J Mol Evol (1981) 17:114–120

Journal of Molecular Evolution

Globin Evolution was Apparently Very Rapid in Early Vertebrates: A Reasonable Case Against the Rate-Constancy Hypothesis

Morris Goodman

Department of Anatomy, Wayne State University School of Medicine, Detroit, Michigan 48201, USA

Abstract. Kimura mistook ambiguous maximum parsimony codons for wrong codons. The maximum parsimony method performed well as judged by the two classes of serine codons (which can not be connected by silent mutations) on comparing the parsimony codons for serines in human, rabbit, and mouse α hemoglobin chains to actual codons determined by nucleotide sequencing. In genealogical reconstructions involving 247 eucaryotic globins, the maximum parsimony distances separating the contemporary sequences show that Kimura's Poisson and Dayhoff's PAM estimates of rate of globin evolution miss most of the superimposed replacements and are therefore seriously in error. Nor is Kimura's constant rate assumption and his belief in a single origin of myoglobin supported. Lamprey myoglobin appears to be most like lamprey hemoglobin, while gnathostome myoglobin seems closest to gnathostome hemoglobin. It was found that the three types of gnathostome globins (Mb, α Hb, β Hb) evolved between the shark-boney vertebrate and bird-mammal ancestors at a much faster rate than from the latter ancestor to the present. The data indicate that rates were exceedingly fast during the origin of these globin chains because a high proportion of substitutions were adaptive. It was concluded that wherever strong stabilizing selection acts on a protein, somewhere in the past positive Darwinian selection must have spread the amino acid substitutions now being preserved.

Key words: Maximum parsimony method – Globin evolution – Superimposed substitutions – Accelerated evolutionary rates – Positive Darwinian selection – Decelerated rates – Stabilizing selection

In an earlier study of globin evolution by the maximum parsimony method (Goodman et al. 1974, 1975) my coworkers and I provided evidence that the values for rates of globin evolution proposed by Zuckerkandl and Pauling (1965), Kimura (1969), and Dickerson (1971), all using the same Poisson correction for hidden substitutions, were gross underestimates of the true values. Our findings of large variations in rates and of substantial numbers of amino acid substitutions that appeared to be adaptive could not readily be explained by Kimura's constant rate-neutral mutation theory. Instead our findings pointed to Darwinian evolution of vertebrate hemoglobin. From the genealogy reconstructed for 55 contemporary globin sequences, it was deduced that positive selection for differentiation of globin molecules had greatly accelerated evolutionary rates in early vertebrates, especially at molecular sites acquiring cooperative functions, and that stabilizing selection then preserved the functional improvements, markedly decelerating rates in the amniote lineages to birds and mammals (Goodman et al. 1975). It also seemed apparent that during the adaptive radiation from which the placental mammalian orders emerged evolutionary rates accelerated on the stem to the Primates and later drastically slowed in groups such as hominoids. In the present note in reply to Kimura (1981) who disputes these findings, I reaffirm their validity.

Kimura (1981) disputes the finding of accelerated globin evolution in early vertebrates, firstly, by arguing that it was based on an erroneous assignment of the time of divergence of vertebrate myoglobin and hemoglobin and, secondly, by taking the discovery of lamprey myoglobin (Romero-Herrera et al. 1979) to mean that the ancestral split between myoglobin and hemoglobin in the gnathostomes (jawed vertebrates) must have preceded the agnathan (lamprey)-gnathostome divergence (i.e. arose from a prior globin gene duplication in a pre-vertebrate stock ancestral to invertebrates and vertebrates). Kimura further argues that the Goodman

0022–2844/81/0017/0114/$01.40

et al. augmentation procedure overestimated the hidden substitutions and that their maximum parsimony method produced fallacious codons. Although disputing the Goodman et al. (1974, 1975) evidence that natural selection caused globin evolution to be very rapid in its early vertebrate stages, Kimura (1981) at least considers the idea plausible but prefers loss of constraint rather than acquisition of a new function when an exceptionally high rate is encountered in molecular evolution.

