ON THE EVOLUTION OF MYOGLOBIN

BY A. E. ROMERO-HERRERA† AND H. LEHMANN, F.R.S.
University Department of Clinical Biochemistry, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 2QR, U.K.
AND K. A. JOYSEY AND A. E. FRIDAY
University Museum of Zoology, Downing Street, Cambridge, CB2 3EJ, U.K.

(*Received* 17 *February* 1977)
[Pullouts 1 and 2]

CONTENTS

In previous studies, particularly of primates, a high degree of concordance was obtained between an evolutionary pattern based on comparative anatomy and another based on a reconstruction of the possible pathway of evolution of the myoglobin molecule. Accordingly, in extending our studies, we included species of uncertain evolutionary kinship, such as the tree shrew, and also increased the representation of species within the mammalian orders already studied, in the hope that it would be possible to resolve some contended issues and to establish the sequence of branching and pattern of relationship between those orders.

As a first step, the phylogenetic pattern within each of the mammalian orders was constrained on zoological grounds according to generally accepted kinship based on the evidence of comparative anatomy and the fossil record. Eight phylogenetic patterns reflecting different possible kinship between the orders were chosen for detailed investigation and for each of these the most economical (parsimonious) pathway which could be obtained for the evolution of myoglobin was reconstructed. The uncertainties which are inherent in such reconstruction were increased by a high incidence of parallel substitutions. Although the eight phylogenetic patterns were widely different from one another it was found that none of the solutions had a

† Present address: Department of Anatomy, Wayne State University, 540 East Canfield Avenue, Detroit, Michigan 48201, U.S.A.

convincing advantage in terms of parsimony over the alternatives. A phylogenetic grouping of the order Artiodactyla with the order Cetacea, rather than with the order Perissodactyla, did not provide a more economical solution and, within the Carnivora, a phylogenetic pattern implying a monophyletic origin of the pinnipeds had very little advantage over a phylogenetic model grouping the seal with the mustelid and the sea lion with the canids.

An amino acid difference matrix is of special interest because it carries information which is more comparable with that obtained by immunological methods, where access to sequence information has not usually been available. Several different clustering procedures were applied to the amino acid difference matrix for myoglobin. As would be expected from the different assumptions on which each is based, the procedures resulted in different branching patterns which varied in their degree of zoological acceptability. Clustering of the harbour seal with the Cetacea suggested that there might be functional reasons, perhaps associated with diving, for several substitutions acquired in parallel by the myoglobins of these mammals. Repeated clustering of the horse with the sportive lemur, and a tendency for the opossum to join the primates, is associated with a high proportion of parallel substitutions. This tends to undermine confidence in phylogenetic inferences which might be drawn from the repeated clustering of the tree shrew (and often the hedgehog) among the primates.

If we assume that there is an overall resemblance of the three-dimensional structure of all myoglobins, and use the crystallographic model of sperm whale myoglobin as a basis, the availability of over 30 vertebrate myoglobin sequences (and that of the mollusc *Aplysia*) has provided an opportunity to consider the functional relevance of certain positions at which the nature of the amino acid residue appears to have remained unchanged, or to have changed only conservatively, during several hundred million years of evolution. Part of this conservation is attributable to the maintenance of the monomeric nature of the molecule, and it seems likely that much of the conservatism can be attributed to functional needs in initiating the folding of the molecule and in the maintenance of its tertiary structure. Concomitantly, consideration is given to the likely functional consequence of some of the substitutions which have been accepted.

Atassi and his co-workers have carried out extensive immunological studies on myoglobin using antisera prepared in rabbit and goat. The availability of the amino acid sequences of rabbit and sheep (in lieu of goat) has made it possible to investigate the relation between amino acid sequence difference and the distance as measured by immunological criteria. It has not been possible to confirm Reichlin's hypothesis that the myoglobin antigenic reactive regions of Atassi are particularly variable compared with the rest of the molecule; on the contrary it appears that differences in the immunological reactive regions occur approximately in proportion to those occurring in the molecule as a whole.

Estimates of the rate of molecular evolution are of interest because its supposed constancy has been a major argument in favour of the hypothesis that a high proportion of fixed mutations are neutral, or nearly so, as far as selection is concerned. Furthermore, a near constant rate of molecular evolution holds promise for a molecular clock which might be used for dating evolutionary events. However, there is considerable variation in the rate of molecular evolution as evidenced by differences in the number of nucleotide substitutions ('hits') in lineages arising from the same phylogenetic branching point or by the examination of the changes at the amino acid level. In order to investigate absolute rates of evolution we sought to establish the best estimate of the date of divergence between the ancestors of the living species included in this study. These dates are based on direct fossil evidence and must be regarded as minimum dates because we have not indulged in open-ended speculations about fossils which have not yet been found. The combination of these dates with the evidence of our cladograms leads us to reaffirm our earlier finding that there are considerable differences in the amount of change in different lineages. For example, whereas one lineage (to gibbon) appears to have accepted no mutations during the past 20 Ma, another lineage (to ox) seems to have fixed seven mutations during the last 18 Ma.

Goodman and others have drawn attention to the apparently low rate of molecular evolution among higher primates; a similar observation applies to the myoglobin of the Old World monkeys so far studied, but neither the myoglobins of New World monkeys nor the prosimian myoglobins share this feature.

After their divergence from one another the two bird lineages included in this study appear to have fixed mutations at rates comparable with those found among mammals during the past 79 Ma. However, the number of differences between the bird and mammal ancestral stems appears to be remarkably low bearing in mind the date of divergence of their ancestors, about 293 Ma ago. Recognizing that this could be an artefact resulting from multiple changes at the same site, from undetectable back mutations and from isosemantic mutations accumulating during a long period of evolution, various procedures were adopted to transform the data in order to estimate the number of such events. None of these procedures, however, eliminated the phenomenon that during the first 214 Ma since their separation the ancestral stems leading eventually to birds and mammals seem to have fixed mutations in their myoglobin at a lower rate than the average rate prevailing during the past 79 Ma, the latter being approximately one mutation in 4 Ma. It is to be expected that sampling error will produce some fluctuations in rate, but even over the relatively long period of 79 Ma the fastest rate of fixation of mutations is about three times the slowest rate and so we are inclined to discard the molecular clock as unreliable for dating divergences, at least within this span of time.

On integrating the various sections of this study we see that the changes in myoglobin have not been at random throughout the molecule. Given the constraints demanded by the functional morphology of the molecule itself and the constraints of the genetic code, it is to be expected that both will contribute to parallel change in different lineages. In the reconstructed pathways of evolution of myoglobin about 50 % of the changes were found in parallel in other lineages. There is an indication, provided by the larger number of arginine residues present in aquatic forms, that some of these parallel changes may be correlated with mode of life. The adaptive significance of the several parallel changes between the cetaceans and the pinnipeds certainly deserves physiological investigation.

Regardless of the causes of parallel evolution at the molecular level this phenomenon has contributed to unexpected similarities between myoglobins and has led to difficulties in phylogenetic reconstruction of a nature already familiar to comparative anatomists.

1. Introduction

In a previous paper (Romero-Herrera, Lehmann, Joysey & Friday 1973) we considered the evidence available from the myoglobin sequences of 18 mammal species, and sought to combine this with a phylogenetic pattern which was itself based on the evidence of comparative anatomy and the fossil record. We reconstructed a possible pathway of the molecular evolution of myoglobin between each of these 18 living species and the hypothetical ancestor common to all. Aiming at keeping the number of changes to a minimum we found that nearly all of the myoglobin evidence could be integrated with the accepted pattern of mammalian phylogeny based on non-molecular evidence, the only exception being that from the Cetacea. We also investigated the rate of fixation of mutations in various lines of descent within the mammals and found that there were fluctuations in rate within lineages and differences in the overall rate between lineages. These different overall rates could themselves be averaged and, in the context of molecular evolution, we proposed that the term 'average' rate is more appropriate than 'constant' rate and suggested that overall rates for separate lineages are likely to be more informative than the average rate.

