The Rate of Molecular Evolution of α-Fetoprotein Approaches That of Pseudogenes

Phillip P. Minghetti,* Simon W. Law,† and Achilles Dugaiczyk*

*Department of Biochemistry, University of California, Riverside; and †National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland

We conducted the present study in an attempt to correlate function with the rate of molecular evolution for serum albumin and α-fetoprotein. We found a high rate of silent substitution (between $5 \times 10^{-9}$ and $7 \times 10^{-9}$/site/year) for both the albumin and α-fetoprotein genes, perhaps the highest so far reported for an expressed nuclear gene. The rates of effective substitution and amino acid changes were also very high, but in contrast to silent substitutions, they are higher for α-fetoprotein than for albumin by ~70%. For α-fetoprotein, the rate of effective substitution ($1.5 \times 10^{-9}$/site/year) may be approaching that for nonfunctional pseudogenes (about $3 \times 10^{-9}$/site/year). Evolutionary divergence was also estimated at the amino acid level. It was found that the rate of change of α-fetoprotein (55% amino acids replaced in 100 Myr) approaches that of the fastest-evolving fibrinopeptides (92% amino acids replaced in 100 Myr). This high rate may indicate that α-fetoprotein can tolerate a great deal of molecular variation without its function being impaired in the process. Albumin evolves at a slower rate (39% amino acids replaced in 100 Myr), although still faster than either hemoglobin (17% amino acids replaced in 100 Myr) or cytochrome c (5% amino acids replaced in 100 Myr). The slower evolutionary rate may indicate that albumin has more refined functional specifications and hence can tolerate fewer mutational changes. The latter conclusion remains, however, to be reconciled with the condition of inherited analbuminemia, where a virtually complete absence of albumin produces surprisingly few symptoms.

Introduction

As a consequence of a certain amount of inaccuracy in the DNA polymerase reaction, mutations accumulate throughout the genome during DNA replication. However, some of these mutations escape our detection because they are not fixed in a population during the process of evolution, or perhaps they give rise to a defective protein and therefore are eliminated by natural selection (Zuckerkandl and Pauling 1965; Dickerson 1971). Thus, in the course of evolution, a protein is free to change its structure as long as the change does not impair its function. And within this framework of limited freedom, a correlation is observed between the rate of evolutionary change and the necessity to conserve a preexisting structure (Dickerson 1971). The rate at which a protein evolves is thus inversely related to the amount of selection imposed on it.

1. Key words: DNA sequence, gene divergence, molecular clock, pseudogene, protein structure and function.

Address for correspondence and reprints: Dr. Achilles Dugaiczyk, Department of Biochemistry, University of California, Riverside, California 92521.

© 1985 by The University of Chicago. All rights reserved.
0737-4038/85/0204-0317$2.00
Serum albumin and α-fetoprotein (AFP) are both major plasma proteins that are synthesized at different times during development by the embryonic yolk sac and liver (Gitlin and Perricelli 1970; Rouslahti et al. 1974; Gitlin 1975). The corresponding genes are closely linked on chromosome 5 in the mouse (D'Eustachio et al. 1981; Ingram et al. 1981), they map within bands q11–22 on chromosome 4 in humans (Harper and Dugaiczyk 1983), and they have been recently recognized as belonging to the same gene family (Law and Dugaiczyk 1981; Gorin et al. 1981; Jagodzinski et al. 1981; Sargent et al. 1981; Beattie and Dugaiczyk 1982; Dugaiczyk et al. 1982). Interestingly, a genetically transmitted disorder, termed analbuminemia, is characterized by a deficiency of albumin in humans (Gitlin and Gitlin 1975) and rats (Nagase et al. 1979). Despite the fact that albumin is a major serum protein, possessing reportedly important physiological functions (Rothschild and Oratz 1976) in normal adults, its absence in analbuminemics produces surprisingly few clinical symptoms (Gitlin and Gitlin 1975; Boman et al. 1976). Because of the questionable functional importance of serum albumin and the elusive function of α-fetoprotein (Tomasi 1977), we thought it would be worthwhile to examine these proteins from the perspective of their evolution. In our present effort we have cloned and sequenced mouse albumin cDNA and compared this sequence to previously published albumin and AFP sequences from rodents and man (Jagodzinski et al. 1981; Law and Dugaiczyk 1981; Beattie and Dugaiczyk 1982; Dugaiczyk et al. 1982; Morinaga et al. 1983).

