Published in final edited form as:
Gene. 2010 January 1; 449(1-2): 95–102. doi:10.1016/j.gene.2009.08.014.

Alpha-fetoprotein related gene (*ARG*): A new member of the albumin gene family that is no longer functional in primates

Sathyabama Naidu, **Martha L. Peterson**[a], and **Brett T. Spear**[a,*]

Department of Microbiology, Immunology & Molecular Genetics, University of Kentucky College of Medicine, Lexington, KY 40536

[a]The Markey Cancer Center, University of Kentucky College of Medicine, Lexington, KY 40536

Abstract

The serum albumin gene family is comprised of *albumin*, *alpha-fetoprotein*, *alpha-albumin* (*afamin*), and the more distantly related *Vitamin D binding protein*. These genes arose from a common ancestor through a series of duplication events, are expressed primarily in the liver and tightly linked in all species where this has been investigated. Here, we describe a fifth member of the albumin gene family that we have named *Alpha-fetoprotein Related Gene* (*ARG*) since it exhibits greatest similarity to this family member. ARG is activated in the liver perinatally, but is expressed at very low levels. The *ARG* gene is present and intact in the mouse, rat, dog and horse genomes. In contrast, the *ARG* gene in human, chimpanzee, rhesus monkey, and marmoset contains a number of mutations common to all four species, indicating that this gene has been an inactive pseudogene in primates for at least 40 million years. Low expression and aberrant splicing of the *ARG* gene in the mouse liver suggests that ARG may have less functional significance than other members of the serum albumin gene family even in species where it is still intact.

Keywords

mouse; evolution; gene duplication; liver-specific transcription; RNA splicing

1. Introduction

The serum albumin gene family is comprised of *albumin* (*Alb*), *alpha-fetoprotein* (*AFP*), *afamin* (*Afm*; also called α-albumin), and *Vitamin D Binding Protein* (*DBP*). These four genes, which arose from a common ancestral gene through a series of duplication events, encode serum proteins that are involved in the transport of numerous endogenous and exogenous ligands and, due to their high concentration, help maintain the serum osmolarity (Alexander et al., 1984; Gibbs et al., 1987; Ray et al., 1991; Belanger et al., 1994; Gibbs et al., 1998). All albumin family genes are expressed primarily in the liver, although their levels and developmental timing of expression are different. AFP is activated early in hepatogenesis and continues to be expressed at high levels in the fetal liver (Belayew and Tilghman, 1982). AFP expression is dramatically repressed at birth, and remains at very low levels in the normal adult

*Corresponding Author: Department of Microbiology, Immunology & Molecular Genetics, University of Kentucky College of Medicine, 800 Rose Street, Lexington, KY 40536-0298, 859.257.5167 (office), 859.257.8994 (fax), bspear@uky.edu.

liver but can be reactivated during periods of liver regeneration and in liver cancer (Abelev, 1971; Tilghman, 1985; Abelev and Eraiser, 1999; Spear, 1999). Alb is activated in parallel with AFP when the liver bud forms and is also highly expressed in the fetal liver; in contrast to AFP, Alb continues to be expressed at high levels in the adult liver (Gauldi et al., 1996). DBP is activated during the mid-gestational period whereas Afm expression increases during the perinatal period; both genes continue to be highly expressed in the adult liver (McLeod and Cooke, 1989; Cooke et al., 1991; Belanger et al., 1994; Lichenstein et al., 1994). Afm is downregulated in hepatocellular carcinomas in a manner that is reciprocal to AFP reactivation (Wu et al., 2000). AFP and albumin are also expressed in the yolk sac at high and low levels, respectively (Tilghman, 1985).

The genes in this small family have remained tightly linked in all animal species that have been studies to date (Juneja et al., 1982; Buetow et al., 1991; Guan et al., 1996). Alb, AFP and Afm are tandemly arranged in the same transcriptional orientation. In human, the Alb-AFP and AFP-Afm intergenic distances are 14.8 and 26.0 Kilobase pairs (Kb), respectively; in mice, these distances are 14.1 and 10.0 Kb, respectively (Fig. 1A). DPB is less tightly linked, located 1.5 megabase pairs (Mb) and 1.0 Mb upstream of the 5′ end of Alb in humans and mice, respectively, and is in the opposite transcriptional orientation. The tight linkage of these genes, particularly Alb, AFP and Afm, and their primary expression in the liver, has led to the suggestion that these genes share common regulatory elements. This idea has recently been tested and the data indicate that the enhancer region between Alb and AFP is required for AFP and Alb activation during hepatogenesis (L. Jin and B.T.S., submitted).

Computer analysis of the region 3′ of mouse *Afm* gene led us to identify a new member of the Alb gene family. Since this new gene is more similar to AFP than to other members of this gene family, we have called it *AFP-Related Gene* (*ARG*). Our analysis reveals that ARG contains 14 exons and is predicted to encode a protein of 620 amino acids. Analysis of the predicted protein is consistent with ARG being a functional member of the serum albumin gene family. However, while this gene is intact in mice, rats, horses and dogs, it can no longer encode a functional protein in primates due to multiple mutations. Despite the fact that this gene became a nonfunctional pseudogene prior to the divergence of humans and marmosets (40 million years ago; Gibbs et al., 2007; Goodman et al., 2005), the gene is still recognizable in primates. While ARG is intact and gives rise to a mature polyadenlyated mRNA in mice, it is expressed at very low levels. Furthermore, ARG exhibits an aberrant pattern of splicing between exons 1 and 2. Taken together, these data suggest that the functional importance of ARG is less than other members of the albumin gene family, even in species where the gene is still intact.

2. Materials and methods

2.1. Computational Analysis

BLAST and BLAT analysis was performed using the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/), University of California Santa Cruz (UCSC; http://genome.ucsc.edu/) and Ensembl (http://www.ensembl.org/index.html) sites using sequence information for mouse (assembly m37), human (Build 36.1), Chimpanzee (Pan troglodytes, v2.1), rhesus macaque (Macaca mulatta, draft assembly 1.0), Marmoset (Callithrix jacchus, draft assembly 2.0.2), Dog (Canis familiaris, v2.0) and horse (Equus caballus EquCab2). Pairwise DNA comparisons and determination of intron/exon boundaries were made using the NCBI website and DNA Strider. Cross-species genomic comparisons were carried out by VISTA analysis (http://genome.lbl.gov/vista/index.shtml) (Frazer et al., 2004).