In that earlier paper we rigorously based our phylogenetic chart on the fossil record and when the available evidence failed to resolve the time sequence of two successive dichotomies,

then the branching points were superimposed on the phylogenetic chart and shown as a trichotomy. It was reassuring to find that in most cases the myoglobin evidence resolved the probable sequence of branching in a manner consistent with the relationships judged from comparative anatomy. For example, among the Primates, the evidence of myoglobin lent support to the classical view that apes and Old World monkeys have characters in common which are not present in New World monkeys. On the other hand, the myoglobin evidence indicated a closer kinship between the Carnivora and those living forms derived through the extinct order Condylarthra than between either of these groups and the Primates. However, this was based on a single member of the Carnivora, the harbour seal, and with the addition of evidence from the myoglobin of other Carnivora a different phylogenetic pattern has now emerged as the most economical solution.

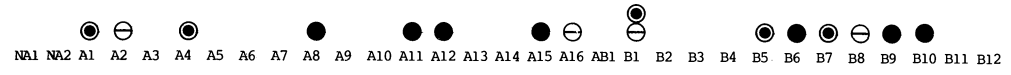
In a more recent publication (Lehmann, Romero-Herrera, Joysey & Friday 1974) we attempted to use the evidence of myoglobin to throw some light on the phylogenetic position of the tree shrew (*Tupaia*) because its systematic position is contended; it is included within the Primates by some and not so by others. The introduction of the evidence of this single animal led to several improvements in our previous reconstruction of the pathways of change in myoglobin, but it was not possible to assign the tree shrew to a definite systematic position because the two alternative interpretations could be arrived at through an identical set of

TABLE 1

Man, *Homo sapiens* (Romero-Herrera & Lehmann 1974*a*).
Chimpanzee, *Pan troglodytes* (Romero-Herrera & Lehmann 1972*a*).
Gorilla, *Gorilla gorilla beringei* (Romero-Herrera *et al.* 1975).
Gibbon, *Hylobates agilis* (Romero-Herrera & Lehmann 1971).
Baboon, *Papio anubis* (Romero-Herrera & Lehmann 1972*b*).
Macaque, *Macaca fascicularis* (Romero-Herrera & Lehmann 1972*b*).
Woolly monkey, *Lagothrix lagothricha* (Romero-Herrera & Lehmann 1973*a*).
Squirrel monkey, *Saimiri sciureus* (Romero-Herrera & Lehmann 1973*a*).
Marmoset, *Callithrix jacchus* (Romero-Herrera & Lehmann 1973*a*).
Galago, *Galago crassicaudatus* (Romero-Herrera & Lehmann 1973*b*).
Potto, *Perodicticus potto* (Romero-Herrera & Lehmann 1975*a*).
Slow loris, *Nycticebus coucang* (Romero-Herrera *et al.* 1976).
Sportive lemur, *Lepilemur mustelinus* (Romero-Herrera & Lehmann 1973*b*).
Sperm whale, *Physeter catodon* (Edmundson 1965).
Dolphin, *Delphinus delphis* (Karadjova *et al.* 1970; Kluh & Bakardjieva 1971).
Porpoise, *Phocaena phocaena* (Bradshaw & Gurd 1969).
Horse, *Equus caballus* (Dautrevaux *et al.* 1969).
Zebra, *Equus burchelli* (Darbre *et al.* 1975).
Ox, *Bos taurus* (Han *et al.* 1970).
Sheep, *Ovis aries* (Han *et al.* 1972; Vötsch & Anderer 1972).
Harbour seal, *Phoca vitulina* (Bradshaw & Gurd 1969).
Sea lion, *Salophus californianus* (Vigna *et al.* 1974).
Dog, *Canis familiaris* (Y. Boulanger, 1975, personal communication).
Hunting dog, *Lycaon pictus* (Romero-Herrera *et al.* 1976).
Badger, *Meles meles* (Ducastaing 1973).
Tree shrew, *Tupaia glis belangeri* (Romero-Herrera & Lehmann 1974*b*).
Hedgehog, *Erinaceus europaeus* (Romero-Herrera *et al.* 1975).
Kangaroo, *Megaleia rufa* (Air & Thompson 1971).
Opossum, *Didelphis marsupialis* (Romero-Herrera & Lehmann 1975*b*).
Chicken, *Gallus gallus* (Deconinck *et al.* 1975).
Penguin, *Aptenodytes forsteri* (Peiffer 1973).
Aplysia limacina (Tentori *et al.* 1973).
Lamprey, *Petromyzon marinus* (Li & Riggs 1970).
Glycera dibranchiata (Imamura *et al.* 1972).

TABLE 2. MYOGLOBIN SEQUENCES

(See footnote on page 68. Symbols as in figure 3.)



