the sequence, and the NOT α/β proteins. Table 1 shows only slight deviations from expected values for the NOT α/β structures but large differences for the α/β proteins, in which there is a very strong preference for the N-terminal 10 residues to include a β-strand and not a helix, whereas the C-terminus is frequently helical, but rarely extended.

We examined the individual amino-terminal strands (28 proteins) and carboxy-terminal helices (33 proteins) to determine whether there were any common characteristics. The most striking difference is that the N-terminal strands are usually buried in the centre of the sheet, in contrast to the C-terminal helices which are predominantly exposed. The data for the amino-terminal strands show that there are less edge strands than expected by chance. A full analysis of the relative solvent accessibilities of the terminal regions is in progress.

Almost half of the proteins in the data base are extracellular and will probably be synthesized with an N-terminal signal peptide, which is later cleaved[7]. These signal sequences have high helical propensities and, if not removed from the protein, might be considered as N-terminal helical regions. However, this observation does not affect the results for the β/α proteins which are almost exclusively intracellular. In addition, the signal peptide is unlikely to influence the native structure of the secreted protein.

Thus, the amino- and carboxy-terminal regions have different conformational probabilities: the amino-terminus preferentially adopts an extended β-strand conformation while the carboxy-terminus is more usually helical. The observed difference derives basically from the α/β proteins, suggesting that the origin of this preference lies not in protection against degradation but in the special structural topology of α/β proteins and the $\beta\alpha\beta$ unit. There are several possible explanations. Perhaps the simplest is that the $(\beta\alpha)_n$ structure is the basic unit in evolution, from which all α/β proteins are derived. In 44 $\beta_1\alpha\beta_2$ units, in which the parallel strands are adjacent in the sheet, there are more than twice as many alpha-carbon close contacts (<6 Å) between β_1 and the intervening helix α, than between β_2 and α. In addition, the secondary structure prediction using the method developed by Robson and co-workers[8], is bettter for $\beta_1\alpha$ than for $\alpha\beta_2$ (W. R. Taylor, personal communication). This reinforces the concept of a basic $\beta\alpha$ structure, as opposed to a $\alpha\beta$ unit. Alternatively, from a sequential folding viewpoint, the most favourable structure for a $\beta\alpha\beta$ sequence is almost certainly parallel β-strands joined by an antiparallel helix. In contrast, an $\alpha\beta\alpha$ sequence would probably maximize the α–α contacts including helix dipole interactions[9] in an antiparallel hairpin joined by an extended region. If folding proceeds from an amino-terminal nucleating core, this terminus would tend to be more hydrophobic and therefore incorporate more β-sheet than an exposed carboxy-terminus. A definitive explanation must await more experimental data on the folding pathways of α/β proteins.

We thank Professor Tom Blundell for his interest and encouragement and William Taylor for useful comments. J.M.T. holds a SERC Advanced Fellowship.

Received 31 March; accepted 13 May 1982.

1. Birktoft, J. J. & Blow, D. M. *J. molec. Biol.* **68**, 187–240 (1972).
2. Huber, R., Kukla, D., Ruhlmann, A. & Steigemann, W. *Cold Spring Harb. Symp. quant. Biol.* **36**, 141–150 (1971).
3. Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C. & Rupley, J. A. in *The Enzymes* Vol. 7, 3rd edn (ed. Bayer, P. D.) 665–868 (Academic, New York, 1972).
4. Levitt, M. & Chothia, C. *Nature* **261**, 552–558 (1976).
5. Bernstein, F. C. *et al. J. molec. Biol.* **112**, 535–542 (1977).
6. Richardson, J. *Adv. Protein Chem.* **34**, 167–339 (1981).
7. Kreil, G. *A. Rev. Biochem.* **50**, 317–348 (1981).
8. Garnier, J., Osguthorpe, D. J. & Robson, B. *J. molec. Biol.* **20**, 97–120 (1978).
9. Hol, W. G. J., Halie, L. M. & Sander, C. *Nature* **294**, 532–536 (1981).

Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes

John Czelusniak*†, Morris Goodman*†‡, David Hewett-Emmett§, Mark L. Weiss‡, Patrick J. Venta§ & Richard E. Tashian§

* Department of Anatomy, Wayne State University School of Medicine, Detroit, Michigan 48201, USA
† Department of Biological Sciences, Wayne State University, Detroit, Michigan 48202, USA
‡ Department of Anthropology, Wayne State University, Detroit, Michigan 48202, USA
§ Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA

Recent years have seen rapid growth in amino acid sequence data on globins and nucleotide sequence data on haemoglobin genes and pseudogenes, and cladistic analysis[1] of these data continues to reveal new facets of globin evolution. Our present findings demonstrate: (1) avian and mammalian embryonic α genes (π and ξ, respectively) had a monophyletic origin involving an α locus duplication about 400 Myr ago soon after the duplication which separated α and β genes; (2) much later in phylogeny, independent β-gene duplications produced the embryonic ρ locus of birds and embryonic ε and fetal γ loci of mammals. This parallels the earlier finding[2] that myoglobins evolved more than once from generalized globin ancestors. Here we support the view[2] that such globin evolution resulted from natural selection acting on mutations in duplicated genes. Thus, our evidence contradicts the neutralist view[3,4] in which almost all amino acid substitutions in descent to extant globins evaded positive selection.