Material and Methods
Cloning of the Mouse Albumin cDNA

Total RNA was isolated from adult BALB/c mouse livers (Chirgwin et al. 1979) and used as a template for AMV reverse transcriptase. Messenger RNA that was converted to double-stranded cDNA was cloned into the PstI site of the plasmid pBR322, as previously described (Law et al. 1980). Recombinant clones were screened with a cDNA probe obtained from immunoprecipitated albumin-containing polysomes (Taylor and Tse 1976).

Identification of the Recombinant Plasmid pMA615

The largest of the positively hybridizing clones, pMA615, was further identified on the basis of a hybrid-arrested in vitro translation (Law et al. 1980) of mouse liver mRNA, which specifically arrested the translation of the albumin mRNA species (fig. 1). The unambiguous identity of this clone was ascertained by DNA sequence analysis according to the method of Maxam and Gilbert (1980), as shown in figure 2. For comparative purposes, the Morris hepatoma 7777 rat albumin sequence (Sargent et al. 1981) is also shown in figure 2, together with the inferred amino acid sequence.

Terminology

We shall follow an accepted terminology of using substitution for nucleotide changes and replacement for amino acid changes fixed during evolution. Nucleotide substitutions that do not change the encoded amino acid are called silent, but those that do change it are called either nonsilent or replacement substitution. There is no consensus as to which of the two terms is preferred, possibly because nonsilent, being a negative term, lacks precision, while replacement substitution is an awkward
Evolution within the Albumin-α-Fetoprotein Family 349

1 + pMALB 615 DNA  
2 - pMALB 615 DNA

Fig. 1.—Hybrid-arrested translation of mouse albumin mRNA. Total mouse liver poly A⁺ mRNA was translated in a reticulocyte lysate containing ³⁵S-methionine (Pelham and Jackson 1976). Translation products were separated electrophoretically in a 12.5% SDS-acrylamide gel (Laemmli 1970). Lane 1: prior to translation, the mRNA was prehybridized (Law et al. 1980) with DNA from the recombinant plasmid pMA615. Lane 2: prior to translation, the mRNA was mock-hybridized under the same conditions as the sample above, except that pMA615 DNA was replaced by water. The arrow indicates the position of albumin.

mixture of the two types of change involved. In the present work, those substitutions that affect the encoded amino acid and bring about its change will be called effective.

Calculating Sequence Divergence

The number of silent and effective sites in a coding sequence was calculated according to the method of Miyata and Yasunaga (1980). However, we have eliminated from consideration the 23 substitutions that generate termination triplets from amino acid codons. The number of substitutions per site ($\lambda_S$) was obtained by dividing the observed number of silent substitutions by the averaged value of silent sites in two diverging sequences (Fitch 1980). Similarly, $\lambda_E$ was calculated by dividing the observed number of effective substitutions by the averaged value of effective sites in two diverging sequences. Rates of evolutionary change were then calculated using the formula $K = -\frac{3}{4} \ln(1 - \frac{\lambda}{\lambda_E})$, according to Kimura (1981) and Jukes and Cantor (1969), where $\lambda$ is the observed number of substitutions per site.