	NA1	NA2	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	AB1	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
MAN			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Val	Trp	Gly	Lys	Val	Glu	Ala	Asp	Ile	Pro	Gly	His	Gly	Gln	Glu	Val	Leu	Ile	Arg
CHIMPANZEE			Gly	(Leu	Ser	Asp	Gly	Glu)	Trp	Gln	(Leu	Val	Leu	Asn	Val	Trp	Gly)	Lys	(Val	Glu	Ala	Asp	Ile	Pro	Gly	His	Gly	Gln	Glu	Val	Leu	Ile)	Arg
GORILLA			(Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Val	Trp)	Gly	Lys	(Val	Glu	Ala	Asp)	Ile	Ser	(Gly	His	Gly	Gln	Glu	Val	Leu)	Ile	Arg
GIBBON			(Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Val	Trp)	Gly	Lys	(Val	Glu)	Ala	Asp	Ile	Pro	Ser	(His	Gly	Gln	Glu	Val	Leu)	Ile	Arg
BABOON			(Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Val	Trp)	Gly	Lys	(Val	Glu	Ala	Asp	Ile	Pro	Ser	His)	Gly	Gln	(Glu	Val)	Leu	Ile	Arg
MACAQUE			(Gly	Leu	Ser	Asp	Gly	Glu)	Trp	(Gln	Leu	Val	Leu	Asn	Val	Trp	Gly)	Lys	(Val	Glu	Ala	Asp	Ile	Pro	Ser	His)	Gly	Gln	(Glu	Val)	Leu	Ile	Arg
WOOLLY MONKEY			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Ile	Trp	Gly	Lys	(Val	Glu	Ala	Asp	Ile	Pro	Ser	His)	Gly	Gln	Glu	Val	Leu	Ile	Ser
SQUIRREL MONKEY			(Gly	Leu	Ser	Asp	Gly	Glu)	Trp	(Gln	Leu	Val	Leu)	Asn	Ile	(Trp	Gly)	Lys	(Val	Glu)	Ala	(Asp	Ile	Pro	Ser	His	Gly	Gln	Glu	Val	Leu	Ile	Ser
MARMOSET			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	(Leu	Val	Leu	Asn	Val	Trp)	Gly	Lys	(Val	Glu)	Ala	Asp	Ile	Pro	Ser	(His	Gly	Gln	Glu	Val	Leu)	Ile	Ser
GALAGO			(Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu)	Lys	Ile	(Trp	Gly)	Lys	(Val	Glu	Ala	Asp)	Leu	Ala	(Gly	His	Gly	Gln)	Asp	Val	Leu	Ile	Arg
POTTO			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Ser	Val	Leu	Asn	Val	Trp	Gly	Lys	(Val	Glu	Ala	Asp)	Leu	Ala	(Gly	His	Gly	Gln	Glu)	Ile	Leu	Ile	Arg
SLOW LORIS			(Gly	Leu	Ser	Asp	Gly	Glu)	Trp	Gln	Ser	(Val	Leu	Asn	Val	Trp	Gly)	Lys	(Val	Glu	Ala	Asp)	Leu	Ala	(Gly	His	Gly	Gln	Glu)	Ile	Leu	Ile	Arg
SPORTIVE LEMUR			(Gly	Leu	Ser	Asp	Gly	Glu)	Trp	Gln	(Leu	Val	Leu	Asn	Val	Trp	Gly)	Lys	(Val	Glu	Ala	Asp)	Val	Gly	(Gly	His	Gly	Gln	Glu	Val	Leu)	Ile	Arg
SPERM WHALE			Val	Leu	Ser	Glu	Gly	Glu	Trp	Gln	Leu	Val	Leu	His	Val	Trp	Ala	Lys	Val	Glu	Ala	Asp	Val	Ala	Gly	His	Gly	Gln	Asp	Ile	Leu	Ile	Arg
DOLPHIN			Val	Leu	Ser	Glu	Gly	Glu	Trp	Gln	Leu	Val	Leu	His	Val	Trp	Ala	Lys	Val	Glu	Ala	Asp	Val	Ala	Gly	His	Gly	Gln	Asp	Ile	Leu	Ile	Arg
PORPOISE			Gly	Leu	Ser	Glu	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Val	Trp	Gly	Lys	Val	Glu	Ala	Asp	Leu	Ala	Gly	His	Gly	Gln	Asp	Val	Leu	Ile	Arg
HORSE			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Gln	Val	Leu	Asn	Val	Trp	Gly	Lys	Val	Glu	Ala	Asp	Ile	Ala	Gly	His	Gly	Gln	Glu	Val	Leu	Ile	Arg
ZEBRA			(Gly	Leu	Ser	Asp	Gly	Glu)	Trp	(Gln	Gln	Val	Leu	Asn	Val	Trp	Gly)	Lys	(Val	Glu	Ala	Asp	Ile	Ala	Gly	His	Gly	Gln	Glu	Val	Leu	Ile)	Arg
OX			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Ala	Val	Leu	Asn	Ala	Trp	Gly	Lys	Val	Glu	Ala	Asp	Val	Ala	Gly	His	Gly	Gln	Glu	Val	Leu	Ile	Arg
SHEEP			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Ala	Trp	Gly	Lys	Val	Glu	Ala	Asp	Val	Ala	Gly	His	Gly	Gln	Glu	Val	Leu	Ile	Arg
HARBOUR SEAL			Gly	Leu	Ser	Asp	Gly	Glu	Trp	His	Leu	Val	Leu	Asn	Val	Trp	Gly	Lys	Val	Glu	Thr	Asp	Leu	Ala	Gly	His	Gly	Gln	Glu	Val	Leu	Ile	Arg
SEA LION			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Ile	Trp	Gly	Lys	Val	Glu	Ala	Asp	Leu	Val	Gly	His	Gly	Gln	Glu	Val	Leu	Ile	Arg
DOG			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Ile	Trp	Gly	Lys	Val	Glu	Thr	Asp	Leu	Ala	Gly	His	Gly	Gln	Glu	Val	Leu	Ile	Arg
HUNTING DOG			(Gly	Leu	Ser	Asp	Gly	Glu)	Trp	Gln	(Leu	Val	Leu	Asn	Ile	Trp	Gly)	Lys	(Val	Glu	Thr	Asp	Leu	Ala	Gly	His	Gly	Gln	Glu	Val	Leu)	Ile	Arg
BADGER			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Val	Trp	Gly	Lys	Val	Glu	Ala	Asp	Leu	Ala	Gly	His	Gly	Gln	Glu	Val	Leu	Ile	Arg
TREESHREW			(Gly	Leu	Ser	Asp	Gly	Glu)	Trp	Gln	(Leu	Val	Leu	Asn	Val	Trp	Gly)	Lys	(Val	Glu	Ala)	Asp	Val	Ala	(Gly	His	Gly	Gln)	Glu	Val	Leu	Ile	Arg
HEDGEHOG			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Val	Trp	Gly	Lys	Val	Glu	Ala	Asp	Ile	Pro	Gly	His	Gly	Gln	Glu	Val	Leu	Ile	Arg
KANGAROO			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Ile	Trp	Gly	Lys	Val	Glu	Thr	Asp	Glu	Gly	Gly	His	Gly	Lys	Asp	Val	Leu	Ile	Arg
OPOSSUM			Gly	Leu	Ser	Asp	Gly	Glu	Trp	Gln	Leu	Val	Leu	Asn	Ala	Trp	Gly	Lys	Val	Glu	Ala	Asp	Ile	Pro	Gly	His	Gly	Gln	Glu	Val	Leu	Ile	Arg
CHICKEN			Gly	Leu	Ser	Asp	Gln	Glu	Trp	Gln	Gln	Val	Leu	Thr	Ile	Trp	Gly	Lys	Val	Glu	Ala	Asp	Ile	Ala	Gly	His	Gly	His	Glu	Val	Leu	Met	Arg
PENGUIN			Gly	Leu	Asn	Asp	Gln	Glu	Trp	Gln	Gln	Val	Leu	Thr	Met	Trp	Gly	Lys	Val	Glu	Ser	Asp	Leu	Ala	Gly	His	Gly	His	Ala	Val	Leu	Met	Arg
APLYSIA			Ser	Leu	Ser	Ala	Ala	Glu	Ala	Asp	Leu	Ala	Gly	Lys	Ser	Trp	Ala	Pro	Val	Phe	Ala	Asn	Lys	Asn	Ala	Asn	Gly	Ala	Asp	Phe	Leu	Val	Ala
LAMPREY Hb	*	Ala	Leu	Ser	Ala	Ala	Glu	Lys	Thr	Lys	Ile	Arg	Ser	Ala	Trp	Ala	Pro	Val	Tyr	Ser	Asn	Tyr	Glu	Thr	Ser	Gly	Val	Asp	Ile	Leu	Val	Lys
		Pro																			Thr											
GLYCERA Hb		Gly	Leu	Ser	Ala	Ala	Gln	Arg	Gln	Val	Ile	Ala	Ala	Thr	Trp	Lys	Asp	Ile	Ala	Gly	Asp	Gly	Ala	Gly	Val	Gly	Lys	Asp	Cys	Leu	Ile	Lys
																				Asn	Asn											
	*	Pro	Ile	Val	Asp	Thr	Gly	Ser	Val																							