In countering Kimura's points, I take advantage of a huge increase in the density of globin sequence data over the past five years and of improvements in the maximum parsimony approach used in my laboratory (Goodman et al. 1979 a; 1979 b; Baba et al. 1981). Genealogical reconstructions have now been carried out involving 247 eucaryotic globin sequences (Goodman and Czelusniak 1980; Goodman 1980). These reconstructions show that the observation of very rapid globin evolution in early vertebrates does not depend on the times chosen for ancestral divergence either between myoglobin and hemoglobin of jawed vertebrates or between Agnatha and Gnathostomata, nor does it depend on using augmented maximum parsimony distances. The unaugmented maximum parsimony distances serve by themselves. Indeed they show that not only Kimura's K_{aa} values, but also Dayhoff's PAM values grossly underestimate the evolutionary change between globin sequences (e.g. in the divergence of mammalian orders, or over the much longer period of time separating α and β hemoglobins). Due to the increased density of sequence data a larger proportion of previously hidden substitutions are now detected by the maximum parsimony method. In this connection shark myoglobin (Fisher and Thompson 1979), shark α hemoglobin (Nash et al. 1976), and shark β hemoglobin (Fisher et al. 1977) are included in the present body of globin amino acid sequence data. This has allowed the rates of evolution of the three types of globin genes to be estimated for one of the earlier vertebrate periods, that between the gnathostome (shark-boney vertebrate) and amniote (bird-mammal) ancestors, i.e. roughly between $\bar{c}$425 MyBP (million years before the present) and $\bar{c}$300 MyBP which are paleontological derived dates generally used by molecular evolutionists for these branch points (e.g. Fig. 6.1 in Dayhoff 1972 and Table 1 in Dayhoff 1978). Even with the unaugmented maximum parsimony distances the rates for this period come out to be about three times faster than the subsequent rates from the amniote ancestor ($\bar{c}$300 MyBP) to the present. Furthermore the new evidence still indicates that myoglobin and hemoglobin arose, independently, in more than one phylogenetic branch. This was previously apparent for mollusc myoglobin versus vertebrate myoglobins and hemoglobins (Goodman et al. 1975) and can now be deduced for lamprey hemoglobins and myoglobin (they appear to be much closer to one another than to other globins). Moreover, the most parsimonious genealogical tree still separates all vertebrate globins from invertebrate ones and has the gnathostome myoglobin branch closer to the gnathostome hemoglobin branch than to Agnathan globins. Thus a reasonable case can be made that the very rapid globin evolution observed in the gnathostome stem resulted from gene duplications and natural selection acting on duplicated genes to differentiate gnathostome myoglobin and hemoglobin from one another and then α and β chains of hemoglobin. In building this case, however, I shall first deal with Kimura's claim that maximum parsimony codons are wrong.

Most of the parsimony codons presented in Tables 1–5 of Goodman et al. (1974) are simply ambiguous rather than wrong.[1] This is apparent from the maximum parsimony algorithm's synonomy deletion rules described in Moore et al. (1973) and stated again in the footnote to Table 1 of Goodman et al. (1974). Maximum parsimony, when applied only to amino acid sequences, is not intended to reveal for the sets of synonymous codons the exact member of each set that was actually translated (amino acid sequences alone usually do not contain enough information for this purpose). The algorithm, however, does do what it is intended to ... it calculates for any given tree the absolute minimum number of nucleotide replacements (NRs), and nothing less than that number, needed to derive the contemporary amino acid sequences from their common ancestor. To count this number of NRs (which can be more than the minimum number of amino acid exchanges counted in Dayhoff's parsimony method) each contemporary amino acid residue must be mapped into corresponding codons, but to save computing time many synonymous codons are deleted without jeopardizing the count. For example, in all cases, a U ending codon means at least U/C (C ending codons are arbitrarily deleted by the algorithm) and in many cases U/C/A/G. Similarly, when methionine and tryptophane are absent, a G ending codon means A/G and in many cases U/C/A/G. The aim of Tables 1–5 of Goodman et al. (1974) was to show representative parsimony codons that allowed the minimum number of NRs to be counted in various regions of the reconstructed glo-

