J Mol Evol (1998) 46:552–561

JOURNAL OF MOLECULAR EVOLUTION

The Molecular Clock Runs at Different Rates Among Closely Related Members of a Gene Family

Peter E.M. Gibbs,[1,2] Werner F. Witke,[1] Achilles Dugaiczyk[1]

[1] Department of Biochemistry, University of California, Riverside, CA 92521, USA
[2] Department of Biochemistry and Biophysics, University of Rochester, Rochester, NY 14642, USA

Received: 28 February 1995 / Accepted: 6 October 1997

Abstract. The serum albumin gene family is composed of four members that have arisen by a series of duplications from a common ancestor. From sequence differences between members of the gene family, we infer that a gene duplication some 580 Myr ago gave rise to the vitamin D–binding protein (DBP) gene and a second lineage, which reduplicated about 295 Myr ago to give the albumin (ALB) gene and a common precursor to α-fetoprotein (AFP) and α-albumin (ALF). This precursor itself duplicated about 250 Myr ago, giving rise to the youngest family members, AFP and ALF. It should be possible to correlate these dates with the phylogenetic distribution of members of the gene family among different species. All four genes are found in mammals, but AFP and ALF are not found in amphibia, which diverged from reptiles about 360 Myr ago, before the divergence of the AFP-ALF progenitor from albumin.

Although individual family members display an approximate clock-like evolution, there are significant deviations—the rates of divergence for AFP differ by a factor of 7, the rates for ALB differ by a factor of 2.1. Since the progenitor of this gene family itself arose by triplication of a smaller gene, the rates of evolution of individual domains were also calculated and were shown to vary within and between family members. The great variation in the rates of the molecular clock raises questions concerning whether it can be used to infer evolutionary time from contemporary sequence differences.

Key words: Albumin/α-fetoprotein/α-albumin/vitamin D–binding protein — Gene duplication — Sequence divergence — Evolutionary tree

Introduction

Mutations are introduced during replication and/or repair of DNA, although not all of them accumulate in the population; some are eliminated because they are deleterious to the organism, and others are lost through stochastic processes. DNA mutations accumulate at different rates in different species (Britten 1986) and at different rates for nuclear and mitochondrial genes within a species (Vawter and Brown 1986). Some of these mutations are silent at the protein level, whereas others replace the encoded amino acid and hence the sequence of the protein. Of the latter, only a fraction alter the phenotype of the species.

Early comparison of protein sequence data suggested that evolutionary changes in a particular protein accumulated at a seemingly constant rate, leading to the concept of a molecular clock (Zuckerkandl and Pauling 1965; Dickerson 1971). Such a clock must perforce be driven by mutation, but there must be processes which act at the phenotypic level and thereby regulate the clock speed. It is of great interest to distinguish and understand the mechanisms which affect the evolution of proteins, since these processes operate at the molecular basis of evolution, of speciation, and conceivably might even affect extinction phenomena.

Correspondence to: A. Dugaiczyk

We have previously reported on the rates of molecular evolution for the homologous albumin and AFP (Minghetti et al. 1985), although few sequences were available, and some of those were incomplete. The present work is based on complete sequences, and includes examples from two additional members of the gene family, DBP and ALF. Members of this gene family are structurally very similar (Law and Dugaiczyk 1981; Cooke and David 1985; Lichenstein et al. 1994) and they are closely linked; in humans, on chromosome 4q (Nishio et al. 1996). The four genes are specifically expressed in the liver; AFP and some ALB are also produced in the yolk sac. The expression of the four genes is developmentally regulated. The production of ALB starts in the fetal liver and is maintained in the adult at high levels. The expression of the AFP gene starts also in the fetal liver but is turned off after birth. DBP production also starts in the fetal liver and continues in the adult, although at a much lower level than albumin. The expression of the ALF gene starts after birth, and the protein continues to be produced in the adult (Belanger et al. 1994). There is abundant evidence that these proteins arose from a common ancestor (Brown 1976; Gibbs and Dugaiczyk 1987), although in the course of evolution they have acquired different functions. It may well be possible to draw biological inferences about this gene family from the course of their evolution and from the rates of their molecular clocks.

Methods

Sequences. The sequences used in this study were obtained from a search of the nonredundant protein database maintained at the NCBI and were obtained by searches using the program BLASTP with the blast network server (Altschul et al. 1990), selecting all sequences related to human serum albumin. The sequences (some of which exist as multiple database entries) used are given in Table 1 and are identified by accession number. In addition, three probably allelic sequences for human DBP were used, and the present results for comparisons are based on the average of the three values. The serum albumin for the lamprey *Petromyzon marinus* was not included in our analysis. This protein actually contains seven domains (Gray and Doolittle 1992), and has clearly undergone a more extensive evolutionary expansion than the three-domain proteins from the other species.

Alignments. All sequences were initially aligned with the program MULTALIN (Corpet 1988), and the alignments so obtained were used as the basis for subsequent analysis. Identification of conservative and nonconservative amino acid replacements was determined for all pairwise alignments, which were obtained using either the FASTA algorithm (Pearson and Lipman 1988) for relatively closely related proteins (all AFPs, DBPs, and mammalian and avian albumin comparisons) or the program GAP, using the end-weighted option, in the University of Wisconsin Genetics Group package (Devereux et al. 1984) for comparisons of more distantly related taxa. From these alignments we calculated the numbers of identities and the numbers of conservative substitutions by the criteria described by Dayhoff et al. (1978). The calculations were made essentially as described by Minghetti et al. (1985), except that for convenience they were performed by computer rather than the graphic method described earlier. The computer method

Table 1. Sources of sequence data[a]