	B13	B14	B15	B16	C1	C2	C3	C4	C5	C6	C7	CD1	CD2	CD3	CD4	CD5	CD6	CD7	CD8	D1	D2	D3	D4	D5	D6	D7	E1	E2	E3	E4	E5
	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62
MAN	Leu	Phe	Lys	Gly	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Ser	Glu	Asp	Glu	Met	Lys	Ala	Ser	Glu	Asp	Leu	Lys
CHIMPANZEE	(Leu	Phe)	Lys	(Gly	His	Pro	Glu	Thr	Leu	Glu)	Lys	(Phe	Asp)	Lys	Phe	Lys	(His	Leu)	Lys	(Ser	Glu	Asp	Glu	Met)	Lys	(Ala	Ser	Glu	Asp	Leu)	Lys
GORILLA	(Leu	Phe)	Lys	(Gly	His	Pro	Glu	Thr	Leu	Glu)	Lys	(Phe	Asp)	Lys	Phe	Lys	(His	Leu)	Lys	(Ser	Glu	Asp	Glu	Met)	Lys	(Ala	Ser	Glu	Asp	Leu)	Lys
GIBBON	(Leu	Phe)	Lys	(Gly	His	Pro	Glu	Thr	Leu	Glu)	Lys	(Phe	Asp	Lys	Phe)	Lys	(His	Leu)	Lys	(Ser	Glu	Asp	Glu	Met)	Lys	(Ala	Ser	Glu	Asp	Leu	Lys)
BABOON	(Leu	Phe)	Lys	(Gly	His	Pro	Glu	Thr	Leu	Glu)	Lys	(Phe	Asp)	Lys	Phe	Lys	(His	Leu)	Lys	(Ser	Glu	Asp	Glu	Met)	Lys	(Ala	Ser	Glu	Asp	Leu)	Lys
MACAQUE	(Leu	Phe)	Lys	(Gly	His	Pro	Glu	Thr	Leu	Glu)	Lys	(Phe	Asp)	Lys	Phe	Lys	(His	Leu)	Lys	(Ser	Glu	Asp	Glu)	Met	Lys	Ala	Ser	Glu	Asp	Leu	Lys
WOOLLY MONKEY	Leu	Phe	Lys	(Gly	His	Pro	Glu	Thr)	Leu	Glu	Lys	(Phe	Asp)	Lys	Phe	Lys	His	Leu	Lys	(Ser	Glu	Asp	Glu)	Met	Lys	Ala	Ser	Glu	Glu	Leu	Lys
SQUIRREL MONKEY	Leu)	Phe	Lys	(Gly	His	Pro	Glu	Thr	Leu	Glu)	Lys	(Phe	Asp)	Lys	Phe	Lys	(His	Leu)	Lys	(Ser	Glu	Asp	Glu	Met)	Lys	(Ala	Ser	Glu	Glu	Leu)	Lys
MARMOSET	Leu	Phe	Lys	(Gly	His	Pro	Glu	Thr	Leu	Glu)	Lys	(Phe	Asp)	Lys	Phe	Lys	(His	Leu)	Lys	(Ser	Glu	Asp	Glu	Met)	Lys	(Ala	Ser	Glu	Glu	Leu	Lys
GALAGO	Leu	Phe	Thr	Ala	(His	Pro	Glu	Thr	Leu	Glu)	Lys	(Phe	Asp)	Lys	Phe	Lys	Asn	Leu	Lys	Thr	Ala	Asp	Glu	Met	Lys	(Ala	Ser	Glu	Asp	Leu)	Lys
POTTO	Leu	Phe	Thr	Ala	(His	Pro	Glu	Thr	Leu)	Glu	Lys	Phe	Asp	Lys	Phe	Lys	(Asn	Leu)	Lys	Thr	Pro	(Asp	Glu	Met)	Lys	(Ala	Ser	Glu	Asp	Leu)	Lys
SLOW LORIS	Leu	Phe	Thr	Ala	(His	Pro	Glu	Thr	Leu)	Glu	Lys	(Phe	Asp)	Lys	Phe	Lys	(Asn	Leu)	Lys	Thr	Pro	(Asp	Glu)	Met	Lys	(Ala	Ser	Glu)	Asp	Leu	Lys
SPORTIVE LEMUR	Leu	Phe	Thr	(Gly	His	Pro	Glu	Thr	Leu	Glu)	Lys	(Phe	Asp)	Lys	Phe	Lys	(His	Leu)	Lys	Thr	Ala	Asp	Glu	Met	Lys	(Ala	Ser	Glu	Asp	Leu)	Lys
SPERM WHALE	Leu	Phe	Lys	Ser	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Arg	Phe	Lys	His	Leu	Lys	Thr	Glu	Ala	Glu	Met	Lys	Ala	Ser	Glu	Asp	Leu	Lys
DOLPHIN	Leu	Phe	Lys	Gly	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Thr	Glu	Ala	Asp	Met	Lys	Ala	Ser	Glu	Asx	Leu	Lys
PORPOISE	Leu	Phe	Lys	Gly	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Thr	Glu	Ala	Glu	Met	Lys	Ala	Ser	Glu	Asp	Leu	Lys
HORSE	Leu	Phe	Thr	Gly	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Thr	Glu	Ala	Glu	Met	Lys	Ala	Ser	Glu	Asp	Leu	Lys
ZEBRA	(Leu	Phe	Thr	Gly	His	Pro	Glu	Thr	Leu	Glu)	Lys	(Phe	Asp)	Lys	Phe	Lys	(His	Leu)	Lys	(Thr	Glu	Ala	Glu	Met)	Lys	(Ala	Ser	Glu	Asp	Leu)	Lys
OX	Leu	Phe	Thr	Gly	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Thr	Glu	Ala	Glu	Met	Lys	Ala	Ser	Glu	Asp	Leu	Lys
SHEEP	Leu	Phe	Thr	Gly	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Thr	Glu	Ala	Glu	Met	Lys	Ala	Ser	Glu	Asp	Leu	Lys
HARBOUR SEAL	Leu	Phe	Lys	Ser	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Ser	Glu	Asp	Asp	Met	Arg	Arg	Ser	Glu	Asp	Leu	Arg
SEA LION	Leu	Phe	Lys	Gly	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Ser	Glu	Asp	Glu	Met	Lys	Arg	Ser	Glu	Asp	Leu	Lys
DOG	Leu	Phe	Lys	Asn	His	Pro	Glu	Thr	Leu	Asp	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Thr	Glu	Asp	Glu	Met	Lys	Gly	Ser	Glu	Asp	Leu	Lys
HUNTING DOG	(Leu	Phe)	Lys	(Asn	His	Pro	Glu	Thr	Leu	Asp)	Lys	(Phe	Asp)	Lys	Phe	Lys	(His	Leu)	Lys	(Thr	Glu	Asp	Glu	Met)	Lys	(Gly	Ser	Glu	Asp	Leu)	Lys
BADGER	Leu	Phe	Lys	Gly	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Ser	Glu	Asp	Glu	Met	Lys	Gly	Ser	Glu	Asp	Leu	Lys
TREESHREW	Leu	Phe	Lys	(Gly	His	Pro	Glu	Thr	Leu	Glu)	Lys	(Phe	Asp)	Lys	Phe	Lys	(His	Leu)	Lys	Thr	(Glu	Asp	Glu	Met)	Lys	(Ala	Ser	Glu	Asp)	Leu	Lys
HEDGEHOG	Leu	Phe	Lys	Asp	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Ser	Glu	Asp	Glu	Met	Lys	Ser	Ser	Glu	Asp	Leu	Lys
KANGAROO	Leu	Phe	Lys	Gly	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Ser	Glu	Asp	Glu	Met	Lys	Ala	Ser	Glu	Asp	Leu	Lys
OPOSSUM	Leu	Phe	Lys	Gly	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys	Ser	Glu	Asp	Glu	Met	Lys	Ala	Ser	Glu	Asp	Leu	Lys
CHICKEN	Leu	Phe	His	Asp	His	Pro	Glu	Thr	Leu	Asp	Arg	Phe	Asp	Lys	Phe	Lys	Gly	Leu	Lys	Thr	Glu	Pro	Asp	Met	Lys	Gly	Ser	Glu	Asp	Leu	Lys
PENGUIN	Leu	Phe	Lys	Ser	His	Pro	Glu	Thr	Met	Asp	Arg	Phe	Asp	Lys	Phe	Arg	Gly	Leu	Lys	Thr	Pro	Glu	Asp	Met	Arg	Gly	Ser	Glu	Asp	Met	Lys
APLYSIA	Leu	Phe	Glu	Lys	Phe	Pro	Asp	Ser	Ala	Asn	Phe	Phe	Ala	Asp	Phe	Lys	Gly		Lys	Ser	Val	Ala	Asp	Ile	Lys	Ala	Ser	Pro	Lys	Leu	Arg
LAMPREY Hb	Phe	Phe	Thr	Ser	Thr	Pro	Ala	Ala	(Gln	Glu)	Phe	Phe	Pro	Lys	Phe	Lys	Gly	Leu	Thr	Thr	Ala	Asp	Gln	Leu	Lys	Lys	Ser	Ala	Asp	Val	Arg
GLYCERA Hb	His	Leu	Ser	Ala	His	Pro	Gln	Met	Ala	Ala	Val	Phe	Gly		Phe	Ser	Gly	Ala	Ser								Asp	Pro	Ala	Val	Ala

TABLE 2 (*cont.*)

