# Silent Nucleotide Substitutions and the Molecular Evolutionary Clock

Thomas H. Jukes

*Summary.* Half of the nucleotide substitutions during the evolutionary divergence of genes in animals, bacteria, and viruses are silent changes. These result from an inherent biochemical property of DNA and are fixed by genetic drift. Evolution may be viewed as a device for protecting DNA molecules from extinction.

Sequences of messenger RNA (mRNA) molecules are becoming available directly or through sequencing of genes in DNA, thus making it possible to reappraise evolutionary comparisons of homologous protein molecules. Numerous such comparisons have been made by aligning amino acid sequences, but these comparisons have not revealed the occurrence of silent nucleotide substitutions, predominantly in the third positions of codons, in the corresponding regions of DNA. These are changes in nucleotide sequences that do not produce corresponding changes in amino acid sequences. One of the first surprises was finding 33 percent of silent substitutions in comparing portions of mRNA's for histone H4 in two species of sea urchin (*1*). This was remarkable, because the amino acid sequences of histone H3 show only a 3 percent difference between species as unrelated as peas and cows. Further results, summarized by Jukes and King (*2*), revealed high proportions of silent nucleotide substitutions in comparisons of the genes of globins and of virus proteins. Nichols and Yanofsky (*3*) found that the nucleotide sequences of *trpA* genes of *Escherichia coli* and *Salmonella typhimurium* were 24.5 percent divergent as compared with an amino acid sequence divergence of 14.9 percent. They concluded that most synonymous codon changes in *trpA* were selectively neutral.

Neutral, or near-neutral, nucleotide substitutions in DNA accumulate during evolutionary divergence as a result of genetic drift (*4*). The changes are initiated by point mutations, many of which are detrimental, and are rejected by natural selection or are corrected by repair mechanisms. Some replacements are adaptive, and there has been a debate as to how many are adaptive and how many are neutral or near-neutral (*2*). In this article certain evolutionary effects of nucleotide substitutions are examined.

Table 1 groups nucleotide substitutions in various genes with respect to their presence in codons and their effect on amino acid replacements. It is a common practice to classify amino acid replacements on the basis of the minimum changes in codons needed to effect the replacement. For example, serine occurring at two homologous locations is recorded as no change when actually one substitution, as in UCU to UCC, two as in UCU to AGU, or three as in UCU to AGC (U, uracil; C, cytosine; G, guanine; A, adenine) may have taken place. Furthermore, as the number of substitutions increases, many sites, for stochastic reasons, will receive more than one "hit," and some of these will result in revertants. These considerations are included in the "random evolutionary hit" theoretical model of molecular evolution (*5*, *6*). Tables 1 and 2 show that a large number of silent nucleotide substitutions occurs during evolutionary divergence. King and I proposed that such replacements are neutral changes (*7*). The alternative explanation would be that they are adaptive for reasons such as necessary species differences in mRNA secondary structure or in transfer RNA (tRNA) utilization. But these reasons would require that closely related species must have marked differences in the manner of expression of each pair of homologous genes, which would impose an enormous burden of concomitant molecular changes. It appears much more likely that genetic drift plays a leading role in evolution, and that it produces a steady accumulation of nucleotide replacements in DNA.

Table 2 shows that, except for examples 1 and 2 (which have diverged only a small amount), the silent substitutions as the percentage of all nucleotide substitutions show a downward trend as the percentages of total nucleotide substitutions and amino acid replacements show an upward trend. Simultaneously, codons with only single-base substitutions tend to decrease, because an increasing number of codons receives more than one "hit." The calculation of minimum base differences per 100 codons reveals an average of less than half of the base differences actually present (*8*).

## Comparisons of Genes

*Mouse β-globins.* The mRNA's for two mouse β-globin genes (major and minor) were found (*9*) to differ by 17 nucleotide substitutions; seven are silent, and the remaining ten have produced nine amino acid replacements. Konkel *et al.*, who made these observations (*9*), concluded that the two genes "presumably evolved from a common ancestral gene as recently as 50 million years ago." In this example, nucleotide substitutions have accumulated without any known change in protein function, and about 40 percent of them are silent. It is difficult to conceive that duplicate genes for variants of the same protein would require different sets of tRNA's in translation, or would give rise to two mRNA's with different secondary structures. The occurrence of genetic drift seems the best explanation for the changes.

However, continued evolutionary divergence of these two genes could lead to one of them acquiring a new function. Duplication thus gives a gene a "fresh start," but the extent to which genetic drift is involved in divergence was not previously measurable. Perhaps mouse β-globin minor will evolve into a globin of the δ type that is now found in human beings and some apes.

*Human and rat hormones.* Human growth hormone (hGH) and human chorionic somatomammotropin (hCS) are homologous, although hGH is secreted by the hypophysis and hCS by the placenta. The primary structures of mRNA's for hCS (*10*) and hGH (*11*) have

The author is a professor in the department of biophysics and medical physics at the University of California, Berkeley and the Space Sciences Laboratory 94720.



0036-8075/80/1128-0973$01.25/0 

been reported. These sequences may be compared with each other and with that of rat growth hormone (rGH) (*12*). The results are in Tables 1 and 2. The divergence between hGH and hCS appears to be much more recent than the divergence between the common ancestor of these two hormones and rGH. It is interesting that, with respect to amino acid sequences, hGH and rGH are more similar than are hCS and rGH, even though the divergences in the nucleotide sequences are almost identical (75 and 77 percent).