[1] Aside from not taking account of the ambiguous nature of the parsimony codons, Kimura listed in his Table 1 two positions where the parsimony codons represented different amino acids than the sequenced or true codons in the rabbit α hemoglobin chain. Goodman et al. (1974) used the von Ehrenstein (1966) amino acid sequence, which is also given in Dayhoff (1972). It has glutamine at position 64 and valine at position 113 in contrast to Heindell et al.'s (1978) glutamic acid (GAA) and histidine (CAC). The maximum parsimony method should not be faulted if it is dealing with a reported sequence that has amino acids that do not appear to be coded for in the sequenced mRNA. Either there were mistakes in one or the other reported data or the sequence variation at the two positions resulted from allelic variation or from divergence of duplicated α genes

bin genealogy. The paper should have left no doubt about the ambiguous nature of such codons. In retrospect, I regret that this may not have been made completely clear for all readers.

The maximum parsimony method used on sequence data is a rigorous way of applying Occam's razor to test genealogical hypotheses for which the data provide empirical evidence. In principle, the genealogical branching patterns and parsimony codons found by the method should become more accurate as the body of sequence data analyzed becomes denser. In this connection, Holmquist (1980) – working with a set of simulated sequences produced by the Poisson model – has provided evidence that maximum parsimony codons are reasonably accurate at nodes in the denser part of the tree but are fallacious in sparse parts of the tree where the simulated sequences diverge so much that the evidence of their evolutionary relationship is lost. With the 247 eucaryotic globin sequences that I am working with significant evolutionary homology is always apparent even among the most anciently separated ones. Now that the actual nucleotide sequences are being detetmined for a growing number of globin genes, it would be valuable to redesign the maximum parsimony algorithm in order to allow both gene nucleotide and protein amino acid sequences to be included as OTUs (operational taxonomic units) in the same data set. With respect to globins, this infusion of actual nucleotide sequence data into the large existing body of amino acid sequence data should eliminate much of the ambiguity of the reconstructed parsimony codons as well as improve their accuracy. An advantage of such an algorithm is that the accuracy of the maximum parsimony reconstruction could be assessed by deliberately replacing the nucleotide sequence by its translated amino acid sequence for an OTU in the dense part of the tree and alternatively in sparser parts of the tree.

The comparison shown in Table 1 of parsimony codons with the actual nucleotide codons of the serines in α hemoglobin chains of man, rabbit, and mouse illustrates about the best that can be expected when the present maximum parsimony algorithm is applied to amino acid sequences. Because silent mutations (from one synonymous codon to another) occur rather frequently, in contrast to substitutions that change the encoded amino acid, a parsimony codon has a fair chance of not representing the real synonymous codon, e.g. a real codon for glycine...GGC...misrepresented as GGR (R = A or G) because glutamic acid (GAR) occurred in neighboring OTUs. The six serine codons, however, can be divided into two classes which can *not* be interconnected by silent mutations alone. To go from a AG beginning serine to a UC beginning serine by the shortest route, a mutation either to a cystine codon (such as UGU or UGC) or a threonine codon (such as ACU or ACC) has to occur. Thus the two classes of serine codons offer a good model for comparing parsimony to the actual codons determined by nucleotide sequencing.