Sequence	Accession number	Reference
Human albumin	M12523	Minghetti et al. (1986)
Rhesus albumin	M90463	Watkins et al. (1993)
Bovine albumin	M73993	Holowachuk and Stoltenborg (1991), unpublished
Sheep albumin	X17055	Brown et al. (1989)
Pig albumin	M36787	Weinstock and Baldwin (1988)
Horse albumin	X74045	Ho et al. (1993)
Rat albumin	V01222	Sargent et al. (1981)
Mouse albumin	M16111	Minghetti et al. (1985); Gibbs and Dugaiczyk (1994)[b]
Cat albumin	X84842	Hilger et al. (1996)
Rabbit albumin	U18344	Sheffield et al. (1995), unpublished
Chicken albumin	X60688	Cassady et al. (1991), unpublished
Cobra albumin	X78598	Havsteen et al. (1994), unpublished
Xenopus albumin [68 kD]	M18350	Moskaitis et al. (1989)
Xenopus albumin [74 kD]	M21442	Moskaitis et al. (1989)
Bullfrog albumin	M38195	Averyhart-Fullard and Jaffe (1990)
Salmon albumin I	X52397	Byrnes and Gannon (1990)
Salmon albumin II	X60776	Byrnes and Gannon (1992)
Human AFP	M16110	Gibbs et al. (1987)
Gorilla AFP	M38272	Ryan et al. (1991)
Chimpanzee AFP	U21916	Nishio et al. (1995)
Rat AFP	X02361;V01254	Jagodzinski et al. (1981); Turcotte et al. (1985)
Mouse AFP	V00743	Law and Dugaiczyk (1981)
Horse AFP	U28947	McDowell et al. (1975), unpublished
Human ALF (afamin)	L32140	Lichenstein et al. (1994)
Rat ALF	X76456	Bélanger et al. (1994)
Human DBP [Gc1]	M12654	Cooke and David (1985)
Human DBP [Gc2]	M11321	Yang et al. (1985)
Human DBP [genomic]	L10641	Witke et al. (1993)
Rat DBP	M12450	Cooke (1986)
Mouse DBP	M55413	Yang et al. (1990)
Rabbit DBP	D29666	Osawa et al. (1994)

[a] Sequences used in this study were obtained by searching the NCBI databases for sequences related to human serum albumin, using the BLAST programs (Altschul et al. 1990)

[b] The bulk of the sequence of mouse albumin was from Minghetti et al. (1985) and additional sequence assembled as described in Gibbs and Dugaiczyk (1994). Two remaining gaps were filled with consensus sequences derived from mouse albumin sequences in the database of expressed sequence tags

yielded the same results as earlier for those comparisons which were repeated in this study.

Distance Calculations. The observed degrees of sequence similarity between pairs of proteins obtained from the sequence alignments were used to estimate quantitatively the actual rates of amino acid replacement by correcting for multiple amino acid changes at each site. We used two approaches to this problem. Back mutations, arising as a consequence of multiple events at the same site, are incorporated in the methods used to calculate genetic distance.

First, an estimate was made assuming that replacements follow a strictly Poisson distribution. If the proportion of nonidentical amino acids in the comparison is p, the number of changes per amino acid site is given by:

$$d_p = -\ln (1 - p)$$

This correction method was used in our earlier study of the evolution of this family (Minghetti et al. 1985). However, it has been demonstrated that where p exceeds 0.2, the estimate made by this method does not correspond well to the genetic distance (Dayhoff et al. 1978). A more accurate method of comparison calculates the distance by assuming that the sequence differences are related by a γ function (Nei et al. 1976). These calculations were made using the program MEGA (Kumar et al. 1993), using the default options for the program.

More recently, Grishin (1995) further refined the method for estimating the number of amino acid replacements when the rate varies among sites, but the results of his methods are practically superimposable with the γ distance (see Fig. 1 of Grishin 1995). Our present results are based on measurements of the γ distance.

Divergence Dates. The dates of divergence of the tetrapod groups were taken from Benton (1990), those for the artiodactyl species from Romero-Herrera et al. (1973), for the primates from Sibley and Ahlquist (1984), for the mouse/rat separation from Britten (1986), and for the duplication of the *Xenopus laevis* genome from Bisbee et al. (1977). Thus, human/chimp separation was taken at 5 Myr; human/gorilla, and chimp/gorilla at 8 Myr; cow/sheep at 18 Myr; rat/mouse at 25 Myr; human/rhesus at 28 Myr; pig/cow and pig/sheep at 55 Myr; duplication of the *Xenopus* genome at 30 Myr; and the remaining mammals at 80 Myr.

Results

Rates of Divergence

A genetic distance requires an estimate of the number of instances where multiple changes have occurred. For closely related sequences, this can be estimated by assuming that sequence changes follow a Poisson distribution. However, where the proportion of different amino acids (p) exceeds 0.2, the Poisson estimate is no longer valid, as the distance is underestimated. For distantly related sequences, more accurate estimates of the genetic distance may be made assuming that substitutions at different sites follow a γ function (Nei et al. 1976). Examples of genetic distances calculated by the two methods are given in Table 2. Our present studies are based on measurements of the γ distance.

The inclusion of more recent sequence information for DBP and ALF extends our earlier findings (Minghetti et al. 1985) to two other members of the gene family. In addition, the availability of new primate sequences, for macaque albumin and for gorilla and chimpanzee AFPs, enables us to expand the data set for the rodent–primate comparisons; the apparent extents of conservation between the orders using these sequences are essentially identical to those obtained using human sequences, suggesting that our earlier data were not a consequence of a statistical quirk resulting from the use of only human sequence on the primate branch. The present genetic distance measures are shown in Table 3.

