1⁴⁸. The common marmoset⁴⁷, the rat⁴⁶ have all sites. The house mouse (Liu et al unpublished, accession number: AAB82001), has both translation start sites but nothing is known about the existence of transcription start sites. As the mouse has mitochondrial and peroxisomal hepatic AGT protein⁴³, it is likely that both transcription start sites exist. The rabbit⁴⁷ and the guinea pig⁴⁹ have lost TL-1, however due to inefficient peroxisomal targeting of guinea pig AGT protein, some is cytosolic. It is unknown whether TS-B exists in the rabbit but its absence would be predicted to have no functional consequences due to the absence of TL-1. As the cat AGT gene has only TS-A, translation usually starts from TL-1 but leaky ribosome scanning results in approximately 5 % of the protein being targeted to the peroxisomes⁵⁹.

protein can be dramatically changed by loss of a translation start site, which can be brought about by just one nucleotide mutation.

1.5.1 The context specificity of peroxisomal targeting of human AGT protein

Human AGT includes a PTS1, KKL, which is necessary for its targeting to peroxisomes⁶⁰. However KKL is not sufficient for peroxisomal targeting of reporter proteins, such as firefly luciferase (FFL)⁶⁰ or CAT⁶¹, to the peroxisomes. Therefore, the human AGT PTS1 is protein context dependent. These observations have led to the hypothesis that there is additional targeting information within human AGT that works with the C-terminal PTS1 to enable peroxisomal targeting of human AGT protein⁴⁵ (figure 6).

1.6 Metabolic function of AGT protein

The dual organellar distribution of AGT is thought to reflect its dual metabolic roles of gluconeogenesis and glyoxylate detoxification (figure 7). AGT, with its cofactor pyridoxal phosphate, catalyses at least two reactions of metabolic significance:





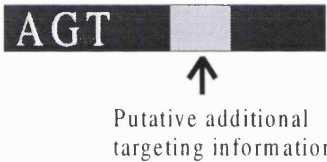






| Protein | Subcellular Distribution |
|---|--------------------------|
|  AGT  SKL | Peroxisomal |
|  AGT  KKL | Peroxisomal |
|  AGT  ↑ Putative additional targeting information | Cytosolic |
|  Reporter Protein  SKL | Peroxisomal |
|  Reporter Protein  KKL | Cytosolic |
|  Reporter Protein | Cytosolic |

Figure 6 - The context specificity of peroxisomal AGT targeting

In contrast to the canonical PTS1 (SKL), KKL is necessary for targeting of human AGT to the peroxisomes⁶⁰ but not sufficient for peroxisomal targeting of a reporter^{60;61}.

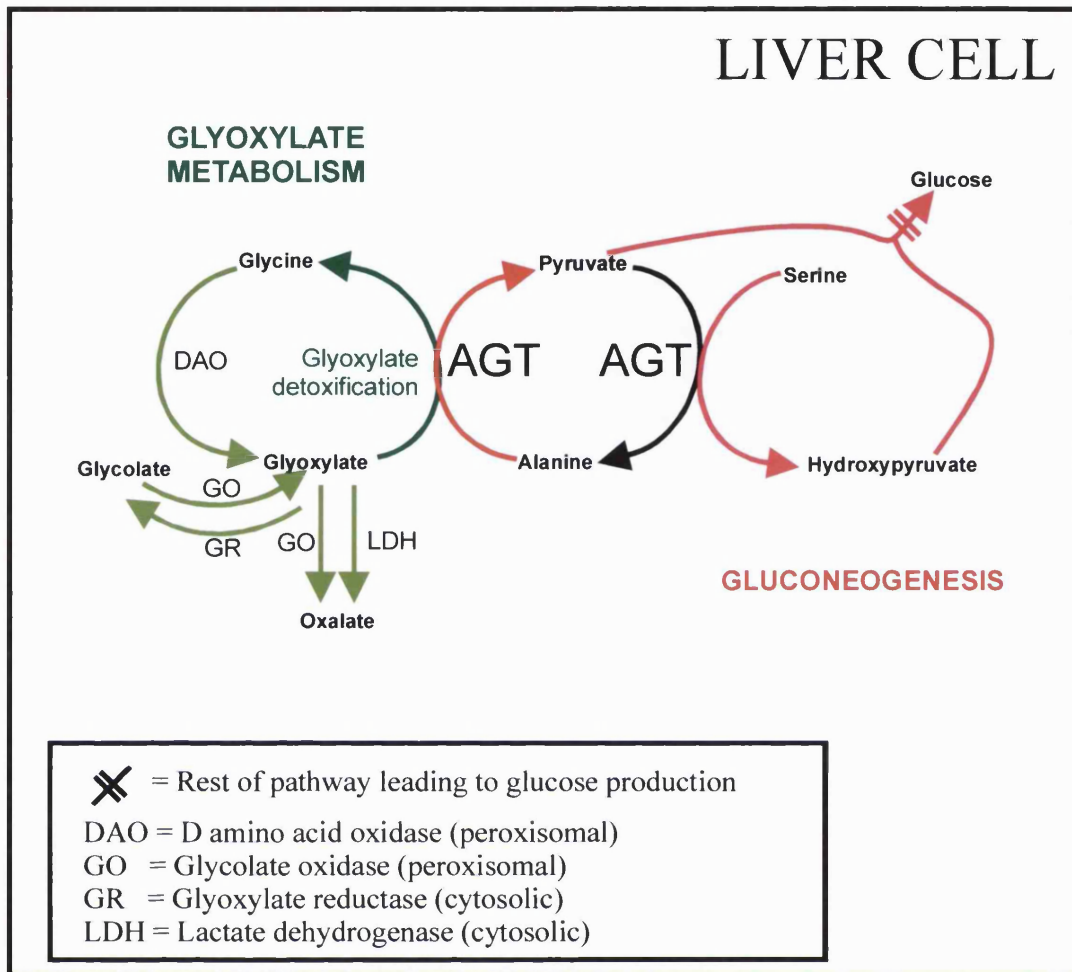
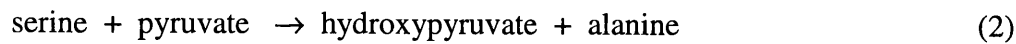
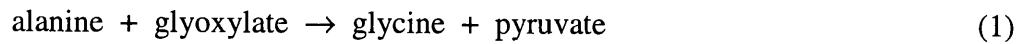
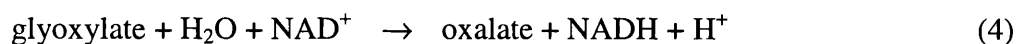
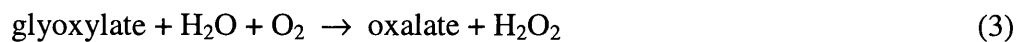


Figure 7 - Metabolic roles of AGT

AGT catalyses the conversion of alanine and glyoxylate to pyruvate and glycine. Failure to convert glyoxylate to glycine allows the glyoxylate to be oxidised to oxalate. AGT also catalyses the conversion of pyruvate and serine to alanine and hydroxypyruvate. This contributes to the process of gluconeogenesis.



Although both reactions 1 and 2 are predicted to be important in gluconeogenesis, reaction 1 can also be considered as glyoxylate detoxification, because a failure to transaminate glyoxylate to glycine allows it to be oxidised to oxalate instead. This is catalysed by the peroxisomal enzyme glycolate oxidase (reaction 3) or the cytosolic enzyme lactate dehydrogenase (reaction 4):



Reactions 3 and 4 are potentially life threatening because the calcium salt of oxalate is highly insoluble and can crystallise out in regions of high concentration, such as the kidney and urinary tract, to form stones⁶². This happens in a rare autosomal recessive disease called primary hyperoxaluria type I (PH1)⁶². PH1 is caused by a functional deficiency of AGT and is characterised by increased urinary excretion of oxalate, and usually, glycolate⁶². Calcium oxalate stones are formed and deposited in the urinary tract (urolithiasis) and renal parenchyma (nephrocalcinosis) and, in severe cases calcium oxalate crystals can deposit throughout the body (systemic oxalosis). Untreated PH1 patients frequently die of chronic renal failure in late childhood and early adulthood⁶².

It is predicted that reaction 1 (the glyoxylate detoxification reaction) would be most efficiently placed in the peroxisomes, which is the major site of glyoxylate synthesis. Proof is amply shown by the observation that one third of PH1 patients have the disease, not because they have no AGT, but because it is mistargeted to the

mitochondria⁶³, instead of its correct localisation in the peroxisomes⁶⁴. This is caused by the combined presence of a PH1-specific Gly₁₇₀Arg mutation and a normally occurring Pro₁₁Leu polymorphism⁶⁶. The Pro₁₁Leu polymorphism substitutes a helix-breaking residue for a helix-forming residue, and seems to allow the N-terminal section of the AGT protein to fold into an α -helix. With the Gly₁₇₀Arg mutation, which only occurs on the background of the Pro₁₁Leu polymorphism, the α -helix is able to act as an MTS^{65;66}. The combination of both substitutions, seems to be required for the cryptic MTS to be fully activated. The Pro₁₁Leu polymorphism present in AGT, causes targeting of 95% of AGT to the peroxisome and only 5% to the mitochondria in normal human hepatocytes, whilst, a combination of Pro₁₁Leu and Gly₁₇₀Arg causes about 90% of AGT to be targeted to the mitochondria in the hepatocytes of PH1 patients homozygous for both⁶⁶. Interestingly, the Pro₁₁Leu polymorphism is present in normal North American and European populations with an allelic frequency of ~20%⁶⁶, why it is maintained at such high population levels is puzzling as its co-occurrence with Gly₁₇₀Arg (present with an allelic frequency of ~0.05%) causes such a devastating disease.

As for reaction 2 (a gluconeogenic reaction) it has been suggested that it would be likely to be placed in the mitochondria or cytosol⁴⁵. This is supported by evidence from a restricted group of rodents (rats, mice and hamsters). which normally have AGT distributed fairly evenly between peroxisomes and mitochondria⁴³. When fed on high protein diets or given gluconeogenic stimuli, such as glucagon, there is a large increase in the synthesis of mitochondrial, but not peroxisomal AGT⁶⁷⁻⁶⁹.

1.7 An adaptive explanation for variable AGT targeting

Despite the great diversity in the pattern of AGT distribution in mammals, there is a marked relationship between the diet of a species and the distribution of AGT in its liver. Herbivorous species tend to have peroxisomal AGT, carnivorous species tend to have mitochondrial AGT and omnivorous species tend to have both mitochondrial and peroxisomal AGT. This correlation is thought to reflect the dual metabolic roles of AGT.

Glyoxylate detoxification would be expected to be most important when the diet is high in oxalate or oxalate precursors. Most of the oxalate in the living world is found in plants, fungi, and lichens; very little is found in animals⁷⁰. Therefore, herbivores are likely to eat high levels of oxalate and its precursors compared to carnivores.

Consequently the most important, metabolic role of AGT in herbivores is predicted to be glyoxylate detoxification. As mentioned above, the glyoxylate detoxification role of AGT is thought to take place in the peroxisomes, the major site of glyoxylate synthesis.

Therefore, in herbivores the optimal subcellular distribution of hepatic AGT is peroxisomal.

Unlike most plant material, animal tissue has a high protein but low carbohydrate content. Therefore, in carnivores gluconeogenesis is especially important. As mentioned before, the site of the gluconeogenic role of AGT is thought to be the mitochondria. ***Therefore, in carnivores the optimal subcellular distribution of hepatic AGT is mitochondrial.***

1.8 Aims of this study

The general aim of the work described in this thesis has been to further elucidate the mechanisms and evolution of intracellular targeting of AGT. This has been pursued, in the following ways:

1. The nature and evolution of AGT targeting in anthropoid primates was examined by cloning of the 5' regions of a number of primate AGT genes, and the resulting sequences were analysed for evidence of selection pressure (Chapter 3).
2. The molecular basis for the mitochondrial and cytosolic distribution of AGT in amphibian liver was determined by the cloning, expression and analysis of the xenopus AGT gene (Chapter 4).
3. The pressures and constraints on the evolution of the targeting sequences of AGT were further investigated by identification of non-mammalian AGT homologues and comparison to the targeting sequences of other proteins (Chapter 5)
4. Candidates for the pressure driving the evolution of differential targeting of AGT in mammals were subjected to formal statistical analyses (Chapter 6).

2. Experimental and Analytical Methods

2.1 Materials

2.1.1 Animal tissues

Primate tissues were obtained by kind gift of Andrew Cunningham (Institute of Zoology, London) or from Zoobank at Institute of Zoology, London (see table 2).

Xenopus tissue was provided by the UCL animal house

2.1.2 *Escherichia coli* genotypes

JM109 strain (used for all cloning steps unless otherwise stated) λ - $\Delta(lac-proAB)$

$[F' traD36, proAB, lacI^q Z\Delta M15]$ (Promega)

SURE 2 strain: $e14^-(McrA^-)\Delta(mcrCB-hsdSMR-mrr)171, endA1, supE44, thi-1, gyrA96, relA1, lac, recB, recJ, sbcC, umuC::Tn5 (Kan^r), uvrC, [F' proAB lacI^q Z\Delta M15 Tn10, (Tet^r), Amy, Cam^r]^a$ (Stratagene)

SOLR strain: $e14^-(McrA^-)\Delta(mcrCB-hsdSMR-mrr)171, sbcC, recB, recJ, uvrC, umuC::Tn5, (Kan^r), lac, gyrA96, relA1, thi-1, endA1, \lambda^R [F' proAB, lacI^q Z\Delta M15] Su^-$ (non-suppressing) (Stratagene)

XL1-Blue MRF' strain: $\Delta(mcrA)183\Delta(mcrCB-hsdSMR-mrr)173, endA1, supE44, thi-1, recA1, gyrA96, relA1, lac, [F' proAB, lacI^q Z\Delta M15, Tn10, (Tet^r)]$ (Stratagene)

Table 2 - Species discussed in this study

| Common Name (used in text) | Latin Name | Code |
|----------------------------|---------------------------------|------|
| Arabidopsis | <i>Arabidopsis thaliana</i> | At |
| Baboon | <i>Papio anubis</i> | Pa |
| Brown lemur | <i>Lemur fulvis</i> | Lf |
| Bullfrog | <i>Rana catesbeiana</i> | Rc |
| Celebes macaque | <i>Macaca nigra</i> | Mn |
| Chimpanzee | <i>Pan troglodytes</i> | Pt |
| Common frog | <i>Rana temporis</i> | Rt |
| Common squirrel monkey | <i>Saimiri sciureus</i> | Ss |
| Cotton-top tamarin | <i>Sanguinus oedipus</i> | So |
| Diana monkey | <i>Cercopithecus diana</i> | Cd |
| Domestic cat | <i>Felix catus</i> | Fc |
| Drosophila | <i>Drosophila melanogaster</i> | Dm |
| Fat tailed dwarf lemur | <i>Cheirogaleus medius</i> | Cm |
| Lily | <i>Fritillaria agrestis</i> | Fa |
| Gibbon | <i>Hylobates lar</i> | Hl |
| Goeldis monkey | <i>Callimico goeldii</i> | Cg |
| Golden lion tamarin | <i>Leontopithecus rosalia</i> | Lt |
| Gorilla | <i>Gorilla gorilla</i> | Gg |
| Guinea pig | <i>Cavia porcellus</i> | Cp |
| House mouse | <i>Mus musculus</i> | Mm |
| Human | <i>Homo sapiens</i> | Hs |
| Japanese macaque | <i>Macaca fuscata</i> | Mf |
| Lesser slow loris | <i>Nycticebus pygmaeus</i> | Np |
| Marmoset | <i>Callithrix jacchus</i> | Cj |
| Orang-utan | <i>Pongo pygmaeus</i> | Ppy |
| Pobblebonk | <i>Limnodynastes dumerillii</i> | Ld |
| Rabbit | <i>Oryctolagus cuniculus</i> | Oc |
| Rat | <i>Rattus norvegicus</i> | Rn |
| Spider monkey | <i>Ateles paniscus</i> | Ap |
| Silvery marmoset | <i>Callithrix argentata</i> | Ca |
| Slender loris | <i>Loris tardigradus</i> | Lt |
| Tobacco | <i>Nicotiana tabacum</i> | Nt |
| White-faced saki monkey | <i>Pithecia pithecia</i> | Pp |
| Xenopus | <i>Xenopus laevis</i> | Xl |

All species are given 2 letter codes, in the cases where they are duplicated, 3 letters are assigned eg. *Pithecia pithecia* (Pp) and *Pongo pygmaeus* (Ppy)). Organisms are referred to by common names, in the text.

2.2 Oligonucleotide primers

The primers used in this study are shown in table 3.

2.2.1 Primate primers

Primate PCR primers were designed to map to regions of high sequence identity. The 5' PCR primer (P1) anneals to the coding strand 53-32 bp upstream of TL-1. Two 3' PCR primers were used: P2 anneals to the non-coding strand 12-41 bp downstream of the intron A/exon 2 border; P3 anneals to the non-coding strand 32-52 bp downstream of TL-2. For positions see figure 10. Amplification of intron A was considered important to minimise the possibility of unrecognised 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.

2.2.2 Xenopus primers

The PCR, sequencing and mutagenic primers used in the xenopus study are detailed in table 3 and figure 8 (page 46).

2.3 Cloning strategies

2.3.1 Primate cloning strategy

5' regions of primate AGT genes were amplified by degenerate PCR. The PCR products were purified and cloned into pGEM-T Easy (Promega) by TA cloning.

Table 3 - Oligonucleotide primers

| Name | sequence | Position | orientation | Restriction site | Use |
|------|--|-----------------|-------------|------------------|----------------------------|
| P1 | 5'- AAGCCCATCCA CCAATCCTCN-3' | H: -97 to -117 | - | | PCR |
| P2 | 5'- CTGGAACACGT ACTGGATCCCT TCCTTGAN-3' | H: 207 to 178 | - | | PCR |
| P3 | 5'- AGGGGCTTGAG CAGGGCCTTN-3' | H: 53 to 33 | - | | PCR |
| P4 | 5'- CATCAT <u>GGTAC</u> <u>CGTCATCAGGT</u> GTAACAGCCTC- 3' | X: 16 to 36 | + | <i>KpnI</i> | PCR, sequencing |
| P5 | 5'- CATCAT <u>CTCGA</u> <u>GCGTCATCTTC</u> AGCAAACAGG- 3' | X: 1499 to 1480 | - | <i>XhoI</i> | PCR, sequencing |
| P6 | 5'- CATCAT <u>GAATT</u> <u>CGGGACTAGGT</u> GAAGGTTCAC-3' | X: 163 to 187 | + | <i>EcoRI</i> | PCR, sequencing |
| P7 | 5'- CATCAT <u>GAATT</u> <u>CGCCTCTCTGCT</u> CCTGTCAG-3 | X: 276 to 295 | + | <i>EcoRI</i> | PCR |
| P8 | 5'- CATCAT <u>GGATC</u> <u>CGATTCTGACA</u> GGAGCAGAGAG -3' | X: 299 to 279 | - | <i>BamHI</i> | PCR |
| P9 | 5'- CTTCCCACAGT CACCCTG-3' | H: 952 to 970 | + | | PCR |
| P10 | 5'- CAACCCTTAGA TGGGCTGGGAG ATCT-3' | X: 786 to 811 | + | | sequencing |
| P11 | 5'- GTTCTGGAGG GGCATTTCAG-3' | X: 946 to 927 | - | | sequencing, library PCR |

Table 3 - Continued

| | | | | | |
|-----|--|-----------------|---|--------------|--|
| P12 | 5'- GACACTGTGCC ATGGAGACTGC- 3' | X: 541 to 562 | + | | sequencing |
| P13 | 5'- GCCATCACATC CCCAATAG-3' | X: 1061 to 1043 | - | | sequencing, library PCR |
| P14 | 5'- GTGGCAAGCGA AGACATG-3' | X: 316 to 299 | - | | sequencing, library PCR, RACE |
| P15 | 5'- GAATCCGAGGA GGTACATTGGA G-3' | X: 408 to 386 | - | | sequencing, RACE |
| P16 | 5'- GTGTCCGGATC CACTCACAGC-3' | X: 545 to 525 | - | | sequencing, library PCR, RACE |
| P17 | 5'- CATCAT GGGCC <u>CTCACAGCTTG</u> CTTTGGGAC -3' | X: 1475 to 1456 | - | <i>Apal</i> | PCR, mutagenises xenopus AGT C terminus |
| P18 | 5'- CATCAT GGGCC <u>CTCACAGCTTTT</u> TTTTGGGAC -3' | X: 1475 to 1456 | - | <i>Apal</i> | PCR, mutagenises xenopus AGT C terminus |
| P19 | 5'- CATCAT GGATC <u>CTCACATCTTCT</u> TCTTGGGGCAG TG -3' | H: 1179 to 1156 | - | <i>BamHI</i> | PCR, mutagenises human AGT C terminus |
| SP6 | 5'- TATTTAGGTGA CAGTATAG-3' | plasmid primer | | | sequencing |
| T7 | 5'- TAATACGACTC ACTATAGGG-3' | plasmid primer | | | sequencing |

Restriction enzyme sites are bold and underlined. Bases differing from the original sequence and designed for mutagenesis are bold and italicised. Mapping positions are calculated from the xenopus (X:) or human (H:) AGT cDNA sequence, the numbers run as figures 11 and 26, i.e. start from TL-2 in human and from beginning of known sequence in xenopus.

2.3.2 *Xenopus* cloning strategy

A fragment of the *xenopus* AGT gene was cloned by expression library screening. This clone was initially named clone 4 but after characterisation was renamed **pXS**.

Xenopus liver cDNA was obtained by reverse transcription of total RNA (5 µg) using an oligo dT primer with SuperscriptII reverse transcriptase (GIBCO BRL), according to the manufacturer's instructions. PCR amplification of the cDNA was performed using primer pairs P4/P5 and P6/P5 to generate products **Xall** and **XL** respectively (figure 8). **XL** was digested with *EcoRI* and *XhoI* and ligated into previously digested pcDNA3, to make **pXL**.

pXL was used, as template, in a PCR reaction with a mutagenic primer at the 3' end. Primer pairs were P7/P17 for **XS-SKL**, P7/P18 for **XS-KKL** and P6/P18 for **XL-KKL**. Products were digested with *EcoRI* and *ApaI* and ligated into previously digested pcDNA3, to make clones **pXS-SKL**, **pXS-KKL** and **pXL-KKL**.

pXL was used as template in a PCR reaction using primer pair P6/P8, the PCR product was digested with *EcoRI* and *BamHI* and ligated into previously digested pEGFP (Clonetech). To make **pXmtsGFP**.

The creation of **pHsAGT-KKM** is described in figure 9. Cloned human AGT was used, as template, in a PCR reaction with primers, P9/P19. P19 is mutagenic, designed to modify the 3' end of the human gene, that encodes KKL to sequence encoding KKM and insert a *BamHI* site. The PCR product was cloned into pGEM-T Easy. This clone was then

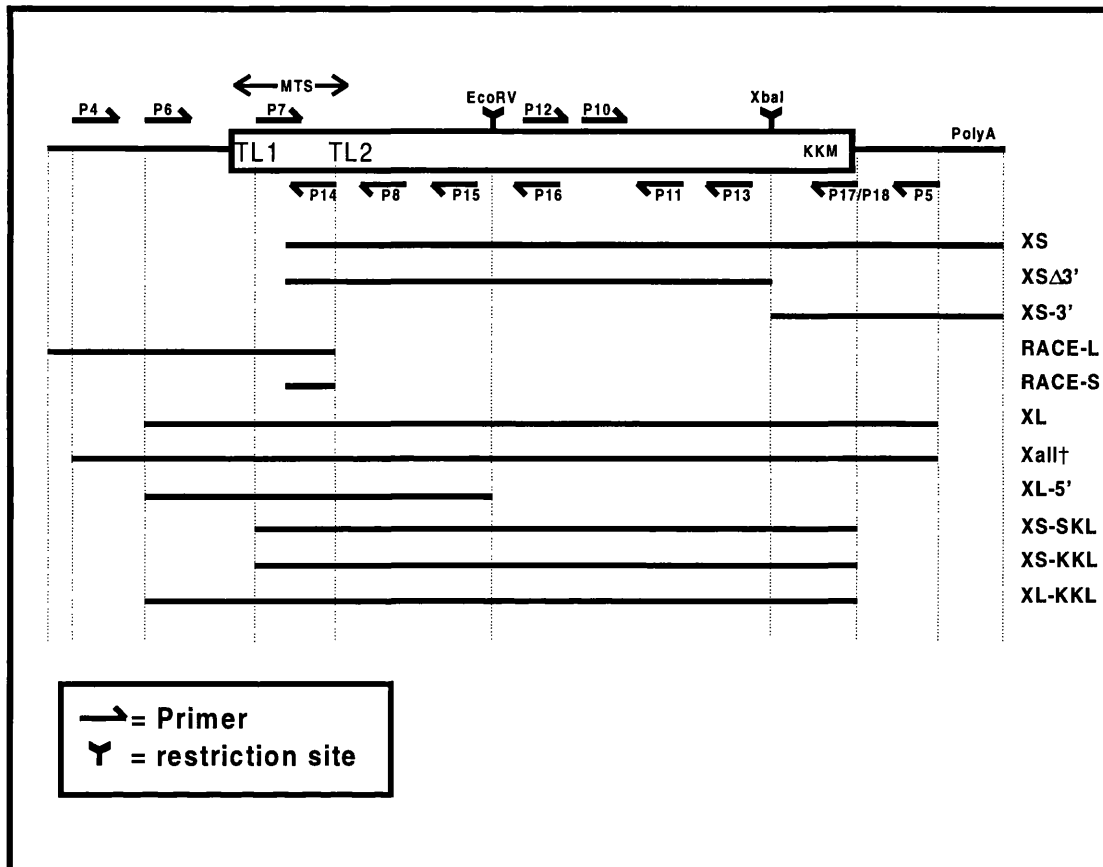


Figure 8- Schematic diagram of xenopus cloning strategy

The line represents the xenopus AGT cDNA, the box the ORF. The primers are numbered P4-P8 and P10-P18, for primer sequences and annealing positions see table 3. The clones are represented aligned to their position of the cDNA diagram and the primers used to generate them. Diagram is not to scale. † Xall has never been cloned and is a PCR product only.

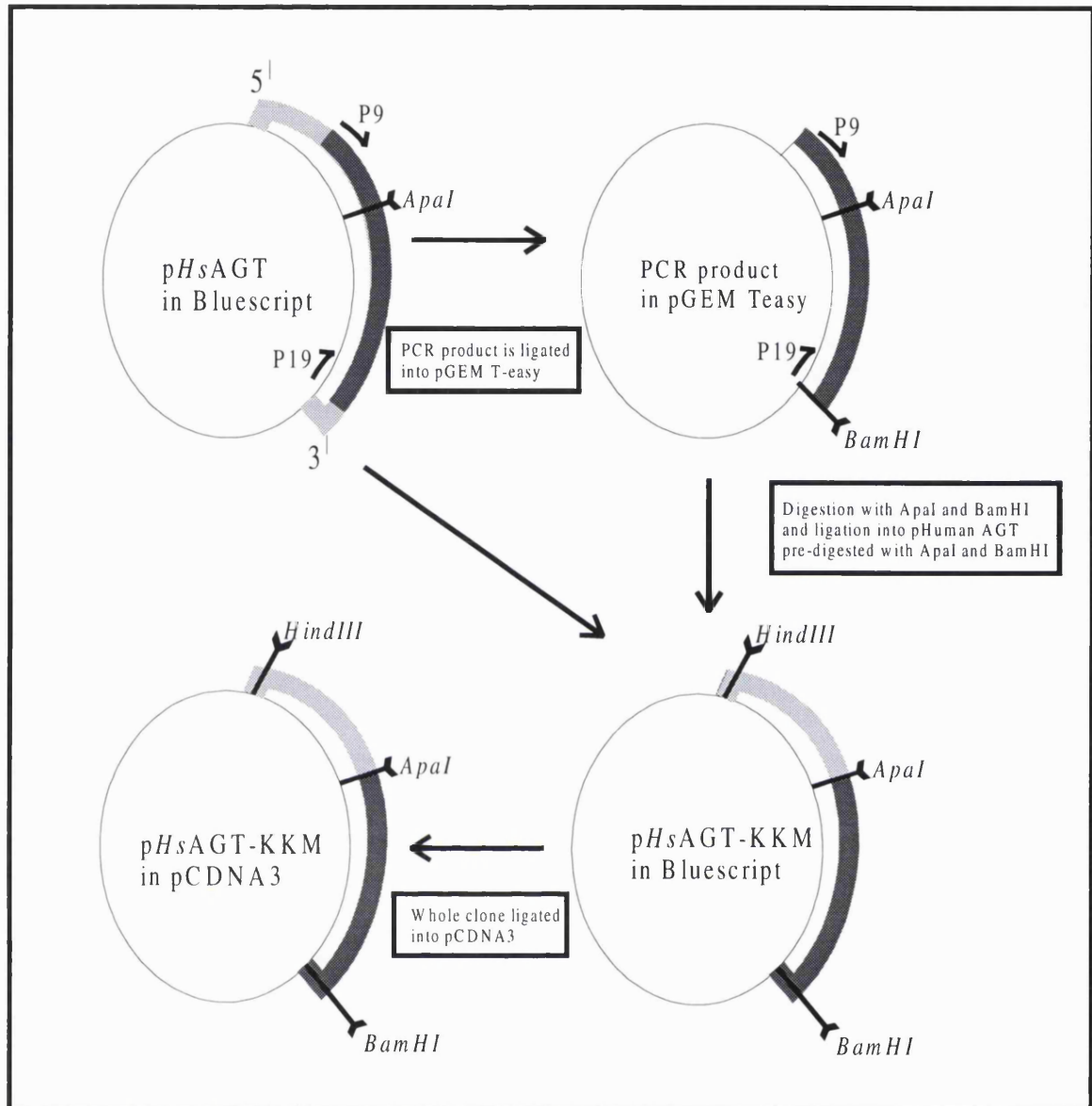


Figure 9 - Cloning strategy for pHsAGT-KKM

pHsAGT was used as template in a PCR reaction with P9/P19. p19 is mutagenic; it annealed to the region of the human gene encoding KKL and primed the product to encode KKM. The product was cloned into pGEM-T Easy and cut with *ApaI* and *BamHI*. The digested product was ligated back into pHsAGT replacing the original region. The whole insert was cut out of pBluescript with *HindIII* and *BamHI* and cloned into pCDNA3 to form clone pHsAGT-KKM.

digested with *ApaI* and *BamHI*, and ligated into pBluescript (Stratagene) containing a human AGT insert, previously digested with *ApaI* and *BamHI*. The whole insert was cut out of the resulting clone with *HindIII* and *BamHI*, the inset was then ligated into previously digested pcDNA3 to make p*Hs*AGT-KKM.

pXL was digested with *EcoRV* and *XhoI*, which cut the insert 3' of TL-2 and at the 3' end. The resulting fragments were treated with Klenow fragment to blunt end them and the larger fragment (with the plasmid sequence) was purified. It was then self ligated to form an intact plasmid with the 5' end of pXL as insert. This clone was called **pXL-5'** and used to generate antisense riboprobes for RNase protection.

2.4 Sequencing Strategies

2.4.1 Primate sequencing strategy

Primate AGT sequences of region 1 and region 2, were determined in both strands, from at least two clones, using a mixture of dideoxy and cycle sequencing. Plasmid primers, SP6 and T7, were used as sequencing primers (table 3).

2.4.2 Xenopus sequencing strategy

The Xall PCR product proved to be very prone to recombination even when grown up in an *E. coli* strain mutated in many recombination pathways (SUREII, Stratagene). Therefore Xall was sequenced directly from the PCR product to avoid recombination and to highlight polymerase errors. Cycle sequencing of Xall was performed using primers P4, P5, P6, P10, P11, P12, P13, P14, P15, and P16, to generate overlapping sequence in both strands. The XL product was also recombinogenic but was stable when cloned into pcDNA3 and grown up in SUREII cells. Cycle sequencing of XL was

performed using plasmid primers SP6 and T7 and internal primers to generate sequence in both strands. The library clone (named pXS) was also cloned into pBluescript and cycle sequenced with plasmid primers T7, T3, and clone specific primers P12, P8 and P16, cycle sequencing failed at the 3' end of the gene, probably due to polymerase "slippage" at the polyA tail. Therefore, the clone was cut with *XbaI* and the 3' fragment was subcloned into pBluescript (**pXS-3'**). The clone with the excised 3' end (**pXΔ3'**) was sequenced using T7, T3, to generate sequence in both strands. pXS-3' was sequenced with primers T3 and T7. The deduced nucleotide sequences for XL and XS were compared with that of Xall and found to be identical. All new clones based on xenopus AGT i.e. pXmtsGFP, pXS-SKL, pXL-KKL, pXS-KKL and pXL-5' were sequenced by cycle sequencing and compared to the Xall sequence, no anomalies were found. p*Hs*AGT-KKM was also sequenced to check for PCR errors, none were found.

2.5 EMBL numbers

The sequences described have been deposited in the EMBL database under accession numbers: AJ237886 (*Pt*), AJ237887 (*Gg*), AJ237888 (*Hl*), AJ237889 (*Pa*), AJ237890 (*Mn*), AJ237891 (*Cd*), AJ237892 (*Lr*), AJ237893 (*Cg*), AJ237894 (*Ss*), AJ237895 (*Pp*), AJ237896 (*Ap*) and AJ278065 (xenopus).

2.6 Experimental methods

Standard methods were used for DNA cloning⁷¹.

2.6.1 Extraction of genomic DNA from animal tissue

Primate and xenopus genomic DNA was isolated from frozen tissue (liver, blood or skin) by proteinase K digestion at 55°C for 5 hours. The digested tissue was then

subjected to phenol/chloroform extraction. The DNA was precipitated by the addition of 1/10 volume of 3M sodium acetate pH 6.8 and 2.5 volumes of ice-cold ethanol. Samples were spun at maximum speed in a microcentrifuge for 15 minutes at 4°C. Pellets were aspirated and washed in 100µl 80% ethanol and centrifuged for a further 5 minutes at 4°C. The ethanol was removed and the pellets were dried for 5 minutes under vacuum.

2.6.2 Extraction of total RNA from animal tissue

Total RNA was extracted from liquid nitrogen snap frozen xenopus liver. The liver was ground to powder in a pre-cooled pestle and mortar under liquid nitrogen.

Approximately 10 mg of frozen, powdered tissue was weighed out and the sample in kit buffer containing β-mecaptoethanol was passed through a 19 gauge needle to homogenise it. RNA was isolated using the RNeasy miniprep kit (Qiagen) and was stored at -20°C. Larger quantities of RNA (for RNase protection) were extracted using TRIZOL (GibcoBRL). 50 mg of frozen tissue powder was added to 1 ml of TRIZOL reagent and processed as per manufacturer's instructions.

2.6.3 Polymerase Chain Reaction (PCR)

PCR was performed using slight variations of this protocol:

50 µl reactions with 1.5 mM magnesium chloride and 100 pmoles of each primer.

Temperature profile: 80°C, 3 mins; (94°C, 30 secs; 55-65°C, 30 secs; 72°C, 45 secs)

x 35; 72°C, 15 minutes. Taq Polymerase (Promega) was added after the first 2 mins of this temperature profile. Annealing temperatures were varied according to the T_M of the primers and the amount of non-specific product generated in previous attempts.

Annealing and extension times were varied according to the size of the expected product.

2.6.4 Dideoxy Sequencing

5-10 µg of plasmid DNA was denatured with a solution containing 200 mM sodium hydroxide and 20 µM EDTA, then ethanol precipitated to remove the alkali. DNA sequencing was performed using Sequenase Version 2.0 T7 DNA polymerase, following the manufacturer's instructions (Amersham Life Science).

2.6.5 Automated Cycle Sequencing

Cycle sequencing was performed using the Big Dye Terminator Cycle Sequencing Ready Reaction kit (PE Biosystems) following the manufacturer's instructions. Sequencing reactions were analysed by the core facility at the Department of Biology, UCL.

2.6.6 Southern Blotting

The PCR products were electrophoresed on a 1.5% agarose gel. The gel was washed in denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 1 hour, then in neutralising solution (0.5 M Tris-HCl, 3 M NaCl, pH 7.0) for 1 hour. The DNA was transferred by overnight capillary blotting onto an uncharged nylon membrane, Biotrans (ICN). The DNA was covalently bound to the membrane by UV-crosslinking (UV Stratalinker 1800, Stratagene). Prehybridisation of the membrane was carried out for 4-5 hours at 65 °C in hybridisation buffer (6 x SSC, 10 x Denhardt's solution, 0.1 % (w/v) SDS, 200µg/ml sonicated herring sperm DNA, 10 % dextran sulphate). [1 x SSC is 0.15 M NaCl, 15mM sodium citrate, pH 7; 1 x Denhardt's solution is 0.02 % (w/v) BSA, 0.02 % (w/v) polyvinylpyrrolidone, 0.02 % Ficoll]

The radiolabelled probe was prepared by random priming in the presence of ^{32}P using the Megaprime DNA labelling system (Amersham Life Science) and heat denatured for 5 min at 100°C prior to its addition to the hybridisation buffer (50% formamide, 6 x SSC, 10 x Denhardt's solution, 0.1 % (w/v) SDS, 200 $\mu\text{g/ml}$ sonicated herring sperm DNA, 10 % dextran sulphate). Hybridisation was carried out overnight at 55°C . Following hybridisation, the membrane was washed in 0.1 % (w/v) SDS, 2 x SSC for 15 min at 65°C . This was followed by a second wash in 0.1 % (w/v) SDS, 1 x SSC for 15 min at 65°C , additional washing was performed when the background was judged to be high by evaluation using a geiger counter. The membrane was exposed overnight at 70°C to Biomax MR-1 film (Kodak).

2.6.7 PCR of xenopus library to clone 5' end of gene

PCR of the cDNA library was performed using the positive strand, plasmid primer (T7) and downstream primers on the negative strand (P14, P16, P11 and P13) (see figures 8 and 22). The reactions were incubated at 94°C for 5 minutes to destroy the virus coats before the Taq polymerase was added. The PCR products were Southern blotted using a probe generated by random priming of pXS in the presence of ^{32}P , Megaprime DNA labelling system (Amersham Life Science).

2.6.8 Rapid amplification of cDNA ends (RACE)

The RACE reactions were performed using the 5'/3' RACE kit (Boehringer Mannheim). The kit was used according to the manufacturer's instructions. First strand cDNA synthesis was carried out using xenopus liver total RNA with P16 (table 3, figure 8). The cDNA was purified using the High Pure PCR Product Purification kit (Boehringer Mannheim), according to the manufacturer's instructions. After tailing of the 5' 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 xenopus specific primer P15. A twentieth of this product was used as template in a second round of PCR amplification using a pair of nested primers (the PCR-anchor primer and xenopus specific P14) (table 3, figure 8).

2.6.9 RNase protection

A clone of the 5' end of pXL was made by digestion with *EcoRV* and religation (pXL-5') (see figure 8 and cloning strategy, section 2.3.2). Following linearisation of this clone with *HindIII*, an antisense riboprobe was generated from the T7 RNA polymerase promoter in the presence of [$\alpha^{32}\text{P}$]UTP (600 Ci/mmol). Total RNA (20 μg) prepared from xenopus liver was hybridised for 16 h at 45 °C with the antisense riboprobe (10^5 cpm). Following RNase digestion, the protected fragments were separated on a 6% denaturing polyacrylamide sequencing gel and visualised by autoradiography using Biomax MR-1 film (Kozak).

2.6.10 *In vitro* transcription and translation

Clones were linearised at the 3' end, leaving a 5' overhang. *In vitro* transcription was carried out using SP6 or T7 RNA polymerase, at 37°C for 1 hour. *In vitro* translation was carried out in rabbit reticulocyte lysate (Promega) and in the presence of ^{35}S labelled methionine at 30°C for 1 hour.

2.6.11 Immunoblotting

Samples were boiled in loading buffer [solution containing 0.125 M TRIS pH6.8, 4 % SDS, 4 % glycerol, 10 % B mecaptoethanol and a trace of bromophenol blue]and

electrophoresed on a 10% denaturing polyacrylamide gel (SDS - PAGE) prior to electroeluting onto nitrocellulose membranes using a semi-dry electrophoretic transfer cell (Trans-Blot SD, Bio-Rad) according to the manufacture's instructions. The nitrocellulose membrane was then washed in a blocking solution: PBS [8 g/l NaCl, 0.2 g/l KCl, 1.15 g/l Na₂HPO₄, 0.2 g/l KH₂PO₄, pH 7.2] containing 5 % milk proteins, for 1 hour at room temperature in order to reduce non-specific binding. The blocked membrane was incubated with rabbit anti-human AGT antibody, diluted in PBS with 0.1% TWEEN20 at concentrations varying between of 1 in 3000 to 1 in 500. *Xenopus* AGT was predicted to have lower affinity for anti-human AGT antibody than human AGT so the primary antibody concentrations were adjusted accordingly. After washing with PBS, the bound antibodies were visualised using horseradish peroxidase-conjugated goat anti-rabbit IgG (Sigma) at a 1 in 3000 dilution, and developed using enhanced chemiluminescence (ECL, Amersham Life Science), according to the manufacturer's instructions, and BioMax MR-1 film (Kodak).

2.6.12 AGT enzyme assay

The AGT assay was performed as described previously⁷², except it was not carried out under nitrogen, as the failure to exclude oxygen has been shown not to affect the outcome (M. Lumb, personal communication). The reaction mixture contained 100 mM potassium phosphate buffer pH7.4, 50 mM L-alanine, 50 mM glyoxylate, 40 µM pyridoxal phosphate and 100 µl sample in a final volume of 300 µl. The mixture was incubated for 60 mins at 37 °C. The reaction was stopped by the addition of 50 µl of 50% (w/v) trichloroacetic acid. After cooling on ice, the precipitates were removed by centrifugation at 4 °C. The pyruvate formed was assayed by adding 100 µl supernatant to 900 µl of a solution containing 0.22 mM NADH and 370 mM Tris-HCl pH 8.0. The

decrease in absorbance (reduction of NADH to NAD) at 340 nm was measured after the addition of 100 μ l lactate dehydrogenase (LDH) (180 U/ ml).

2.6.13 Expression library screening

A xenopus liver cDNA expression library, in the Uni-ZAP XR vector was purchased from Stratagene. The Uni-ZAP XR vector is based on pBluescript and allows *in vivo* excision of the pBluescript phagemid. The inserts are cloned into the phagemid via the *EcoRI* and *XhoI* sites. Using an *E. coli* host strain (X1-1 Blue MRF⁺), 300,000 plaques were screened. Nitrocellulose lifts were prepared and hybridised to the human AGT antibody. Positively hybridising regions were cut out of the plates and the λ bacteriophages were extracted from the agar and used to infect more SOLR (Stratagene) cells and used to perform a secondary screen. A tertiary screen was performed, allowing isolation of single plaques. *In vivo* excision and recircularization of cloned inserts to form a pBluescript phagemid, was performed by infecting SOLR cells simultaneously with the Uni ZAP XR λ bacteriophage and the ExAssist helper bacteriophage (Stratagene). The ExAssist bacteriophage produces proteins which recognise the origin of replication, initiation and termination sites, cloned separately into the Uni ZAP XR genome, they nick the DNA at the initiation site and then replicate the sequence between the two sites which includes the pBluescript sequence and the insert sequence. The single stranded DNA molecule produced by the ExAssist polymerase, is circularised by more ExAssist proteins and produces the pBluescript vector with insert. SOLR cells are non-suppressing and so do not allow the replication of ExAssist due to the amber mutation within ExAssist's genome, avoiding contamination of preps with ExAssist DNA.

2.6.14 Tissue culture

SV40 transformed monkey kidney cells (COS-1) were grown in DMEM (GibcoBRL) supplemented with 10 % fetal calf serum, glutamine and penicillin-streptomycin mix (10,000 µg/ml), at 37°C under 5 % CO₂. These cells were suitable for AGT transfection studies because endogenous AGT expression is undetectable and endogenous catalase acts as a peroxisomal marker. Cell lines were maintained by passaging every three days or as soon as they had reached 90 % confluence.

2.6.15 Transfection of COS cells

COS cells were transiently transfected using Superfect (Qiagen), according to the manufacturer's instructions. The optimal quantity of DNA was found to be 5 µg per 60 mm dish of 40 - 60% confluent COS cells. The optimal quantity of Superfect was found to be 25 µl.

2.6.16 Harvesting of protein from COS cells and animal tissues

Frozen animal tissues were ground under liquid nitrogen to form a tissue powder. They were then suspended in homogenisation buffer (250mM sucrose, 10mM TRIS pH7.4, 1mM EDTA and 0.1% Triton X-100). COS cells were added directly to this buffer. A Protease inhibitor cocktail was added (Roche). The samples were then sonicated with two 5 second blasts and centrifuged for 10 minutes at 4°C.

2.6.17 Immunofluorescence microscopy

Two days after transfection, immunofluorescence microscopy was carried out as described previously⁴⁴. The cells were washed in PBS and fixed in freshly prepared 3 % (w/v) paraformaldehyde for 15 minutes at room temperature, followed by permeabilisation with 1 % Triton X-100 for 15 minutes at room temperature. The cells

were then processed for immunofluorescence using rabbit anti-human AGT and fluorescein isothiocyanate-conjugated goat anti-rabbit IgG (Sigma) or guinea pig anti-human catalase followed by anti-guinea pig biotinylated antibody and avidin Texas Red (Vector Laboratories). Sometimes the cells were stained with MitoTracker (Molecular Probes) before fixation. Cells were grown in MitoTracker (1/50,000) containing media for 15 minutes and then washed and grown in media without MitoTracker for a further 15 minutes.

All incubations were performed at room temperature for 15 minutes in PBS containing 3 % BSA. After each incubation the cells were washed extensively in PBS. The cover slips were mounted onto glass slides in Mowiol (Harlow Chemical Co. Ltd) containing diazabicyclo[2,2,2]octane (DABCO, Sigma). The fluorescent staining pattern was visualised in a BioRad MRC1000 confocal laser-scanning microscope and the images were digitally recorded.

2.7 Analytical methods

Nucleotide and peptide sequences were manipulated using GCG (version 10) from the Genetics Computer Group Inc., available through the UK Human Genome Mapping Project (HGMP) resource centre server, <http://www.hgmp.mrc.ac.uk/>. Percentage identities were determined using the GAP program which is within the GCG suite.

2.7.1 Bioinformatics

BLAST searching was also carried out within HGMP. The main database used was Entrez (<http://www3.ncbi.nlm.nih.gov/Entrez/>). Multiple alignments were performed using CLUSTAL W ⁷³. Clustering relationships were determined using Pileup which

uses neighbour joining. Expected Protein molecular weights, from sequence data, were determined using PepPlot. Both Pileup and PepPlot can be found within GCG.

2.7.2 Maximum likelihood

Maximum likelihood statistical methods were used in chapters 3 and 6 of this study. Likelihood is suited to the study of evolutionary phenomena as it describes, not the probability that the events under study happened, but the probability that a given process as opposed to some other is responsible for the observed data ⁷⁴ i.e. it treats the extant sequence data as a given. Likelihood tests the differences between models of evolution, put forward to explain the data. They are tested by the likelihood ratio statistic (LR), defined as $LR = -2\log[H_0/H_1]$ where H_0 is the null hypothesis and H_1 is the alternative hypothesis. If the hypotheses are special cases of one another, then LR approximates to a χ^2 statistic with degrees of freedom equal to the difference in the number of free parameters in the two models. ⁷⁴

2.7.3 Ancestral reconstruction

Ancestral traits were reconstructed using a codon-based maximum likelihood analysis using marginal reconstruction of ancestral sequences ⁷⁵.

2.7.4 d_N/d_S ratios

The pairwise estimates of synonymous and nonsynonymous mutations per synonymous and nonsynonymous site (d_S and d_N respectively) were calculated by the maximum likelihood method of Goldman and Yang ⁷⁶. Estimation of d_N/d_S (ω) among lineages by maximum likelihood was performed by the method of Yang ⁷⁷ using the computer package PAML (version 2, freely available at

<http://abacus.gene.ucl.ac.uk/software/paml.html>). Alternative methods used included that of Li (1993)⁷⁸ and Nei and Gojobori (1986)⁷⁹.

2.7.5 Comparative analysis

Comparative analysis was performed by a maximum likelihood method^{74;80} and using the computer package for the algorithm, by kind permission of Mark Pagel, University of Reading, UK.

3. Molecular Adaptation of AGT Targeting in Primates

[The work presented in this chapter has been published, in part, in Holbrook et al (2000)⁸³]

3.1 Introduction

During mammalian evolution, the subcellular distribution of hepatic AGT protein has changed within orders and between closely related species (figure 3). One group for which this is apparent is primates, which has provided the focus for this study. The distribution of hepatic AGT in primates appears to clearly separate the platyrrhines and catarrhines. With the exception of the saki monkey the platyrrhines have the putatively more ancestral AGT distribution (i.e. mitochondrial and peroxisomal), which is also found in the prosimians⁴³. However, all the catarrhines so far studied, have an exclusively peroxisomal AGT distribution. From previous studies on the molecular basis of AGT targeting in the human⁴⁸ and common marmoset⁵⁹, it has been suggested that the loss of TL-1 in the human ancestral line (figure 5) probably occurred soon after the separation of the platyrrhine and catarrhine lineages⁴⁵.