The parsimony codons shown in Table 1 were derived from the parsimonious genealogical tree constructed for

Table 1. Positions where serine occurs in one or more of these α hemoglobin sequences

		3	8	19	35	49	52[a]	63	68	81[a]	84[b]
Human α	True codon	UCU			UCC	AGC	UCU			UCC	AGC
	Parsimony codon	UCX			UCX	AGY	UCR			UCX	*UC*X
Rabbit α	True codon	UCU		AGC			UCU	UCC		UCU	AGC
	Parsimony codon	UCX		AGY			UCR	UCH		UCY	*UC*X
Mouse α	True codon	UCU	AGC		AGC	AGC	UCU		AGU	UCU	AGC
	Parsimony codon	UCX	AGY		*UC*X	AGY	UCR		AGY	UCY	*UC*X
		102	111	115	124	131	133	138[a]			
Human α	True codon	AGC			UCC	UCU	AGC	UCC			
	Parsimony codon	AGY			UCR	UCY	AGY	UCR			
Rabbit α	True codon	UCC		AGU	UCC		AGC	UCC			
	Parsimony codon	*AGY*		*UCY*	UCR		AGY	UCR			
Mouse α	True codon	AGC	AGC		UCU	UCU	AGC	UCC			
	Parsimony codon	AGY	AGY		UCR	UCU	AGY	UCR			

[a]All mammalian α hemoglobin amino acid sequences have only serine at positions 52, 81, and 138

[b]All α hemoglobin amino acid sequences, except that of shark, have only serine at position 84

Y = U or C; R = A or G; H = U or C or A; X = U or C or A or G

The nucleotide sequence for human α hemoglobin chain is that reported by Wilson et al. (1980), for rabbit by Heindell et al. (1978), and for mouse by Nishioka and Leder (1979)

the amino acid sequences of 77 gnathostome α hemoglobin chains. Since the maximum parsimony algorithm employed did not determine the alternatives that could serve for the U, A, and G ending codons produced by the synonymy deletion rules, they were determined by inspection of the distribution of amino acids in neighboring branches. The symbols used in Table 1 to depict this nucleotide ambiguity are Y = U or C, R = A or G, H = U or C or A, and X = U or C or A or G. The parsimony codons are divided into the two serine classes (UC- and AG-), and those in the wrong class when compared to the sequenced codon are italicized. The human nucleotide sequence is that reported by Wilson et al. (1980), the rabbit by Heindell et al. (1978), and the mouse by Nishioka and Leder (1979). There are 17 alignment positions in which serine occurs lat least once in these three mammalian α hemoglobin chains. The parsimony codons are in the correct class in 10 of the 11 human serine positions, 8 of the 11 rabbit, and 12 of the 14 mouse, or, altogether the correct choice of parsimony codons was made 14.17 times at the 17 alignment positions on equally weighting each of these 17 positions. This is good accuracy especially when reasons for the errors are considered. For example at position 84, the shark has alanine (GCX) whereas all the other gnathostomes have serine. Thus although the maximum parsimony solution for this serine must be UCX, it is evident that more than enough evolutionary time had elapsed for UCY and AGY to be equally probable. Similarly at positions 52, 81, and 138 all the mammals (including the anciently separated monotremes) have serine. The parsimony codons chosen belong to the correct class, but in 2 of the 3 positions failed to detect silent mutations (UCR compared to UCU at position 52 and UCR compared to UCC at position 138). The results shown in Table 1 support my conviction that for genealogical reconstructions from sequence data the parsimony method is the rational way to make use of the available information and that the reconstructions should become more accurate as the sequence data improve in quality and increase in density.

The maximum parsimony results show that the Poisson model which Kimura uses to estimate evolutionary change in protein sequences is seriously in error. Dayhoff's PAM values also appear to be substantial underestimates although not as low as Kimura's. The maximum parsimony algorithm when applied solely to amino acid sequences can indeed detect silent mutations in codons but only those for which there is evidence from the configuration of amino acid changes in surrounding regions of the tree. Thus the maximum parsimony NR distances can validly be compared to Kimura's K_{aa} values and Dayhoff's PAM values because the K_{aa} and PAM values are based on models which are supposed to include in the estimates of divergence the superimposed substitutions associated with the observed amino acid differences. Even the unaugmented NR distances can be much larger than the K_{aa} and PAM values. To support this claim let me give examples of results obtained from genealogical trees constructed by the maximum parsimony method for several of the data sets of globin sequences (in particular the 77 α's, 95 β's, and 159 eucaryotic globins including many of the gnathostome sequences).

