# *THE RATE OF MOLECULAR EVOLUTION CONSIDERED FROM THE STANDPOINT OF POPULATION GENETICS*

BY MOTOO KIMURA*

DEPARTMENT OF BIOLOGY, PRINCETON UNIVERSITY

*Communicated by James F. Crow, May 26, 1969*

*Abstract.*—The rate of amino acid substitutions in the evolution of homologous proteins is remarkably constant. Furthermore, estimated rates of amino acid substitutions based on comparisons of the alpha hemoglobin chains of various mammals with that of the carp are about the same as those based on comparisons of the carp alpha and mammalian beta or the alpha and beta chains in mammals. These uniformities are regarded as evidence for the hypothesis that a majority of amino acid substitutions that occurred in these proteins are the result of random fixation of selectively neutral or nearly neutral mutations.

Two implications of this possibility are discussed: (*a*) Random gene frequency drift is playing an important role in determining the genetic structure of biological populations and (*b*) genes in "living fossils" may be expected to have undergone as many DNA base (and therefore amino acid) substitutions as corresponding genes (proteins) in more rapidly evolving species.

---

Although since Darwin a great deal of knowledge has accumulated concerning evolution at the phenotypic level, the molecular changes are only beginning to be understood. Recent developments in the comparative study of protein sequence, however, have provided a powerful tool by which evolution at the molecular level may be investigated.

A few years ago, Zuckerkandl and Pauling[1] claimed that the evolutionary rate is approximately constant in most polypeptide chains. In hemoglobins, according to their estimate, about one substitution occurs in every 800 million years per amino acid site. Although the constancy of the molecular evolutionary rate has been accepted in many writings other than Zuckerkandl and Pauling, no systematic check on this thesis has been made taking due account of the statistical variations involved. For many biologists who are accustomed to think of evolution in terms of natural selection, this thesis may not readily be accepted.

Since extensive information is now available concerning amino acid arrangements in various proteins, as admirably compiled by Dayhoff and Eck,[2] I shall first examine in some detail the rate of amino acid change in hemoglobin $\alpha$ and $\beta$ chains. The method of estimation is as follows: Let $n_{aa}$ be the total number of amino acid sites in two polypeptide chains compared with each other (preferably excluding deletions and insertions), and let $d_{aa}$ be the number of sites in which they are different. If we denote by $K_{aa}$ the mean number of substitutions per amino acid site over the whole evolutionary period that separated these two polypeptides, then assuming independence of substitution, we have

$$d_{aa} = n_{aa}\,(1 - e^{-K_{aa}}),$$

so that $K_{aa}$ may be estimated from

$$K_{aa} = -2.30 \log_{10} (1 - p_d), \qquad (1)$$

where $p_d = d_{aa}/n_{aa}$ is the fraction of different sites. An equivalent formula has also been used by Zuckerkandl and Pauling.[1] The standard error of this estimate may be obtained by

$$\sigma_K = \sqrt{\frac{p_d}{(1 - p_d)n_{aa}}}. \quad (2)$$

The rate of substitution per amino acid site *per year* may then be obtained from

$$k_{aa} = K_{aa}/(2T), \quad (3)$$

where $T$ is the number of years that have elapsed since divergence from a common ancestor.

For example, if we compare the hemoglobin $\alpha$ chain of the carp with that of man, excluding insertions or deletions amounting to 3 amino acids, we have $n_{aa} = 140$ and $d_{aa} = 68$. Thus, from formulas (1) and (2), we obtain $K_{aa} = 0.665 \pm 0.082$. From paleontological evidence, we are reasonably sure that the common ancestor of the carp and man lived in the Devonian period, the age of fishes (about 350–400 million years ago). During this period, most of the basic differentiation of fishes occurred, and in its later part crossopterygians gave rise to amphibians (see, for example, refs. 3 and 4). So, we may assume that divergence of lines leading to the carp and man occurred around the middle of Devonian (Fig. 1), and take $2T = 750 \times 10^6$. This gives the rate of substitution $k_{aa} = 8.9 \times 10^{-10}$ per amino acid site per year. Table 1 lists the result of similar calculations in which the $\alpha$ chain of the carp is compared with the human, mouse, rabbit, horse, and bovine $\alpha$ chains. The mammals probably diverged from their common ancestor some 80 million years ago (see Fig. 1).