The mRNA sequences of human preproinsulin and rat preproinsulin I were compared by Bell *et al.* (*13*). Silent changes were found in all regions. The findings are summarized in Tables 1 and 2.

*Sea urchin histones.* Sequences of histone genes of two species of sea urchins, *Strongylocentrus purpuratus* and *Psammechinus miliaris* (*14–16*) show a greater percentage of silent nucleotide substitutions than has been found in any other comparisons (Tables 1 and 2). There are only ten amino acid replacements in 404 codons compared. The complete DNA coding sequences of histone 2A from both species were reported, and partial sequences of the codons of histone 2B (96 codons), histone 3 (134 codons), and histone 4 (50 codons) were available for comparison.

There are 37 silent substitutions per 100 codons in the comparison of the two sea urchin histone genes. These two species belong to the same subclass of Echinodermata, and it would seem unlikely that two such closely related species would have so many substitutions for functional reasons.

*TrpA synthetases.* Comparison of *trpA* genes of *S. typhimurium* and *E. coli* shows a high proportion of silent substitutions, most of which are single-base changes in codons. Nichols and Yanofsky (*3*) found no evidence that codon selection or RNA structure was of major selective value in their comparison of these genes. The silent substitutions are a remarkable 77 percent.

Table 3 shows the number of each type of substitutions in silent third-base positions in codons plus first-base substitutions (C ⇋ U) in leucine codons (for example, CUA and CUG to UUA and UUG), in arginine codons (C ⇋ A) (for example, CGA to AGA and CGG to AGG), and two bases in serine codons (UC ⇋ AG) (for example, UCU to AGU and UCC to AGC). There is a net increase of 13 percent in the combined C + G content in the *S. typhimurium* group of silent changes as compared with *E. coli* silent changes, but it is not unidirectional; it is the difference between 94 changes toward the C + G, and 74 changes toward A + T (T, thymine).

*Globin genes.* By comparing the two $\alpha$- and four $\beta$-globin mRNA's, we can for the first time examine the end result of a molecular divergence that is estimated to have occurred about 500 million years ago (*17*). The base composition of silent substitutions is summarized in Table 3. Most prevalent are C and U interchanges. Rabbit $\alpha$-globin has the highest C content (53 percent) in third codon positions, and hence 57 percent of all substitutions are of C in comparisons of rabbit $\alpha$-globin with other globins. In general, the participation of the four nucleotides in silent third-position substitutions (U, 29 percent; C, 38 percent; A, 8 percent; G, 25 percent) of all six globins compared in Table 3 has some resemblance to their average respective abundances in third positions of all codons in these six globins (U, 22 percent; C, 40 percent; A, 6 percent; G, 32 percent). A similarity would be expected if silent nucleotide substitutions are randomized. More detailed exploration of these evolutionary comparisons is obviously possible. Changes involving G seem fewer than would be expected from its occurrence.

The $\alpha$- and $\beta$-globin mRNA's are translated consecutively on ribosomes, so that an $\alpha$-$\beta$ dimer is formed, starting during the translation process. This would seem to make it probable that the same group of tRNA's takes part in the translation of both mRNA's.

The species differences in $\alpha$-globins and in $\beta$-globins among rabbits, mice, and humans (Tables 1 and 2) date from the differentiation of the placental mammals, estimated as 100 million years ago (see below). The predominant silent substitutions are between C and U (Table 3). The differences between chicken and mammalian $\beta$-globins (Tables 1 and 2) date from an estimated 215 million years.

*Viruses (coliphages $\phi$X174 and G4).*

Table 1. Comparison of coding regions of homologous mRNA's and genes, grouped into types of substitutions in corresponding codons in the aligned nucleotide sequences [as explained in (*2, 5*)].

| Substitutions in codons | Example | | Column* | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Codon | Amino acid† | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Single-base | | | | | | | | | | | | | | | | | |
| (a) Degenerate | UUU/UUC | Phe/Phe | 7 | 12 | 132 | 114 | 40 | 84 | 97 | 30 | 101 | 47 | 39 | 487 | 182 | 109 | 86 |
| (b) Recognizable | UUU/AUU | Phe/Ile | 8 | 27 | 4 | 12 | 14 | 28 | 54 | 10 | 28 | 43 | 46 | 160 | 200 | 37 | 41 |
| Two-base | | | | | | | | | | | | | | | | | |
| (c) Degenerate | CUC/UUA | Leu/Leu | | | 5 | 6 | 1 | 1 | 1 | 1 | 5 | | | 31 | 6 | 10 | 7 |
| (d) Recorded as single-base | UUU/AUC | Phe/Ile | | 4 | 5 | 20 | 7 | 31 | 45 | 5 | 22 | 20 | 25 | 151 | 229 | 53 | 59 |
| (e) Recognizable | UUU/GAU | Phe/Asp | 1 | 1 | 1 | 2 | 1 | 9 | 18 | 1 | 11 | 7 | 8 | 80 | 79 | 32 | 40 |
| Three-base | | | | | | | | | | | | | | | | | |
| (f) Degenerate | UCU/AGC | Ser/Ser | | | | | | | 5 | | 3 | | | 3 | 1 | 1 | 1 |
| (g) Recorded as single-base | UCU/GGC | Ser/Gly | | | | 2 | 2 | 1 | 3 | | 5 | 2 | 3 | 29 | 14 | 6 | 8 |
| (h) Recorded as two-base | UCU/GUA | Ser/Val | | | 1 | 4 | 2 | 4 | 16 | 3 | 12 | 4 | 2 | 124 | 111 | 31 | 30 |
| (j) Recognizable | UUU/AAG | Phe/Lys | | | | | | | | | 2 | | | 6 | 2 | 1 | 1 |