2.2 Tissues and RNA Analysis

Tissues were removed from adult mice (4-6 weeks old). Livers were also obtained from embryonic day 18.5 fetuses, postnatal day 1 and postnatal day 7 mice. Tissues were used immediately or frozen in liquid nitrogen and stored at -80°C until use. Total RNA was isolated using Trizol (Invitrogen, Inc) following manufacturers instructions or using the Lithium Chloride procedure (Long and Spear, 2004). RT reactions were carried out using the Omniscript RT Kit with random hexamers (Qiagen). Standard RT-PCR reactions were carried out using the GeneAmp 9700 Thermal Cycler (Applied Biosystems) with cDNA from various adult tissues. Real-time RT-PCR was performed using the MyiQ Thermal Cycler (BioRad). The following primer pairs were used for amplification: AFP (AGCGAGGAGAAATGGTCCGG and GGACATCTTCACCATGTGG, amplicon = 544 bp); Alb (AAGACCCCAGTGAGTGAGCATG and GCTTGTGCTTCACCAGCTCAGC, amplicon = 214 bp); Afm (CAACCTGTTGCACTCTCAGTGACG and GCTCAAAGCAGTGTCTTCTGAAGG, amplicon = 159 bp); ARG (CATTTGCAACAACCAAGGCCTG and ACTTGCTGGATAAGATGGCCTG, amplicon = 650 bp); β-actin (TTTGCAGCTCCTTCGTTGCC and CGGTTGGCCTTAGGGTTCAGGGGGG, amplicon = 391 bp). For all genes, primers were from different exons to distinguish cDNA products from potentially contaminating genomic products. For Northern analysis, polyA+ mRNA was obtained from adult mouse liver and brain tissue using oligo dT columns. RNA was resolved using formaldehyde gel electrophoresis and transferred to nitrocellulose. Blots were incubated with ^{32}P-labeled probes for albumin and ARG, washed, and subjected to autoradiography. For albumin, a 213 bp probe containing exons 12 and 13 was amplified from mouse liver cDNA. For ARG, a 604 bp probe spanning exons 4-8 was amplified from mouse liver cDNA. Both 5′ RACE and 3′ RACE were carried out using the RLM-RACE kit (Ambion Bioscience, Austin, TX). To characterize cDNA products spanning ARG intron 1, cDNA from adult liver was amplified by PCR were amplified using forward primer from exon 1 (CGGCGGAACTTCATCTGAAACAATG) and reverse primer from exon 2 (TCCTAAGTTCTCCTCCAGGTGATC). These PCR amplicons, 5′ RACE and 3′ RACE products were cloned into pGEMT-Easy and sequenced.

3. Results

3.1 Identification of ARG as a novel member of the serum albumin gene family

Analysis of the DNA downstream of the mouse Afm gene, in an effort to identify conserved regions that might indicate the presence of enhancers or other control elements, revealed several regions that exhibited similarity to AFP exons. Since this DNA was not predicted to contain any genes, we explored further this region. A more detailed computational analysis indicated the presence of 13 putative exons that correspond in size and sequence similarity to exons 1-13 of the 15-exon AFP gene. All thirteen exons were flanked by GA/GT canonical splice sites. To determine whether this gene was functional and encoded a transcript, we performed RT-PCR with liver RNA using forward and reverse primer pairs from predicted exons four and eight, respectively. This resulted in an amplicon of roughly 600 base pairs; sequence analysis of the 605 bp product was identical to what would be predicted based on our computer analysis, confirming that this gene did give rise to an RNA product (see Fig. 3 below). Since the putative protein product of this gene is more similar to AFP than to other members of the serum albumin gene family (see Table I, below), we have named this novel gene *AFP-Related Gene* (*ARG*). Both 5′ RACE and 3′ RACE were also carried out to identify the ends of this gene. Based on this data, RT-PCR amplification using primers from different putative exons, and computational analysis of the mouse genome, we conclude that this gene contains 14 exons.

The primordial member of the albumin gene family is thought to have arisen originally from a seven-exon gene through several uneven crossing-over events (Brown, 1976; Alexander et

al., 1984). These events gave rise to the three-domain structure that is now seen in albumin gene family members (Fig.1B). When compared to other members of the albumin family, ARG exhibits a similar exon structure (Fig. 1C). ARG has 14 exons whereas AFP, Afm and Alb have 15 exons; however, the 15th exon in these three latter genes is non-coding. In contrast, DBP has only 13 exons due to a deletion that removes two exons that correspond to exons 12 and 13 of the other members of this family. The 3′ RACE indicates that the 14th exon in ARG is substantially larger than the 14th exon in AFP, Alb, and Afm; this is not unexpected since the 14th exon in ARG is the terminal exon and therefore does not have the size constraints that are often seen with internal exons. The predicted coding regions of ARG exons 1 and 2 are 85 and 52 nucleotides, respectively, which are identical to that seen with AFP but different from other members of this family, supporting the idea that ARG is more related to AFP than to other albumin family genes. ARG exon 3 is 139 base pairs in length, which is a different size than exon 3 from any of the other members of this gene family; based on sequence analysis this exon is more divergent than any of the other coding exons in this family with the exception of largely non-coding exon 14.

The ARG gene is 36.8 Kb in length, which would make it the largest gene in the albumin gene family although it is only slightly larger than Afm (Fig. 1A). The small cluster of ARG transcription start sites (see below) are located 7.6 Kb downstream of the 3′ end of Afm, and the 3′ end of ARG exon 14 is only 5.2 Kb away from the 3′ end of RASSF6.

3.2 Comparison of the predicted ARG protein with other members of the albumin family

Sequence analysis indicates that ARG is an intact, functional gene in mice. We therefore analyzed the protein predicted to be encoded by this gene. Translation from an ATG in exon 1 (translation start site based on comparisons with the other proteins in the albumin family) indicates that ARG encodes a 620 amino acid protein (Fig. 2). A pair-wise comparison indicates that ARG is more slightly more similar to AFP than to Alb and Afm, although the similarity between ARG and other family members is comparable to other pair-wise combinations within this group of related proteins (Table 1). A hallmark of this family of proteins (not including DBP) is the presence of twenty-eight cysteine residues involved in the formation of 14 disulfide bonds that are important for the tertiary structure of these proteins (Brown, 1976). We have found that all twenty-eight cysteine residues are present in ARG (Fig. 2), suggesting that this protein would have a similar domain structure as other proteins in this family.

3.3 Developmental and tissue-specific ARG expression

Members of the albumin gene family are expressed primarily in the liver, although developmental patterns of expression do vary. AFP and alb are expressed early in liver development; Alb continues to be expressed at high levels in the adult liver whereas AFP exhibits a dramatic decline in expression at birth. DBP and Afm are activated late during fetal development and soon after birth, respectively. RT-PCR was used to monitor the developmental profile of ARG expression in the liver. This analysis indicated that ARG was expressed at low levels in the fetal liver and was activated late in gestation (Fig. 3A). This pattern of expression is similar to that seen for Afm. ARG is not expressed in the yolk sac, an extraembryonic site where AFP is highly expressed, Alb and DBP are expressed at low levels, and Afm is transcriptionally silent (McLeod and Cooke, 1989). While an increase in ARG expression is observed during the perinatal period, ARG mRNA levels are substantially lower than other albumin family members in the adult liver; quantitation of real-time data normalized to β-actin indicates that ARG is expressed at levels that are roughly 20- and 5-fold lower that Alb and Afm, respectively, in the adult liver. A survey of adult mouse tissues indicates that in addition to the liver, ARG is also expressed at low but detectable levels in the kidney but not in any of the other tissues examined (Fig. 3B). Other studies have found low but detectable levels of albumin, AFP and DBP in the kidney (McLeod and Cooke, 1989).