Sequence Data

Sequence changes were scored only in the regions of mature proteins because albumin and AFP differ in their leader peptides. Nascent albumin has both a pre- and a propeptide, whereas nascent AFP is devoid of the prosequence. The total number of amino acid residues in human albumin is 585, the last amino acid being leucine (TTA). In the rat albumin sequence, this codon was found to be a termination codon (TAA), and hence the protein is shorter by one amino acid. We have eliminated this 585th codon from our analysis. In the present study, the mouse
Fig. 2.—Nucleotide sequence of murine albumin mRNA (M) determined from cloned cDNA. DNA sequence analysis was performed according to the method of Maxam and Gilbert (1980). The amino acid sequence was inferred from the nucleotide sequence. The nucleotide sequence of rat albumin mRNA (R) (Sargent et al. 1981) is shown below the mouse sequence to indicate sequence homology. The amino acids indicate replacements.

albumin sequence extended 481 codons. The codons correspond to amino acids 75–492 of the human and rat albumin sequence, where the first 74 and the last 92 residues are missing. We have calculated rates of substitution and replacement not only in the complete sequences but also in the 75–492 region that corresponds to the incomplete mouse albumin sequence.
The total number of amino acid residues in mouse AFP is 586; the corresponding human sequence has 590 residues. The extra four in the latter constitute a stretch of four consecutive amino acids, and they have been eliminated from our comparison with the corresponding rodent data. On reexamination of our mouse AFP sequence data (Law and Dugaiczyk 1981), we discovered a reading mistake in one sequencing gel. The reported 579th codon GAA (Glu) should read AAA (Lys), and the corrected sequence is used in the present study.

Results

α-Fetoprotein Accumulates More Amino Acid Changes

A comparison of the accumulated silent and effective substitutions as well as amino acid changes is given in table 1. The comparisons are made in the full-size sequence (584 codons for albumin; 586 codons for AFP) and in the shorter region of 418 codons (residues 75–492) for which the mouse albumin sequence is available. As can be seen in table 1, the results in the shorter (418) region are very similar (within 8%) to those obtained in the full-size molecule. Thus, the shorter (418) sequence reflects fairly accurately the overall changes that have accumulated in the whole molecule. Differences between species in this (418) region are therefore taken to reflect evolutionary changes in the human and rodent albumin and AFP molecules. In the rodent species, these differences indicate that effective substitutions and amino acid replacements accumulate at a rate ~70% faster for AFP than for albumin. Although only 36% more amino acid changes accumulated in AFP than in albumin during the human-rodent separation, it is clear from the overall data that α-fetoprotein diverges appreciably faster than the closely related albumin.

### Table 1

**Nucleotide and Amino Acid Changes in the Evolution of Albumin and α-Fetoprotein**

<table>
<thead>
<tr>
<th></th>
<th>SILENT SUBSTITUTIONS</th>
<th>EFFECTIVE SUBSTITUTIONS</th>
<th>AMINO ACID CHANGES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_s$</td>
<td>$K_s$</td>
<td>$K_s$/year ($\times 10^{-9}$)</td>
</tr>
<tr>
<td><strong>Albumin:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human/Rat (584)</td>
<td>0.608</td>
<td>1.249</td>
<td>7.4</td>
</tr>
<tr>
<td>Human/Rat (418)</td>
<td>0.604</td>
<td>1.230</td>
<td>7.2</td>
</tr>
<tr>
<td>Human/Mouse (418)</td>
<td>0.516</td>
<td>0.873</td>
<td>5.1</td>
</tr>
<tr>
<td>Rat/Mouse (418)</td>
<td>0.293</td>
<td>0.371</td>
<td>6.2</td>
</tr>
<tr>
<td><strong>α-Fetoprotein:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat/Mouse (418)</td>
<td>0.319</td>
<td>0.416</td>
<td>6.9</td>
</tr>
<tr>
<td>Human/Mouse (418)</td>
<td>0.496</td>
<td>0.811</td>
<td>4.8</td>
</tr>
<tr>
<td>Human/Rat (418)</td>
<td>0.554</td>
<td>1.006</td>
<td>5.9</td>
</tr>
<tr>
<td>Human/Mouse (586)</td>
<td>0.492</td>
<td>0.802</td>
<td>4.7</td>
</tr>
</tbody>
</table>