*The column numbers in Tables 1, 2, and 3 refer to the following items (see *8*): 1, Mouse $\beta$-globin major and minor genes (*9*); 2, human growth hormone (hGH) and human chorionic somatomammotropin (hCS) mRNA's (*10, 11*); 3, histones 2A, 2B, 3, and 4 genes, *S. purpuratus* and *P. miliaris* (*14–16*); 4, *trpA* synthetase genes, *E. coli* and *S. typhimurium* (*3*); 5, rabbit and mouse $\alpha$-globin genes (*35*); 6, rabbit $\beta$-, human $\beta$-, and mouse $\beta$-globin genes (major) (*35, 36*); 7, chicken $\beta_3$ compared to rabbit, human, and mouse $\beta$-major globin genes (*35–37*); 8, human and rat preproinsulin mRNA's (*13*); 9, BK and SV40 viruses, VP1 genes (*23*); 10, hGH and rGH mRNA's (*10, 11*); 11, hCS and rGH mRNA's (*10, 12*); 12, $\phi$X174 and G4 coliphages, genes A, C, D, F, G, H, and J (*18*); 13, rabbit $\alpha$ and mouse $\alpha$ versus rabbit $\beta$, human $\beta$, mouse $\beta$-major, and chicken $\beta_3$ globin genes (*35, 37*); 14, SV40 and polyoma viruses, VP1 genes (*21*); 15, BK and polyoma viruses, VP1 genes (*22, 23*).
†Abbreviations for the amino acid residues are as follows: Asp, aspartic acid; Gly, glycine; Ile, isoleucine; Leu, leucine; Lys, lysine; Phe, phenylalanine; Ser, serine; Val, valine.

These viruses are homologous, and each has ten protein-coding genes, three of which are read by frame-shifts, that is, they are read twice, from different starting points (*18, 19*). These three genes (B, E, and K) are not included in the comparisons (Table 1) because this would mean that their nucleotides were compared twice. The figures in column 5 (Tables 1 and 2) are the totals for genes A, C, D, F, G, H, and J. More than half of the 1660 substitutions are silent changes.

The genes of phage $\phi$X174 are notably high in their content of T in third codon positions (*20*). The totals in these positions for the genes compared in Tables 1 and 2 are 45 percent T for phage $\phi$X174, and 34 percent T for phage G4. Changes (for $\phi$X174 to G4) from T to C (~ 238) predominate over those from C to T (~ 112), and changes from T to A (~ 91) predominate over those from A to T (~ 28); but G to T changes (~ 56) are more common than T to G (~ 38).

The participation of third-base nucleotides in changes (T, 34 percent; C, 29 percent; A, 19 percent; G, 18 percent) is fairly concordant with the average representation of the four bases in third positions in codons in the two viruses (T, 39 percent; C, 24 percent; A, 19 percent; G, 18 percent).

*Viruses (mammalian) SV40, polyoma, and BK.* Simian virus 40 (SV40), polyoma, and BK are small (mammalian) DNA viruses; SV40 multiplies in primate cells, and polyoma causes cell transformation in vitro and tumors in newborn hamsters. BK is a papovavirus of human origin (*21*). Homology in these viruses, including protein coding regions, has been demonstrated by Friedmann and co-workers (*21, 22*), and by Yang and Wu (*23*).

The genes for protein VP1 in the three viruses are compared in Tables 1 to 3. Total nucleotide substitutions, per 100 codons, are about the same in comparisons of polyoma with the other two viruses, as in the case of the comparisons between $\alpha$- and $\beta$-globin genes. Since vi-

Table 2. Summary derived from data in Table 1, showing the relation between nucleotide substitutions and amino acid replacements in evolutionary comparisons of homologous genes and mRNA's.

| Item | Column* | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Nucleotides compared | 441 | 651 | 1212 | 804 | 426 | 1293 | 1308 | 330 | 1086 | 648 | 645 | 5094 | 3330 | 1074 | 1074 |
| Nucleotide substitutions, per 100 codons | 12 | 23 | 40 | 75 | 59 | 48 | 81 | 57 | 75 | 75 | 77 | 98 | 126 | 127 | 128 |
| Silent nucleotide substitutions, per 100 codons | 5 | 7 | 37 | 57 | 39 | 29 | 42 | 36 | 46 | 35 | 33 | 52 | 51 | 64 | 58 |
| Amino acid replacements, per 100 codons | 6 | 15 | 3 | 15 | 18 | 17 | 31 | 17 | 23 | 35 | 39 | 32 | 57 | 45 | 50 |
| Codons with only one-base substitutions as percent of all codons with substitutions | 94 | 89 | 92 | 79 | 81 | 71 | 63 | 80 | 68 | 73 | 69 | 60 | 46 | 52 | 47 |
| Silent substitutions as percent of all nucleotide substitutions | 41 | 33 | 92 | 77 | 65 | 59 | 52 | 63 | 62 | 46 | 43 | 54 | 41 | 50 | 45 |

*Column numbers are explained in a Table 1 footnote.