Typical results are afforded by the evolutionary divergence between human α and β hemoglobin chains. Over their 139 shared amino acid positions, there are 75 amino acid differences. The mean number of substitutions per amino acid site over the whole evolutionary period that separated these two polypeptides, according to the Poisson model used by Kimura (1969), i.e. the K_{aa} value, comes out to 0.776. The PAM model (accepted point mutations per 100 amino acid residue

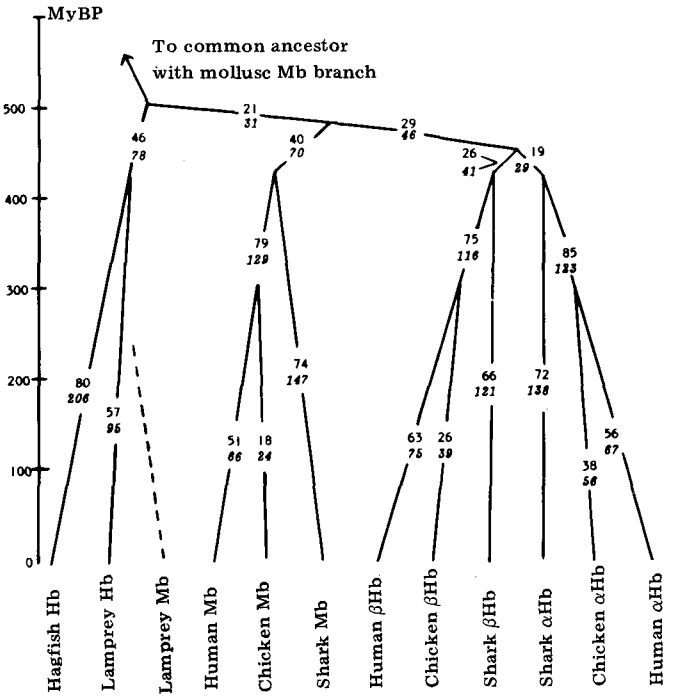

Fig. 1. Some of the vertebrate lines of descent from the genealogical tree constructed by the maximum parsimony method for 159 eucaryotic globin sequences. Numbers of nucleotide replacements (NRs) estimated to have occurred during descent are shown on the branches. Top numbers are the original or unaugmented NR values found by the maximum parsimony algorithm. Bottom italicized numbers are the augmented NR values found by an algorithm which uses the empirical distribution of mutations on the tree to correct for superimposed replacements. The dashed line depicts from present evidence the evolutionary position of the lamprey myoglobin discovered by Romero-Herrera et al. (1979). Mb = myoglobin; Hb = hemoglobin; MyBP = millions of years before the present as inferred from paleontological evidence on the branch times of the species lineages depicted in the tree. The nodes for the Mb-Hb and αHb-βHb gene duplications were placed on this time scale by equating the NR length between the vertebrate and gnathostome Hb ancestors (averaged for α and β) to 75 million years (500 MyBP–425 MyBP) and then subdividing this time by proportionality first to the NR length between the vertebrate and gnathostome Mb-Hb ancestors and next between the latter and the αHb-βHb ancestor

positions) of Dayhoff (1972) yields the distance of 98 or 0.98 mutations per position. In contrast, in my parsimonious genealogical tree for the 159 globins, representative branches of which are shown in Fig. 1, the unaugmented parsimony distance between human α and β hemoglobin chains comes out to 2.26 NRs per amino acid position, and the augmented parsimony distance[2] comes out to 3.15 NRs per position. Thus the unaugmented parsimony distance is roughly 3 times larger than Kimura's K_{aa} distance and at least twice as large as the PAM distance, with the augmented parsimony distance being roughly 4 and 3 times larger respectively. Indeed the full number of superimposed substitutions that are reflected in amino acid exchanges is probably larger than the augmented parsimony distance. This is suggested by the results of simulation experiments (Czelusniak et al. 1978). They show that the augmented maximum parsimony distances, rather than being overestimates, tend to be underestimates of the true distances.