Table 2. Comparison of genetic distance measures

Human albumin vs	Poisson distance ± S.E.	Gamma distance ± S.E.
Rhesus albumin	6.72 ± 1.08	6.84 ± 1.11
Bovine albumin	26.65 ± 2.24	28.51 ± 2.56
Rat albumin	30.99 ± 2.44	33.51 ± 2.85
Chicken albumin	73.51 ± 4.22	88.83 ± 6.10
Xenopus 74-kD albumin	93.13 ± 5.05	118.61 ± 8.04
Cobra albumin	112.19 ± 5.84	150.47 ± 10.23
Salmon albumin I	128.83 ± 6.61	180.87 ± 12.58

Genetic distances were calculated as described in Methods and are expressed as % changes/site

Figure 1 shows the amino acid replacements for albumin, AFP, ALF, and DBP, excluding the signal sequence, as a function of divergence time. It is apparent that the rates are not constant over time. The slopes of the albumin and DBP are similar (0.404 changes/site/100 Myr vs 0.359), although they do differ, while that for AFP is greater (0.504) and for ALF higher still (0.542). Since the dates of divergence actually estimate only half the divergence time, the slopes of the lines in Fig. 1 are in fact twice as large as the true number of changes. Thus albumin accumulates 20.2% amino acid changes per 100 Myr, DBP 18.0%, ALF 27.1%, and AFP 25.2%. This is in general agreement with earlier estimates which have consistently shown that AFP evolves more rapidly than albumin (Minghetti et al. 1985; Nardelli-Haefliger et al. 1989). However, the calculated rates of divergence seen here are somewhat different than those previously published, reflecting the different methods used to calculate distances, a larger set of sequences, and possibly different dates used to calibrate the clock.

Evolutionary History of the Gene Family

Using the genetic distance measures from Table 5 and the rates of change (slopes) from the data in Fig. 1, we calculated the approximate divergence dates for members of the gene family. Thus for the AFP–ALF divergence: Given that the average distance is 1.306 (Table 5) and the rates for AFP and ALF are 0.504 and 0.542 changes/site/100 Myr, respectively (Fig. 1), then, using the average of the rates, the date estimate for this pair is

Table 3. Genetic distance measures for mammalian family members[a]

A. Albumins

	Human	Rhesus	Rat	Mouse	Rabbit	Cat	Cow	Sheep	Pig	Horse
Human		6.87	33.50	36.00	32.14	20.92	29.56	30.87	30.35	28.78
Rhesus	1.11		31.60	34.32	31.34	21.85	28.53	30.08	29.56	27.00
Rat	2.92	2.80		11.59	36.00	30.81	38.37	39.25	34.67	34.39
Mouse	3.06	2.96	1.51		37.13	34.32	40.14	40.74	38.66	37.22
Rabbit	2.84	2.79	3.06	3.13		30.81	36.64	35.23	33.29	37.79
Cat	2.15	2.21	2.76	2.96	2.76		26.25	27.76	24.53	26.50
Cow	2.68	2.62	3.21	3.31	3.10	2.48		8.42	24.77	32.75
Sheep	2.76	2.71	3.26	3.35	3.02	2.57	1.27		26.75	30.35
Pig	2.73	2.68	2.99	3.22	2.91	2.38	2.39	2.51		29.56
Horse	2.64	2.53	2.97	3.14	3.17	2.50	2.87	2.73	2.68	

B. AFPs

	Human	Chimp	Gorilla	Mouse	Rat	Horse
Human		1.04	0.69	45.59	46.55	23.46
Chimp	0.42		1.39	45.59	46.55	23.46
Gorilla	0.35	0.49		45.59	46.88	23.70
Mouse	3.64	3.64	3.64		19.43	46.23
Rat	3.69	3.69	3.71	2.06		48.52
Horse	2.31	2.31	2.32	3.67	3.81	

C. DBPs

	Human	Mouse	Rat	Rabbit
Human		29.06	28.07	21.37
Mouse	3.01		10.48	33.27
Rat	2.94	1.62		31.54
Rabbit	2.47	3.29	3.17	

D. ALF

Human *vs.* Rat	43.36 ± 3.52

[a] Genetic distances between mature protein sequences (i.e., lacking the leader peptides) were calculated using the γ function of Nei et al. (1976) and are expressed as % changes/site. Genetic distances are shown above the diagonal and the standard errors below. The data are graphed in Fig. 1A and B, except for the rhesus albumin comparisons, which are very similar to the human data

$$130.6/0.523 = 249.7 \text{ Myr}$$

Prior to that date, ALB and an AFP–ALF progenitor diverged. To calculate that divergence date, we first calculated the distance between all ALB/AFP pairs and all ALB/ALF pairs at 250 Myr ago using an average divergence rate as above. Today's distances for ALB–ALF and ALB–AFP are 1.506 and 1.292, respectively, and the ALB slope is 0.404. Therefore, 250 Myr ago, the ALB–ALF distance was

$$150.6 - (249.7 \times 0.473) = 32.5$$

and the ALB–AFP distance was

$$129.2 - (249.7 \times 0.454) = 15.8$$

Since there are 20 ALB–ALF distances and 60 ALB–AFP distances, this gives a weighted average distance at 250 Myr before present of 20.0. Using an average rate of $(0.404 + 0.523)/2 = 0.464$, this distance translates to $20/0.464 = 43.1$ Myr. Thus the date for albumin divergence from the ALB/ALF progenitor is

$$249.7 + 43.1 = 292.8 \text{ Myr ago}$$

In a similar manner we calculated the distances at 250 Myr ago, using the average of the slopes from Fig. 1:

$$\text{ALB–DBP: } 224.2 - 249.7 \times (0.404 + 0.359)/2 = 224.2 - 95.3 = 128.9$$
$$\text{AFP–DBP: } 258.5 - 249.7 \times (0.504 + 0.359)/2 = 258.5 - 107.7 = 150.8$$
$$\text{ALF–DBP: } 254.9 - 249.7 \times (0.542 + 0.359)/2 = 254.9 - 112.5 = 142.4$$

Similarly, the divergence of DBP from the other lineage was calculated to have occurred 291.9 Myr before the previously calculated split, i.e., 249.7 + 43.1 + 291.9 = 584.7 Myr ago. The results are graphically depicted in Fig. 2.