Table 3. Nucleotide substitution in silent positions in codons of sequences that are compared in Tables 1 and 2. These are third-position substitutions, except for certain Arg/Arg, Leu/Leu, and Ser/Ser codon comparisons. Arg, arginine.

| Item | Column* | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| T (U) ⇆ C | 3 | 8 | 71 | 53 | 24 | 51 | 81 | 10 | 57 | 31 | 25 | 350 | 222 | 59 | 39 |
| A ⇆ G | 3 | 4 | 31 | 37 | 9 | 22 | 16 | 15 | 32 | 18 | 15 | 135 | 45 | 39 | 37 |
| C ⇆ G | | 3 | 8 | 27 | 10 | 10 | 29 | 6 | 8 | 5 | 9 | 62 | 152 | 20 | 18 |
| T (U) ⇆ G | 1 | 2 | 7 | 17 | 3 | 8 | 14 | 6 | 15 | 6 | 7 | 94 | 72 | 27 | 20 |
| A ⇆ C | | 1 | 19 | 9 | 2 | 6 | 9 | 1 | 12 | 9 | 9 | 62 | 23 | 32 | 40 |
| A ⇆ T (U) | | | 12 | 7 | 1 | 10 | 2 | 3 | 29 | 3 | 3 | 119 | 14 | 43 | 36 |

*Column numbers are explained in a Table 1 footnote.

Table 4. Differences between globins as percent substitutions in amino acid sequences; differences between $\alpha$-globins are below the diagonal line and differences between $\beta$-globins are above the diagonal line with the same numbers used to designate species.

| $\alpha$ chain | $\beta$ chain | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 1. Human | | 7 | 19 | 10 | 10 | 17 | 26 | 21 | 23 | 31 | | 44 | | 62 |
| 2. Loris | 6 | | 17 | 10 | 13 | 19 | 26 | 21 | 21 | 31 | | 44 | | 60 |
| 3. Mouse | 11 | 13 | | 20 | 20 | 27 | 30 | 27 | 29 | 34 | | 42 | | 62 |
| 4. Rabbit | 18 | 17 | 19 | | 14 | 20 | 25 | 21 | 23 | 33 | | 41 | | 62 |
| 5. Dog | 16 | 15 | 16 | 20 | | 19 | 24 | 24 | 25 | 29 | | 42 | | 64 |
| 6. Bovine | 12 | 17 | 13 | 18 | 20 | | 30 | 25 | 28 | 37 | | 45 | | 63 |
| 7. Kangaroo | 19 | 21 | 22 | 26 | 23 | 18 | | 27 | 27 | 32 | | 48 | | 62 |
| 8. Echidna | 26 | 26 | 30 | 35 | 30 | 30 | 35 | | 10 | 29 | | 42 | | 62 |
| 9. Platypus | 28 | 28 | 30 | 33 | 32 | 31 | 33 | 6 | | | | 45 | | 62 |
| 10. Chicken | 25 | 28 | 28 | 30 | 31 | 27 | 29 | 34 | 33 | | | 46 | | 59 |
| 11. Viper | 35 | 37 | 37 | 37 | 38 | 38 | 38 | 47 | 45 | 40 | | | | |
| *12. Amphibians | 44 | 45 | 45 | 49 | 46 | 45 | 48 | 50 | 46 | 45 | 51 | | | 60 |
| 13. Carp | 49 | 47 | 49 | 51 | 48 | 46 | 51 | 54 | 53 | 51 | 52 | 53 | | |
| 14. Shark | 56 | 57 | 56 | 54 | 57 | 54 | 57 | 59 | 57 | 59 | 59 | 60 | 61 | |

*Newt, $\alpha$; frog, $\beta$.

Table 5. Amino acid differences between certain $\alpha$- and $\beta$-globin sequences (percent). The numbers at the top of the table designate species as in the first column.

| $\alpha$ chain | $\beta$ chain | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | Frog | 10 |
| 1. Human | 54 | 55 | 56 | 61 | 57 | 58 | 64 | 65 |
| 2. Mouse | 53 | 53 | 53 | 60 | 56 | 59 | 66 | 65 |
| 3. Rabbit | 57 | 60 | 59 | 62 | 60 | 62 | 69 | 65 |
| 4. Kangaroo | 57 | 57 | 57 | 60 | 58 | 59 | 64 | 69 |
| 5. Opossum | 63 | 60 | 64 | 66 | 61 | 63 | 71 | 71 |
| 6. Echidna | 55 | 53 | 56 | 58 | 55 | 58 | 63 | 66 |
| 7. Chicken | 61 | 59 | 60 | 63 | 62 | 63 | 63 | 68 |
| 8. Viper | 59 | 60 | 58 | 62 | 59 | 61 | 67 | 70 |
| 9. Carp | 56 | 51 | 58 | 58 | 58 | 62 | 64 | 65 |
| 10. Shark | 66 | 65 | 65 | 65 | 63 | 66 | 68 | 63 |

ruses do not leave fossils, one cannot place their evolution on a time scale, but the VP1 genes in SV40 and BK seem to be diverging from polyoma VP1 at approximately equal rates, as measured by total nucleotide substitutions (*23a*). BK and SV40 are more closely related to each other than to polyoma. The average base composition of silent changes in the comparisons of polyoma to BK and polyoma to SV40 is T, 31 percent; C, 37 percent; A, 8 percent; and G, 24 percent. In contrast, the base composition of silent changes in the comparison of BK to SV40 is T, 33 percent; C, 25 percent; A, 24 percent; G, 18 percent—with a high percentage of A to T or T to A interchanges (Table 3). The average composition of third codon bases in the three viral genes is T, 29 percent; C, 19 percent; A, 30 percent; G, 21 percent.