The table reveals a remarkable fact that these chains have differentiated in relation to carp's $\alpha$ chain roughly to the same extent. Actually, most of the $K_{aa}$ values agree with each other within the limit of statistical error. The standard errors attached to the values of $K_{aa}$ are valid for comparisons of independent estimates, but, when two different $K_{aa}$ values in the table are compared with each other, we must note that they are correlated in the sense that they share a common ancestor about 800 million years ago. Thus, the standard error appropriate to such comparisons is

$$\sigma_K^* = \sigma_K\sqrt{1 - r}, \quad (4)$$

where $r = (2T - T_0)/(2T)$ in which $T_0$ is the number of years after divergence. Then the standard error of the difference is $\sqrt{2}\sigma_K^*$, which is approximately 0.05. The agreement of the $K_{aa}$ values for different mammals as compared with the carp appears to be even more remarkable if we consider the fact that these $\alpha$ chains of mammals also differ from each other roughly in 20 amino acid sites on the average. The results given in Table 1 are combined and summarized in the first row of Table 2 as comparison 1. Assuming $2T = 7.5 \times 10^8$ as before, and, disregarding the error involved in the estimation of the time parameter, we obtain the rate of substitution; $k_{aa} = (8.9 \pm 0.5) \times 10^{-10}$ per amino acid site per year.

Comparisons 2 and 3 in the same table summarize the results of similar analysis with respect to $\alpha$ chains among mammals, taking the human $\alpha$ chain as

the reference point in the former and that of the mouse in the latter. They give $k_{aa} = (8.8 \pm 0.9) \times 10^{-10}$ and $k_{aa} = (10.9 \pm 0.9) \times 10^{-10}$. These values are not only alike but they also agree well with the corresponding value derived above in comparison 1 in which the carp $\alpha$ is compared with various mammalian $\alpha$'s.

Similar calculations for $\beta$ chains are summarized as comparisons 4 and 5 in Table 2.

Much more interesting and significant results are obtained by comparing the $\alpha$ and $\beta$ chains. It is generally accepted that they have originated by gene duplication in the remote past. In Table 3, the $\beta$ chain of man is compared with human, mouse, rabbit, horse, bovine, and carp $\alpha$ chains. In these comparisons, nonmatching parts are excluded between the $\alpha$ and $\beta$ chains due to deletion or insertion that amount to 9 or 10 amino acids. The table shows that each of these $\alpha$ chains had diverged from the human $\beta$ chain almost to the same extent. Actually, their $K_{aa}$ values (expected numbers of substitutions per amino acid site) agree with each other within the limit of statistical error. In Table 2, they are combined to give $K_{aa} = 0.799 \pm 0.038$ as comparison 6. They indicate that the two structural genes corresponding to the $\alpha$ and $\beta$ chains, after their origination by duplication, have diverged from each other independently and to the same extent on whatever evolutionary line they are placed so that the amount of divergence is the same, irrespective of whether the compared $\alpha$ and $\beta$ chains are taken from the same organism (man) or from two different organisms, such as man and carp, which have evolved independently over 350 million years. To my knowledge, the significance of this remarkable fact in relation to the

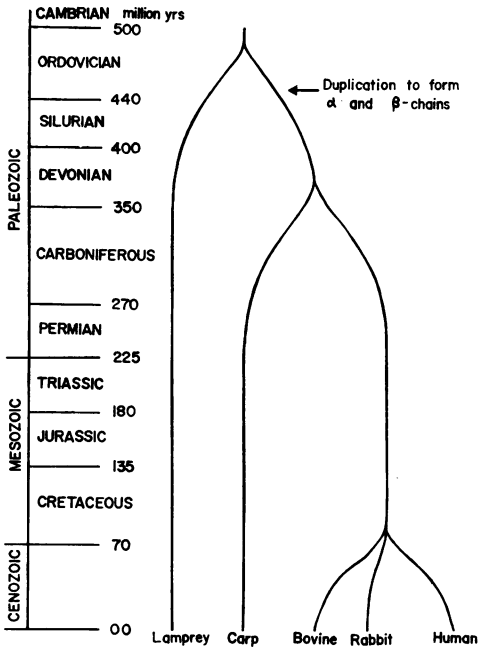


FIG. 1.—A phylogenetic tree of the vertebrate evolution.