Our albumin AFP divergence date at 293 Myr before present is similar to the 280 Myr calculated by Nardelli-

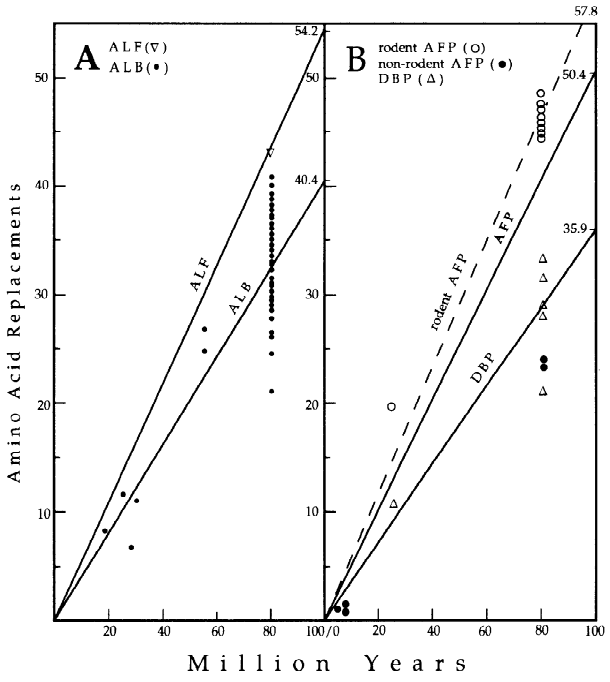

Fig. 1. Amino acid replacements per 100 sites for albumin and ALF **(A)**, and for AFP and DBP **(B)** as a function of divergence time between species. The genetic distances for the mature protein (i.e., after removal of the leader peptide sequences) are taken from Table 3. An additional point at 30 Myr is that for the duplicated *Xenopus* albumins (Bisbee et al., 1977; Moskaitis et al., 1989); the distance was 11.04%. Divergence dates of species within a single mammalian order are as described in the text. Species in different orders are considered as diverging since the mammalian radiation, 80 Myr ago. The *lines* were fitted by linear regression, although it is apparent that a single straight line cannot provide a reasonable fit for all points (except for ALF, which is based on one experimental point). The AFP line is based on all the data points, including rodent (○) and nonrodent (●) comparisons. The alternative line for AFP (''rodent AFP'') is based only on comparisons involving at least one rodent species, using the data points indicated by open circles (○).

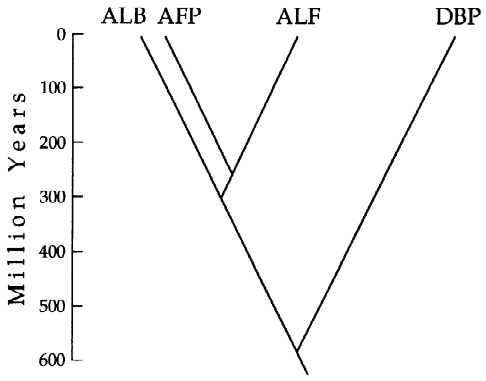

Fig. 2. Simplified phylogeny of members of the albumin gene family determined from the calculated divergence dates. The approximate dates of separation were calculated from the slopes of the regression lines in Fig. 1 and they are indicated by the *scale at left.* Even though our data are a poor fit to a clock, it was possible to use them to infer a phylogenetic tree for the gene family. The topology of the tree is not in dispute, as it is supported by a bootstrap analysis to a high degree of confidence. However, the reliability of the divergence dates is a different question.

Haefliger et al. (1989) and somewhat less than the estimate of Gray and Doolittle (1992). It should be noted that the estimated date of divergence of the reptile–avian line from mammals is approximately 290 Myr and that Lindgren et al. (1974) have detected an AFP-like protein in chicken. AFP and ALF show a mean distance of 1.306 and a divergence date of 250 Myr; since this is the most recent date, it would be expected that the most recent duplication in the gene family gave rise to AFP and ALF. This conclusion has been supported by phylogenetic reconstruction using the neighbor joining method (Saitou and Nei 1987), where the AFPs and ALFs segregate as a monophyletic clade in a tree reconstructed from all of the sequences used, and supported by a bootstrap analysis of this phylogeny, where the ALFs and AFPs were found on the same lineage in 100% of the replicates. Our calculated date for the albumin DBP divergence at 585 Myr ago is in contrast to the estimated date of the lamprey–bony fish separation date of 450 Myr and the observation that lampreys lack a vitamin D–binding protein (Hay and Watson 1976). The calculated value does agree well with the estimate of 560–600 Myr of Nardelli-Haefliger et al. (1989). Several possibilities present themselves to explain these apparent discrepancies, although it is most probable that the clock calibrations are inaccurate. The most likely sources of such inaccuracy lie in the divergence dates, but also in the possibility that the genes do not show true clocklike behavior. Figure 2 shows a simplified phylogeny for the four members of the family.

How Reliable Is the Clock?

The rate estimates and calculated divergence dates rest upon the assumption that there is a true molecular clock for each gene, an assumption which is open to question. Indeed, data generated in the course of this study indicate that certain comparisons do not fit the molecular clock assumptions. For serum albumins (Fig. 1A), the rates of divergence differ by a factor of 2.0, the data being scattered from a high 40.7% (mouse/sheep) to a low 20.9% (human/cat); the human/rhesus point is even lower. For AFP (Fig. 1B), the high point of 48.52% (horse/rat) differs from the low point of 23.46% (human/horse) by a factor of 2.1, and from the human/gorilla divergence by a factor of 7.0. The DBP data are scattered within a smaller range, differing by a factor of 1.6, from a low 21.4% for human/rabbit to a high 33.3% for mouse/rabbit, but this could be due to the fact that fewer species have been compared. The ALF line has only one experimental point, reflecting the human/rat genetic distance. The distances for all known mammalian sequences are shown in Table 3. It can be seen that those comparisons which include a rodent sequence yield consistently higher distances than those which do not. Therefore, in order to make a meaningful comparison in evolutionary

rate between ALF and AFP, we drew an alternative line for AFP, using only the experimental points that include rodent/rodent and rodent/primate sequences, obtaining a higher (57.8% change/site/100 Myr) rate for AFP. Although not included in Fig. 1, other variations in rate can be noted. The reptiles and birds are a monophyletic clade, yet their albumin sequences yield significantly different values in comparisons with other species; e.g., comparing chicken and human albumin sequences yields a γ distance of 88.8 ± 6.1, whereas a cobra/human comparison gives a distance 150.5 ± 10.2 (Table 2). These distances are clearly significantly different and indicate that the rates of evolutionary change in certain lines of descent can show dramatic variation.