Since RT-PCR was used for our analysis of ARG expression, we could not be certain that a full-length ARG mRNA was synthesized. Our genomic analysis would predict that the ARG gene would encode an mRNA of roughly 2400 nucleotides in length. To confirm the presence of this product, northern analysis was performed with oligo-dT selected mRNA from adult mouse liver. We observed a message of the correct size that was detected with an ARG probe in liver samples but not in mRNA from brain tissue (Fig. 3C). This confirms that a full-length, polyadenylated product was generated by the mouse ARG gene and confirms RT-PCR data indicating that the ARG mRNA was significantly less abundant than albumin mRNA.

3.4 Variation in transcription initiation and intron 1 splicing of ARG

The 5′ UTR region of other members of the mouse albumin gene family range from 34 - 47 bp (Fig. 1C). To characterize the 5′ UTR of ARG and identify the transcription start site, we performed 5′ RACE. This resulted in three predominant amplicons, suggesting the presence of several transcription start sites. The size of these products indicated that the start sites were roughly 100-200 bp upstream of the ATG translation start site. The 5′ RACE products were cloned into pGEM-T Easy and sequenced. Two clones ended at -187 (relative to the “A” of the putative ATG translation start site being designated as +1), three clones ended at -162, one ended at -122, and four ended at -95 (Fig. 4). This data would suggest that multiple transcription start sites are utilized for initiation of ARG transcription, which is in contrast to other members of this family in which a single transcription start site is primarily used.

The intron-exon structure of the mouse ARG gene was deduced from mouse genome sequence analysis and comparison to the AFP intron-exon structure. Most of the ARG splice junctions were confirmed by sequencing of RT-PCR products and were consistent with the predicted boundaries. However, we found products of two different sizes, of roughly equal abundance, when we amplified across intron 1. To determine the basis for this data, RT-PCR products were cloned and sequenced. Four of the clones were spliced at the predicted AG/GT boundaries (Fig. 4). However, five clones were spliced at a GT that was 33 bp upstream of the predicted 5′ splice site to an AG that was 4 bp downstream of the predicted 3′ splice site; one clone used the predicted 5′ splice site and the downstream 3′ splice site (Fig. 4). These aberrant transcripts would all shift the reading frame and therefore be unable to encode a functional ARG protein.

3.5 ARG is a functional gene in rat, horse and dog but not in primates

The presence of ARG in mice led us to consider whether this gene was present in other species. Analysis of the rat, horse, and dog genome databases identified an intact ARG gene, highly similar to the mouse gene, in all three of these species. Translation of these three genes would encode proteins that are highly similar to the predicted mouse ARG protein (Fig. 5). Importantly, all the 28 highly conserved cysteine residues that are involved in disulfide bond formation are present. The amino acid conservation (those that are identical and those that are similar) of ARG between mouse and rat, mouse and dog and mouse and horse is 90%, 79% and 78%, respectively. This is comparable to the cross-species conservation of AFP and Afm, and less than that seen for Alb (Table 2).

The presence of an intact ARG gene in several species led us to ask whether ARG was present in humans and other primates. Computer analysis of the human, chimp, rhesus and marmoset genomes revealed the presence of the ARG gene downstream of Afm. However, detailed analysis revealed that ARG has become a nonfunctional pseudogene in humans and the other primates analyzed. An in-frame stop codon exists in exon 1 in all four species; this mutation in humans was confirmed by our own sequence analysis. Furthermore, frameshifts occur in six exons and seven of the twenty existing splice junctions contain inactivating mutations (Fig. 6 and supplemental data). The counterparts to exons 4, 7 and 12 are absent in primates. Furthermore, primate exons 6 and 11 contain a non-LTR Line and non-LTR Sine elements,

respectively. RT-PCR analysis of human adult liver RNA found no evidence of ARG transcripts, providing additional evidence that this pseudogene is no longer expressed in primates (data not shown).

4. Discussion

Previous studies have demonstrated that the albumin gene family in mammals contains four members. These genes – Alb, AFP, Afm and DBP – arose from a series of duplications of an ancestral gene and share similar exon structures, and encode structurally and functionally similar proteins (Alexander et al., 1984; Gibbs et al., 1987; Ray et al., 1991; Belanger et al., 1994; Gibbs et al., 1998). It has been predicted that the initial duplication event gave rise to DBP and a second gene; duplication of this second gene gave rise to Alb and the precursor to AFP and Afm (Brown, 1976; Gibbs et al., 1998). Here, we describe a new gene, ARG, that is the fifth member of this gene family. ARG is the largest gene in this family and is located 3′ of Afm. While ARG is clearly a member of this family, several notable differences with other family members were observed. First, 5′ RACE indicates that the 5′ UTR of ARG exon 1 is substantially longer than the 5′ UTR of other genes in this family. An additional difference is that ARG contains multiple transcription start sites that span 100 bp; other genes in this family have a single start site or several adjacent start sites. Also, ARG is the only member of this family that lacks exon 15. Since this in a non-coding exon in other family members, the absence of this exon does not impact the protein coding capacity of ARG compared to other family members. The unusually large exon 14 of ARG is not surprising since this is no longer an internal exon. Despite these differences, the mouse ARG exon structure is very similar to other albumin family genes.

The mouse ARG gene is predicted to encode a protein of 620 amino acids. ARG is more similar to AFP than to other family members. Significantly, the presence of all 28 cysteine residues that are involved in the formation of 14 disulfide bonds suggests that ARG would have a similar structure to other albumin family proteins (Brown, 1976; Alexander et al., 1984). Orthologues to the mouse ARG gene can be found in other mammals. When the predicted ARG proteins are compared between different species, ARG is as conserved as AFP and Afm, but less conserved that Alb (Table 2).

While the ARG protein is most similar to AFP, expression of ARG is most similar to Afm. AFP and Alb are activated early in hepatogenesis; AFP is silenced at birth whereas Alb continues to be expressed in the adult liver (Belayew and Tilghman, 1982). Afm is activated late in gestation and continues to be expressed in the adult liver (Belanger et al., 1994; Lichenstein et al., 1994). We found that ARG is also activated later in fetal development, similarly to Afm. However, ARG is expressed at substantially lower levels than other albumin family genes. ARG is also expressed at low levels in the kidney, a tissue where AFP and Alb are also expressed at low levels. Despite its low expression, primary ARG transcripts are processed to a mature polyadenylated mRNA that can be detected by northern analysis.