In this study, the molecular evolution of AGT targeting in anthropoid primates was investigated further. The nucleotide sequence of the 5' part of the AGT gene, including the regions containing both of the ancestral translation initiation sites (figure 10) and the rest of exon 1, were determined for eleven primates. For the purposes of this study the region of the AGT gene encoding the ancestral mitochondrial targeting sequence, including TL-1 and running to, but not including TL-2, is defined as region 1 and the region encoding exon 1, downstream of TL-2, is defined as region 2 (figure 10). Both

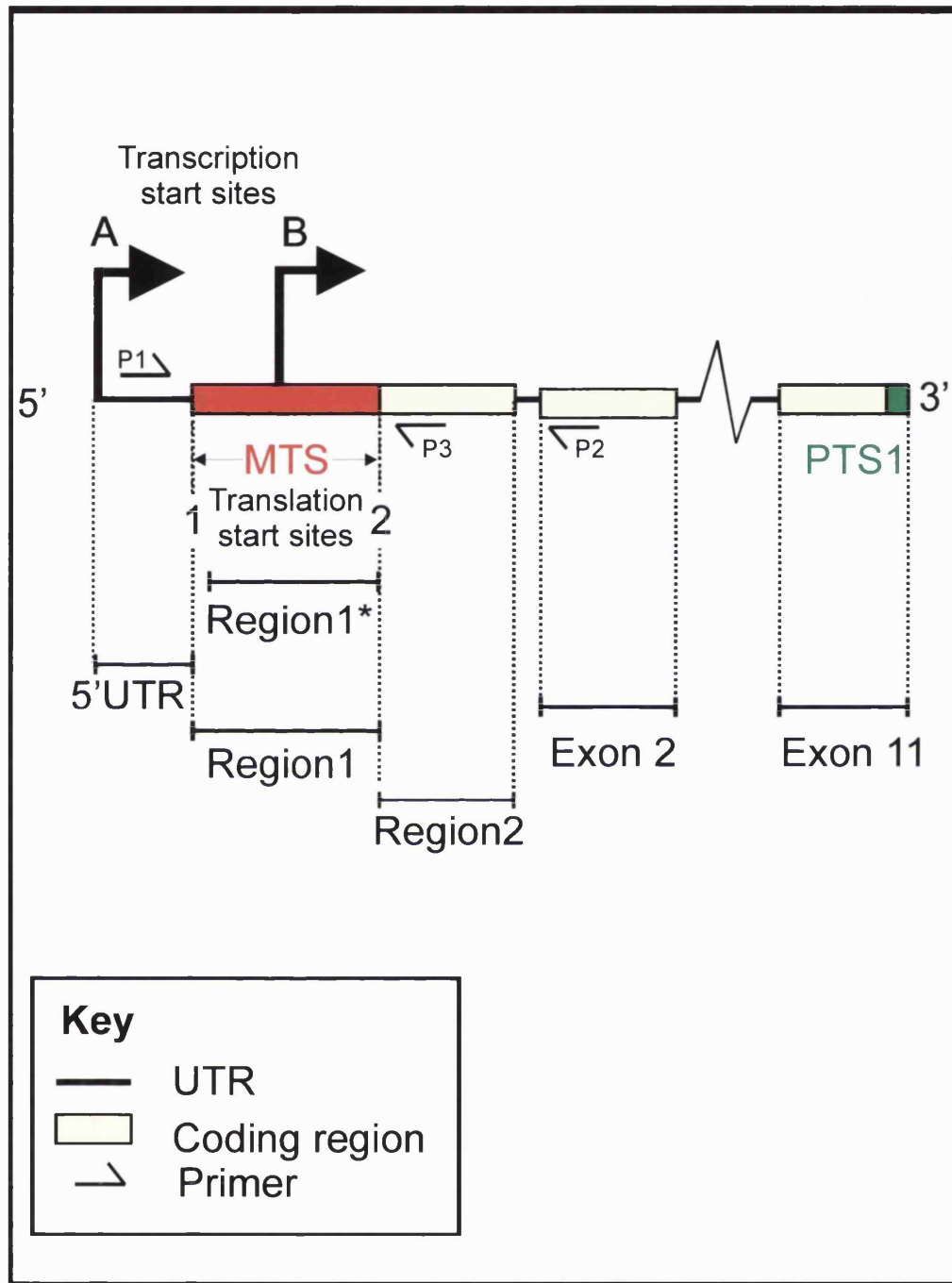


Figure 10 - Archetypal mammalian AGT gene with regions 1 and 2 indicated

Region 1 refers to the part of the gene running from TL-1 to TL-2. TL-1 is included in region 1 but not included in region 1*, TL-2 is not included in either. Region 2 refers to the part of the gene within exon 1 and downstream of (and including) TL-2. The mapping positions of PCR primers P1-P3 are indicated.

known AGT genes (i.e. human ⁴⁸ and marmoset ⁴⁷), were subjected to a variety of analyses. The analyses attempted to determine the nature of the mutational events that have lead to the differences in AGT targeting within anthropoid primates and their temporal relationships.

3.2 Results

3.2.1 Cloning and sequencing of the 5' region of the primate AGT genes

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 figure 10 and table 3 - page 43). 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 PCR products (excluding intron A and exon 2) and the P1/P3 PCR products are shown in figure 11, and their deduced amino acid sequences are shown in figure 12. The baboon and diana monkey clones did not sequence readably all the way through. As expected, the amplified regions were very similar to each other at both the nucleotide and amino acid levels. Following pairwise comparisons (figure 13), the nucleotide identity of region 1 varied between 78% and 100 %, while the amino acid identity varied between 62% and 100%. In this region, there was an average 61% nucleotide identity and a 33% amino acid identity across all

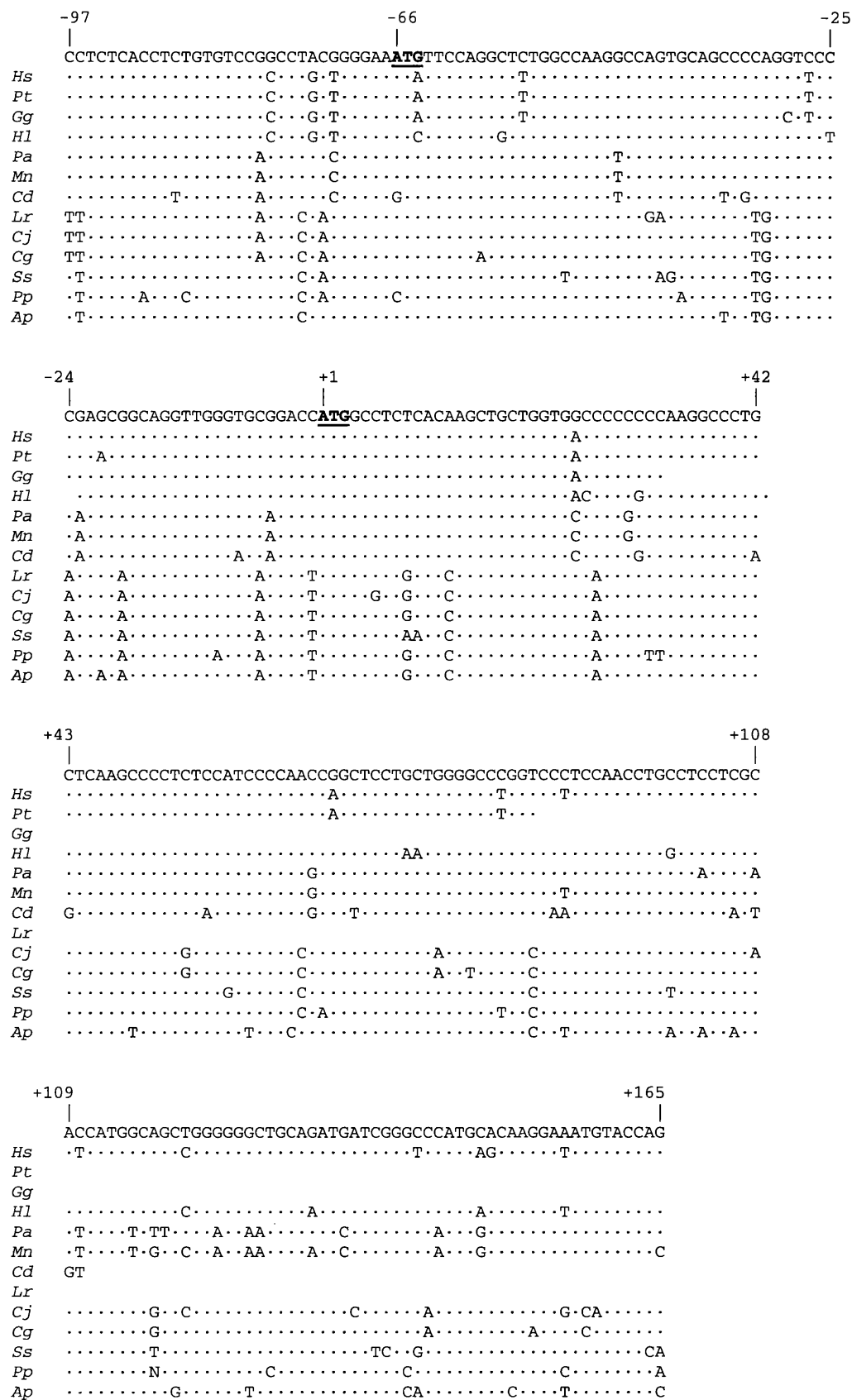


Figure 11 - Nucleotide sequence of the 5' region of the AGT gene in primates

See next page for legend

Figure 11 - Nucleotide sequence of the 5' region of the AGT gene in primates

The top 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 bold and underlined. The numbering starts from the first base of the codon for TL-2. For four species (*Pt*, *Gg*, *Cd* and *Lr*) only incomplete sequence was available. Species abbreviations are defined in table 2. Note that TL-1 has been lost in the human, chimpanzee, gorilla, gibbon, diana monkey and saki monkey.

| | | -22 | REGION 1 | +1 | REGION 2 | +55 |
|--------------|------------------|-----|--|----|---|-----|
| | <i>Consensus</i> | | <u>M</u> FQALAKASAA ^p <u>L</u> GPRAAGWVRT | | <u>M</u> ASHKLLVAPPKALLKPLSIPTRLLLPGGPSNLPPRTMAAGGLQMIGPMHKEMYQ | |
| <u>Hom.</u> | <i>Hs</i> | | I.....P.S..... | |T.....NQ.....I.....S.S.D... | |
| | <i>Pt</i> | | I.....P.S.T..... | |T.....NQ..... | |
| | <i>Gg</i> | | I.....PAS..... | |T.. | |
| | <i>Hl</i> | | I.G.....P..... | |P.....N..M.....A.....N.D... | |
| <u>Cerc.</u> | <i>Pa</i> | |S..P..Q...Q. | |P.....K.....T.I.SS..M.I...D..... | |
| | <i>Mn</i> | |S..P..Q...Q. | |P.....K.....I.S..M.I...D...H | |
| | <i>Cd</i> | | V.....S..VA..Q...EQ. | |P.A..V..Y..K.F...Q...S | |
| <u>Call.</u> | <i>Lr</i> | |D..L..... | | ...Q..... | |
| | <i>Cj</i> | |L..... | | ...Q.....L.H...T.. | |
| | <i>Cg</i> | |L..... | | ...Q.....H...T.. | |
| | <i>Ss</i> | | ...V..K..L..... | | ...NQ.....V.....S.....RA.....P | |
| <u>Ceb.</u> | <i>Pp</i> | | L.....T.L.....* | | ...Q...L.....Q.....P.....D... | |
| | <i>Ap</i> | |VL..T..... | | ...Q.....S..H.....TTS..G.....T..QD..H | |

Figure 12 - Deduced amino acid sequences encoded by the 5' region of the AGT gene in 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 methionines encoded by TL-1 and TL-2 are underlined in the consensus sequence. The numbering starts from the initiating methionine of TL-2. The deduced amino acid sequences from residues -22 to -1 assumes the presence of this region in the ORF, even though in many cases it is clearly not, see text for details. For species codes see table 2. Note the initiating methionine encoded at TL-1 has been lost in the human, chimpanzee, gorilla, gibbon, diana monkey and saki monkey. Hom. = Hominoidea, Cerc = Cercopithecidae, Call = Callitrichidae, Ceb = Cebidae

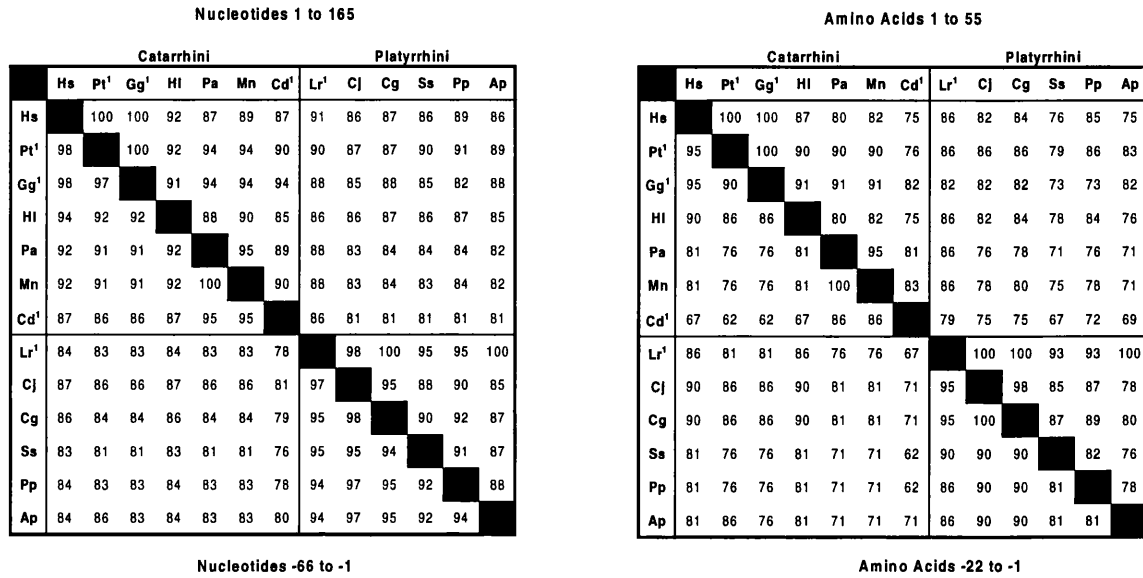


Figure 13 - Pairwise percentage identities of the 5' region of primate AGT genes

Percentage identities were calculated for each of the 78 individual interspecies pairwise comparisons, of region 1 (bottom left of grids, nucleotides -66 to -1 and amino acids -22 to -1), and region 2 (top right of grids, nucleotides 1 to 165 and amino acids 1 to 55).

See table 2 (page 41) for species codes.

¹ full-length sequence was not always available for the region 2 comparisons.

thirteen primates. Pairwise comparisons of region 2 (figure 13) showed that the nucleotide identity varied between 81% and 100% and amino acid identity varied between 67% and 100%. Excluding the missing nucleotides and residues in the chimpanzee, gorilla, diana monkey and tamarin sequences, the overall nucleotide identity was 62% and the overall amino acid identity was 48%. Overall, the level of conservation (i.e. sequence identity) of regions 1 and 2 appears to be similar, despite the likelihood that region 1 is frequently excluded from the ORF. In fact, based on the absence of TL-1 (figures 11 and 12), the ancestral MTS would appear to be excluded from the ORF in five out of the seven catarrhines studied (i.e. the human, chimpanzee, gorilla, gibbon and diana monkey) and one out of the six platyrrhines (i.e. the saki monkey).

3.2.2 Immunoblot of human and baboon livers

As TL-1 is present in the baboon AGT gene sequence, the explanation for the loss of mitochondrial targeting in the human (figure 5) does not explain the solely peroxisomal distribution of AGT in baboon liver. Another explanation for loss of mitochondrial AGT targeting could be a non-functional MTS. The arginine residue at position -2, which is the site for MTS cleavage in the other sequences, is not found in the baboon peptide sequence. This suggests that the baboon MTS is not cleaved. An uncleaved MTS would make the baboon AGT protein larger than its orthologues. To test if this is the case, an immunoblot was performed on human and baboon liver homogenates. Figure 14 shows that the AGT protein in human and baboon liver is the same size. Therefore, the mature baboon AGT protein does not have an uncleaved presequence.

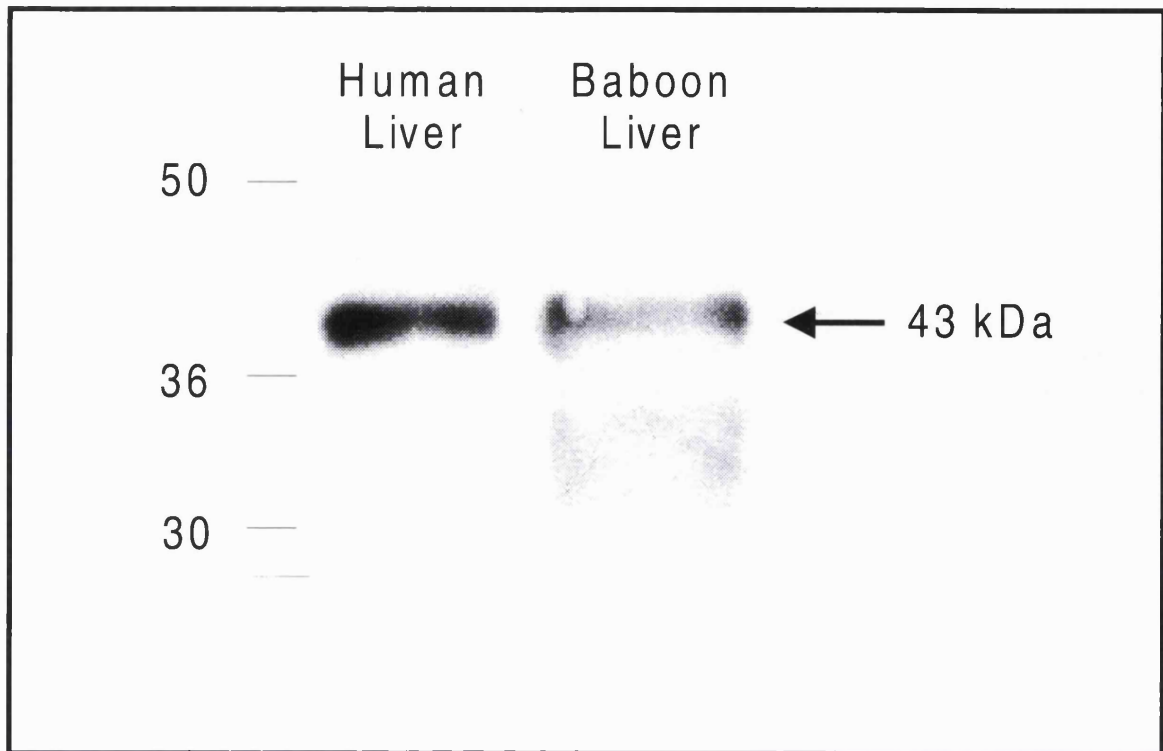


Figure 14 - Immunoblot of baboon and human livers

Homogenates of the livers were run on a SDS-PAGE gel and the proteins were transferred to a nitrocellulose membrane. The membrane was probed with a rabbit anti-human AGT antibody. The human and baboon AGT proteins seemed to be of identical size (~ 43 kD).

3.2.3 Ancestral reconstruction of region 1

An attempt was made 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 sequences⁷⁵. The results are shown in figure 15 (page 71). Numerous mutations have occurred that would be expected to abolish mitochondrial AGT targeting. For example, the analysis suggests that TL-1 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 residues⁴, the loss of two basic amino acids in branch 8 and the acquisition of two acidic amino acids, one in branch 10 and one in branch 19 would be expected to decrease markedly the effectiveness of region 1 as a 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.

3.2.4 The d_N/d_S analysis of region 1* and region 2

Because ancestral reconstruction of the region 1 sequence indicated that mitochondrial targeting of AGT could have been lost on at least four occasions (figure 15), an attempt was made to determine whether this was likely to be an adaptive response to positive selection pressure for change. Therefore, using a maximum likelihood method⁷⁶, the relative number of non-synonymous differences per non-synonymous site (d_N) compared with the number of synonymous differences per synonymous site (d_S) for each pairwise species comparison, have been determined (figure 16 - page 73). The

d_N/d_S ratio (ω) is frequently taken as a measure of the selection pressure having acted on genes or parts of genes⁸¹.

- ◆ Thus, $d_N/d_S < 1$ suggests that there has been pressure for conservation,
- ◆ $d_N/d_S > 1$ 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 ORF) and region 2 (which would be expected to be under continuous functional constraint because it is always included in the ORF) are shown in figure 16. TL-1 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 ORF or not, for rationale see Discussion (section 3.3.4.3).

3.2.4.1 Region 2

Several issues emerge from this analysis. When region 2 is compared between all thirteen 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.

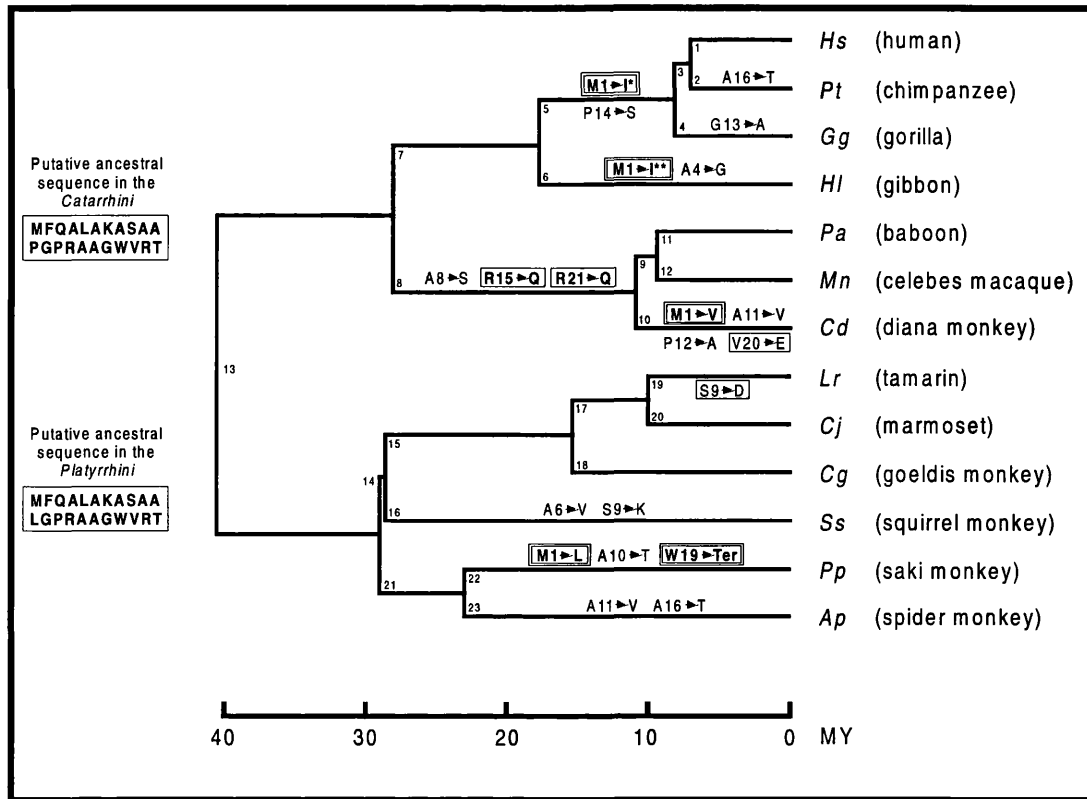


Figure 15 - Ancestral reconstruction of amino acid substitutions occurring in region 1 in primate lineages

The reconstructions are based on a maximum likelihood method using the computer program PAML (see Analytical methods). The phylogenetic tree and timescale are derived from Purvis (1995)⁸². The branches are arbitrarily labelled 1-23. The amino acid substitutions shown assume that region 1 remains within the ORF even though those following the loss of TL-1, clearly do not. Mutations that will abolish mitochondrial targeting are shown contained within a double box, and those that are predicted to diminish if not abolish mitochondrial targeting are shown contained within a single box. The loss of the TL-1 in branch 5 (*) is a separate mutational event from its loss in branch 6 (**). Not shown on this tree is the putative loss of TS-A in the baboon and macaque lineage (branch 8 or 9). The only difference between the reconstructed catarrhine and platyrrhine ancestral sequences is residue 12 (P in the former, L in the latter). As the residue in 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 means values for d_S and d_N in region 2 were 0.43 ± 0.30 and 0.09 ± 0.04 , respectively, giving an overall d_N/d_S ratio of 0.20, see table 4. To test if this value was representative of the whole codon region, the d_N/d_S ratios for the exons, of the previously cloned AGT genes i.e. the human, marmoset, rat, rabbit, cat and guinea pig AGT genes, were calculated (figure 17). The d_N/d_S ratio varies between exons. As expected the MTS is the least conserved, as the sequence requirements on a MTS tend to be degenerate (see introduction) and the MTS is often excluded from the ORF. Exon 4 is the most conserved with a d_N/d_S ratio of 0.08, exon 7 has the highest d_N/d_S ratio (0.55) and the average value for all eleven exons is 0.25. The values for all the exons are well below 1. Therefore, region 2 in primates is affected by the same selection pressures as the rest of the mammalian AGT genes (0.20 cf. 0.25). d_S does vary among exons from 0.712 - 0.361 and this can affect the d_N/d_S ratio, for example exon 5 has a lower than average d_N but because d_S is also low it has a higher than average d_N/d_S ratio.

3.2.4.2 Region 1*

When d_N/d_S is determined for region 1* a rather different picture emerges, depending on whether distant or close relatives are compared (figure 16, table 4). Thus, when the catarrhines are compared with the platyrrhines, $d_N/d_S < 1$ in 30/42 cases, and $d_N/d_S > 1$ in 12/42 cases, six of which involve comparisons with the same species (i.e. the diana monkey). The presence of ratios both less than and more than 1 might be indicative of variable functional constraint and positive selection pressure on a sequence that is predicted to be sometimes within the ORF and sometimes not. When the catarrhines are compared with each other, $d_N/d_S > 1$ in 19/21 cases, and $d_N/d_S < 1$ in only 1/21 cases. In the remaining one case $d_N = d_S = 0$. When the platyrrhines are compared with each other, $d_N/d_S > 1$ in 13/15 cases and $d_N/d_S < 1$ in only 2/15 cases. The high

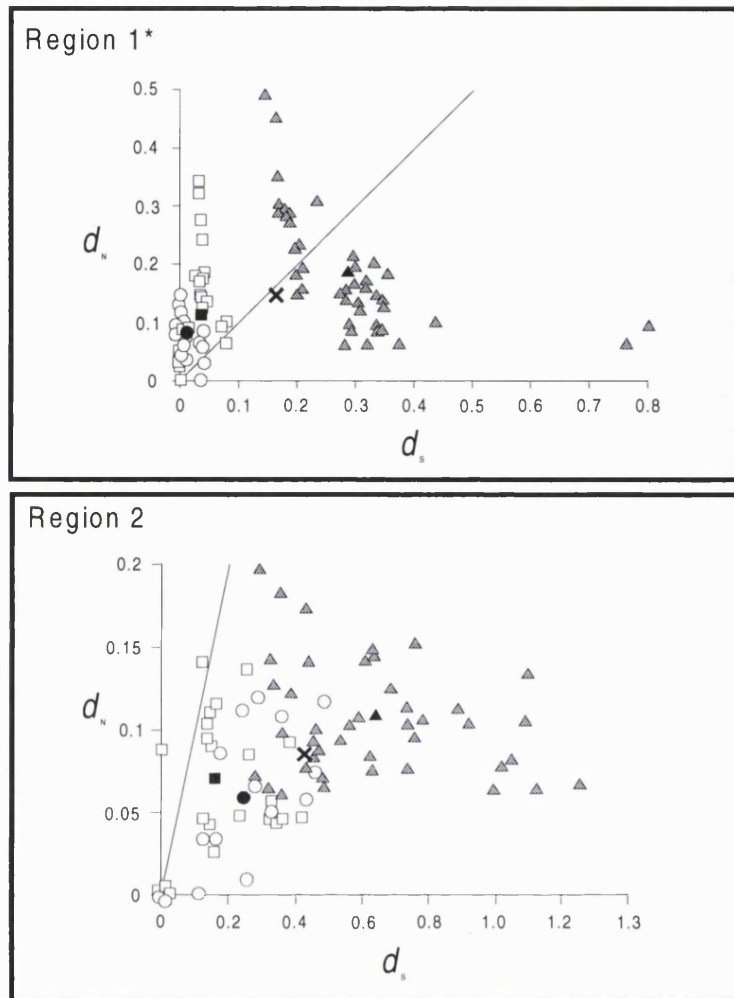


Figure 16 - The d_N/d_S substitution rates of the 5' region of the AGT gene in primates

The synonymous (d_S) and nonsynonymous (d_N) substitution rates were calculated using the maximum-likelihood method of Goldman and Yang (1994)⁷⁶ for each of the 78 individual interspecies pairwise nucleotide comparisons. There are 21 catarrhine-catarrhine comparisons (open squares), 15 platyrrhine-platyrrhine comparisons (open circles), and 42 catarrhine-platyrrhine comparisons (shaded triangles). The upper and lower panels show the plots of d_N against d_S for region 1* (excluding TL-1) and region 2, respectively (see figure 10). The black symbols show the means for each group of comparisons, and the cross indicates the overall mean (see table 4). The diagonal line indicates $d_N = d_S$. In region 1*, d_N was greater than d_S in 90 % of the catarrhine-catarrhine comparisons and in 87 % of the platyrrhine-platyrrhine comparisons, but was only more than 1 in 29 % of the catarrhine-platyrrhine comparisons (half of which were comparisons against the diana monkey (Cd)). The equivalent values for region 2 were 9.5 %, 0 %, and 0 %, respectively.

Table 4 - Mean values for d_N and d_S in pairwise comparisons

| | N | Mean d_N (\pm SD) | Mean d_S (\pm SD) | Ratio of mean d_N /mean d_S † |
|---------------------------|----|---------------------------|---------------------------|--------------------------------------|
| Region 1* | | | | |
| catarrhine v catarrhine | 21 | 0.141 \pm 0.093 | 0.031 \pm 0.046 | 4.54 |
| platyrrhine v platyrrhine | 15 | 0.075 \pm 0.039 | 0.012 \pm 0.019 | 6.25 |
| catarrhine v platyrrhine | 42 | 0.185 \pm 0.101 | 0.290 \pm 0.133 | 0.64 |
| Overall | 78 | 0.152 \pm 0.099 | 0.167 \pm 0.166 | 0.91 |
| Region 2 | | | | |
| catarrhine v catarrhine | 21 | 0.067 \pm 0.043 | 0.174 \pm 0.124 | 0.39 |
| platyrrhine v platyrrhine | 15 | 0.058 \pm 0.045 | 0.247 \pm 0.153 | 0.23 |
| catarrhine v platyrrhine | 42 | 0.106 \pm 0.035 | 0.632 \pm 0.261 | 0.17 |
| Overall | 78 | 0.086 \pm 0.044 | 0.434 \pm 0.302 | 0.20 |

† The means of the d_N/d_S ratios cannot be calculated because some individual comparisons give $d_S = 0$. Region 1* excludes TL-1.

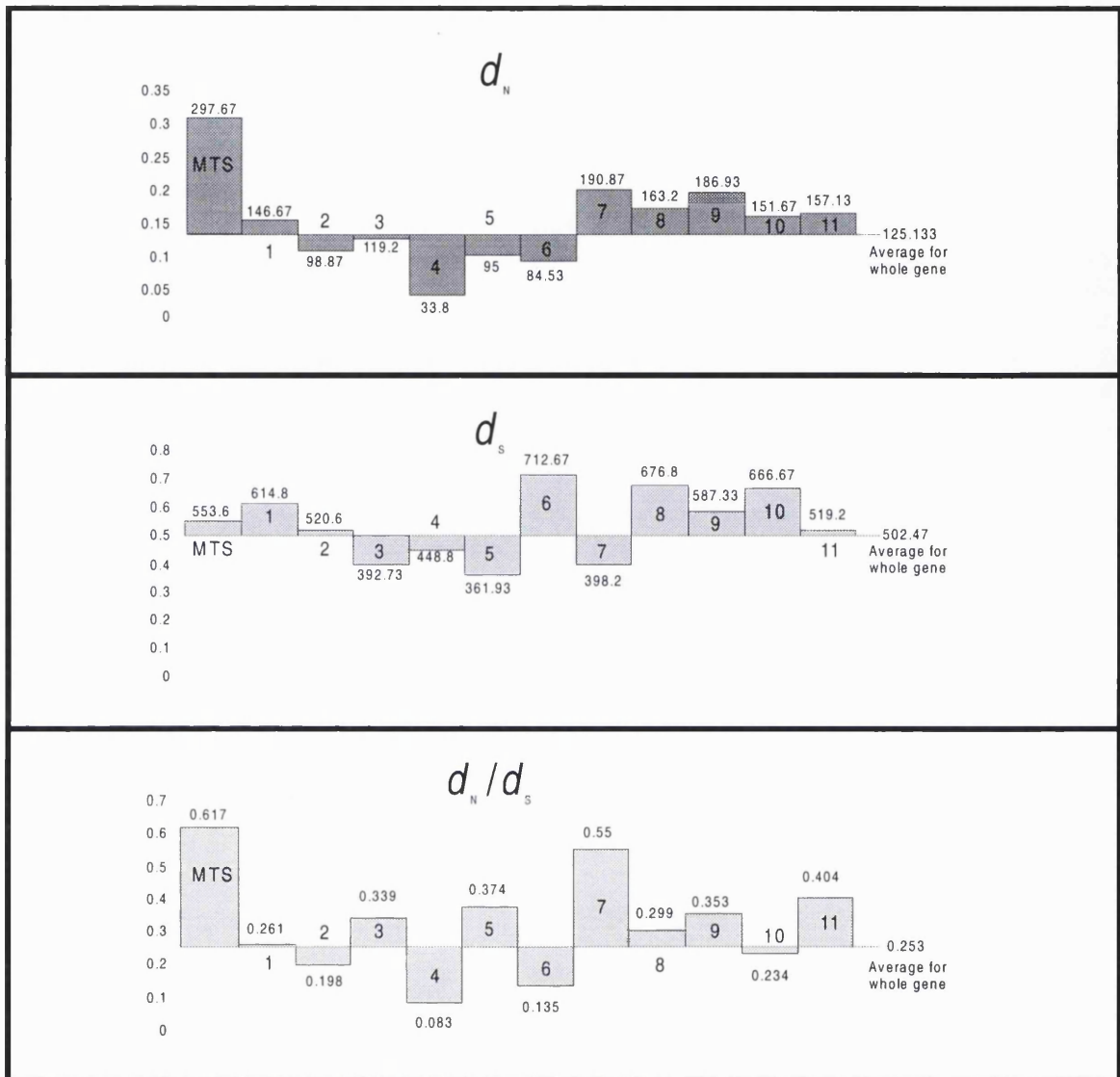


Figure 17 - Exon specific conservation of mammalian AGT genes

The pairwise d_N , d_S and d_N/d_S ratios of each exon was compared between human, marmoset, cat, rat, rabbit and guinea pig AGT. The average for each exon was calculated. The average value for the whole gene was computed and represented as the horizontal line on the histogram. Exon 1 was divided into MTS without TL-1 (region 1*) and the rest of the exon downstream of TL-2 (region 2). Region 1* was assumed to always be within the ORF for the purposes of the calculation although this is not the case for the human, rabbit and guinea pig. The values for each exon are placed near each bar. The MTS (region 1*) had the highest d_N/d_S ratio, which is unsurprising considering the degeneracy of MTSs and the frequency of its exclusion from the ORF. Exon 7 had a high d_N/d_S ratio which is partly due to a low d_S rate. Exon 4 has been noted to be very conserved before⁴⁹.

preponderance of short-range comparisons in which $d_N/d_S > 1$ (i.e. 32/36) is strongly suggestive of positive selection pressure for change acting on region 1*. Although only the analysis of region 1* is described for reasons detailed in the Discussion (Section 3.3.4.3), when region 1 (i.e. including TL-1) is analysed the results show the same pattern (figure A1 and table A1, in appendix).

3.2.5 Analysis of the d_N/d_S ratios among lineages

In order to determine the evolutionary basis of the differences in d_N/d_S ratios obtained by pairwise comparisons (see above), the likelihood ratio test was used to calculate d_N/d_S ratios among the different lineages⁷⁷ (figure 18). The data was fitted to various models of sequence evolution and the models were compared by the likelihood ratio test using the χ^2 approximation (table 5).

3.2.5.1 Region 1 *

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*, predictions of the ω ratios are > 1 in the majority (8 out of 12) of the branches in the phylogeny with non-zero estimated lengths (i.e. branches 2, 4, 8, 10, 16, 19, 22 and 23) (figure 18B). Eleven of the branches in region 1* have no substitutions at all and, therefore, have zero length and cannot be included in the analyses (figure 18). Of the four branches in region 1* in which ω is < 1 (i.e. 5, 6, 13 and 18), two of the more recent ones (i.e. 5 and 6) have ratios approaching one (i.e. 0.78), whereas the long ancestral branch 13 joining the catarrhines and platyrrhines has a ratio of only 0.13. Again when TL-1 was included in

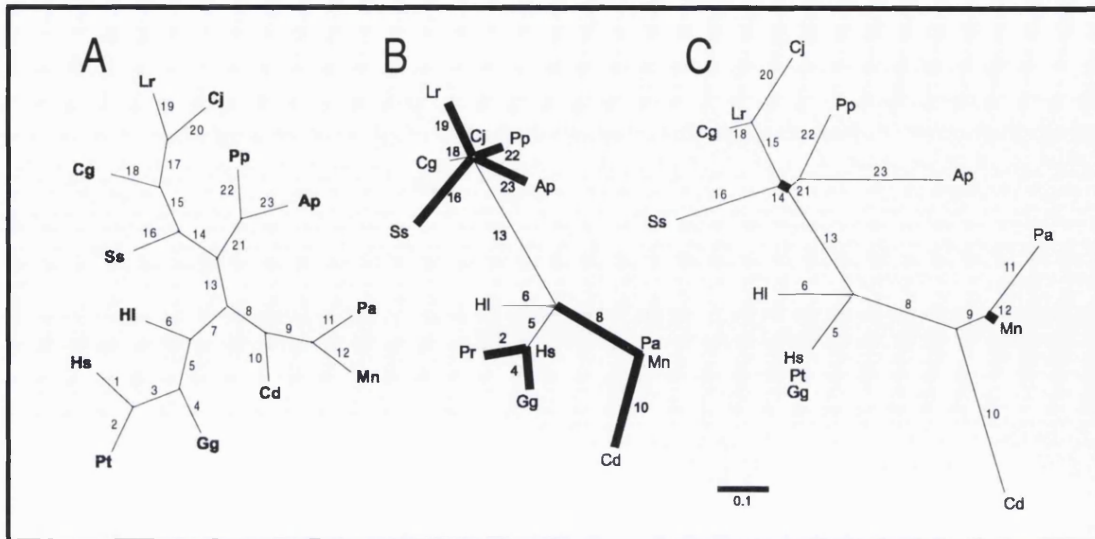


Figure 18 - d_N/d_S among lineages in the 5' region of the AGT gene in primates

Three trees are shown. The phylogenetic tree A is based on that of Purvis (1995)⁸² but with all branches (numbered 1-23 as in fig 15) drawn equal length. The derived trees B and C were generated using the computer program PAML (see Chapter 2, section 2.7.4) and drawn using the program TREEVIEW⁸⁴. The branch lengths were calculated using the one-ratio model and the d_N/d_S ratios calculated using the free-ratios model⁸⁵. The bar represents the expected number of nucleotide substitutions per codon estimated using the one-ratio model⁸⁵. 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 platyrrhines and catarrhines. Results of the analysis of region 1* 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 2) are attached to terminal nodes. The calculated d_N and d_S values (to 3 decimal places) along various lineages for region 1* (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 5 - Log-likelihood values and maximum likelihood estimates of parameters under different models

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

L = log likelihood value; $\omega = d_N/d_S$ (ω_0 is for ancestral branch 13, ω_1 is for all other branches, see figure 18) κ = transition/transversion rate ratio.

the analysis, the results showed a similar pattern. In fact, the region 1 results have $\omega > 1$ for 10 out of 12 of the branches in the phylogeny with non-zero estimated lengths. The long ancestral branch 13 has a ratio of 0.113 (figure A2, in appendix)

Another model that assumes the same ω ratio for all branches (the “one-ratio” model) gave $\omega = 2.00$ for Region 1*. To test whether this ω ratio is significantly greater than 1, it was compared with a one-ratio model in which ω was fixed at 1. When the likelihood ratio statistic ($2\Delta L$) was compared with the χ^2 distribution, the difference between the models was not significant ($P = 0.110$) (see table 6).

In order to test the hypothesis (based on the subjective assessment of the data to date) that positive selection has been driving only the recent evolution of region 1*, a “two-ratio” model, which assumes that the ancestral branch joining the catarrhines with the platyrrhines (i.e. branch 13) has a ω ratio (ω_0) that is different from all the other (more recent) branches (ω_1) was fitted to the data. 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, it was found that ω_1 for these recent lineages is significantly different from (greater than) 1 ($P = 0.0038$) (see table 6).

5.2.5.2 Region 2

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 < 1 except for two branches whose estimated lengths are very small (fig. 18C). The short branches contain less information about the ω ratio. Seven branches in region 2 had no substitutions and, therefore, have zero length and cannot be included in the analysis.

Table 6 - Test of evolutionary hypothesis

| Hypothesis being tested (H_0) | Models compared (see Table 5) | Test statistic ($2\Delta L$) | <i>P</i> value |
|-----------------------------------|-------------------------------|--------------------------------|----------------|
| Region 1* | | | |
| a) $\omega = 1$ | (1) & (2) | 2.56 | 0.110 |
| b) $\omega_0 = \omega_1$ | (1) & (3) | 10.16 | 0.0014 |
| c) $\omega_1 = 1$ | (3) & (4) | 8.38 | 0.0038 |
| Region 2 | | | |
| d) $\omega = 1$ | (5) & (6) | 36.46 | < 0.001 |

Hypotheses being tested: a) Tests if there is the same the d_N/d_S ratio in region 1* for all the lineages, is it different from one. b) Tests whether the d_N/d_S ratio in region 1 for the ancient branch (branch 13 in figure 18) is different from the ratio for all other lineages. c) Tests whether the d_N/d_S ratio in the more recent lineages are different from one, i.e. more than 1 for region 1*. d) Tests, if there is the same d_N/d_S ratio in region 2 for all lineages, is it different from one, i.e. less than 1.

The model assuming one ω ratio for all lineages gives the maximum likelihood estimate $\omega = 0.258$. This is significantly different from (smaller than) 1, with $P < 0.001$.

3.3 Discussion

3.3.1 The presence or absence of TL-1 is often, but not always, correlated with AGT protein distribution

Previous evidence suggested that the intracellular compartmentalisation of AGT is dependent on the presence or absence of TL-1 (i.e. the codon at nucleotide positions –66 to –64) (see figures 4 and 10). In its absence, all translation would be expected to start at the TL-2 (nucleotide positions +1 to +3). Although one or other of the ancestral transcription start sites (see figure 10) can also be lost, this has usually been considered to be a neutral secondary event⁴⁵. For example, previous studies have shown that the presence of ATG at the TL-1 in the common marmoset⁴⁷ and its mutation to ATA in the human ancestral line⁴⁸ is entirely responsible for the presence of both mitochondrial and peroxisomal AGT in the former, yet only peroxisomal AGT in the latter⁴⁴.

In addition to the absence of TL-1 in human⁴⁸ and saki monkey⁸⁶, the present study has shown that it is also absent in the following catarrhines: chimpanzee, gorilla, and diana monkey (see figures 11 and 12). Of these, the human, chimpanzee, gorilla and saki monkey are known to have a peroxisomal distribution of AGT⁴³ (and Danpure et al, unpublished observations). The subcellular distribution of AGT in the gibbon and diana monkey is not known. In addition to the presence of TL-1 in the marmoset⁴⁷ and baboon⁸⁶, it has been shown in this study that it is also present in the spider monkey, golden lion tamarin, goeldis monkey, and squirrel monkey (all platyrrhines), as well as

the celebes macaque (a catarrhine) (see figures 11 and 12). 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 AGT⁴³, it is likely that the golden lion tamarin, goeldis monkey and common squirrel monkey do as well. In addition, it is likely that celebes macaque has peroxisomal AGT as its close relative the japanese macaque does⁸⁷. Unfortunately, there are no clues to the subcellular distribution of AGT in the spider monkey.

With the notable exception of two of the Cercopithecidae, the baboon and the celebes macaque, the sequences at TL-1 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 TL-1 and hence the exclusion of the region encoding the MTS from the ORF, as suggested previously⁴⁵. Clearly, this cannot be the case for the baboon and macaque which must have lost mitochondrial AGT targeting by a different mechanism.

There are at least two possible mechanisms by which the baboon and macaque could have lost the ability to target AGT to the mitochondria without loss of TL-1. They could have lost the TS-A (figure 10), or they could have accumulated non-synonymous mutations in the MTS that prevent it from functioning as such. Loss of TS-A, which is upstream of both translation start sites, would result in the exclusion of region 1 (encoding the MTS) from the ORF as does the loss of TL-1. TS-A 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 TL-1⁴⁹.

On the assumption that region 1 is contained within the ORF in the baboon and 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 -helices⁴. For example, R(-2) and R(-8) have both been replaced by Q (figure 12). This would not only decrease substantially the net positive charge of the sequence, but also the loss of R(-2) would be predicted to interfere with presequence cleavage⁸⁸. In addition, the presence of three juxtaposed helix breakers, P(-11), G(-10) and P(-9), would call into question the ability of this region to fold into an α -helix.

Although both of the mechanisms proposed above for the loss of mitochondrial targeting in the baboon and macaque are possible, the former is the more probable (i.e. the loss of the TS-A). This is because the SDS-PAGE (figure 14) shows that AGT in the baboon liver is similar in size to human AGT (i.e. it lacks the 22 amino acid mitochondrial leader sequence). If the MTS was included within the ORF it would not be expected to cleave (due to the lack of a cleavage site) and therefore it would result in a larger protein. If these conclusions are correct then this is the first example of the loss of mitochondrial AGT targeting caused by loss of TS-A.

3.3.2 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 (i.e. the saki monkey) having lost mitochondrial AGT targeting and this is clearly independent from the loss of mitochondrial AGT targeting in the catarrhines (figure 15). In addition, the saki monkey is the only platyrrhine in which TL-1 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 TL-1 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 My ago⁸² (see figure 15).

The situation in the catarrhines appears to be much more complicated than in the platyrrhines. Although, as is so far known, the distribution of AGT in the catarrhines is uniform (i.e. peroxisomal), the molecular analysis and ancestral reconstruction suggests the possibility of four different reasons, namely mutation of the TL-1 to ATA in the human, chimpanzee and gorilla, to ATC in the gibbon, and to GTG in the diana monkey, and the loss of the TS-A in the baboon and macaque. The difficulty is determining which events in the evolutionary history of the catarrhines 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 figure 15). This is especially the case with the diana monkey, the more recent ancestral branches (i.e. branches 8 and 10 in figure 15) of which contain six non-synonymous mutations as well as the loss of the TL-1. Superimposed on this already complex situation is the predicted loss of TS-A in the baboon/macaque ancestral lineage

that presumably could have occurred in branches 8 or 9 (figure 15). 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.

3.3.3 Positive selection pressure to lose mitochondrial targeting has been widespread in the evolutionary history of anthropoids

Due to the large number of events that appear to have led to loss of mitochondrial targeting of AGT in primates, an effort was made to understand the nature of, and to quantify, the selection pressures that might have led to the changes. All the analyses point to region 1 (i.e. the MTS or the region encoding the MTS) behaving very differently from region 2 (figures 16 and 18).

When region 1 is included in the ORF 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 ORF, 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 ORF (i.e. diminish its effectiveness as a 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 > 1 in region 1* than in region 2 (figure 16). In addition, $d_N/d_S > 1$ is much more common when region 1* is compared between closely related species than 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, there was ancient pressure to conserve mitochondrial AGT targeting but this changed to more recent pressure to lose it. If generally true, then the catarrhines would appear to have been more “successful” than the platyrrhines at succumbing to that pressure.

Estimation of d_N/d_S ratios among lineages gives a compatible picture (figure 18). In region 1*, eight of the more recent branches have $d_N/d_S > 1$, three branches have $d_N/d_S < 1$ (although two of these have d_N/d_S ratios approaching one), and eleven branches have no substitutions at all. Significantly, the ancestral branch joining the catarrhines to the platyrrhines had a d_N/d_S ratio very much lower than one (i.e. 0.128). Statistical analysis confirmed the ancient/recent divide (table 6). 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 lysozyme⁸¹, but with the adaptive changes being concentrated in more recent lineages.

3.3.4 Theoretical problems with these analyses

There are five major problems with such analyses, some unique to the MTS of AGT and some more generally applicable to the comparison of short closely related sequences.

3.3.4.1 Problem 1 (chronology)

Pressure to lose or diminish mitochondrial AGT targeting could result in two consequences at the sequence level:

- ◆ loss of the TL-1, or
- ◆ changes to 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 e.g. Messier and Stewart (1997)⁸¹. However, loss of TL-1 and subsequent loss of mitochondrial targeting, would be an instantaneous switch and would not necessarily be reflected in the d_N/d_S ratios of region 1 as a whole. Therefore, high d_N/d_S ratios resulting from positive selection pressure can only occur while the MTS and TL-1 still exists within the ORF. Once TL-1 has been lost, any changes to the region (except perhaps those that might reinstate the MTS, or effect mRNA stability, translation or transcription efficiency) would be expected to be neutral.

It is possible that in branches where TL-1 has been predicted to have been lost (i.e. branches 5, 6, 10 and 22 - see figure 15) it was the last event to occur on that lineage and 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 TL-1 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, Problem 2).

3.3.4.2 Problem 2 (threshold effect)

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* or 66 nucleotides in the case of region 1) in closely related species where only relatively few mutational events have occurred. This is especially noticeable when region 1* is compared within the platyrrhines where d_S in most comparisons is zero (figure 16).

This is not found in region 2. Although marked differences in d_S across individual genes has been found before^{89;90}, the reasons remain unclear. As the total number of non-synonymous sites is usually greater than the number of synonymous sites, a threshold effect could artefactually elevate the d_N/d_S ratio in very closely related short sequences, i.e. the inequality of nonsynonymous and synonymous sites becomes a factor when the number of mutation events is low because the ones that have occurred will be more likely to be nonsynonymous and if no synonymous events occur, dividing by the amount of sites will not sufficiently correct for the inequality. Although this could conceivably contribute to the high d_N/d_S ratio when the platyrrhines (which often have $d_S = 0$) are compared with each other, it is less likely to contribute significantly to the high d_N/d_S ratio when the catarrhines (which have higher d_S values) are compared with each other.

3.3.4.3 Problem 3 (region 1 vs region1*)

It was uncertain whether to include TL-1 in the d_N/d_S analysis of region 1. The selection pressure acting on both region 1* and TL-1 is probably the same. However the consequences of change from the ancestral condition are dramatically different in each case. If TL-1 changes from the ancestral condition, it would exclude region 1 from the ORF and prevent mitochondrial targeting of AGT. Nonsynonymous mutations in region 1* may have functional consequences but, in most cases, are unlikely to abolish

mitochondrial targeting completely. Nonsynonymous mutations in region 1* will contribute to an incremental process of loss of mitochondrial targeting. The analyses were performed both with and without TL-1 (region 1 and region 1*, respectively). As excluding TL-1 from the open reading frame is considered to be more conservative, these are the analyses presented in the results section. However, analyses of both region 1 and region 1* produced similar results. The results for analyses with the whole of region 1 can be found in the appendix, figures A1 and A2 and table A1.

3.3.4.4 Problem 4 (some substitutions are more nonsynonymous than others)

As detailed in Chapter 1, some information is known about the sequence requirements for MTSs. However, the functional consequences of all the amino acid substitutions seen in figure 15 cannot be accurately predicted. Therefore, a nonsynonymous mutation in region 1* may not necessarily affect targeting. Therefore the d_N/d_S ratio for region 1* may overestimate the molecular adaptation occurring. However, as can be seen in figure 15, 9/21 of the peptide substitutions that have been predicted to occur are likely to negatively affect or abolish mitochondrial targeting, the effect of the other 11 substitutions is unknown e.g. substitution of W by a stop codon (branch 22) will abolish mitochondrial targeting, substitution of S by D (branch 19) will negatively affect mitochondrial targeting, and the effect of substitution of A by T (branch 2) is not known.

3.3.4.5 Problem 5 (methodology)

Estimates of pairwise d_N/d_S ratios were obtained using the maximum-likelihood method of Goldman and Yang⁷⁶. As this is a controversial field and a method of d_N/d_S estimation has not become standard, two other methods were also used Li (1993)⁷⁸ and

Nei and Gojobori (1986)⁷⁹. Although, the results (shown in the Appendix, figures A3 and A4) of the analyses varied, the overall pattern remained the same, with $d_N/d_S > 1$ for most short range comparisons of region 1 and $d_N/d_S < 1$ for all other comparisons of regions 1 and 2. Therefore, the overall pattern of d_N/d_S ratios is independent of the method used, for this data.

3.3.5 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 occurred⁹¹. 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 pressure⁸⁹. Some of the best known examples of genes, or parts of genes, that do show such pressure include primate lysozymes⁸¹, the antigen recognition sites of multiple histocompatibility complex genes⁹², the surface antigens of parasites and viruses⁸⁹, and abalone fertilisation proteins⁹³. Even if sequence comparisons provide strong evidence of positive selection pressure, it is not always easy to identify the nature of the selection pressure.