The facets of globin evolutionary history shown in Fig. 1 come from a genealogical reconstruction carried out on amino acid sequence data from 195 globin chains of 87 vertebrate and 19 non-vertebrate species. Figure 2 shows the genealogical reconstruction for 40 α- and β-haemoglobin genes and pseudogenes of 8 vertebrate species carried out on nucleotide sequence data representing coding regions of the expressed genes and corresponding regions of the unexpressed pseudogenes. Previous studies of globins[1] and other protein families, such as calcium binding proteins[5], cytochromes *c* (ref. 6), α-lens crystallins[7], carbonic anhydrase isozymes[8] and ribonucleases[9], indicate that genealogical reconstructions by the maximum parsimony procedure provide an effective way of extracting cladistic information from sequence data. The use of nucleotide sequences allows the numbers of nucleotide replacements which cause amino acid changes and those which do not (silent replacements) to be calculated for each branch of such a tree (see Fig. 2 and legend). The ratio of amino acid changing to silent substitutions R_{cs} value obtained from this calculation is one of three parameters that we used to evaluate the role of natural selection in globin evolution, the other two being rate of amino acid substitutions and distributions of substitutions in functionally different parts of the haemoglobin molecule. The changes in values of these parameters during different evolutionary periods provide evidence that a substantial proportion of amino acid substitutions were adaptive.

Myoglobin, β-haemoglobin and embryonic and adult α-haemoglobins of extant birds and mammals had their origins during early vertebrate phylogeny (Fig. 1). Assuming that the agnathan from gnathostome (jawed vertebrate) and shark from teleost–tetrapod globin-lineage splittings were congruent with the corresponding species-lineage splittings and that paleontological views on early vertebrates are not too inaccurate, rates of amino acid substitutions in these diverging early-vertebrate globin lineages were exceptionally fast, with rates as high as 110 NR% (nucleotide replacements per 100 codons per