**Note.** $K$, the number of substitutions per site in two diverging sequences, was calculated using the formula $K = -\frac{3}{4} \ln(1 - \frac{1}{3} \lambda)$, according to Kimura (1981), where $\lambda$ is the observed number of differences per site in the two sequences. $K$/year $= K/2T$, where $T$ = number of years since divergence from a common ancestor. The value $n$ is the number of observed amino acid differences divided by the number of amino acid sites compared. These values were then corrected ($m$) for multiple changes at the same site, according to Poisson: $m = -\ln(1 - n)$. The underlined values were used for plotting data in fig. 4 and for comparing the relative rate of AFP vs. albumin evolution (0.182/0.106 = 1.72; or 0.463/0.342 = 1.35).
α-Fetoprotein Accumulates More Drastic Amino Acid Changes

In addition to the faster accumulation of amino acid replacements, α-fetoprotein accumulates more drastic amino acid changes, which augment the rapid evolution of this protein. This can be seen on a matrix of amino acids arranged according to their relative similarity, such as according to size, solubility, pKₐ, and degree of polar or nonpolar character of side groups (Doolittle 1979). Conservative changes are those that involve two amino acids within a small group of (four) consecutive amino acids with similar properties. Such changes will cluster (within four squares) along the diagonal of the matrix (fig. 3). More drastic changes involve more dissimilar amino acids, and these lie farther apart on the matrix. Such changes will be dispersed farther away from the diagonal.

Amino acid changes that have accumulated between rat and mouse were plotted on such a similarity matrix (fig. 3). For albumin, it was found that only 14% (6/42) of the replacements represent drastic amino acid changes. The remaining 86% (36/42) are considered conservative changes, and the ratio of conservative to drastic events is 36/6 = 6. However, a similar plot for the AFP replacements showed that a much larger proportion, 36% (25/70), were drastic amino acid changes (fig. 3). The remaining 64% (45/70) represent conservative changes, and the ratio of conservative to drastic replacements is only 45/25 = 1.8.

The classification of conservative versus radical change of amino acids also can be made on the basis of their relative occurrence in nature. The amino acid changes that occur most frequently can be considered conservative; while being tolerated by selection, they must have introduced little or no phenotypic changes. Although there are 190 possible amino acid exchanges, some of them are never observed. Of the 1,572 exchanges observed on a variety of closely related proteins (Dayhoff et al. 1978), the 30 most commonly observed were taken and marked as shaded areas in figure 3. These most commonly occurring amino acid changes represent nature's conservative events.

Amino acid changes that have accumulated between rat and mouse can thus be viewed on the same matrix (fig. 3), by distinguishing their location on shaded or nonshaded squares. In such a comparison for albumin, only 17% (7/42) of the replacements are found to represent drastic amino acid changes. The remaining 83% (35/42) are conservative changes, and the ratio of conservative to drastic events is 35/7 = 5. For AFP, however, 43% (30/70) of the replacements are drastic amino acid changes. An almost equal proportion, 57% (40/70), are conservative changes, and the ratio of conservative to drastic events is only 40/30 = 1.3.