In the case of AGT, 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 whole⁴³ (Chapter 1), but the connection between the putative dietary selection pressure and the apparent frequent loss of mitochondrial targeting in anthropoid primates (especially the catarrhines) is much less clear. For example, the dietary variability in primates is restricted, as most are opportunistic omnivores (figure 19 - page 92)⁹⁴. Therefore a dietary selection pressure would not be expected to have a strong effect on primate AGT. However, the 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 primate lineages is compatible with various hypotheses which suggest that the diet of anthropoid primates, have indeed changed over the past 30-40 My or so. Body size of primates is closely correlated to diet⁹⁵; small primates tend to be insectivorous and frugivorous whilst large primates tend to be herbivorous and frugivorous. Extant anthropoid primates are believed to be much larger than ancestral primates, the fossil record shows an increase in anthropoid primate body size in the last 30-40 My putatively in response to competition from rodents⁹⁶. A shift in size would be expected to be accompanied by a shift from a more insectivorous diet to a more herbivorous diet^{94;97;98}. Changes in the dentition of the Cercopithecidae also indicate relatively recent changes in diet⁹⁵.

So, a hypothesis to explain the data presented in this study is: mitochondrial targeting of AGT protein was conserved whilst primate diets were insectivorous. However at the same time as anthropoid primate body sizes increased (in response to competition from rodents) diets became more herbivorous⁹⁹. The optimal AGT distribution became peroxisomal (being the best AGT distribution for herbivores) and there was selection pressure to increase peroxisomal AGT at the expense of mitochondrial AGT by removing the MTS from the ORF or accumulating nonsynonymous substitutions in the MTS (see figure 20 - page 93).

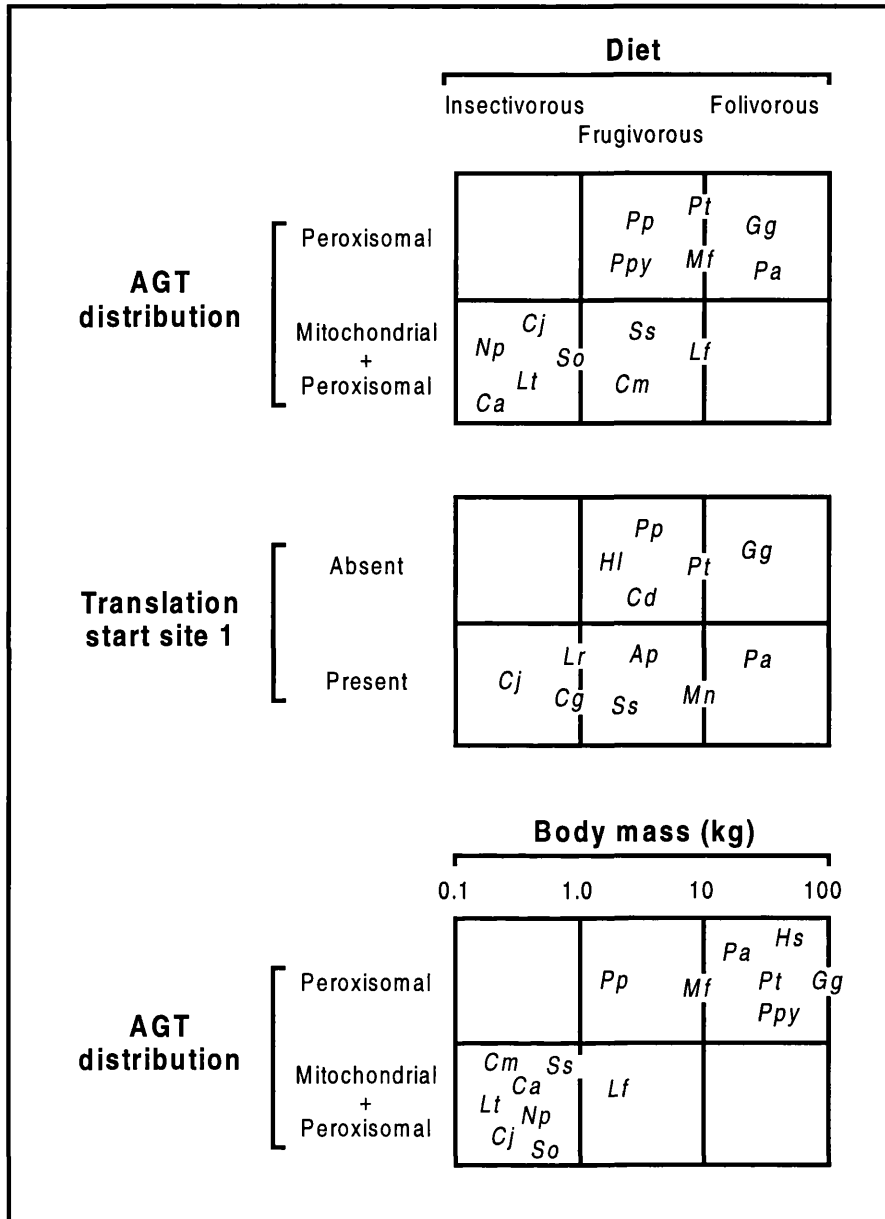


Figure 19 - AGT distribution and presence or absence of TL-1 in primates compared with diet and body mass.

Extant diets of primates are divided into three categories, mainly insectivorous, mainly frugivorous and mainly folivorous^{95;100}. Body masses were taken from Smith and Jungers (1997)¹⁰¹. Species abbreviations are defined in table 2. The human has been omitted from the dietary analysis. Large primates tend to have peroxisomal AGT, whereas the smaller primates (mainly platyrrhines and prosimians) tend to have both mitochondrial and peroxisomal AGT. Note that the saki monkey (Pp) is the largest platyrrhine studied and is the only one with a peroxisomal AGT distribution.

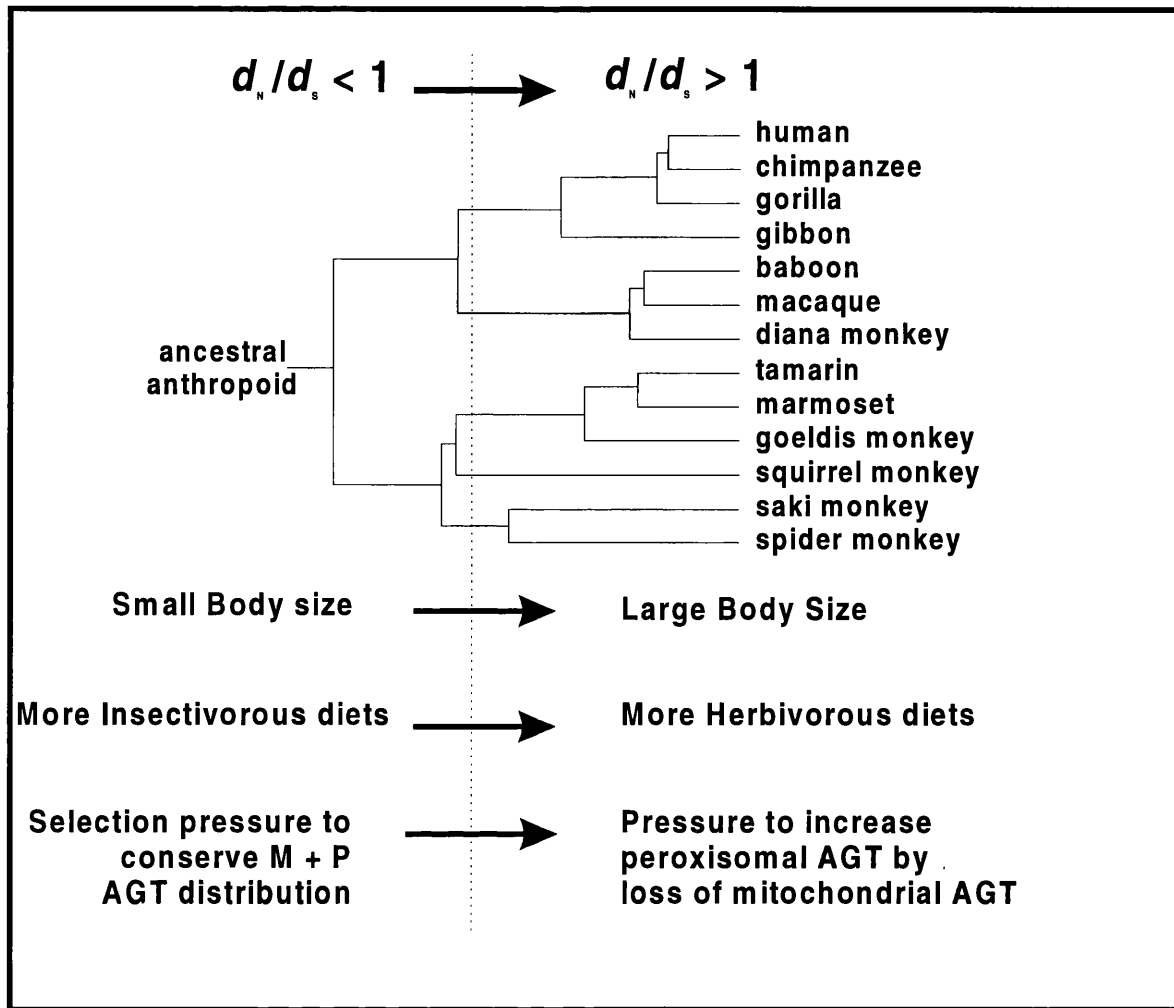


Figure 20 - Representation of adaptive hypothesis of primate AGT evolution

Changes over an evolutionary time scale have been manifested in the MTS of primate AGT by a change from $d_N/d_S < 1$, indicating selection pressure for conservation, to $d_N/d_S > 1$, selection pressure to change. This change in the pressures on primate AGT may be related to a change in body size known to happen during primate evolutionary history. An increase in body sizes would have been accompanied by a shift to more herbivorous diets. The optimal distribution of AGT protein in the livers of herbivores is thought to be peroxisomal; therefore a change in selection pressure from conservation of the MTS of AGT to selection to lose the MTS may have occurred. This would explain the change in d_N/d_S ratios. The largest primates have been the most successful at losing mitochondrial AGT, i.e. the catarrhines and the saki monkey (see figure 19).

Additional support for this hypothesis comes from the fact that catarrhines tend to be larger than platyrrhines¹⁰¹ (see figure 19), and the observation that the catarrhines have been rather more 'successful' at losing mitochondrial AGT than have the platyrrhines. Also, the saki monkey is the largest of the platyrrhines and the only member of the group known to have lost mitochondrial AGT targeting.

Thus, it is clear that the variable compartmentalisation of AGT is a unique and remarkable example of molecular adaptation as a consequence of positive (probably dietary) selection pressure.

4. A Molecular Explanation for the Unusual Distribution of AGT in Amphibian Liver

4.1 Introduction

In amphibian species, including the newt, common frog, pobblebonk, bullfrog and xenopus, hepatic AGT is mitochondrial and cytosolic in all developmental stages (figure 21), despite the diet of many amphibian species being herbivorous in the tadpole and carnivorous in the adult. The only precedent for cytosolic AGT was discovered in guinea pig liver⁴⁹, in which inefficient targeting of guinea pig AGT protein leads to a peroxisomal and cytosolic distribution of AGT⁴⁹. It is unusual to find a single gene product located in both the mitochondria and the cytosol⁴²; inefficient mitochondrial targeting has been found in the case of *Sacchromyces cerivisae* fumerase¹⁰² but remains very rare. So, the distribution of xenopus AGT may be the result of a novel molecular mechanism for AGT targeting.

Previous to this project only mammalian AGT genes had been cloned and studied (see Chapters 1 and 3). Consequently, the amphibians represent an outgroup for the mammalian studies, which is valuable as it may provide information about the AGT in the ancestral mammal.

To these ends, xenopus AGT was cloned and functionally characterised. The xenopus AGT gene was chosen as a representative of amphibian AGT, due to the availability of a xenopus cDNA library.

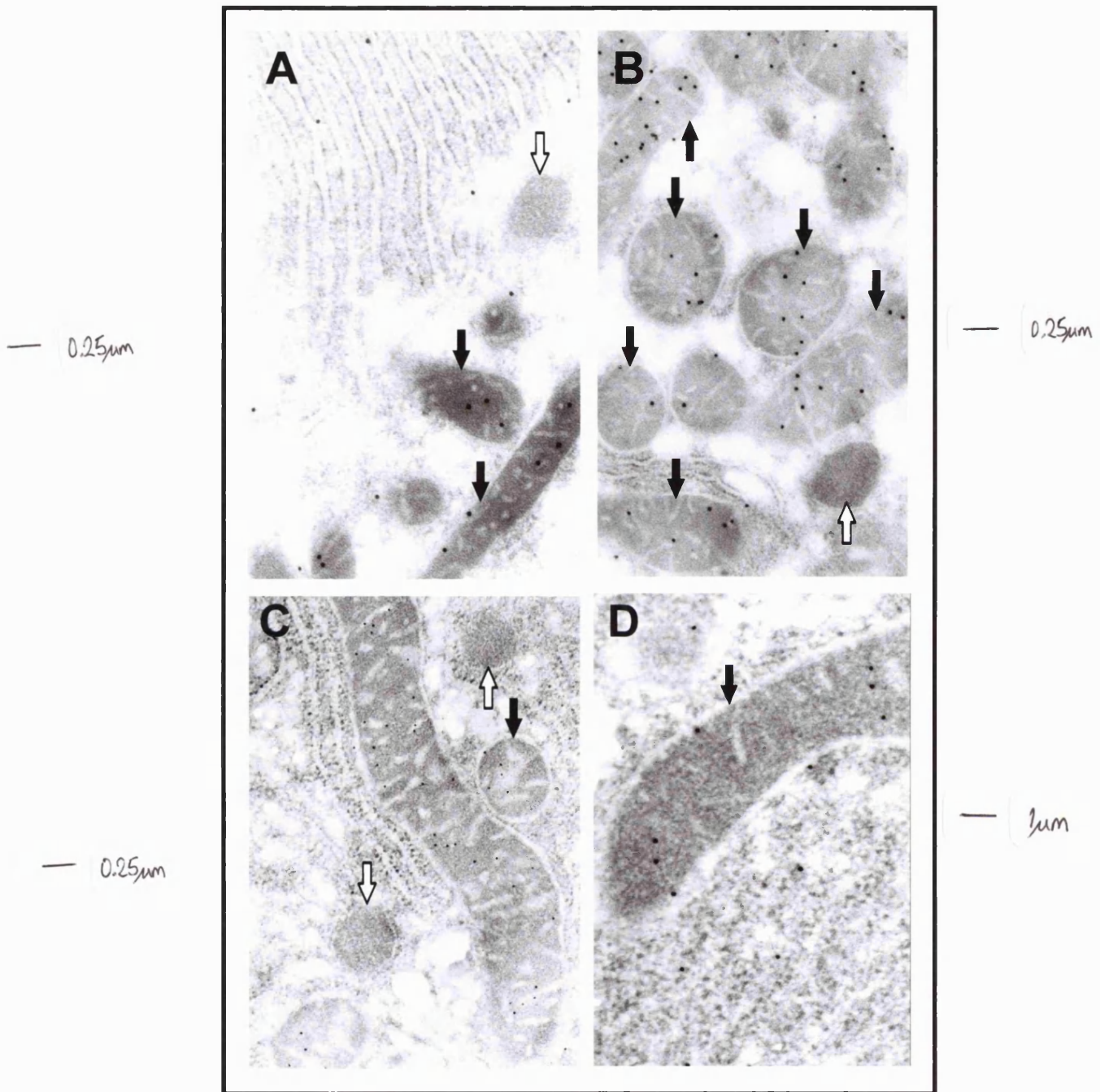


Figure 21 - Immuno-electron micrographs of amphibian livers

Sections of newt tadpole liver (A) (Danpure et al, unpublished), adult common frog liver (B)⁴³, pobblebonk tadpole liver (C) (Gallagher, unpublished) and adult xenopus liver (D) (Gallagher unpublished) were labelled for AGT with protein A gold.

Mitochondria are indicated with black arrows and peroxisomes are indicated with white arrows. AGT labelling is mitochondrial and cytosolic but not peroxisomal.

4.2 Results

4.2.1 Cloning the xenopus AGT gene

Attempts to clone the xenopus AGT gene by RT-PCR, using primers designed by alignment of previously known AGT sequences were not successful. It was considered that it may be due to a high level of divergence between the nucleotide sequences of the human and xenopus AGT genes. Therefore “DNA-based” strategies were abandoned and a “protein-based” method using the anti-human AGT antibody was attempted. It was known that the anti human AGT antibody recognised the xenopus AGT protein as immuno-electron microscopy (figure 21) had been performed successfully on amphibian tissue.

4.2.1.1 Screening an expression library

An anti human AGT antibody was used to screen a xenopus liver cDNA expression library constructed in Uni ZAP XR (Stratagene) (see Chapter 2, section 2.6.12). 300,000 plaques were screened. Three sequential screens were performed on the library and the tertiary screen yielded seven plaques that bound the antibody. The DNA from the plaques was recovered and sequenced (see Chapter 2 for sequencing strategy, section 2.4.2). The resulting sequences were subjected to BLAST searching (table 7). Clone 4 revealed a homology to cat AGT and seemed to be an AGT homologue but was non-identical to any other AGT genes previously cloned. Clone 4 was named pXS (see figure 8 - page 46). It aligned against the human AGT nucleotide sequence, but the 5' end was 3' of TL-1. As no putative MTS could be identified in the clone 4 sequence, the clone was suspected of not being full length.

Table 7 - BLAST searching of library screen positives

| Plaque Number | Closest Relative found in BLAST search | Significance Rating* |
|---------------|--|----------------------|
| 1 | <i>Papio hamadryas</i> Cyclophilin A | $1e^{-8}$ |
| 2 | <i>Trichuris trichiura</i> Tt52 mRNA | 0.048 |
| 3 | Unidentified sequence on Human Chromosome 7 | 0.1 |
| 4 | <i>Felix cattus</i> AGT | $3e^{-8}$ |
| 5 | pBluescript | $2e^{-62}$ |
| 6 | <i>Caenorhabditis elegans</i> cosmid F25B4 | 0.19 |
| 7 | <i>Xenopus laevis</i> SSB2 gene | $3e^{-6}$ |

* The significance rating is the probability that a comparison between the query sequence and a random sequence would achieve a score greater than or equal to the score attributed to the match between the query sequence and the returned sequence.

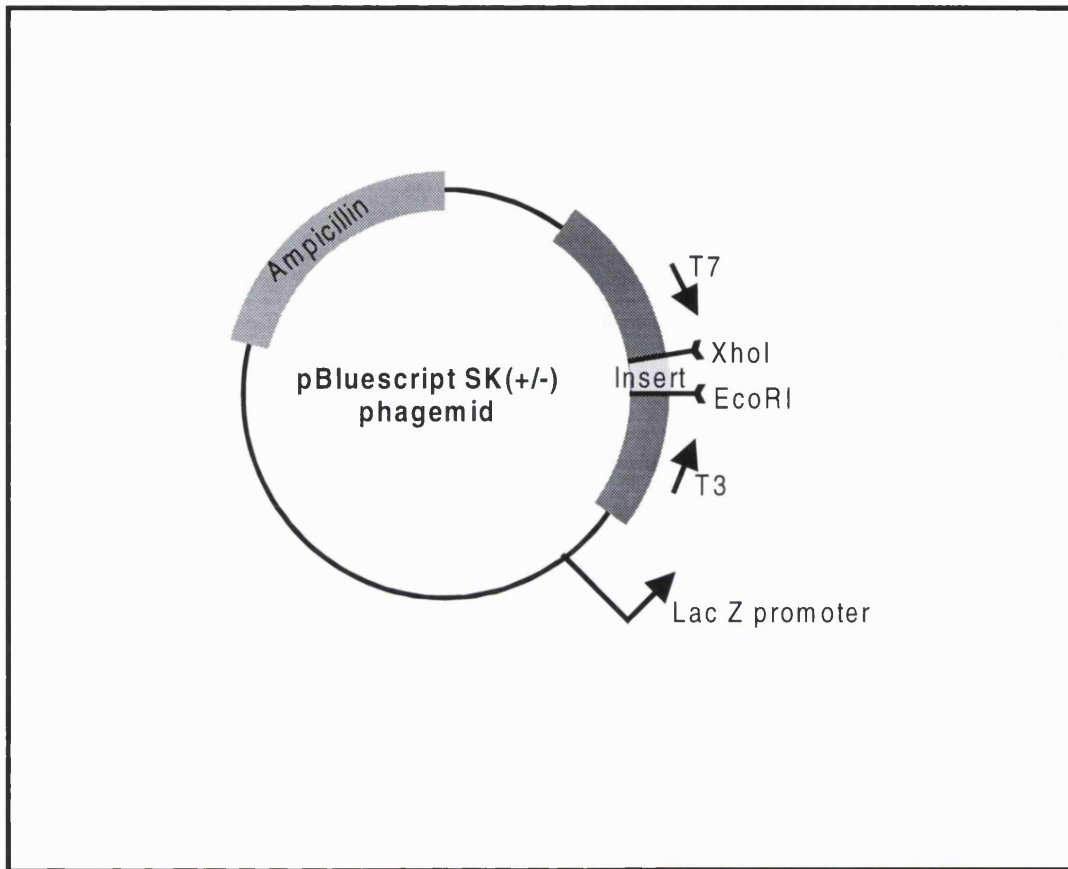


Figure 22 - Cloning vector of xenopus cDNA library

Inserts were directionally cloned with *EcoRI* and *XhoI*. The inserts are under the influence of the LacZ promoter, therefore the sense strand of the inserts should run from 5' at T3 to 3' at T7.

Surprisingly, it was noted that pXS seemed to be in the opposite orientation to the one that would have been expected given the cloning strategy used to generate the library and the direction of action of the LacZ transcriptional promoter (figure 22). This was very puzzling as the clone was detected by antibody binding and could not have been detected if expressed in the wrong orientation.

4.2.1.2 PCR and Southern blotting analysis of the xenopus expression library

To investigate the orientation anomaly and to determine if there were clones with more 5' sequence in the library, the library was used as template in a PCR reaction. The PCR reactions used the plasmid primers T3 and T7 (figure 23) and clone specific primers. The PCR reactions were subjected to Southern blot analysis using pXS as the probe. The reactions using T3 as the plasmid primer and primers P14, P16, P11 and P13 (figure 8 and table 3) failed to yield any products, but the reactions using T7 as the plasmid primer and P14, P16, P11 and P13 yielded products that hybridised to pXS (see figure 22, for positions of T3 and T7 primers). The products were the same size as would be expected for a cDNA identical to pXS (figure 23). PCR using P14 and T7 did not yield a product but this was unsurprising as P14 hybridises to only 22 nt 3' to the start of the pXS sequence (see figure 8 and table 3 for primers). Therefore it seems that the library contained inserts in the same orientation as pXS. Therefore there was no switch in orientation during recovery of the bacteriophage DNA. It was also inferred that the library did not contain any inserts with more 5' sequence so 5' RACE was carried out to acquire more 5' sequence.

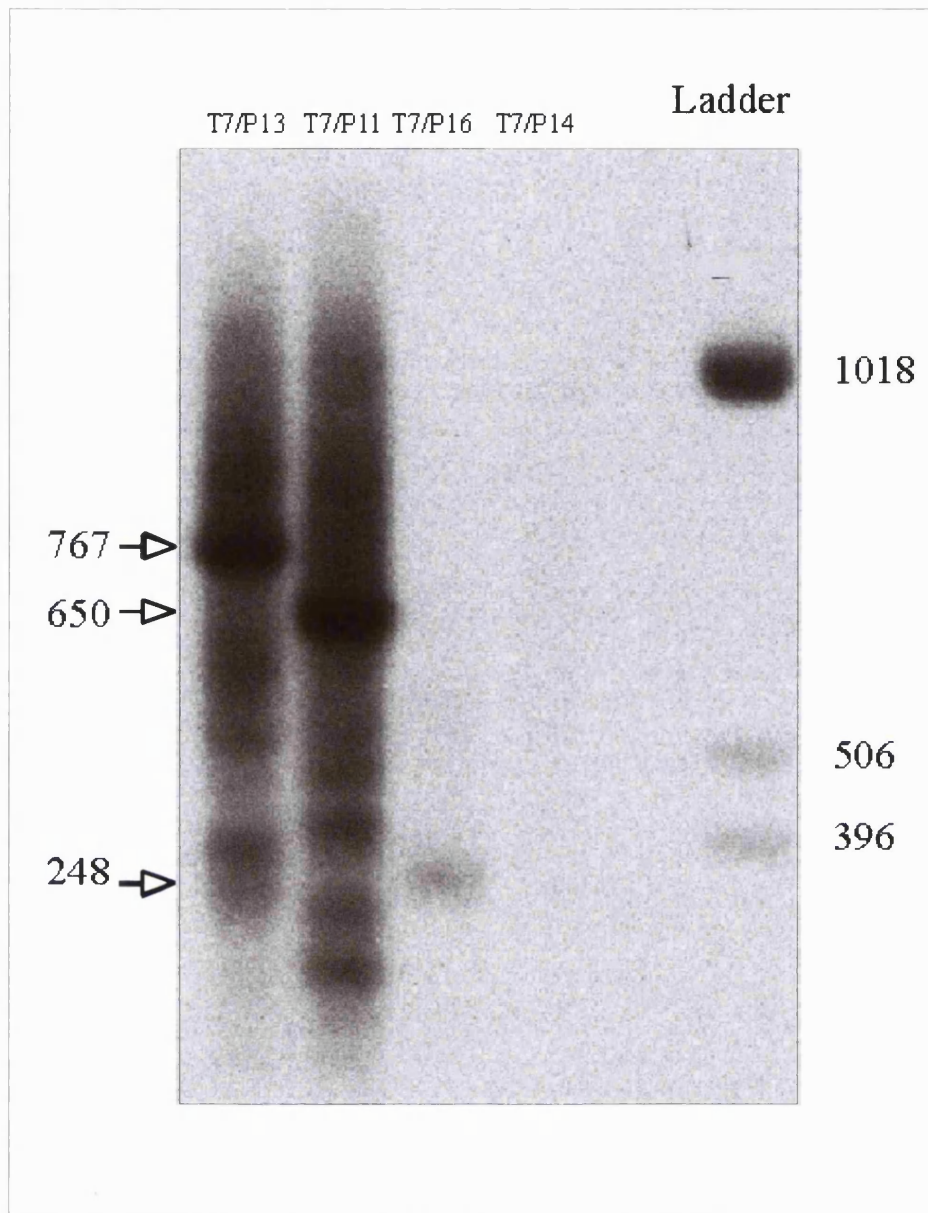


Figure 23 - Southern blot of PCR products from xenopus library PCR

Four primer pairs were used, T7 & P13, T7 & P11, T7 & P16 and T7 & P14, in a PCR reaction with the xenopus cDNA library as template. The reactions were predicted to generate products of size: 767 nt, 650 nt, 248 nt and 22 nt respectively, from the known sequence. The PCR products were transferred on to a nylon membrane and probed using a radiolabelled probe generated from random priming of pXS.

4.2.1.3 5' RACE of pXS

5' RACE was performed using semi-nested primers P14, P15 and P16 (figure 8).

Agarose gel analysis of the third round PCR showed two bands of approximately 350 nt and 100 nt (figure 24). These were cloned and sequenced. The smaller cDNA clone (pRACE-S) was found to correspond exactly to the 5' end of pXS. The larger cDNA clone (pRACE-L) was found to extend 280 nt upstream of the 5' end of pXS.

It was noted that pRACE-L, cloned into the pGEM-T Easy plasmid, was prone to rearrangement when passaged through JM109 *E. coli* cells. However, pRACE-L could be grown up in SUREII *E. coli* cells.

The pRACE-L predicted amino acid sequence, included stop codons at its N terminus suggesting it included the whole 5' end of the ORF. The predicted peptide sequence also contained two in frame potential translation start sites, 24 amino acids apart, roughly corresponding to the alternative translation start sites of the cloned mammalian AGT cDNAs (figure 4).

4.2.1.4 RT-PCR of the xenopus AGT

To verify that pXS was a genuine transcript and to obtain full-length clones, RT-PCR was performed, using xenopus liver total RNA as template. Two primer pairs were used: P4/P5 and P6/P5 (figure 8). Both were expected to yield a PCR product that included both putative translation start sites. Primers P4 and P5 produced a product, which included nearly all of the known sequence, this was named Xall. When pXall was cloned and passaged through *E. coli* cells it rearranged in both the JM109 and

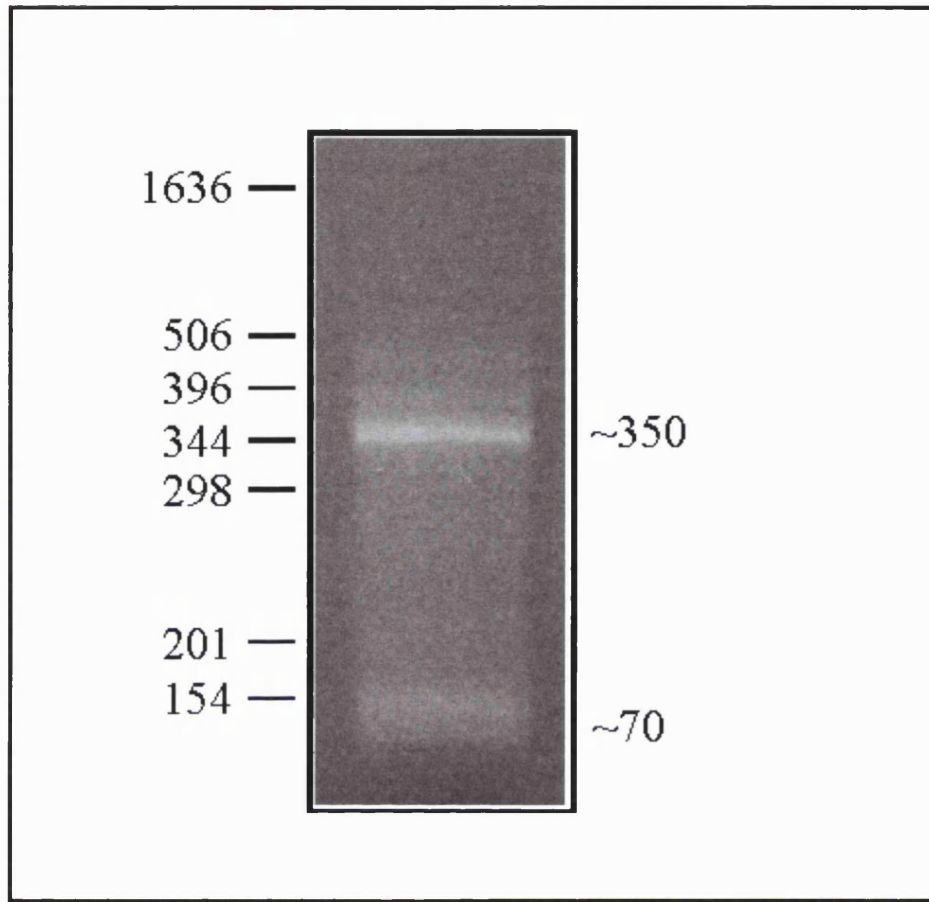


Figure 24 - Products of the 5' RACE reaction

RACE produced two products of approximately 350 and 70 nt in size as estimated by electrophoretic separation. The sequence of the smaller band corresponded exactly to the 5' sequence of pXS, the larger band yielded ~ 280 nt of upstream sequence.

SUREII strains. Primers P6 and P5 produced a product, which was thought to exclude some of the theoretical recombination sites (shown in figure 25), and named XL. When pXL was cloned it was found to be recombinogenic in JM109 cells but did not seem prone to rearrangement in SUREII cells.

4.2.2 Analysis of the xenopus AGT sequence

The Xall PCR product was sequenced (for strategy see Chapter 2, section 2.4.2) from the PCR product. The clones pXL and pXS were also sequenced and the sequences were compared with Xall and the overlapping sequences were found to be identical. The nucleotide and deduced peptide sequence of the xenopus AGT gene is shown in figure 25. The ORF (excluding the putative MTS, i.e. the 24 amino acids between the 2 in-frame methionines) of the xenopus AGT gene is 64% identical to the human AGT ORF at the nucleotide level and 66% identical and 75% similar to the human AGT at the amino acid level (figure 26). The peptide sequence between TL-1 and TL-2 was predicted to be an MTS, as it includes arginine residues and no negatively charged residues, and maps to roughly the same position as the sequence encoding the MTSs in other mammalian AGT genes.

4.2.3 *In vitro* transcription and translation of the xenopus clones

The pXL, pXS, p*Hs*AGT (normal human AGT) and p*Hs*AGT⁺ (human AGT with TL-1 reinstated) clones were *in vitro* transcribed and translated and the radiolabelled products were separated on an SDS PAGE gel (figure 27A). The *in vitro* translated pXL clone resulted in two major bands, one was larger than the *Hs*AGT (known to be 43 kD). The

```

1  gtcggaaagggtttttgtcatcagggtgaaacagcctctcagtcacagagcacaggtcagaa 60
61  ccccgtagcctctgggttggaggaggattccagaattcagagaaagagatagggtggaaaaa 120
121 gagcggcatgcattagacaagccagtgtaggaccagaacaacaaatagggactagggtgaa 180
181 ggttcactggggtcatttatctgagaccttttagaagttaacaagatgcaagggtcag 240
      M Q G S V
241 ttagaagcattttcttctctctctctctctctctctctctctctctctctctctctctca 300
      R S I S S S L L C A A R L S A P V R I M
301 tggctctcgtcttgccactatccctccaccttcagctcttcagcggcccttgaatgtcccc 360
      S S L A T I P P P S A L Q R P L N V P Q
361 aacgctcatgcttggccccggaccctccaatgtacctcctcggattcaagcagcagggg 420
      R L M L G P G P S N V P P R I Q A A G G
421 gcttacagcttattggtcacatgcaccctgaaatgttccagataatggatgatatcaaac 480
      L Q L I G H M H P E M F Q I M D D I K Q
481 aagggatacaaatatgctttccagacaaaaaacctgacatttgctgtgagtgatccg 540
      G I Q Y A F Q T K N N L T F A V S G S G
541 gacactgtgccatggagactgcaatttttaacgtggtggagaaaggagatggtgctcttg 600
      H C A M E T A I F N V V E K G D V V L V
601 ttgctgtaaaaggaatttgggggaaagagccgggtgatattgcagagcgaatagggtgagg 660
      A V K G I W G E R A G D I A E R I G A D
661 atgtacgatatgtgtcaaaaccagtaggtgaagcctttacacttaaggatgtagaaaagg 720
      V R Y V S K P V G E A F T L K D V E K A
721 ctttagcagagcacaagccatcgctcttcttcttaccacggagagtcacccagtgagg 780
      L A E H K P S L F F I T H G E S S S G V
781 tggttcagcccttagatggcctgggagatctctgtcaccggtataactgtctactgcttg 840
      V Q P L D G L G D L C H R Y N C L L L V
841 tagactcctgtagcctctctaggtggagctccaatctacatggacaaacaggggaattgaca 900
      D S V A S L G G A P I Y M D K Q G I D I
901 tttatattctggatctcagaagggttctgaatgccctccaggaactgctcccatttct 960
      L Y S G S Q K V L N A P P G T A P I S F
961 tcagtgaagctgcaagtaagaagatggttggccgtaaaacaaaacctccatctttgtatg 1020
      S E A A S K K M F G R K T K P P S L Y V
1021 tggatataaaattggcttgcaaaactattggggatgtgatggcaaaccaagaatctaccacc 1080
      D I N W L A N Y W G C D G K P R I Y H H
1081 aactggaccagtgactaatttcttcccttgagagaaggactggcaattcttctgctgagc 1140
      T G P V T N F F T L R E G L A I L A E L
1141 tgggattggaacgctcctgggctgtgcaccaagagaatgcactgaagctgcataaaggat 1200
      G L E R S W A V H Q E N A L K L H K G L
1201 tggagcactaggaataaagctttttgtaaaggaccagcactgcgctccccaccgctca 1260
      E A L G I K L F V K D P A L R L P T V T
1261 caacaattagcgttccaaatgggtatgaatggaagacataactacatttatcatgaaga 1360
      T I S V P N G Y E W K D I T T F I M K N
1321 atcacgccattgaaatcacaggaggcctggggccatctactggaagggttctgcgtattg 1380
      H A I E I T G G L G P S T G K V L R I G
1381 gcctcatgggctataactcaacgcaactgaatgtagatcgtgttctagaggctctgcgtg 1440
      L M G Y N S T Q L N V D R V L E A L R D
1441 atgctctgcagcaatgtcccaaaaaaagatgtgaataacctgtttgctgaagatgacgg 1500
      A L Q Q C P K K K M *
1501 ctccataaaattgtcattaaatccaaactgtaacttcaactcattcatcattctcatacaat 1560
1561 atatgaatcactctcttctactatcatcacagatacctcaattaaataactaaaaataaca 1620
1621 aaaaaaaaaaaaaaaaaaaaaaacctcgtgccc 1652

```

Figure 25 - Nucleotide and predicted amino acid sequence of xenopus AGT

Legend on next page.

Figure 25 - Nucleotide and predicted amino acid sequence of xenopus AGT

The sequence is a composite of the sequence of Xall and the 3' end of pXS. Predicted amino acid sequence is shown underneath the first base of every codon of the nucleotide sequence. Translation start sites are underlined and bolded. The putative poly-adenylation signal is underlined. The putative pyridoxal binding site is boxed. The A and G rich regions in the 5' UTR, which are suggested to be sites of recombination are bolded. Codons which would code for stops are shaded. Arrowheads mark the start of pXL and pXS.

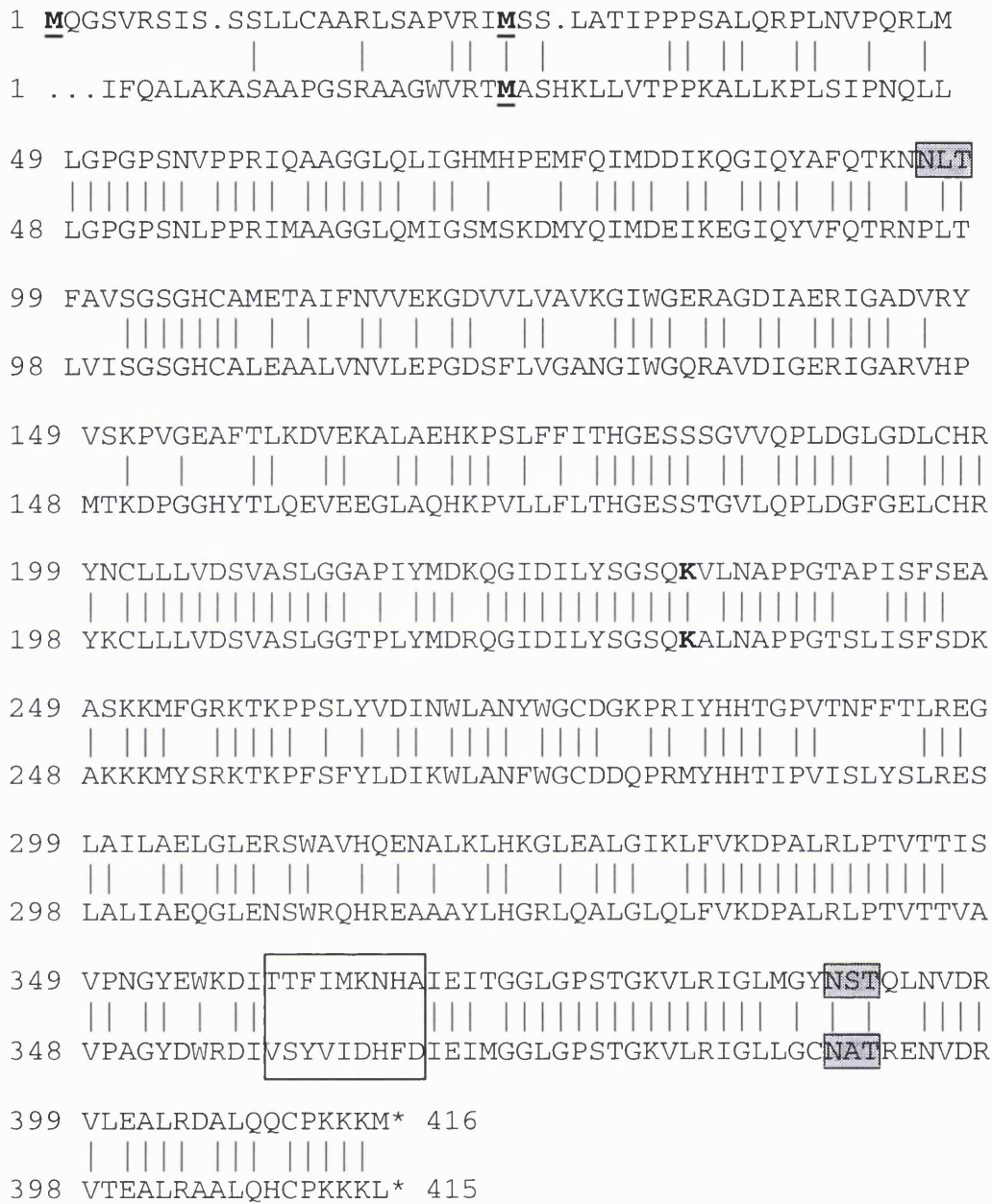


Figure 26 - Alignment of xenopus AGT peptide sequence and human AGT peptide sequence

The xenopus sequence runs above the human sequence. Potential translation start sites are bold and underlined. A region of low homology is boxed. Potential glycosylation sites are also boxed and shaded. Pyridoxal phosphate binding sites are bolded.

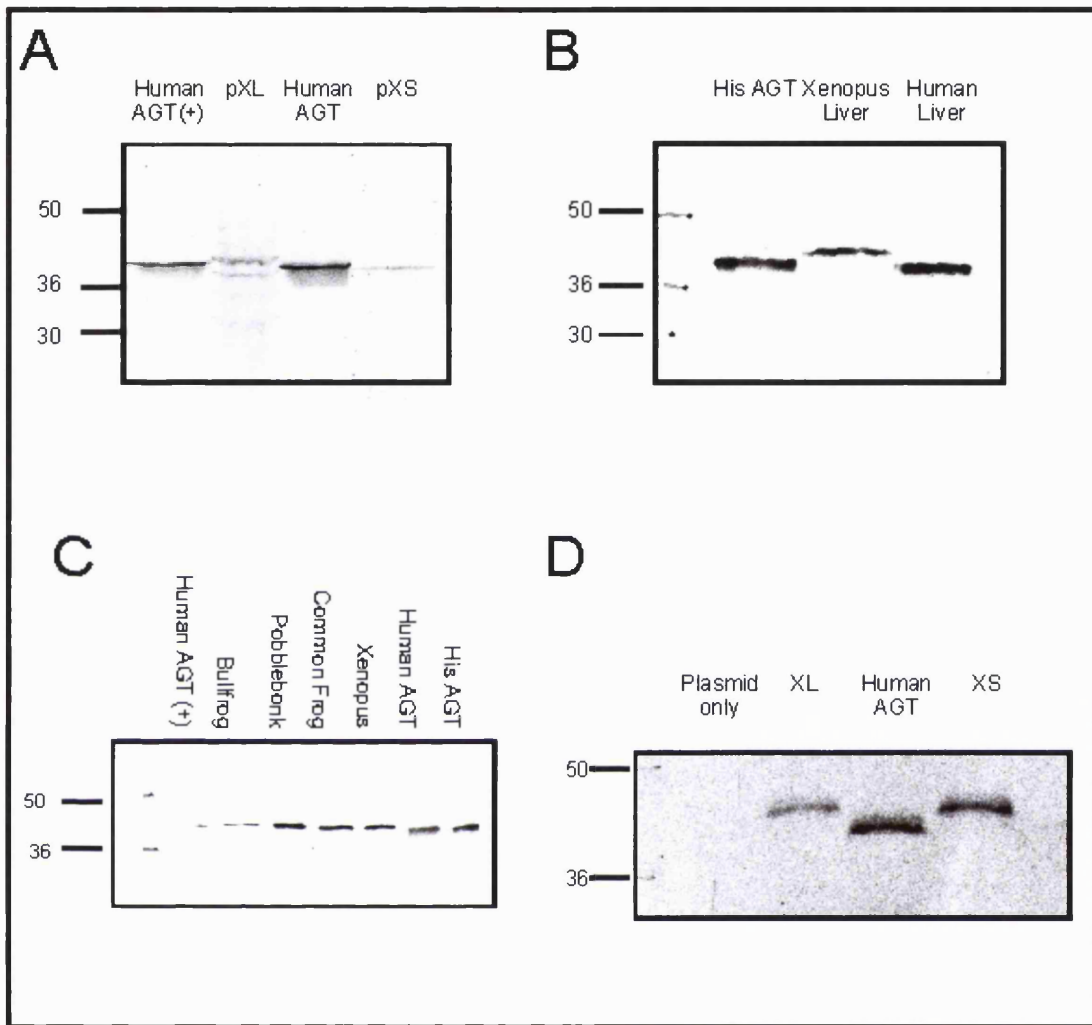


Figure 27 - Characteristics of xenopus clones and proteins

Legend on next page

Figure 27 - Characteristics of xenopus clones and proteins

A - *In vitro* transcription and translation of xenopus and human AGT clones. The clones were transcribed from the T7 promoter of pcDNA3. Translation of p*HsAGT*⁺ shows one main band that corresponds to *HsAGT* with a 22 amino acid MTS. pXL translates into two major protein products, the larger band is larger than *HsAGT*⁺ and the smaller band is smaller than *HsAGT*. pXL shows some evidence of internal translation or degradation resulting in smaller bands. The translation of pXS yields one protein product, which is smaller than *HsAGT*.

B - Immunoblot of human and xenopus livers with rabbit anti human AGT antibody, compared with pure AGT protein with a HIS tag. Both livers reveal one band but the xenopus liver AGT is bigger than the 43kD human liver AGT and His tagged AGT protein (44 kD).

C - Immunoblot of amphibian and human livers with rabbit anti human AGT antibody compared to purified *HsAGT*⁺ and His tagged AGT protein. The amphibian livers reveal AGT bands of the same size as that of the xenopus and bigger than that of the human and the purified *HsAGT*⁺ and His tagged AGT protein.

D - Immunoblot of COS cells transfected with xenopus and human clones, with rabbit anti human AGT antibody including a negative control of COS cells transfected with expression plasmid only. Both the pXL and pXS transfected cells show one band of the same size in both. The plasmid only lane was blank.

smaller of the two bands was smaller than *HsAGT*. There were other smaller bands, which were thought to be internal translation or degradation products of the larger band. Translation of pXS, however, resulted in one band, which was smaller than *HsAGT*. The computer program Peplot predicted molecular weights of 42.5 kD and 45 kD for the pXS and pXL sequences, respectively. Therefore, the sizes of the *in vitro* translated XL and XS products roughly match their predicted sizes.

4.2.4 Immunoblotting of amphibian liver

The livers of amphibian species: xenopus, common frog, bullfrog and pobblebonk, were immunoblotted, along with human liver, with the anti-human AGT antibody (figure 27B and 27C). AGT protein in the amphibian livers appeared larger than human liver AGT protein, their sizes were estimated as ~ 2 kD larger than *HsAGT*. The reason for this size difference was not clear.

4.2.5 Immunoblotting of COS Cells Transfected with AGT Clones

To investigate whether it was the transcripts corresponding to pXL or pXS that resulted in the larger size of xenopus AGT compared with *HsAGT*, COS cells were transfected with the xenopus and human AGT clones, harvested and lysed (see Chapter 2, section 2.6.16). The crude homogenates were run on a SDS PAGE gel and immunoblotted with anti human AGT antibody. This method revealed a single band for both pXL and pXS that seemed to be of identical size and was estimated to be ~ 45 kD in size. Both are larger than the human AGT band (figure 27D).

Table 8 - AGT activities of transfected COS cells

| Construct in transfected cells | AGT activity $\mu\text{mol/ hour / mg}$ cellular protein |
|---------------------------------------|--|
| pXL | 71 |
| pXS | 51 |
| pHsAGT | 10 |
| Plasmid only control | 0 |

| Livers | |
|---------------|----|
| Xenopus liver | 53 |
| Human liver | 11 |

4.2.6 AGT enzymatic activity in transfected COS cells

The pXL, pXS and p*Hs*AGT constructs were expressed in COS cells. The lysates were assayed for AGT enzymatic activity along with a xenopus liver homogenate. The activities were compared with cells transfected with a plasmid only control and the previously known activity of AGT in human liver (Brink, personal communication) (table 8 - page 111). As shown, the proteins encoded by both pXL and pXS are catalytically active. The xenopus AGT protein seems to be 5-7 times more active or more plentiful than human AGT in both transfected COS cells and liver.

4.2.7 RNase protection

To ascertain if transcripts corresponding to the clone pXL and pXS were expressed in xenopus liver, RNase protection was carried out using xenopus liver total RNA. The 5' end of pXL was used to form a new clone (pXL-5', see figure 8 - page 46) which included both translation start sites and sequence extending 5' to sequence encoding stop codons (figure 8). RNase protection revealed two kinds of protected fragments of 259 and 209 nt (figure 28 - page 114) corresponding to a transcript 39 bp downstream from the 5' end of XL and to a transcript 89 bp downstream from the 5' end of XL. This result suggests there are two types of mRNA in xenopus liver, the 5' end of the longer form maps upstream of TL-1 while the 5' of the shorter form maps between TL-1 and TL-2.

4.2.8 Immunofluorescence of xenopus AGT in COS cells

As it was known that transcripts corresponding to both pXL and pXS were expressed in xenopus liver, an attempt was made to determine if the clones pXL and pXS could account for the mitochondrial and cytosolic distribution of AGT in xenopus liver. The clones were transiently expressed in COS cells and the subcellular distribution of AGT was visualised by confocal immunofluorescent microscopy. To help identify the subcellular compartments, the cells were either double labelled for endogenous catalase, or stained with MitoTracker, a vital dye that identifies mitochondria. As a control p*Hs*AGT and p*HS*AGT⁺, clones known to encode peroxisomal and mitochondrial products respectively, were studied. When p*Hs*AGT was expressed in COS cells (figure 29) AGT had a punctate distribution that co-localised with catalase. This is compatible with previous studies that have found *Hs*AGT to be a peroxisomal protein^{48;60}.

When p*Hs*AGT⁺ was expressed in COS cells (figure 30), AGT was distributed in elongated structures that co-localised, almost completely with MitoTracker. This is compatible with previous studies that have found *Hs*AGT⁺ to be mitochondrial⁴⁴. Therefore, from the use of these control constructs it is clear that mitochondrial and peroxisomal AGT can be clearly distinguished.

When pXL was expressed in COS cells (figures 31 and 32- pages 119 and 120), AGT distributed in elongated structures that colocalised almost completely with MitoTracker, no colocalisation with catalase was discernible. When pXS was expressed in COS cells (figures 33 and 34) AGT had a diffuse distribution that did not co-localise with either catalase or MitoTracker.

