J Mol Evol (1981) 17:110–113

Journal of Molecular Evolution

Letters to the Editor

Was Globin Evolution Very Rapid in its Early Stages?: A Dubious Case Against the Rate-Constancy Hypothesis*

Motoo Kimura

National Institute of Genetics, Mishima, 411 Japan

Summary. Goodman et al's (1975) claim of accelerated evolution in the early stages of globin evolution is based on an erroneous assignment of the time of divergence of vertebrate myoglobin and hemoglobin. When this is corrected, there is no basis for their claim. The data are much more consistent with the nearly constant rate expected on the neutral mutation-random drift hypothesis than with the uneven rates expected if most amino acid changes were caused by substitution of favorable mutants through Darwinian selection. In addition, the majority of the codons determined by their maximum parsimony method have turned out to be wrong when compared to the actual nucleotide sequences of rabbit α and human β hemoglobins determined by direct sequencing.

Key words: Molecular evolutionary clock – Maximum parsimony method – Neutral mutation random drift hypothesis

The near-constancy of amino acid substitution rate for the same protein in different phylogenies over long time periods, which Zuckerkandl and Pauling (1965) referred to as "a molecular evolutionary clock", is one of the most controversial subjects in molecular evolution. Particularly, it has been debated in relation to the date of human-ape divergence (see, Wilson et al. 1977) and to evidence for the neutral theory of molecular evolution (see Kimura 1979).

An especially strong claim of non-constancy has been made by Goodman and his associates in their study of globin evolution (Goodman 1976, 1978; Goodman et al. 1974, 1975). They maintain that mutant substitutions occurred at a very high rate in the early stage of globin evolution, soon after duplication to form myoglobin and α and β hemoglobins, and that this was followed by a markedly reduced rate during the last 300 million years from the ancestral amniote to the present. According to Goodman et al. (1975), from 500 to 400 million years ago the genes descending from "the basal vertebrate ancestor through the hemoglobin-myoglobin and β-α ancestors to the teleost-tetrapod α ancestor evolved at the rate of 109 NR %", where NR % means the number of nucleotide replacements per 100 codons per 10^8 years. This corresponds to $k_{aa} = 10.9 \times 10^{-9}$ per amino acid per year in my terminology (Kimura 1969), and it is some 10 times as high as my estimate of about 10^{-9} per amino acid site per year. On the other hand, for the last 300 million years, their estimated rate is only 15 NR % (i.e. $k_{aa} = 1.5 \times 10^{-9}$) which is not very different from mine. Together with their plausible explanation that when new functions emerged in duplicated globin genes rapid mutant substitutions occurred by *positive* Darwinian selection, Goodman et al.'s (1975) claim of non-constancy has been quoted widely as evidence against the theory (Kimura 1968, King and Jukes 1969) that most molecular evolution occurs by random drift of mutations that are nearly equivalent selectively (the neutral theory, for short) (see, for example, Vogel and Motulsky 1979). Referring to the discrepancy between the evolutionary rates estimated by them and others, Goodman (1976 and elsewhere) writes that both Dickerson (1971) and I (Kimura 1969) failed to account sufficiently for superimposed mutations in our calculations of distances between contemporary globin sequences, and that therefore we more grossly underestimated the evolutionary distances between the more anciently separated globins.

*Contribution No. 1314 from the National Institute of Genetics, Mishima, 411 Japan.

0022–2844/81/0017/0110/$01.00

The statistical method used by Goodman's group is detailed in Goodman et al. (1974). They make use of an extensive series of computer programs, termed MMUTD, UWPGM, PSLNG, PSITR, PSANC etc. to assign mutations to various branches of the phylogenetic tree, to determined ancestral sequences at evolutionary forks, etc. These are all based on the principle of "maximum parsimony". They determine the geological dates of the branch points from paleontological information and from their own opinions. They correct for undetected mutant substitutions by using their "augmentation" method (considered to be particularly important for unpopulated branches in the "parsimony tree").

The main purpose of this note is to point out that the relatively high evolutionary rates obtained by Goodman et al. (1974, 1975) are not the result of their claimed superior statistical method, but rather of wrong assignment of geological dates to duplication events in the early history of globin evolution. I shall also point out a surprising fact; namely, that the great majority of globin codons determined by Goodman et al. (1974) using their maximum parsimony method have turned out to be wrong.