0028–0836/82/280297—04$01.00

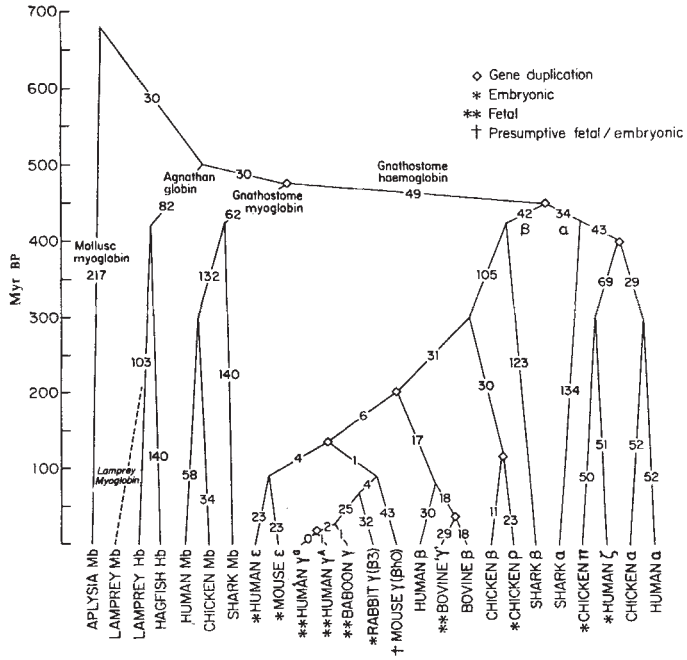

Fig. 1 Representative lineages from the most parsimonious globin phylogenetic tree found for 195 globin amino acid sequences. The strategy for finding among the alternative trees examined the one with fewest genic changes was the same as that used in a previous cladistic analysis[1] of 172 globin sequences. The present analysis provides clear evidence on the phylogenetic origins of bird and mammal embryonic haemoglobin (Hb) sequences. For example, grouping the chicken embryonic α or π branch with the reptilian–avian adult α branch while grouping the human embryonic α or ζ branch with the mammalian adult α branch, rather than having the monophyletic origin of π and ζ sequences shown in the figure, adds 35 NRs to the tree. Conversely, having in the basal amniotes a monophyletic origin for mammalian embryonic β-like or ε chains and avian embryonic β-like or ρ chains, rather than the phlogenetically later independent origins shown in the figure, adds 30 NRs to the tree. The numbers of NRs shown on the lines of descent have been corrected for superimposed mutations by our augmentation algorithm[6]. The ordinate scale in Myr BP is based on palaeontological views, as previously used[1], concerning the ancestral separations of the organisms from which the 195 globins came. If the full tree were shown, it would be apparent why the gene duplications are placed in their particular positions on the ordinate scale. For example, in the full tree the embryonic αs (π and ζ) and teleost αs are joined together and diverge from the lineage to tetrapod adult αs. Thus, the duplication from which the embryonic and adult α loci emerged is deduced to have occurred just before the ancestral divergence of teleosts and tetrapods, that is at ~400 Myr BP. As another example, the divergence of the mammalian ε–γ line from the β line is placed at ~200 Myr BP because it is followed in the β line by the divergence of monotremes from therian mammals. Lamprey myoglobin (Mb) was not included in the actual computer reconstruction. However, the dashed line shows, from evidence reviewed elsewhere[16], that this myoglobin is cladistically closer to lamprey haemoglobin than to any other globins. Sequences not previously catalogued[1,16] are chicken α[17], human ζ[2] (ref. 18), chicken ρ (ref. 19), mouse γ (βh0), (C. A. Hutchinson, S. J. Phillips, A. Hill, S. C. Hardies and M. H. Edgall, personal communication), rabbit γ (β3)[20] and mouse ε[21]. Computer simulations establish that our phylogenetic tree reconstruction algorithms capture such variables of the simulation as the frequencies of different types of nucleotide substitutions and the differential rates of evolutionary change in different tree regions and at different alignment positions (R. Holmquist, T. Conroy, J. C. and M.G., in preparation). The results also support the finding of an earlier study[22] that the denser the phylogenetic tree the more accurate the maximum parsimony reconstruction. The density of sequences used for the present amino acid tree was sufficient to yield a reliable reconstruction.

100 Myr). Between gnathostome and amniote (bird–reptile–mammal) ancestors (from ~425–300 Myr BP), rates were also fast, averaging 59 NR% in the four major globin lineages. However, later in vertebrate phylogeny from the amniote ancestor to present day birds and mammals, amino acid substitutions occurred at a much slower rate, averaging ~15 NR%. During the early period of fast globin evolution, the highest concentrations of amino acid substitutions were at residue positions which acquired new or altered functions. This was particularly true during the initial evolution of heterotetrameric haemoglobin following the gene duplication which produced separate α and β loci. The fastest rates seem to have occurred at those sites which became responsible for the subunit cooperativity permitting rapid unloading of oxygen in respiring tissues (Table 1). However, later in descent these cooperative sites, that is the $\alpha_1\beta_2$ contacts, Bohr effect sites and 2,3-diphosphoglycerate (DPG)-binding sites[10,11], along with the haem contacts, evolved at the slowest rate. It follows that natural selection, rather than random drift of selectively neutral mutations, was responsible for this nonrandom pattern of amino acid substitutions. A shift from higher to lower R_{cs} values during descent of the expressed α- and β-haemoglobin genes is possibly a further indication that positive selection for adaptive substitutions became stabilizing selection preserving perfected adaptations.

A standard for evaluating the R_{cs} values of lineages of expressed haemoglobin genes is provided by the R_{cs} values of lineages of unexpressed α- and β-haemoglobin pseudogenes. The various eutherian lineages of these α- and β-pseudogenes depicted in Fig. 2 have an average R_{cs} value of 2.2, somewhat less than that of randomly mutating codons ($R_{cs} = 3.0$). Previous calculations regarding the interval of time between the production by gene duplications of a proto-pseudogene lineage and its silencing suggest that the lower the proportion of amino acid changing to silent base replacements, the longer the interval before silencing[12]. However, doubt must now be cast on the assumptions used in such calculations. Consider the two cladistically related goat ψ–β sequences: each possesses the same chain terminating mutation (thus indicating that they both originated from an ancestral sequence already silenced[13]), yet they have an average R_{cs} value of only 1.9. Perhaps structural constraints operating on the chromosomal segments containing the expressed haemoglobin genes also operate on the unexpressed pseudogenes, preventing a strictly random accumulation of mutations. Moreover, the same nonrandom usage of bases in the third position of codons (for example, excesses of C-ending alanine, asparagine, glycine, threonine and tyrosine codons) found in α- and β-haemoglobin genes are found in the pseudogenes. Nevertheless, the high rate of evolution of eutherian haemoglobin pseudogenes indicates that they are under much less stringent selection for preservation of coding sequence structure than are expressed haemoglobin genes. In contrast to pseudogene lineages, expressed lineages frequently show R_{cs} values less than 1. This indicates that silent replacements are less constrained by stabilizing selection than are amino acid-changing replacements. However, when positive natural selection is transforming the sequence structure of a protein, elevated R_{cs} values in the range of pseudogenes and randomly mutating codons can be expected.