	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15	E16	E17	E18	E19	E20	EF1	EF2	EF3	EF4	EF5	EF6	EF7	EF8	F1	F2	F3	F4	F5	F6	F7
	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92
MAN	Lys	His	Gly	Ala	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Ile	Lys	Pro	Leu	Ala	Gln	Ser
CHIMPANZEE	Lys	His	Gly	Ala	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Ile	Lys	Pro	Leu	Ala	Gln	Ser
GORILLA	Lys	His	Gly	Ala	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Ile	Lys	Pro	Leu	Ala	Gln	Ser
GIBBON	Lys	His	Gly	Ala	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Ile	Lys	Pro	Leu	Ala	Gln	Ser
BABOON	Lys	His	Gly	Ala	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Ile	Lys	Pro	Leu	Ala	Gln	Ser
MACAQUE	Lys	His	Gly	Val	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Ile	Lys	Pro	Leu	Ala	Gln	Ser
WOOLLY MONKEY	Lys	His	Gly	Val	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	Gln	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
SQUIRREL MONKEY	Lys	His	Gly	Thr	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	Gln	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
MARMOSET	Lys	His	Gly	Val	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
GALAGO	Lys	His	Gly	Val	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	Gln	His	Glu	Ala	Glu	Ile	Lys	Pro	Leu	Ala	Gln	Ser
POTTO	Lys	His	Gly	Val	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Ile	Lys	Pro	Leu	Ala	Gln	Ser
SLOW LORIS	Lys	His	Gly	Val	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	Gln	His	Glu	Ala	Glu	Ile	Lys	Pro	Leu	Ala	Gln	Ser
SPORTIVE LEMUR	Lys	His	Gly	Thr	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	Gln	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
SPERM WHALE	Lys	His	Gly	Val	Thr	Val	Leu	Thr	Ala	Leu	Gly	Ala	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
DOLPHIN	Lys	His	Gly	Asp	Thr	Val	Leu	Thr	Ala	Leu	Gly	Ala	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Asp	Ala	Glx	Leu	Lys	Pro	Leu	Ala	Gln	Ser
PORPOISE	Lys	His	Gly	Asn	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Asn	Leu	Lys	Pro	Leu	Ala	Gln	Ser
HORSE	Lys	His	Gly	Thr	Val	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
ZEBRA	Lys	His	Gly	Thr	Val	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
OX	Lys	His	Gly	Asn	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Val	Lys	His	Leu	Ala	Glu	Ser
SHEEP	Lys	His	Gly	Asn	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Val	Lys	His	Leu	Ala	Glu	Ser
HARBOUR SEAL	Lys	His	Gly	Asn	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
SEA LION	Lys	His	Gly	Lys	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Asp	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
DOG	Lys	His	Gly	Asn	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
HUNTING DOG	Lys	His	Gly	Asn	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
BADGER	Lys	His	Gly	Asn	Thr	Val	Leu	Thr	Ala	Leu	Gly	Ala	Ile	Leu	Lys	Lys	Lys	Gly	His	Gln	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
TREESHREW	Lys	His	Gly	Asn	Thr	Val	Leu	Ser	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	Gln	His	Glu	Ala	Glu	Ile	Lys	Pro	Leu	Ala	Gln	Ser
HEDGEHOG	Lys	His	Gly	Thr	Thr	Val	Leu	Thr	Ala	Leu	Gly	Gly	Ile	Leu	Lys	Lys	Lys	Gly	Gln	His	Glu	Ala	Gln	Leu	Ala	Pro	Leu	Ala	Gln	Ser
KANGAROO	Lys	His	Gly	Ile	Thr	Val	Leu	Thr	Ala	Leu	Gly	Asn	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
OPOSSUM	Lys	His	Gly	Ala	Thr	Val	Leu	Thr	Ala	Leu	Gly	Asn	Ile	Leu	Lys	Lys	Lys	Gly	Asn	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ala	Gln	Ser
CHICKEN	Lys	His	Gly	Gln	Thr	Val	Leu	Thr	Ala	Leu	Gly	Ala	Gln	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Asp	Leu	Lys	Pro	Leu	Ala	Gln	Ser
PENGUIN	Lys	His	Gly	Val	Thr	Val	Leu	Thr		Leu	Gly	Gln	Ile	Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Leu	Lys	Pro	Leu	Ser	Gln	Thr
APLYSIA	Asp	Val	Ser	Ser	Arg	Ile	Phe	Thr	Arg	Leu	Asn	Glu	Phe	Val	Asn	Asp	Ala	Ala	Asn	Ala*	Ser	Ala	Met	Leu	Ser	Gln	Phe	Ala	Lys	Glu
LAMPREY Hb	Trp	His	Ala	Glu	Arg	Ile	Ile	Asn	Ala	Val	Asn	Asp	Ala	Val	Ala	Ser	Met	Asp	Asp	Thr	Glu	Lys	Ser	Met	Lys	Asn	Leu	Ser	Gly	Lys
GLYCERA Hb	Asp	Leu	Gly	Ala	Lys	Val	Leu	Ala	Glx	Ile	Gly	Val	Ala	Val				Ser	His	Leu	Gly	Asp	Glx	Met	Lys	Ala	Val	Gly	Val	Arg