Now, if we examine carefully the data used by Goodman et al. (1974, 1975), we find that their estimate of the high evolutionary rate (109 NR %) comes almost entirely from their assumption that the gene duplication leading to lamprey globin on the one hand and, on the other, to the ancestral globin that produced myoglobin and the α and β globins, occurred 500 million years ago (point *a* on the phylogenetic tree; see Fig. 1 of Goodman et al. 1975). To facilitate discussion, I present the main duplication events postulated by them as Figure 1 in this article. Goodman et al. further assume that the gene duplication leading to myoglobin vs. β-α hemoglobin divergence occurred about 470 million years ago (point *b*). In addition, they assume that the duplication which produced the α and β hemoglobins occurred about 425 million years ago (point *c*) and that the ancestor of carp α and chicken α (equivalently, carp α and human α) diverged 400 million years ago (point *d*). They allocate 50 mutations between point *a* and *b*, 58 between *b* and *c*, and 51 between *c* and *d*. In other words, for the time span of 100 million years (i.e. between points *a* and *d*), they assume the establishment of two gene duplications and allocate a total of 159 mutations (among about 145 amino acid sites).

I believe that their positioning of gene-duplication points in time contains some errors, of which the most serious is their assumption that the duplication responsible for the myoglobin vs. hemoglobin divergence occurred about 470 million years ago (point *b*), after the ancestor of jawed fishes diverged from the ancestor of the lamprey. In other words, they assume that the history of myoglobin is younger than that of the jawless fish (agnatha).

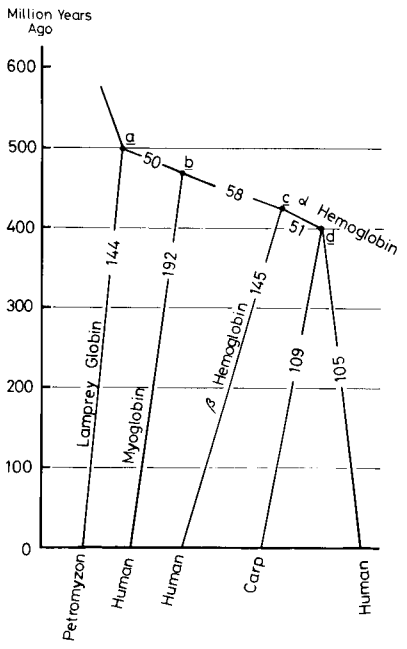

Fig. 1. A phylogenetic tree to show the main assumptions made by Goodman et al. (1975) on the duplication events in the course of globin evolution of vertebrates. The number in each branch of the tree is the number of mutant substitutions allocated by them

It is much more likely that the gene duplication that led to the lamprey (*Petromyzon marinus*) globin whose sequence was used in their analysis and the ancestral molecule from which myoglobin and α-β globins have descended is much older than they postulate. It is likely that the duplicated globin genes already existed in the ancestral agnatha when they appeared about 500 million years ago (point *a*). A similar point has been made by Zuckerkandl (1976; see page 300 of his paper). According to Hendrickson and Love (1971) the lamprey globin that has been sequenced is only one of six hemoglobins present in *P. marinus* blood; it is hemoglobin V, the most prevalent of the six. This suggests that gene duplication occurs rather frequently in evolution. Much more important is the recent finding by Romero-Herrera et al. (1979) that true myoglobin exists in the red muscle of the lamprey heart. Although the amino acid sequence is yet to be determined, this finding completely invalidates Goodman et al.'s assumption that myoglobin originated only some 470 million years ago, *after* the jawed-fish diverged from the agnatha, and that therefore myoglobin is newer than Hb V of the lamprey. In this respect, Hunt et al. (1978) were more judicious in placing the lamprey globin branch on the opposite side of the phylogenetic tree from the other vertebrate globins, stating that the gene duplication responsible for myoglobin and lamprey globin must have occurred at a very ancient time. In Hunt et al. 's figure (see Dayhoff 1978, p 229) this duplication point is depicted as being one billion or so years old. I would like to emphasize that if the gene duplication that led to myo-

globin and lamprey Hb V occurred earlier than postulated by Goodman et al., for example, late in the Pre-Cambrian, some 800 million years ago, evidence for a very high evolutionary rate that Goodman et al. find breaks down completely. Hunt et al. also place the divergence point of α and β globins approximately 500 million years ago, which is a much more appropriate time than that of Goodman et al. who place it about 425 million years ago.