Local Clock Differences Within One Gene

Since the evolutionary rate of change was found to differ significantly, both between members of the family in the same species and for the same protein across different species, we wanted to see whether a ''local'' clock could be found that would run at a more constant rate than that for the entire protein. The individual domains of the proteins were a logical choice for such a focused view. Each protein is composed of three almost identical domains, which have been recognized by Brown (1976), based on their structural similarity. A total of 18 symmetrically placed disulfide bonds constrain the polypeptide chain, folding it into a characteristic serpentine structure. This two-dimensional, serpentine structure has been recognized for each member of the human gene family (Brown 1976; Cooke and David 1985; Minghetti et al. 1986; Gibbs et al. 1987; Lichenstein et al. 1994), and also for the chimpanzee and gorilla AFP (Nishio et al. 1995; Ryan et al. 1991), for mouse AFP (Law and Dugaiczyk 1981), rat ALB and AFP (Jagodzinski et al. 1981), and for *Xenopus* ALB (Nardelli-Haefliger et al. 1989). Despite the accumulated structural data, it would be premature, however, to say something about the role of the individual domains in the overall function of the protein, since the functions of ALB and DBP are not very precisely defined, in molecular terms, and those of AFP and ALF are actually unknown. In fact, it was our hope to make biological inferences about their relative function from the rate of their evolutionary change.

A comparison of the divergence rates for individual domains of the four proteins is given in Table 4. It can be seen that rates for individual domains actually differ more than those for the whole protein. Rates of evolution do vary significantly for domain II with the rates for AFP > DBP ≈ ALF > ALB. The rates of divergence for the albumin domains II and III are very similar. The rates for domain I vary in the order AFP > ALF > ALB > DBP. The rate for AFP domain I is clearly the greatest rate for any domain in the gene family, and that for DBP domain I is the smallest. Since the three domains within these proteins are of almost equal length, it is apparent that individual domains do make a substantial contribution to the global rates of evolution of individual family members.

Table 4. Divergence rates for mature proteins and their individual domains[a]

	ALB	AFP	ALF	DBP
Mature protein	40.4	50.4	54.2	35.9
Domain I	47.2	64.6	53.5	25.1
Domain II	37.4	51.9	42.5	44.9
Domain III	37.3	37.2	68.8	45.5

[a] Genetic distances for orthologous domains were calculated as described in Table 3. Rates of amino acid replacement of individual domains were also determined by linear regression and are expressed as % change/site/100 Myr

Table 5. Genetic distance measures for mammalian family members[a]

Proteins compared	Number of comparisons	Mean distance
AFP–ALF	12	1.306
ALB–AFP	60	1.292
ALB–ALF	20	1.506
ALB–DBP	40	2.242
AFP–DBP	24	2.585
ALF–DBP	8	2.549

[a] Genetic distances were calculated using the γ function and are expressed as substitutions/site for the mature proteins from various species

Population Genetics and Gene Evolution

One approach which might be advantageous in estimating the potential for divergence could be examination of the diversity of these proteins within species, since a pool of sequence variants is a prerequisite for evolution of the protein. In this respect, ALF is at present uninformative, since its discovery is recent and no population studies have yet been initiated. Despite its clinical utility, no population studies have been made for AFP either, limiting its potential use in this regard. In contrast, albumin and DBP have been the subject of some of the most extensive human genetic studies. Each has numerous variants described (e.g., Constans et al. 1983; Madison et al. 1991). However, with the exception of the two alleles of DBP (Constans et al. 1983), these variants occur at a very low frequency.

Discussion

According to the concept of the molecular clock, DNA or protein sequence differences between extant species would measure the time elapsed since they diverged from a common ancestor. We have studied the serum

albumin family molecular clocks using a variety of approaches, some of which are possible because of unique features of this family. First, there are several homologous proteins in the family in a variety of species. Second, since each of the proteins was derived from an ancient triplication event, the structure of the proteins lends itself to analyses within a single line of evolutionary descent. Based on such an analysis of individual proteins, we have shown elsewhere (Gibbs and Dugaiczyk 1994) that some proteins (ceruloplasmin) evolve at the same rate in different lineages (human and rat), while members of the albumin gene family evolve at distinctly different rates in the same two lineages. In the present study, we have demonstrated that the individual proteins of the family, and the domains therein, each evolve at different rates.

Considering that members of this gene family are closely linked in the human genome, conceivably in as little as 200 kb of DNA (Nishio et al. 1996), and since the error rates of DNA replication in such a small chromosomal region are likely to be invariant, it is reasonable to assume that the rate of mutation in this family is likely to be constant, or nearly so. Therefore, the different rates of evolution of the proteins encoded by this gene family probably reflect the different biological constraints imposed on the individual proteins. Of the four proteins discussed herein, only DBP has a well-characterized function, that of transport of vitamin D in the circulation. The relatively slow rate of change in the first domain of DBP may reflect this function. Although several different functions have been attributed to albumin, its absence in the adult is not deleterious to survival (Gitlin and Gitlin 1975). No function has as yet been attributed to AFP, which is expressed only in fetal life, nor to ALF. The selective pressures exerted on these gene products are therefore unknown.