The rates of sequence evolution of hemoglobin chains calculated by Kimura and Dayhoff are derived largely from comparisons between eutherian mammalian orders. Thus Kimura's value of 10^{-9} per amino acid site per year or 10 fixed mutations per 100 amino acid sites per 10^8 years and Dayhoff's (1978) value of 12 PAMs per 10^8 years can be compared to my NR% values (nucleotide replacements per 100 codons per 10^8 years) for the period of eutherian evolution (when I take the date for the most ancestral eutherian node in the globin genealogical trees as 90 MyBP). From the unaugmented maximum parsimony distances the eutherian evolutionary rate for α hemoglobin comes out to 23 NR% and for β hemoglobin 25 NR%; while from the augmented maximum parsimony distances these rates are 29 and 30 NR% respectively. About the same NR% values are obtained for a range of alternative parsimonious or nearly parsimonious trees all of which depict genealogical arrangements in accord with the evidence on eutherian phylogeny. Thus, it is again apparent that the K_{aa}-Poisson and PAM models tend to grossly underestimate the true amounts of globin evolution.

Nor is Kimura's constant rate assumption supported by the rates calculated from the genealogical reconstructions. The average rate from the amniote (bird-mammal) ancestor (at $\bar{c}$300 MyBP) to the present is about one half the eutherian rate and is especially slow on the path from the amniote to the eutherian ancestor. In contrast the earlier vertebrate rate from the gnathostome to the amniote ancestor is about three to four times the amniote rate. An example of this fast-slow evolution is provided by the rate for the 125 or so million years from the gnathostome ancestor to the amniote ancestor compared to the rate for the 300 million years from the amniote ancestor to the present using the NR values in Fig. 1 on the α hemoglobin, β hemoglobin, and myoglobin lineages to man. From the unaugmented NR values the earlier fast/later slow rates are 48 NR%/13 NR% for α hemoglobin, 41 NR%/14 NR% for β hemoglobin, and 41 NR%/11 NR% for myoglobin; from augmented NR values, 70 NR%/16 NR%, 64 NR%/17 NR%, and 67 NR%/14 NR% respectively[3].

Several facts indicate that the rate of globin evolution was even faster in the earliest vertebrates on the stem to the gnathostomes. Firstly, the data on lamprey myoglobin (Romero-Herrera et al. 1979) provide *no* evidence that a gene duplication in a pre-vertebrate stock produced the ancestral separation between the myoglobin and hemoglobin genes found in vertebrates. Rather this data provide evidence that lamprey myoglobin is much closer to lamprey hemoglobin than to any non-lamprey globins. Dr. A.E. Romero-Herrera has calculated from the tentative tryptic peptide alignment of lamprey myoglobin that it shows 60% similarity (amino acid homology) with lamprey hemoglobin and 33% similarity with hagfish hemoglobin. In contrast, each of the two sequenced lamprey hemoglobins (Braunitzer and Fujuki 1969; Li and Riggs 1970) while showing 43% similaritiy with hagfish hemoglobin (Liljeqvist et al. 1979) shows only 25 to 26% similarity with gnathostome myoglobins and hemoglobins. These comparisons certainly suggest that a gene duplication within the lamprey lineage after divergence of the hagfish lineage produced the gene for lamprey myoglobin, but of coursse, for a rigorous analysis of whether or not

[2]The augmentation procedure uses only the empirical distribution of mutations over the maximum parsimony tree to provide corrections for hidden superimposed mutations. It propagates mutational information from the paths in the tree that are denser in intervening ancestors to those that are sparser. The algorithm currently employed, written by J. Czelusniak, was first introduced in Goodman et al. (1979 b) and is described in detail in Baba et al. (1981). This algorithm differs from the earlier, most populous path (MPP) algorithm (Goodman et al. 1974; Moore 1977) by averaging the results of many paths rather than focusing on a single most populous path, in calculating each augmentation increment. Although the current average algorithm does not correct in computer simulation experiments for missing mutations on the larger links as well as the MPP algorithm, it yields augmentation values which increase at a smoother rate and, when faced with changes in the branching arrangement of a tree, remain stable

[3]Kimura (1979) has about 160 million years separating the gnathostome and amniote ancestors by placing the former (shark-boney vertebrate split) at about 440 MyBP and the latter (bird-mammal split) at 280 MyBP. With this time scale the earlier rate is still about two to three times faster than the latter rate.