Met

Gly Lys Met

*APLYSIA HAS AN INSERTION OF Gly Lys Met

	F8	F9	FG1	FG2	FG3	FG4	FG5	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	GH1	GH2	GH3	GH4	GH5
	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123
MAN	His	Ala	Thr	Lys	His	Lys	Ile	Pro	Val	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Glu	Cys	Ile	Ile	Gln	Val	Leu	Gln	Ser	Lys	His	Pro	Gly	Asp	Phe
CHIMPANZEE	His	Ala	Thr)	Lys	(His	Lys	Ile	Pro)	Val	Lys	(Tyr	Leu	Glu	Phe	Ile	Ser	Glu)	Cys	(Ile	Ile	Gln)	Val	Leu	His	Ser	Lys	(His	Pro	Gly	Asp	Phe
GORILLA	His	Ala	Thr)	Lys	His	Lys	(Ile	Pro	Val)	Lys	(Tyr	Leu	Glu	Phe	Ile	Ser	Glu	Cys	Ile	Ile	Gln	Val	Leu	Gln	Ser)	Lys	(His	Pro	Gly	Asp	Phe
GIBBON	His	Ala	Thr)	Lys	(His	Lys	Ile	Pro)	Val	Lys	(Tyr	Leu	Glu	Phe	Ile	Ser	Glu)	Cys	(Ile	Ile	Gln	Val	Leu	Gln	Ser)	Lys	(His	Pro	Gly	Asp	Phe
BABOON	His	Ala	Thr)	Lys	(His	Lys	Ile	Pro)	Val	Lys	Tyr	(Leu	Glu	Leu	Ile	Ser	Glu)	Ser	Ile	Ile	Gln	(Val	Leu	Gln	Ser)	Lys	(His	Pro	Gly	Asp	Phe
MACAQUE	His	Ala	Thr)	Lys	(His	Lys	Ile	Pro)	Val	Lys	Tyr	(Leu	Glu)	Leu	(Ile	Ser	Glu)	Ser	Ile	Ile	Gln	Val	Leu	(Gln	Ser)	Lys	(His	Pro	Gly	Asp	Phe
WOOLLY MONKEY	His	Ala	Thr)	Lys	(His	Lys	Ile	Pro)	Val	Lys	Tyr	(Leu	Glu)	Phe	Ile	Ser	Asp	Ala	Ile	Ile	His	Ala	(Leu	Gln)	Lys	Lys	(His	Pro	Gly	Asp	Phe
SQUIRREL MONKEY	His	Ala	Thr)	Lys	(His	Lys	Ile	Pro)	Val	Lys	Tyr	Leu	Glu	Leu	(Ile	Ser	Asp)	Ala	Ile	Val	His	Val	Leu	Gln	Lys	Lys	(His	Pro	Gly	Asp	Phe
MARMOSET	His	Ala	Thr)	Lys	(His	Lys	Ile	Pro)	Val	Lys	(Tyr	Leu	Glu	Phe)	Ile	Ser	Asp	Ala	Ile	Val	His	Val	(Leu	Gln)	Lys	Lys	(His	Pro	Gly	Asp	Phe
GALAGO	His	Ala	Thr)	Lys	(His	Lys	Ile	Pro)	Val	Lys	(Tyr	Leu	Glu	Phe	Ile	Ser	Glu)	Ala	Ile	Ile	His	Val	Leu	Gln	Asn	Lys	His	Ser	(Gly	Asp)	Phe
POTTO	His	Ala	Thr)	Lys	His	Lys	(Ile	Pro	Val)	Lys	Tyr	(Leu	Glu	Phe	Ile	Ser	Glu)	Ala	Ile	Ile	His	Val	Leu	(Gln	Ser)	Lys	(His	Pro	Gly	Asp	Phe
SLOW LORIS	His	Ala	Thr)	Lys	(His	Lys	Ile	Pro	Val)	Lys	Tyr	(Leu	Glu)	Phe	Ile	Ser	Gly	Ala	Ile	Ile	His	(Val	Leu	Gln	Ser)	Lys	(His	Pro	Gly	Asp)	Phe
SPORTIVE LEMUR	His	Ala	Thr)	Lys	(His	Lys	Ile	Pro)	Ile	Lys	(Tyr	Leu	Glu	Phe)	Ile	Ser	Asp	Ala	Ile	Val	His	Val	Leu	(His	Ser)	Lys	His	Pro	Ala	Glu	(Phe
SPERM WHALE	His	Ala	Thr	Lys	His	Lys	Ile	Pro	Ile	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Glu	Ala	Ile	Ile	His	Val	Leu	His	Ser	Arg	His	Pro	Ala	Gln	Phe
DOLPHIN	His	Ala	Thr	Lys	His	Lys	Ile	Pro	Ile	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Glu	Ala	Ile	Ile	His	Val	Leu	His	Ser	Arg	His	Pro	Ala	Glu	Phe
PORPOISE	His	Ala	Thr	Lys	His	Lys	Ile	Pro	Ile	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Glu	Ala	Ile	Ile	His	Val	Leu	His	Ser	Arg	His	Pro	Ala	Glu	Phe
HORSE	His	Ala	Thr	Lys	His	Lys	Ile	Pro	Ile	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Asp	Ala	Ile	Ile	His	Val	Leu	His	Ser	Lys	His	Pro	Gly	Asp	Phe
ZEBRA	His	Ala	Thr)	Lys	His	Lys	(Ile	Pro	Ile)	Lys	Tyr	(Leu	Glu)	Phe	(Ile	Ser	Asp)	Ala	(Ile	Ile	His	Val	Leu	His)	Ser	Lys	(His	Pro	Gly	Asp	Phe
OX	His	Ala	Asn	Lys	His	Lys	Val	Pro	Ile	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Asp	Ala	Ile	Ile	His	Val	Leu	His	Ala	Lys	His	Pro	Ser	Asn	Phe
SHEEP	His	Ala	Asn	Lys	His	Lys	Val	Pro	Ile	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Asp	Ala	Ile	Ile	His	Val	Leu	His	Ala	Lys	His	Pro	Ser	Asn	Phe
HARBOUR SEAL	His	Ala	Thr	Lys	His	Lys	Ile	Pro	Ile	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Glu	Ala	Ile	Ile	His	Val	Leu	Gln	Ser	Lys	His	Pro	Gly	Asp	Phe
SEA LION	His	Ala	Thr	Lys	His	Lys	Ile	Pro	Ile	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Glu	Ala	Ile	Ile	His	Val	Leu	Gln	Ser	Lys	His	Pro	Gly	Asp	Phe
DOG	His	Ala	Thr	Lys	His	Lys	Ile	Pro	Val	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Asp	Ala	(Ile	Ile	Gln	Val)	Leu	Gln	Ser	Lys	His	Ser	Gly	Asn	Phe
HUNTING DOG	His	Ala	Thr)	Lys	His	Lys	(Ile	Pro	Val)	Lys	Tyr	(Leu	Glu	Phe	Ile	Ser	Asp	Ala	Ile	Ile	Gln	Val	Leu)	Gln	Asn	Lys	(His	Ser	Gly	Asp)	Phe
BADGER	His	Ala	Thr	Lys	His	Lys	Val	Pro	Ile	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Asp	Ala	Ile	Ala	Gln	Val	Leu	Gln	Ser	Lys	His	Pro	Gly	Asn	Phe
TREESHREW	His)	Ala	Thr	Lys	(His	Lys	Ile	Pro)	Val	Lys	(Tyr	Leu	Glu	Phe	Ile	Ser	Glu)	Ala	(Ile	Ile	Gln	Val	Leu)	Gln	Ser	Lys	(His	Pro	Gly	Asp)	Phe
HEDGEHOG	His	Ala	Asn	Lys	His	Lys	Ile	Pro	Val	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Glu	Ala	Ile	Ile	Gln	Val	Leu	Lys	Ser	Lys	His	Ala	Gly	Asp	Phe
KANGAROO	His	Ala	Thr	Lys	His	Lys	Ile	Pro	Val	Gln	Phe	Leu	Glu	Phe	Ile	Ser	Asp	Ala	Ile	Ile	Gln	Val	Ile	Gln	Ser	Lys	His	Ala	Gly	Asn	Phe
OPOSSUM	His	Ala	Thr	Lys	His	Lys	Ile	Ser	Val	Gln	Phe	Leu	Glu	Phe	Ile	Ser	Glu	Ala	Ile	Ile	Gln	Val	Ile	Gln	Ser	Lys	His	Pro	Gly	Asp	Phe
CHICKEN	His	Ala	Thr	Lys	His	Lys	Ile	Pro	Val	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Glu	Val	Ile	Ile	Lys	Val	Ile	Ala	Glu	Lys	His	Ala	Ala	Asp	Phe
PENGUIN	His	Ala	Thr	Lys	His	Lys	Val	Pro	Val	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Glu	Ala	Ile	Met	Lys	Val	Ile	Ala	Gln	Lys	His	Ala	Ser	Asn	Phe
APLYSIA	His	Val	Gly		Phe	Gly	Val	Gly	Ser	Ala	Gln	Phe	Glu	Asn	Val	Arg	Ser	Met	Phe	Pro	Gly	Phe	Val	Ala	Ser	Val	Ala	Ala	Pro	Pro	Ala
LAMPREY Hb	His	Ala	Lys	Ser	Phe	Gln	Val	Asp	Pro	Gln	Tyr	Phe	Lys	Val	Leu	Ala	Ala	Val	Ile	Ala	Asp	Thr	Val	Ala	Ala						
GLYCERA Hb	His	Lys	Asn	Lys	His		Ile	Lys	Gly	Gln	Tyr	Phe	Glu	Pro	Leu	Gly	Ala	Ser	Leu	Leu	Ser	Ala	Met	Glu	His	Arg	Ile	Gly	Gly	Lys	Met