The amino acid-changing and silent base-replacement values recorded on the branches of the nucleotide sequence tree (Fig. 2), and the R_{cs} values calculated from them suggest that during differentiation of α- and β-haemoglobin loci there were many more amino acid-changing replacements than silent ones. A caveat is necessary because other methods[14] imply that saturation of silent substitutions leads to high R_{cs} ratios between distantly related sequences. This suggests that the estimated R_{cs} values between amniote α and β ancestors, at the apex of the tree, do not reflect the actual R_{cs} values which occurred. We have tested, by simulation, that under the null hypothesis of a uniformly faster rate of silent to amino acid-changing substitutions, high R_{cs}values at the tree's apex would result from artefacts of our method. Results to be described elsewhere show that the apex R_{cs} values given in Table 1 and Fig. 2 are two to three times higher than those found in the simulations, thus indicating that during differentiation of α- and β-haemoglobin loci there was indeed a higher rate of amino acid-changing substitutions to silent. Moreover, the codons for sites which have the strongest functions in fully evolved

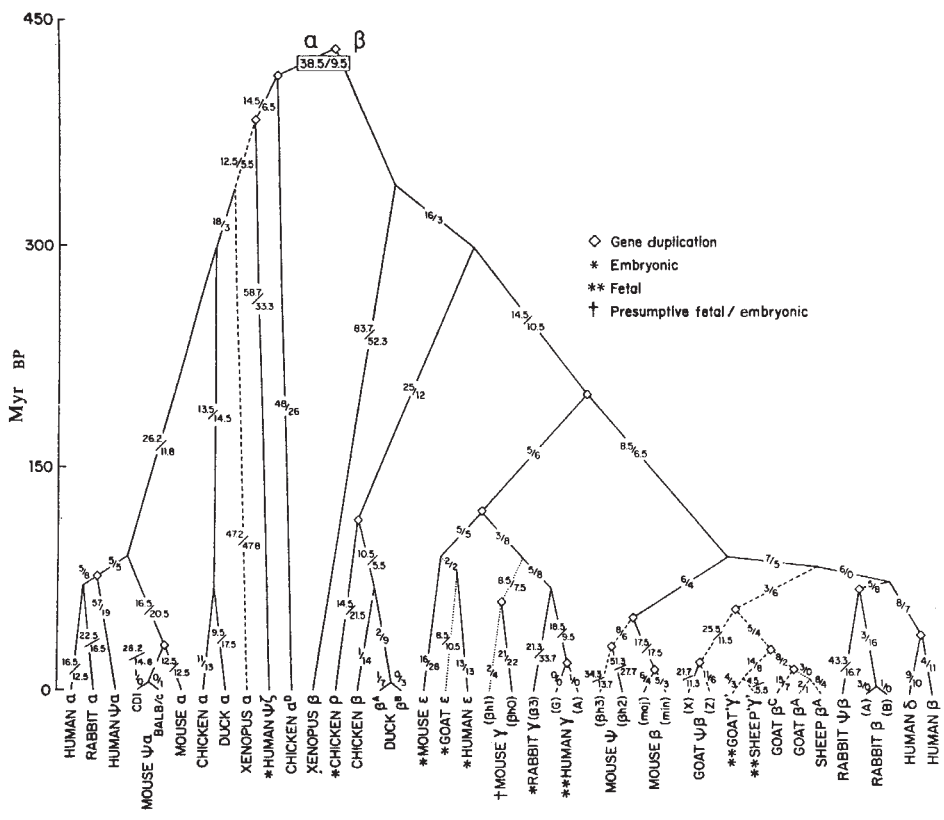

Fig. 2 Most parsimonious phylogenetic tree for 40 haemoglobin genes and pseudogenes. The nucleotide sequences were aligned to give 444 nucleotide positions; intervening sequences (introns), 3′ and 5′ non-coding DNA and insertions present only in the pseudogene sequences were not included in the alignment. The method used to construct the most parsimonious tree and calculate the branch lengths was that described by Goodman (ref. 1 refs therein) modified for use with nucleotide sequences. The numbers shown as a fraction on each branch or link are, in the numerator, the number of amino acid-changing base replacements and, in the denominator, the number of silent base replacements. Our computer algorithm carried out this calculation for each pair of ancestral and descendant-codons at each two adjacent nodes of the tree; if there is more than one mutation separating the two codons, all possible minimum pathways of getting through the genetic code from one codon to the other are examined, the number of amino acid-changing nucleotide replacements for each path are calculated, and the average is taken. For calculating the ratios of amino acid-changing to silent replacements (that is, R_{cs} values), the numbers recorded on the links for the two types of point mutations are originally observed or unaugmented NR values. Because relatively few contemporary sequences represent the long period of evolutionary time involved (from 450 Myr BP to the present), the numbers of both the amino acid-changing and silent NRs are almost certainly underestimated, especially at the apex of the tree and on the branches which start near the apex and descend to the present without intervening nodes. Nevertheless, the results of computer simulation tests of the accuracy of our maximum parsimony algorithm with regard to measuring the relative proportions of the two types of NRs (see text) suggest that the shift from higher R_{cs} values on earlier links of the tree to lower R_{cs} values within the amniotes is not an artefact of the maximum parsimony reconstruction but rather reflects a real evolutionary trend. The following complete sequences were used: human α[23], rabbit α[24], human $\psi\alpha$[25], mouse α[26], mouse $\psi\alpha$ strains CD1 and BALB/c[27,28], chicken α[17], duck α[29], human $\psi\zeta$1, (N. J. Proudfoot personal communication), chicken α^D (ref. 17), *Xenopus* β[30,31], chicken ρ[19], chicken β[32], duck $^A\beta$ (ref. 33), duck $^B\beta$ (ref. 34), mouse ε[21], human ε[35], mouse $\gamma(\beta h0)$ (C. A. Hutchinson, S. J. Phillips, A. Hill, S. C. Hardies and M. H. Edgall, personal communication), rabbit $\gamma(\beta 3)$[20], human $^G\gamma$ and $^A\gamma$ (ref. 36), mouse $\psi\beta$h2 (C. A. Hutchinson, S. J. Phillips, A. Hill, S. C. Hardies and M. H. Edgall, personal communication), mouse β-maj and β-min[37], goat $\psi\beta^x$ (ref. 38), rabbit $\psi\beta$[39], rabbit $^A\beta$ and $^B\beta$ alleles[40,41], human δ[42] and human β[43]. Also two sequences were used which had, due to extensive deletions, less than the full per cent of coding blocks: goat $\psi\beta^Z$ (76%)[13] and mouse $\psi\beta$h3 (69%) (C. A. Hutchinson, S. J. Phillips, A. Hill, S. C. Hardies and M. H. Edgall, personal communication). In addition, the following partial sequences (portions of each sequence not having been determined by the time of our study) were used: *Xenopus* α (81%)[44], goat ε (35%)[45,46], mouse γ (βh1) (30%) (C. A. Hutchinson, S. J. Phillips, A. Hill, S. C. Hardies and M. H. Edgall, personal communication), goat 'γ' (64%)[45,46], sheep 'γ' (79%)[47], goat $^C\beta$ (62%)[45,46], goat $^A\beta$ (49%)[45,46] and sheep $^A\beta$ (72%)[47]. Dotted lines; links on which the NR values represented very incomplete portions (in the range 30–49%) or the α or β alignment. Dashed lines: links on which the NR values represent more extensive portions (in the range 62–81%) of the α or β alignment. Solid lines: links on which the NR values represent all or almost all the α or β alignment. Note that the lineage to human $\psi\zeta$1 was for most of its history an expressed ζ lineage as the human $\psi\zeta$1 sequence shows only six changes (a deleted residue and five base differences due to four amino acid differences) from the expressed human ζ2 sequence. Note also that certain of the indicated gene duplications (◇) might be gene conversions[36] as well as duplications. For example, the closer cladistic relationships of human $^A\gamma$ and $^G\gamma$ chains to one another than to baboon γ (Fig. 1) could be due to gene conversion in the hominoid lineage in as much as cercopithecoids like hominoids have two γ loci[48]. Similarly, the closer cladistic relationship of bovid 'γ' genes to bovid β genes than to γ genes in primates, lagomorphs and rodents could reflect a conversion of the older γ locus by the bovid β locus.

heterotetrameric haemoglobin (haem contacts and cooperative sites) clearly show, with average (av.) $R_{cs} > 5$, this higher rate of amino acid-changing replacements. Conversely, within the amniotes, not only did far fewer amino acid changing to silent replacements (R_{cs} often <1) occur in the expressed α- and β-gene lines, but the codons for haemoglobin sites with the strongest functions accumulated the smallest proportion of amino acid-changing base replacements (av. $R_{cs} = 0.35$). Thus, we conclude that stringent selection in the amniotes was preserving the adaptive amino acid substitutions which had occurred during differentiation of α and β chains. This view on selection seems more logical than that of the neutralists[3,4,12], who believe that selection serves only as a conservative force. The extreme neutralist's view fails to explain the evolutionary origins of adaptive sequence structures—as though proteins were created with almost all of their adaptive features firmly in place.