It should also be noted that Kimura (1969, 1979) assumed that the gene duplication producing separate α and β hemoglobin genes occurred about 450 MyBP. Yet now after p 229 of Dayhoff's *Atlas*, 1978, Kimura (1981) would have the duplication placed at approximately 500 million years ago, which he says "is a much more appropriate time than that of Goodman et al. who place it about 425 million years ago". Perhaps we could compromise at 450 MyBP, the approximate times that I use in Fig. 1

this is true, the actual amino acid sequence of lamprey myoglobin is needed. The second fact is that the most parsimonious arrangement for the major globin branches still places, as depicted in Fig. 1, gnathostome myoglobin next to gnathostome α and β hemoglobin and then joins this group to the agnathan (hagfish-lamprey) lineage. Thus, the parsimony arrangement depicts the duplication from which gnathostome myoglobin and hemoglobin genes arose and then the duplication for α and β hemoglobin genes as both having occurred in the primitive vertebrate lineage to the Gnathostomata. Thirdly, on taking the earliest date for the Agnatha-Gnathostomata divergence deduced from the fossil record (about 500 MyBP) and the unaugmented NR link values, the rates of globin evolution on the lineages in which these gene duplications occurred in descent to the gnathostome ancestor (about 425 MyBP) are found to be in the range of 56 to 71 NR%, or on taking the augmented link values, 93 to 110 NR%.

From the neutralist standpoint such exceptionally fast molecular evolution would be caused by loss of constraint in duplicated genes. From the Darwinian standpoint, the duplicated genes provided an opportunity for the acquisition of new functions. This is illustrated by the reconstructed genealogy for 159 globin sequences (Goodman and Czelusniak 1980; Goodman (1981). For example, in the differentiation of α- and β-hemoglobin chains in the early vertebrates, the sequence positions which acquired cooperative functions ($\alpha_1\beta_2$ contacts, Bohr effect, DPG-binding) evolved 4 times more rapidly than the exterior positions not implicated in the important functions of heterotetrameric hemoglobin, whereas later from the bird-mammal ancestor to the present the set of positions with cooperative functions were extremely conservative and evolved at less than one fourth the rate of the exterior positions not implicated in important functions. Similarly in the gnathostome myoglobin branch the categories of functional sites (e.g. salt-bridges) that had earlier evolved in the primitive vertebrates at an elevated rate later evolved at a very slow rate. In general, wherever strong stabilizing selection acts on a protein, somewhere in the past positive Darwinian selection must have spread the amino acid substitutions which are now being preserved. Even initially neutral substitutions could become selectively favored, later, when coadaptively advantageous substitutions occurred at other molecular sites. Natural selection would spread such coadapted genes by favoring the deme or species possessing them over competing groups possessing less beneficial genes.

Acknowledgement. I thank Alejo E. Romero-Herrera for giving me the results of his comparison of lamprey myoglobin to lamprey and hagfish hemoglobins and for valuable discussion. This work was supported in part by NSF grant DEB 78–10717.