TABLE 1. *Comparison of the hemoglobin α chain of carp with those of various mammalian species, showing the number of different amino acid sites.*

| Comparison | $d_{aa}$ | $n_{aa}$ | $K_{aa}$ |
|---|---|---|---|
| Carp α—Human α | 68 | 140 | 0.665 ± 0.082 |
| " —Mouse α | 68 | 140 | 0.665 ± 0.082 |
| " —Rabbit α | 72 | 140 | 0.722 ± 0.087 |
| " —Horse α | 67 | 140 | 0.651 ± 0.081 |
| " —Bovine α | 65 | 140 | 0.624 ± 0.079 |

$K_{aa}$ stands for the mean number of substitutions per amino acid site that separate two chains compared. Note smaller errors appropriate for comparisons between different mammals as described in the text.

TABLE 2. *Summary of the results of comparisons involving hemoglobin α, β, and globin chains in various vertebrate species regarding amino acid substitution in evolution.*

| No. | Comparison | | $K_{aa}$ | $2T \times 10^{-8}$ | $k_{aa} \times 10^{10}$ |
|---|---|---|---|---|---|
| 1 | Carp α vs. | Human, mouse, rabbit, horse and bovine α's | 0.665 ± 0.037 | 7.5 | 8.9 ± 0.5* |
| 2 | Human α vs. | Horse, bovine, pig, and sheep α's | 0.141 ± 0.014 | 1.6 | 8.8 ± 0.9 |
| 3 | Mouse α vs. | Human, horse, bovine, pig, rabbit and sheep α's | 0.175 ± 0.015 | 1.6 | 10.9 ± 0.9 |
| 4 | Human β vs. | Horse, pig, sheep and bovine β's | 0.190 ± 0.016 | 1.6 | 11.9 ± 1.0 |
| 5 | Mouse β vs. | Human, rabbit, horse, pig and bovine β's | 0.225 ± 0.019 | 1.6 | 14.0 ± 1.2 |
| 6 | Human β vs. | Human, mouse, rabbit, horse, bovine and carp α's | 0.799 ± 0.038 | 9.0 | 8.9 ± 0.4 |
| 7 | Rabbit β vs. | Human, mouse, rabbit, horse, bovine and carp α's | 0.829 ± 0.039 | 9.0 | 9.2 ± 0.4 |
| 8 | Human β vs. | Lamprey globin | 1.281 ± 0.135 | 10.0 | 12.8 ± 1.4 |

$T$ stands for the length of time elapsed since the evolutionary divergence of the two chains compared. For details, see text.

* Error due to inaccuracy of time parameter is not included.

mechanism of evolution has never been stressed in print before. The amino acid substitutions have proceeded at the same rate throughout the diverse lines of vertebrate evolution. This supports the hypothesis that the changes are largely fortuitous.

As to the time of gene duplication that formed the α and β chains, it certainly must have occurred before the evolutionary divergence of the human and carp lines, but after the divergence of the human and lamprey lines, since the lamprey globin found in the blood is a monomer. Recently, Ohno *et al.*[5] suggested, based on cytological evidence, that gene duplication started to occur widely at the jawless stage of vertebrate evolution. So, we may assume that the formation of α and β chains occurred about 450 million years ago toward the end of Ordovician period (Fig. 1). Assuming this and taking $2T = 9 \times 10^8$ (years), we obtain $k_{aa} = (8.9 \pm 0.4) \times 10^{-10}$ per amino acid site per year, in good agreement with the previous values.