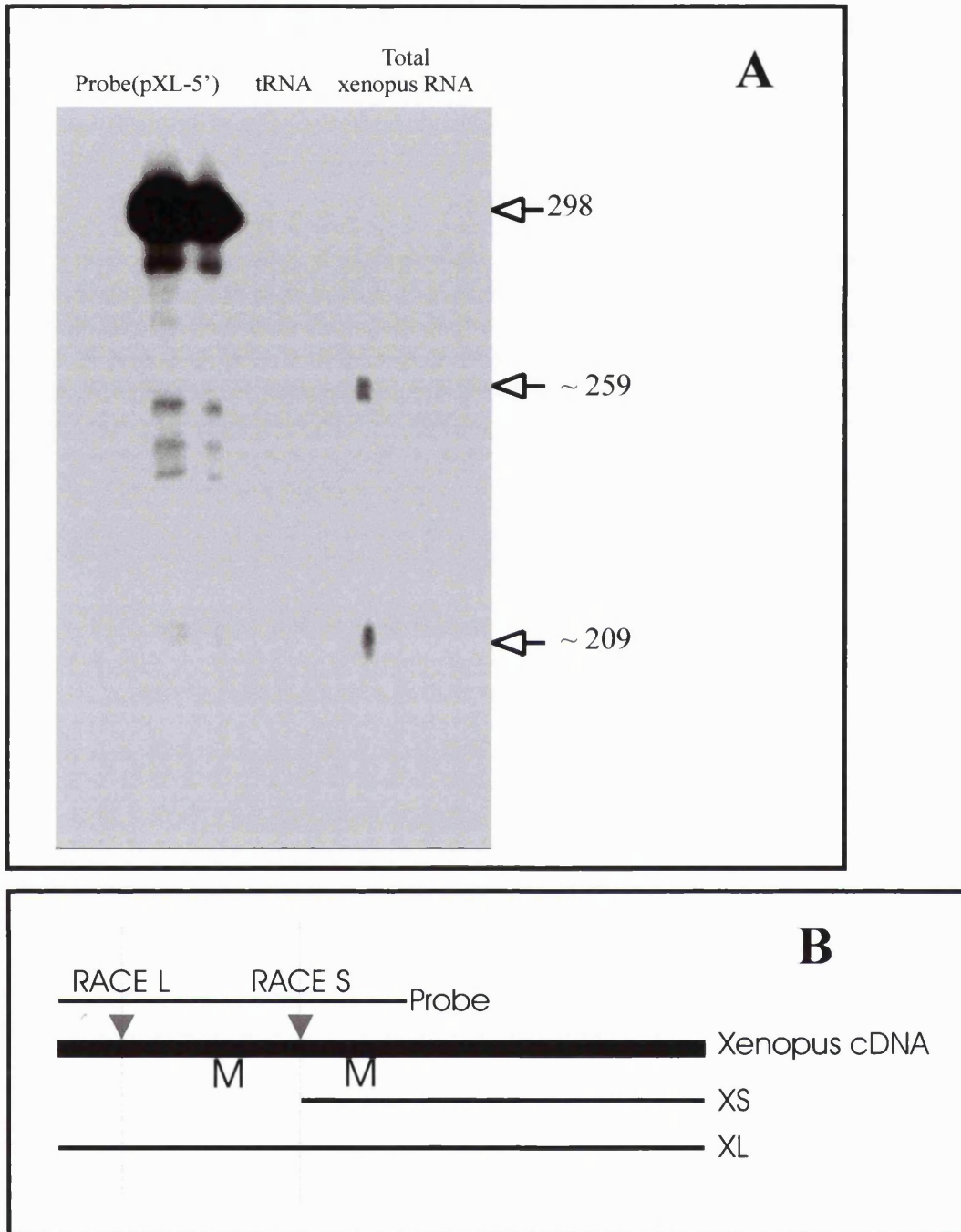


Figure 28 - RNase protection of the xenopus AGT gene

Two RNA species hybridised to the antisense riboprobe (A). The larger was estimated to be 39 nt smaller than the probe and the smaller was estimated to be 99 nt smaller than the probe. This suggests there are two transcripts, one which includes the MTS and one that does not (B). Some internal transcription of the probe was noted but the intensity of these bands was at least 20 times less than the full-length probe.

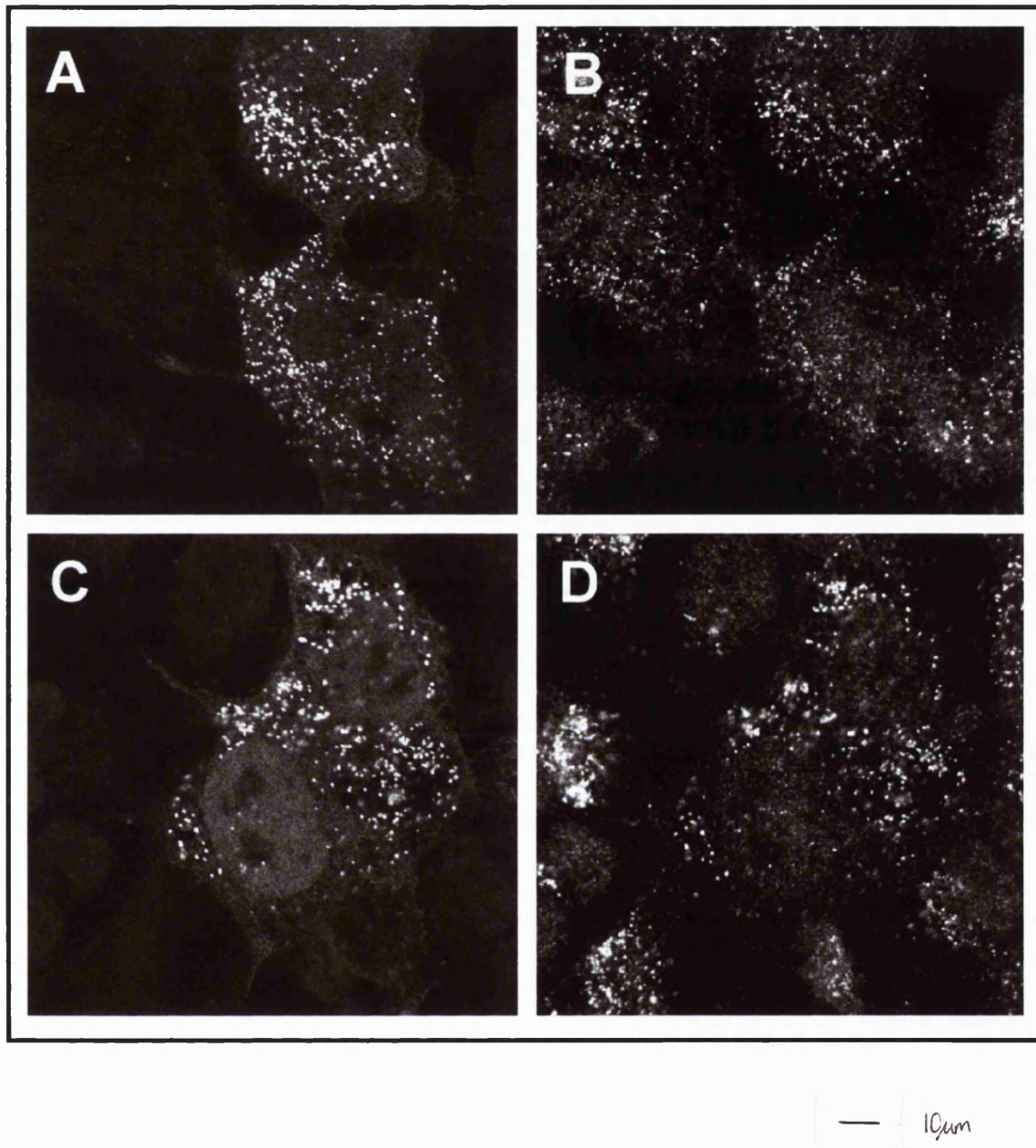
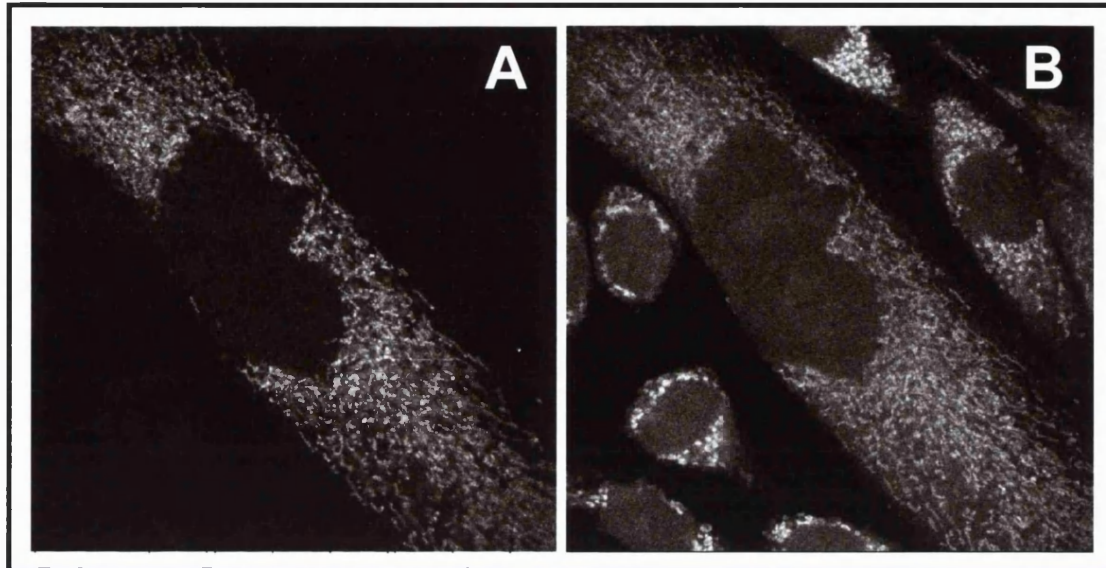


Figure 29 – AGT distribution in COS cells following transfection with *pHsAGT*

COS cells were transiently transfected with *pHsAGT*. Cells were double labelled for AGT (A and C) and catalase (B and D). AGT labelling was punctate and co-localised with endogenous catalase



— 10µm

Figure 30 – AGT distribution in COS cells following transfection with $pHsAGT^+$
COS cells were transiently transfected with $pHsAGT^+$. Cells were labelled for AGT (A) and stained with MitoTracker (B). AGT labelling co-localised with MitoTracker.

4.2.9 The subcellular distribution of proteins encoded by various artificial xenopus AGT constructs

To test if the polypeptide encoded by the sequence between the two in frame translation start sites of xenopus AGT could act as an MTS, it was fused to green fluorescent protein (GFP), to result in pXmts-GFP (Chapter 2, Section 2.3.2). This clone was expressed in COS cells and the protein encoded by the clone distributed in elongated structures that colocalised with MitoTracker (figure 35).

In order to determine if xenopus AGT had a PTS1, p*Hs*AGT was mutagenised to encode the xenopus AGT terminal tripeptide, KKM (Chapter 2, Section 2.3.2) in place of its original C-terminal tripeptide, and named p*Hs*AGT-KKM. When p*Hs*AGT-KKM was expressed in COS cells, it encoded a protein with a subcellular distribution that was diffuse and did not co localise with catalase (figure 36).

In order to determine if xenopus AGT was capable of being directed to the peroxisomes by a PTS1, pXS was mutagenised to encode the PTS1 tripeptide, SKL, at its 3' end by PCR (Chapter 2, Section 2.3.2), in place of its original C-terminal tripeptide, and named pXS-SKL. The clone was expressed in COS cells and the protein it encoded had a punctate subcellular distribution and co-localised with the catalase (figure 37).

In order to determine if xenopus AGT could be targeted to the peroxisomes by a context dependent PTS1 like human AGT, pXS was also mutagenised to encode the human terminal tripeptide, KKL, in place of its original C-terminal tripeptide, at its 3' end (Chapter 2, Section 2.3.2), named pXS-KKL. The clone was expressed in COS cells, the protein it encoded had a diffuse subcellular distribution that did not colocalise with either catalase or MitoTracker (figure 38).

In order to determine if KKL prevented the proper folding of xenopus AGT, pXL was mutagenised to encode the human terminal tripeptide, KKL, in place of its original C-terminal tripeptide, at its 3' end, (Chapter 2 and Section 2.3.2), named pXL-KKL. The clone was expressed in COS cells, the protein it encoded distributed in elongated structures, that colocalised with MitoTracker (figure 39).

A summary to the targeting results is shown in table 9 (page 132). Composite color images of some of the immunofluorescence results are presented in the Appendix (figure A5 - page 210). These were performed when it was thought to be useful to further examine the co-localisation of AGT and peroxisomal markers.

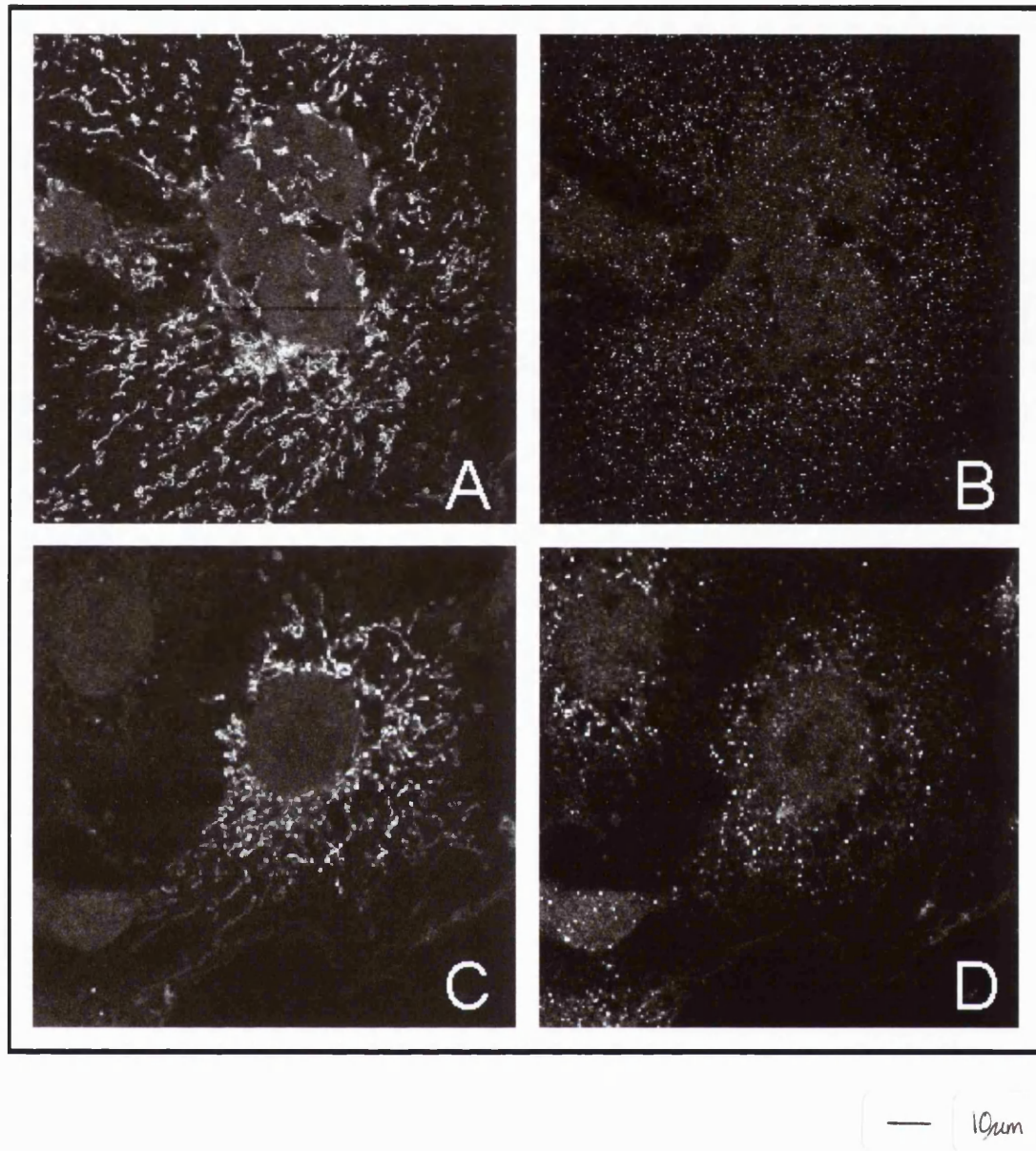
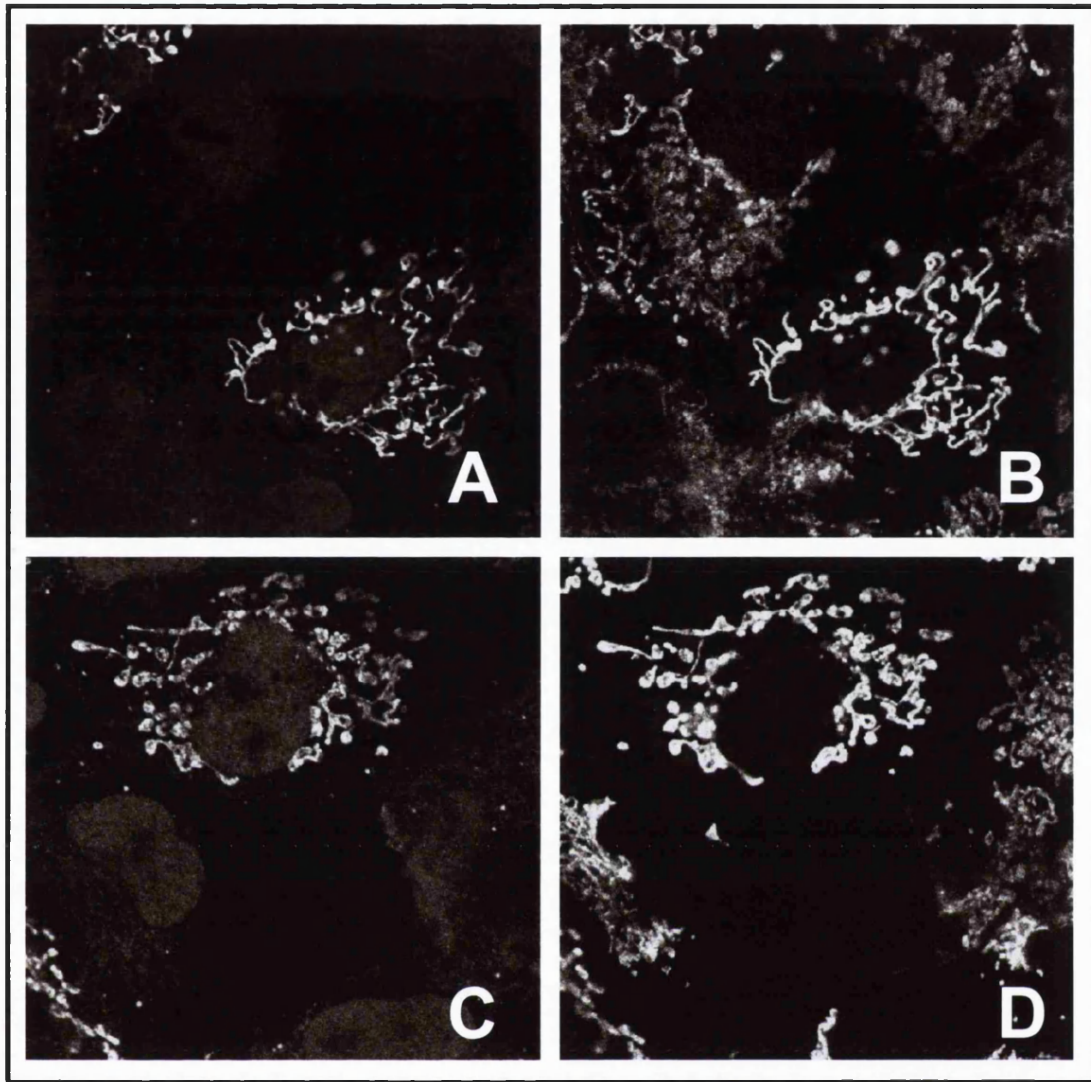


Figure 31 – AGT distribution in COS cells following transfection with pXL (1)

COS cells were transiently transfected with pXL. Cells were double labelled for AGT (A and C) and catalase (B and D). AGT labelling did not co-localise with catalase.



— 10 μ m

Figure 32 – AGT distribution in COS cells following transfection with pXL (2)

COS cells were transiently transfected with pXL. Cells were labelled for AGT (A and C) and stained with MitoTracker (B and D). AGT labelling colocalised with MitoTracker.

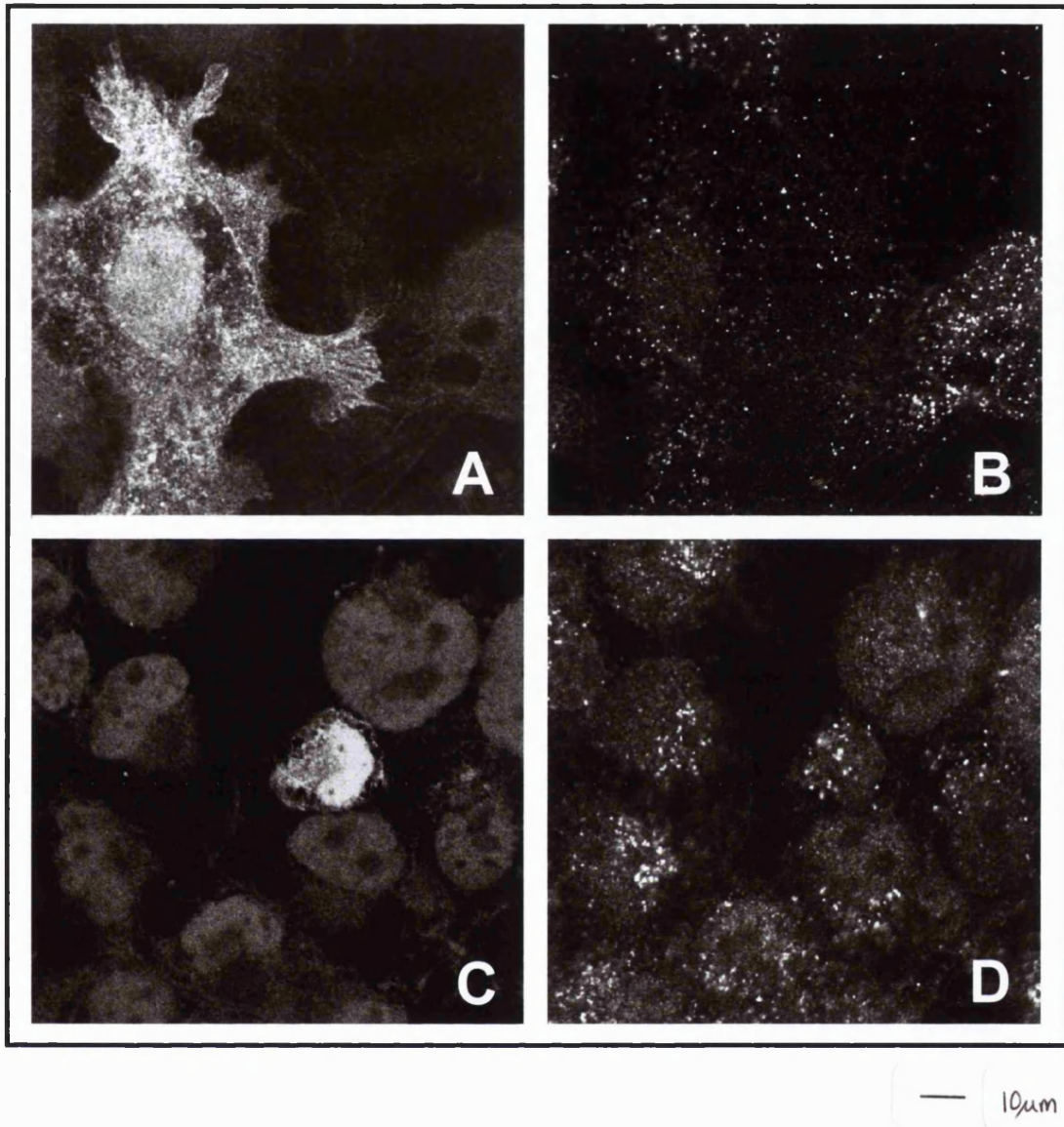


Figure 33 – AGT distribution in COS cells following transfection with pXS (1)

COS cells were transiently transfected with pXS. Cells were double labelled for AGT (A and C) and catalase (B and D). AGT labelling was diffuse and did not co-localise with catalase.

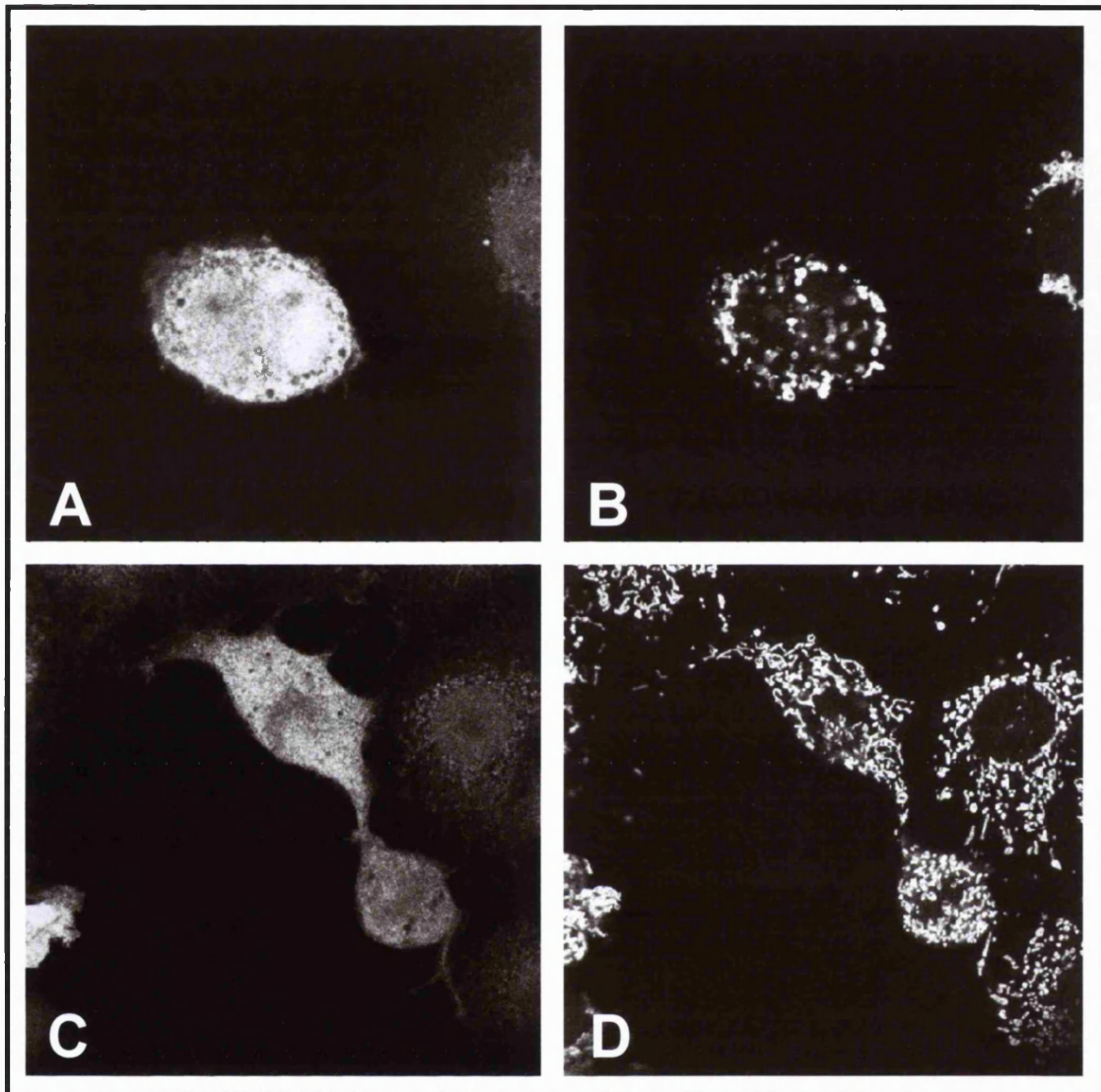
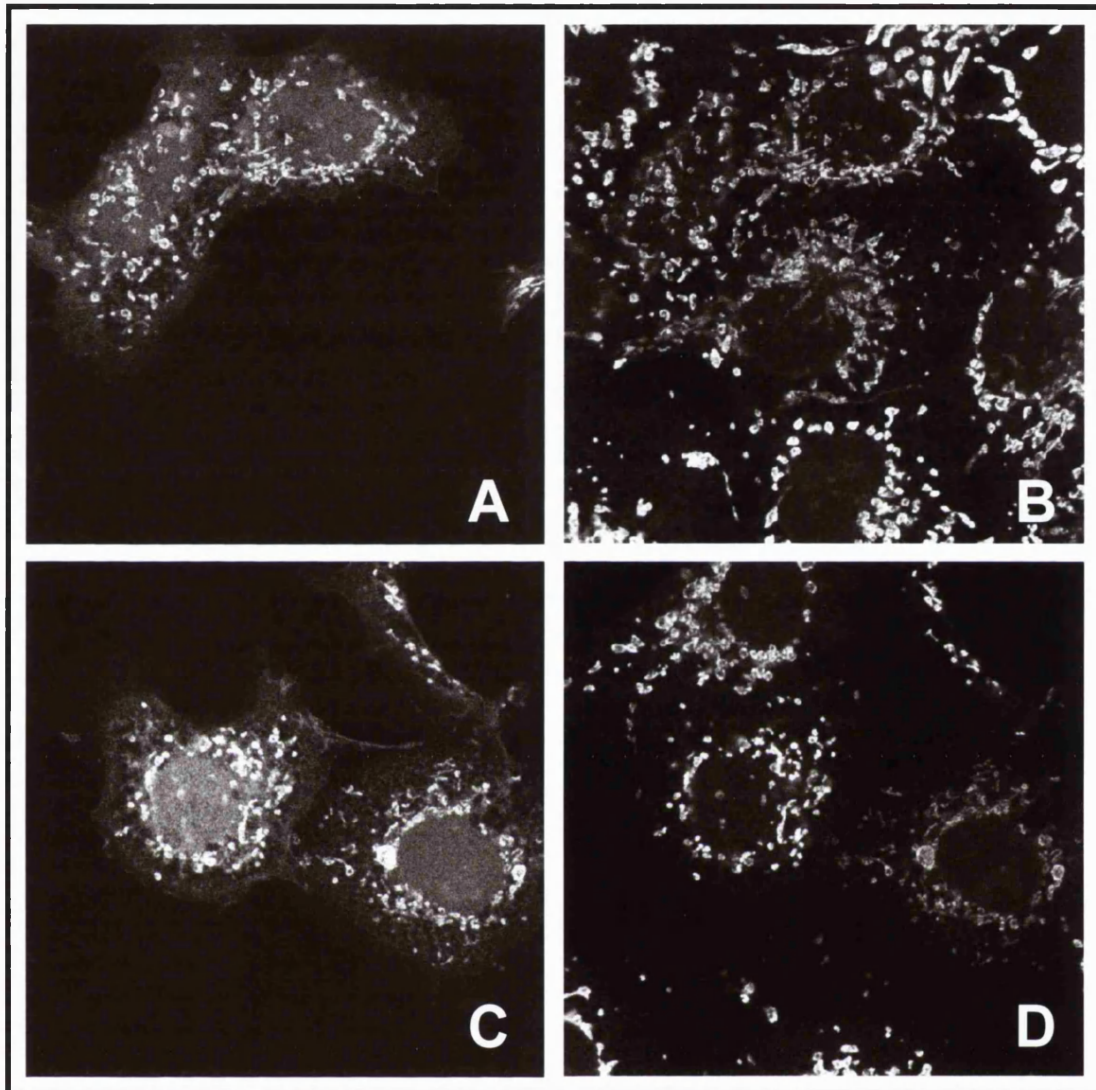


Figure 34 – AGT distribution in COS cells following transfection with pXS (2)

COS cells were transiently transfected with pXS. Cells were labelled for AGT (A and C) and stained with MitoTracker (B and D). AGT labelling was diffuse and did not co-localise with MitoTracker.



— 10 μm

Figure 35 – GFP distribution in COS cells following transfection with pXmtsGFP

COS cells were transiently transfected with pXmtsGFP. Staining is shown for GFP (A and C) and MitoTracker (B and D). GFP staining colocalised with MitoTracker.

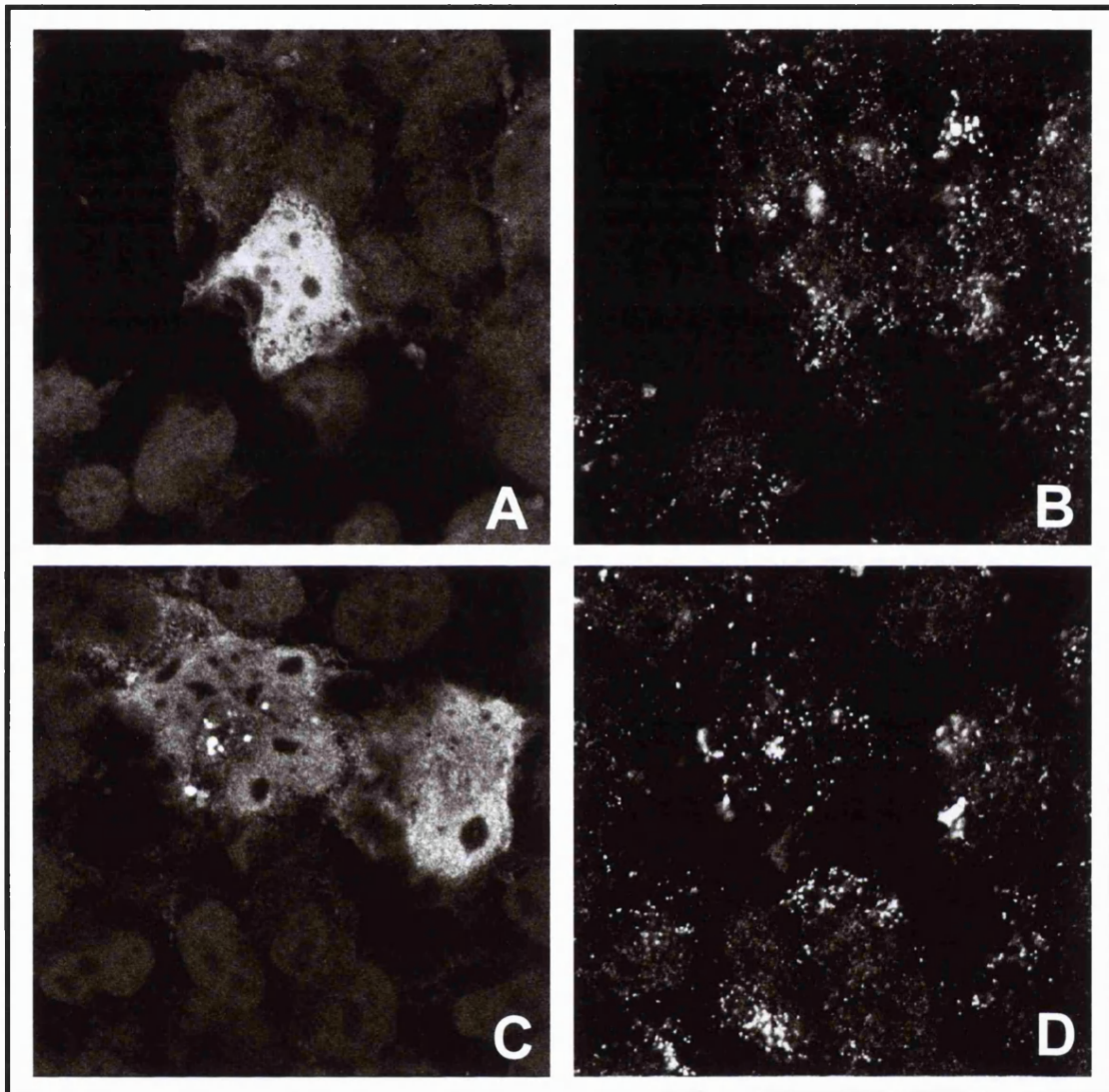


Figure 36 – AGT distribution in COS cells following transfection with *pHsAGT-KKM*

COS cells were transiently transfected with *pHsAGT-KKM*. Cells were double labelled for AGT (A and C) and catalase (B and D). AGT labelling was diffuse and did not co-localise with catalase. Some aggregation was seen (C).

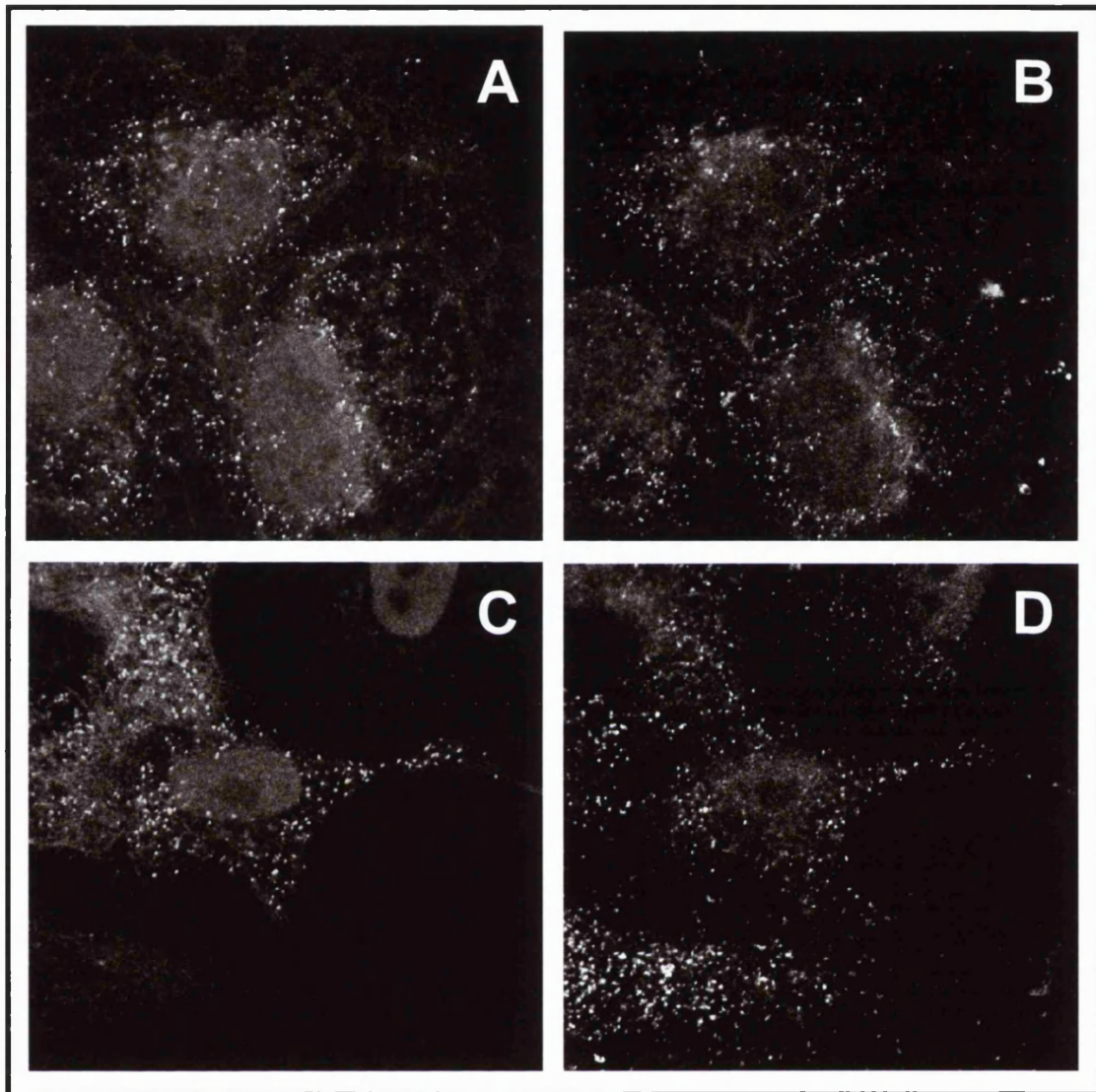
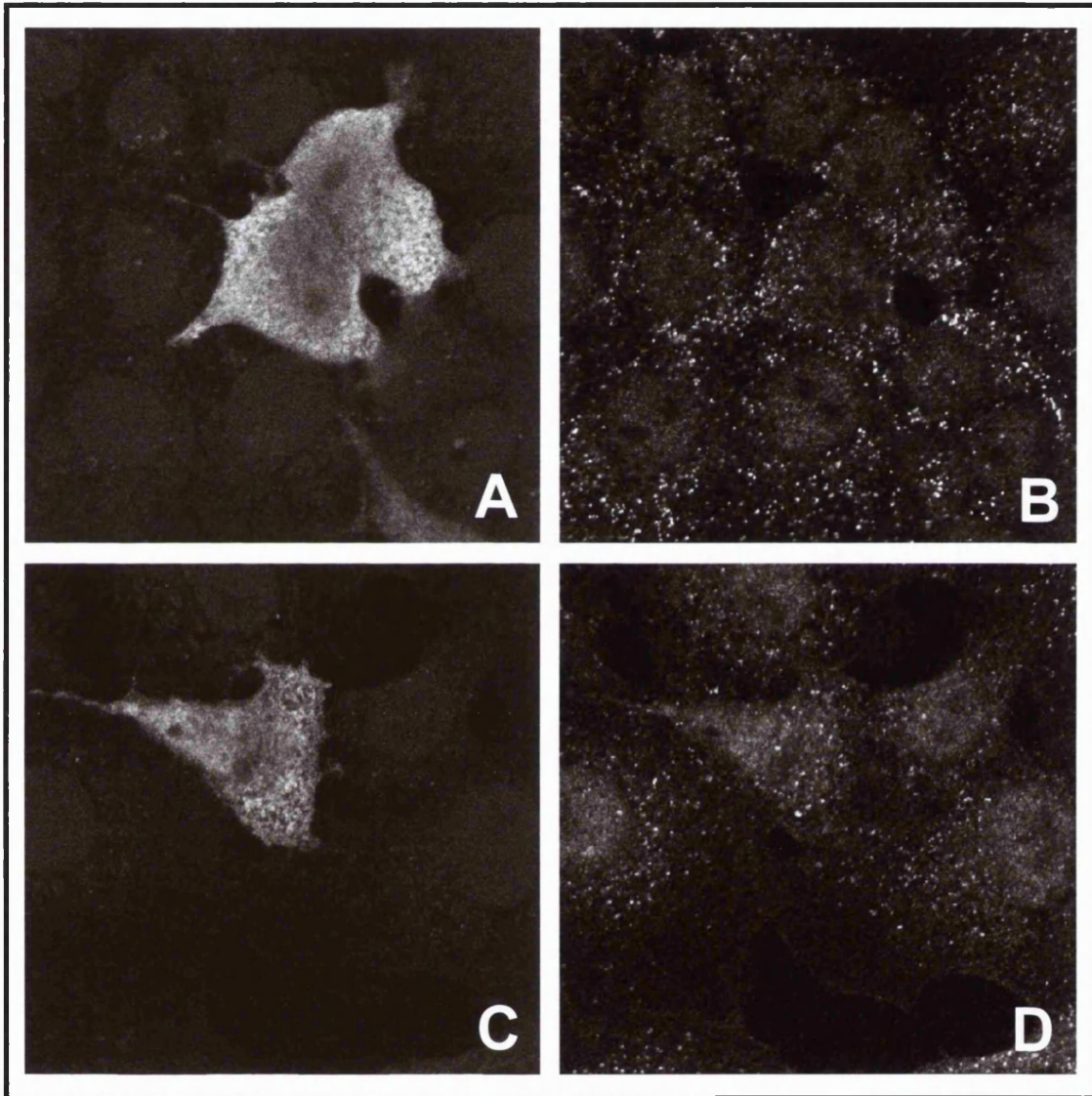


Figure 37 – AGT distribution in COS cells following transfection with pXS-SKL

COS cells were transiently transfected with pXS-SKL. Cells were double labelled for AGT (A and C) and catalase (B and D). AGT labelling was punctate and co-localised with catalase.



— 10 μ m

Figure 38 – AGT distribution in COS cells following transfection with pXS-KKL

COS cells were transiently transfected with pXS-KKL. Cells were double labelled for AGT (A and C) and catalase (B and D). AGT labelling was diffuse and did not colocalise with catalase

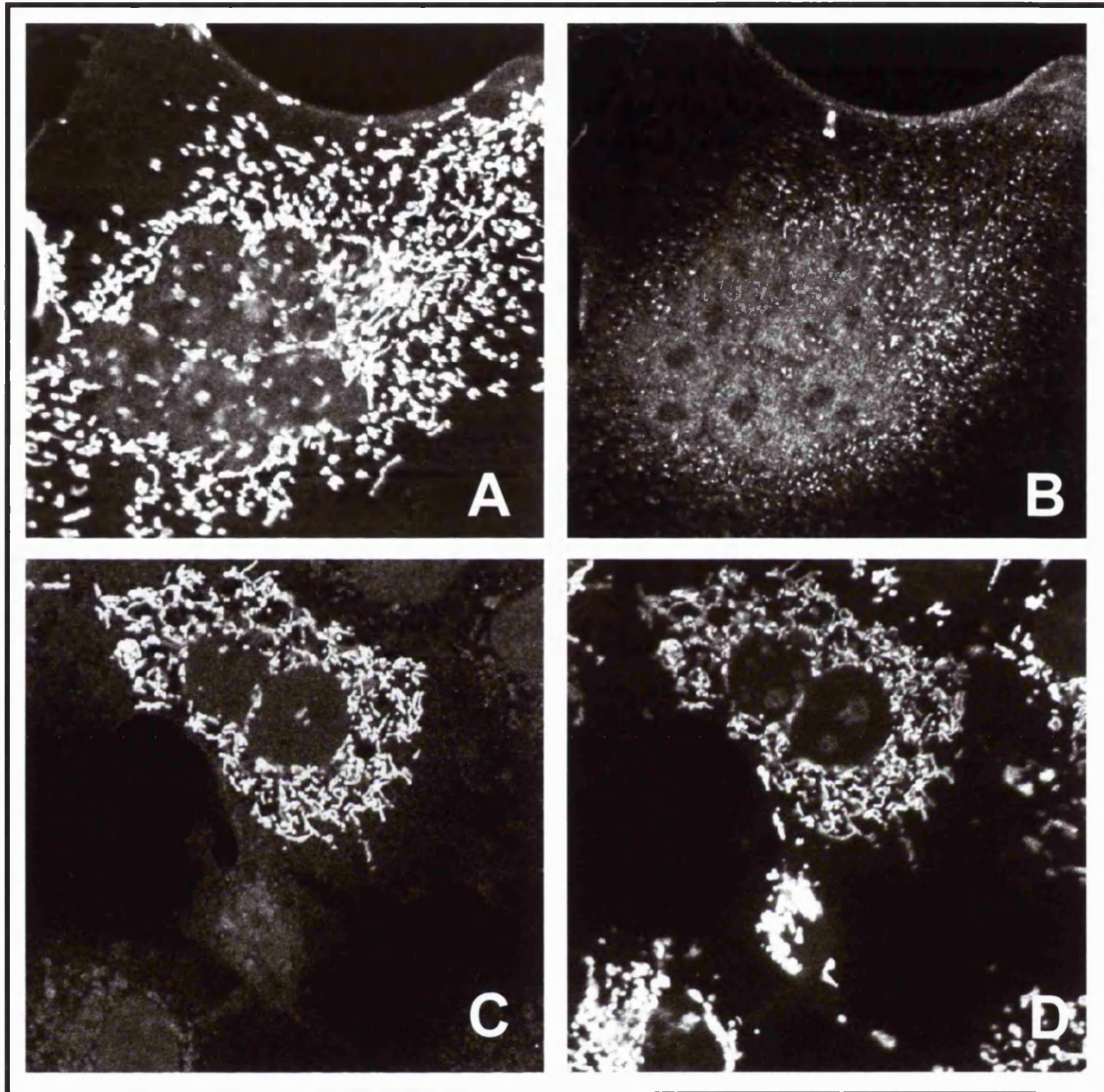


Figure 39 – AGT distribution in COS cells following transfection with pXL-KKL

COS cells were transiently transfected with pXL-KKL. Cells were labelled for AGT (A and C) and stained with MitoTracker (B and D). AGT labelling co-localised with MitoTracker.

4.3 Discussion

4.3.1 The unexpected orientation of pXS may be due to recombination

The presumed sense strands (i.e. the orientation that produced a highly significant match in BLAST searching and encoded no stop codons) of the other cDNAs rescued from the library (table 7) ran from 5' at T3 and 3' at T7, as expected (figure 22 - page 99). As mentioned previously this was not the case with pXS, it seems to be only pXS that is anomalous. PCR and Southern blot analysis of the library (figure 23) shows that the switch in orientation did not occur during rescue of the cDNA. Therefore it must have been in this orientation within the library. When the full length clone was obtained by RT-PCR it was found to recombine in *E. coli* cells, in fact, Xall could not be cloned due to its incompatibility with even strains that have been mutagenised in many recombination pathways (Chapter 2). Therefore it seems likely that the library clone recombined, to change its orientation relative to the promoter sequences T3 and T7. As it was still possible to identify the xenopus AGT expressing plaque with an anti-human AGT antibody, the sense strand of cDNA must still be under the control of the Lac Z promoter. Therefore the recombination event must have rearranged the promoter sites as well as the cDNA insert to produce a conformation compatible with transcription activation of a sense strand RNA from LacZ.

Only the smaller cDNA (pXS, see figure 8) was represented within the library (figure 23). It could be that the library originally contained the longer transcript but the insert recombined to lose its 5' end, or both cDNAs were represented but the highly recombinogenic larger cDNA was rearranged in such a way as to prevent its expression from LacZ. It should be noted that, pXS has not been seen to be recombinogenic so the

theory that the pXL recombined in the library to form pXS may be more probable.

Nevertheless pXS represents a genuine transcript as shown by the reverse transcription of an identical cDNA from xenopus RNA. The 5' end of the xenopus AGT gene includes four regions of A and G nucleotides (figure 25), these regions may bind and cause mispairing of the cDNA strands and recombination.

4.3.2 pXall contains the entire reading frame of the xenopus AGT gene

Sequence of the xenopus AGT gene was determined by combining the overlapping sequences of pXS and the longest 5' RACE product (pRACE-L). To confirm that the sequence was a true representation an existing transcript it was re-amplified from total xenopus liver RNA, by RT-PCR and sequenced. The sequences were compared and were identical to each other. The known sequence encodes stop codons at its 5' and 3' ends and is therefore thought to be full length. When translated *in vitro* both pXS and pXL encode proteins of approximately the expected size (figure 27A) and in addition, when expressed in COS cells encode proteins that have been shown to be immunoreactive (figure 27D) and enzymatically active (table 8).

There is no direct evidence that xenopus AGT is encoded by a single gene. However, indirect evidence is provided by sequencing of Xall directly from the RT-PCR product, this yielded “clean” sequence with no conflicting bands which would suggest there is no similar transcripts present in xenopus total RNA, also all clones, long and short, had identical overlapping sequence. Only one band was present on the immunoblot of xenopus liver (figure 27B and C). Mammalian homologues of xenopus AGT that have been studied, i.e. human, marmoset, rabbit, rat, guinea pig and cat AGT, have been shown to be the only AGT gene in those genomes^{47-49;59}. Also the *Drosophila*

melanogaster AGT homologue (Accession number: CAA58024) is the only closely homologous gene identified from BLAST searching the whole *D. melanogaster* genome (this study, Chapter 5).

The xenopus AGT peptide sequence seems to have two in frame potential translation start sites at the N terminus of the polypeptide (figure 25). One clearly corresponds to TL-2 present in the archetypal AGT gene (figure 4), but the other is two amino acids N-terminal to the corresponding TL-1 in the mammalian homologues. Therefore, the sequence between TL-2 and TL-1 is two amino acids longer (24 instead of 22 residues) than the sequence length between the translation start sites in mammals. This sequence has similarities to a MTS. There is an arginine two residues N-terminal to TL-2, there are three arginine residues in the peptide sequence but no negatively charged residues. That it can function as a MTS, is shown by the fact that when fused to the N terminus of GFP it targets GFP to the mitochondria (figure 35). The C-terminus of the xenopus peptide sequence consists of the tripeptide KKM. This is similar to the human PTS1 KKL. However, it does not function as a PTS1 on the human AGT sequence (figure 36) as the *pHsAGT-KKM* encoded protein is cytosolic in COS cells.