While the ARG gene can be found in different species, we identified a large variety of mutations that have rendered ARG a non-functional pseudogene in primates. Of the 14 ARG exons, six have small insertions or deletions that would lead to frameshifts. Three exons are completely absent. Retrotransposons have integrated into two exons. Seven splice sites have been changed from the canonical AG/GT sequences. Finally, a stop codon exists in exon 1. Nearly all of these mutations exist in the ARG gene from marmoset, rhesus, chimp, and humans, which would indicate that these changes occurred at least 40 million years ago (Goodman et al., 2005). It seems reasonable to assume that a single mutation inactivated the primate ARG gene, and subsequent mutations accumulated. Since many of the mutations are found in all four species, it is not possible to determine the initial mutation. Analysis of the tarsier and lemur genome

assemblies failed to identify ARG sequences, but it will be of interest to determine if ARG is also inactive in these and other prosimians to elucidate the progression of events that led to ARG inactivation.

α1,3galactosyltransferase (GGTA1) is another examples of a pseudogene in humans that is still intact in mice. The human GGTA1 gene contains numerous frame shifts and point mutations (Koike et al., 2002). GGTA1 gene sequences have been analyzed in a number of different species (Koike et al., 2007). This analysis revealed that the GTTA1 gene is still intact in all species investigated, including prosimians and New World monkeys (including marmoset), but has become an inactive pseudogene in Old World Monkeys and Apes (including humans). Since most mutations in the primate ARG genes are found in marmoset, the initial ARG inactivating mutation occurred earlier in evolution (at least 40 million years ago) than in GGTA1 (between 40 and 25 million years ago).

While ARG does appear to be intact in mouse, rat, dog, and horse, analysis of mouse ARG expression suggest that this gene may have little functional significance in these species. First, ARG is expressed at very low levels in the mouse liver in contrast to other albumin family genes. In this regard, it is interesting to note that the Alb, AFP, Afm and DBP promoters all contain Hepatocyte Nuclear Factor 1 (HNF1) binding sites that are known to be important for promoter activity (Cereghini et al., 1988; Feuerman et al., 1989; Song et al., 1998; H. Liu and B.T.S., manuscript in preparation), but analysis of region upstream of the mouse ARG gene did not identify any consensus HNF1 sites (B.T.S., unpubl. obs). The aberrant splicing across intron 1 of the mouse ARG gene, due to the use of incorrect splice donor and acceptor sites, would decrease the proportion of ARG transcripts that encode ARG protein. Our analysis revealed correct splicing of several other ARG introns, although we have not test all introns for aberrant splicing. Finally, our preliminary analysis of the bovine genome draft assembly database suggests that the cow ARG gene is intact but contains several frameshifts and point mutations, including some that encode premature stop codons; these mutations are different than those seen in the primate genes but would suggest that the cow ARG is also a pseudogene (B.T.S., unpubl. obs.).

In summary, we have identified ARG as a new member of the albumin gene family. ARG can no longer encode a functional protein in primates and its functional significance may be less than other members of the albumin gene family, even in species where the structural gene is still intact. This may be due to the lack of selective pressure to maintain ARG function, which in turn could be explained by functional redundancy between different albumin family members. The possibility of this overlap is suggested by the fact that the congenital absence of albumin has been observed in humans, and spontaneously occurring albumin-deficient rats are viable (Nagase et al., 1979; Ruffner and Dugaiczyk, 1988). Mice lacking AFP, by targeted deletion of the AFP gene, are also viable although homozygous AFP-deficient females are infertile (Gabant et al., 2002). Despite the fact that ARG is not functional in primates, this gene continues to be highly conserved in mammals, suggesting selective pressure and raising the possibility that it does play an important, although not yet identified, role in the species where it remains functional. Further analysis in additional species should help elucidate the phylogenetic relationship of ARG to other albumin genes and the molecular events that have lead to its inactivation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Michelle Glenn and Amanda Ribble for technical assistance, and members of the Peterson and Spear labs for helpful discussion. This work was supported by Public Health Service Grants DK-51600 and DK-074816.

References

Abelev GI. Alpha-fetoprotein in oncogenesis and its association with malignant tumors. Adv Cancer Res 1971;14:295–358. [PubMed: 4107670]

Abelev GI, Eraiser TL. Cellular aspects of alpha-fetoprotein reexpression in tumors. Semin Canc Biol 1999;9:95–107.

Alexander F, Young PR, Tilghman SM. Evolution of the albumin: alpha-fetoprotein ancestral gene from the amplification of a 27 nucleotide sequence. J Mol Biol 1984;173:159–73. [PubMed: 6200600]

Belanger L, Roy S, Allard D. New albumin gene 3′ adjacent to the alpha-fetoprotein locus. J Biol Chem 1994;269:5481–5484. [PubMed: 7509788]

Belayew A, Tilghman SM. Genetic analysis of α-fetoprotein synthesis in mice. Mol Cell Biol 1982;2:1427–1435. [PubMed: 6186903]

Brown JR. Structural origins of mammalian albumin. Fed Proc 1976;35:2141–4. [PubMed: 947793]

Buetow KH, Shiang R, Yang P, Nakamura Y, Lathrop GM, White R, Wasmuth JJ, Wood S, Berdahl LD, Leysens NJ, et al. A detailed multipoint map of human chromosome 4 provides evidence for linkage heterogeneity and position-specific recombination rates. Am J Hum Genet 1991;48:911–25. [PubMed: 1673289]

Cereghini S, Blumenfeld M, Yaniv M. A liver-specific factor essential for albumin transcription differs between differentiated and dedifferentiated rat hepatoma cells. Genes Dev 1988;2:957–74. [PubMed: 3169549]

Cooke NE, McLeod JF, Wang XK, Ray K. Vitamin D Binding Protein: genomic structure, functional domains, and mRNA expression in tissues. J Steroid Bochem Mol Biol 1991;40:787–793.

Feuerman MH, Godbout R, Ingram RS, Tilghman SM. Tissue-specific transcription of the mouse α -fetoprotein gene promoter is dependent on HNF-1. Mol Cell Biol 1989;9:4204–4212. [PubMed: 2479822]

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res 2004;32:W273–9. [PubMed: 15215394]

Gabant P, Forrester L, Nichols J, Van Reeth T, De Mees C, Pajack B, Watt A, Smitz J, Alexandre H, Szpirer C, Szpirer J. Alpha-fetoprotein, the major serum protein, is not essential for embryonic development but is required for female fertility. Proc Natl Acad Sci, USA 2002;99:12865–12870. [PubMed: 12297623]

Gauldi R, Bossard P, Zheng M, Hamada Y, Coleman JR, Zaret KS. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. Genes and Development 1996;10:1670–1682. [PubMed: 8682297]

Gibbs PE, Witke WF, Dugaiczyk A. The molecular clock runs at different rates among closely related members of a gene family. J Mol Evol 1998;46:552–561. [PubMed: 9545466]

Gibbs PE, Zielinski R, Boyd C, Dugaiczyk A. Structure, polymorphism, and novel repeated DNA elements revealed by a complete sequence of the human alpha-fetoprotein gene. Biochemistry 1987;26:1332–43. [PubMed: 2436661]

Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, et al. Evolutionary and biomedical insights from the rhesus macaque genome. Science 2007;316:222–34. [PubMed: 17431167]

Goodman M, Grossman LI, Wildman DE. Moving primate genomics beyond the chimpanzee genome. Trends in Genetics 2005;21:511–517. [PubMed: 16009448]