It is evident from the above information that essentially the same results were obtained by two different criteria used for evaluating the radical or conservative nature of changes accumulating in evolution. But this should not be surprising, since most of the frequently observed changes involve amino acids with similar chemical properties (see fig. 3). What is perhaps surprising is the large number of drastic amino acid changes occurring in AFP. In fact, some of them, like the Pro/Asp or the Lys/Tyr change, occur very infrequently (1 in 1,572) in a vast number of proteins (Dayhoff et al. 1978). In AFP the frequency is 1 in 70. We have previously observed that AFP has two disulfide bonds fewer than the 17 that maintain the secondary structure of albumin (Law and Dugaiczyk 1981). It is notable that AFP can tolerate even the loss of such structural features as disulfide bonds and still retain its putative function.
Fig. 3.—Matrix of amino acid exchanges. Amino acids are arranged along the axes according to the relative similarity of their chemical properties (Doolittle 1979). The observed exchanges are plotted on the matrix for albumin and α-fetoprotein; a total of 418 amino acids of each species is compared. Each dot represents one of the 42 amino acid differences observed between rat and mouse albumin, or one of the 70 differences observed between rat and mouse α-fetoprotein. Conservative changes are considered to be those that occur within a group of four consecutive amino acids. Their corresponding dots are located within four squares along the diagonal. In albumin, only 14% (6/42) of the changes are classified as radical, while in α-fetoprotein (AFP) the proportion reaches 36% (25/70). Conservative changes were also classified on the basis of their frequency of occurrence in nature (Dayhoff et al. 1978); amino acid changes determined as occurring most frequently in nature are indicated by shaded squares. According to this criterion, only 17% (7/42) in albumin are radical changes, while the proportion increases to 43% (30/40) in AFP. The two numerator values, displayed as \( \frac{40-45}{70} = 61\% \), were obtained by the two methods (Dayhoff et al. 1978; Doolittle 1979) of evaluating the conservative or radical nature of replaced amino acids. Therefore, the 61% figure represents an average of data procured by these two methods. Note that compared to albumin, a significantly larger proportion of dots (replacements) is found away from the diagonal and outside of shaded squares for AFP.
The Molecular Clock

(a) Silent Substitutions

The data in table 1 reveal a high rate of silent substitution (between $5 \times 10^{-9}$ and $7 \times 10^{-9}$/site/year) for both the albumin and AFP genes, perhaps the highest so far reported for an expressed nuclear gene. Only two other estimates are comparable with our values. One ($7 \times 10^{-9}$/site/year) is the rate for the silent substitution in the C-peptide region of the preproinsulin genes (Perler et al. 1980), and the other ($6 \times 10^{-9}$/site/year) is that of the rapidly evolving fibrinopeptides (Kafatos 1977).

Although the silent substitution rates between albumin and AFP do not differ by much, more substitutions appear to have accumulated in the rat than in the mouse lineage. If this differential accumulation of substitutions in one lineage is taken into account, then the rate of silent substitution is fairly linear with evolutionary time for the albumin locus. Thus, during the rat/mouse divergence $K_s$/year = $6.2 \times 10^{-9}$. For the human/mouse divergence, this value drops to $5.1 \times 10^{-9}$, but it increases to $7.2 \times 10^{-9}$ for the human/rat divergence, giving an average $K_s$/year of $6.2 \times 10^{-9}$ for the human/rodent divergence. In other words, the rate of silent substitution is the same in the first 30 Myr of the rat/mouse separation as it is during the 85 Myr period of the human/rodent separation.

(b) Effective Substitutions and Amino Acid Changes

Since effective substitutions are the underlying cause of amino acid replacements, the two types of change closely reflect each other. For both genes investigated, the rates of effective substitution and amino acid changes are quite high, but in contrast to silent substitutions, effective substitutions are higher for $\alpha$-fetoprotein than for albumin (table 1). For AFP the rate of effective substitution/site/year ($1.5 \times 10^{-9}$) could be approaching that of globin pseudogenes at $3-4 \times 10^{-9}$ (Li et al. 1981; Li and Gojobori 1983). The high rate of evolutionary change of $\alpha$-fetoprotein is perhaps best illustrated by comparing it to that of some well-characterized proteins. The underlined data shown in table 1 were therefore taken in order to plot the accumulated amino acid changes versus separation time of species (fig. 4). As can be seen from this graph, the rate of change of $\alpha$-fetoprotein (55% amino acids replaced in 100 Myr) approaches that of fibrinopeptides (92% amino acids replaced in 100 Myr). Fibrinopeptides are known for having the fastest rate of evolution and for their lack of a specific function (except for being a fragment that is discarded from fibrinogen when the latter is activated to form fibrin during blood clotting). AFP can tolerate more amino acid changes than can any other protein with a known specific function. This suggests that there are few, if any, functional constraints imposed on AFP, so that the vast quantity of molecular variation remains indifferent to the putative function of this protein. On the other hand, albumin appears to be less tolerant of structural alteration because it has accumulated fewer and less drastic changes in the course of its evolution.