Gly Tyr Gly

TABLE 2 (*cont.*)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23	H24	H25	H26	HC1	HC2	HC3	HC4
	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153
MAN	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Asn	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Lys	Asp	Met	Ala	Ser	Asn	Tyr	Lys	Glu	Leu	Gly	Phe	Gln	Gly
CHIMPANZEE	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Asn)	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	Lys(Asp	Met	Ala	Ser	Asn	Tyr)	Lys(Glu	Leu	Gly	Phe	Gln	Gly)			
GORILLA	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Asn)	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	Lys(Asp	Met	Ala	Ser	Asn	Tyr)	Lys(Glu	Leu	Gly	Phe	Gln	Gly)			
GIBBON	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Asn)	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	(Lys	Asp	Met	Ala	Ser	Asn	Tyr)	Lys	(Glu	Leu	Gly	Phe	Gln	Gly	
BABOON	Gly	Ala	Asp	Ala	Gln	Gly	Ala)	Met	Asn	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	Asn	Asp	Met	Ala	Ala	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)	
MACAQUE	Gly	Ala	Asp	Ala	Gln	Gly)	Ala	Met	Asn	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	Asn	Asp	Met	Ala	Ala	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)	
WOOLLY MONKEY	Gly	Ala	Asp	Ala	Gln	Gly	Ala)	Met	Lys	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	Asn	Asp	Met	Ala	Ala	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)	
SQUIRREL MONKEY	Gly	Ala	Asp	Ala	Gln)	Gly	Ala	Met	Lys	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	Asn(Asp	Met	Ala	Ala)	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)		
MARMOSET	Gly	Ala	Asp	Ala	Gln	Gly	Ala)	Met	Lys	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	Asn	Asp	Met	Ala	Ala	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)	
GALAGO	Gly	Thr	Asp	Val	(Gln	Gly)	Ala	Met	Ser	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	Asn	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)	
POTTO	Gly	Ala	Asp	Ala	Gln	Gly	Ala)	Met	Asn	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	Asn	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)	
SLOW LORIS	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Ser	Lys(Ala	Leu	Glu)	Leu	Phe	Arg	Asn	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)	
SPORTIVE LEMUR	Gly	Ala	Asp	Ala	Gln	Ala)	Ala	Met	Lys	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	Asn	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)	
SPERM WHALE	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Asn	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Lys	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	Glu	Leu	Gly	Tyr	Gln	Gly
DOLPHIN	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Asx	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Lys	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	Glu	Leu	Gly	Phe	His	Gly
PORPOISE	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Asn	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Lys	Asp	Ile	Ala	Thr	Lys	Tyr	Lys	Glu	Leu	Gly	Phe	His	Gly
HORSE	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Thr	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asn	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	Glu	Leu	Gly	Phe	Gln	Gly
ZEBRA	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Thr)	Lys(Ala	Leu	Glu	Leu	Phe)	Arg	(Asn	Asp	Ile	Ala	Ala)	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)	
OX	Ala	Ala	Asp	Ala	Gln	Gly	Ala	Met	Ser	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asn	Asp	Ala	Ala	Glu	Lys	Tyr	Lys	Val	Leu	Gly	Phe	His	Gly
SHEEP	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Ser	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asn	Asp	Met	Ala	Ala	Glu	Tyr	Lys	Val	Leu	Gly	Phe	Gln	Gly
HARBOUR SEAL	Gly	Ala	Asp	Ala	Gln	Ala	Ala	Met	Lys	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asn	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	Glu	Leu	Gly	Phe	His	Gly
SEA LION	Gly	Ala	Asp	Thr	His	Ala	Ala	Met	Lys	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asn	Asp	Ile	Ala	Ala	Lys	Tyr	Arg	Glu	Leu	Gly	Phe	Gln	Gly
DOG	His	Ala	Asp	Thr	Gln	Ala	Ala	Met	Lys	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asn	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	Glu	Leu	Gly	Phe	Gln	Gly
HUNTING DOG	His	(Ala	Asp	Thr	Glu)	Ala	Ala	Met	Lys	Lys	(Ala	Leu	Glu	Leu	Phe)	Arg	(Asn	Asp	Ile	Ala	Ala)	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)
BADGER	Ala	Ala	Glu	Ala	Gln	Gly	Ala	Met	Lys	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asn	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	Glu	Leu	Gly	Phe	Gln	Gly
TREESHREW	Gly	(Ala	Asp	Ala	Gln)	Ala	Ala	Met	Ser	Lys(Ala	Leu	Glu)	Leu	Phe	Arg	Asn	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	(Glu	Leu	Gly	Phe	Gln	Gly)	
HEDGEHOG	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Ser	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asn	Asp	Ile	Ala	Ala	Lys	Tyr	Lys	Glu	Leu	Gly	Phe	Gln	Gly
KANGAROO	Gly	Ala	Asp	Ala	Gln	Ala	Ala	Met	Lys	Lys	Ala	Leu	Glu	Leu	Phe	Arg	His	Asp	Met	Ala	Ala	Lys	Tyr	Lys	Glu	Phe	Gly	Phe	Gln	Gly
OPOSSUM	Gly	Gly	Asp	Ala	Gln	Ala	Ala	Met	Gly	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asn	Asp	Met	Ala	Ala	Lys	Tyr	Lys	Glu	Leu	Gly	Phe	Gln	Gly
CHICKEN	Gly	Ala	Asp	Ser	Gln	Ala	Ala	Met	Lys	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asp	Asp	Met	Ala	Ser	Lys	Tyr	Lys	Glu	Phe	Gly	Phe	Gln	Gly
PENGUIN	Gly	Ala	Asp	Ala	Gln	Glu	Ala	Met	Lys	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asn	Asp	Met	Ala	Ser	Lys	Tyr	Lys	Glu	Phe	Gly	Phe	Gln	Gly
APLYSIA	Gly	Ala	Asp	Ala				Trp	Thr	Lys	Leu	Phe	Gly	Leu	Ile	Ile	Asp	Ala	Leu	Lys	Ala	Ala	Gly	Lys						
LAMPREY Hb				Gly	Asp	Ala	Gly	Phe	Glu	Lys	Leu	Arg	Met	Ile	Cys	Ile	Leu		Leu	Arg	Ser	Ala	Tyr							
GLYCERA Hb	Asn	Ala	Ala	Ala	Lys	Asp	Ala	Trp	Ala	Ala	Ala	Tyr	Ala	Asp	Ile	Ser	Gly	Ala	Leu	Ile	Ser	Gly	Leu	Gln	Ser					

TABLE 3. COMPOSITION OF MYOGLOBINS BY AMINO ACIDS

	MAN	CHIMPANZEE	GORILLA	GIBBON	BABOON	MACAQUE	WOOLLY MONKEY	SQUIRREL MONKEY	MARMOSET	GALAGO	POTTO	SLOW LORIS	SPORTIVE LEMUR	SPERM WHALE	DOLPHIN	PORPOISE	HORSE & ZEBRA	OX	SHEEP	HARBOUR SEAL	SEA LION	DOG	HUNTING DOG	BADGER	TREESHREW	HEDGEHOG	KANGAROO	OPOSSUM	CHICKEN	PENGUIN	APLYSIA	LAMPREY Hb	GLYCERA Hb
Aspartic acid	8	8	8	8	8	8	8	8	8	9	8	8	8	7	9	6	8	7	7	8	9	9	10	7	8	9	9	8	12	8	9	10	8
Glutamic acid	14	14	14	14	14	14	14	14	14	12	13	12	13	14	12	14	13	14	14	14	13	12	13	14	14	13	13	14	12	12	4	6	4
Asparagine	3	3	3	3	3	3	2	2	2	3	4	3	2	1	1	4	2	5	5	3	2	5	5	4	3	3	3	4	0	3	8	4	4
Glutamine	7	6	7	7	7	7	7	7	6	7	6	7	6	5	5	4	6	3	4	3	5	7	6	8	8	8	7	8	8	8	3	4	6
Lysine	20	20	20	20	20	20	22	22	22	20	19	19	20	19	20	20	19	19	18	19	21	21	21	21	20	20	21	19	20	19	12	13	11
Arginine	2	2	2	2	2	2	1	1	1	2	2	2	2	4	3	3	2	2	2	5	4	2	2	2	2	2	2	2	3	5	4	5	3
Histidine	9	10	9	9	9	9	9	9	10	8	9	8	10	12	13	12	11	13	12	13	11	10	10	8	8	8	10	8	10	9	1	2	6
Threonine	4	4	4	4	4	4	4	5	4	7	6	6	7	5	5	6	7	5	5	5	5	7	7	4	4	4	5	4	6	7	2	9	1
Serine	7	7	8	8	8	8	7	7	7	6	6	7	5	6	5	5	5	6	6	7	6	6	5	6	7	8	6	7	6	7	13	12	10
Glycine	15	15	15	14	14	14	14	14	14	14	14	15	14	11	11	14	15	13	14	12	14	13	13	14	14	14	14	15	12	13	11	6	20
Alanine	12	12	12	12	13	12	14	13	13	14	15	15	16	17	18	15	15	19	17	14	12	12	12	16	15	14	14	15	15	12	28	21	28
Valine	7	7	7	7	7	8	6	7	9	8	7	7	8	8	7	6	7	9	9	6	6	6	6	7	8	7	6	6	7	8	10	12	10
Leucine	17	17	17	17	18	18	18	19	18	18	17	17	18	18	18	19	17	16	17	19	19	19	19	19	17	18	16	17	15	14	10	9	11
Isoleucine	8	8	8	8	8	8	8	7	6	9	9	9	7	9	9	8	9	6	6	8	9	8	8	6	8	8	9	8	7	3	4	8	8
Methionine	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2	3	3	4	8	3	4	5
Cysteine	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	
Proline	5	5	4	5	5	5	5	5	5	3	5	5	4	4	4	4	4	3	3	4	4	3	3	4	4	4	3	4	4	4	6	6	3
Tryptophan	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Phenylalanine	7	7	7	7	6	6	7	6	7	7	7	7	7	6	7	7	7	7	7	7	7	7	7	7	7	7	9	8	8	8	15	8	3
Tyrosine	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	2	0	4	3
	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	153	152	145	146	147	

mutations from the ancestral myoglobin chain, and the tree shrew appeared to fit equally well as either a primate or a non-primate.

Our work on myoglobin sequences has continued, and when additional information has become available from other laboratories it has been incorporated in our phylogenetic consideration. The myoglobin sequences of 29 mammals and 2 birds, listed in table 1, are shown in table 2.† Their respective amino acid compositions are given in table 3.

Not all the 31 sequences shown in table 2 have been established in the strictly chemical sense. There are some alignments by homology, a procedure which does not exclude 'sequence inversions'. Segments based on homology are enclosed in parentheses. However, differences in amino acid sequence have almost invariably been established by accepted chemical procedures such as the comparison of overlapping peptides and sequential peptide chain degradation.

In all three tables we have included for comparison the myoglobin of the mollusc *Aplysia*, the major monomeric component of the haemoglobin of the polychaete worm *Glycera*, and the monomeric haemoglobin V of the lamprey *Petromyzon*.