Since Goodman (1978) and Goodman et al. (1974, 1975) have criticized my claim of the constancy of the evolutionary rate by saying that I failed to consider multiple mutations in my estimation procedure (that uses Poisson correction for undetected mutant substitutions), it is appropriate to examine this aspect of their work. Their method consists of first estimating the number of mutant substitutions using the principle of "maximum parsimony" and then modifying the estimate, sometimes by an enormous amount, using their "augmentation procedure". Take, for example, the evolutionary change of α hemoglobin along the two lineages, one leading to the carp and another leading to man, from their common ancestor some 400 million years back. According to Goodman et al. (1974) the number of mutations accumulated along the lineage leading to the carp amounts to 64, while the corresponding number accumulated along the lineage to man is 56, both being estimated by the maximum parsimony method as "A-solution" values (see their Fig. 2). These two numbers certainly do not show that the intrinsic rates of mutant substitutions are significantly different in these two lineages. Then, they augment these values by adding 48 mutations to the lineage leading to the carp and 33 mutations to that leading to man. So the final estimates of mutant substitutions turn out to be 112 and 89 respectively for these two lineages (see Fig. 6 of Goodman et al. 1974). Although no statistical errors are available for these estimates, presumably they are very large, and the difference between these two values can not be statistically significant (not significant if the distribution is Poisson). This can be seen clearly by the fact that in another paper published the next year (Goodman et al. 1975), they allocate 109 and 105 mutations to these two lineages (see their Fig. 1 and also Fig. 1 of this article) instead of 112 and 89 as mentioned above. That their augmentation procedure itself contains serious statistical problems has been pointed out by Tateno and Nei (1978) and also by Fitch (1980). In addition, it is likely that the maximum parsimony method they used is inherently error-ridden, for there is no guarantee that evolution proceeds through the shortest path. In fact one might argue that a simple Poisson correction more closely reflects the true randomness of the natural process than a principle requiring that nature somehow choose the shortest path.

According to Goodman et al. (1974), their maximum parsimony method can often discriminate between different codons for the same amino acid and this discriminating power increases its utility in reconstructing phylogeny (see page 4 of Goodman et al. 1974). They have published extensive globin codon sequences (in terms of RNA code) which they determined by this method. So, it is of interest to check if their determinations are valid now that the actual nucleotide sequence has become available for a few globin genes, thanks to the new technology of amplifying gene copies in bacterial plasmids followed by rapidly sequencing of DNA by the Maxam-Gilbert method. Recently, the sequence of rabbit α-globin messenger RNA (mRNA) has been determined by Heindell et al. (1978) using these procedures. This allows us to examine the corresponding maximum parsimony codon sequence determined by Goodman et al. (1974).

Table 1. Comparison between the codons determined by Goodman et al. (1974) using the "maximum parsimony" procedure for rabbit α globin and the corresponding codons determined by Heindell et al. (1978) using the Maxam-Gilbert method. Wrong bases assigned by Goodman et al. are underlined

	4	5	8	10	12	13	15	17	19	21
Amino Acid	Pro	Ala	Thr	Ilu	Thr	Ala	Glu	Ilu	Ser	Gly
True Codon	CCC	GCU	ACC	AUC	ACU	GCC	GAA	AUC	AGC	GGU
Parsimony Codon	CCU	GCG	ACU	AUU	ACU	GCU	GAG	AUU	AGU	GGA

22	23	24	26	29	30	32	34	35	37	48	49
Gly	Glu	Tyr	Ala	Val	Glu	Met	Leu	Gly	Pro	Phe	Thr
GGC	GAG	UAU	GCC	GUG	GAG	AUG	UUG	GGC	CCC	UUC	ACC
GGU	GAG	UAU	GCG	GUG	GAA	AUG	CUA	GGU	CCU	UUU	ACU

50	53	55	57	60	63	64	65	67	68	70	71
His	Glu	Ilu	Ala	Lys	Ser	Glu	Ala	Thr	Lys	Val	Gly
CAC	GAG	AUC	GCC	AAG	UCC	GAA	GCC	ACC	AAG	GUG	GGC
CAU	GAG	AUU	GCG	AAA	UCU	CAG	GCU	ACU	AAG	GUU	GGG

73	76	78	82	85	89	104	107	111	113	115	116
Leu	Leu	Gly	Thr	Asp	His	Cys	Val	Asn	His	Ser	Glu
CUG	CUG	GGC	ACU	GAC	CAC	UGC	GUG	AAC	CAC	AGU	GAA
CUA	CUU	GGU	ACG	GAU	CAU	UGU	GUU	AAU	GUU	UCU	GAG