Further evidence of positive selection of adaptive amino acid substitutions followed by stabilizing selection is provided by changes in rates of amino acid substitutions and in R_{cs} values which occurred during shorter spans of evolutionary time and closer to the present than for the previous evidence presented. A good example of this is the pattern of changes on the lineage to catarrhine primate γ chains. From the primate–lagomorph ancestral node to catarrhine ancestral node (~70–25 Myr BP), the rate of amino acid-changing base replacements, as calculated from the amino acid sequence tree (Fig. 1), was six to seven times faster than from the catarrhine ancestral node to present (40 NR% compared with only ~6 NR%). During the earlier period of accelerated evolution in this primate γ line, three base replacements occurred at DPG-binding sites: valine at helical position NA1 mutated to glycine, and histidine at helical position H21 mutated through two amino acid-changing base replacements to serine. The loss of DPG-binding capacity brought about by these amino acid substitutions favoured the ability of fetal haemoglobin to capture oxygen from maternal haemoglobin. This may have helped make possible the relatively long period of fetal life coupled with extensive prenatal brain development distinguishing catarrhine primates from other eutherian mammals. That almost all the amino acid-

Table 1 Amino acid changing (C) and silent (S) nucleotide replacement rates: NR/codon position/100 Myr × 10

Evolutionary lineages	Functional groups (no. of codon positions)										Total NRs All (148)		
	HC (23)		Coop (24)		$\alpha_1\beta_1$ (14)		IP (19)		Other (68)				
	C	S	C	S	C	S	C	S	C	S	C	S	R_{cs}
a, α–β link	1.21	0.29	4.00	0.71	1.50	0.50	1.14	1.14	1.57	0.21	38.5	9.5	4.05
	HC (19)		Coop (16)		$\alpha_1\beta_1$ (15)		IP (21)		Other (70)		All (141)		
b, Stem to amniote α anc	0.42	0.00	0.50	0.50	3.92	1.08	2.00	2.00	3.67	0.83	45	15	3.00
c, Amniote αs	0.06	0.82	0.27	1.01	1.53	1.05	1.08	1.20	1.57	1.23	133.2	131.8	1.01
d, Bird αs	0.00	1.14	0.43	1.36	1.43	0.93	1.36	1.71	1.29	1.79	20.5	30.5	0.67
e, Eutherian αs	0.20	1.28	0.00	2.52	3.20	1.60	1.52	2.08	2.96	2.40	73	75	0.97
f, Eutherian $\psi\alpha$s	3.45	1.82	5.36	1.55	4.82	1.18	4.73	1.73	6.55	2.91	86.2	34.8	2.48
	HC (23)		Coop (19)		$\alpha_1\beta_1$ (15)		IP (19)		Other (70)		All (146)		
g, Stem to amniote β anc	1.00	0.00	5.25	0.00	3.25	0.00	1.25	0.00	2.75	1.00	16	3	5.33
h, Amniote ρ-ε-γ-βs	0.53	0.99	0.29	1.00	1.64	1.60	1.13	1.19	1.52	1.52	276.9	317.1	0.87
i, Bird βs	0.00	1.73	0.00	0.73	0.87	0.87	0.00	1.40	0.20	1.80	4	33	0.12
j, Eutherian ε-γ-βs	0.86	1.49	0.34	1.59	2.33	2.41	1.64	1.41	2.20	2.14	186.9	211.1	0.89
k, Eutherian $\psi\beta$s	3.82	2.88	7.29	2.88	8.71	5.00	4.18	1.24	6.35	3.00	149.8	73.2	2.05