References

Baba ML, Darga LL, Goodman M, Czelusniak J (1980) Evolution of cytochrome *c* investigated by the maximum parsimony method. J Mol Evol (In press)

Braunitzer G, Fujuki H (1969) Zur evolution der vertebraten die konstitution und tertiärstruktur des hämoglobins des flußneunauges. Naturwissenschaften 56:322:323

Czelusniak J, Goodman M, Moore GW (1978) On investigating the statistical properties of the populous path algorithm by computer simulation. J Mol Evol 11:75–85

Dayhoff MO (1972) Atlas of protein sequence and structure, vol 5. National Biomedical Research Foundation, Washington, DC

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

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

Fisher WK, Nash AR, Thompson EOP (1977) Haemoglobins of the shark, *Heterodontus portusjacksoni*. III. Amino acid sequence of the β chain. Aust J Biol Sci 30:487–506

Fisher WK, Thompson EOP (1979) Myoglobin of the shark *Heterodontus portusjacksoni*: isolation and amino acid sequence. Aust J Biol Sci 32:277–294

Goodman M (1981) Decoding the pattern of protein evolution. Prog Biophys Mol Biol (in press)

Goodman M, Czelusniak J (1980) Mode, tempo, and role of natural selection in the evolution of heme proteins. Protides of the Biological Fluids 28:57–60

Goodman M, Moore GW, Barnabas J, Matsuda G (1974) The phylogeny of human globin genes investigated by the maximum parsimony method. J Mol Evol 3:1–48

Goodman M, Moore GW (1975) Darwinian evolution in the genealogy of haemoglobin. Nature 253:603–608

Goodman M, Czelusniak J, Moore GW, Matsuda G (1979a) Fitting the gene lineage into its species lineage: A parsimony strategy illustrated by cladograms constructed from globin sequences. Syst Zool 28:132–163

Goodman M, Pechére JF, Haiech J, Demaille JG (1979b) Evolutionary diversification of structure and function in the family of intracellular calcium-binding proteins. J Mol Evol 13:331–352

Heindell HC, Liu A, Paddock GV, Studnicka GM, Salser WA (1978) The primary sequence of rabbit α-globin in RNA. Cell 15:43–54

Holmquist R (1979) The method of parsimony: an experimental test and theoretical analysis of the adequacy of molecular restoration studies. J Mol Biol 135:939–958

Kimura M (1969) The rate of molecular evolution considered from the standpoint of population genetics. Proc Natl Acad Sci USA 63:1181–1188

Kimura M (1979) The neutral theory of molecular evolution. Sci Am 241 (No. 5, Nov.):94–104

Kimura M (1981) Was globin evolution very rapid in its early stages?: A dubious case against the rate-constancy hypothesis. J Mol Evol 17:110–113

Li SL, Riggs A (1970) The amino acid sequence of hemoglobin V from the lamprey *Petromyzon marinus*. J Biol Chem 245: 6149–6169

Liljeqvist G, Braunitzer G, Palétus S (1979) Hämoglobine, XXVII Die sequenz der monomeren hämoglobine III von *Myxine* glutinosa L: ein neurer hämkomplex: E7 glutamin, E11 isoleucin. Hoppe-Seyler's Z Physiol Chem 360:125–135

Moore GW (1977) Proof of the populous path algorithm for missing mutations in parsimony trees. J Theor Biol 66:95–106

Moore GW, Barnabas J, Goodman M (1973) A method for constructing maximum parsimony ancestral amino acid sequences on a given network. J Theor Biol 38:459–485

Nash AR, Fisher WK, Thompson EOP (1976) Haemoglobins of the shark, *Heterodontus portusjacksoni*. II. Amino acid sequence of the α-chain. Aust J Biol Sci 29:73–97

Nishioka Y, Leder P (1979) The complete sequence of a chromosomal mouse α-globin gene reveals elements conserved throughout vertebrate evolution. Cell 18:875–882

Romero-Herrera AE, Lieska N, Nasser S (1979) Characterization of the myoglobin of *Petromyzon marinus*. 14:259–266

Von Ehrenstein G (1966) Translational variations in the amino acid sequence of the α-chain of rabbit hemoglobin. Cold Spring Harbour Symp Quant Biol 31:705–714

Wilson JT, Wilson LB, Reddy VB, Cavallesco C, Ghosh PK, de Riel JK, Forget BG, Weissman SM (1980) Nucleotide sequence of the coding portion of human α globin messenger RNA. J Biol Chem 255:2807–2815

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

Received September 29, 1980