## Evolution of Globins

Molecular differences in hemoglobins and myoglobins are regarded by some as being exquisitely adapted to the physiological idiosyncrasies of the individual species of animal, structure being keyed to function (*24, 25*). Another viewpoint is that amino acid substitutions are introduced into protein molecules at a constant rate by genetic drift (*4, 26*). The first conclusion is based on biochemical studies, such as species-specific differences in the release of oxygen from hemoglobin at various *p*H values, or resistance to dissociation by urea in elasmobranch blood (*27*). The second view is supported by the striking relation between phylogeny and amino acid differences. It is indeed difficult to understand why this should be so constant if differences between hemoglobins are ascribable solely to differences in physiological function.

Table 4 shows percentage differences in amino acids between $\alpha$-globin sequences (left-hand part of the table) and between $\beta$-globin sequences (right-hand part of table). The difference in amino acid sequence between the $\alpha$ chains of the two fish, carp and shark (61 percent), is as great as the differences between shark and mammals (54 to 59 percent). The range of differences between shark $\beta$ chain and the other $\beta$ chains is 59 to 64 percent, and the average is 61.5 percent.

In contrast, the range of differences between the various placental mammal $\beta$ chains is 10 to 27 percent, and the average is 17.5 percent. The other comparisons show a decrease in scatter as divergence widens, and the same is true for the $\alpha$ chain comparisons. A "smoothing out" of differences seems to occur as phylogenetic differences increase.

Table 5 compares differences between $\alpha$ and $\beta$ chains, arranged in both directions. For example, rabbit $\alpha$ is 57 percent different from human $\beta$, and rabbit $\beta$ is 56 percent different from human $\alpha$. The uniformity is quite remarkable; even though human $\alpha$ and shark $\alpha$ differ by 56 percent, and human $\beta$ and shark $\beta$ differ by 62 percent (Table 4), yet human $\alpha$ and shark $\beta$ differ by 65 percent and human $\beta$ and shark $\alpha$ by the same amount: 66 percent. We may conjecture from Table 2, columns 12 and 13, that the $\alpha : \beta$ amino acid differences in Table 5 are probably each accompanied by about 40 to 45 percent of silent nucleotide substitutions.

Table 6 summarizes the amino acid differences in vertebrate hemoglobin chains. There is a steady divergence when the $\alpha$ and $\beta$ differences are averaged. This corrects the anomalies for chicken $\alpha$ and monotreme $\beta$ seen in the first two columns. Dickerson (*17*) has plotted the rates of evolution of hemoglobins as amino acid changes per 100 residues, against millions of years since divergence. He assumes that the separation of fish from higher vertebrates was 400 million years ago. On this basis, and correcting, as he did, for multiple hits, we obtain dates for divergence as shown in Table 6.

## Molecular Evolutionary Clock

When amino acid sequences of homologous proteins in two different species are compared, evolutionary replacements can be perceived. The time since the two descended from a common ancestor is estimated from the fossil record, and the time per amino acid replacement may be calculated for a protein. Some proteins evolve more rapidly than others (*7, 17*). However, even in the same protein, the rate of change may vary at different times and in different species (*24, 27*).

Attempts have been made to interpret evolutionary amino acid replacements in terms of equivalent numbers of nucleotide substitutions, by relating the replacements to the genetic code. There may be multiple substitutions at the same site, including revertant changes (*28*). Efforts to correct for these possibilities have varying degrees of credibility. This may be inferred from Table 2, which shows that the number of measurable nucleotide substitutions per amino acid replacement may range from 1.8 to 5.3. However, a "clock" based on amino acid replacements measures the actual phenotypic divergence, and Wilson *et al.* (*29*) drew attention to the concept of the molecular evolutionary clock, in their words, "the most significant result of research in molecular evolution."

The findings in Table 1 can be viewed in terms of the statement (*7*) that "the genome becomes virtually saturated with such changes as are not thrown off through natural selection." If we regard DNA as being under constant pressure from nucleotide replacements caused by point mutations, it may be possible that "molecular evolutionary clock" is a result of the impact of this pressure, and of genetic drift, on the countervailing force of natural selection.