Is the “clock” geared to generation time or to absolute time?

There are two types of homologous genes, the distinction arising from the initial mechanism by which they diverged from one another. Orthologous genes diverged after speciation, and they perform essentially the same function (cytochrome c) in the new species. Paralogous genes, on the other hand, diverged from one another after gene duplication (and possibly before speciation), and they perform
somewhat different functions (various hemoglobins) within the same species. While a comparison of orthologous genes always involves two species, which often have different generation times and other biological parameters, a comparison of paralogous genes eliminates all these variables because it is performed within one species. Orthologous genes reveal only differences that have accumulated between two diverging lineages, and one can never tell what proportion of the difference can actually be attributed to each individual lineage. Paralogous genes, on the other hand, reveal direct changes that have accumulated within a single line of evolution.

A comparison of the amino acid sequence for the third domain of the albumin and AFP genes shows them to be paralogous. Even without gaps, 44% of the amino acids are identical when the comparison is between the human sequences and 36% for the rat sequences. Nucleotide sequence comparison leads to a similar conclusion. The number of sequence differences between albumin and AFP that have accumulated since the duplication of the ancestral gene is thus greater in the rodent lineage (64%) than in the human lineage (56%). The shorter generation time of the rodents could be a factor contributing to the faster divergence of their genomes.

Discussion

The concept of a molecular clock (Zuckerkandl and Pauling 1965; Dickerson 1971) is like a unified theory of modern biology. It maintains that the rate of change
over evolutionary time has been essentially constant for a given protein. It is further argued (Dickerson 1971) that the mechanism driving the clock is geared to the function of a protein, so that one can predict the functional constraints, or physiological importance, of a given protein from its evolutionary clock.

The recently accumulated cDNA information for albumin and α-fetoprotein from three species provided us with the opportunity to test the aforementioned concept with new experimental data. Specifically, we wanted to probe the function of albumin and α-fetoprotein by looking at the molecular clock of each protein. It was found that AFP evolves at a very fast rate, one approaching that of the fastest-evolving fibrinopeptides and globin pseudogenes. The results seem to imply that the structure of AFP can change significantly without making any difference in its putative function. Most of the variation appears to be indifferent to natural selection. The results could also be interpreted as reflecting an adaptive process to a rapidly changing function of α-fetoprotein. However, it is difficult to see the need for a significant change in function of AFP in closely related species such as rat and mouse.

Albumin evolves at a slower rate than does AFP by a factor of ~1.7. According to the arguments about the molecular clock, a slower rate of evolution implies that a moderate amount of selection is being imposed on the structure of albumin. This interpretation would comply with the concept of the evolutionary clock, were it not for the fact that albumin appears to be dispensable and therefore should seemingly diverge faster than is observed. Surprisingly, the absence of serum albumin produces no serious clinical symptoms in humans and rats with inherited analbuminemia (Gitlin and Gitlin 1975; Nagase et al. 1979), and the mutation leading to the disorder has not been eliminated by natural selection.

Acknowledgment

This work was supported by NIH grant 1RO1 HD 19281-01.

LITERATURE CITED


DUGAICZYK, A., S. W. LAW, and O. E. DENNISON. 1982. Nucleotide sequence and the


MASATOSHI NEI, reviewing editor

Received July 7, 1983; revision received March 19, 1985.