The same clear indication of constancy in the rate of amino acid substitution in evolution is obtained if we compare the rabbit β chain with human, mouse, rabbit, horse, bovine, and carp's α chains. It gives $k_{aa} = (9.2 \pm 0.4) \times 10^{-10}$ per amino acid site per year (Table 2, comparison 7).

I shall conclude this type of analysis by comparing human β hemoglobin with lamprey globin. Excluding insertions or deletions amounting to 15 amino acids, $n_{aa} = 144$ and $d_{aa} = 104$. This gives $K_{aa} = 1.281 \pm 0.135$ as listed in the last row of Table 2. If we assume that the divergence of lines leading to lamprey

TABLE 3. *Comparison of the β chain of man with the α chains of various vertebrate species.*

| Comparison | $d_{aa}$ | $n_{aa}$ | $K_{aa}$ |
|---|---|---|---|
| Human β—Human α | 75 | 139 | 0.776 ± 0.092 |
| " —Mouse α | 75 | 139 | 0.776 ± 0.092 |
| " —Rabbit α | 79 | 139 | 0.840 ± 0.098 |
| " —Horse α | 77 | 139 | 0.807 ± 0.094 |
| " —Bovine α | 76 | 139 | 0.791 ± 0.093 |
| " —Carp α | 77 | 139 | 0.807 ± 0.094 |

(a jawless fish of today) and man occurred in the earlier part of Ordovician period some 500 million years ago, $2T = 10^9$ (years), we obtain $k_{aa} = (12.8 \pm 1.4) \times 10^{-10}$ per amino acid site per year, giving once again a similar value for the rate of substitution. From the uniformity in the rates of substitution as shown in this table we may take $k_{aa} = 10^{-9}$ (per amino acid site per year) as a representative figure for hemoglobins.

I believe that the present analysis supports the hypothesis that at least in the hemoglobin $\alpha$ and $\beta$ chains, amino acid substitutions and the underlying nucleotide substitutions have proceeded at a constant rate and in a fortuitous manner throughout the diverse lines of vertebrate evolution during the past 500 million years. It is extraordinary that they mainly depend on *time measured in years* but are almost independent of generation time, living conditions, or even the genetic background.

In my previous paper[6] on the rate of molecular evolution, in order to estimate the rate of nucleotide substitution per genome per generation in mammalian species, I computed the average rate of amino acid substitution, using three proteins: hemoglobins, cytochrome *c*, and triosephosphate dehydrogenase. This gave the rate of about one substitution in $28 \times 10^6$ years for a polypeptide chain consisting of 100 amino acids, that is, roughly $\bar{k}_{aa} = 0.4 \times 10^{-9}$ per amino acid site per year. Recently, King and Jukes[7] obtained a more reliable figure, $\bar{k}_{aa} = 1.6 \times 10^{-9}$, by averaging seven proteins. According to them, the rates of substitution differ among various proteins, ranging from $0.33 \times 10^{-9}$ (insulins) to $4.29 \times 10^{-9}$ (fibrinopeptide A).

In this context, it might be convenient to coin a new word, for example, a *pauling*, for a unit of evolutionary rate at the molecular level, defined as the rate of substitution of $10^{-9}$ per amino acid site per year. Accordingly, the hemoglobin rate is nearly one pauling, while the rates for the seven proteins range from 33 centipaulings (insulins) to 4.3 paulings (fibrinopeptide A). This unit is analogous to the *darwin*, the unit of evolutionary rate at the phenotypic level, representing an increase or decrease of quantitative measurement, such as the length of a tooth, at the rate of 1/1000 per 1000 years.[8]

Because of the estimated high rate of nucleotide substitution per genome per generation and from the consideration of the accompanying substitutional load, I concluded[6] that the majority of molecular mutations due to base substitution must be neutral or almost neutral for natural selection. The present analysis on the rate of amino acid substitution in hemoglobins seems to offer additional support for this conclusion. The remarkable constancy per year is most easily understood by assuming that in diverse vertebrate lines the rate of production of neutral mutations per individual per year is constant.