The globins of vertebrates have a common evolutionary origin, and their sequences are homologous. Their polypeptide chains contain 140 to 160 amino acid residues that are folded into a specialized tertiary structure that is essentially the same for all globins. Within these conditions there is an extraordinarily wide range of different amino acid sequences, so that only about five sites are occupied by the same amino acid in nearly all vertebrate globins. These differences are accompanied by variations of oxygen affinity, strength of binding of diphosphoglycerate, sensitivity to hydrogen ion, and other functional properties that have adaptive implications. Many vertebrates have duplicate genes for one or more globins. The rates of evolution-

Table 6. Average amino acid differences in hemoglobin chains of species from various classes of the vertebrates, compared with estimated times of divergence.

| Animals | Differences per 100 codons | | Average $\alpha + \beta$ | Corrected* ($m$) | Million years† |
|---|---|---|---|---|---|
| | $\alpha$ chain | $\beta$ chain | | | |
| Placental mammals‡, inter alia (except human versus loris) | 16.1 | 16.7 | 16.4 | 17.9 | 100 |
| Kangaroo versus placental mammals | 21.7 | 26.9 | 24.3 | 27.8 | 160 |
| Monotremes§ versus theria | 30.5 | 24.5 | 27.5 | 32.2 | 190 |
| Chicken versus mammals | 29.6 | 31.7 | 30.6 | 36.5 | 215 |
| Viper versus warm-blooded | 39.2 | | | 49.6 | 290 |
| Amphibians‖ versus terrestrial | 46.7 | 48.9 | 47.8 | 65.0 | 380 |
| Bony fish ¶ versus tetrapods | 49.3 | 49.6 | 49.5 | 68.3 | 400 |
| Shark versus bony vertebrates | 57.5 | 63.8 | 60.6 | 93.0 | 545 |
| Shark $\alpha$ versus all $\beta$'s in Table 5 | | | 66.7 | 110.0 | 625 |
| Shark $\beta$ versus all $\alpha$'s in Table 5 | | | 65.1 | 104.0 | |

*Corrected for multiple hits (17) by $m/100 = -\ln[1 - (n/100)]$. †Million years since divergence, if it is assumed that carp separation was 400 million years ago (*17*). ‡Human, loris, mouse, rabbit, dog, bovine. §Echidna, platypus. ‖Newt $\alpha$-globin, bullfrog $\beta$-globin. ¶Carp and goldfish.

ary change have been stated to be different at different times (*26*). Nevertheless, the globins are one of the best examples of the molecular evolutionary clock, and, when comparisons of distantly related vertebrates are made, there is a "levelling in the number of [amino acid replacements] despite differences in rate during some periods of time" (*30*).

### Types of Substitution

Are there two different classes of evolutionary substitutions of nucleotides, one silent and the other amino acid altering? Our evaluation is that there is no such demarcation in bacteria, where evolutionary pressures may greatly increase the C + G content of DNA (*7*). This may be accompanied by (i) filling the third positions of codons with C and G and (ii) changing the amino acid content of proteins. In those bacteria with DNA high in C + G content, alanine, arginine (CGN codons), and glycine may tend to displace isoleucine, leucine, phenylalanine, and tyrosine, as shown by Sueoka (*31*). We assume such displacements to be neutral. Indeed, only by step (ii) can the C + G composition of DNA be raised above 67 percent, and C + G in some bacteria (such as *Micrococcus lysodeikticus*) is as high as 72 percent. Conversely, other bacterial species (such as *Bacillus cereus*) contain only 34 percent C + G.

The comparison between *trpA* genes for *S. typhimurium* and *E. coli* recalls the finding by Cox and Yanofsky (*32*) that, in *E. coli*, the presence of the *mut T* gene produces, upon repeated subculture, a trend toward DNA with C + G content higher than that of the original stock.

The C + G content of vertebrate DNA is about 42 percent, and the mRNA for the six globins compared in Tables 1 and 2 is about 57 percent C + G. This level would not impose constraints on the amino acid content of globins.

Evolution is traditionally presented in terms of descriptions of phenotypes, because these pervade the biosphere and hence dominate the perception of biology. But phenotypes have short lifespans. In contrast, the continuity of DNA molecules, reaching through billions of years, is essential for the existence of life. The new findings with nucleic acid sequences (Tables 1 and 2) should be examined in terms of their contribution to the continuity of DNA. The strategy of nature for safeguarding this continuity is that DNA, by recombination, has the property of duplication, as distinct from replication, and that DNA can be translocated, and is subject to point mutations that produce heritable nucleotide substitutions, some of which become fixed. About half of the nucleotide substitutions that accumulate during evolution are silent (and probably selectively neutral). The pattern of evolution is such as to provide for persistence of DNA, and we must conclude that neutral changes are a part of this. Surplus ("selfish" or "junk") DNA accumulates in so-called higher organisms "by the spreading of sequences that had little or no effect on the phenotype" (*33*). Such events are another illustration of a neutral or near-neutral change in evolution.

### Conclusions

The steady divergence that is expressed by the inexorable molecular evolutionary clock must be reconciled with the conclusion that the primary structure of a protein results from functional requirements that are specific for each organism. Two primary properties of DNA are the acquisition of point mutations and the tendency of DNA sequences to duplicate. The most important role of DNA is self-perpetuation. This concept has been emphasized in the thinking of sociobiologists, who regard phenotypes as existing to shelter DNA molecules from an adverse environment (*34*). A molecule of DNA cannot perpetuate itself indefinitely, for it accumulates mutations that lead to blind alleys of extinction. Extinction can be circumvented by duplication and subsequent independent divergence, not only of genes, but also of entire genomes, as in the appearance of new species.