When human and xenopus AGT are aligned they show a high level of identity. The region in exon 4 (amino acids 176-197 in figure 26) which has been noted previously to have high sequence conservation (figure 17 and Birdsey and Danpure (1998)⁴⁹) is almost totally conserved between xenopus and human (figure 26). One region which does not show conservation between human and xenopus AGT is the 5' end. The putative MTS of xenopus AGT bears little resemblance to the MTSs of the mitochondrial AGT genes that have been cloned, so far

in mammals. Two opposing theories could explain this: either the xenopus and mammalian AGT MTSs have a common ancestor and have diverged due to lax and /or episodic constraints on MTS sequence evolution; or mitochondrial targeting (and hence the MTS) of AGT has been acquired separately in mammals and amphibians. These have consequences for the likely ancestral distribution of AGT protein, this is discussed fully in chapter 5.

4.3.3 The molecular explanation for the mitochondrial and cytosolic distribution of AGT in xenopus liver

The xenopus AGT gene has the potential to encode two proteins. Transcript analysis showed that there are two major RNA transcripts (figure 28). Surprisingly, a larger band of the same size as the probe (pXL-5') was not detected using this methodology, although it is known that there must be a RNA species equivalent to the size of the probe as the RACE-L product extends further 5' than the probe. However as both the predicted transcript size from the RNase protection (figure 28) and the larger transcript size from the 5' RACE (figure 24) start 5' of TL-1 and a stop codon, they would both code for the same protein. As TL-1 is included in the transcript, the MTS would be included within the ORF. pXL includes TL-1 and when expressed in COS cells is mitochondrial (figures 31 and 32 and table 9).

The smaller transcript size visualised on figure 28, corresponds exactly to the smaller product of the RACE reaction (figure 24) and the sequence of the library clone, pXS. This transcript size does not include TL-1 and consequently the MTS would be excluded from the ORF. The C-terminal tripeptide (KKM) does not appear to be a PTS1

Table 9 - Subcellular distribution of clones in COS cells, as determined by immunofluorescent microscopy

| Construct | Subcellular distribution in COS cells |
|---------------------|---------------------------------------|
| pXL | Mitochondrial |
| pXS | Cytosolic |
| GFP | Cytosolic |
| pXmts-GFP | Mitochondrial |
| pXS-KKL | Cytosolic |
| pXL-KKL | Mitochondrial |
| pXS-SKL | Peroxisomal |
| p <i>Hs</i> AGT-KKM | Cytosolic |
| p <i>Hs</i> AGT | Peroxisomal |

Clones were expressed in COS cells under the control of the CMV promoter of pcDNA3. The C-terminal tripeptide of *Hs*AGT is KKL whilst in xenopus AGT it is KKM.

Table 10 - Estimated sizes of xenopus AGT proteins

| Method | XL(kD) | XS (kD) | HsAGT (kD) |
|--|--------|---------|------------|
| Predicted molecular weight | 45 | 42.5 | 43 |
| <i>In vitro</i> translation | > 43 > | < 43 | 43 |
| Immunoblot of liver | > 43 | > 43 | 43 |
| Immunoblot of transfected COS cells | > 43 | > 43 | 43 |

and the distribution of pXS in COS cells is cytosolic (figures 33 and 34 and table 9).

Therefore, the xenopus AGT gene (figure 40) has two transcription and two translation start sites and so is similar to that of the rat and marmoset AGT (figure 4, table 1), except that xenopus AGT does not have a C-terminal PTS1 and the sequence encoding the MTS is 6 nucleotides (2 amino acids) longer than that of mammalian AGTs.

4.3.4 The size of xenopus AGT protein is larger *in vivo* than *in vitro*

When pXL and pXS were *in vitro* transcribed and translated (figure 27A and table 10) pXS yields one band on a SDS PAGE gel, which is slightly smaller than human AGT. pXL translation yields two major bands, one larger and one smaller than human AGT, the larger pXL band is also larger than p Hs AGT⁺ which includes the “ancestral” MTS,. Therefore the *in vitro* transcription and translation of pXL and pXS yielded proteins of approximately the expected size as predicted from their amino acid sequence (table 10). However when xenopus liver was immunoblotted, one band was seen which was larger than AGT in human liver (figure 27B and table 10). Presumably, there was one band because the MTS of the XL protein is cleaved resulting in a product the same size as the XS protein. The larger than human size of both xenopus AGT proteins was unexpected. The liver homogenates were prepared by boiling in β -mecaptoethanol and run on a denaturing gel in the presence of SDS. Therefore differential folding or the ionic binding of additional components should not affect the running properties. When pXL and pXS were expressed in COS cells and the homogenates immunoblotted (figure 27D and table 10) the proteins resulting from XL and XS expression were the same size. This is stronger evidence that the MTS of XL protein is cleaved. However, there

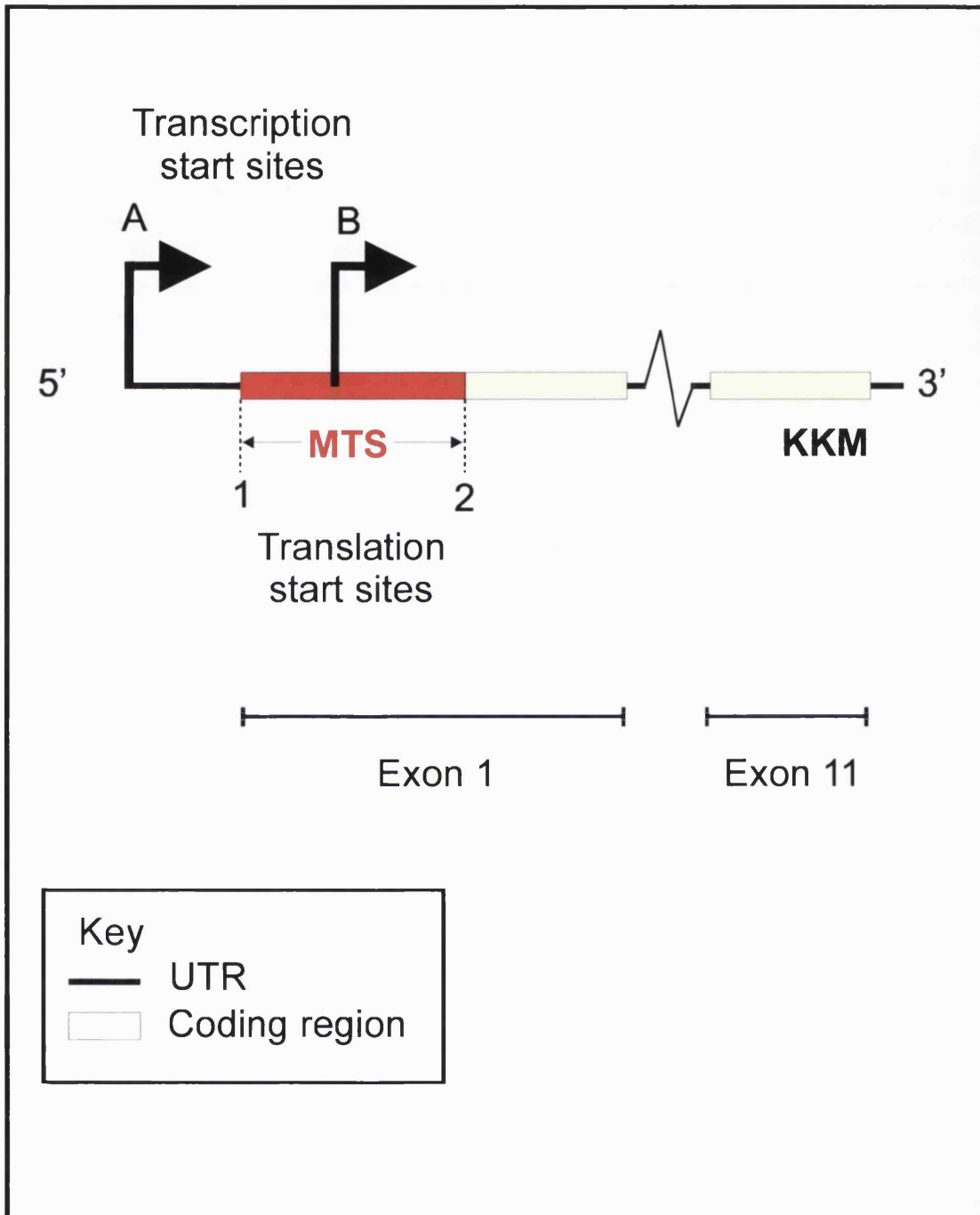


Figure 40 - Schematic diagram of xenopus AGT gene

The xenopus AGT gene encodes two RNA transcripts one of which encodes a mitochondrial protein and one of which encodes a cytosolic protein. The mitochondrial product is transcribed from TS-A and translated from TL-1; the cytosolic product is transcribed from TS-B and translated from TL-2. The main difference between the archetypal mammalian AGT gene and the xenopus AGT gene is the presence of 3' sequence encoding a PTS1 in the former and absence of such sequence in the latter.

is still a size difference between the xenopus AGT proteins and the human AGT proteins. There seems to be an increase in the size of xenopus AGT protein *in vivo*, but not *in vitro* (table 10). The most likely explanation is post-translational modification. There are many possibilities for ways in which xenopus AGT could have been post-translationally modified. Two candidates that would yield the observed size difference on a SDS PAGE gel, (approximately 2-3 kD) are phosphorylation and glycosylation. Although, phosphorylation only changes the size of a protein by 80 Da per phosphate group¹⁰³ it does alter the migration pattern of the phosphorylated protein¹⁰⁴. Glycosylation of proteins changes their size by 892-2770 Da per carbohydrate side chain¹⁰³, which would fit the observed pattern. The oligosaccharides attached to glycoproteins may assist their proper folding and help protect the mature proteins from proteolysis¹⁰⁵. A suggestion that the xenopus AGT protein may need protection from proteolysis comes from the work of Brink and Danpure (manuscript in preparation) which shows that human AGT protein lacking a PTS1 is less stable in the cytosol than wild type human AGT protein. The human AGT protein contains one potential glycosylation site (figure 26) which has been shown not to be glycosylated¹⁰⁶. The newly elucidated xenopus AGT peptide sequence contains two potential glycosylation sites (figure 26). However, protein glycosylation reactions occur either in the lumen of the endoplasmic reticulum (ER) or the lumina of the Golgi¹⁰⁵. There is no evidence for the trafficking of xenopus AGT through these compartments. Therefore of the possible types of post translational modification that present themselves, none appear likely.

Immunoblots of the AGT in the cat, rat, marmoset, rabbit and guinea pig livers reveal AGT of the same size as human AGT⁴³ as does the immuno-blot of baboon liver (figure

14). Therefore, there is no evidence of post translational modification in the proteins encoded by the other AGT genes that have been cloned.

The estimated size for the immunoblotted XL protein, in transfected COS cells, is the same as the estimated size for the larger band of the pXL *in vitro* translation (table 10). This could suggest that the xenopus MTS is not cleaved from the XL protein, leading to its larger size *in vivo*, whilst the XS protein is post translationally modified to make it the same size as the XL protein. The XL protein would escape modification as it is imported into the mitochondria before it folds.

4.3.5 Enzymatic activity of xenopus AGT protein is higher than that of human

At face value, the xenopus AGT protein appears to be more enzymatically active than its human homologue (table 8). When xenopus AGT is expressed in COS cells, the homogenates have higher AGT activity than homogenates expressing human AGT. The activities are a rough estimates that have several sources of error. pXL, pXS and human AGT were expressed by the CMV promoter from the same expression plasmid (pcDNA3) and transfected in the same quantity of DNA. However, firstly, the transfection rates are not controlled for, so the xenopus proteins could have been transfected into more cells than the human AGT. Secondly, transient transfection tends to lead to high expression rates and the effects of toxicity were not accounted for. The amounts of cells in the transfection reactions were only roughly estimated. However the activity (per mg of cell protein) of the COS cells expressing pXL is approximately seven times greater than those expressing human AGT and this may reflect a genuine greater activity of xenopus AGT protein than human AGT protein. More confidence may be held in the comparisons of the AGT activities of human and xenopus livers, again AGT activity is 5 times greater in xenopus liver, although importantly, the level

of expression is not known. The greater activity of xenopus AGT compared to human AGT is compatible with the observed trend for AGT proteins of species with mainly mitochondrial AGT to have greater activity than AGT proteins of species that have peroxisomal AGT protein (discussed in Chapter 6). The higher xenopus AGT activity does not seem to be compartment dependent because the activity of pXS expressing cells is also high but the subcellular distribution of pXS is cytosolic. Maybe, the putative post translational modification of xenopus AGT is responsible for the higher activity. However, this is not compatible with the observation that other mitochondrial AGT proteins also have higher activity because, as mentioned above, there is no evidence of post translational modification in other AGT proteins. The levels of AGT in human and xenopus liver seem roughly equal (figure 27B), however the levels of other metabolic enzymes are not known. Maybe AGT makes more contribution to gluconeogenesis in xenopus liver cells due to lower levels of other gluconeogenic enzymes.

4.3.6 Xenopus AGT lacks the additional targeting information present in the human AGT protein

The xenopus AGT gene is the first AGT gene cloned that has no potential for peroxisomal import. Cat AGT was previously the most mitochondrial AGT studied⁵⁹ but 5 % of cat AGT protein is imported to the peroxisomes via the cat AGT PTS1 (NKL) and an artificial construct of cat AGT excluding the MTS encodes an entirely peroxisomal product when expressed in COS cells⁴⁴. However, uniquely, xenopus AGT has no functional PTS1 (figures 33, 34 and 36).

The peroxisomal targeting of human AGT is not just dependent on its C terminal PTS1⁶¹ as the C-terminal tripeptide of human AGT (KKL) is necessary but not sufficient for peroxisomal targeting⁶⁰. This has led to the suggestion that there is additional targeting information within human AGT that works with the PTS1 (KKL) to target the protein to the peroxisomes (see figure 6). As xenopus AGT does not have a PTS1, it may not include the additional targeting information putatively within human AGT. However, if the additional targeting information is an element that is also important for AGT function (i.e. catalytic activity) or folding it would be expected to be present in xenopus AGT. Also if the ancestral distribution of amphibian AGT had been peroxisomal, and peroxisomal targeting has been lost quite recently in the xenopus, the additional targeting information may still be present. If the loss was more ancient could have been lost due to lack of functional constraint after the loss of the PTS1 or the result of selection pressure to lose peroxisomal targeting.

The canonical C-terminal tripeptide SKL, is necessary and sufficient to target human AGT^{60;61} or a reporter protein such as CAT²⁴ to peroxisomes. When pXS is mutagenised to include SKL at its C-terminus (pXS-SKL), it is imported to the peroxisomes when expressed in COS cells (figure 37 and table 9) therefore xenopus AGT protein does not fold to obscure the C-terminus and is capable of peroxisomal import. When pXS is mutagenised to include the human AGT PTS1 (KKL) at its C-terminus (pXS-KKL) its distribution in COS cells is cytosolic (figure 38 and table 9).

The observation that KKL is not sufficient to target xenopus AGT protein to the peroxisomes leads to the conclusion that xenopus AGT does not include the additional targeting information that is present in human AGT. Comparison of human and xenopus AGT may yield information about the nature of the additional targeting

information. The alignment of human and xenopus AGT (figure 26- page 107) does shows a 9 amino acid region of reduced homology. If the nature of the additional targeting information were to be mapped by making fusions of the human and xenopus AGTs , this would be a good starting point. Candidates for the additional targeting information and PTS1s that are dependent on additional information are more fully discussed in chapter 5.

4.4 Summary

The mitochondrial and cytosolic distribution of AGT in xenopus liver cells is due to the use of two transcription and two translation start sites in the gene, which result in a mitochondrial and a cytosolic protein. The protein may be post translationally modified and is more active than human AGT. Xenopus AGT does not include the additional targeting information that is predicted to be present in human AGT and may be used as a tool to discover the nature of the additional targeting information.

5. Evolution of Eukaryotic AGT and its Targeting Signals

5.1 Introduction

In learning more about the molecular events that have led to the variable targeting of mammalian AGT, it would be of great use to know something about AGT in the ancestral mammal. In Chapter 3 it was suggested, that the ancestral distribution of AGT to anthropoid primates was mitochondrial and peroxisomal. This is the distribution in prosimians and it was assumed that the loss of mitochondrial targeting is more likely than *de novo* gain, as the former requires only one nucleotide substitution (in some cases) whilst the latter might require many amino acid substitutions. All three possible organellar locations for AGT (i.e. mitochondrial, peroxisomal and both mitochondrial and peroxisomal) are found in mammalian species, and following the above idea that loss is more parsimonious than gain, the hypothetical ancestral state would be AGT in possession of both targeting sequences. However, xenopus AGT is an outgroup for the mammalian AGTs, and xenopus AGT is mitochondrial and cytosolic (Chapter 4), so perhaps mitochondrial and cytosolic AGT is ancestral to mammals. This would necessitate the mammalian lineage gaining a PTS1 prior to the evolution of the AGT distributions we see in extant species.

As mentioned in Chapter 1 (Section 1.4), the subcellular distribution of most proteins does not vary. Therefore the targeting information would also be well conserved. As the subcellular distribution of AGT does vary, unusually the evolution of at least some of the targeting information is likely to be affected by episodic, positive and negative selection pressure. This is suggested by targeting sequence variability, for example the lack of similarity of the xenopus AGT MTS to the human MTS (Chapter 4 - figure 26),

brings into question their common origin and the PTS1s of mammalian AGT orthologues are often different from one another. The variation in the targeting sequences of AGT could indicate what types of changes are permissible in a MTS or PTS1 and what types abolish the targeting sequence's function.

To answer the questions raised by these speculations the variation in AGT mitochondrial and peroxisomal targeting sequences was analysed. These were compared to those of other proteins to highlight unique properties of their evolution. Information about many more extant AGT sequences was required. In order to acquire this information, a systematic search for AGT homologues (i.e. genes that share evolutionary history with human AGT) was undertaken. Only orthologues of AGT, e.g. rabbit AGT, would be expected to share the differential subcellular distribution of AGT protein and hence the unique evolutionary history of its targeting sequences. The targeting sequence of paralogues, e.g. aminotransferases other than AGT, would have completely different evolutionary histories. Therefore at high levels of sequence divergence it becomes important to discriminate between orthologues and paralogues. The working definition of orthologues and paralogues used in this study, is not consistent with the strict definition of a paralogue i.e. genes arising from a gene duplication event; the definition is extended to cover all genes that, although similar to AGT, are not the most homologous gene to AGT in a given species' genome. Therefore, methods were developed to identify what part of its evolutionary history a homologue shared with human AGT.

5.2 Results and Discussion

5.2.1 The MTS of AGT

Figure 41 shows helical analysis of the N-terminal amino acid sequences (equivalent to region 1 in Chapter 3) of the known AGT orthologues. Some of these, such as rat, marmoset and cat MTS, have been shown to act as MTSs, but some, such as rabbit, human and guinea pig, are not included in the ORF. However they all, except possibly the guinea pig, have a cluster of positively charged amino acids at one side of the helical wheel, this is suggestive of MTS functionality. The topology of the mammalian AGT MTSs varies considerably and they show no discernible topological or sequence similarity to the xenopus AGT MTS. This variability could suggest multiple origins or lax functional constraint.

In order to identify which residues in the MTS of mammalian AGT orthologues, were free to vary, the number of different amino acids found at each position in a multiple alignment of AGT MTSs, was scored (figure 42). The analysis was restricted to mammals due to the uncertainty about the relationship between the xenopus and mammalian MTSs (they have little similarity, see Chapter 4). From figure 41, it can be seen that the number of amino acids found at the residue positions varied between 4 and 1. The residue position 1 (TL-1) has changed more than all other positions and no position showed especial degeneracy, therefore it seemed that the evolution of the MTS of AGT is not confined by important site factors. The data was divided into those residues in the extant MTSs (below line) and those that are found in species with the

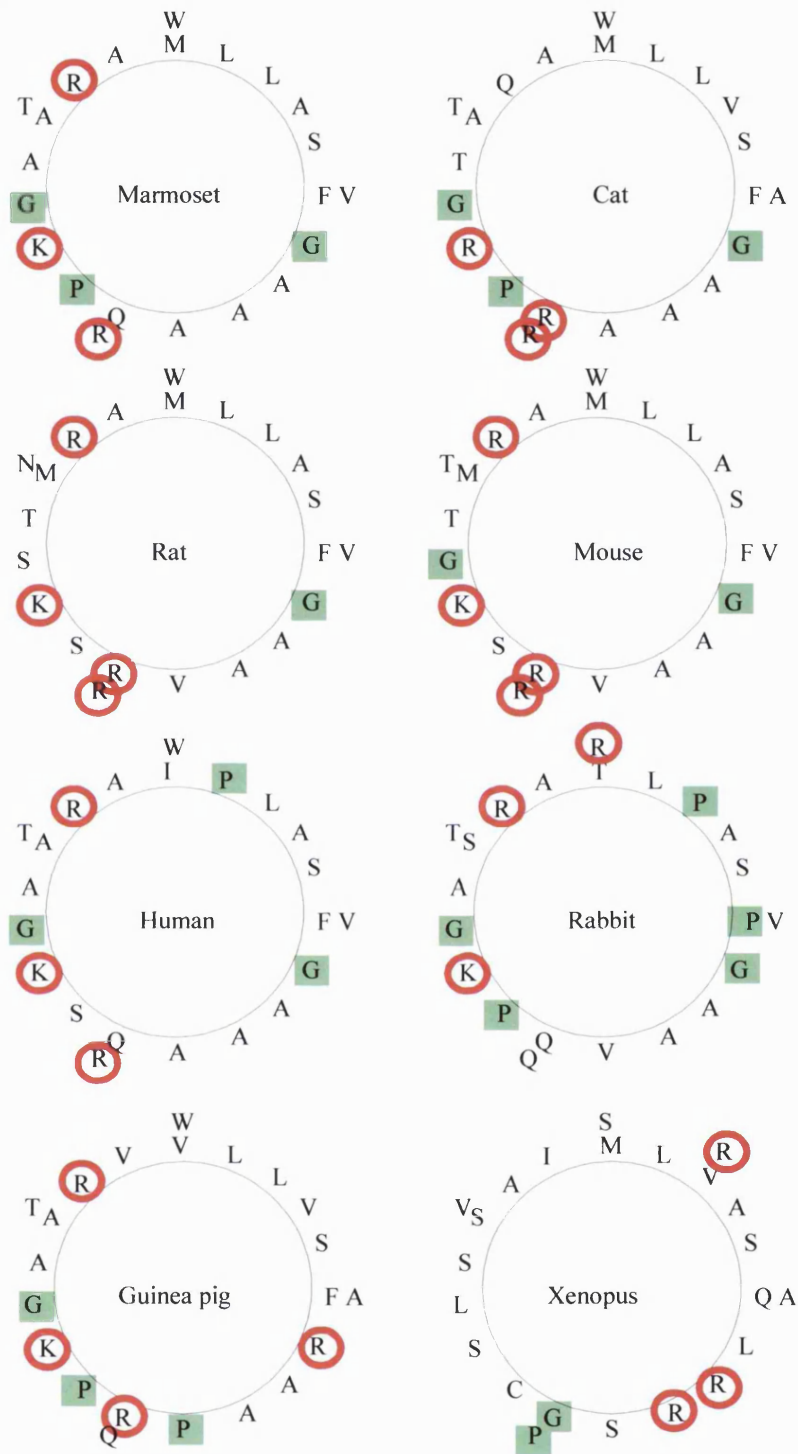


Figure 41 - Helical analysis of N-terminal amino acids of known AGT orthologues

The amino acid sequences are arranged in a helical format (clockwise, 1st, 12th, 5th, 16th, 9th, 2nd, 13th, 6th, 17th, 10th, 3rd, 14th, 7th, 18th, 11th, 4th, 15th, 8th).

Positively charged residues (K, R and H) are ringed in red, helix breaking residues (G and P) are shaded in green.

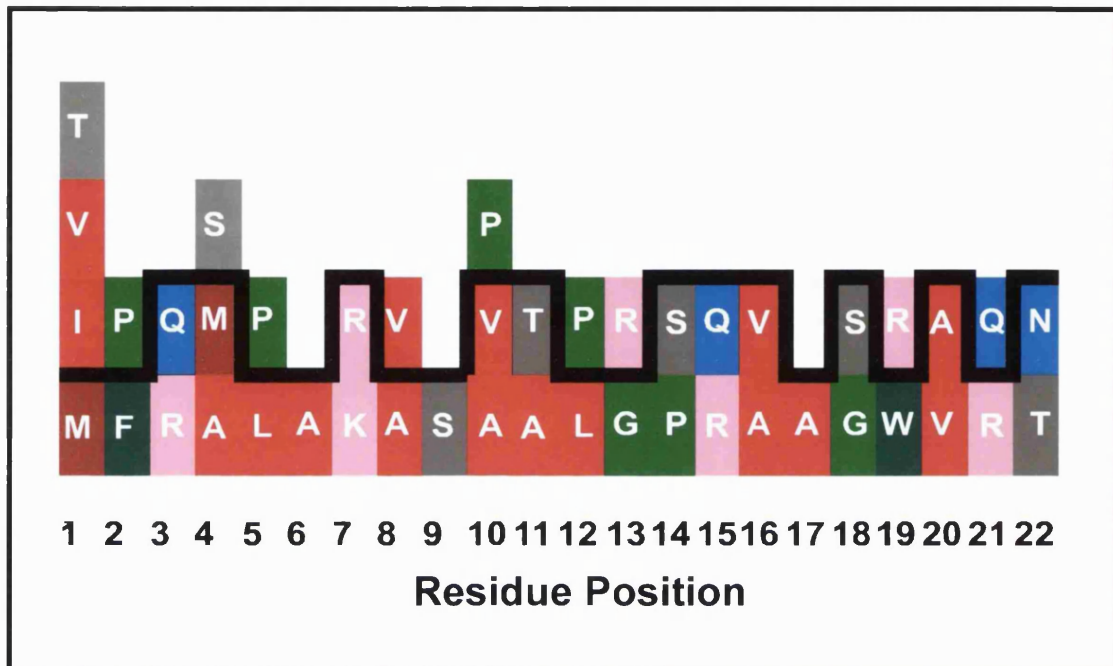


Figure 42 - The position specific constraints on AGT MTS evolution in mammals

Residues predicted to be encoded by the MTS or ancestral MTS, now 5'UTR, of human, rabbit, marmoset, rat, mouse, guinea pig and cat AGT gene, were arranged in residue positions. The amino acid boxes were coloured for amino acid type, for key see figure 46. The residues above the line are only in the predicted amino acid sequence of the 5'UTR of the human, rabbit and guinea pig genes corresponding to the ancestral MTS.

MTS now in the 5'UTR. No other residue positions were found to have a proportionately greater amount of changes. These results might suggest that mammalian AGT MTSs have not been under the same positive selection pressure to loose mitochondrial targeting as has been seen in the primates (Chapter 3). It was also noted that, the evolution of the MTS is somewhat confined, as expected, by amino acid type. None of the mammalian AGT MTSs contained negatively charged amino acids (D and E) or cysteines (C).

5.2.2 The PTS1 of AGT

PTS1s are extraordinarily degenerate. Many tripeptide combinations can be involved in peroxisomal targeting, but their effects can be context dependent at both the cell and protein level. Very rarely can the potential of a C-terminal tripeptide to peroxisomally target be authoritatively predicted unless it follows the conservative consensus²³²⁴ or has been tested. Work on AGT evolution sometimes requires the ability to predict subcellular location on the basis of sequence. Also, as mentioned in chapter 4, the PTS1 of human AGT is protein context dependent. This has lead to the hypothesis that there is additional targeting information within human AGT protein. Protein context dependence of PTS1s has been suggested for other PTS1s and proteins¹⁰⁷ but the phenomenon has not been entirely defined.

For these reasons, an extensive literature and sequence database survey of C-terminal tripeptides, which have some characteristics of PTS1s, was performed. Characteristics of PTS1s were defined as:

- shown to be necessary for the targeting of a protein to the peroxisomes,
- shown to target a reporter protein to the peroxisomes

- shown to bind to Pex5 (the PTS1 receptor)
- present in a peroxisomal matrix protein

The terms used to search by keyword in PubMed, in various combinations, included:

"Pex5", "PTS1", "peroxisomal", "tripeptide", "import", "targeting", "peroxisome", "reporter", "C-terminus", "Pas8", "Pas10", "PXR1", "PER3", "PTS1R", "carboxyl terminus", and "peroxin".

The Entrez protein database was searched using the terms "peroxisomal", "peroxisomal targeting", "PTS1", "peroxisomal targeting sequence, type 1".

Information on 78 tripeptides that have some characteristics of a PTS1 was obtained (table 11). The category for which the survey was probably least comprehensive is the C-terminal tripeptides found naturally on peroxisomal proteins, as the presence of a PTS1 in a cloned gene is rarely the subject of a publication and the sufficiency of a PTS1 to direct a reporter to the peroxisomes is very rarely tested. Also, studies of PTS1 function tend to test known or suspected PTS1s, similar to the consensus sequence. Therefore, there may be tripeptides dissimilar to the consensus, that are capable of binding Pex5 and directing a protein to the peroxisomes, that are not included in table 11 because they have never been tested. Negative information was included when the C-terminal tripeptide was known to function as a PTS1 in other contexts but C-terminal tripeptides were not included when the only information known about them was negative.

Table 11 - C terminal tripeptides with at least some properties of PTS1s

| Tri-peptides | Natural Peroxisomal Protein | In 2 hybrid assay interacts with:- | | Sufficient to target reporter protein to peroxisomes in:- | | | Necessary for import of natural protein in:- | |
|--------------|--|---|--|---|---------------------------------------|--|--|---|
| | | Yeast Pex 5 | Human Pex 5 | yeast cells | mammalian cells | plant cells | yeast cells | mammalian cells |
| AEL | | | | MDH3 in <i>Sc</i> ^M | | | | |
| AHM | | | Random polypeptide ^A | | Not CAT in COS cells ^B | | | |
| AKF | | <i>Sc</i> , random polypeptide ^A | | | | | | |
| AKI | <i>Ct</i> trifunctional enzyme ^O . BHD-IV ^R | | On <i>Ss</i> BHD-IV ^R | | | | | Necessary for <i>Ct</i> trifunctional enzyme ^O and BDH-IV ^R |
| AKL | Human Acyl-CoA: dihydroxyacetonephosphate acyltransferase ^G Human alcohol dehydrogenase ^V Human carnitine acyltransferase ^Z Human dihydroxyacetone phosphate acyltransferase ^{A,A,A} Human NADP+ dependent isocitrate dehydrogenase ^{II} Human trans 2-enoyl CoA reductase ^{JJ} | <i>Sc</i> , random polypeptide ^A | Random polypeptide ^A . <i>Nt</i> cDNA ^K | | CAT in COS cells ^B . | | | |
| AKM | FFL ^{BBB} | <i>Sc</i> , random polypeptide ^A | | | Poor on CAT in COS cells ^B | | | |
| AKV | <i>Ct</i> isocitrate lyase ^{CCC} | | | | | | | |
| ALL | | <i>Sc</i> , random polypeptide ^A | random polypeptide ^A | | | | | |
| ANL | Human and rat Catalase ^{BB} Spinach 2-hydroxyacid oxidase ^{ZZ} | | random polypeptide ^A | Weak on MDH3 in <i>Sc</i> ^M | | Very weak on CAT in <i>Nt</i> ^Q | | |
| ARF | <i>Hp</i> and <i>Pp</i> alcohol oxidase ^I . | | | | | | | |
| ARL | Spinach glycolate oxidase ^H | <i>Sc</i> , random polypeptide ^A | Random polypeptide ^A . <i>Nt</i> cDNA ^K . | | CAT in COS cells ^B . | | | |
| ARM | Castor bean and cotton isocitrate lyase ^F | | Random polypeptide ^A . <i>Nt</i> cDNA ^K | Poor on CAT in COS cells ^B | | | | |
| ASL | Human alpha methylacyl CoA racemase ^W | | | | | | | |
| CHL | | <i>Sc</i> , random polypeptide ^A | Not on random polypeptide ^A | Not CAT-CHL ^B | Not CAT in COS cells ^B | | | |
| CKL | | | Random polypeptide ^A . <i>Nt</i> cDNA ^K | | CAT in COS cells ^B | CAT in <i>Nt</i> ^Q | | |
| CLL | | <i>Sc</i> , random polypeptide ^A | Random polypeptide ^A | | | | | |

Table 11 - C terminal tripeptides with at least some properties of PTS1s

| Tri-peptides | Natural Peroxisomal Protein | In 2 hybrid assay interacts with:- | | Sufficient to target reporter protein to peroxisomes in:- | | | Necessary for import of natural protein in:- | |
|--------------|--|-------------------------------------|--|---|--|------------------------------------|--|-----------------|
| | | Yeast Pex 5 | Human Pex 5 | yeast cells | mammalian cells | plant cells | yeast cells | mammalian cells |
| CNL | | Sc, random polypeptide ^A | Random polypeptide ^A | | | | | |
| CRL | | Sc, random polypeptide ^A | Random polypeptide ^A <i>Nt</i> cDNA. | | | | | |
| CRM | | | Random polypeptide ^A . <i>Nt</i> cDNA ^K | | Not CAT in COS cells ^B | | | |
| CVL | | Very weak with Sc ^A | Random polypeptide ^A | | | | | |
| EEL | | | | MDH3 in Sc ^M | | | | |
| EKL | | | | MDH3 in Sc ^M | | | | |
| FFF | | | | Very weak on MDH3 in Sc ^M | | | | |
| FKL | | | | MDH3 in Sc ^M | | Not on CAT in tobacco ^Q | | |
| FRL | | Sc, random polypeptide ^A | | | | | | |
| GKL | | Sc, random polypeptide ^A | Random polypeptide ^A | MDH3 in Sc ^M | | CAT in <i>Nt</i> ^Q . | | |
| GRL | | | Random polypeptide ^A . <i>Nt</i> cDNA ^K | | | | | |
| HRL | Guinea pig AGT ^{FFF} | Sc, random polypeptide ^A | Very weak ^A . | | | | | |
| IKL | | | | MDH3 in Sc ^M | | | | |
| KKL | Human AGT ^{VY} Marmoset AGT ^{GGG} | Sc, not on AGT ^{HHH} | Not on AGT ^{HHH} | MDH3 in Sc ^M | Not on CAT, GFP or FFL in COS cells ^S or Xenopus AGT ^{ooo} | | | |
| LKL | Human very long chain acyl CoA synthetase ^{oo} | Sc, random polypeptide ^A | | MDH3 in Sc ^M | | | | |
| MKL | | | | MDH3 in Sc ^M | | | | |
| NKL | <i>Hp</i> Dihydroxyacetone synthase ^I Cat ^{GGG} , rat ^{III} and mouse AGT ^{III} | | | MDH3 in Sc ^M | | | | |
| PKI | Acyl CoA oxidase homolog Pox1p in <i>Pp</i> ^C | | | GFP in <i>Pp</i> ^C | | | | |
| PKL | | Sc, random polypeptide ^A | Random polypeptide ^A | MDH3 in Sc ^M | | Not CAT in <i>Nt</i> ^Q | | |
| PKM | | | Random polypeptide ^A | | | | | |
| PNL | | | Random polypeptide ^A | | | | | |

Table 11 - C terminal tripeptides with at least some properties of PTS1s

| Tri-peptide | Peroxisomal Protein | In 2 hybrid assay interacts with:- | | Sufficient to target reporter protein to peroxisomes in:- | | | Necessary for import of natural protein | |
|-------------|---|--|--|--|--|---|---|-----------------|
| | | Yeast Pex 5 | Human Pex 5 | yeast cells | mammalian cells | plants cells | yeast cells | mammalian cells |
| PRL | | | Random polypeptide ^A . <i>Nt</i> cDNA ^K | | | | | |
| PRM | Cucumber MFP II ^E | | | | | | | |
| PSL | | Not <i>Sc</i> on random polypeptide ^A | Random polypeptide ^A | | | | | |
| PSM | Maize catalase ^{MMM} | | | | | | | |
| PTM | Sweet potato catalase ^{NNN} | | | | | | | |
| QKL | | <i>Sc</i> , random polypeptide ^A | | | | | | |
| SAL | | Not <i>Sc</i> on random polypeptide ^A Not <i>Nt</i> on <i>Nt</i> cDNA ^K | Random polypeptide ^A <i>Nt</i> cDNA ^K | Weak on MDH3 in <i>Sc</i> ^M | | | | |
| SEL | | | | Weak on MDH3 in <i>Sc</i> ^M MDH3 in <i>Sc</i> ^M . | | | | |
| SFL | | | | On MDH3 in <i>Sc</i> ^M | On CAT in COS cells ^B . | Targets CAT in <i>Nt</i> ^Q | | |
| SHL | D-amino acid oxidase in human, mouse pig and rabbit ^{DD} Fruitfly urate oxidase ^{XX} | | On random polypeptide ^A . On <i>Nt</i> cDNA ^K . | On MDH3 in <i>Sc</i> ^M | On CAT in COS cells ^B . | Targets CAT in <i>Nt</i> ^Q | | |
| SKD | | | | Not sufficient to target MDH3 in <i>Sc</i> ^M | | Is not sufficient to target CAT in <i>Nt</i> ^Q . | | |
| SKE | | | | Very weak on MDH3 in <i>Sc</i> ^M | | | | |
| SKF | Fruitfly catalase ^{WW} | | | On Catalase in <i>Sc</i> . On MDH3 in <i>Sc</i> ^M | | CAT in <i>Nt</i> ^Q | | |
| SKI | <i>Hp.</i> Catalase ^I Human HAOX1 ^{MM} Rat epoxide hydrolase ^{LLL} Human and mouse glycolate oxidase ^{PPP} | Not <i>Nt</i> Pex5p on <i>Nt</i> cDNA ^K | Not on <i>Nt</i> cDNA ^K | | Weak on rat epoxide hydrolase ^{LLL} | CAT in <i>Nt</i> ^Q | | |

Table 11 - C terminal tripeptides with at least some properties of PTS1s

| Tri-peptide | Natural Peroxisomal Protein | In 2 hybrid assay interacts with:- | | Sufficient to target reporter protein to peroxisomes in:- | | | Necessary for import of natural protein in:- | |
|-------------|---|---|---|---|--|--|--|--------------------------|
| | | Yeast Pex 5 | Human Pex 5 | yeast cells | mammalian cells | plant cells | yeast cells | mammalian cells |
| SKL | FFL ^N Rat Acyl-CoA oxidase ^{SS} Cucumber and cotton malate synthase ^{TT} Soybean urate oxidase ^{UU} Click beetle luciferase ^{VV} Sc Citrate synthase ^G Human Acyl-CoA thioesterase ^F Human Acyl CoA oxidase ^P Human acyl CoA oxidase 3 ^U Human Bifunctional enzyme ^X Human branched chain CoA oxidase ^Y Human D3,D2-enoyl CoA isomerase ^{CC} Human DBI-related protein 1 ^{EE} Human fatty acyl-CoA oxidase ^{FF} Human pristanoyl-CoA oxidase ^{GG} Human soluble epoxide hydrolase ^{NN} | Sc random polypeptide ^A <i>Nt</i> , <i>Nt</i> cDNA clone ^K . | random polypeptide ^A <i>Nt</i> cDNA clone ^K . | FFL in <i>Sc</i> . MDH3 in <i>Sc</i> ^M . | FFL in COS cells. CAT in COS cells ^B | On fusion in tobacco ^Q | | Necessary ^N . |
| SKM | | | | | Poor on CAT in COS cells ^B | | | |
| SKN | | | | Very weak on MDH3 in <i>Sc</i> ^M | | | | |
| SKP | | | | Very weak on MDH3 in <i>Sc</i> ^M | | | | |
| SKV | | | | MDH3 in <i>Sc</i> ^M | | | | |
| SKR | | | | Not MDH3 in <i>Sc</i> ^M | | | | |
| SKS | | | | Very weak MDH3 in <i>Sc</i> ^M | | | | |
| SKT | | | | Very weak targeting of MDH3 in <i>Sc</i> ^M | | | | |
| SKY | | | | | | Targets CAT in tobacco ^Q | | |
| SLK | <i>Hp</i> malate synthase ^{RR} | | | | | | | |
| SLL | | Not <i>Sc</i> on random polypeptide ^A Not <i>Nt</i> cDNA ^K | Random polypeptide ^A <i>Nt</i> cDNA ^K | | | Very weak on CAT in tobacco ^Q | | |
| SML | | Very weak <i>Sc</i> on random polypeptide ^A <i>Nt</i> , <i>Nt</i> cDNA clone ^K | On random polypeptide ^A <i>Nt</i> cDNA clone ^K | | | | | |

Table 11 - C terminal tripeptides with at least some properties of PTS1s

| Tri-peptide | Natural Peroxisomal Protein | In 2 hybrid assay interacts with:- | | Sufficient to target reporter protein to peroxisomes in:- | | | Necessary for import of natural protein in:- | |
|-------------|--|--|---|--|--|--|--|---------------------------|
| | | Yeast Pex 5 | Human Pex 5 | yeast cells | mammalian cells | plant cells | yeast cells | mammalian cells |
| SNM | 2-hydroxyphytanoyl-CoA lyase ^E | | | | | | | |
| SQL | Human putative antioxidant enzyme ^{LL} Rabbit AGT ^{KKK} | Not <i>Sc</i> , random polypeptide ^A Not <i>Nt</i> , <i>Nt</i> cDNA ^K | Random polypeptide ^A . <i>Nt</i> cDNA ^K | | | | | |
| SRL | Rat hydroxyacid oxidase ^{QQ} Castor bean, pumpkin and rape malate synthase ^G Human HAOX3 ^{HH} Human phosphomevalonate kinase ^{KK} | <i>Sc</i> , random polypeptide ^A <i>Nt</i> , <i>Nt</i> cDNA ^K | Random polypeptide ^A <i>Nt</i> cDNA ^K | MDH3 in <i>Sc</i> ^M . | CAT in COS cells ^B . DCR in COS cells ^R | CAT in <i>Nt</i> ^Q . On fusion in <i>Nt</i> ^K | | |
| SRM | Oilseed rape, cotton seed, cucumber, tomato, castor bean and lob lolly pine isocitrate lyase ^F | Not <i>Sc</i> , random polypeptide ^A Not <i>Nt</i> , <i>Nt</i> cDNA ^K | Random polypeptide ^A On <i>Nt</i> cDNA ^K | | Poor on CAT in COS cells ^B | fusion in <i>Nt</i> ^K | | |
| SSL | PGK-C ^{DDDD} | <i>Nt</i> , <i>Nt</i> cDNA ^K | Random polypeptide ^A <i>Nt</i> cDNA ^K | Weak on MDH3 in <i>Sc</i> ^M . Not PGK-C in <i>Sc</i> ^{DDDD} | Not CAT in COS cells ^B AGT in COS cells ^T Not FFL in COS cells ^{DDDD} | | | Necessary ^{DDDD} |
| SYL | | | Random polypeptide ^A | On MDH3 in <i>Sc</i> ^M | | | | |
| THL | Human carnitine O-octanoyltransferase ^{AA} | | | | | | | |
| TKL | | <i>Sc</i> , random polypeptide ^A | | MDH3 in <i>Sc</i> ^M | | CAT in <i>Nt</i> ^Q | | |
| VKL | | <i>Sc</i> , random polypeptide ^A | | | | | | |
| VRI | Rat Xanthine dehydrogenase ^{PP} | | | | | | | |
| WKL | | <i>Sc</i> , random polypeptide ^A | Very weak ^A | | | | | |
| WRL | | <i>Sc</i> , random polypeptide ^A | | | | | | |
| YKL | | <i>Sc</i> , random polypeptide ^A | | | | | | |

Species abbreviations were as follows: *Sc* = *Saccharomyces cerevisiae*, *Nt* = *Nicotiana tabacum*, *Pp* = *Pichia pastoris*, *Hp* = *Hansenula polymorpha*, *Tb* = *Trypanosoma brucei*, *Ct* = *Candida tropicalis*, *Pc* = *Penicillium chrysogenum*, *Cb* = *Candida boidinii*, *Lm* = *Leishmania mexicana*, *Ss* = *Sus scrofa*. Protein abbreviations were as follows: CAT = Chloramphenicol acetyltransferase, GFP = Green fluorescent protein, MDH3 = malate dehydrogenase, PK = Glycosomal phosphoglycerate kinase, BDH-IV = Porcine 17beta-hydroxysteroid dehydrogenase type IV, MFP II = cucumber multifunctional β -oxidation protein, HAOX3 = medium chain 2-hydroxy acid oxidase 3, HAOX1 = short chain 2-hydroxy acid oxidase FFL = Firefly luciferase. Information was referenced (Appendix). Where literature references were not available accession numbers were given.

5.2.2.1 A new Consensus

The residues present within the -1, -2 and -3 positions in the tripeptides contained in table 11 were counted (table 12). Each tripeptide was counted once, regardless of how many proteins it is found within or the information pertaining to it. Some residues are very often seen in a certain position, e.g. there are 13 tripeptides in table 11 with alanine in the -3 position. The more frequent residues are: S, A, P and C at -3, K and R at -2, and L and M at -1. This combination is similar to the consensus PTS1²³²⁴ (Chapter 1), which is, S/A/C-K/R/H-L/M. With regard to the less frequent residues: the neighbouring two residues are recorded (table 12) and these have a strong tendency to be popular residues (in 46 out of 54 cases). *C-terminal tripeptides with some properties of a PTS1 have two residues from the combination (S/A/P/C - K/R - L/M) and can tolerate one less frequent residue.* Tripeptides with all types of amino acid, except aspartic acid and arginine at the -3 position are included in table 11, as are tripeptides with 14 types of amino acid residue in the -2 position and 14 types of residue in the -1 position.

5.2.2.2 Context dependence of PTS1s

The canonical PTS1 (SKL) is made up of 3 frequent residues as are other tripeptides that work across all species and protein contexts e.g. AKL, CKL, CRL and SRL (table 11). For instance, SKL binds *ScPex5*, *NiPex5* and human Pex5; it is sufficient to target MDH3 and luciferase to *S. cerevisiae* peroxisomes, luciferase, GFP and CAT to COS cell peroxisomes and CAT into glycosomes of trypanosomes (tables 11 and 13B). However another class of PTS1s can be discerned that have 2 frequent residues and one infrequent one. These tend to be species and/or protein context dependent. An example

Table 12 - The residues present within the -1, -2 and -3 positions in the tripeptides contained in table 11**-3 Position**

| Residue | Freq | Context | Residue | Freq | Context | Residue | Freq | Context |
|---------|------|---------|---------|------|---------|---------|------|---------|
| S | 27 | | G | 2 | -KL | H | 1 | -RL |
| A | 13 | | | | -RL | I | 1 | -KL |
| P | 9 | | T | 2 | -HL | K | 1 | -KL |
| C | 7 | | | | -KL | L | 1 | -KL |
| E | 2 | -KL | V | 2 | -KL | M | 1 | -KL |
| | | -EL | | | -RI | N | 1 | -KL |
| F | 3 | -KL | W | 2 | -KL | Q | 1 | -KL |
| | | -RL | | | -RL | Y | 1 | -KL |
| | | -FF | | | | | | |

-2 Position

| Residue | Freq | Context | Residue | Freq | Context | Residue | Freq | Context |
|---------|------|---------|---------|------|---------|---------|------|---------|
| K | 35 | | H | 4 | A-M | V | 1 | C-L |
| R | 14 | | | | C-L | T | 1 | P-M |
| N | 3 | C-L | | | S-L | A | 1 | S-L |
| | | P-L | | | T-L | F | 2 | S-L |
| | | S-M | L | 4 | A-L | | | F-F |
| S | 4 | A-L | | | C-L | M | 1 | S-L |
| | | P-L | | | S-K | Q | 1 | S-L |
| | | P-M | | | S-L | Y | 1 | S-L |
| | | S-L | E | 2 | A-L | | | |
| | | | | | E-L | | | |

-1 Position

| Residue | Freq | Context | Residue | Freq | Context | Residue | Freq. | Context |
|---------|------|---------|---------|------|---------|---------|-------|---------|
| L | 46 | | I | 3 | AK- | T | 1 | SK- |
| M | 11 | | | | SK- | S | 1 | SK- |
| F | 4 | AR- | | | VR- | R | 1 | SK- |
| | | AK- | V | 2 | AK- | P | 1 | SK- |
| | | SK- | | | SK- | N | 1 | SK- |
| | | FF- | Y | 1 | SK- | E | 1 | SK- |
| | | | K | 1 | SL- | D | 1 | SK- |

All residues found in the tripeptides of table 11 are included along with the number of different tripeptides in which they occur at that position. For residues that are considered infrequent the neighbouring residues are recorded. Freq = the number of tripeptides (not proteins) in which it occurs.

Table 13 - C-terminal tripeptides in functional in human, *S. cerevisiae* and trypanosome systems**A - human and *S. cerevisiae***

| Human | | | <i>S. cerevisiae</i> | | |
|----------|---------------------------------|--------|----------------------|--|--------|
| Position | Residues that occur | Number | Position | Residues that occur | Number |
| -3 | A, C, S, T, G, P | 6 | -3 | A, C, E, S, T, F, G, H, I, K, L, M, N, P, Q, V, W, Y | 18 |
| -2 | H, K, L, N, R, S, Q, V, A, M, Y | 11 | -2 | E, K, L, N, R, H, F, Y | 8 |
| -1 | M, L | 2 | -1 | L, F, M, V | 4 |

B - Tripeptides capable of directing a reporter protein to Glycosomes

| 0-19% | | 20-39% | | 40-59% | | 60-79% | | 80-100% | |
|-------|-----|--------|-----|--------|-----|--------|-----|---------|-----|
| SKS | IKL | SKV | VK | TKL | SNL | GKL | PKL | SKY | CKL |
| SKT | KKL | QKL | YKL | | | HKL | SHL | SKL | NKL |
| SKW | LKL | SCL | SIL | | | SSL | | AKL | SML |
| DKL | MKL | SYL | SKF | | | | | SQL | SKI |
| EKL | RKL | | | | | | | SRL | SKM |
| FKL | WKL | | | | | | | | |
| SDL | SAL | | | | | | | | |
| SEL | SLL | | | | | | | | |
| SFL | SPL | | | | | | | | |
| SGL | STL | | | | | | | | |
| SKA | SVL | | | | | | | | |
| SKC | SWL | | | | | | | | |
| SKD | SKG | | | | | | | | |
| SKE | SKH | | | | | | | | |
| SKN | SKK | | | | | | | | |
| SKP | SKR | | | | | | | | |
| SKQ | | | | | | | | | |

Column headings refer to % efficiency of import measured by % luciferase activity in the pellet fraction.

of species dependent tripeptide is CHL, which can bind to *ScPex5* but not *HsPex5*. The consequence of this is that GFP-CHL is imported into *S. cerevisiae* peroxisomes but CAT-CHL is not imported into COS cell peroxisomes (table 11). Species specificity was investigated further by tabulating the residues that are acceptable in PTS1 positions in mammalian and *S. cerevisiae* systems (table 13A). In the *S. cerevisiae* system the number of residues that are acceptable in the -3 position are more numerous than in the human system. In fact, the only residues which are not seen in the -3 position are arginine and aspartic acid (table 13A). Another example of species difference is the glycosomes of trypanosomes, which are thought to be related to peroxisomes¹⁰⁸. The acceptable tripeptide range in glycosomally targeted protein in trypanosomes is different again from the range acceptable in mammalian or yeast cells (table 13B).