Only lineages to complete sequences are included in the calculation of nucleotide replacement rates for the different functional groups of codon positions. The time scale for the rate calculations is that used in Fig. 2. Nodes representing gene duplications are placed on this time scale by extrapolation of the link-length NR values connecting such nodes to those which represent the species-splittings dated by palaeontological evidence. *a*, Time of α–β duplication is placed at 450 Myr BP, the first node in descent of the α stem (where chicken α^D joins other αs) at 420 Myr BP, and the *Xenopus*–amniote ancestral β node at 340 Myr BP; this yields a time of 140 Myr. *b*, Times of the connecting links from the first α node after the α–β duplication to the amniote α ancestral node, placed at 300 Myr BP, add up to 120 Myr. *c*, Times of the connecting links from the amniote α ancestral node to the extant complete sequences add up to 830 Myr. *d*, Times of the connecting links from the chicken–duck ancestral α node (placed at 70 Myr BP) to the present add up to 140 Myr. *e*, Times of connecting links from the eutherian α ancestral node (placed at 90 Myr) to the extant complete α sequences add up to 250 Myr. *f*, Times of the links from the ancestral nodes of these $\psi\alpha$ sequences with their nearest expressed α lineages to the $\psi\alpha$ sequences add up to 110 Myr. *g*, Time of the link from the Xenopus–amniote ancestral β node to amniote ancestral ρ–ε–γ–β node is 40 Myr. *h*, Times of the connecting links from the amniote ancestral ρ–ε–γ–β node to the present day ρ, ε, γ and β complete sequences add up to 1,630 Myr. *i*, Times of the connecting links from the duck–chicken ancestral β node to the present day duck and chicken βsequences add up to 150 Myr. *j*, Times of the connecting links from eutherian ε, eutherian γ and eutherian β ancestral nodes (each placed at 90 Myr BP) to present day complete ε, γ and β sequences, respectively, add up to 760 Myr. *k*, Times of the links from the ancestral nodes of these $\psi\beta$ sequences with their nearest expressed β lineages to the $\psi\beta$ sequences add up to 170 Myr. The functional groups of codon positions follow the scheme used for amino acid sequence data[1]: HC haem contacts; Coop: the cooperative sites $\alpha_1\beta_2$ contact, Bohr effect, and for lineages *a* and *g*–*k* also DPG; $\alpha_1\beta_1$, $\alpha_1\beta_1$ contacts; IP, interior positions; Other, remaining positions. The amino acid changing NR rates calculated for these functional groups over the evolutionary lineages are smaller (probably because of the tree's low density of nucleotide sequence data) than the rates previously found in reconstructions of globin phylogeny from amino acid sequence data. Nevertheless, the trends observed for the amino acid-changing NR rates among the different functional groups and between early vertebrate and later amniote lineages generally agree with those found[1] on a dense body of amino acid sequence data. in addition, the present analysis explores a new parameter, comparison of silent NR rates to amino acid-changing ones. It also compares these rates between unexpressed pseudogene lineages and expressed gene lineages. Total NRs, the total numbers of amino acid-changing and silent NRs over all codon positions for each group of evolutionary lineages, rather than nucleotide replacement rates.

changing base replacements on this early primate γ line, not just the three at DPG-binding sites, also helped to perfect this adaptation for fetal life, is indicated by the very slow rate of evolution of catarrhine γ chains.

Complimentary evidence that natural selection directed the evolution of primate γ genes is provided by the R_{cs} values of ε and γ lineages of the nucleotide sequence tree (Fig. 2). Many more silent replacements than amino acid-changing ones occurred in the stem to the ε–γ ancestral node and also in lineages descending from this node to human, goat and mouse ε genes and to mouse and rabbit γ genes. However, after the divergence of primate and rabbit γ lines, most base replacements on the primate γ line were amino acid-changing ones rather than silent. This supports the inference that positive selection acted at the DPG-binding sites and many other moderating amino acid positions of the pre-catarrhine γ chains to fix adaptive amino acid substitutions.

An earlier part of this work has been reported elsewhere[15].

We thank Drs N. J. Proudfoot, P. Leder, C. A. Hutchinson III, S. J. Phillips, A. Hill, S. C. Hardies and M. Edgell for supplying us with nucleotide sequence data before publication. This work was supported by NSF grant DEB 78-10717 (M.G.) and USPHS grant GM 24681 (R.E.T.).

Received 17 February; accepted 28 April 1982.