Guan XJ, Arhin G, Leung J, Tilghman SM. Linkage between vitamin D-binding protein and alpha-fetoprotein in the mouse. Mamm Genome 1996;7:103–106. [PubMed: 8835525]

Juneja RK, Sandberg K, Andersson L, Gahne B. Close linkage between albumin and vitamin D binding protein (gc) loci in chicken: a 300 million year old linkage group. Genet Res 1982;40:95–8. [PubMed: 6183167]

Koike C, Fung JJ, Geller DA, Kannagi R, Libert T, Luppi P, Nakashima I, Profozich J, Rudert W, Sharma SB, Starzl TE, Trucco M. Molecular basis of evolutionary loss of the alpha 1,3-galactosyltransferase gene in higher primates. J Biol Chem 2002;277:10114–20. [PubMed: 11773054]

Koike C, Uddin M, Wildman DE, Gray EA, Trucco M, Starzl TE, Goodman M. Functionally important glycosyltransferase gain and loss during catarrhine primate emergence. Proc Natl Acad Sci U S A 2007;104:559–64. [PubMed: 17194757]

Lichenstein HS, Lyons DE, Wurfel MM, Johnson DA, McGinley MD, Leidli JC, Trollinger DB, Mayer JP, Wright SD, Zukowski MM. Afamin is a new member of the albumin, alpha-fetoprotein, and vitamin D-binding protein gene family. J Biol Chem 1994;269:18149–54. [PubMed: 7517938]

Long L, Spear BT. FoxA proteins regulate H19 endoderm enhancer E1 and exhibit developmental changes in enhancer binding in vivo. Molec Cell Biol 2004;24:9601–9. [PubMed: 15485926]

McLeod JF, Cooke NE. The vitamin D-binding protein, alpha-fetoprotein, albumin multigene family: detection of transcripts in multiple tissues. J Biol Chem 1989;264:21760–9. [PubMed: 2480956]

Nagase S, Shimamune K, Shumiya S. Albumin-deficient rat mutant. Science 1979;205:590–1. [PubMed: 451621]

Ray K, Wang XK, Zhao M, Cooke NE. The rat vitamin D binding protein (Gc-globulin) gene. Structural analysis, functional and evolutionary correlations. J Biol Chem 1991;266:6221–9. [PubMed: 2007578]

Ruffner DE, Dugaiczyk A. Splicing mutation in human hereditary analbuminemia. Proc Natl Acad Sci U S A 1988;85:2125–9. [PubMed: 3353369]

Song YH, Ray K, Liebhaber SA, Cooke NE. Vitamin D-binding protein gene transcription is regulated by the relative abundance of hepatocyte nuclear factors 1alpha and 1beta. J Biol Chem 1998;273:28408–18. [PubMed: 9774468]

Spear BT. Alpha-fetoprotein gene regulation: Lessons from transgenic mice. Seminars in Cancer Biology 1999;9:109–116. [PubMed: 10202132]

Tilghman SM. The structure and regulation of the mouse α-fetoprotein and albumin genes. Oxford Surveys in Eukaryotic Genes 1985;2:160–206.

Wu GX, Lin YM, Zhou TH, Gao H, Pei G. Significant down-regulation of alpha-albumin in human hepatoma and its implication. Cancer Letters 2000;160:229–236. [PubMed: 11053653]

Abbreviations

Arg	alpha-fetoprotein related gene
Alb	albumin
AFP	alpha-fetoprotein
Afm	Alpha-albumin
DBP	vitamin D binding protein
kb	kilobase pairs
Mb	megabase pairs
bp	base pair
RT	Reverse Transcriptase
PCR	Polymerase Chain Reaction
RACE	rapid amplification of cDNA ends
UTR	untranslated region
HNF-1	hepatocyte nuclear factor 1