124	129	130	131	137
Ser	Leu	Ala	Asn	Thr
UCC	CUG	GCC	AAC	ACC
UCU	CUU	GCG	AAU	ACU

It turns out that, of 51 codons predicted by Goodman et al. for rabbit α-globin (see Table 4 of their 1974 paper), 44 are wrong and only 7 codons are correct. This is shown in Table 1, where mistakenly assigned bases are underlined. This means that only 14% of the maximum parsimony codons are valid. This is even more surprising because most of the indeterminancy is at the third position of the codon and even random assignment of one of four nucleotide bases to this position can achieve 25% agreement on the average. A similar examination can be made for the maximum parsimony β-globin codon sequence of man (see Table 5 of Goodman et al. 1974), since the entire sequence of human β-globin mRNA has been determined by Marotta et al. (1977). In this case, it turns out that about half of the maximum parsimony codons are wrong, the worst error being the assignment of AGU for the serine at amino acid position 44 where the actual codon is UCC; all three letters are wrong.

All these results show clearly that Goodman et al.'s (1974, 1975) work is fraught with errors and uncertainty, and contrary to their claim, there is no clear-cut evidence that evolutionary amino acid substitutions were very rapid in the early stages of vertebrate globin evolution. Therefore, their claim that there was a period of fast evolution due to positive Darwinian evolution, although plausible, has no support from the data. I believe that it is much more likely, as we have pointed out before (Kimura and Ohta 1974) that, if high rates occur, they are caused by the removal of a preexisting functional constraint, allowing previously harmful mutants to become selectively neutral.

From the standpoint of the neutral theory I should like to suggest that whenever an exceptionally high rate is encountered in molecular evolution, we should suspect loss of constraint rather than acquisition of a new function.

This does not mean of course that no adaptive mutant substitutions have occurred in the course of globin evolution. But we should be careful not to adopt uncritically the facile explanation that each case is adaptive without examining the evidence for it. In my opinion, such definitely adaptive substitutions are much less frequent than selectively neutral or nearly neutral substitutions caused by random frequency drift.

Acknowledgement. I thank Dr. J.F. Crow who read the first draft and made many useful suggestions for improving the presentation.

References

Dayhoff MO (1978) Atlas of protein sequence and structure, vol 5, suppl 3. National Biomedical Research Foundation, Washington, DC

Dickerson RE (1971) J Mol Evol 1:26–45

Fitch WM (1980) J Mol Evol 16:153–209

Goodman M (1976) Protein sequences in phylogeny. In: Ayala FJ (ed) Molecular evolution. Sinauer, Sunderland, MA, p 141

Goodman M (1978) Substitutional trends and non-random changes in rates during protein evolution. In: Matsubara H, Yamanaka T (eds) Evolution of protein molecules. Tokyo, Japan Scientific Societies Press, p 17

Goodman M, Moore GW, Barnabas J, Matsuda G (1974) J Mol Evol 3:1–48

Goodman M, Moore GW, Matsuda G (1975) Nature 253:603–608

Heindell HC, Liu A, Paddock GV, Studnicka GM, Salser WA (1978) Cell 15:43–54

Hendrickson WA, Love WE (1971) Nature New Biol 232:197–203

Hunt LT, Hurst-Calderone S, Dayhoff MO (1978) In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 3. National Biomedical Research Foundation, Washington DC, p 229

Kimura M (1968) Nature 217:624–626

Kimura M (1969) Proc Natl Acad Sci USA 63:1181–1188

Kimura M (1979) Scientific American 241 (No 5, Nov): 94–104

Kimura M, Ohta T (1974) Proc Natl Acad Sci USA 71:2848–2852

King JL, Jukes TH (1969) Science 164:788–798

Marotta CA, Wilson JT, Forget BG, Weissman SM (1977) J Biol Chem 252:5040–5053

Romero-Herrera AE, Lieska N, Nasser S (1979) J Mol Evol 14: 259–266

Tateno Y, Nei M (1978) J Mol Evol 11:67–73

Vogel F, Motulsky AG (1979) Human genetics. Springer, Berlin Heidelberg New York

Wilson AC, Carlson SS, White TJ (1977) Ann Rev Biochem 46: 573–639

Zuckerkandl E (1976) J Mol Evol 7:269–311

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

Received April 18, 1980