In this scheme, the specialized properties of proteins arise as part of the vast accumulative process of mutational change. Amino acid replacements that are incorporated as neutral or near-neutral can occasionally become adaptive as a result of further evolutionary changes in a protein molecule.

The progress of mutational change is now revealed as being more extensive than previously conjectured, by the new findings on silent changes. These show that genetic drift plays a leading role in evolution.

#### References and Notes


1. M. Kimura, *Nature (London)* **267**, 275 (1977); W. Min Jou and W. Fiers [*J. Mol. Biol.* **106**, 1047 (1976)] compared translated regions of MS2, f2, and R17 viruses. They found 25 "neutral point mutations" and 9 amino acid-altering changes.
2. T. H. Jukes and J. L. King, *Nature (London)* **281**, 605 (1979).
3. B. P. Nichols and C. Yanofsky, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5244 (1979).
4. M. Kimura, *Sci. Am.* **241**, 98 (November 1979).
5. R. Holmquist, C. Cantor, T. Jukes, *J. Mol. Biol.* **64**, 145 (1972).
6. T. Jukes and R. Holmquist, *ibid.*, p. 163.
7. J. L. King and T. H. Jukes, *Science* **164**, 788 (1969).
8. Comparisons exclude the initiator codon and any preceding nucleotides, but include the terminator codon. Five codons are missing from the human $\beta$-globin sequence. Overlapping genes B, E, and K are not included in column 12. The values for silent substitutions in Table 2 were calculated as follows: In Table 1, (a) + 2(c) + (d) + 3(f) + 2(g) + (h). The ex-

ample CUC to UUA, for instance, contains silent changes of C to U and C to A; and UCU to AGC is assumed to have evolved from a single-nucleotide replacement ACC to AGC. Minimum base differences are the smallest number of nucleotide substitutions that will bring about the designated amino acid replacements, and substitutions in excess of these numbers are counted as silent. Most of the silent changes are synonymous (degenerate) third-base substitutions, but in some cases the steps in silent substitutions cannot be precisely identified. For example, AGA→UUG may have been AGA→AGG→CGG→CUG→UUG or AGA→CGA→UUA→UUG. For this reason, the totals in Table 3 are lower than the totals of silent nucleotide substitutions in Table 2. In column 6, Table 3, each item contains the sum of three pairwise comparisons of three globin genes. The values in Table 2, lines 2 to 6, have been rounded to the nearest integer, for simplification. More precise values may be calculated from the data in Table 1.

9. D. A. Konkel, J. V. Maizel, Jr., P. Leder, *Cell* **18**, 865 (1979).
10. J. Shine, P. H. Seeburg, J. A. Martial, J. D. Baxter, H. M. Goodman, *Nature (London)* **270**, 464 (1977); N. E. Cooke, D. Coit, J. Shine, J. D. Baxter, J. A. Martial, *J. Biol. Chem.*, in press.
11. J. A. Martial, R. A. Hallewell, J. D. Baxter, H. M. Goodman, *Science* **205**, 602 (1979).
12. P. H. Seeburg, J. Shine, J. A. Martial, J. D. Baxter, H. M. Goodman, *Nature (London)* **270**, 486 (1977); N. E. Cooke, D. Coit, R. I. Weiner, J. D. Baxter, J. A. Martial, *J. Biol. Chem.* **255**, 6502 (1980).
13. G. I. Bell, W. F. Swain, R. Pictet, B. Cordell, H. M. Goodman, W. J. Rutter, *Nature (London)* **283**, 525 (1979).
14. I. Sures, J. Lowry, L. H. Kedes, *Cell* **15**, 1033 (1978).
15. W. Schaffner, G. Kunz, H. Daetwyler, J. Telford, H. O. Smith, M. L. Birnstiel, *ibid.* **14**, 655 (1978).
16. M. Grunstein and J. Grunstein, *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1093 (1977).
17. R. E. Dickerson, *J. Mol. Evol.* **1**, 26 (1971).
18. G. N. Godson, B. G. Barrell, R. Staden, J. C. Fiddes, *Nature (London)* **276**, 236 (1978).
19. J. C. Fiddes and G. N. Godson, *J. Mol. Biol.* **133**, 19 (1979).
20. F. Sanger *et al.*, *Nature (London)* **265**, 687 (1979).
21. T. Friedmann, A. Esty, P. La Porte, P. Deininger, *Cell* **17**, 715 (1979).
22. P. Deininger, A. Esty, P. La Porte, T. Friedmann, *ibid.* **18**, 771 (1979).
23. R. C. A. Yang and R. Wu, *Science* **206**, 456 (1979).

23a. M. Verhoeyen, R. Fang, W. Min Jou, R. Devos, D. Huylebroeck, E. Saman, and W. Fiers [*Nature (London)* **286**, 771 (1980)] compared H3 haemagglutinin genes of two strains of Hong Kong influenza virus. There were 63 nucleotide substitutions, 34 of which were silent, in 1701 nucleotide sites (567 codons) compared, and 28 amino acid replacements. The authors conclude that the divergences "accumulated over a 7-year period (1968–75)."