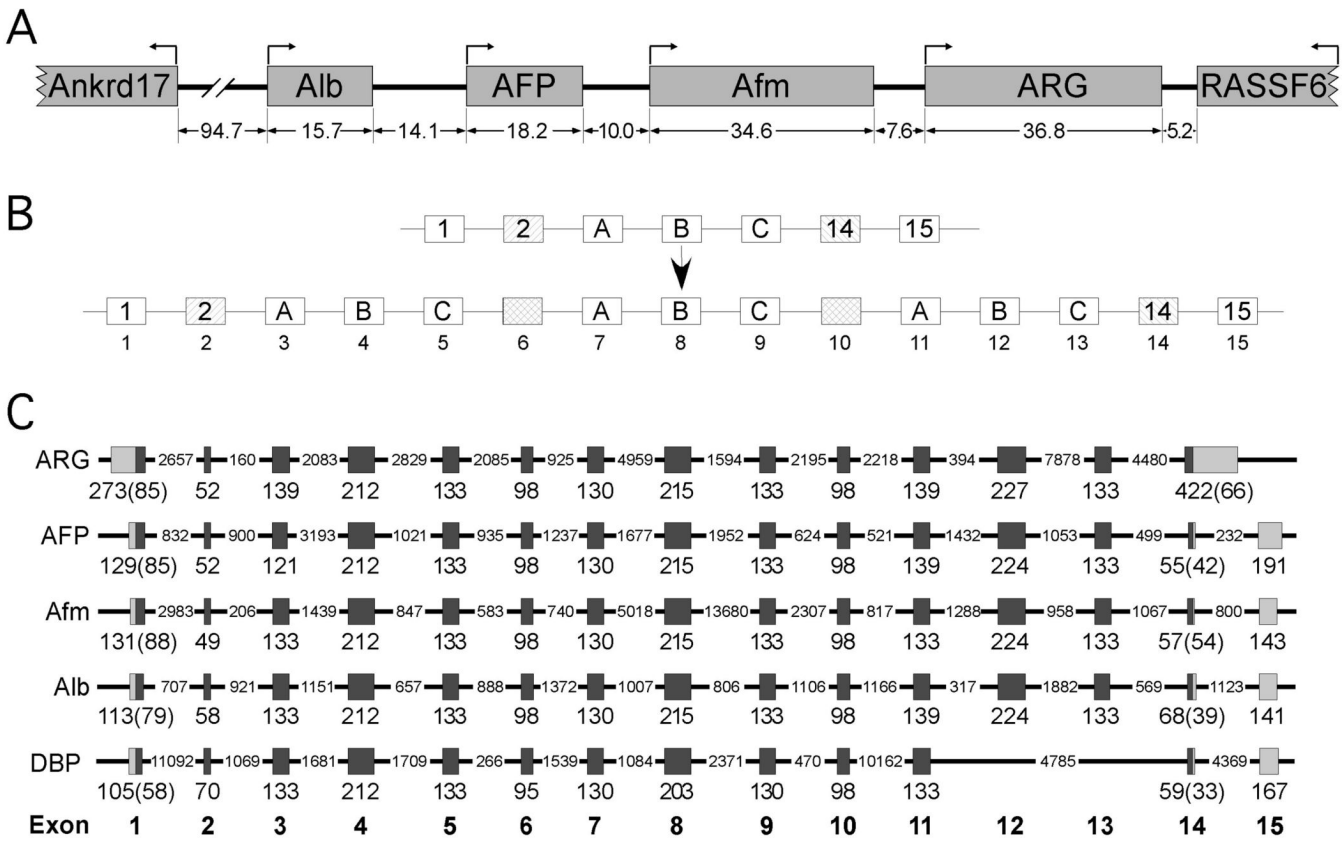

Figure 1. Structure and genomic organization of the albumin family of genes

(A) Organization of the mouse albumin locus. The relative location of the albumin genes are shown, numbers below indicate the size of the genes and intergenic regions (in kb). The arrows represent the start of transcription. The genes flanking this locus are Ankrd17 and RASSF6; the DBP gene is located approximately 1.5 Mb upstream of the Alb gene and is in the opposite transcriptional orientation. ***(B)*** The original 15-exon gene of the albumin gene family is predicted to have arisen from a primordial 7-exon gene through a series of duplications. This model proposes that exons A, B and C of the primordial gene gave rise to exons 3/7/11, 4/8/12, and 5/9/13, respectively (Brown, 1976; Alexander et al., 1984). ***(C)*** Exon-intron comparison of members of the mouse albumin gene family. The names of the five members of this gene family are on the left. The exons are designated as gray boxes, with dark gray representing coding regions. The numbers below the exons indicate the exon size (in bp); numbers in parentheses for exons 1 and 14 represent the number of coding nucleotides. The numbers between exons represent intron length (in bp). Exons 12 and 13 have been deleted from DBP, the most distantly related member of the albumin gene family.

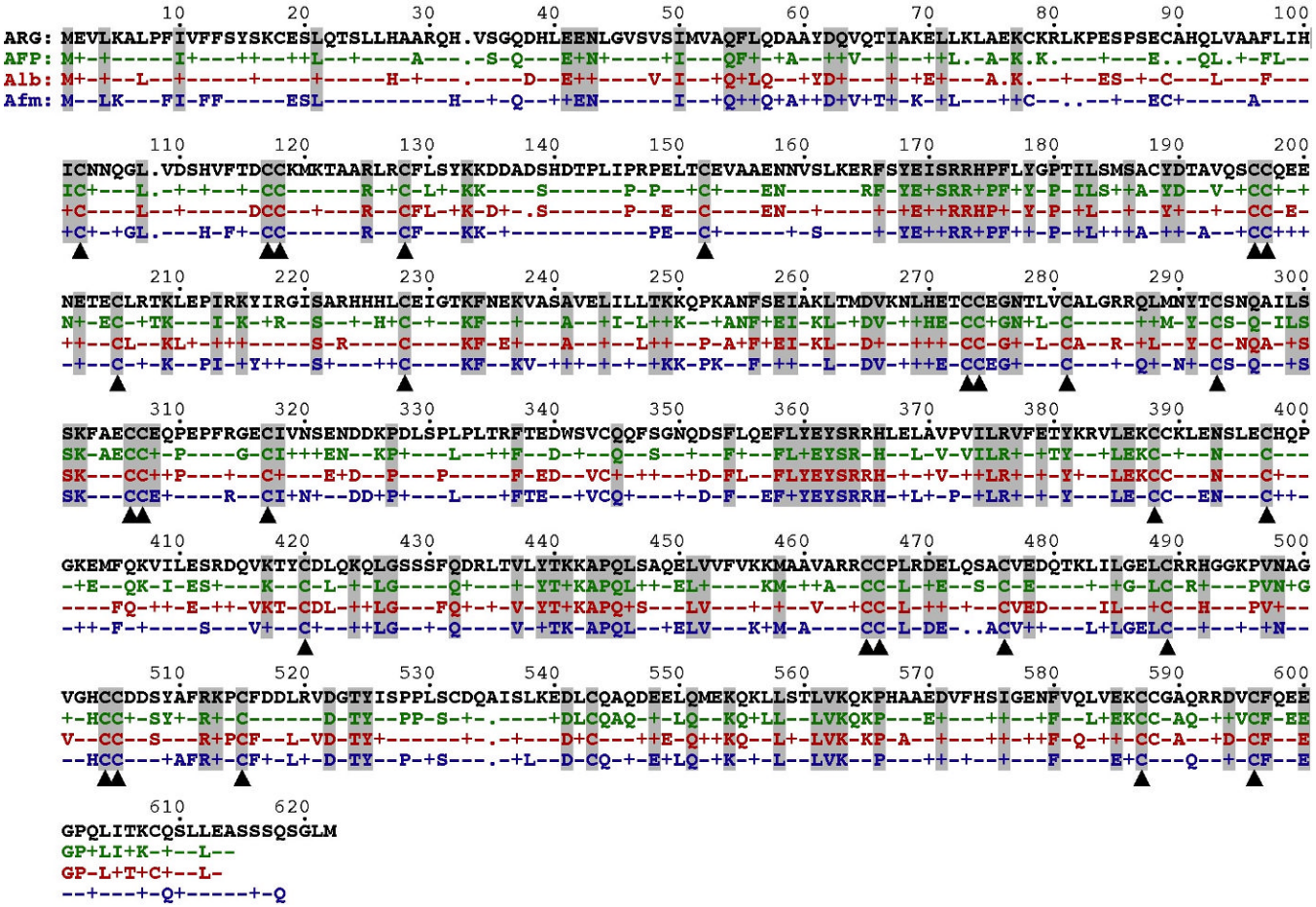

Figure 2. Alignment of ARG, AFP, Alb and Afm proteins in mice

The amino acids in AFP (green), Alb (red) and Afm (blue) that are identical to those in ARG (black) are designated by the one letter code; those that are similar are designated by "+". Residues that are identical or similar in all four proteins are highlighted in grey. The twenty-eight arrowheads designate the cysteine residues that are required for the fourteen disulfide bonds involved in maintaining the tertiary structure of these proteins.