In addition to species dependence, and as has been seen for the C-terminal tripeptide of human AGT (KKL), some tripeptides act in a protein context dependent manner. They are necessary for the targeting of the natural protein but not sufficient to target reporter proteins. Other examples of the phenomenon have been found in table 11. The C terminal tripeptide of typanosome phosphoglycerate kinase (PK) SSL, is necessary for PK peroxisomal targeting and it is also sufficient to target human AGT to the peroxisomes in COS cells. However, GFP-SSL and luciferase-SSL are cytosolic in COS cells (table 11). Similarly the C-terminal tripeptide of human soluble epoxide hydrolase (SKM) is presumably important in the targeting of the enzyme but is not sufficient for the peroxisomal targeting of CAT in COS cells.

Protein context dependence may be due to an additional element in the protein.

Possible functions of this element are:

- binding Pex5 directly,
- allosterically affecting the PTS1 binding site of Pex5,
- binding an adapter molecule, or
- interaction with downstream components of the import pathway.

It is unknown if the additional element is the same in all proteins that are targeted by protein-context dependent PTS1s, or it is a different type of element with a different function in each protein. No obvious shared motif was observed in an alignment of the human soluble epoxide hydrolase, trypanosome phosphoglycerate kinase or AGT protein sequences (data not shown). As peroxisomal proteins can be imported in a folded conformation^{38;39;109} it is likely that the additional element is structural.

Gatto and co-workers (2000)¹¹⁰ have proposed a model of Pex5-SKL binding. The structure of the region of the Pex5 protein known to be involved in PTS1 binding, the tetratricopeptide repeats or TPR domain^{25;111} was predicted by the mapping of the amino acid sequence on to a close homologue. The tripeptide appears to sit in the groove formed by two adjacent TPRs. There seems to be enough space within this region of Pex5 to accommodate larger side chains at the -1 position, than those of the leucine residue, but the fit for the -2 and -3 positions is more constrained¹¹⁰. From table 12, the more frequently found residues seem to have the similar side chains, i.e. at -2 position arginine is the most frequent residue after lysine and both residues have large aliphatic positively charged side chains. At the -3 position the most frequent residue after serine is alanine, alanine has a methyl group side chain whilst serine has a methyl alcohol side chain. However if the model proposed by Gatto and co-workers is correct, many of the tripeptides in table 11 would

be unable to bind to the TPR domain of Pex5, unless some allosteric interaction occurred with another part of the PTS1 containing protein to change the conformation of the binding groove. This may be the function of the additional targeting information.

5.2.2.3 Variability of PTS1s

Although the PTS1 of mammalian AGT seems to vary considerably, the expected amount of variation was difficult to ascertain. So, the C-terminal tripeptides of other orthologous groups of mammalian peroxisomal proteins were investigated (table 14). The C-terminal tripeptides were much less variable in most other peroxisomal enzymes than the AGT C-terminal tripeptides. The C-terminal tripeptides in seven out of the ten enzymes in table 14 do not vary in different species. In the three that do vary, unfortunately, only two different species were available on the database and so the extent of the variability is unknown. D-amino acid oxidase is a very good example as it has been cloned in many of the same mammalian species as AGT. The C-terminus of D-amino acid oxidase between mouse, rabbit, human, guinea pig and rat, is invariable; whilst the C-terminus of AGT varies four times in the same species. This analysis highlights the remarkable variability of the AGT PTS1 and leads to the question, why should the AGT PTS1 be so variable?

Table 14 - The variability of PTS1s found on mammalian peroxisomal enzymes

| Gene | Species | Accession Number | C-terminal tripeptide |
|---|------------|------------------|-----------------------|
| Urate Oxidase | Mouse | P25688 | SRL |
| | Rat | O44111 | SRL |
| | Rabbit | P11645 | SRL |
| | Pig | P16164 | SRL |
| | Baboon | P25689 | SRL |
| Glycolate oxidase | Mouse | NP_062418 | SRL |
| | Human | AAF63219 | SKI |
| D amino acid oxidase | Mouse | BAA01063 | SHL |
| | Rabbit | JX0132 | SHL |
| | Human | S01340 | SHL |
| | Pig | OXPGBA | SHL |
| | Guinea pig | CAA07616 | SHL |
| | Rat | O35078 | SHL |
| | Hamster | BAA74715 | SHL |
| Peroxisomal trans 2-enoyl CoA reductase | Human | NP_060911 | AKL |
| | Mouse | AAF69800 | AKL |
| | Guinea pig | AAF69799 | AKL |
| Sarcosine oxidase | Mouse | NP_032978 | AHL |
| | Rabbit | AAB48443 | AHL |
| Antioxidant enzyme B166 | Human | AAF03750 | SQL |
| | Rat | AAF03751 | SQL |
| Acyl CoA oxidase | Mouse | BAA86870 | SKL |
| | Rat | NP_059036 | SKL |
| | Human | JC2066 | SKL |
| Pristanoyl-CoA oxidase | Human | NP_003492 | SKL |
| | Rat | Q63448 | SQL |
| Catalase | Mouse | P24270 | ANL |
| | Human | P04040 | ANL |
| | Dog | BAA36420 | ANL |
| | Guinea pig | CAB57222 | ANL |
| | Rat | NP_036652 | ANL |
| D aspartate oxidase | Human | Q99489 | SNL |
| | Cow | CAA64622 | SKL |

The C-terminal tri-peptides of mammalian enzymes, recovered from the database are recorded.

5.2.3 Recovery of homologues from the database

Possible homologues of AGT were identified by BLAST searching Genbank in May 2000, with human and xenopus AGT peptide sequences. Hits for both sequences are listed in table 15. Other database entries, such as the plant and fungal sequences were added to the data set, even though they were not returned from the BLAST search. The additional sequences had been returned on preliminary searches and were subjected to further analysis to test their validity as homologues. The final data set included 28 possible homologues, including sequences from bacterial and archaeal organisms, as well as eukaryotes.

5.2.4 Recovery of human paralogues from the database

Human paralogues were obtained from Entrez and TrEMBL by searching with the human AGT peptide sequence and using the search terms "human aminotransferase" and "human pyridoxal phosphate dependent enzyme" 168 hits were returned which included 13 full-length, human paralogue sequences (table 16).

5.2.5 Classification of homologues

It is probable that, at some time, all the homologues arose from one ancestral enzyme. Therefore, it is unclear whether a gene identified as being homologous to human AGT, is an orthologue or a paralogue. To distinguish orthologues from paralogues, strict criteria were applied.

Table 15 - AGT homologues recovered from BLAST searches

| Species | abbrev- iation | Acc number (Entrez) | Significance rating against human | Significance rating against Xenopus |
|---|-------------------|------------------------|---|---|
| <i>Aeropyrum pernix</i> | Ap | Q9Y962 | 6e-37 | 9e-42 |
| <i>Anabaena cylindrica</i> | Ac | P16421 | E=1e-26 | 1e-26 |
| <i>Arabidopsis thaliana</i> | At | AAD28669 | not returned | not returned |
| <i>Archaeoglobus fulgidus</i> | Af | O28855 | E=2e-33 | not returned |
| <i>Bacillus subtilis</i> | Bs | O32148 | 3e-52 | 1e-51 |
| <i>Caenorhabditis elegans</i> | Ce | Q94055 | 2e-87 | 3e-98 |
| <i>Callithrix jacchus</i> | Cj | P31029 | 0 | 1e-153 |
| <i>Cavia Porcellus</i> | Cp | O35504 | 1e-176 | 1e-141 |
| <i>Deinococcus radiodurans</i> | Dr | AAF10920 | 7e-45 | 1e-55 |
| <i>Drosophila melanogaster</i> | Dm | CAA58024 | 8e-43 | 3e-47 |
| <i>Felis catus</i> | Fc | P41689 | 0 | 1e-150 |
| <i>Fruittillaria agrestis</i> | Fa | AAB95218 | not returned | not returned |
| <i>Homo sapiens</i> | Hs | NP_000021 | 0 | 1e-149 |
| <i>Hyphomicrobium methylovorum</i> | Hm | BAA19919 | not returned | not returned |
| <i>Methanobacterium thermoautotrophicum</i> | Mta | O27638 | 4e-39 | 1e-34 |
| <i>Methanobacterium thermoformicum</i> | Mt | Q59569 | 2e-40 | 1e-35 |
| <i>Methanococcus jannaschii</i> | Mj | P55819 | 5e-38 | 4e-38 |
| <i>Mus muscules</i> | Mm | O35423 | 1e-171 | 1e-140 |
| <i>Oryctolagus cuniculus</i> | Oc | P31030 | 0 | 1e-147 |
| <i>Pyrococcus abyssi</i> | Pa | CAB49753 | 4e-30 | 4e-30 |
| <i>Pyrococcus horikoshii</i> | Ph | O59033 | 2e-30 | 2e-30 |
| <i>Rattus norvegicus</i> | Rn | P09139 | 0 | 1e-145 |
| <i>Sacchromyces cerivisiae</i> | Dc | NP_011157 | not returned | not returned |
| <i>Streptomyces coelicolor</i> | Sc | Q9XBE7 | 1e-28 | 1e-28 |
| <i>Synechococcus PCC 6716</i> | Sp7 | P14776 | 1e-35 | 1e-34 |
| <i>Synechocystis PCC 6803</i> | Sp8 | P74281 | 6e-39 | 2e-29 |
| <i>Thermotoga maritima</i> | Tm | Q9X1C0 | 5e-43 | 6e-41 |
| <i>Xenopus laevis</i> | Xl | | not in database | not in database |

The significance rating is the probability that a comparison between the submitted sequence and a random sequence would achieve a score greater than or equal to the score attributed to the match between the query sequence and the returned sequence. Not returned = not returned from this search, added to the data set because they have been returned on preliminary searches.

Table 16 - AGT paralogues and their percentage identity to human AGT

| Name | Accession Number (Entrez) | Code | % Identity to human AGT |
|--|---------------------------|------------|-------------------------|
| 4-Aminobutyrate aminotransferase | NM_000663 | ABAT | 31 |
| Alanine aminotransferase | NM_005309 | A2O | 36 |
| α -Methylacyl-CoA racemase | AF158378 | AMR | 34 |
| Aspartate aminotransferase 2 | NM_002080 | AAT2 | 28 |
| Aspartate aminotransferase 1 | NM_002079 | AAT1 | 28 |
| Branched chain aminotransferase 1 | NM_005504 | BCAT1 | 30 |
| Cysteine conjugate- β -lyase | NM_004059 | CBL | 32 |
| Glutamine-fructose-6-phosphate transaminase 1 | NM_002056 | GFP1 | 31 |
| Kynurenine 3-monooxygenase | NM_003679 | KMO | 33 |
| L-Kynurenine/ α -amino adipate aminotransferase | AF097994 | CAA | 31 |
| Ornithine aminotransferase | M12267 | OAT | 35 |
| Phosphoserine aminotransferase | AF113132 | PSA | 30 |
| Tyrosine aminotransferase | X52520 | TAT | 32 |

Identity scores determined by comparison of full-length sequences, "end-gaps" are not penalised. Alanine aminotransferase is the most similar to human AGT and is shaded.

Firstly, the clustering relationships of the homologous sequences was determined using the Pileup program which uses progressive pairwise alignments. The resulting tree can be seen in figure 43. The paralogue sequence, alanine aminotransferase, is the most divergent sequence in the tree, suggesting that the other homologues are orthologues to AGT. If this were true, the clustering relationships would be expected to follow the taxonomic distribution. However the three kingdoms do not always cluster together, so some of the sequences may be paralogous. Therefore the classification between orthologues and paralogues is conflicting. Notwithstanding the above, as different cellular (but not enzymatic) roles for AGT have been suggested for mammals and plants, orthologues could be highly divergent due to recent diversifying selection pressure. As, correct categorisation of orthologues and paralogues is very important for the purposes of this study, further tests to identify orthologues were designed.

5.2.5.1 Test 1

- **To qualify as an orthologue, a homologue must have a higher percentage identity to human AGT than the human paralogue most identical to human AGT.**

To this end, the GAP pairwise alignment program (using the algorithm of Needleman and Wunsch¹¹²) was used to determine the percentage identities of human paralogues to AGT. Gap creation penalties were set at 2 as were gap extension penalties, as this generated the best alignments. Higher gap creation penalties resulted in failed alignments. The percentage identity scores are tabulated in table 16. The paralogue with the highest identity to human AGT is alanine aminotransferase, (shaded in table 16) being 36 % identical. Therefore, the criterion used in the first Orthologue test was that its identity score to human AGT should be higher than 36 %. Orthologue identity scores were measured

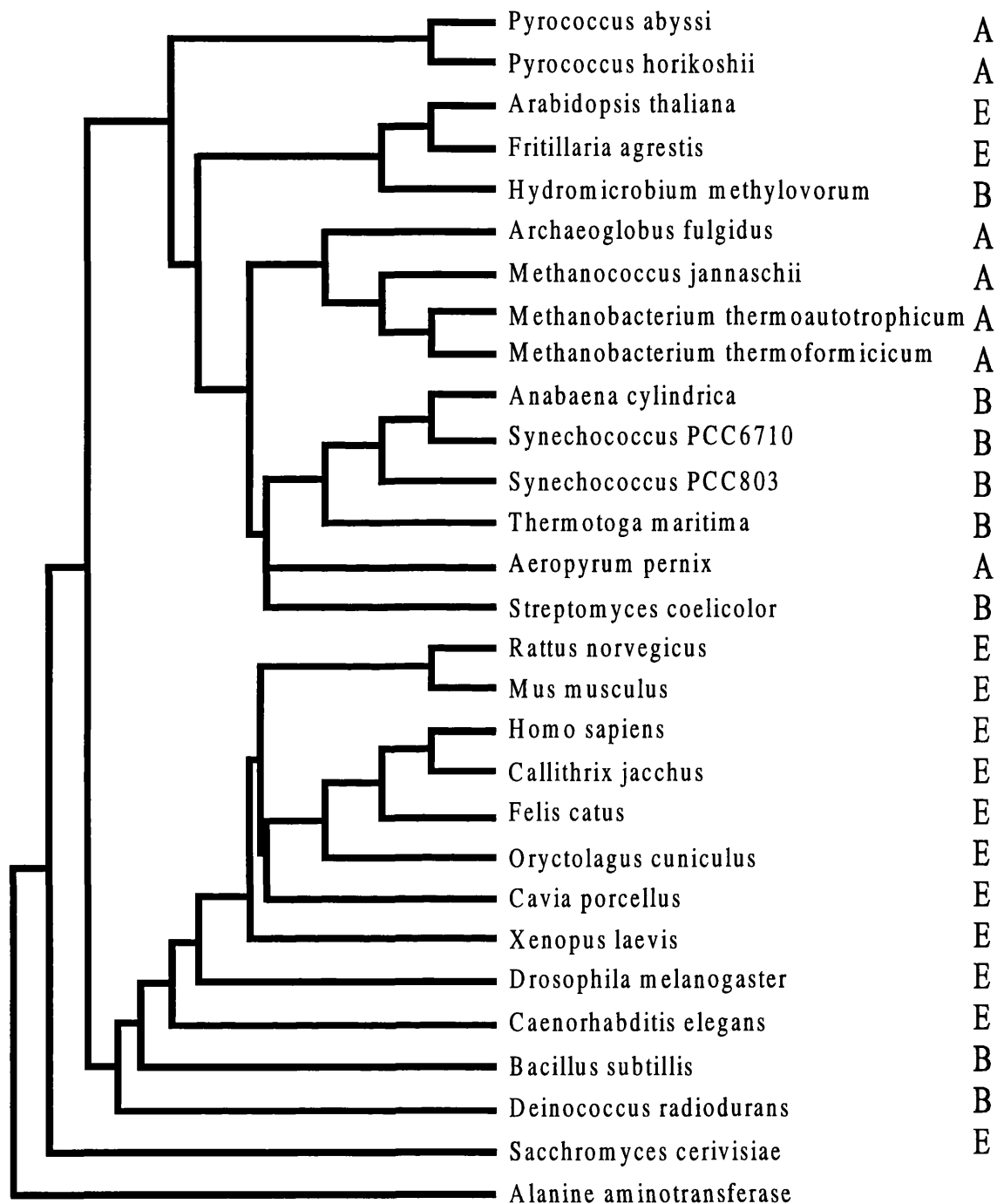


Figure 43 - Clustering relationships of AGT homologues

All sequences returned from the homologue search (Section 5.2.1 and table 12) are phylogenetically analysed using Pileup. Branch lengths are not to scale. The kingdom of life to which each species belongs is noted in the right hand column using these abbreviations: A = Archaeal, B = Bacterial (Eubacterial) and E = Eukaryotic. Alanine aminotransferase was returned from the paralogue search (Section 5.2.2 and table 13).

with the default settings of a gap creation penalty of 8 and a gap extension penalty of 2; these conditions were used to make the test stringent. By this test all the vertebrate, the *D. melanogaster* and the *C. elegans* sequences were classified as orthologues, but the plant, fungal, bacterial and archaeal homologues failed the test.

5.2.5.2 Test 2

Test one was judged to be very stringent, especially as the gap creation penalties differed between the orthologue and paralogue comparisons. Therefore, a second test was applied, whereby the potential orthologue was compared directly to human paralogues at a gap penalty of 2:

- **To be cautiously treated as an orthologue a homologue must have a higher percentage identity to human AGT than it does to any of the other human paralogues**

See appendix table A2 (page 217) for the test scores. This test resulted in the categorisation of both the plant and some of the bacterial and archaea homologues as putative orthologues.

When the homologues classified as putative orthologues are phylogenetically analysed, the kingdoms of life cluster together much more (compare figure 43 and figure 44). The exceptions are the archaea sequence from *Aeropyrum pernix* which clusters with the bacteria and the plant sequences (*Arabidopsis thaliana* and *Fritillaria agrestis*) which do not cluster with the other eukaryotic sequences but instead seem to have more in common with the archaeal and bacterial sequences. As mentioned above, this could indicate that the plant AGTs perform a different role to that of the other eukaryotic AGTs.

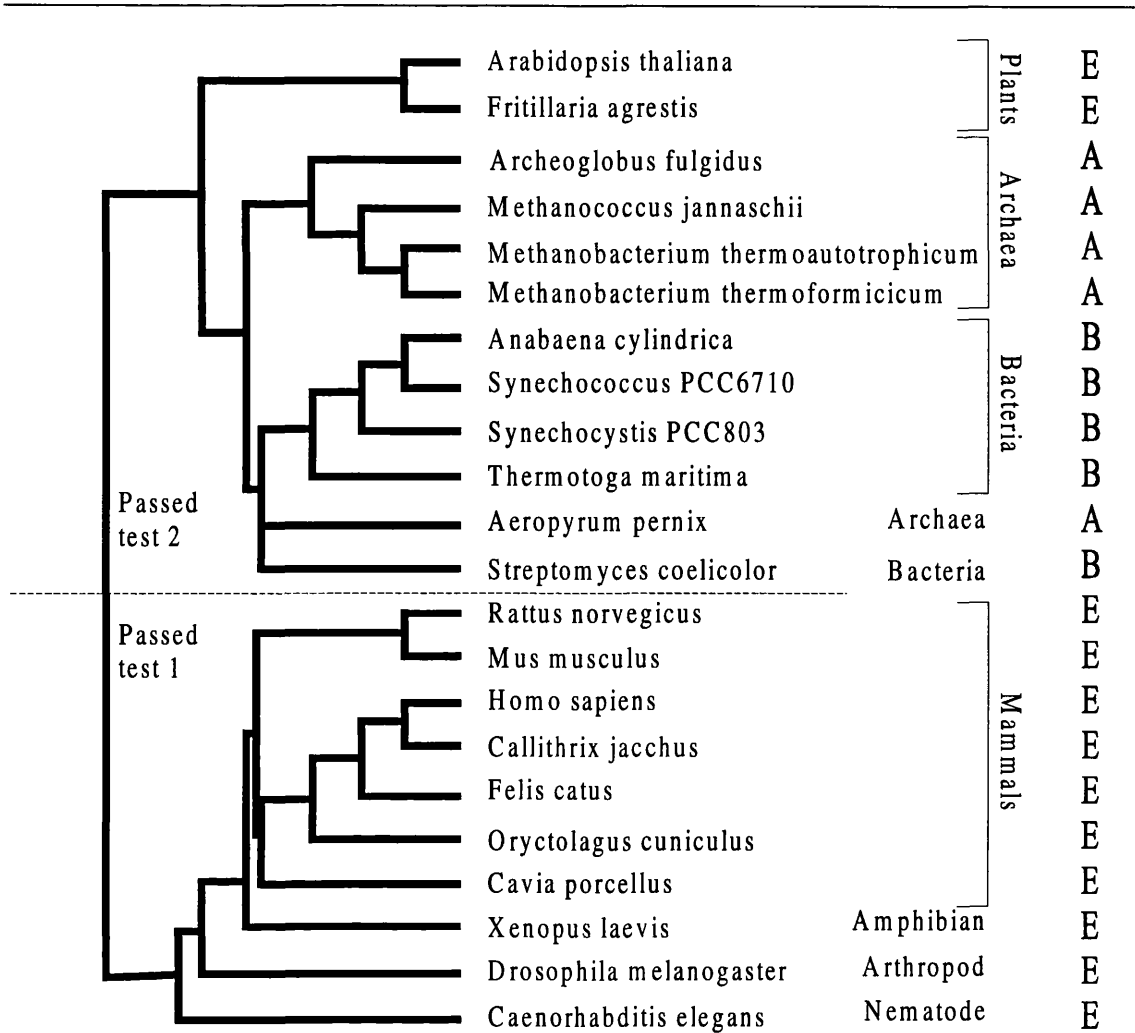


Figure 44 - Clustering relationships of the putative AGT orthologues

All putative orthologue sequences identified from tests 1 and 2 are phylogenetically analysed. Above the dotted line are the putative orthologues that passed test 2 and below the dotted line are the putative orthologues that passed test 1. Branch lengths are not to scale. The kingdom of life to which each species belongs is noted in the right hand column using these abbreviations: A = Archaeal, B = Bacterial (Eubacterial) and E = Eukaryotic. The taxonomic groups for each of the putative eukaryotic are indicated.

5.2.6 Targeting information of eukaryotic AGT orthologues

The putative targeting sequences of the eukaryotic putative AGT orthologues are listed in table 17. *C. elegans* has an N-terminal sequence, between two putative translation start sites, that may be a MTS. This sequence contains 2 arginine residues and no negatively charged amino acids; however, it also contains three helix breakers and no obvious cleavage site. None of the other putative orthologues include an easily identifiable MTS. In contrast, all of the putative orthologues have C-terminal tripeptides that may potentially be PTS1s.

The PTS1s of the AGT orthologues were referred to in table 11, this information is summarised in table 17. The C-terminal tripeptides of the *C. elegans* and *D. melanogaster* sequences, SKI, targets CAT to tobacco glyoxysomes (table 11). The putative plant AGT orthologues have SRI as a C-terminal tripeptide. Although no information is available on this tripeptide (table 11), the residues are often found in these positions in tripeptides that have some characteristics of PTS1s (table 12).

Therefore, the plant AGT orthologues may well be peroxisomally targeted. Both SRI and SKI have 2 frequent residues in the -3 and -2 positions but a less popular residue in the -1 position. This could mean they are species or protein context dependent. SKI is definitely species context dependent as it seems to work in plants, less well in mammalian cells and not at all in yeast. Its inefficient targeting of rat epoxide hydrolase could mean that it is also inefficient at targeting *C. elegans* and *D. melanogaster* AGT and the protein is peroxisomal and cytosolic in these species as it is the guinea pig⁴⁹. The putative *C. elegans* AGT sequence has 22 N-terminal amino acids between two methionines. Helical wheel analysis (figure 45) of these amino acids show a cluster of 3 positive residues at one side of the helical wheel. Therefore this region of the putative *C. elegans* AGT polypeptide could be a MTS.

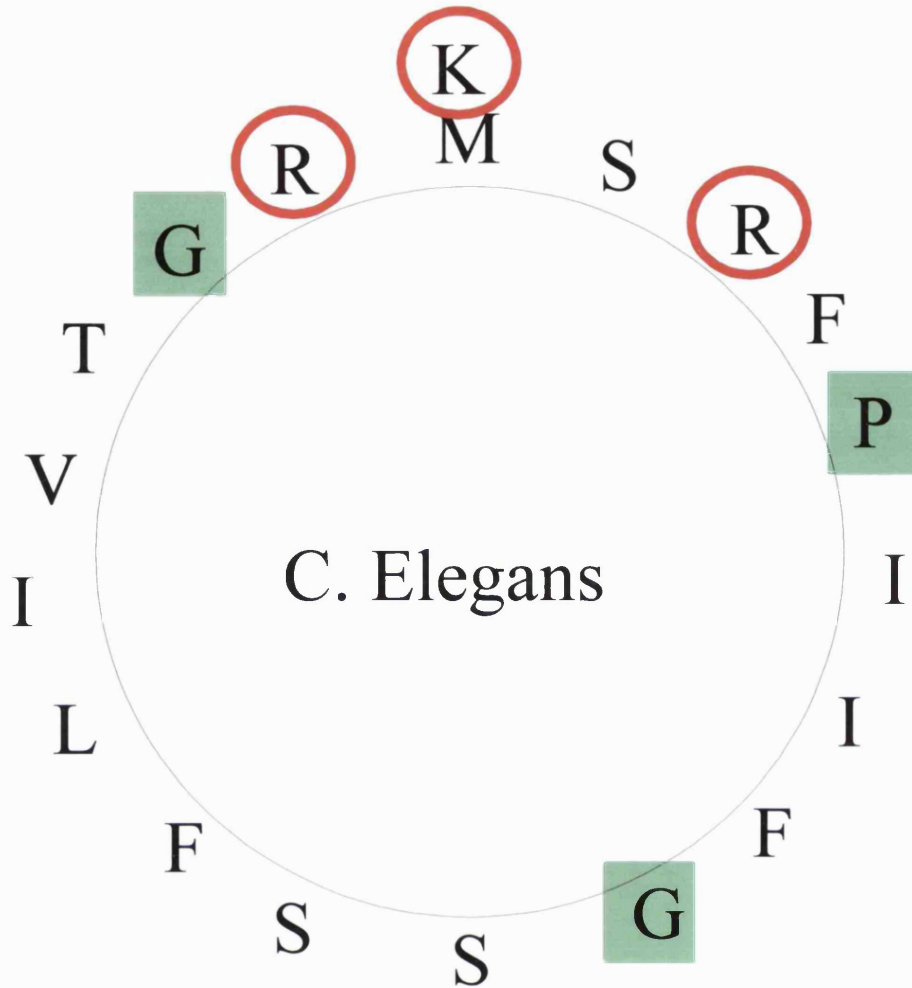


Figure 45 - Helical wheel analysis of the N-terminal amino acids of putative

***C. elegans* AGT**

The amino acid sequences are arranged in a helical format (clockwise, 1st, 12th, 5th, 16th, 9th, 2nd, 13th, 6th, 17th, 10th, 3rd, 14th, 7th, 18th, 11th, 4th, 15th, 8th).

Positively charged residues (K, R and H) are ringed in red, helix breaking residues (G and P) are shaded in green.

If this is true, the predicted subcellular distribution of AGT in *C. elegans* is mitochondrial and peroxisomal, due to the existence of a potential MTS and PTS1 in the deduced amino acid sequence. The distribution of AGT in *D. melanogaster*, *A. thaliana* and *F. agrestis* could be predicted to be peroxisomal, due to the presence of a potential PTS1 but no apparent MTS, in the putative amino acid sequences. As mentioned before (chapter 4) the xenopus and mammalian MTSs have no obvious similarity and this is also the case for the possible MTS of *C. elegans* (table 17). This could mean that the MTSs have arisen independently in separate lineages. Accordingly, a theory for the ancestral distribution of AGT could be postulated whereby peroxisomal AGT is the ancestral state for both plants and animals. Theoretical scenarios for the evolution of extant AGT distributions in these species from a peroxisomal ancestor can be seen in figure 46.

The C-terminal tripeptides of the AGTs are much more variable than would be expected given the invariability of the PTS1s of other orthologous proteins e.g. D-amino acid oxidase (table 14). The variability of the AGT PTS1 may be due to:

- less selective pressure on the PTS1 of AGT due to the additional targeting information thought to be involved in its import or
- episodic selection pressure, similar to the pressure acting on the MTS of AGT in primates (Chapter 3).

Table 17 - The targeting information of C-terminal tripeptides of AGT orthologues

| Orthologue | MTS | C-terminal tripeptide | Tripeptide PTS1 status | AGT distribution |
|------------------------|----------------------------|-----------------------|--|--|
| Rabbit | | SQL | Binds human Pex5p but not <i>S. cerevisiae</i> or <i>N. tabacum</i> Pex5p. Can target GFP to peroxisomes in HeLa cells | Peroxisomal |
| Guinea pig | | HRL | Binds <i>S. cerevisiae</i> Pex5p. Has very weak interaction with human Pex5p. | Peroxisomal and cytosolic |
| Human | | KKL | Can target MDH3 to peroxisomes in <i>S. cerevisiae</i> . | Peroxisomal |
| Marmoset | MFQALAKASAALGPRAAGWVRT(M) | KKL | Can target MDH3 to peroxisomes in <i>S. cerevisiae</i> . | Mitochondrial and peroxisomal |
| Cat | MFRALARASATLGPQVAGWART(M) | NKL | Can target MDH3 to peroxisomes in <i>S. cerevisiae</i> . | Mitochondrial |
| Rat | MFRMLAKASVTLGSRAASWVRN(M) | NKL | Can target MDH3 to peroxisomes in <i>S. cerevisiae</i> . | Mitochondrial and peroxisomal |
| Mouse | MFRMLAKASVTLGSRAAGWVRT(M) | NKL | Can target MDH3 to peroxisomes in <i>S. cerevisiae</i> . | Mitochondrial and peroxisomal |
| Xenopus | MQGSVRSISSLLCAARLSAPVRI(M) | KKM | Is not sufficient to take human AGT to peroxisomes in COS cells. | Mitochondrial and cytosolic |
| <i>D. melanogaster</i> | | SKI | Does not bind human or <i>N. tabacum</i> Pex5p. Targets catalase to <i>S. cerevisiae</i> peroxisomes and CAT to <i>N. tabacum</i> peroxisomes but is not sufficient to target luciferase in COS cells. Targets its natural protein rat epoxide hydrolase very inefficiently, mainly cytosolic. | Unknown ?peroxisomal? |
| <i>C. elegans</i> | MISTRFLRPSVSIFGFIKSS(M) | SKI | Does not bind human or <i>N. tabacum</i> Pex5p. Targets catalase to <i>S. cerevisiae</i> peroxisomes and CAT to <i>N. tabacum</i> peroxisomes but is not sufficient to target luciferase to COS cells. Targets its natural protein rat epoxide hydrolase very inefficiently, mainly cytosolic. | Unknown ?mitochondrial and peroxisomal? |
| <i>A. thaliana</i> | | SRI | No Information | Unknown ?peroxisomal? |
| <i>F. agrestis</i> | | SRI | No Information | Unknown ?peroxisomal? |

All data from table 11 (pages 147-151) MDH3 = malate dehydrogenase, GFP = Green fluorescent protein, CAT = Chloramphenicol acyltransferase

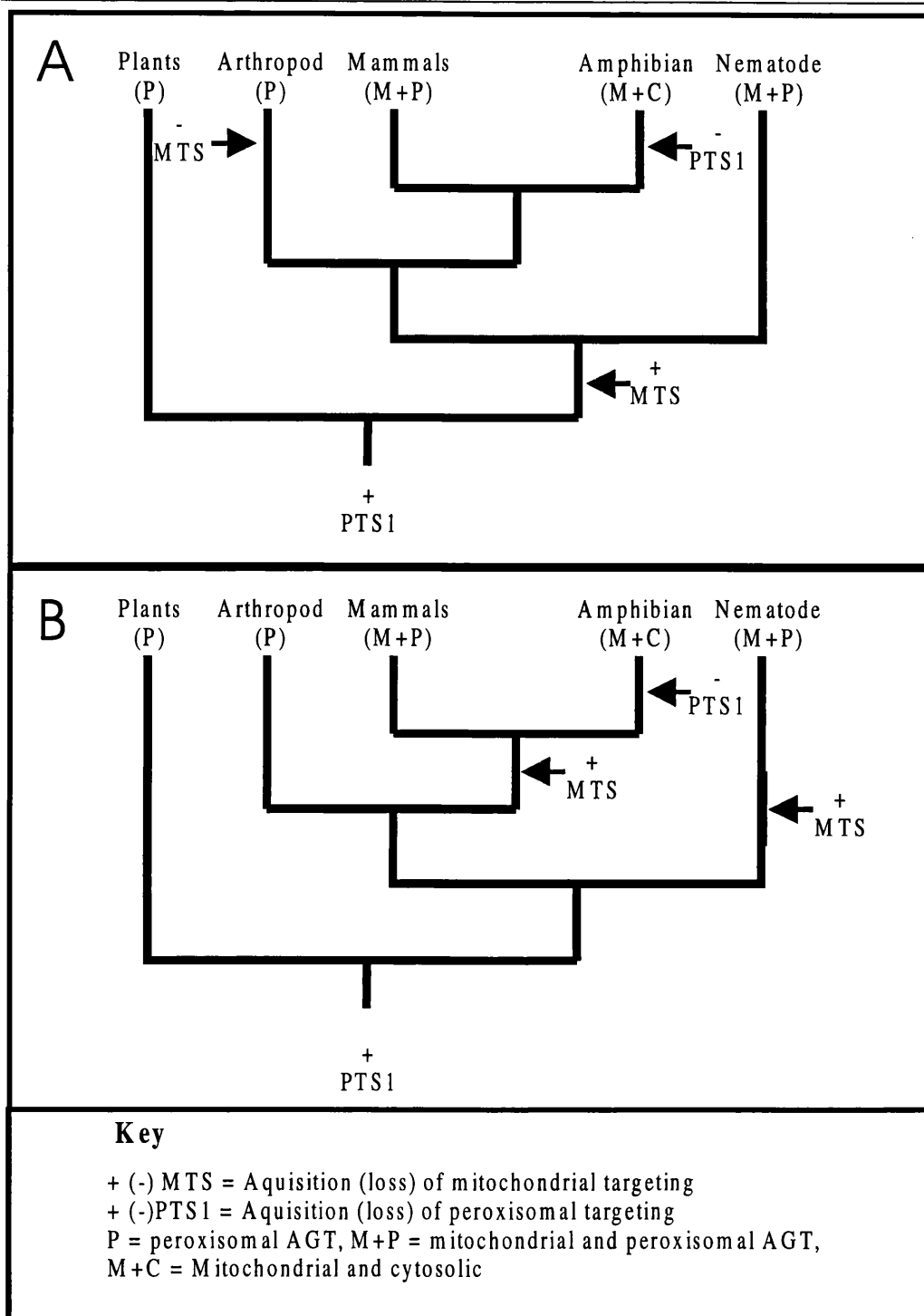


Figure 46 - Theoretical evolution of AGT targeting

This theoretical scenarios describing explanations for real and predicted extant AGT distribution. Mammals are grouped with a mitochondrial and peroxisomal AGT distribution as the most probable ancestral distribution in mammals. The arrows denote evolutionary events. The two scenarios are equally parsimonious, each suggesting the occurrence of 3 evolutionary events, but in A the MTS arises once and in B the MTS arises on 2 separate occasions. The phylogenetic relationships are taken from "The tree of life" (<http://phylogeny.arizona.edu/tree/>).

5.2.8 Alignment of AGT orthologues

The putative AGT orthologues are weighted heavily in favour of eukaryotic sequences, especially vertebrates and even more especially mammals; fungi are completely absent. To reduce some of the sampling bias, the orthologues were grouped into taxonomic categories (for categories see figure 44), for subsequent analysis. Some categories contained several sequences (i.e. mammals, plants, archaea and bacteria) and some only one (i.e. amphibian, nematode, arthropod). This strategy may decrease the sampling bias in favour of mammals but it in no way eliminates the problem all together. This tactic had the added advantage of allowing the data to be put in a manageable format to inspect sequence alignments across the three kingdoms of life. A consensus sequence, generated by simple majority vote was used to represent each group. The archaeal sequences were so divergent, a meaningful consensus could not be generated from them and so they were left out of the analysis. Figure 47 shows the alignment of the remaining consensus sequences.

One expectation of this alignment is that the additional targeting information, mentioned in Chapter 1, Chapter 4 and in this Chapter, may be discernible. This information might be expected to be within the mammal, arthropod, nematode and plant sequence; but not in the amphibian and bacteria sequence. As the former four are hypothesised to have protein-context dependent PTS1s, whilst xenopus AGT has been shown not to have a PTS1 or the additional targeting information (Chapter 5) and bacteria, which obviously do not have peroxisomes, have no need of targeting information. Disappointingly, there is no very obvious region that has this pattern of conservation, although the candidate region suggested in figure 26 (page 107) has a residue position that shows a similar pattern. The residue in the second position in this region is small, neutral and hydrophobic (i.e. A,V,I,L) in the mammal, nematode,

MFRXLAKASXTLGX.RAAGW...VRTMASHQLL.VAPPEALLKPLSIPN
 MQGS.VR.SISSLLCAARLSAPVRIMSSLAT..IPPPSALQRPLNVPO
 MIST.RFLRPSVSIFGFGIKSS....MSS.....RAPPKALLQDMVVPF
ME.....VPPPLVLKRPLYVPS
MDYXY....GPRX..HLFV...
MDDKMLMIPGPT.....

RLLLGPGPSNLAPRVLAAGGLQMI GHMHKEMX.QIMDEIKOGIQYVFOT
 RLMLGPGPSNVPPRIQAAGGLQLIGHMHEMF.QIMDDIKOGIQYAFOT
 RQLFGPGPSNMADSI AETQSRNLLGHLHPE.FVQIMADVRLGLQYVFKT
 KTLMGPGPSNCSHRVLEAMSNEVLGHMHEC.LQIMDEVKEGIKYIFOT
PGPVNIPEFVIRAMN.RNNEDYRXPAIPALTKTLLEDVKKXFKT
FVPEAVLLALAKHPIGHRSGEFSQIMA.EXTXNLKWLHOTEND.

RNPLTLVISGSGHCA.LEXALFNLLEPGDSFLV GANGIWGOR.AAEIGER
 KNNLTFAVSGSGHCA.METAIFNVVEKGDVVLVAVKGIWGER.AGDIAER
 DNKYTFAVSGTGHS.GMECAMVNLEPGDKFLVVEIGLWGOR.AADLANR
 LNDATMCISGAGHS.GMEAALCNLIEDGDVVMGITGVWHR.AGDMARR
 T.G.TPFXXPTTGTGAWESALTNTLSPGD XIVSFLIGQFSLLWI.DQQXR
 VLMLT..ASGTGAX...EAXIINFLSXGDRVLV GXNGKFGXRWV.EVGXA

IGARVHPMIK..DPGSHYTLQEEVEGLAQ..HKE.VLLFLTHGESSTGVL
 IGADVRYVSK..PVGEAFTLKDVEKALAE..HKE.SLFFITHGESSGVV
 MGIEVKKITA..PQGQAVFVEDIRKAIAD..YKE.NLVFVCQGDSSTGVA
 YGAEVHYVEASFG.RALSH.EEITFAF.EG.HRP.KVFFIAQGDSSTGII
 LXFVVDVXESXWGQGANLXXLASKLX.XDXXHTI.KAXCIVHNETATGVT
 FGLXVEXIXAEWGOPLDE..DDFKQLLXADXXKEXKAVIITHSETSTGVX

QP..LDGFG.ELCHRYXCLLLV.DSVASLGGVPIY.MDQQGIDILYSGSQ
 QP..LDGLG.DLCHRYNCLLLV.DSVASLGGAPIY.MDKQIDILYSGSQ
 QP..LETIG.DACREHGALFLV.DTVASLGGTP.FAADDLKVDCVYSATQ
 QQNIRE.LG.ELCRKYDC.FLIVDTVASLGGTE.FLMDEWKVDVAYTGSQ
 NXXXAVRXLLDXHXHF..ALLVDGVSSICALD.FRMDEWGXDVALTGSQ
 NDLVAINXHVKXHX..ALIIVDAVTSLGAXXV.PIDELGLDVVASGSQ

*
 KVLNAPPGTSL.....ISFSDKA.KKKXYSRKTKPKS FYLDIKYLANLW
 KVLNAPPGTAF.....ISFSEAA.SKKMFGRKTKPFSLYVDINWLANYW
 KVLNAPPGLAP.....ISFSORAMEKIRNRKQVA.SFYFDAIELGNYW
 KS..LGGPA.GLTP...ISFSKRALTRIRKRK.TPKVYYFDILLIGQYW
 KA..LSXF.TGX.GXXCASP..KALE.AX..KTXXSXXVFFDWN DYLKFY
 KGYMIPPGLAFVS....VSX..KAWE.A.YXTAXLFX.XYLDLXKYRKST

Positively charged region

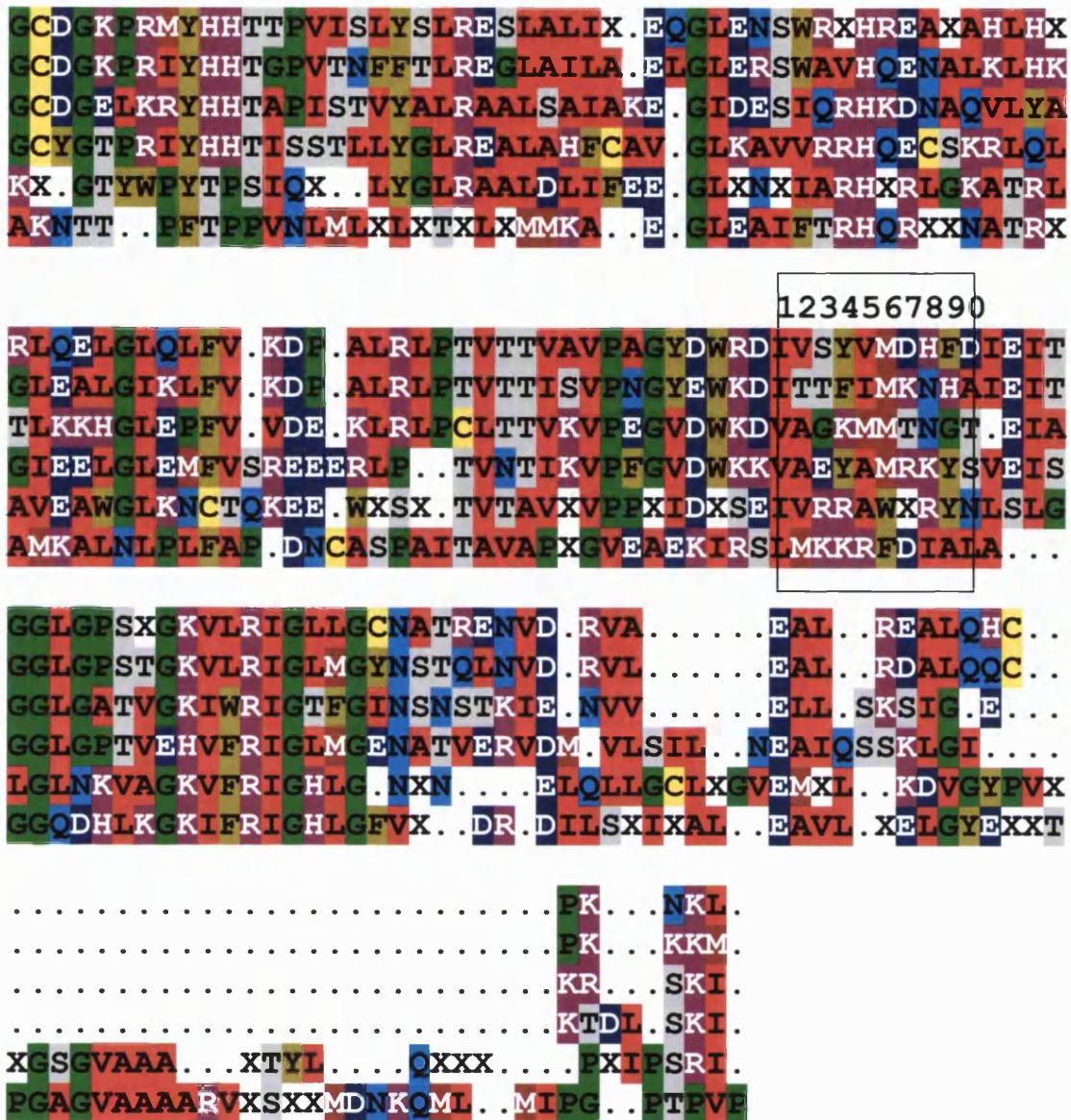


Figure 47- Multiple alignment of consensus orthologue sequences

| | | |
|---------|--------------------------------------|--------------------|
| K R H | Leucine, Arginine, Histidine | Positively charged |
| D E | Aspartic acid, Glutamic Acid | Negatively charged |
| N Q | Asparagine, Glutamine | Amides |
| F Y W | Phenylalanine, Tyrosine, Tryptophan | Large aromatic |
| A V I L | Alanine, Valine, Isoleucine, Leucine | Small neutral |
| S T | Serine, Threonine | Hydroxyl group |
| G P | Glycine, Proline | Helix Breakers |
| C | Cysteine | |
| M | Methionine | |

Order of phylogenetic groups : Mammals, amphibian, nematode, arthropod, plants and bacteria . Xs represent positions where there was no residue was in the majority. Boxes are around the suggested region of additional targeting information from Chapter 4, a region of positively charged residues and the pyridoxal binding site.

arthropod and plant sequences whilst the xenopus sequence is serine in this position and the bacteria sequence is methionine.

The pyridoxal phosphate-binding site (starred and boxed in figure 47), is conserved in all the sequences, suggesting that the enzymatic function of AGT is similar in the enzyme encoded by all of the sequences.

As can also be seen in figure 47, there is a region of the sequences (boxed in figure 47) that has a disproportionate amount of positive residues. This region has previously been suggested to be a cryptic nuclear targeting sequence¹²⁰.

The lack of similarity between the MTSs of the AGT orthologues is highlighted in figure 47, as there is little similarity of amino acid types. This further supports the theory the MTSs of the AGT orthologues arose separately and suggests that figure 46B is closer to the evolutionary past of AGT than figure 46A.

6. Comparative Analysis of AGT and Life Style Traits

6.1 Introduction

Central to the study of the evolution of AGT intracellular targeting is the assumption that the subcellular distribution of hepatic AGT protein changes due to the influence of selective pressure. The best evidence that the change in distribution of AGT is not a selectively neutral event that occurs many times due to genetic drift comes from the study of the human disease PH1. PH1 was described in Chapter 1 and in that chapter it was observed that, at least in humans, the wrong subcellular distribution of AGT can have serious, even fatal, consequences⁴⁵.

As also described in Chapter 1, a relationship between AGT distribution and natural diet has been noted. Thus carnivores tend to have mitochondrial AGT, herbivores tend to have peroxisomal AGT and omnivores tend to have both. In primates, evidence suggests that in recent history there has been selection pressure to lose or diminish mitochondrial AGT targeting and that this pressure was a change in diet (Chapter 3).

Also, an inverse correlation of aminotransferase activity to adult body size has been reported¹¹⁴, (i.e. small animals tend to have high aminotransferase activity and large animal tend to have low aminotransferase activity¹¹⁴). The authors proposed that aminotransferase activity may have a role in the stimulation of heat production.

However, the relationships between AGT variables and lifestyle traits have never been formally statistically analysed. Closely related species will tend to have similar traits due to phylogenetic niche conservation and lack of time since divergence from their common ancestor⁸⁰; this is called phylogenetic inference. Therefore, the apparent

relationship between AGT distribution and diet could be due to phylogenetic inference and sampling bias. Clearly, a statistical test that controls for phylogeny is needed.

In order to analyse the relationships between AGT and life style variables, a maximum likelihood based method was used^{74;80}. This was one of the relatively new methods that have been developed¹¹⁵ to study adaptation of species traits. These analyses will help discover if there is evidence for adaptation in AGT evolution, elsewhere than in primates.

6.2 Results and Discussion

6.2.1 Acquisition the data set

The data set used for this analysis was compiled using published sources. However, some of the information was not as complete or accurate as would be desired, for example some information on diet was ambiguous and some of the AGT activity data showed inter-sample variation. The traits studied are shown in table 18. Due to the requirement of the method for continuous data, the discrete data was scored in a continuous fashion (table 18). The data on total AGT enzymatic activity (not corrected for crossover activity of glutamate:glyoxylate aminotransferase¹¹⁶,

Table 18 - AGT and life style traits examined by comparative analysis

| Trait | Discrete or continuous | Categories if treated as discrete |
|-------------------------------------|-------------------------------|--|
| AGT distribution | Discrete* | 1 = mainly mitochondrial 2 = mitochondrial and peroxisomal 3 = peroxisomal |
| AGT catalytic activity | Continuous | |
| Level of AGT immunoreactive protein | Discrete* | 0 = none or very low 1 = low 2 = medium 3 = high |
| Diet | Discrete* | 1 = predominantly carnivorous 2 = omnivorous 3 = predominantly herbivorous |
| Body size | Continuous | |
| Gestation length | Continuous | |

*Actually, all the listed traits are continuous but the restrictions of the methods used to measure them or the incompleteness of the information made their categorisation necessary.

has only been published in partial form⁴³. Because of doubt over the quality of the liver samples being assayed and variability of the intra species samples, the data was classified, by the authors, into high, medium or low. In this study, the problem is dealt with in another way. In the original data (Danpure, personal communication) there are several measurements for each species. The inter-sample variation is thought to be due to poor sample quality. As this will negatively affect the AGT activity assay, the lowest values for each species are more likely to be erroneous. It is difficult to understand how sample quality could artificially elevate the result of the activity assay. Consequently the highest value for each species was added to the data set. As immunoreactivity is more robust than enzymatic activity, data on the levels of AGT immunoreactivity were added to the analysis.

Other life style traits as well as diet, were added to the analysis because they were expected to have relationships to AGT distribution of less significance than that of diet to AGT distribution. As seen in Chapter 3, body size can be related to diet but its relationship to AGT is thought to be indirect. Therefore body size and gestation length were added as forms of negative controls to give more confidence in significant results obtained from comparison of the other traits. There are many other parameters that would have been interesting to add to the data set. Possible candidates were body temperature, basic metabolic rate and nocturnal/diurnal lifestyle. However data on these traits for all species in the analysis were not readily available. The full data set used in this test is tabulated in the appendix, table A4.