24. E. L. Smith, *Harvey Lect. Ser.* **62**, 231 (1967).
25. M. Goodman, G. W. Moore, G. Matsuda, *Nature (London)* **253**, 603 (1975).
26. M. Kimura, *ibid.* **217**, 624 (1968).
27. W. K. Fisher, A. R. Nash, E. O. P. Thompson, *Aust. J. Biol. Sci.* **30**, 487 (1977).
28. T. H. Jukes and R. Holmquist, *Science* **177**, 530 (1972).
29. A. C. Wilson, S. S. Carlson, T. J. White, *Annu. Rev. Biochem.* **46**, 573 (1977).
30. E. O. P. Thompson, in *Evolution of Protein Structure and Function* (Academic Press, New York, in press).
31. N. Sueoka, *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1141 (1961).
32. E. C. Cox and C. Yanofsky, *ibid.* **58**, 1895 (1967).
33. L. E. Orgel and F. H. C. Crick, *Nature (London)* **284**, 604 (1980).
34. R. Dawkins, *The Selfish Gene* (Oxford Univ. Press, London, 1975).
35. Y. Nishioka and P. Leder, *Cell* **18**, 875 (1979); F. C. Kafatos, A. Efstratiadis, B. G. Forget, S. M. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5618 (1977); C. A. Marotta, J. T. Wilson, B. G. Forget, S. M. Weissman, *J. Biol. Chem.* **252**, 5040 (1977).
36. H. C. Heindell, A. Liu, G. V. Paddock, G. M. Studnicka, W. A. Salser, *Cell* **15**, 43 (1978); J. Van Dan Berg, A. Van Ooyen, N. Matei, A. Schambock, G. Grosveld, R. A. Flavell, C. Weissman, *Nature (London)* **276**, 37 (1978); D. A. Konkel, S. M. Tilghman, P. Leder, *Cell* **15**, 1125 (1978).
37. A. Efstratiadis, F. C. Kafatos, T. Maniatis, *Cell* **10**, 571 (1977); R. I. Richards, J. Shine, A. Ullrich, J. R. E. Wells, H. M. Goodman, *Nucleic Acids Res.* **7**, 1137 (1979).
38. Supported by NASA grant NGR 05-003-460.

# Prostaglandins, Arachidonic Acid, and Inflammation

Frederick A. Kuehl, Jr., and Robert W. Egan

The prostaglandins (PG's) are a complex group of oxygenated fatty acids that have been detected in virtually all mammalian tissues thus far examined. They include some of the most potent natural substances known, and are important both as bioregulators and as participants in pathological states. The prostaglandins are not stored free in tissues, but are synthesized as a result of membrane perturbations that cause the release of free fatty acids, generally arachidonic acid, from esterified lipid sources. The release of arachidonic acid can be brought about by a wide variety of hormones either directly or indirectly (*1*), as well as by inflammatory or immunological stimuli (*2*, *3*), calcium ionophores (*4*), ultraviolet light (*5*), melittin, the membrane active component of bee venom (*6*), tumor promoting agents (*7*), and even mechanical agitation (*8*). The free arachidonic acid then reacts with prostaglandin cyclooxygenase, the first enzyme of the prostaglandin biosynthetic sequence. This enzyme oxygenates arachidonic acid to the endoperoxide intermediate, $PGG_2$, which is then converted to a variety of other biologically active products, the nature of which is determined by the enzyme content of the tissue under consideration. For example, platelets make primarily thromboxane $A_2$ ($TXA_2$) whereas the aorta forms prostacyclin ($PGI_2$).

The nature of these products is depicted in Fig. 1 with special reference to $PGE_2$, an oxygenated fatty acid with a *cis* double bond, a *trans* double bond, two hydroxyl groups, a carbonyl, and a five-membered ring joined to the two side chains at carbons-8 and -12. Prostaglandins $E_3$ and $E_1$ are derived from other eicosanoids (fatty acids with 20 carbon atoms), with $PGE_3$ having an additional *cis* double bond at carbon-17, and $PGE_1$ being devoid of the double bond at carbon-5. However, the $E_2$ prostaglandins, derived from arachidonic acid, are the most abundant. Variations on the ring give $PGD_2$ by interchanging the carbonyl and hydroxyl or $PGF_{2\alpha}$ with a hydroxyl group at carbon-9. Compounds with other rings and similar side chains (Fig. 1) such as the endoperoxides ($PGG_2$ and $PGH_2$), $TXA_2$, and $PGI_2$, are also physiologically active. Strictly speaking, $TXA_2$, devoid of a prostanoic acid skeleton, is not a prostaglandin. These structures are reviewed elsewhere in detail (*9*).

Dr. Kuehl is senior scientist and head, and Dr. Egan is senior research biochemist, in the Department of Biochemistry of Inflammation, Merck Institute for Therapeutic Research, Rahway, New Jersey 07065.

## Biosynthetic Pathways for Arachidonic Acid Oxygenation

*Cyclooxygenase*. The precise mechanism for releasing free fatty acid precursors of prostaglandins has not been elucidated. It is generally agreed, however, that they originate largely from phospholipid reserves in cell membranes. Although phospholipase $A_2$ has been recognized as an important enzyme in the release of these precursor acids, recent studies with platelets implicate a phosphatidyl inositol-specific phospholipase C, yielding diacylglycerides and, subsequently, arachidonic acid (*10*). However, the importance of this new pathway in other cell types has not yet been established. Of the three substrate fatty acids, *cis*-5,8,11,14-eicosatetraenoic acid (arachidonic acid), *cis*-8,11,14-


0036-8075/80/1128-0978$01.25/0 