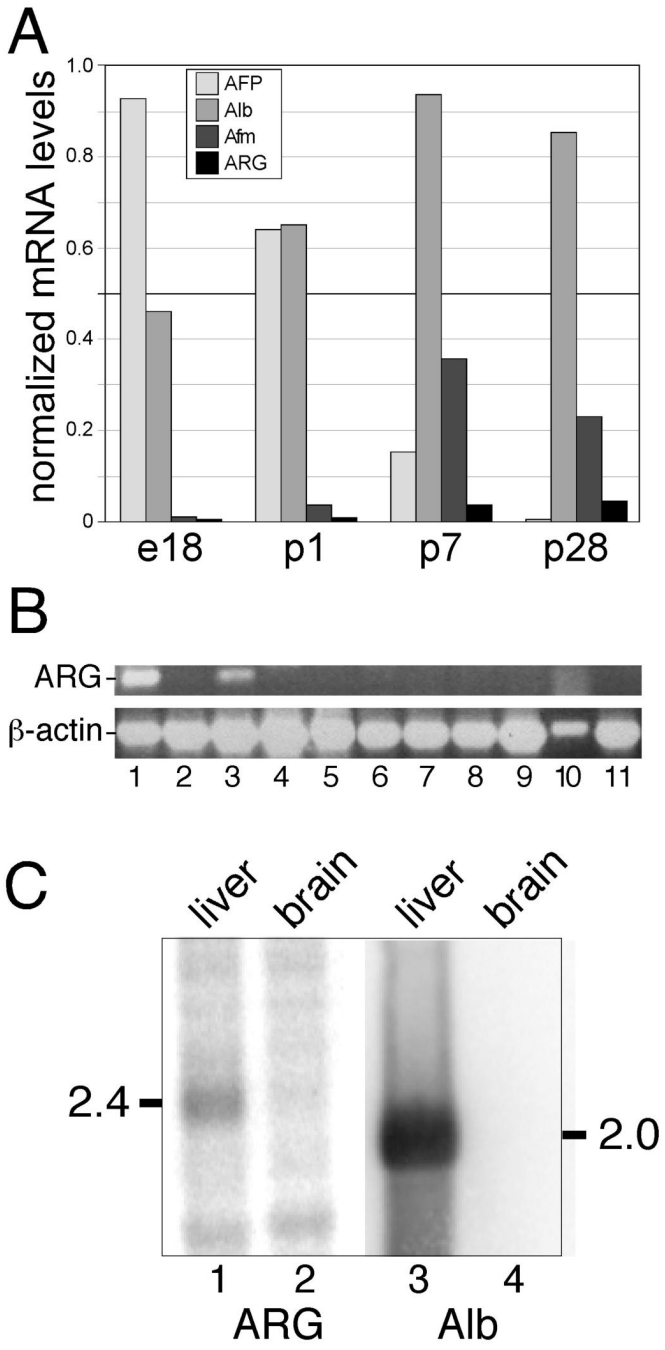

Figure 3. Tissue-specific and developmental expression of the mouse ARG gene

(A) Developmental changes in the expression of ARG, AFP, Alb, and Afm in the perinatal liver. Total RNA was isolated from the livers of embryonic day 18 (e18) and postnatal day 1 (p1), p14 and p28 livers. The levels of mRNA were analyzed by real-time RT-PCR and normalized against β-actin levels. AFP levels show a dramatic decline after birth, whereas Alb exhibits a modest increase during this period. ARG and Afm are both activated during the perinatal period; ARG levels are substantially lower that the levels of AFP, Alb, and Afm. ***(B)*** ARG expression in a panel of adult mouse tissues. RT-PCR was used to monitor ARG expression in liver (lane 1), spleen (lane 2), kidney (lane 3), adipose (lane 4), skeletal muscle (lane 5), lung (lane 6), heart (lane 7), pancreas (lane 8), brain (lane 9), thymus (lane 10), and

small intestine (lane 11). ARG transcripts were detected only in liver and kidney. ***(C)*** Northern analysis of ARG expression in the adult liver and brain. PolyA$^+$ mRNA from liver (lanes 1 and 3) and brain (lanes 2 and 4) was resolved using formaldehyde agarose gel electrophoresis. The RNA was transferred to nitrocellulose and hybridized to cDNA probes for ARG (lanes 1 and 2) or Alb (lanes 3 and 4). Bands corresponding to the predicted full-length ARG mRNA (~2400 nucleotides) and full-length Alb mRNA (2028 nucleotides) are detected in liver but not in brain samples.

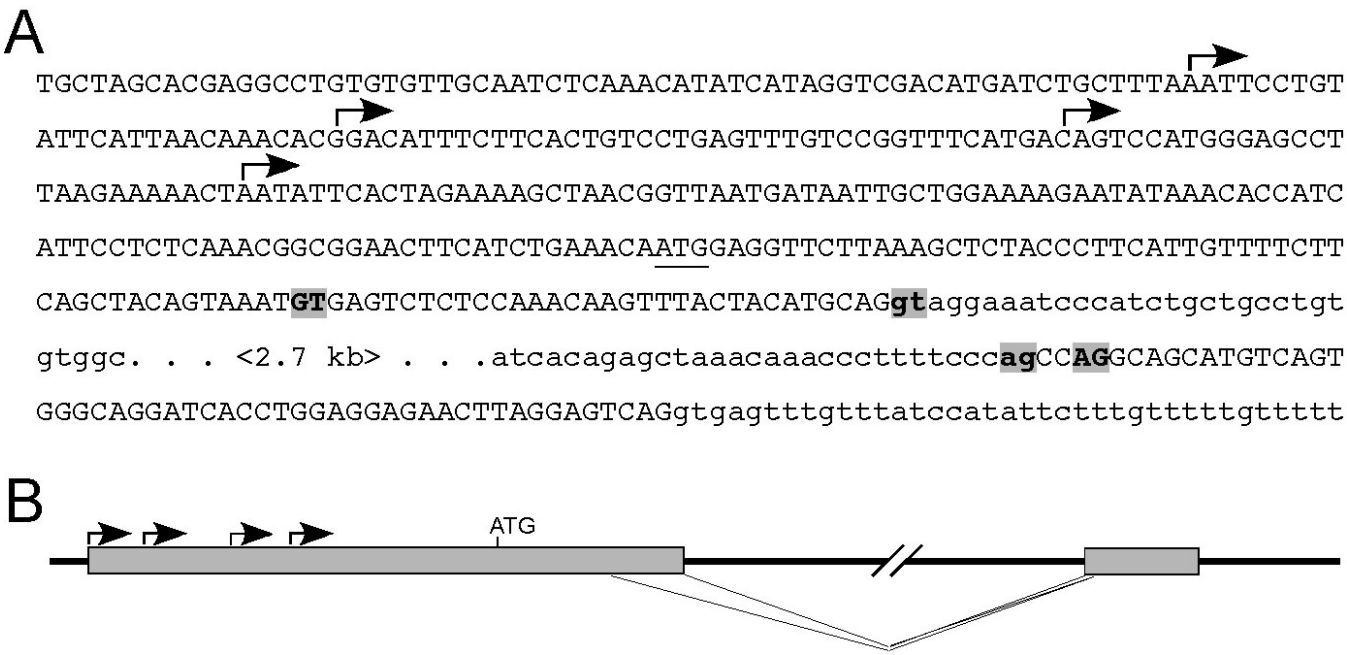

Figure 4. Organization of transcripts arising from the 5′ end of the ARG gene

(A) DNA sequence of the 5′ end of the ARG gene. The upper case letters correspond to exons 1 and 2; lower case letters correspond to introns 1 and 2 (the entire ~2.7 kb intron 1 is not shown). The arrows represent the transcription start sites as determined by 5′ RACE. The underlined ATG represents the predicted translation start site. The lower case “gt” and “ag” residues on the 3′ end of exon 1 and 5′ end of exon 2, respectively (bold and in gray boxes) correspond to the predicted splice junctions based on comparisons with other members of the Albumin gene family; sequencing of RT-PCR products indicate that these splice acceptor and donor sites are used. The “GT” and “AG” residues in exons 1 and 2, respectively (bold and in gray boxes) represent additional splice donor and acceptor sites as determined by sequencing of RT-PCR products. ***(B)*** Diagram of exons 1 and 2 (gray boxes) of the 5′ of the ARG gene, showing the multiple start sites (arrows) and patterns pf splicing. Differential use of transcription start sites and splice junctions give rise to several mRNA variants of the ARG gene.