The above inference is based on a simple principle[6] that for neutral mutations the rate of gene substitution in a population is equal to the rate of production of new mutations per gamete, because for such a mutation, the probability of gene fixation is equal to the initial frequency. Also, it was shown by Kimura and Ohta[9] that it takes about $4N_e$ generations for a neutral mutant to reach fixation through random drift in a population of effective size $N_e$, if we exclude the cases in which it is lost from the population by chance.

The above arguments bring us to inquire about the mutation rate of the

hemoglobin genes at the molecular level. In other words, what is the mutation rate per codon per year in hemoglobin $\alpha$ and $\beta$ cistrons? The assumption of neutral substitutions requires that it is at least about $10^{-9}$ per year. From the observed frequencies of rare hemoglobin variants and by statistical treatment based on the theory of the number of generations until extinction of mutant genes,[10] the estimated mutation rate per amino acid site per generation in man turns out to be $u_{aa} = 4.4 \times 10^{-8}$. (Details will be published elsewhere.[11])

Within the past 15 generations or so, average generation time must have been roughly 20 years for man, so that the mutation rate per year per amino acid site amounts approximately to $U_{aa} = 2.2 \times 10^{-9}$. Considering many uncertainties involved in the process of estimation both of the rate of amino acid substitution in evolution and the rate of mutation per amino acid site per year, we may consider that the estimated mutation rate $U_{aa} = 2.2 \times 10^{-9}$ is at least as high as the rate of substitution $k_{aa} = 10^{-9}$. This indicates that the observed mutation rate is sufficiently high to accommodate the assumed neutral mutations.

Other evidence for a considerable degree of randomness in DNA base-pair substitution during evolution is the observation that the amino acid composition of proteins can be predicted from a knowledge of the code and DNA base ratios.[12] This observation also suggests that "genic" and "nongenic" DNA do not differ in composition.

As to the differences in rates of evolution among various proteins such as insulin and fibrinopeptide A, it may be that different fractions of the amino acid substitutions are neutral, as pointed out by King and Jukes.[7] One must also admit the possibility of intrinsic differences in mutation rates. Alpha and beta hemoglobins should be similar in both respects, hence the similar rate of evolution.

In man, the total number of nucleotide pairs making up the haploid chromosome set is estimated to be about $3 \sim 4 \times 10^9$ (cf. refs. 13 and 14). This number is roughly the same among different species of mammals. If we assume that the average rate of amino acid substitution is $\bar{k}_{aa} = 1.6 \times 10^{-9}$ (per amino acid site per year), as estimated by King and Jukes[7] for seven proteins, and note that about 20 per cent of base substitutions are synonymous,[12] that is, code for the same amino acid, then we obtain the figure that in the evolutionary history of mammals nucleotide substitution has proceeded at the rate of about 2 or 2.5 per year. This is some four or five times higher than the corresponding figure obtained in the previous report.[6] In the line leading to man, average generation time was probably longer than 10 years. This gives at least the rate of nucleotide substitution per generation of about 20, making the contrast still larger with Haldane's estimate[15] of 1/300 per generation as the standard rate of gene substitution in evolution. Considering the amount of selective elimination that accompanies the process of gene substitution (i.e., the cost of natural selection or the substitutional load[6]), the most natural interpretation is, I believe, that a majority of molecular mutations due to DNA base substitution are almost neutral in natural selection.

In this connection it is appropriate to mention the remarkable experiment conducted by Cox and Yanofsky[16] using Treffers' mutator gene in *E. coli*. This

gene is known to cause preferentially the transversion from an AT pair to a CG pair in the genome of *E. coli*. According to them, the estimated rate of mutation is $3.5 \times 10^{-6}$ per AT pair per generation. In a strain containing this gene, they observed an 0.2–0.5 per cent increase in the GC content of the DNA after 80 subcultures, which corresponds to 1200–1600 cell generations. On the other hand, the expected increase in GC content based on the above mutation rate and assuming some 50 per cent GC in the genome of *E. coli* turns out to be about 0.21–0.28 per cent. Thus, the agreement between the observed and the expected amount of increase in GC content is satisfactory. It shows that the rate of mutation per genome is approximately equal to the rate of mutant (nucleotide pair) substitution in the population and suggests that a majority of the base substitutions are selectively neutral, though in this case most of the mutants that are destined to reach fixation may have not yet been fixed. This agrees also with their observation that the strain is fully viable after accumulation of more than 7000 base substitutions in the course of the experiment.