It is apparent from the work described herein and our earlier studies (Gibbs and Dugaiczyk 1994) on the evolution of members of this gene family that not only do rates of evolution vary for individual proteins but that rates of divergence for the same protein can differ across mammalian lineages. When all the available sequence data were used, the replacement rate for AFP was found to be 5.0×10^{-12} site/year; for ALF, 5.4×10^{-12} site/year; for ALB, 4.0×10^{-12}/site/year; and for DBP, 3.6×10^{-12}/site/year. Thus there is a 1.5-fold variation in rate between ALF and DBP. Moreover, visual inspection of Fig. 1, particularly the values at 80 Myr, indicates that there are clear differences in rates for a single protein. For example, for ALB, the maximum divergence at 80 Myr is 40.7%, or 5.1×10^{-12}/site/year (mouse/sheep), while the minimum is 20.9% (human/cat); the rate is even lower (2.45×10^{-12}/site/year) for the human/rhesus divergence. For AFP, the maximum divergence at 80 Myr is 48.5%, or 6.1×10^{-12}/site/year (horse/rat), while the minimum is 23.5% (horse/human); the human/gorilla replacement rate is still lower (8.63×10^{-13}/site/year). The rates for DBP vary by a factor of up to a 1.6, although fewer species were compared. These variations in rate raise questions as to the general usefulness of the molecular clock to determine elapsed evolutionary time of diverging species. In addition, similar variations are found within nonfunctional sequences. Thus rates for pseudogenes have been reported as high as 12.6×10^{-9}/site/year for an α-globin pseudogene (Miyata and Yasunaga 1981) and as low as 10^{-9}/site per year for β- and η-globin pseudogenes (Chang and Slightom 1984; Goodman et al. 1984). The age of a human enolase pseudogene has been reported as 14 Myr, based on sequence divergence (Feo et al. 1990), or 25–30 Myr, based on its phylogenetic distribution among primate species (Minghetti and Dugaiczyk 1993), a hardly satisfactory result in measuring evolutionary time in the divergence of primates.

Three biological variables are generally invoked to account for the variability in the rate of the molecular clock. (1) Species-dependent differences in the fidelity of DNA replication, (2) population size and differences in generation time between species, and (3) changes in function of a protein along a line of genetic descent. A higher substitution rate will put additional pressure on every gene in the species to mutate faster. A shorter generation time will shorten the time to fixation of new mutations; a larger population size, on the other hand, will increase the time to fixation of new mutations. The molecular clock should run faster in organisms with shorter generation time and more slowly in large populations. Indeed, comparing rat, bovine, and human sequences, Ohta (1993) reported that the rodent line is the most divergent among the three lineages for synonymous substitutions, but not necessarily for non-synonymous substitutions. Ohta also suggested that nonsynonymous substitutions (amino acid replacements) will more closely follow the clock because the generation-time effect is likely to be reduced by cancellation with the population-size effect (Ohta 1993). Higher rates in rodents than in primates have been also reported by Gu and Li (1992), and faster clocks in monkeys and apes than in humans by Li and Tanimura (1987).

Our present results on ALB, AFP, and DBP tend to support the notion of a faster rate in rodents, but there are noticeable incongruities. Although the rat/mouse AFP distance of 19.43 (Table 1) is high above the AFP line in Fig. 1, the rat/mouse ALB distance of 11.59 (Table 3) is only slightly above the ALB line in Fig. 1, and so is the rat/mouse DBP distance of 10.48 (Table 3) only slightly above the DBP line in Fig. 1. The lack of rodent/rodent acceleration in the ALB and DBP divergence is particularly noteworthy, because rodent/nonrodent distances are high above the lines in Fig. 1. These incongruities cannot be explained with a single hypothesis, such as a different generation time or a different accuracy in the DNA rep-

lication mechanism. We have previously noted that ceruloplasmin appears to evolve as fast in human as it does in rat, based on deterioration of internal symmetry of a gene within single evolutionary lineages (Gibbs and Dugaiczyk 1994). Thus, shorter generation times in rodents do not always translate into a faster evolutionary rate in the organism. Similarly, the human and rhesus ALB comparison with eight other species (Table 3) shows no indication of a faster rate in the rhesus lineage, which has a shorter generation time than human. In fact in seven out of the eight species compared, the human distance is always by about 1% higher than that of the rhesus monkey. The human, chimp, and gorilla AFP comparison with three nonprimate species gave practically identical results (Table 3), again giving no support for the notion that a longer generation time in humans results in a slower rate of divergence. On the other hand, glycerol-3-phosphate dehydrogenase evolves at several times different rates in different *Drosophila* species that propagate with the same generation time (Ayala et al. 1996). Thus the number of examples where generation times cannot be correlated with divergence rates is too large to be comfortably accommodated into a single model governing evolution.

A possible linkage between the molecular clock and functional variation of a protein is even more difficult to disentangle. In the albumin gene family, the functions of AFP and ALF are not entirely clear, but they are likely to be different from that of albumin. Should their roles be less subject to selective pressure than that of albumin, then perhaps their accelerated rate is to be expected. But then, what is the role of albumin when its absence in the adult is not a life-threatening condition (Gitlin and Gitlin 1975)? Considering the apparent dispensability of albumin, it is difficult to argue that the almost continuous variation in its evolutionary rate (see the large column of dispersed experimental points at 80 Myr in Fig. 1A) reflects a fine tuning to its functional variation along the lines of genetic descent. Why would the function of sheep albumin be so much different from that of cat albumin? They differ from each other by 50%, as compared to human (Table 3). The reduced rate of evolution of DBP might be attributable to two events—gain of functions relative to albumin (i.e., vitamin D transport, binding to actin) and loss of a significant portion of domain III of the protein. Together, these may have limited the paths by which DBP was free to evolve. The cobra albumin appears to have evolved at a much accelerated rate, and the protein has acquired an entirely new function, that of self-protection against conspecific venom. Zuckerkandl (1976) proposed the concept of a functional density, which was the fraction of a polypeptide chain involved in specific functions. Chains having a higher proportion of amino acids involved in specific functions would have a higher functional density and therefore would be more constrained against evolutionary drift. If a new function arises in a polypeptide chain, its functional density would be expected to increase. The altered rate of evolution of the snake albumin might be a consequence of altered functional density, subject to positive selection that ultimately leads to a change in function. In their covarion hypothesis, Fitch and Markowitz (1970) proposed that at any given stage in evolution, only a limited subset of the amino acids in a protein is subject to replacement, although this subset may change at a later stage in some lineages. By assuming this covarion hypothesis, Ayala et al. (1996) could account for a fivefold decrease in the rate of superoxide dismutase as being consistent with the molecular clock hypothesis, but the same covarion hypothesis could not explain a 15-fold rate difference in the glycerol phosphate dehydrogenase evolution in the different drosophilids (Ayala et al. 1996).