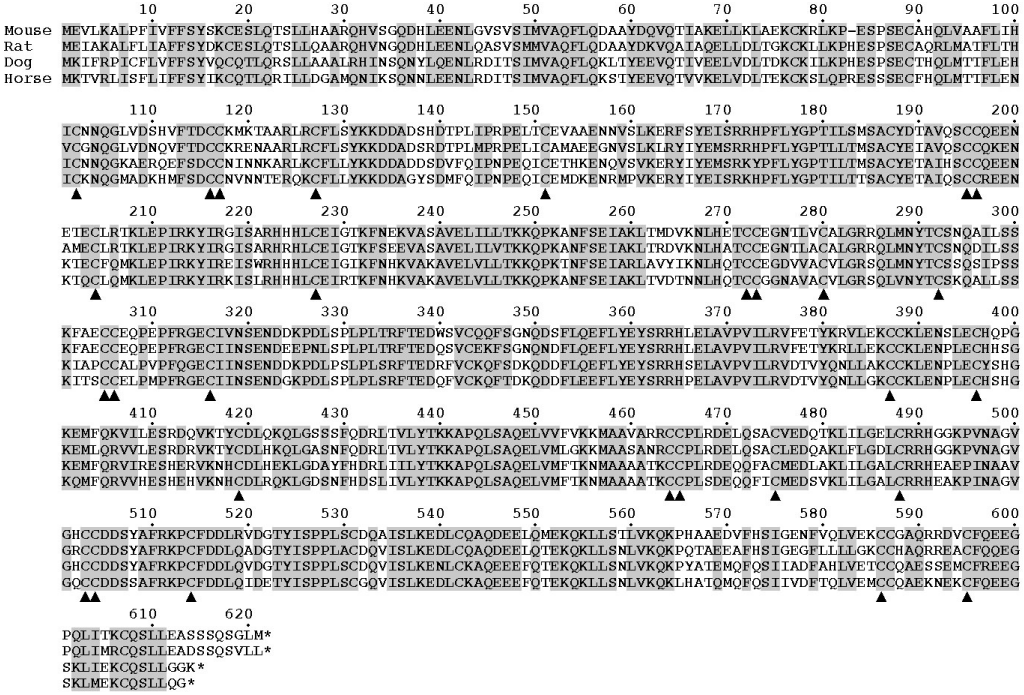

Figure 5. Alignment of ARG protein from mouse rat, dog and horse

The ARG amino acids that are conserved (or similar) between all four species are highlighted in gray. The twenty-eight arrowheads designate the cysteine residues that conserved between Alb, AFP, Afm and ARG proteins.

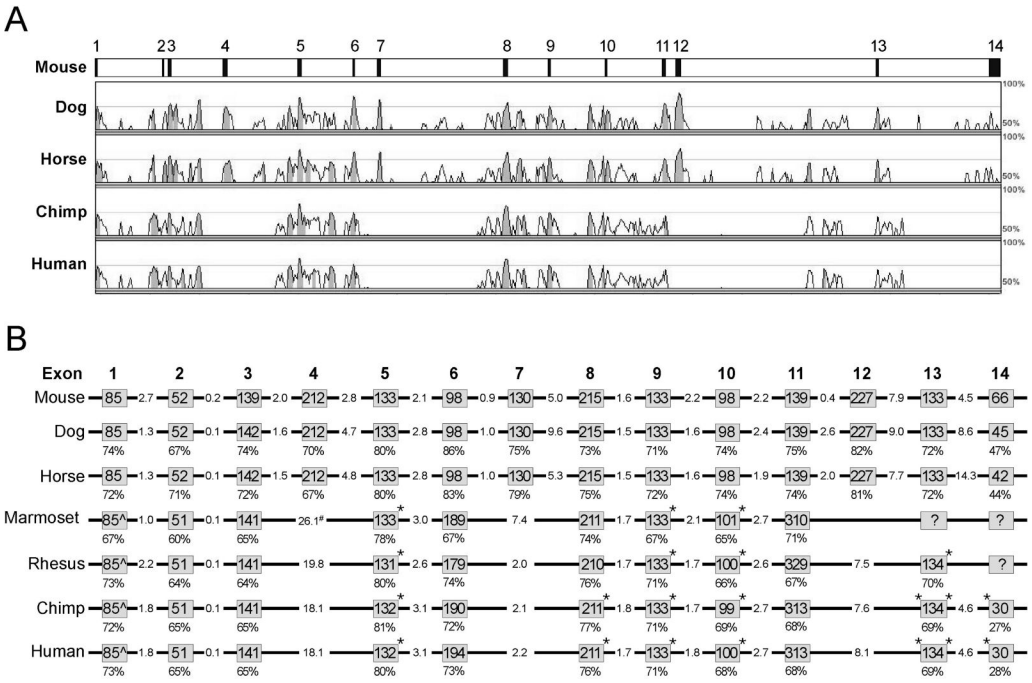

Figure 6. Comparison of ARG gene between different species

(A) The dog, horse, chimp and human ARG genes were aligned to the mouse ARG gene using VISTA analysis. Exons are conserved between all species, although exons 4, 7 and 12 are absent from primates. ***(B)*** Exon-intron regions were compared between mouse, dog, horse, marmoset, rhesus, chimp and human. Gray boxes represent exonic regions; numbers inside boxes correspond to exon length (in bp). The numbers between exons represent intron length (in Kb). The numbers below each exon represent sequence the percent identity of each exon to the corresponding mouse exon. The asterisk indicates that the canonical AG or GT splice sites have been mutated. The increased lengths of primate exons 6 and 11 are due to retrotransposon insertions.

Table 1

Pairwise comparison of mouse albumin family proteins

Protein Pair	% Identical	% Similar	% Total
ARG - AFP	37.3	16.9	54.2
ARG - Alb	31.4	17.0	48.4
ARG - Afm	31.2	18.6	50.7
AFP - Alb	34.3	19.7	54.0
AFP - Afm	32.2	19.2	51.4
Alb - Afm	29.4	17.1	46.5

Pair-wise alignment of predicted full-length mouse proteins was performed. The % of amino acid residues that were identical or similar were determined; these numbers were added to give the % total amino acid identity.

Table 2

Comparison of albumin family proteins between species

Pair	ARG	AFP	Afm	Alb
Mouse - Rat	82	82	87	90
Mouse - Dog	67	65	64	72
Mouse - Horse	66	64	64	71
Mouse - Human	N.D.	65	67	71

Pair-wise alignment of predicted full-length proteins in the albumin family from different species was performed. The % of amino acid residues that were identical in each is shown. N.D.: not determined.