The data were summarised in the form of scatter graphs (figures 48-52) for the more important variables, namely AGT subcellular distribution and AGT activity. Figure 48 shows

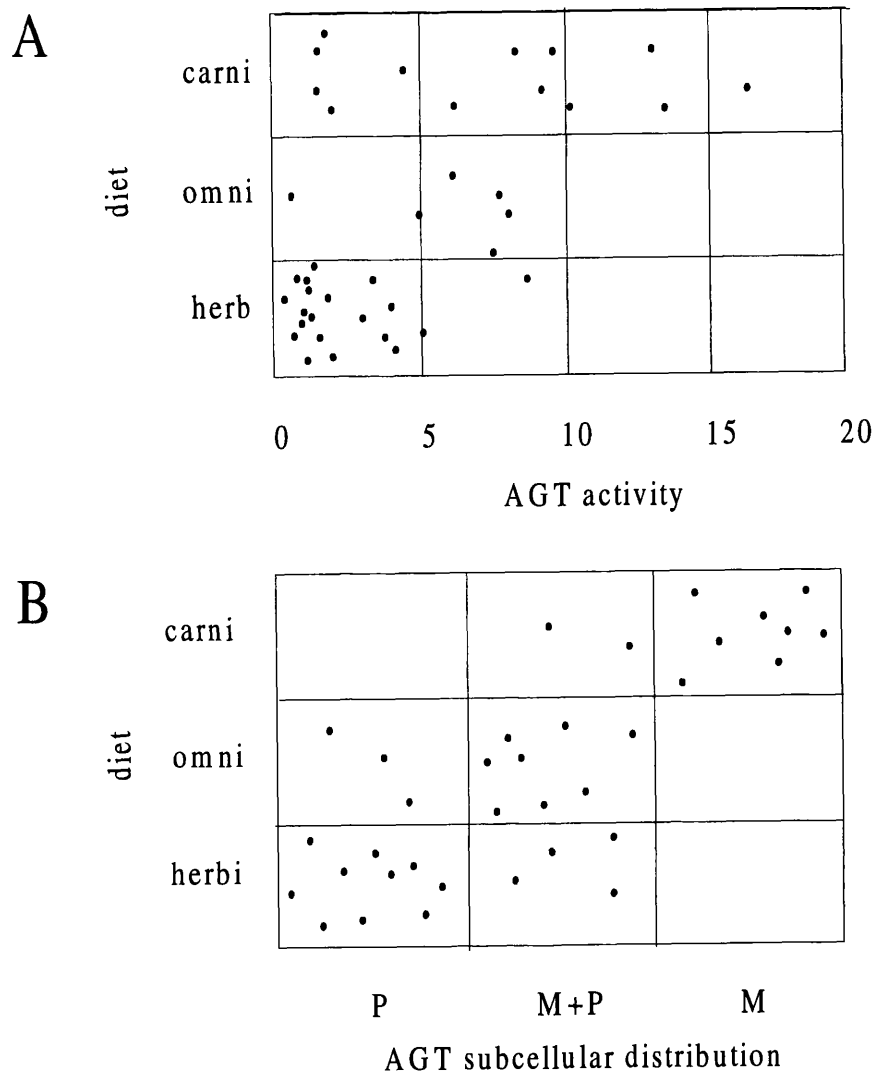


Figure 48 - The relationships between diet and AGT activity or subcellular distribution

The natural diet of each species is plotted against AGT catalytic activity (A) in $\mu\text{mol}/\text{hour}/\text{mg}$ of total protein or AGT subcellular distribution (B). For all data see appendix table A4. Primary data sources are Danpure et al (1994)⁴³ for AGT activity and distribution and Nowak (1991)¹⁰⁰ for diet. carni = carnivore, omni = omnivore, herb = herbivore. M =mitochondrial AGT, M+P = mitochondrial and peroxisomal AGT, P = peroxisomal AGT. AGT activity is measured as total activity in μmol of pyruvate formed / hour / mg total liver protein

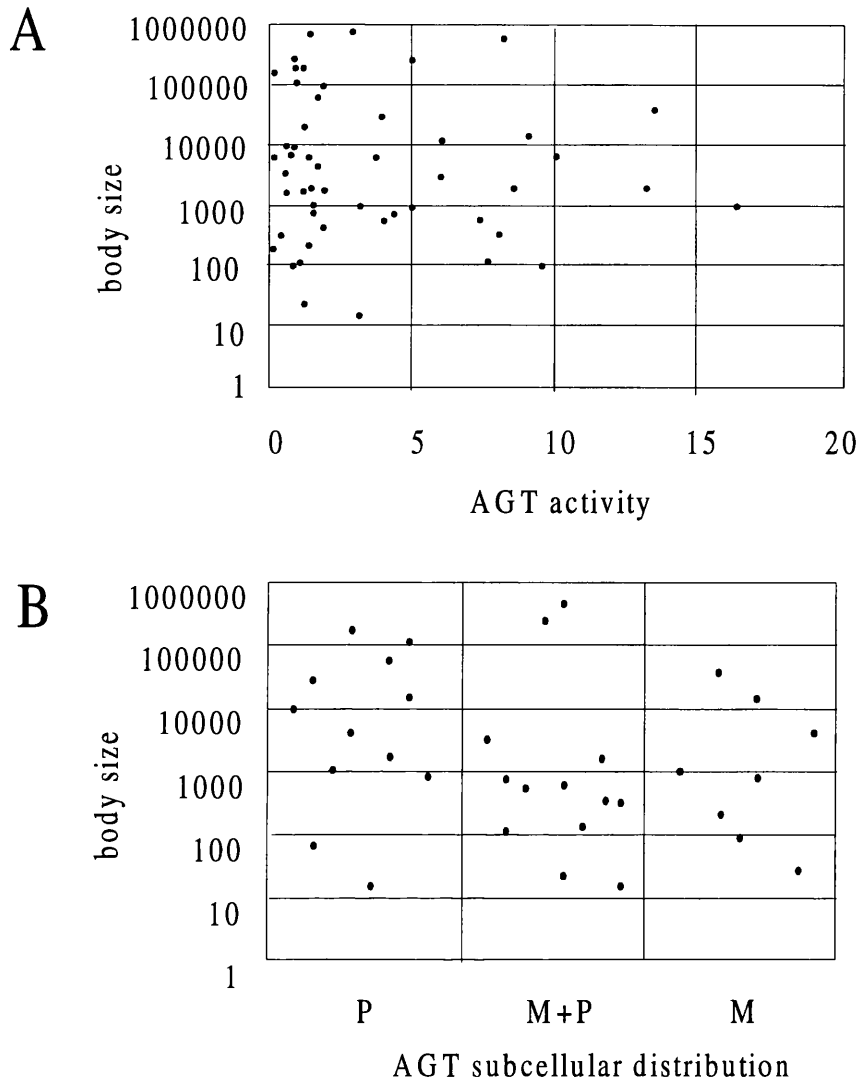


Figure 49 - The relationships between body size and AGT activity or subcellular distribution

The body size of each species is plotted against AGT catalytic activity (A) or AGT subcellular distribution (B). Body sizes are plotted on log scales. For all data see appendix table A4. Primary data sources are Danpure et al (1994)⁴³ for AGT activity and distribution and Nowak (1991)¹⁰⁰ for body size. M = mitochondrial, M+P = mitochondrial and peroxisomal, P = peroxisomal.

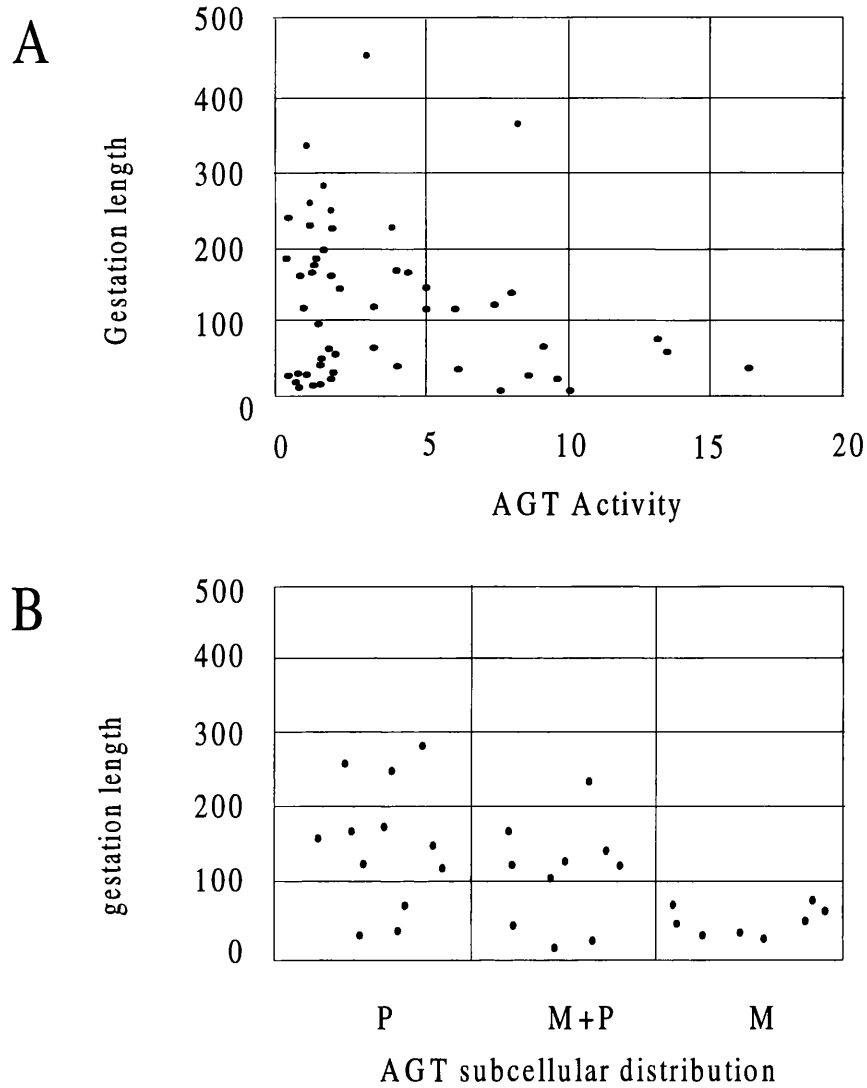


Figure 50 - The relationships between length of gestation and AGT activity or subcellular distribution

The gestation length of each species is plotted against AGT catalytic activity (A) or AGT subcellular distribution (B). For all data see appendix table A4. Primary data sources are Danpure et al (1994)⁴³ for AGT activity and distribution and Nowak (1991)¹⁰⁰ for gestation length. M = mitochondrial, M+P = mitochondrial and peroxisomal, P = peroxisomal.

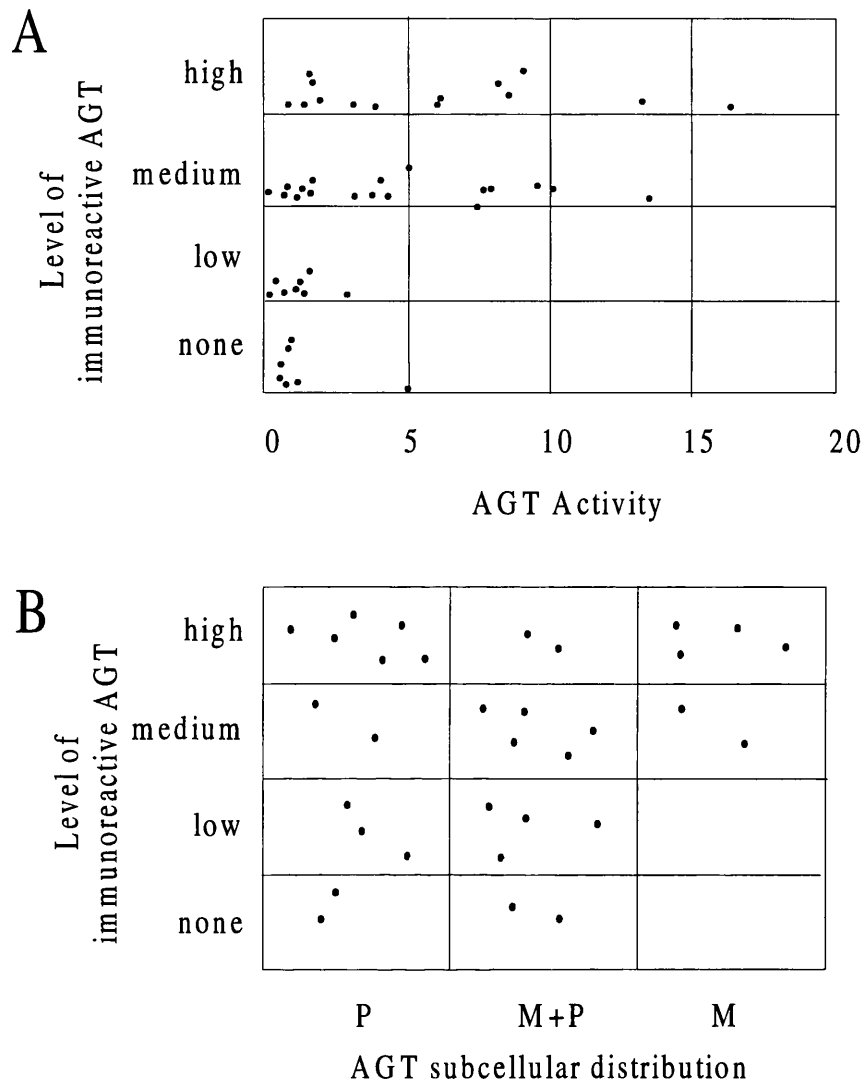


Figure 51 - The relationships between the levels of AGT immunoreactive protein and AGT activity or subcellular distribution

The level of immunoreactive AGT protein of each species is plotted against AGT catalytic activity (A) or AGT subcellular distribution (B). For all data see appendix table A4. Primary data source is Danpure et al (1994) ⁴³.

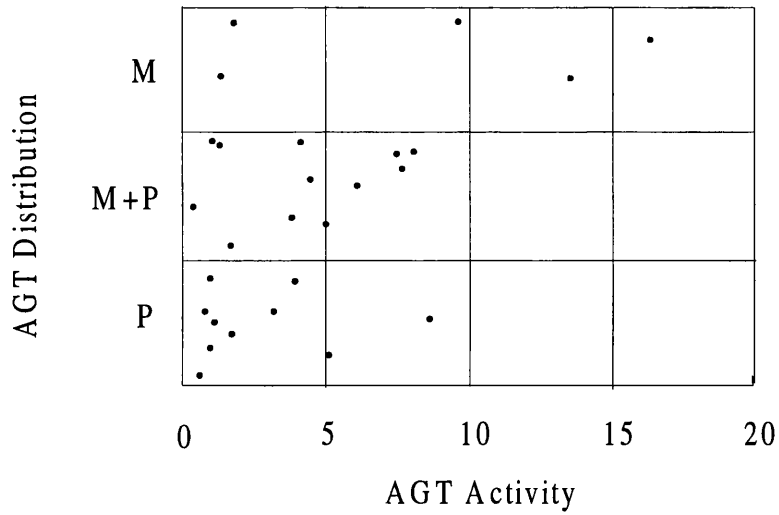


Figure 52 - The relationship between AGT activity and subcellular distribution

AGT catalytic activity is plotted against AGT subcellular distribution. For all data see appendix table A4. Primary data source is Danpure et al (1994)⁴³.

the relationship between these variables and diet. The, previously noted ⁴⁵, relationship between AGT subcellular distribution and diet is clearly visible (figure 48B). Also, a possible relationship between diet and AGT activity can be seen (figure 48A), although less pronounced. Figure 49, shows the relationships between the two AGT variables and body size. Body size does not seem to be correlated to any variable, given the distributions pictured in figure 49. Figure 50 shows the relationships between gestation lengths and AGT activity and subcellular distribution, again there seems to be no correlation. Figure 51, shows the relationships between levels of AGT immunoreactive protein and AGT activity or subcellular distribution. There does seem to be a positive relationship between levels of AGT immunoreactive protein and AGT activity (figure 51A). However a stronger relationship was expected, as the level of AGT protein must be a factor in the total AGT activity. The distribution suggests that there may be some figures in the data set from samples in which the enzymatic AGT had decayed. These would be the points in the top left of the scatter which have high levels of immunoreactive AGT but low AGT activity. There may also be a trend between level of AGT immunoreactive protein and AGT subcellular distribution. There are no species with mitochondrial AGT and low or very low levels of AGT immunoreactive protein. Figure 52, shows the relationship between AGT subcellular distribution and activity. Given that these two variables both show a relationship with species diet, a correlation between them would be expected and a trend can be seen. Again there are some points of low AGT activity which do not fit the trend.

6.2.2 The Comparative analysis test controlling for phylogeny

Due to the difficulty of finding a phylogeny that included all the species studied, a composite tree was created (Appendix figure A6) from publications on the phylogenetic analysis of mitochondrial DNA^{50-55;57;58}. As this process is prone to error and no phylogeny is totally accurate, the analysis was repeated using a taxonomic tree (Appendix figure A7), in which the soft (non-burificating) nodes were arbitrarily resolved. As these trees were quite different, this approach should control for errors resulting from incorrect phylogeny.

An important parameter of the test is the phylogeny scaling parameter (λ)⁷⁴. It indicates how dependent the trait values for each species, is on the phylogeny (i.e. could the trait variance be predicted by the shape of the tree). λ can vary between 1 and 0. $\lambda = 0$ indicates that a trait is independent of phylogeny and does not vary in a predictable phylogenetic fashion, whilst $\lambda = 1$ suggests the trait follows phylogeny exactly. In this analysis, λ is fixed to the value optimal (as determined by maximum likelihood) for the null hypothesis (the null hypothesis being, that the species traits are not correlated). This is the most conservative approach as the value of λ biases the test in the favour of the null hypothesis⁷⁴.

For each comparison, two models are used. There is a null hypothesis model, in which the traits are constrained not to co-vary, and a test model, in which the variance of the traits is not constrained. A likelihood value is obtained for each model. These are used to compute the log likelihood difference, which is approximated to the χ^2 distribution, to obtain a p value. This p value is the probability that the test model fits the data better

than the null hypothesis model. The results of the analysis, using the mitochondrial DNA phylogeny (Appendix figure A6), are tabulated in figure 53.

6.2.3 AGT distribution and diet

AGT subcellular distribution was found to be significantly related to diet, in a manner that is independent of phylogeny ($p = 0.00147$). This result supports the dietary hypothesis for AGT evolution⁴⁵ and suggests that the evolution of the subcellular distribution of AGT shows adaptation to diet in mammals as a whole, as well as in primates in particular (Chapter 3). Given that many traits are related to diet, the chronological order of the trait changes is very difficult to determine. However, the subcellular distribution of AGT can be changed by the mutation of just one nucleotide (e.g. chapter 3). Therefore AGT distribution could be a very plastic trait, as opposed to a trait like diet in which, usually, many complicated physiological and behavioural traits must be altered to give rise to a change. Because of this, changes in AGT subcellular distribution are probably a result rather than a cause of dietary change. One group in which the dietary hypothesis does not explain the data is in the ruminants such as sheep and cattle. AGT is peroxisomal in sheep but has very low activity in the sheep, bison and oryx and has undetectable activity in the kudu⁴³. As these mammals are classic herbivores, peroxisomal AGT would be expected to be very important. It has been suggested that ruminants are hosts to bacteria in their foreguts that may breakdown oxalate (and probably its precursors) and hence diminish the need for glyoxylate detoxifying AGT⁴⁵. However, this result is not so surprising, as this study has shown that AGT activity tends to be high in carnivores and low in herbivores (figures 52 and 53 and next section).

| | AGT distribution | AGT activity | Species diet | Body size | Gestation length | Level of AGT IR-protein |
|--------------------------|------------------|--------------|------------------------|-----------|-------------------------|-------------------------|
| AGT distribution | | 0.0799 | <u>0.00147</u> | 0.347 | 0.145 | 0.306 |
| AGT activity | 1.00 | | <u>0.000815</u> | 0.634 | 0.133 | 0.516 |
| Species diet | 0.473 | 0.909 | | 0.682 | 0.223 | 0.218 |
| Body size | 0.092 | 0.530 | 0.801 | | <u>0.0000288</u> | 0.717 |
| Gestation length | 0.674 | 0.725 | 0.931 | 0.600 | | 0.577 |
| Levels of AGT IR-protein | 0.945 | 0.754 | 0.982 | 0.608 | 0.796 | |

Figure 53 - Results of comparative analysis of mammalian species traits

The top right of the grid records p values for the comparisons, to 3 significant figures. The bottom left of the grid records the maximum likelihood estimated value of the phylogeny scaling factor (λ) optimal for the null hypothesis, to 3 significant figures. λ was set at this value for both hypotheses in the likelihood ratio test, which generated the p value. p values bolded and underlined are significant to < 1 % level. IR = Immunoreactive.

6.2.4 AGT activity and diet

AGT activity in different species was very significantly related to diet ($p = 0.000815$). This is a new observation and even more significant than the more established relationship between AGT distribution and species diet. When the scatter graph for AGT activity and diet (figure 48A) is examined, the AGT activities of the carnivores are evenly distributed between 2 and 16 $\mu\text{mols} / \text{hour} / \text{mg protein}$, whilst the AGT activities of the herbivores nearly are all less than 5 $\mu\text{mols} / \text{hour} / \text{mg protein}$. It could be that the lower AGT activities of the carnivores are due to poor sample quality, but there is no way of determining this without more samples. Given the arrangement of traits on the scatter graph, it seems that AGT activity can be high or low for carnivores but is usually low in herbivores. Perhaps high AGT activity levels in the peroxisomes (AGT is peroxisomal in herbivores) are toxic.

It is easy to understand why high AGT activity levels are desirable in carnivores. Carnivores have a high protein diet and gluconeogenesis is important (Chapter 1 - Section 1.7 - page 38). Therefore high levels of gluconeogenic enzymes, of which AGT is one, could be advantageous. However, it is less obvious why high levels of AGT activity might be toxic for herbivores. One possibility is that AGT activity produces hydroxypyruvate through its serine:pyruvate aminotransferase activity (figure 7 - page 35). Hydroxypyruvate is an oxalate precursor and high levels of oxalate can be fatal (Chapter 1 - Section 1.6 - page 36). On average, herbivores are likely to eat much higher levels of oxalate and its precursors than carnivores and so maybe can not tolerate hydroxypyruvate production as well. AGT activity was not found to be significantly related to body size despite the prediction of Rowsell et al (1979)¹¹⁴.

The only other variable, apart from diet, found to show any relationship to AGT subcellular distribution is AGT activity, species which have mitochondrial AGT tend to have higher AGT activity than species with peroxisomal AGT. The correlation between these two variables just misses significance at the 5 % level ($p = 0.0799$). This is surprising given the strong relationship between the two variables and species diet, the lack of significance is probably due to a reduced sample size ($n = 28$).

6.2.5 Levels of immunoreactive AGT and AGT catalytic activity

The levels of immunoreactive AGT were not found to be correlated to AGT subcellular distribution ($p = 0.306$). Surprisingly, they were also not found to be correlated to AGT activity ($p=0.516$). As mentioned before, it is likely that samples with low AGT activity but high levels of AGT immunoreactive protein are those in which some degradation has occurred and these species data points may be erroneously influencing the result.

If there was more confidence in the quality of the liver samples the lack of a correlation between levels of AGT immunoreactive protein and AGT catalytic activity could suggest another hypothesis. It may be evidence that AGT performs another function not connected to its aminotransferase activity and so the high levels of the enzyme are maintained for this reason. Other proteins have been discovered recently with two unrelated functions¹¹⁷. According to the theory expressed in the previous chapter, AGT is peroxisomal in plants. A role in glyoxylate detoxification for AGT is less likely in plants as oxalate has not been shown to be toxic. Perhaps this other role of AGT is the principal role in plants such as a role in carbon retrieval during photorespiration.

Unfortunately, very little information is available about the expression and activity levels of other enzymes in the same range of species in which the AGT data is known.

The catalytic roles of many enzymes overlap with those of AGT and they could be compensating for AGT and obscure the true relationships between AGT variables. For example, AGT2¹¹⁸ is located only in the mitochondria and is specific for L-alanine and glyoxylate, at least in rats¹¹⁹. Therefore in some species (such as rats) AGT2 could make a more significant contribution to gluconeogenesis than AGT1. The expression levels of AGT2 are very variable between species. AGT2 is highly expressed in rat and mouse liver, is expressed at low levels in dog and cat liver and is absent in human liver¹¹⁹.

6.3 Summary

This study constitutes the first formal analysis of the relationship between the subcellular distribution of AGT in the livers of mammals and diet. The correlation was seen to be significant and independent of phylogeny. Also, despite limitations of the data, a significant correlation between AGT activity and diet is reported, for the first time.

7. Contributions and Future Directions

7.1 Contributions of this study to our understanding of the evolution of AGT targeting

7.1.1. Selection pressure

The variable subcellular distribution of hepatic AGT is a remarkable example of molecular adaptation. Before this study, a relationship between the subcellular distribution of hepatic AGT protein and diet had been observed⁴³ and a metabolic explanation to account for this phenomenon had been postulated⁴⁵. However there was no evidence that the subcellular distribution of AGT was changing over evolutionary time, due to the effect of selection pressure and not due to lack of functional constraint. This study records the first evidence that the subcellular distribution of hepatic AGT protein is indeed under the influence of positive selection pressure to change. Firstly, in recent primate history (Chapter 3) positive selection pressure has been shown to have acted on the 5' region of primate AGT genes to abolish or diminish mitochondrial targeting, possibly in response to a shift in diet from frugivorous and insectivorous to frugivorous and folivorous. Secondly, in mammals as a whole (Chapter 6), a positive correlation, independent of phylogeny, between the subcellular distribution of hepatic AGT and diet was found, whereby carnivores tend to have mitochondrial AGT, herbivores tend to have peroxisomal AGT and omnivores tend to have both. This is strongly suggestive of evolutionary adaptation. Also found was a new observation about the significant correlation between AGT enzymatic activity levels and species diet (chapter 6).

7.1.2 Mechanisms of AGT targeting

Three new molecular mechanisms for the protein targeting of AGT to different subcellular compartments were discovered or suggested in this study. Namely:

1. The loss of TS-A resulting in solely peroxisomal AGT protein in baboon liver (Chapter 3),
2. Accumulation of mutations in the MTS incompatible with its role as a MTS, resulting in a solely peroxisomal AGT protein in primate liver (Chapter 3) and
3. The absence of a PTS1 resulting in a mitochondrial and cytosolic distribution of AGT protein in xenopus liver (Chapter 4).

A survey of C-terminal tripeptides with some characteristics of a PTS1, led to a more detailed description of protein context dependent PTS1s (Chapter 5). The cloning of xenopus AGT cDNA (Chapter 4) furthered understanding of context dependent peroxisomal targeting. This represents a molecular tool, which can be used to identify the putative additional targeting information involved in context dependent peroxisomal targeting and so advance our understanding of this phenomenon (see Section 7.2.).

7.2 Directions for the future

1. The additional targeting information thought to be contributing to peroxisomal targeting by context specific PTS1s is present in human AGT protein but not in the xenopus AGT protein. This was shown by the different subcellular distributions of these proteins with KKL at the C-terminus (i.e. *pHsAGT* and *pXS-KKL*). Therefore, this additional targeting information could be mapped by creating fusions between the xenopus and human AGT cDNAs and expressing them in COS

cells. As the human and xenopus AGT proteins are homologous, fusions between them are likely to fold correctly and this could be verified by measuring the enzymatic activity.

2. Expression of newly identified putative AGT orthologues, such as the *D. melanogaster* and *C. elegans* homologues discussed in Chapter 5, in tissue culture cells could provide much useful information. Expression of the cDNAs in, for example, COS cells followed by immunofluorescence analysis would suggest their natural subcellular distribution. Fusions of the potential targeting sequences with GFP and subsequent expression and immunofluorescence analysis may provide more information about PTS1 and MTS degeneracy.
3. It is unknown whether the additional targeting information necessary for targeting of human AGT by the PTS1, KKL, is shared by other protein dependent PTS1s or whether the additional targeting information is different in every case. Expression and immunofluorescence analysis of human AGT cDNA constructs mutagenised to encode other PTS1s at their C termini, in tissue culture cells, could help to answer this question; as could expression of other peroxisomal proteins mutagenised to encode C terminal KKL instead of their natural PTS1s.

Bibliography

1. Omura T: Mitochondria-targeting sequence, a multi-role sorting sequence recognized at all steps of protein import into mitochondria. *Journal of Biochemistry* 123:1010-1016, 1998
2. Schatz G: Protein transport - the doors to organelles. *Nature* 395:439-440, 1998
3. Neupert W: Protein import into mitochondria. *Annual Review of Biochemistry* 66:863-917, 1997
4. Von Heijne G: Mitochondrial targeting sequences may form amphiphilic helices. *EMBO Journal* 5:1335-1342, 1986
5. Roise D, Theiler F, Horvath SJ, et al: Amphiphilicity is essential for mitochondrial presequence function. *EMBO Journal* 7:649-653, 1988
6. Hammen PK, Weiner H: Mitochondrial leader sequences: structural similarities and sequence differences. *Journal Of Experimental Zoology* 282:280-283, 1998
7. Lumb MJ, Drake AF, Danpure CJ: Effect of N-terminal α -helix formation on the dimerization and intracellular targeting of alanine:glyoxylate aminotransferase. *Journal Of Biological Chemistry* 274:20587-20596, 1999
8. Gaume B, Klaus C, Ungermann C, et al: Unfolding of preproteins upon import into mitochondria. *EMBO Journal* 17:6497-6507, 1998
9. Deshaies RJ, Koch BD, Werner-Washburne M, et al: A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. *Nature* 332:800-804, 1988
10. Hachiya N, Alam R, Sakasegawa Y, et al: A mitochondrial import factor purified from rat liver cytosol in an ATP - dependent conformational modulator for precursor proteins. *EMBO Journal* 12:1579-1586, 1993
11. Hachiya N, Komiya T, Alam R, et al: MSF, a novel cytoplasmic chaperone which functions in precursor targeting to mitochondria. *EMBO Journal* 13:5146-5154, 1994
12. Komiya T, Schatz G, Mihara K: Interaction of mitochondrial targeting signals with acidic receptor domains along the protein import pathway: evidence for the acid chain hypothesis. *EMBO Journal* 17:3886-3898, 1998
13. Dekker PJT, Martin F, Maarse AC, et al: The Tim core complex defined the number of mitochondrial translocation contact sites and can hold arrested preproteins in the absence of matrix Hsp70-Tim44. *EMBO Journal* 16:5408-5419, 1997
14. Schatz G: The protein import system of mitochondria. *Journal of Biological Chemistry* 271:31763-31766, 1996

15. Ryan KR, Jensen RE: Protein translocation across mitochondrial membranes: what a long strange trip it is. *Cell* 83:517-519, 1995
16. Ogishima T, Niidome T, Shimokata K, et al: Analysis of elements in the substrate required for processing by mitochondrial processing peptidase. *Journal Of Biological Chemistry* 270:30322-30326, 1995
17. Ito A: Mitochondrial processing peptidase: Multiple site recognition of precursor proteins. *Biochemical And Biophysical Research Communications* 265:611-616, 1999
18. Hendrick JP, Hodges PE, Rosenberg LE: Survey of amino-terminal proteolytic cleavage sites in mitochondrial precursor proteins: Leader peptides cleaved by two matrix proteases share a three-amino acid motif. *Proceedings of the National Academy of Sciences, USA* 86:4056-4060, 1989
19. De Hoop M, Ab G: Import of proteins into peroxisomes and other microbodies. *Biochemical Journal* 286:657-669, 1992
20. Crookes WJ, Olsen LJ: Peroxin puzzles and folded freight: Peroxisomal protein import in review. *Naturwissenschaften* 86:51-61, 1999
21. Danpure CJ: Targeting and import of peroxisomal proteins, in Hurlley SM. (ed): *Protein Targeting*. Oxford, IRL Press, 1992, pp 63-100
22. Subramani S: Components involved in peroxisome import, biogenesis, proliferation, turnover, and movement. *Physiological Reviews* 78:171-188, 1998
23. Gould SJ, Keller GA, Hosken N, et al: A conserved tripeptide sorts proteins to peroxisomes. *Journal of Cell Biology* 108:1657-1664, 1989
24. Swinkels BW, Gould SJ, Subramani S: Targeting efficiencies of various permutations of the consensus C-terminal tripeptide peroxisomal targeting signal. *FEBS Letters* 305:133-136, 1992
25. Dodt G, Braverman N, Wong C, et al: Mutations in the PTS1 receptor gene, *PXRI*, define complementation group 2 of the peroxisome biogenesis disorders. *Nature Genetics* 9:115-125, 1995
26. Fransen M, Brees C, Baumgart E, et al: Identification and characterisation of the putative human peroxisomal C-terminal targeting signal import receptor. *Journal Of Biological Chemistry* 270:7731-7736, 1995
27. Swinkels BW, Gould SJ, Bodnar AG, et al: A novel, cleavable peroxisomal targeting signal at the amino-terminus of the rat 3-ketoacyl-CoA thiolase. *EMBO Journal* 10:3255-3262, 1991
28. Flynn CR, Mullen RT, Trelease RN: Mutational analysis of a type 2 peroxisomal targeting signal that is capable of directing oligomeric protein import into tobacco BY-2 glyoxysomes. *Plant Journal* 16:709-720, 1999

29. Girzalsky W, Rehling P, Stein K, et al: Involvement of Pex13p in Pex14p localization and peroxisomal targeting signal 2-dependent protein import into peroxisomes. *Journal of Cell Biology* 144:1151-1162, 1999
30. Braverman N, Dodt G, Gould SJ, et al: An Isoform of Pex5p, the human PTS1 receptor, is required for the import of PTS2 proteins into peroxisomes. *Human Molecular Genetics* 7:1195-1205, 1998
31. Dodt G, Warren D, Yahraus T, et al: Peroxisomal PTS1 and PTS2 Protein Import is Mediated by Distinct Domains of the human PTS1 Receptor Pex5. *Molecular Biology Of The Cell* 9:2026-2026, 1998
32. Arnold FH: Metal-affinity separations: a new dimension in protein processing. *Biotechnology* 9:151-156, 1991
33. Brocard C, Kragler F, Simon MM, et al: The tetratricopeptide repeat domain of the PAS10 protein of *Saccharomyces cerevisiae* is essential for binding the peroxisomal targeting signal - SKL. *Biochemical And Biophysical Research Communications* 204:1016-1022, 1994
34. Goebel M., Yanagida M: The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends In Biochemical Sciences* 16:173-177, 1991
35. Lamb J, Tugendreich S, Hieter P: Tetratricopeptide repeat interactions: to TPR or not to TPR. *Trends In Biochemical Sciences* 20:257-259, 1995
36. Iwahashi J, Yamazaki S, Komiya T, et al: Analysis of the functional domain of the rat liver mitochondrial import receptor Tom20. *Journal Of Biological Chemistry* 272:18467-18472, 1997
37. Gatto GJ, Geisbrecht BV, Gould SJ, et al: A proposed model for the PEX5-peroxisomal targeting signal-1 recognition complex. *Proteins* 38:241-246, 2000
38. McNew JA, Goodman JM: An oligomeric protein is imported into peroxisomes *in vivo*. *Journal of Cell Biology* 127:12451994
39. Leiper JM, Oatey PB, Danpure CJ: Inhibition of alanine:glyoxylate aminotransferase 1 dimerization is a prerequisite for its peroxisome-to-mitochondrion mistargeting in primary hyperoxaluria type 1. *Journal of Cell Biology* 135:939-951, 1996
40. Glover JR, Andrews DW, Rachubinski RA: *Saccharomyces cerevisiae* peroxisomal thiolase is imported as a dimer. *Proceedings Of The National Academy Of Sciences Of The USA*. 91:105411994
41. Imanaka T, Small GM, Lazarow PB: Translocation of acyl-CoA oxidase into peroxisomes requires ATP hydrolysis but not a membrane potential. *Journal of Cell Biology* 105:29151987
42. Danpure CJ: How can the products of a single gene be localized to more than one intracellular compartment. *Trends in Cell Biology* 5:230-238, 1995

43. Danpure CJ, Fryer P, Jennings PR, et al: 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. *European Journal of Cell Biology* 64:295-313, 1994
44. Oatey PB, Lumb MJ, Danpure CJ: Molecular basis of the variable mitochondrial and peroxisomal localisation of alanine:glyoxylate aminotransferase. *European Journal of Biochemistry* 241:374-385, 1996
45. Danpure CJ: Variable peroxisomal and mitochondrial targeting of alanine:glyoxylate aminotransferase in mammalian evolution and disease. *Bioessays* 19:317-326, 1997
46. Oda T, Funai T, Miura S: *In vitro* association with peroxisomes and conformational change of peroxisomal serine:pyruvate / alanine:glyoxylate aminotransferase in rat and human livers. *Biochemical And Biophysical Research Communications* 228:341-346, 1996
47. Purdue PE, Lumb MJ, Danpure CJ: Molecular evolution of alanine:glyoxylate aminotransferase 1 intracellular targeting. Analysis of the marmoset and rabbit genes. *European Journal of Biochemistry* 207:757-766, 1992
48. Takada Y, Kaneko N, Esumi H, et al: Human peroxisomal L-alanine: glyoxylate aminotransferase. Evolutionary loss of a mitochondrial targeting signal by point mutation of the initiation codon. *Biochemical Journal* 268:517-520, 1990
49. Birdsey GM, Danpure CJ: Evolution of alanine:glyoxylate aminotransferase intracellular targeting: structural and functional analysis of the guinea pig gene. *Biochemical Journal* 331:49-60, 1998
50. Arnason U, Gullberg A, Janke A: Molecular timing of primate divergences as estimated by two nonprimate calibration points. *Journal Of Molecular Evolution* 47:718-727, 1998
51. Arnason U, Gullberg A, Janke A: Phylogenetic analyses of mitochondrial DNA suggests a sister group relationship between Xenarthra (Edentata) and Ferungulates. *Molecular Biology And Evolution* 14:762-768, 1997
52. D'Erchia AM, Gissi C, Pesole G, et al: The guinea-pig is not a rodent. *Nature* 381:597-600, 1996
53. Janke A, Xu X, Arnason U: The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among Monotremata, Marsupialia, and Eutheria. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 94:1276-1281, 1997
54. Ledje C, Arnason U: Phylogenetic relationships within carniform carnivores based on analyses of the mitochondrial 12S rRNA gene. *Journal Of Molecular Evolution* 43:641-649, 1996

55. Mouchaty SK, Gullberg A, Janke A, et al: The phylogenetic position of the Talpidae within eutheria based on analysis of complete mitochondrial sequences. *Molecular Biology And Evolution* 17:60-67, 2000
56. Muir C, Arnason U: Is there sufficient evidence to elevate the orangutan of Borneo and Sumatra to separate species? *Journal Of Molecular Evolution* 46:378-381, 1998
57. Rasmussen AS, Janke A, Arnason U: The mitochondrial DNA molecule of the hagfish (*Myxine glutinosa*) and vertebrate phylogeny. *Journal Of Molecular Evolution* 46:382-388, 1998
58. Ursing BM, Arnason U: The complete mitochondrial DNA sequence of the pig (*Sus scrofa*). *Journal Of Molecular Evolution* 47:302-306, 1998
59. Lumb MJ, Purdue PE, Danpure CJ: Molecular evolution of alanine/glyoxylate aminotransferase 1 intracellular targeting: Analysis of the feline gene. *European Journal of Biochemistry* 221:53-62, 1994
60. Motley A, Lumb MJ, Oatey PB, et al: 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. *Journal of Cell Biology* 131:95-109, 1995
61. Oatey PB, Lumb MJ, Jennings PR, et al: Context dependency of the PTS1 motif in human alanine:glyoxylate aminotransferase 1. *Annals Of The New York Academy Of Sciences* 804:652-653, 1995
62. Danpure CJ, Purdue PE: Primary hyperoxaluria, in Scriver CR, Beaudet AL, Sly WS, et al. (eds): *The metabolic and molecular bases of inherited disease*. (ed7). New York, St Louis, San Francisco, Auckland, Nogota, Caracas, Lisbon, London, Madrid, Mexico City, Milan, Montreal, New Dehli, San Juan, Singapore, Sydney, Tokyo, Toronto, McGraw-Hill Inc, 1995, pp 2385-2424
63. Danpure CJ, Cooper PJ, Wise PJ, et al: An enzyme trafficking defect in two patients with primary hyperoxaluria type 1: peroxisomal alanine/glyoxylate aminotransferase rerouted to mitochondria. *Journal of Cell Biology* 108:1345-1352, 1989
64. Cooper PJ, Danpure CJ, Wise PJ, et al: Immunocytochemical localization of human hepatic alanine:glyoxylate aminotransferase in control subjects and patients with primary hyperoxaluria type 1. *Journal of Histochemistry and Cytochemistry* 36:1285-1294, 1988
65. Purdue PE, Allsop J, Isaya G, et al: Mistargeting of peroxisomal L-alanine:glyoxylate aminotransferase to mitochondria in primary hyperoxaluria patients depends upon activation of a cryptic mitochondrial targeting sequence by point mutation. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 88:10900-10904, 1991

66. Purdue PE, Takada Y, Danpure CJ: Identification of mutations associated with peroxisome-to mitochondrion mistargeting of alanine/glyoxylate aminotransferase in primary hyperoxaluria type 1. *Journal of Cell Biology* 111:2341-2351, 1990
67. Miyajima H, Oda T, Ichiyama A: Induction of mitochondrial serine:pyruvate aminotransferase of rat liver by glucagon and insulin through different mechanisms. *Journal Of Biochemistry (Tokyo)* 105:500-504, 1989
68. Oda T, Yanagisawa M, Ichiyama A: Induction of serine:pyruvate aminotransferase in rat liver organelles by glucagon and a high protein diet. *Journal Of Biochemistry (Tokyo)* 91:219-232, 1982
69. Hayashi S, Sakuraba H, Noguchi T: Response of hepatic alanine:glyoxylate aminotransferase 1 to hormone differs among mammalia. *Biochemical And Biophysical Research Communications* 165:372-376, 1989
70. Hodgkinson A: *Oxalic acid in biology and medicine*. New York, Academic Press, 1977,
71. Sambrook J, Fritsch EF, Maniatis T: *Molecular cloning: a laboratory manual*, New York, Cold Spring Harbour Laboratory Press, 1989,
72. Danpure CJ, Jennings PR: Further studies on the activity and subcellular distribution of alanine:glyoxylate aminotransferase in the livers of patients with primary hyperoxaluria type 1. *Clinical Science* 75:315-322, 1988
73. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22:4673-4680, 1994
74. Pagel M: Inferring the historical patterns of biological evolution. *Nature* 401:877-884, 1999
75. Yang Z, Kumar S, Nei M: A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141:1641-1650, 1995
76. Goldman N, Yang Z: A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology And Evolution* 11:725-736, 1994
77. Yang Z: Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology And Evolution* 15:568-573, 1998
78. Li WH: Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *Journal Of Molecular Evolution* 36:96-99, 1993
79. Nei M, Gojobori T: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology And Evolution* 3:418-426, 1986

-
80. Harvey P, Pagel M: The comparative method in evolutionary biology, Oxford New York Tokyo, Oxford University Press, 1991,
 81. Messier W, Stewart CB: Episodic adaptive evolution of primate lysozymes. *Nature* 385:151-154, 1997
 82. Purvis A: A composite estimate of primate phylogeny. *Philosophical Transactions Of The Royal Society Of London Series B- Biological Sciences* 348:405-421, 1995
 83. Holbrook JD, Birdsey GM, Yang Z, et al: Molecular adaptation of alanine:glyoxylate aminotransferase targeting in primates. *Molecular Biology And Evolution* 17:387-400, 2000
 84. Page RD: GeneTree: comparing gene and species phylogenies using reconciled trees. *Bioinformatics* 14:819-820, 1998
 85. Yang Z, Nielsen R: Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal Of Molecular Evolution* 46:409-418, 1998
 86. Birdsey, G. M. Molecular analysis of the peroxisomal targeting of guinea pig alanine:glyoxylate aminotransferase. 1998. University of London, PhD Thesis.
 87. Takada Y, Noguchi T: Subcellular distribution, and physical and immunological properties of hepatic alanine: glyoxylate aminotransferase isoenzymes in different mammalian species. *Comparative Biochemistry and Physiology B* 72:597-604, 1982
 88. Schneider G, Sjoling S, Wallin E, et al: Feature-extraction from endopeptidase cleavage sites in mitochondrial targeting peptides. *Proteins* 30:49-60, 1998
 89. Endo T, Ikeo K, Gojobori T: Large-scale search for genes on which positive selection may operate. *Molecular Biology And Evolution* 13:685-690, 1996
 90. Alvarez-Valin F, Jabbari K, Bernardi G: Synonymous and nonsynonymous substitutions in mammalian genes: Intragenic correlations. *Journal Of Molecular Evolution* 46:37-44, 1998
 91. Golding GB, Dean AM: The structural basis of molecular adaptation. *Molecular Biology And Evolution* 15:355-369, 1998
 92. Hughes AL, Ota T, Nei M: Positive darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Molecular Biology And Evolution* 7:515-524, 1990
 93. Vacquier VD, Swanson WJ, Lee Y: Positive darwinian selection on two homologous fertilization proteins: What is the selective pressure driving their divergence? *Journal Of Molecular Evolution* 44 (Suppl 1):S15-S22, 1997

-
94. Kay RF: On the use of anatomical features to infer foraging behaviours in extinct primates, in Rodman PS, Cant JG. (eds): Adaptations for foraging in nonhuman primates. New York, Columbia University Press, 1984, pp 21-53
 95. Fleagle JG: Primate adaptation and evolution, San Diego, Academic Press, 1988,
 96. Gibbons A: Primate evolution - new study points to eurasian ape as great ape ancestor. *Science* 281:622-623, 1998
 97. Fleagle JG, Kay RF: The paleobiology of catarrhines, in Delson E. (ed): Ancestors: the hard evidence. New York, Alan R. Liss, 1985, pp 23-36
 98. Kay RF, Simons EL: The ecology of oligocene African Anthropoidea. *International Journal Of Primatology* 1:21-37, 1980
 99. Harvey PH, Martin RD, Clutton-Brock TH: Life histories in common perspective, in Smuts BB, Cheney DL, Seyfarth RM, et al. (eds): Primate societies. Chicago, University of Chicago Press, 1987, pp 181-196
 100. Nowak P: Walker's mammals of the world (ed5). Baltimore and London, The John Hopkins University Press, 1991,
 101. Smith RJ, Jungers WL: Body mass in comparative primatology. *Journal of Human Evolution* 32:523-559, 1997
 102. Knox C, Sass E, Neupert W, et al: Import into mitochondria, folding and retrograde movement of fumarase in yeast. *Journal Of Biological Chemistry* 273:25587-25593, 1998
 103. Krishna RG, Wold F: Identification of common post-translational modifications, in Creighton TE. (ed): Protein Structure - a practical approach. (edsecond). Oxford and New York, Oxford University Press, 1997, pp 91-116
 104. Kraft R, Wittmann-Liebold B: Proteins: postsynthetic modifications, Macmillan Publishers Ltd., 1999, pp A000051.html
 105. Lodish H, Berk A, Zipursky SL, et al: Protein sorting: Organelle biogenesis and Protein Secretion, in Tenney S. (ed): Molecular Cell Biology. (ed4). New York, W.H.Freeman and Company, 1999, pp 675-750
 106. Lee ISM, Takio K, Kido R, et al: Purification and amino- and carboxy- terminal amino acid sequences of alanine-glyoxylate transaminase 1 from human liver. *Journal Of Biochemistry (Tokyo)* 116:12-17, 1994
 107. Elgersma Y, Vos A, van de Berg M, et al: Analysis of the carboxyl-terminal peroxisomal targeting signal 1 in a homologous context in *Sacchromyces cerevisiae*. *Journal Of Biological Chemistry* 271:26375-26382, 1996

108. Sommer JM, Cheng QL, Keller GA, et al: In vivo import of firefly luciferase into the glycosomes of *Trypanosoma brucei* and mutational analysis of the c-terminal targeting signal. *Molecular Biology Of The Cell* 3:749-759, 1992
109. Flynn CR, Mullen RT, Trelease RN: Mutational analyses of a type 2 peroxisomal targeting signal that is capable of directing oligomeric protein import into tobacco by-2 glyoxysomes. *Plant Journal* 16:709-720, 1998
110. Gatto GJ, Geisbrecht BV, Gould SJ, et al: A proposed model for the PEX5-peroxisomal targeting signal-1 recognition complex. *Proteins* 38:241-246, 2000
111. Terlecky SR, Nuttley WM, McCollum D, et al: The *Pichia pastoris* peroxisomal protein PAS8p is the receptor for the C-terminal tripeptide peroxisomal targeting signal. *EMBO Journal* 14:3627-3634, 1995
112. Needleman SB, Wunsch CD: A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* 48:443-453, 1970
113. Danpure CJ, Jennings PR: Peroxisomal alanine:glyoxylate aminotransferase deficiency in primary hyperoxaluria type I. *FEBS* 201:20-24, 1986
114. Rowsell EV, Carnie JA, Wahbi SD, et al: L-serine dehydratase and L-serine pyruvate aminotransferase activities in different animal species. *Comparative Biochemistry and Physiology* 63B:543-555, 1979
115. Martins EP: Adaptation and the comparative method. *Trends in Ecology and Evolution* 15:296-299, 2000
116. Danpure CJ, Jennings PR, Watts RW: Enzymological diagnosis of primary hyperoxaluria type 1 by measurement of hepatic alanine: glyoxylate aminotransferase activity. *Lancet* 1:289-291, 1987
117. Jeffery C: Moonlighting proteins. *Trends In Biochemical Sciences* 24:8-11, 1999
118. Lee IS, Muragaki Y, Ideguchi T, et al: Molecular cloning and sequencing of a cDNA encoding alanine-glyoxylate aminotransferase 2 from rat kidney. *Journal Of Biochemistry (Tokyo)* 117:856-862, 1995
119. Noguchi T: Amino acid metabolism in animal peroxisomes, in Fahimi HD, Sies H. (eds): *Peroxisomes in biology and medicine*. (ed1). Berlin, Heidelbergm New York, Springer-Verlag, 1987, pp 234-243
120. Danpure CJ, Fryer P, Griffiths S, et al: Cytosolic compartmentalization of hepatic alanine:glyoxylate aminotransferase in patients with aberrant peroxisomal biogenesis and its effect on oxalate metabolism. *Journal of Inherited Metabolic Disease* 17:27-40, 1994

Appendix

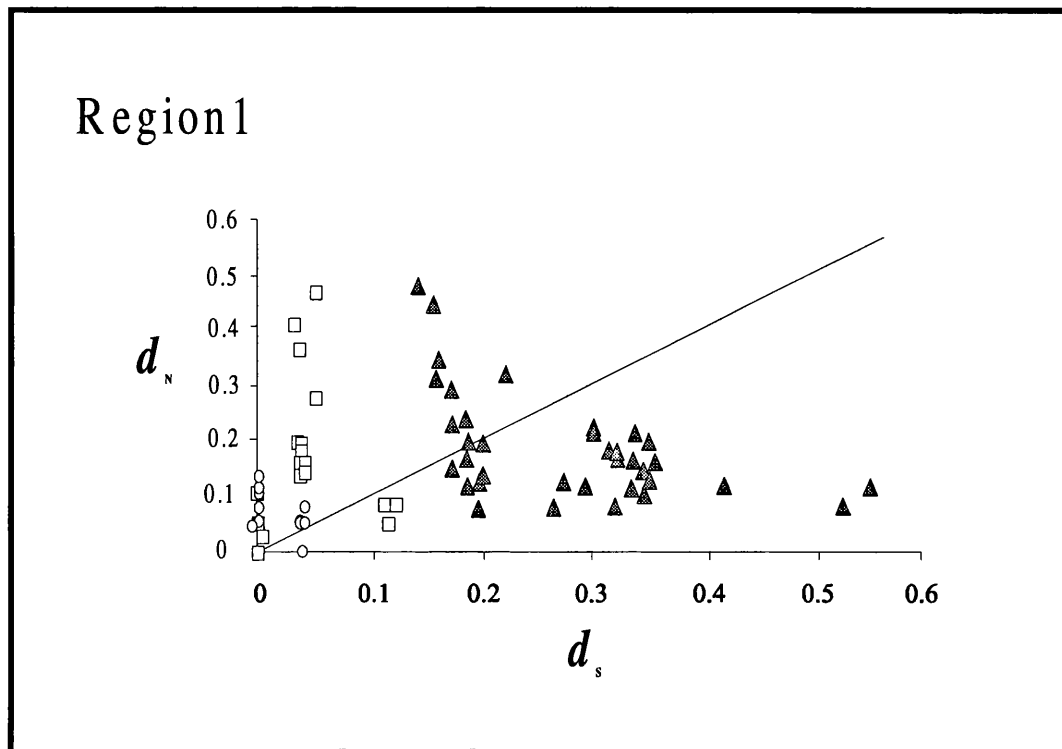


Figure A1 - The d_N/d_S substitution rates of region 1 of the AGT gene in primates (with TL-1 included in the analysis)

The synonymous (d_S) and nonsynonymous (d_N) substitution rates were calculated using the maximum-likelihood method of Goldman and Yang (1994)⁷⁶ for each of the 78 individual interspecies pairwise nucleotide comparisons. There are 21 catarrhine-catarrhine comparisons (open squares), 15 platyrrhine-platyrrhine comparisons (open circles), and 42 catarrhine-platyrrhine comparisons (shaded triangles). Compare with figure 16, page 73.