1. Goodman, M. *Prog. Biophys. molec. Biol.* **38,** 105–164 (1981).
2. Goodman, M. *J. molec. Evol.* **17,** 114–120 (1981).
3. Kimura, M. *Nature* **217,** 624–626 (1968).
4. Kimura, M. *J. molec. Evol.* **17,** 110–113 (1981).
5. Goodman, M. in *Calcium-Binding Proteins: Structure and Function* (eds Siegel, F. L., Carafoli, E., Kretsinger, R. H., MacLennan, D. H. & Wasserman, R. H.) 347–354 (Elsevier, New York, 1980).
6. Baba, M. L., Darga, L. L., Goodman, M. & Czelusniak, J. *J. molec. Evol.* **17,** 197–213 (1981).
7. De Jong, W. W., Zweers, A, & Goodman, M. *Nature* **292,** 538–540 (1981).
8. Tashian, R. E., Hewett-Emmett, D. & Goodman, M. in *Protides of the Biological Fluids* Colloq. 28 (ed. Peeters, H.) 153–156 (Pergamon, Oxford, 1980).
9. Beintema, J. *et al. J. molec. Evol.* **10,** 49–71 (1977).
10. Ladner, R. C., Heidner, E. J. & Perutz, M. F. *J. molec. Biol.* **114,** 385–414 (1977).
11. Fermi, G. & Perutz, M. F. *J. molec. Biol.* **114,** 421–431 (1977).
12. Li, W-H., Gojóbori, T. & Nei, M. *Nature* **292,** 237–239 (1981).
13. Cleary, M. L., Schon, E. A. & Lingrel, J. B. *Cell* **26,** 181–190 (1981).
14. Efstratiadis, A. *et al. Cell* **21,** 653–668 (1980).
15. Hewett-Emmett, D. *et al. Fedn Proc.* **40,** 1591 (1981).
16. Goodman, M., Romero-Herrera, A. E., Czelusniak, J., Dene, H. & Tashian, R. E. in *Macromolecular Sequences in Systematic and Evolutionary Biology* (ed. Goodman, M) (Plenum, New York, in the press).
17. Dodgson, J. B., McCune, K. C., Rusling, D. J., Krust, A. & Engel, J. D. *Proc. natn. Acad. Sci. U.S.A.* **78,** 5998–6002 (1981).
18. Aschauer, J., Sanguansermsri, T. & Braunitzer, G. *Hoppe-Seyler's Z. physiol. Chem.* **362,** 1159–1162 (1981).
19. Roninson, I. B. & Ingram, V. M. *Proc. natn. Acad. Sci. U.S.A.* **78,** 4782 (1981).
20. Hardison, R. C. *J. biol. Chem.* **2566,** 11780–11786 (1981).
21. Hansen, J. N., Konkel, D. A. & Leder, P. *J. biol. Chem.* **257,** 1048–1052 (1982).
22. Holmquist, R. *J. molec. Biol.* **135,** 929–938 (1979).
23. Wilson, J. T. *et al. J. biol. Chem.* **255,** 2807–2815 (1980).
24. Heindell, H. C. *et al. Cell* **15,** 43–54 (1978).
25. Proudfoot, N. J. & Maniatis, T. *Cell* **21,** 537–544 (1980).
26. Nishioka, Y. & Leder, P. *Cell* **18,** 875–882 (1979).
27. Vanin, E. F., Goldberg, G. I., Tucker, P. W. & Smithies, O. *Nature* **286,** 222–226 (1980).
28. Nishioka, Y., Leder, A. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **77,** 2806–2809 (1980).
29. Paddock, G. V. & Gaubatz, J. *Eur. J. Biochem.* **117,** 269–273 (1981).
30. Richardson, L., Cappello, J., Cochran, M. D., Armentrout, R. W. & Brown, R. D. *Devl Biol.* **78,** 161–172 (1980).
31. Williams, J. G., Kay, R. M. & Patient, R. K. *Nucleic Acids Res.* **8,** 4247–4258 (1980).
32. Richards, R. I. *et al. Nucleic Acids Res.* **7,** 1137–1146 (1980).
33. Niessing, J. *Biochem. Int.* **2,** 113–120 (1981).
34. Haupe, A., Therwath, A., Soriano, P. & Galibert, F. *Gene* **14,** 11–21 (1980).
35. Barelle, F. E., Shoulders, C. C. & Proudfoot, N. J. *Cell* **21,** 621–626 (1980).
36. Slightom, J. L., Blechl, A. E. & Smithies, O. *Cell* **21,** 627–638 (1980).
37. Konkel, D. A., Maizel, J. V. Jr & Leder, P. *Cell* **18,** 865–873 (1981).
38. Cleary, M. L., Haynes, J. R., Schon, E. A. & Lingrel, J. B. *Nucleic Acids Res.* **8,** 4791–4802 (1980).
39. Lacy, E. & Maniatis, T. *Cell* **21,** 545–553 (1980).
40. Van Ooyen, A., van den Berg, J., Mantai, N. & Weissman, C. *Science* **206,** 337–344 (1979).
41. Hardison, R. C. *et al. Cell* **18,** 1285–1297 (1979).
42. Spritz, R. A., De Riel, J. K., Forget, G. B. & Weissman, S. *Cell* **21,** 639–646 (1980).
43. Lawn, R. M., Efstratiadis, A., O'Connell, C. & Maniatis, T. *Cell* **21,** 647–651 (1980).
44. Partington, G. A. & Baralle, F. E. *J. molec. Biol.* **145,** 463–470 (1981).
45. Haynes, J. R. *et al. J. biol. Chem.* **255,** 6355–6367 (1980).
46. Haynes, J. R., Rosteck, P. Jr & Lingrel, J. B. *Proc. natn. Acad. Sci. U.S.A.* **77,** 7127–1731 (1980).
47. Kretschmer, P. J. *et al. J. biol. Chem.* **256,** 1975–1982 (1981).
48. Barrie, P. A., Jeffreys, A. J. & Scott, A. F. *J. molec. Biol.* **149,** 319–336 (1981).