Even if the molecular clock runs at a constant rate but the constancy is in reference to an altering function of a protein, the clock is intractable. It is subject to the same vagaries as the rest of biology. Models are only models; they are only as good as the underlying assumptions. And if the number of assumptions (unknowns) is greater than the number of equations, a rigorous solution is but an illusion. This seems to be the case with the molecular clock.

Acknowledgments. We thank Dr. Emile Zuckerkandl for valuable comments and Lars Dugaiczyk for computer graphics of the figures. The work was supported by National Science Foundation grant SBR-9602480

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Averyhart-Fullard V, Jaffe RC (1990) Cloning and thyroid hormone regulation of albumin mRNA in *Rana catesbeiana* tadpole liver. Mol Endocrinol 4:1556–1563

Ayala FJ, Barrio E, Kwiatkowski J (1996) Molecular clock or erratic evolution? A tale of two genes. Proc Natl Acad Sci 93:11729–11734

Belanger L, Roy S, Allard D (1994) New albumin gene 3′ adjacent to the α1-feto-protein locus. J Biol Chem 269:5481–5484

Benton MJ (1990) Phylogeny of the major tetrapod groups: morphological data and divergence dates. J Mol Evol 30:409–424

Bisbee CA, Baker MA, Wilson AC, Hadzi-Azimi J, Fischberg M (1977) Albumin phylogeny for clawed frogs (*Xenopus*). Science 195:785–787

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398

Brown JR (1976) Structural origins of mammalian albumin. Fed Proc 35:2141–2144

Brown WM, Dziegielewska KM, Foreman RC, Saunders NR (1989) Nucleotide and deduced amino acid sequence of sheep serum albumin. Nucleic Acids Res 17:10495–10495

Byrnes L, Gannon F (1990) Atlantic salmon (*Salmo salar*) serum albumin: cDNA sequence, evolution, and tissue expression. DNA Cell Biol 9:647–655

Byrnes L, Gannon F (1992) Sequence analysis of a second cDNA for Atlantic salmon (*Salmo salar*) serum albumin. Gene 120:319–320

Chang LYE, Slightom JL (1984) Isolation and nucleotide sequence analysis of the β-type globin pseudogene from human, gorilla and chimpanzee. J Mol Biol 180:767–784

Constans J, Cleve H, Dykes D, Fischer M, Kirk RL, Papiha SS, Scheffran W, Scherz R, Thyman M, Weber W (1983) The polymorphism of the vitamin D-binding protein (Gc); isoelectric focusing in 3 M urea as additional method for identification of genetic variants. Hum Genet 65:176–180

Cooke NE (1986) Rat vitamin D binding protein: determination of the full-length primary structure from cloned cDNA. J Biol Chem 261: 3441–3450

Cooke NE, David EV (1985) Serum vitamin D-binding protein is a third member of the albumin and alpha fetoprotein gene family. J Clin Invest 76:2420–2424

Corpet F (1988) Multiple sequence alignments with hierarchical clustering. Nucleic Acids Res 16:10881–10890

Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 3. The National Biomedical Research Foundation, Silver Spring, MD, pp 345–352

Devereux J, Haerberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Dickerson RE (1971) The structure of cytochrome c and the rates of molecular evolution. J Mol Evol 1:26–45

Feo S, Oliva D, Arico B, Barba G, Cali L, Giallongo A (1990) The human genome contains a single processed pseudogene for α-enolase located on chromosome 1. DNA Sequence-J. DNA sequencing and mapping, vol 1. Harwood Academic Publishers GmbH, United Kingdom, pp 79–83

Fitch WM, Markowitz E (1970) An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem Genet 4:579–593

Gibbs PEM, Dugaiczyk A (1987) Origin of structural domains of the serum-albumin gene family and a predicted structure for vitamin D-binding protein. Mol Biol Evol 4:364–379

Gibbs PEM, Dugaiczyk A (1994) Reading the molecular clock from the decay of internal symmetry of a gene. Proc Natl Acad Sci USA 91:3413–3417

Gibbs PEM, Zielinski R, Boyd C, Dugaiczyk A (1987) Structure, polymorphism, and novel repeated DNA elements revealed by a complete sequence of the human α-fetoprotein gene. Biochemistry 26: 1332–1343

Gitlin D, Gitlin JD (1975) Genetic alterations in the plasma proteins of man. In: Putnam FW (ed) The plasma proteins, vol 2. Academic Press, New York, pp 321–374

Goodman M, Koop BF, Czelusniak J, Weiss ML, Slightom JL (1984) The η-globin gene: Its long evolutionary history in the β-globin gene family of mammals. J Mol Biol 180:803–823

Gray JE, Doolittle RF (1992) Characterization, primary structure and evolution of lamprey plasma albumin. Protein Sci 1:289–302

Grishin NV (1995) Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. J Mol Evol 41:675–679

Gu X, Li W-H (1992) Higher rates of amino acid substitution in rodents than in humans. Mol Phylogenet Evol 1:211–214

Hay AWM, Watson G (1976) The plasma transport proteins of 25-hydroxy-cholecalciferol in fish, amphibians, reptiles and birds. Comput Biochem Physiol 53B:167–172

Hilger C, Grigioni F, Hentges F (1996) Sequence of the gene encoding cat (*Felis domesticus*) serum albumin. Gene 169:295–296

Ho JX, Holowachuk EW, Norton EJ, Twigg PD, Carter DC (1993) X-ray and primary structure of horse serum albumin (*Equus caballus*) at 0.27 nm resolution. Eur J Biochem 215:205–212