Table A1 - Mean values for d_N and d_S in pairwise comparisons of region 1, including TL-1

| | N | Mean d_N (\pm SD) | Mean d_S (\pm SD) | Ratio of mean d_N /mean d_S † |
|---------------------------|----|---------------------------|---------------------------|--------------------------------------|
| Region 1* | | | | |
| catarrhine v catarrhine | 21 | 0.166 \pm 0.039 | 0.039 \pm 0.039 | 4.23 |
| platyrrhine v platyrrhine | 15 | 0.077 \pm 0.037 | 0.014 \pm 0.02 | 5.50 |
| catarrhine v platyrrhine | 42 | 0.191 \pm 0.091 | 0.284 \pm 0.095 | 0.67 |
| Overall | 78 | 0.162 \pm 0.103 | 0.166 \pm 0.148 | 0.98 |

† The means of the d_N/d_S ratios cannot be calculated because some individual comparisons give $d_S = 0$. NB This analysis includes TL-1.

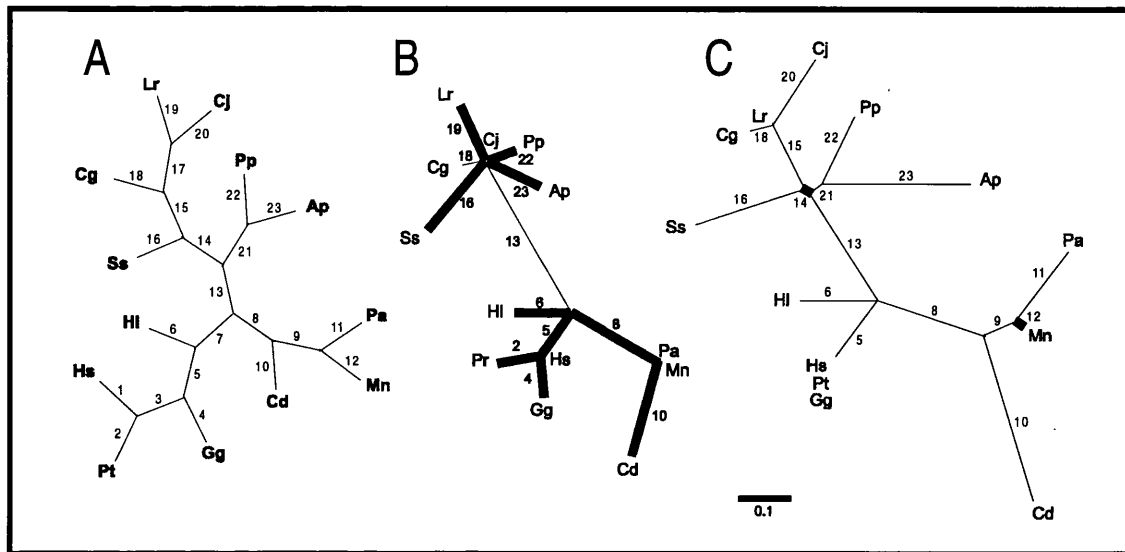


Figure A2 - d_N/d_S among lineages in the 5' region of the AGT gene in primates

Three trees are shown. The phylogenetic tree A is based on that of Purvis (1995)⁸² but with all branches (numbered 1-23 as in fig 15) drawn equal length. The derived trees B and C were generated using the computer program PAML (see Chapter 2, section 2.7.4) and drawn using the program TREEVIEW⁸⁴. The branch lengths were calculated using the one-ratio model and the d_N/d_S ratios calculated using the free-ratios model⁸⁵. The bar represents the expected number of nucleotide substitutions per codon estimated using the one-ratio model⁸⁵. 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 platyrrhines and catarrhines. Results of the analysis of region 1 (including TL-1) 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 2) are attached to terminal nodes. The calculated d_N and d_S values (to 3 decimal places) along various lineages for region 1 (tree B) are as follows: 1 = 0 (d_N), 0 (d_S); 2 = 0.026, 0; 3 = 0, 0; 4 = 0.027, 0; 5 = 0.053, 0.034; 6 = 0.054, 0.004; 7 = 0, 0; 8 = 0.082, 0.001; 9 = 0, 0; 10 = 0.115, 0; 11 = 0, 0; 12 = 0, 0; 13 = 0.026, 0.23; 14 = 0, 0; 15 = 0, 0; 16 = 0.077, 0; 17 = 0, 0; 18 = 0, 0.038; 19 = 0.051, 0; 20 = 0, 0; 21 = 0, 0; 22 = 0.053, 0; 23 = 0.051, 0. The d_N and d_S values along lineages for region 2 (tree C) are as Chapter 3 (figure 18 - page 77).

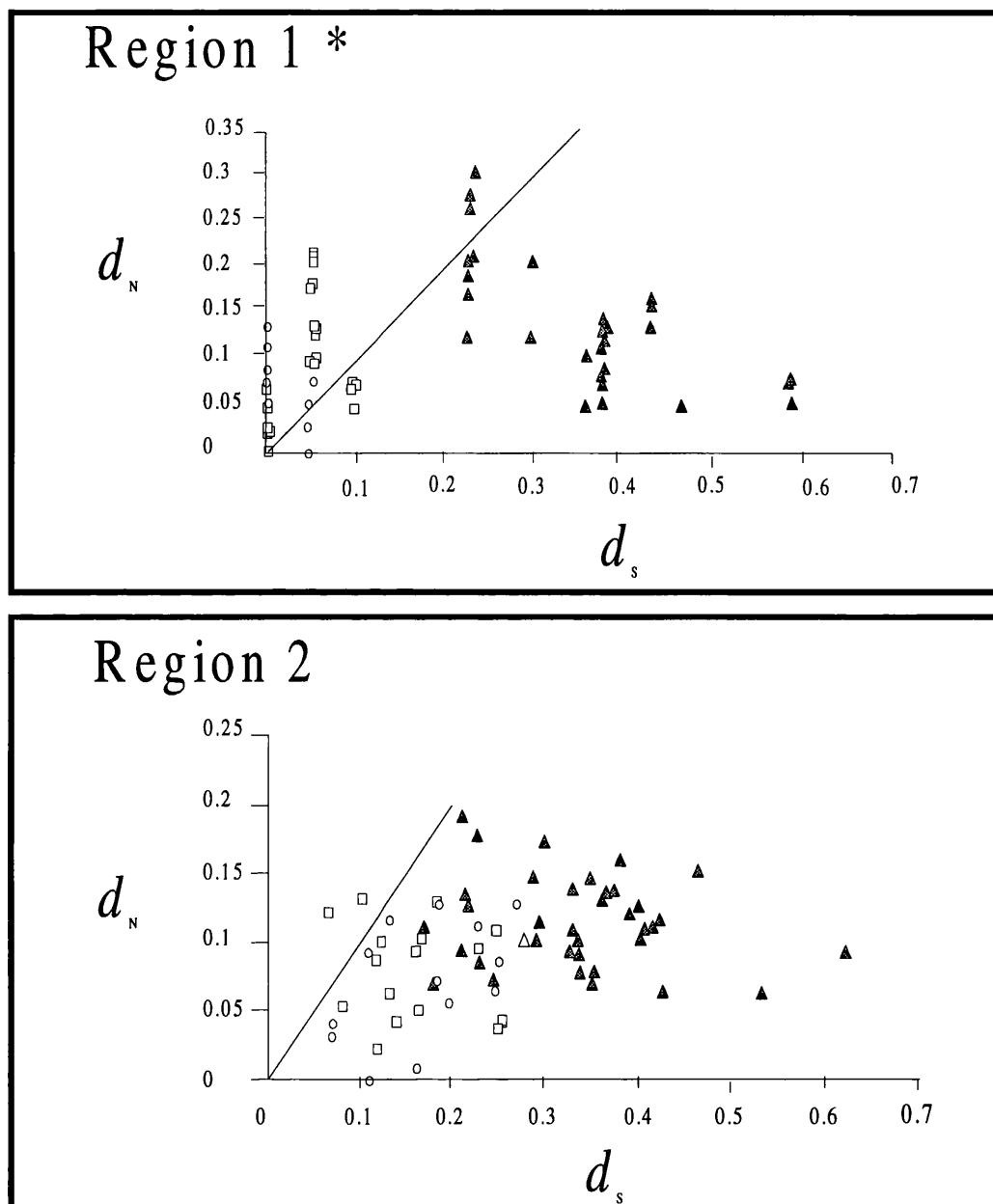


Figure A3 - The d_N/d_S substitution rate of the 5' region of the AGT gene in primates, as estimated by the method of Li (1993)⁷⁸

The synonymous (d_S) and nonsynonymous (d_N) substitution rate were calculated using the method of Li (1993)⁷⁸ for each of the 78 individual interspecies pairwise nucleotide comparisons. There are 21 catarrhine-catarrhine comparisons (open squares), 15 platyrrhine-platyrrhine comparisons (open circles), and 42 catarrhine-platyrrhine comparisons (shaded triangles). The upper and lower panels show the plots of d_N against d_S for region 1* (excluding TL-1) and region 2, respectively. The diagonal line indicates $d_N = d_S$. Compare with figure 16, page 73.

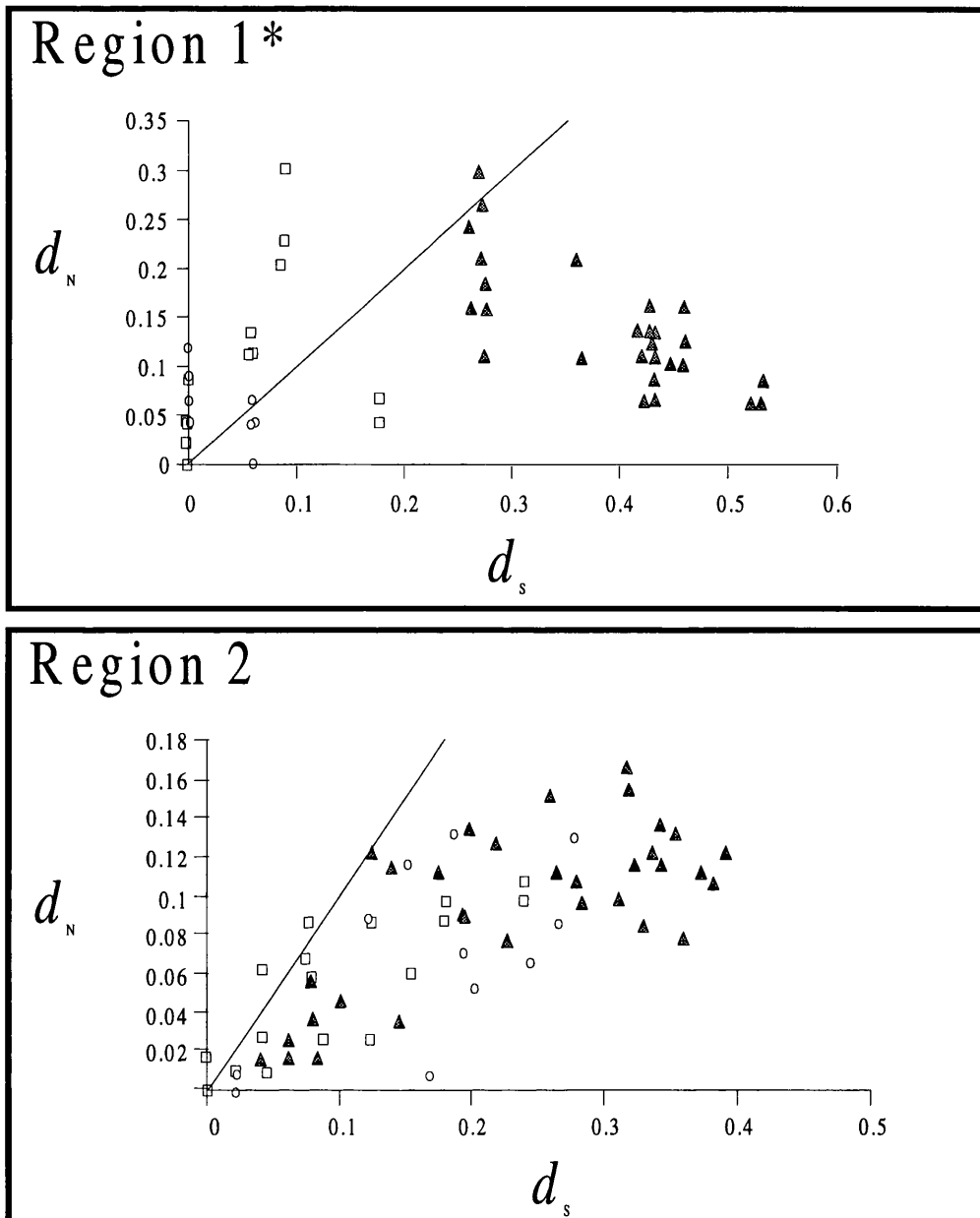


Figure A4 - The d_N/d_S substitution rate of the 5' region of the AGT gene in primates, as estimated by the method of Nei and Gojobori (1986)⁷⁹

The synonymous (d_S) and nonsynonymous (d_N) substitution rate were calculated using the method of Nei and Gojobori (1993)⁷⁹ for each of the 78 individual interspecies pairwise nucleotide comparisons. There are 21 catarrhine-catarrhine comparisons (open squares), 15 platyrrhine-platyrrhine comparisons (open circles), and 42 catarrhine-platyrrhine comparisons (shaded triangles). The upper and lower panels show the plots of d_N against d_S for region 1* (excluding TL-1) and region 2, respectively. The diagonal line indicates $d_N = d_S$. Compare with figure 16, page 73.

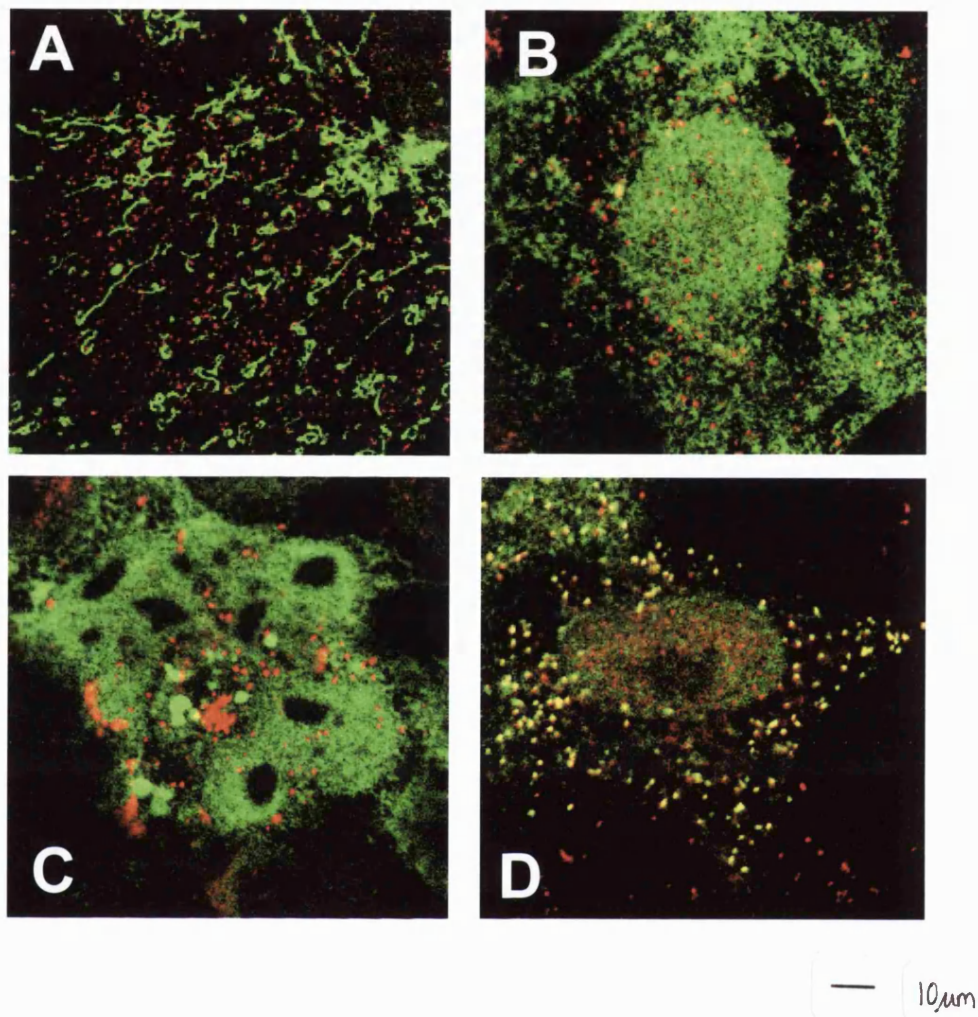


Figure A5 - Colour images of the dual labelled immunofluorescence experiments

The red and green fluorescence channels for selected images in figures 31, 33, 36 and 37 have been merged to generate composite colour images. AGT labelling is coloured green and catalase labelling is coloured red. A, AGT distribution in COS cells following transfection with pXL. It can be seen that the pXL product is entirely mitochondrial and does not co-localise with catalase at all. B, AGT distribution in COS cells following transfection with pXS. It can be seen that, in spite of the somewhat granular nature of the AGT labelling, it does not co-localise with catalase. C, AGT distribution in COS cells following transfection with *pHsAGT-KKM*. It can be seen that the aggregated AGT does not co-localise with catalase. D, AGT distribution in COS cells following transfection with pXS-SKL. It can be seen that the punctuate AGT labelling co-localises with the catalase.

Reference List for Table 11

- A. Lametschwandtner G, Brocard C, Fransen M, et al: The difference in recognition of terminal tripeptides as peroxisomal targeting signal 1 between yeast and human is due to different affinities of their receptor Pex5p to the cognate signal and to residues adjacent to it. *Journal Of Biological Chemistry* 273:33635-33643, 1998
- B. Swinkels BW, Gould SJ, Subramani S: Targeting efficiencies of various permutations of the consensus C-terminal tripeptide peroxisomal targeting signal. *FEBS Letters* 305:133-136, 1992
- C. Koller A, Spong AP, Luers GH, et al: Analysis of the peroxisomal acyl-CoA oxidase gene product from *Pichia pastoris* and determination of its targeting signal. *Yeast* 15:1035-1044, 1999
- D. Wimmer C, Schmid M, Veenhuis M, et al: The plant PTS1 receptor: similarities and differences to its human and yeast counterparts. *Plant Journal* 16:453-464, 1998
- E. Foulon V, Antonenkov VD, Croes K, et al: Purification, molecular cloning and expression of 2-hydroxyphytanoyl-CoA lyase, a peroxisomal thiamine pyrophosphate-dependent enzyme that catalyzes the carbon-carbon bond cleavage during alpha oxidation of 3-methyl branched fatty acids. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 96:10039-10044, 1999
- F. AAD27616
- G. Ofman R, Hettema EH, Hogenhout EM, et al: Acyl-CoA: dihydroxyacetonephosphate acyltransferase: cloning of the human cDNA and resolution of the molecular basis in rhizomelic chondrodysplasia punctata type 2. *Human Molecular Genetics* 7:847-853, 1998
- H. Volokita M: The carboxy-terminal end of glycolate oxidase directs a foreign protein into tobacco leaf peroxisomes. *Plant Journal* 1:361-366, 1991
- I. Didion T, Roggenkamp R: Targeting signal of the peroxisomal catalase in the methylotrophic yeast *Hansenula polymorpha*. *FEBS Letters* 303:113-116, 1992
- J. Hansen H, Didion T, Thiemann A, et al: Targeting sequences of the two major peroxisomal proteins in the methylotrophic yeast *Hansenula polymorpha*. *Molecular & General Genetics* 235:269-278, 1992
- K. Kragler F, Lametschwandtner G, Christmann J, et al: Identification and analysis of the plant peroxisomal targeting signal 1 receptor NtPEX5. *Proceedings of the National Academy of Sciences of the United States of America* 95:13336-13341, 1998

Reference List for Table 11 - cont.

- L. Amery L, Brees C, Baes M, et al: C-terminal tripeptide SNL of human D-aspartate oxidase is a functional peroxisome-targeting signal. *Biochemical Journal* 336:367-371, 1998
- M. Elgersma Y, Vos A, van de Berg M, et al: Analysis of the carboxyl-terminal peroxisomal targeting signal 1 in a homologous context in *Saccharomyces cerevisiae*. *Journal Of Biological Chemistry* 271:26375-26382, 1996
- N. Gould SJ, Keller GA, Hosken N, et al: A Conserved Tripeptide Sorts Proteins to Peroxisomes. *Journal of Cell Biology* 108:1657-1664, 1989
- O. Aitchison JD, Murray WW, Rachubinski RA: The carboxyl terminla tripeptide: ala-lys-ile is essential for targeting *Candida tropicalis* trifunctional enzyme to yeast peroxisomes. *Journal Of Biological Chemistry* 266:23197-23203, 1991
- P. Aoyama T, Tsushima K, Souri M, et al: Molecular cloning and functional expression of a human peroxisomal acyl-coenzyme A oxidase. *Biochemical And Biophysical Research Communications* 198:1113-1118, 1994
- Q. Mullen RT, Lee MS, Flynn R, et al: Diverse amino acid residues function within the type 1 peroxisomal targeting signal. *Plant Physiology* 115:881-889, 1997
- R. Moller G, Luders J, Markus M, et al: Peroxisome targeting of porcine 17beta-hydroxysteroid dehydrogenase type IV/D-specific multifunctional protein 2 is mediated by its C-terminal tripeptide AKI. *Journal of cell biochemistry* 73:70-78, 1999
- S. Motley A, Lumb MJ, Oatey PB, et al: 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. *Journal of Cell Biology* 131:95-109, 1995
- T. Oatey PB, Lumb MJ, Jennings PR, et al: Context Dependency of the PTS1 Motif in Human Alanine:Glyoxylate Aminotransferase 1. *Annals Of The New York Academy Of Sciences* 1995
- U. Vanhove GF, VanVeldhoven PP, Fransen M, et al: The CoA esters of 2-methyl-branched chain fatty acids and of the bile acid intermediates di- and trihydroxycoprostanic acids are oxidized by one single peroxisomal branched chain acyl-CoA oxidase in human liver and kidney. *Journal Of Biological Chemistry* 268:10335-10344, 1993
- V. CAB89810

Reference List for Table 11 - cont.

- W. Ferdinandusse S, Denis S, Clayton PT, et al: Mutations in the gene encoding peroxisomal alpha-methylacyl-CoA racemase cause adult-onset sensory motor neuropathy. *Nature Genetics* 24:188-191, 2000
- X. Hoefler G, Forstner M, McGuinness MC, et al: cDNA cloning of the human peroxisomal enoyl-CoA hydratase: 3-hydroxyacyl-CoA dehydrogenase bifunctional enzyme and localization to chromosome 3q26.3-3q28: a free left Alu Arm is inserted in the 3' noncoding region. *Genomics* 19:60-67, 1994
- Y. CAB65596
- Z. Corti O, Finocchiaro G, Rossi E, et al: Molecular cloning of cDNAs encoding human carnitine acetyltransferase and mapping of the corresponding gene to chromosome 9q34.1. *Genomics* 23:94-99, 1994
- AA. Ferdinandusse S, Mulders J, IJlst L, et al: Molecular cloning and expression of human carnitine octanoyltransferase: evidence for its role in the peroxisomal beta-oxidation of branched-chain fatty acids. *Biochemical And Biophysical Research Communications* 263:213-218, 1999
- BB. Quan F, Korneluk RG, Tropak MB, et al: Isolation and characterization of the human catalase gene. *Nucleic Acids Research* 14:5321-5335, 1986
- CC NP_006108
- DD. Momoi K, Fukui K, Watanabe F, et al: Molecular cloning and sequence analysis of cDNA encoding human kidney D-amino acid oxidase. *FEBS Letters* 238:180-184, 1988
- EE. Suk K, Kim YH, Hwang DJ, et al: Molecular cloning and expression of a novel human cDNA related to the diazepam binding inhibitor. *Biochemical And Biophysical Research Communications* 1454:126-131, 1999
- FF. AAA18595
- GG. CAA72214
- HH. Jones JM, Morrell JC, Gould SJ: Identification and characterization of HAOX1, HAOX2, and HAOX3, three human peroxisomal 2-hydroxy acid oxidases. *Journal Of Biological Chemistry* 275:12590-12597, 2000
- II. Geisbrecht BV, Gould SJ: The human PICD gene encodes a cytoplasmic and peroxisomal NADP(+)-dependent isocitrate dehydrogenase. *Journal Of Biological Chemistry* 274:30527-30533, 1999
- JJ AAF69798

Reference List for Table 11 - cont.

- KK. Chambliss KL, Slaughter CA, Hofmann GE, et al: Molecular cloning of human phosphomevalonate kinase and identification of a consensus peroxisomal targeting sequence. *Journal Of Biological Chemistry* 271:17330-17334, 1996
- LL. AAF17200
- MM. AAF40199
- NN. Beetham JK, Tian T, Hammock BD: cDNA cloning and expression of a soluble epoxide hydrolase from human liver. *Archives of Biochemistry and Biophysics* 305:197-201, 1993
- OO. O14975
- PP. Amaya Y, Yamazaki K, Sato M, et al: Proteolytic conversion of xanthine dehydrogenase from the NAD-dependent type to the O₂-dependent type. Amino acid sequence of rat liver xanthine dehydrogenase and identification of the cleavage sites of the enzyme protein during irreversible conversion by trypsin. *Journal Of Biological Chemistry* 265:14170-14175, 1990
- QQ. Belmouden A, Le KH, Lederer F, et al: Molecular cloning and nucleotide sequence of cDNA encoding rat kidney long-chain L-2-hydroxy acid oxidase: Expression of the catalytically active recombinant protein as a chimera. *European Journal of Biochemistry* 214:17-25, 1993
- RR. Bruinenberg PG, Blaauw M, Kazemier B, et al: Cloning and sequencing of the malate synthase gene from *Hansenula polymorpha*. *Yeast* 6:245-254, 1990
- SS. Miyazawa S, Hayashi H, Hijikata M, et al: Complete nucleotide sequence of cDNA and predicted amino acid sequence of rat acyl-CoA oxidase. *Journal Of Biological Chemistry* 262:8131-8137, 1987
- TT. Graham IA, Smith LM, Brown JW, et al: The malate synthase gene of cucumber. *Plant Molecular Biology* 13:673-684, 1989
- UU. A25776
- VV. Wood KV, Lam YA, Seliger HH, et al: Complementary DNA coding click beetle luciferases can elicit bioluminescence of different colors. *Science* 244:700-702, 1989
- WW. Orr WC, Orr EC, Legan SK, et al: Molecular analysis of the *Drosophila* catalase gene. *Archives of Biochemistry and Biophysics* 330:251-258, 1996
- XX. Wallrath LL, Burnett JB, Friedmann TB: Molecular characterization of the *Drosophila melanogaster* urate oxidase gene, an ecdysone-repressible gene expressed only in the malpighian tubules. *Molecular Cell Biology* 10:5114-5127, 1990

Reference List for Table 11 - cont.

- YY. Takada Y, Kaneko N, Esumi H, et al: Human peroxisomal L-alanine: glyoxylate aminotransferase. Evolutionary loss of a mitochondrial targeting signal by point mutation of the initiation codon. *Biochemical Journal* 268:517-520, 1990
- ZZ. Volokita M, Somerville CR: The primary structure of spinach glycolate oxidase deduced from the DNA sequence of a cDNA clone. *Journal Of Biological Chemistry* 262:15825-15828, 1987
- AAA. AAB88177
- BBB. De Wet JR, Wood KV, DeLuca M, et al: Firefly luciferase gene: structure and expression in mammalian cells. *Molecular Cell Biology* 7:725-737, 1987
- CCC. Gould SJ, Keller GA, Subramani S: Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. *Journal of Cell Biology* 105:2931-1987
- DDD. Atomi H, Ueda M, Taniguchi N, et al: Peroxisomal isocitrate lyase of the n-alkane-assimilating yeast *Candida tropicalis*: gene analysis and characterization. *Journal Of Biochemistry* 107:262-266, 1990
- EEE. Blattner J, Swinkels BW, Dorsam H, et al: Glycosome assembly in trypanosomes: Variations in the acceptable degeneracy of a COOH-terminal microbody targeting signal. *Journal of Cell Biology* 119:1129-1136, 1992
- FFF. Birdsey GM, Danpure CJ: Evolution of alanine:glyoxylate aminotransferase intracellular targeting: structural and functional analysis of the guinea pig gene. *Biochemical Journal* 331:49-60, 1998
- GGG. Purdue PE, Lumb MJ, Danpure CJ: Molecular evolution of alanine:glyoxylate aminotransferase 1 intracellular targeting. Analysis of the marmoset and rabbit genes. *European Journal of Biochemistry* 207:757-766, 1992
- HHH. Knott TG, Birdsey GM, Sinclair KE, et al: The peroxisomal targeting sequence type 1 receptor, Pex5p, and the peroxisomal import efficiency of alanine:glyoxylate aminotransferase. *Biochemical Journal* 352:409-418, 2000
- III. Oda T, Funai T, Miura S: *In vitro* association with peroxisomes and conformational change of peroxisomal serine:pyruvate / alanine:glyoxylate aminotransferase in rat and human livers. *Biochemical And Biophysical Research Communications* 228:341-346, 1996
- JJJ. AAB82001
- KKK. Purdue PE, Lumb MJ, Danpure CJ: Molecular evolution of alanine:glyoxylate aminotransferase 1 intracellular targeting. Analysis of the marmoset and rabbit genes. *European Journal of Biochemistry* 207:757-766, 1992

Reference List for Table 11 - cont.

LLL Arand M, Knehr M, Thomas H, Zeller, HD, Oesch F: An impaired peroxisomal targeting sequence leading to an unusual bicompartamental distribution of cytosolic epoxide hydrolase. FEBS Letters 294:19-22, 1991

MMM S71455

NNN S07124

OOO This study

PPP Williams E, Cregeen D, Rumsby G: Identification and expression of a cDNA for human glycolate oxidase. Biochemical and Biophysical Acta 1493: 246-248, 2000

Table A2 - Homologue test 2

| Sp | AGT | | ABAT | | A20 | | AMR | | AAT2 | | AAT1 | | BCAT1 | | CBL | | GFP1 | | KMO | | KAA | | OAT | | PSA | | TAT | |
|------|-----|----|------|----|-----|----|-----|----|------|----|------|----|-------|----|-----|----|------|----|-----|----|-----|----|-----|----|-----|----|-----|----|
| | I | S | I | S | I | S | I | S | I | S | I | S | I | S | I | S | I | S | I | S | I | S | I | S | I | S | I | S |
| At | 41 | 51 | 33 | 44 | 31 | 42 | 34 | 44 | 33 | 46 | 32 | 40 | 30 | 42 | 30 | 41 | 41 | 43 | 31 | 41 | 30 | 39 | 36 | 49 | 32 | 42 | 32 | 45 |
| Fa | 39 | 47 | 34 | 43 | 30 | 41 | 29 | 39 | 35 | 45 | 34 | 47 | 31 | 41 | 32 | 42 | 30 | 41 | 30 | 42 | 32 | 43 | 33 | 47 | 29 | 42 | 33 | 46 |
| Sc | 32 | 43 | 32 | 41 | 34 | 45 | 32 | 45 | 33 | 40 | 30 | 40 | 30 | 41 | 36 | 47 | 29 | 42 | 32 | 43 | 30 | 38 | 33 | 44 | 35 | 48 | 29 | 40 |
| Hm | 36 | 46 | 36 | 44 | 31 | 42 | 32 | 45 | 32 | 42 | 32 | 42 | 35 | 45 | 31 | 43 | 33 | 44 | 33 | 43 | 31 | 43 | 32 | 43 | 32 | 44 | 31 | 44 |
| Mj | 39 | 53 | 35 | 52 | 32 | 47 | 29 | 44 | 31 | 41 | 37 | 49 | 31 | 43 | 28 | 44 | 35 | 50 | 33 | 48 | 33 | 45 | 34 | 47 | 36 | 48 | 33 | 44 |
| Dr | 37 | 48 | 37 | 47 | 35 | 45 | 29 | 42 | 32 | 41 | 35 | 44 | 34 | 44 | 38 | 46 | 32 | 43 | 30 | 43 | 33 | 44 | 32 | 44 | 33 | 45 | 32 | 44 |
| Bs | 36 | 48 | 35 | 45 | 33 | 43 | 33 | 46 | 35 | 46 | 29 | 41 | 36 | 46 | 29 | 43 | 31 | 43 | 32 | 46 | 30 | 42 | 33 | 43 | 31 | 42 | 29 | 42 |
| Ap | 37 | 47 | 30 | 43 | 30 | 38 | 36 | 45 | 34 | 46 | 33 | 41 | 36 | 47 | 35 | 49 | 32 | 43 | 30 | 42 | 31 | 44 | 35 | 48 | 32 | 45 | 30 | 43 |
| Tm | 36 | 49 | 32 | 46 | 28 | 39 | 31 | 45 | 28 | 42 | 29 | 40 | 33 | 41 | 31 | 47 | 34 | 46 | 30 | 40 | 34 | 45 | 35 | 48 | 32 | 41 | 32 | 44 |
| Mtf | 38 | 50 | 30 | 45 | 31 | 40 | 31 | 40 | 32 | 44 | 27 | 40 | 37 | 47 | 27 | 42 | 32 | 43 | 34 | 44 | 31 | 41 | 33 | 45 | 34 | 43 | 31 | 44 |
| S716 | 37 | 46 | 29 | 41 | 34 | 45 | 33 | 41 | 32 | 42 | 32 | 44 | 35 | 45 | 29 | 40 | 34 | 47 | 31 | 45 | 35 | 45 | 32 | 44 | 33 | 45 | 35 | 46 |
| Mta | 38 | 50 | 31 | 44 | 28 | 42 | 31 | 45 | 31 | 43 | 26 | 40 | 34 | 44 | 26 | 41 | 33 | 43 | 36 | 47 | 31 | 40 | 31 | 44 | 34 | 44 | 34 | 44 |
| Af | 38 | 48 | 30 | 46 | 37 | 48 | 30 | 42 | 34 | 43 | 34 | 47 | 31 | 41 | 32 | 44 | 29 | 45 | 32 | 42 | 31 | 44 | 32 | 43 | 29 | 44 | 30 | 43 |
| Ph | 29 | 43 | 35 | 48 | 33 | 44 | 30 | 46 | 33 | 45 | 33 | 42 | 33 | 46 | 31 | 44 | 27 | 40 | 32 | 47 | 31 | 47 | 30 | 45 | 33 | 44 | 32 | 43 |
| Pa | 29 | 42 | 32 | 47 | 33 | 47 | 34 | 49 | 31 | 45 | 33 | 43 | 30 | 44 | 31 | 43 | 32 | 47 | 34 | 48 | 30 | 44 | 30 | 46 | 34 | 45 | 33 | 47 |
| S803 | 39 | 48 | 32 | 42 | 32 | 42 | 31 | 44 | 32 | 42 | 33 | 46 | 31 | 42 | 29 | 40 | 32 | 43 | 30 | 41 | 34 | 46 | 31 | 41 | 36 | 45 | 33 | 44 |
| Scl | 39 | 47 | 27 | 39 | 32 | 42 | 35 | 44 | 31 | 41 | 31 | 39 | 33 | 45 | 33 | 43 | 33 | 41 | 34 | 44 | 32 | 42 | 32 | 41 | 33 | 45 | 32 | 44 |
| Ac | 35 | 47 | 32 | 43 | 32 | 42 | 32 | 41 | 33 | 44 | 33 | 44 | 31 | 41 | 30 | 40 | 32 | 45 | 33 | 43 | 32 | 43 | 33 | 45 | 32 | 41 | 33 | 41 |

Test 2: Percentage identities of peptide sequences of candidate orthologues to peptide sequences of human paralogues. Gap penalty is 2 as is gap extension penalty. Sequences that score higher identity to any paralogue than to AGT, fail test 2 and are shaded grey in the table. For species codes see table 12. For paralogue codes see table 13.

Table A3 - Data for Comparative analysis

| species | AGT activity | Body mass | Gestation length | Species diet | AGT subcellular distribution | Expression level |
|------------------------|--------------|-----------|------------------|--------------|------------------------------|------------------|
| agouti | 0.83 | 9150 | 118 | 1 | 1 | 0 |
| baboon | 3.95 | 28000 | 171.5 | 1 | 1 | 3 |
| bear | | 425000 | 233 | 1 | 2 | |
| bettong | 0.57 | 1675 | 21 | 1 | | 0 |
| bison | 1.43 | 675000 | 285 | 1 | | 1 |
| cupuchin | 1.09 | 1705 | 180 | 1 | | |
| cat | | 3900 | 72 | 3 | 3 | 3 |
| cavy | 1.93 | 350 | 60 | 1 | | 3 |
| chimpanzee | | 48000 | 230 | 1 | 1 | |
| chipmunk | 0.83 | 106 | 31 | 1 | | 2 |
| civit | 1.61 | 1900 | 64 | 3 | | 2 |
| de brazza monkey | 0.18 | 6150 | 182 | 1 | | |
| dog | 13.48 | 40000 | 63 | 3 | 3 | 2 |
| douracouli | 1.9 | 800 | 151.5 | 1 | | |
| dunnart | | 13 | 13 | 3 | 2 | |
| echidna | 10.08 | 6250 | 18 | 3 | | 2 |
| fat tailed dwarf lemur | | | | 2 | 2 | 2 |
| ferret | 16.3 | 958 | 42 | 3 | 3 | 3 |
| giraffe | 2.89 | 800000 | 457 | 1 | | 1 |
| gorilla | 0.97 | 173000 | 258 | 1 | 1 | 3 |
| guinea pig | 3.18 | 1000 | 68 | 1 | 1 | 2 |
| hamster | 1.09 | 105 | 19 | 1 | 2 | 1 |
| hedgehog | 1.79 | 750 | 35 | 3 | 3 | 3 |
| horse | 0.89 | 250000 | 335 | 1 | | 0 |
| house shrew | | 25 | 23 | 3 | 3 | |
| human | | 64000 | 279 | 2 | 1 | |
| koala | 0.67 | 9300 | 35 | 1 | 1 | 1 |
| kudu | 1.16 | 187500 | 186 | 1 | | 0 |
| langur | 1.4 | 6550 | 200 | 1 | | 3 |
| lemur | 6.02 | 2900 | 117 | 2 | 2 | 3 |
| loris | 4.38 | 683 | 167.5 | 3 | 2 | 2 |
| macaque | | 15000 | 165 | 2 | 1 | |
| mandrill | 0.94 | 18250 | 232.5 | 1 | | 2 |
| margay | 6.14 | 11300 | 42 | 3 | | 3 |
| marmoset | 8.06 | 342 | 140 | 2 | 2 | 2 |
| meerkat | 13.2 | 1800 | 77 | 3 | | 3 |
| mole | 9.59 | 93 | 28 | 3 | 3 | 2 |
| mole rat | 0.14 | 200 | 30 | 1 | | 2 |
| mouse | 1.29 | 21 | 20 | 1 | 2 | 1 |
| ocelot | 9.16 | 13550 | 70 | 3 | 3 | 3 |
| opossum | 7.682 | 115 | 13 | 2 | 2 | 2 |
| Orang utan | 1.73 | 60000 | 249 | 1 | 1 | 3 |
| oryx | 0.23 | 155000 | 240 | 1 | | 1 |

| | | | | | | |
|-----------------|--------|--------|------|---|---|---|
| Patas monkey | 0.76 | 8500 | 167 | 1 | | 2 |
| pig | 5 | 250000 | 120 | 2 | 2 | 0 |
| porcupine | 1.21 | 1750 | 105 | 1 | 2 | 2 |
| possum | 0.62 | 3150 | 17 | 1 | | 0 |
| rabbit | 8.615 | 1800 | 31.5 | 1 | 1 | 3 |
| rat | 0.41 | 300 | 22 | 2 | 2 | 1 |
| reindeer | 1.89 | 95500 | 228 | 1 | | |
| saki monkey | 5.08 | 1200 | 150 | 1 | 1 | 2 |
| sealion | 8.23 | 563500 | 364 | 3 | | 3 |
| seba's bat | 3.18 | 15 | 124 | 1 | 1 | 3 |
| sheep | 0.97 | 110000 | 174 | 1 | 1 | 0 |
| spider monkey | 3.78 | 6000 | 229 | 1 | | 2 |
| squirrel | 4.04 | 543 | 44 | 1 | 2 | 2 |
| squirrel monkey | 1.7 | 925 | 162 | 1 | | 2 |
| tamarin | 7.44 | 563 | 128 | 2 | 2 | 2 |
| tree shrew | 1.33 | 200 | 50 | 3 | 3 | 2 |
| wallaby | 1.7325 | 4250 | 31 | 1 | 1 | 1 |

AGT activity was taken from primary data, published in abridged from in Danpure and Fryer (1994)³³. If more than one test result from one liver was available, the highest figure was used. Body mass (in grams) was taken from Nowak (1997)⁹⁷. Gestation length (in days) was taken from Nowak (1997)⁹⁷. Species diets were taken from Nowak (1997)⁹⁷ and categorised into 1 = mainly herbivorous, 2 = omnivorous and 3 = mainly carnivorous. AGT subcellular distribution data was taken from Danpure and Fryer (1994)³³ 1 = mainly peroxisomal, 2 = mitochondrial and peroxisomal, 3 = mainly mitochondrial. Expression levels were taken from Danpure and Fryer (1994)³³ 0 = none or very low, 1 = low, 2 = medium, 3 = high.

Figure A6 - Results of comparative analysis of mammalian species traits using a phylogeny generated from a taxonomy (figure A6)

| | AGT distribution | AGT activity | Species diet | Body size | Gestation length | AGT expression |
|------------------|------------------|-----------------------|-----------------------------------|-------------------|-----------------------------------|-------------------|
| AGT distribution | | 0.0535657 (n = 28) | <u>0.00430</u> (n = 34) | 0.434 (n = 34) | 0.0569 (n = 34) | 0.471 (n = 31) |
| AGT activity | 1.00 | | <u>0.0135</u> (n = 42) | 0.410 (n = 42) | 0.134 (n = 42) | 0.148 (n = 41) |
| Species diet | 0.473 | 0.823 | | 0.171 (n = 42) | <u>0.0121</u> (n = 42) | 0.330 (n = 41) |
| Body size | 0.567 | 0.800 | 1 | | <u>0.00288</u> (n = 42) | 0.367 (n = 41) |
| Gestation length | 0.936 | 1 | 1 | 1 | | 0.103 (n = 41) |
| AGT expression | 1.00 | 0.907 | 1 | 1 | 1 | |

The top right hand corner of the table records p values for the comparisons, to three significant figures, and the sample sizes (n). The bottom left hand corner of the table records the maximum likelihood estimated value of the phylogeny scaling factor (λ) optimal for the null hypothesis, to three significant figures. λ was set at this value for both hypothesis in the likelihood ratio test, which generated the p value. p values bolded and underlined are significant to < 2 % level. The pattern of significance is very similar to figure 21, the relationship between AGT activity and diet is slightly less significant and the relationship between diet and gestation length has become significant.

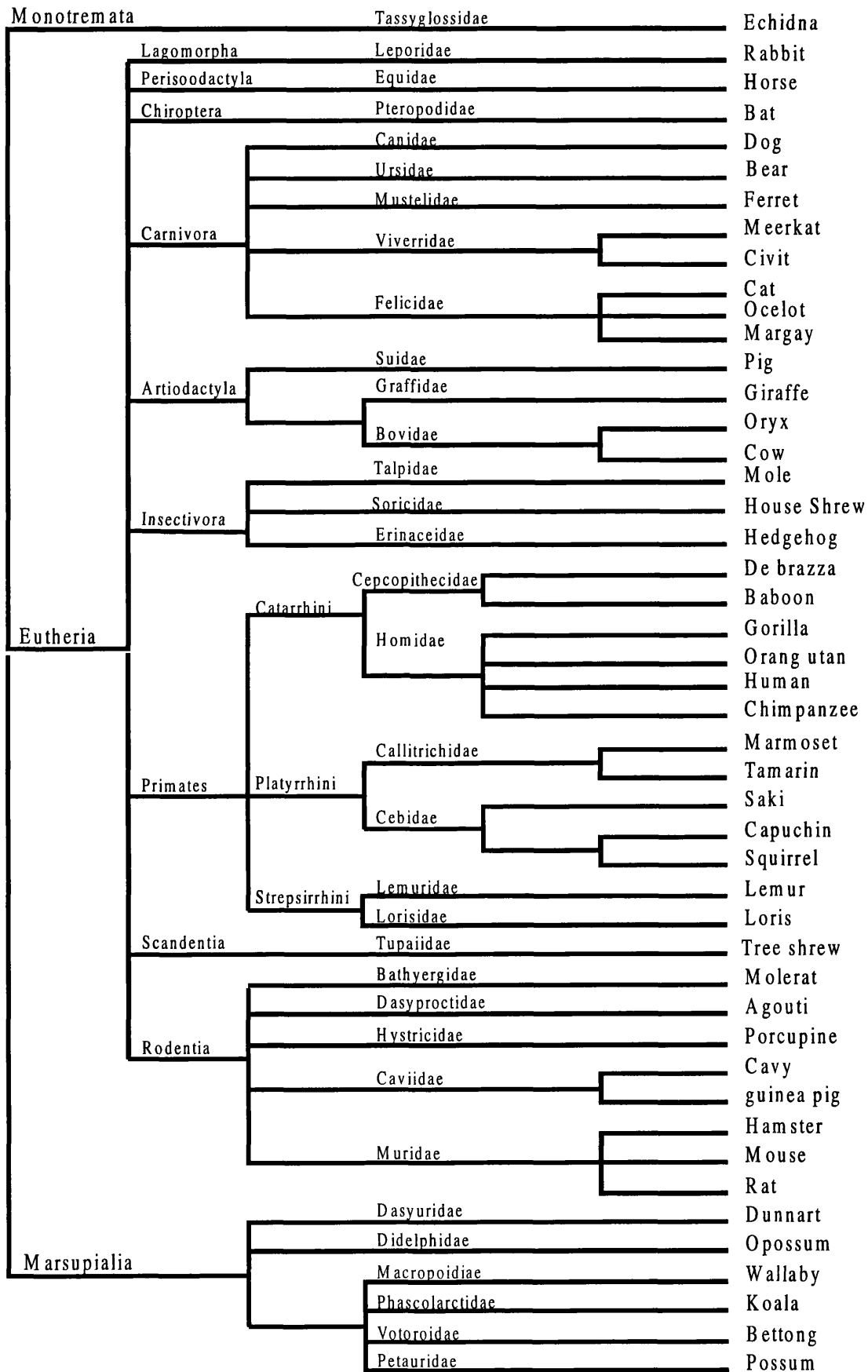


Figure A7 - Taxonomy of mammal species

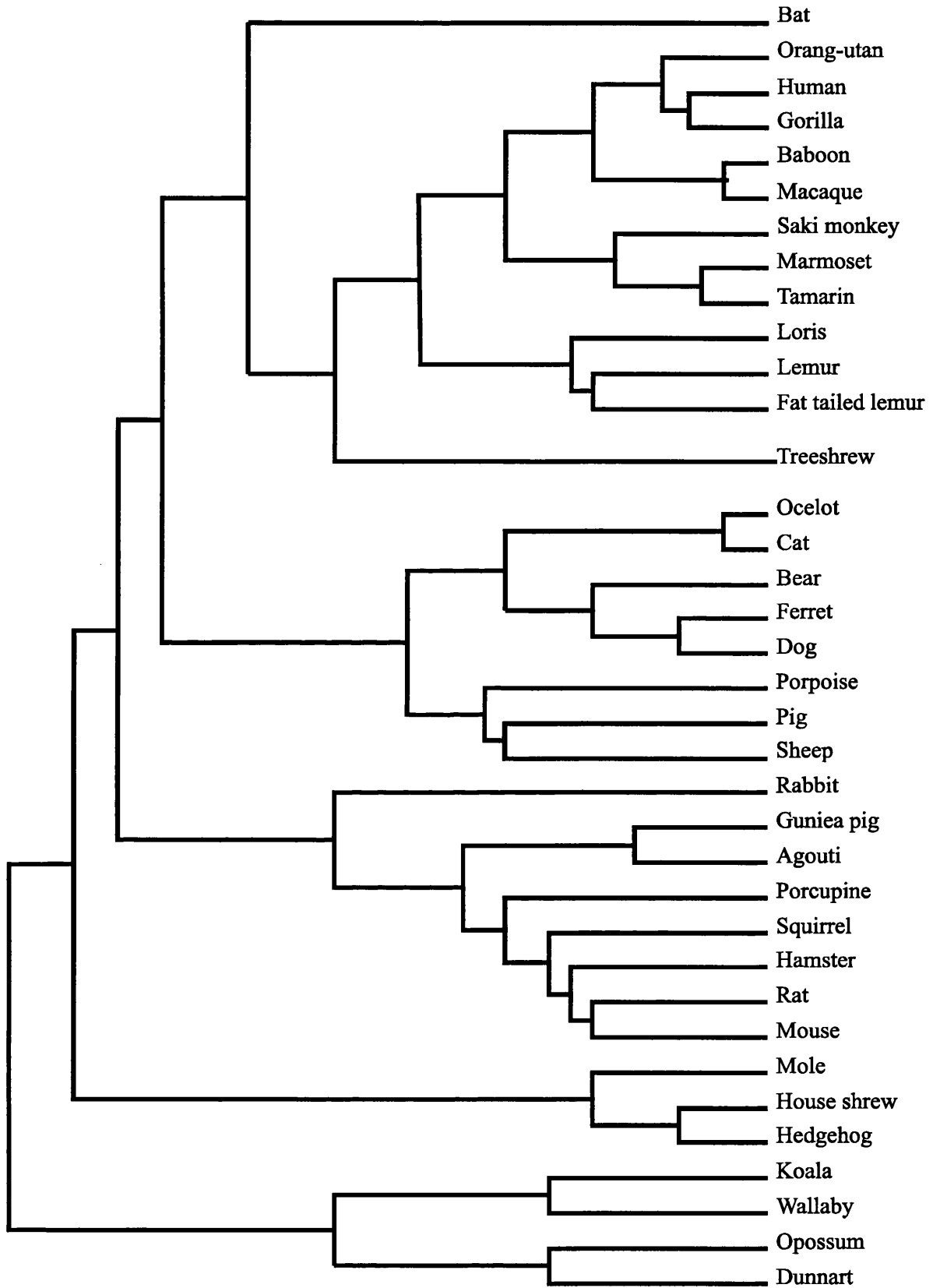


Figure A8 - Composite phylogeny of mammalian species