Under such a circumstance, Sueoka's[17] "effective base conversion rates" are indeed equal to mutation rates.

It should be noted, however, that it is at present unknown what fraction of the total DNA in the genome of a higher organism is used for protein synthesis. If some (conceivably a large proportion) of the base pairs do not so act, substitutions among these might have even a larger probability of being neutral, and the number of changes would be at least as great.

These observations suggest that, particularly if all the DNA is taken into account, random drift is a much more important factor in evolution than has commonly been believed.

If amino acid changes are often due to chance, then these should be established as frequently in evolutionary conservative species as in those that undergo rapid changes in morphology. While we recognize that a constant morphology does not necessarily reflect a constant internal physiology, it is nonetheless likely that living fossils such as coelacanths, horseshoe crabs, and *Lingula* probably have fewer changes in internal function than more rapidly evolving animals. It would support the hypothesis of this paper if hemoglobins and other proteins show the same rate of amino acid substitution in such living fossils as in rapidly evolving species.

None of the above arguments are intended to imply that natural selection is not important in evolution. What we have postulated is that, surrounding the adaptive changes that occur by selection, there is a great deal of random noise from near-neutral random changes.

I would like to express my thanks to Dr. J. F. Crow for valuable suggestions and help in composing the manuscript. Thanks are also due Drs. E. Dempster and J. L. King for many helpful and stimulating discussions.

* On leave from the National Institute of Genetics, Mishima, Japan. This paper constitutes Contribution no. 720 from the National Institute of Genetics, Mishima, Shizuoka-ken, Japan. Aided in part by a grant-in-aid from the Ministry of Education, Japan. This work was also supported in part by the National Institutes of Health (GM 15422).

[1] Zuckerkandl, E., and L. Pauling, in *Evolving Genes and Proteins*, ed. V. Bryson and H. J. Vogel (New York: Academic Press, 1965), pp. 97–166.

[2] Dayhoff, M. O., and R. V. Eck, *Atlas of Protein Sequence and Structure 1967–68* (Silver Spring: National Biomedical Research Foundation, 1968).
[3] Simpson, G. G., C. S. Pittendrigh, and L. H. Tiffany, *Life: An Introduction to Biology* (London: Routledge and Kegan Paul, 1958).
[4] de Beer, G., *Atlas of Evolution* (London: Thomas Nelson and Sons, 1964).
[5] Ohno, S., U. Wolf, and N. B. Atkin, *Hereditas*, **59**, 169–187 (1968).
[6] Kimura, M., *Nature*, **217**, 624–626 (1968).
[7] King, J. L., and T. H. Jukes, *Science*, **164**, 788–798 (1969).
[8] Haldane, J. B. S., *Evolution*, **3**, 51–56 (1949).
[9] Kimura, M., and T. Ohta, *Genetics*, **61**, 763–771 (1969).
[10] Kimura, M., and T. Ohta, *Genetics*, in press.
[11] Kimura, M., in preparation.
[12] Kimura, M., *Genet. Res.*, **11**, 247–269 (1968).
[13] Muller, H. J., *Bull. Am. Math. Soc.*, **64**, 137–160 (1958).
[14] Vogel, F., *Nature*, **201**, 847 (1964).
[15] Haldane, J. B. S., *J. Genet.*, **55**, 511–524 (1957).
[16] Cox, E. C., and C. Yanofsky, these PROCEEDINGS, **58**, 1895–1902 (1967).
[17] Sueoka, N., these PROCEEDINGS, **48**, 582–592 (1962).