Jagodzinski LL, Sargent TD, Yang M, Glackin C, Bonner J (1981) Sequence homology between RNAs encoding rat α-fetoprotein and rat serum albumin. Proc Natl Acad Sci USA 78:3521–3525

Kumar S, Tamura K, Nei M (1993) MEGA: molecular evolutionary genetics analysis, version 1.0. The Pennsylvania State University, University Park, PA

Law SW, Dugaiczyk A (1981) Homology between the primary structure of α-fetoprotein, deduced from a complete cDNA sequence, and serum albumin. Nature 291:201–205

Li W-H, Tanimura M (1987) The molecular clock runs more slowly in man than in apes and monkeys. Nature 326:93–96

Lichenstein HS, Lyons DE, Wurfel MW, Johnson DA, McGinley MD, Leidli JC, Trollinger DB, Mayer JP, Wright SD, Zukowski MM (1994) Afamin is a new member of the albumin, α-fetoprotein, and vitamin D-binding protein gene family. J Biol Chem 269:18149–18154

Lindgren J, Vaheri A, Ruoslahti E (1974) Identification and isolation of a foetoprotein in the chicken. Differentiation 2:233–236

Madison J, Arai K, Sakamoto Y, Feld RD, Kyle RA, Watkins S, Davis E, Matsuda Y, Amaki I, Putnam FW (1991) Genetic variants of serum albumin in Americans and Japanese. Proc Natl Acad Sci USA 88:9853–9857

Minghetti PP, Law SW, Dugaiczyk A (1985) The rate of molecular evolution of α-fetoprotein approaches that of pseudogenes. Mol Biol Evol 2:347–358

Minghetti PP, Ruffner DE, Kuang W-J, Dennison OE, Hawkins JW, Beattie WG, Dugaiczyk A (1986) Molecular structure of the human albumin gene is revealed by nucleotide sequence within q11-22 of chromosome 4. J Biol Chem 261:6747–6757

Minghetti PP, Dugaiczyk A (1993) The emergence of new DNA repeats and the divergence of primates. Proc Natl Acad Sci USA 90:1872–1876

Miyata T, Yasunaga T (1981) Rapidly evolving mouse α-globin-related pseudogene and its evolutionary history. Proc Natl Acad Sci USA 78:450–453

Moskaitis JE, Sargent TD, Smith LH Jr, Pastori RL, Schoenberg DR (1989) *Xenopus laevis* serum albumin: sequence of the cDNAs encoding the 68- and 74-kilodalton peptides and the regulation of albumin gene expression by thyroid hormone during development. Mol Endocrinol 3:464–473

Nardelli-Haefliger D, Moskaitis JE, Schoenberg DR, Wahli W (1989) Amphibian albumins as members of the albumin, α-fetoprotein, vitamin D-binding protein multigene family. J Mol Evol 29:344–354

Nei M, Chakraborty R, Fuerst PA (1976) Infinite allele model with varying mutation rate. Proc Natl Acad Sci USA 73:4164–4168

Nishio H, Gibbs PEM, Minghetti PP, Zieliski R, Dugaiczyk A (1995) The chimpanzee α-fetoprotein-encoding gene shows structural similarity to that of gorilla but distinct differences from that of human. Gene 162:213–220

Nishio H, Heiskanen M, Palotie A, Belanger L, Dugaiczyk A (1996) Tandem arrangement of the human serum albumin multigene family in the sub-centromeric region of 4q: evolution and chromosomal direction of transcription. J Mol Biol 259:113–119

Ohta T (1993) An examination of the generation time effect on molecular evolution. Proc Natl Acad Sci USA 90:10676–10680

Osawa M, Tsuji T, Yukawa N, Saito T, Takeichi S (1994) Cloning and sequence analysis of cDNA encoding rabbit vitamin D-binding protein (GC globulin). Biochem Mol Biol Int 34:1003–1009

Pearson WJ, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448

Romero-Herrera AE, Lehmann H, Joysey KA, Friday AE (1973) Molecular evolution of myoglobin and the fossil record: a phylogenetic synthesis. Nature 246:389–395

Ryan SC, Zielinski R, Dugaiczyk A (1991) Structure of the gorilla α-fetoprotein gene and the divergence of primates. Genomics 9: 60–72

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sargent TD, Yang M, Bonner JJ (1981) Nucleotide sequence of cloned rat serum albumin messenger RNA. Proc Natl Acad Sci USA 78: 243–246

Sibley CE, Ahlquist JE (1984) The phylogeny of the hominid primates as indicated by DNA hybridization. J Mol Evol 20:2–15

Turcotte B, Guertin M, Chevrette M, Bèlanger L (1985) Rat α-1-fetoprotein messenger RNA: 5′-end sequence and glucocorticoid-suppressed liver transcription in an improved nuclear run-off assay. Nucleic Acids Res 13:2387–2398

Vawter L, Brown WM (1986) Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science 234:194–196

Watkins S, Sakamoto Y, Madison J, Davis E, Smith DG, Dwuley J, Putnam FW (1993) cDNA and protein sequence of polymorphic macaque albumins that vary in bilirubin binding. Proc Natl Acad Sci USA 90:2409–2413

Weinstock J, Baldwin GS (1988) Nucleotide sequence of porcine liver albumin. Nucleic Acids Res 16:9045

Witke WF, Gibbs PEM, Zielinski R, Yang F, Bowman BH, Dugaiczyk A (1993) Complete structure of the human Gc gene: differences and similarities between members of the albumin gene family. Genomics 16:751–754

Yang F, Brune JL, Naylor SL, Cupples RL, Naberhaus KH, Bowman BH (1985) Human group-specific component (Gc) is a member of the albumin family. Proc Natl Acad Sci USA 82:7994–7998

Yang F, Bergeron JM, Linehan LA, Lalley PA, Sakaguchi AY, Bowman BH (1990) Mapping and conservation of the group-specific component gene in mouse. Genomics 7:509–516

Zuckerkandl E (1976) Evolutionary process and evolutionary noise at the molecular level. I. Functional density in proteins. J Mol Evol 7:167–183

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, pp 97–166