The information presented in table 2 lends itself to the application of the method of maximum parsimony, whereby the number of mutations required to account for the differences is minimized. In another investigation we have applied this method using a computer program kindly made available by M. Goodman and G. W. Moore. Instead of initiating the procedure with a tree provided by a clustering method we provided the initial phylogenetic pattern and at each subsequent stage of introducing economies we edited the output. If alternative patterns were available then the one preferred on zoological grounds was resubmitted to the computer program. Thus we constrained the parsimony method within the limits of what we regarded as acceptable on the criteria of comparative anatomy and the fossil record, and the outcome of this procedure will be published elsewhere.

In the present work we have constructed, without aid from the computer program, a series of different cladograms, each based on the parsimony criterion and constrained by zoological information other than from the myoglobin sequences. Indeed, many of the starting trees submitted for investigation were based on alternative disputed phylogenies, and it was hoped that the direct and detailed comparison of these might provide support for one hypothesis compared to another. As is well recognized by comparative anatomists, shared ancestral characters are unreliable phylogenetic indicators, and this same problem has been encountered at the molecular level. One of the most striking aspects of molecular evolution revealed by this part of the study is the extent to which parallel mutations contribute to the difficulties of reconstructing phylogeny. Undetectable mutational events further complicate the issue.

We have investigated several approaches currently suggested for adjusting the totals of sequence differences to take account of the number of mutational events which are not directly

† Subsequently to the completion of the greater part of this work several amendments to sequences have become available as follows: Sheep 99 Ile, 101 Val, 145 Gln (Vötsch & Anderer 1975). Dog 122 Asp, 128 Glu (Dumur, Dautrevaux & Han 1976). There has been some contradiction in the literature regarding position 9 in dog, sometimes quoted as Leu and sometimes as Ile; only the former has been confirmed in Cambridge. Badger 74 Gly, 99 Ile, 101 Val (Tetaert *et al.* 1974), also 21 Val, 53 Ala, 113 His, 122 Asp (Dumur *et al.* 1976). Kangaroo 21 Val, 26 Gln, 122 Asp (Fisher & Thompson 1976).

After this paper had been accepted for publication we found that the sequence which we have used for the dolphin (*Delphinus delphis*), copied from table 2 of Kluh & Bakardjieva (1971), is erroneous. According to table 1 in the same paper the necessary amendments are 1 Gly, 4 Asp, 12 Asn, 15 Gly, 26 Glu.

detectable. Two methods are based on a recently developed stochastic model (Holmquist 1972*a;* Holmquist, Cantor & Jukes 1972; Jukes & Holmquist 1972). Using the amino acid difference matrix we have explored a variety of clustering procedures.

By examining all the sequences it has become apparent that during about 300 Ma of evolutionary divergence leading to birds and mammals from their most recent common ancestor one region of the myoglobin molecule appears to have remained invariant and another has sustained only a few changes, and these of a conservative nature.

From the functional point of view we have initially classified the residues of the myoglobin molecule as helical and non-helical, external and internal, the latter including most of the haem contacts (Perutz 1965; Perutz, Kendrew & Watson, 1965; Perutz 1969). We have also taken into account the known bonding structure of the molecule, i.e. the various salt bridges and hydrogen bonds which, together with the van der Waals forces, maintain its tertiary conformation and stability (Watson 1969; Englander & Staley 1969).

There is no doubt about the physiological role of myoglobin both as a short-term oxygen store (Hill 1936; Millikan 1939; Åstrand, Åstrand, Christensen & Hedman 1960*a*, *b*) and as a long-term oxygen store (Theorell 1934; Irving 1939; Scholander, Irving & Grinnell 1942; Scholander 1961; Irving 1964; Scholander 1964; Kooyman 1969). Myoglobin is also involved in facilitating the diffusion of oxygen (Wittenberg 1959; Scholander 1960; Hemmingsen 1963; Hemmingsen 1965; Scholander 1965; Wittenberg 1966; Wyman 1966), and some information is now becoming available about those residues responsible for these functions (T. Takano 1975, personal communication). Compared with the information for haemoglobin our knowledge of the precise participation of particular residues of myoglobin in physiological processes is minimal. There are, for example, only four known variants (Boyer, Fainer & Naughton 1963; Boulton, Huntsman, Lorkin & Lehmann 1969; Boulton *et al.* 1971*a*, *b*, *c*) of human myoglobin compared with over 200 for human haemoglobin (Lehmann & Huntsman 1974). Variants have been of great value in investigating the function of haemoglobin, but sadly this has not been possible for myoglobin.

Much information on the antigenic determinants of sperm whale myoglobin has accumulated in recent years (Atassi 1973). With the sequences now available to us, it has been possible to examine the regions implicated in antigenicity in related species.

We have drawn on the evidence of comparative anatomy and the fossil record to provide possible dates of divergence between the ancestors of the living species in order to investigate the relationship between the number of fixed mutations and time. A number of alternative methods have been applied to throw light on differences in rate within and between lineages.

2. Phylogenetic background

In retrospect we now feel that one particular aspect of the phylogenetic chart presented in our 1974 paper was misleading. At some points of branching we indicated lines leading to particular groups (e.g. Primates and Condylarthra) and we left other groups combined in a single line (e.g. 'other eutherians'). In consequence, the subsequent divergence of the Carnivora from the 'other eutherians' left a line representing 'other eutherians', including the Tupaiidae, leading to the tree shrew (which was convenient as there are no undoubted fossil tupaiids). This method of presentation gives the impression that the Carnivora and tree shrew

shared a common ancestry within the single line labelled 'other eutherians'. This is most unlikely, even though the possibility cannot be ruled out on the present evidence. Hence, the line labelled 'other eutherians' should not be interpreted as a single stock because it is probable that splitting into several stocks had already occurred; it represents a bundle of lines of separate descent artificially bound together by a convention of presentation. As this phylogenetic pattern was used as a basis for reconstructing the pathway of change in myoglobin it becomes questionable whether a mutation shared by the tree shrew and the harbour seal is a product of shared ancestry or of parallel change in separate lines.

In the present study we have chosen to discuss the pattern of phylogenetic relationship separately from the problems of assigning dates to points of divergence.

Figure 1 shows a pattern of phylogenetic kinship between 30 living species and attempts to summarize the consensus of opinion based on the evidence of comparative anatomy of both living and fossil forms. Starting from the living species, we have arranged the pattern of branching geometrically to indicate the probable sequence of divergences along any given line leading back to their common ancestor, but the length of the branches is unrelated to absolute time. Thus pairs of lines linked at the same level of branching may not represent divergences occurring at the same time. A substantial part of this phylogenetic pattern reflects the systematic classification of the species under consideration and it is sometimes salutary to be obliged to re-examine the assumptions on which it is based. If we refer to figure 1, starting at the top, the chimpanzee and gorilla are linked as representatives of the family Pongidae and they are jointly linked with man as the only living representative of the family Hominidae. The family Hominidae originated within the family Pongidae: we do not accept that it is necessary to change rank at every point of dichotomy. We are aware that some might prefer to link one of the pongids with man and that others might prefer to represent the relationship between these three species as an unresolved trichotomy.

The hominid–pongid group is linked with the gibbon as a representative of the family Hylobatidae, all being members of the superfamily Hominoidea. Baboon and macaque are linked as representatives of the family Cercopithecidae, within the superfamily Cercopithecoidea, and they are linked with the Hominoidea, all belonging to the infraorder Catarrhini.

The South American woolly monkey and squirrel monkey are linked as representatives of the family Cebidae, and are jointly linked with the marmoset as a representative of the Callitrichidae, all three species being members of the superfamily Ceboidea within the infraorder Platyrrhini, but we are fully aware that the nature of the relationship between the two families is obscure. The Catarrhini and Platyrrhini have been linked, both belonging to the suborder Anthropoidea.

Among the prosimians the slow loris and potto are linked in the subfamily Lorisinae, and are jointly linked with the galago as a representative of the subfamily Galaginae, all three being members of the family Lorisidae. This group is linked with the sportive lemur as a representative of the family Lemuridae. The Anthropoidea have been linked with the prosimian stem because all are included within the order Primates, but it is a widely held view that the Anthropoidea arose from within the Tarsiiformes and their relationship to the Lorisiformes and Lemuriformes is obscure.

Among the Cetacea, all three representatives are toothed whales (suborder Odontoceti). We have retained the conventional link between the porpoise and dolphin, both being representatives of the superfamily Delphinoidea, and shown them as being jointly linked to the sperm

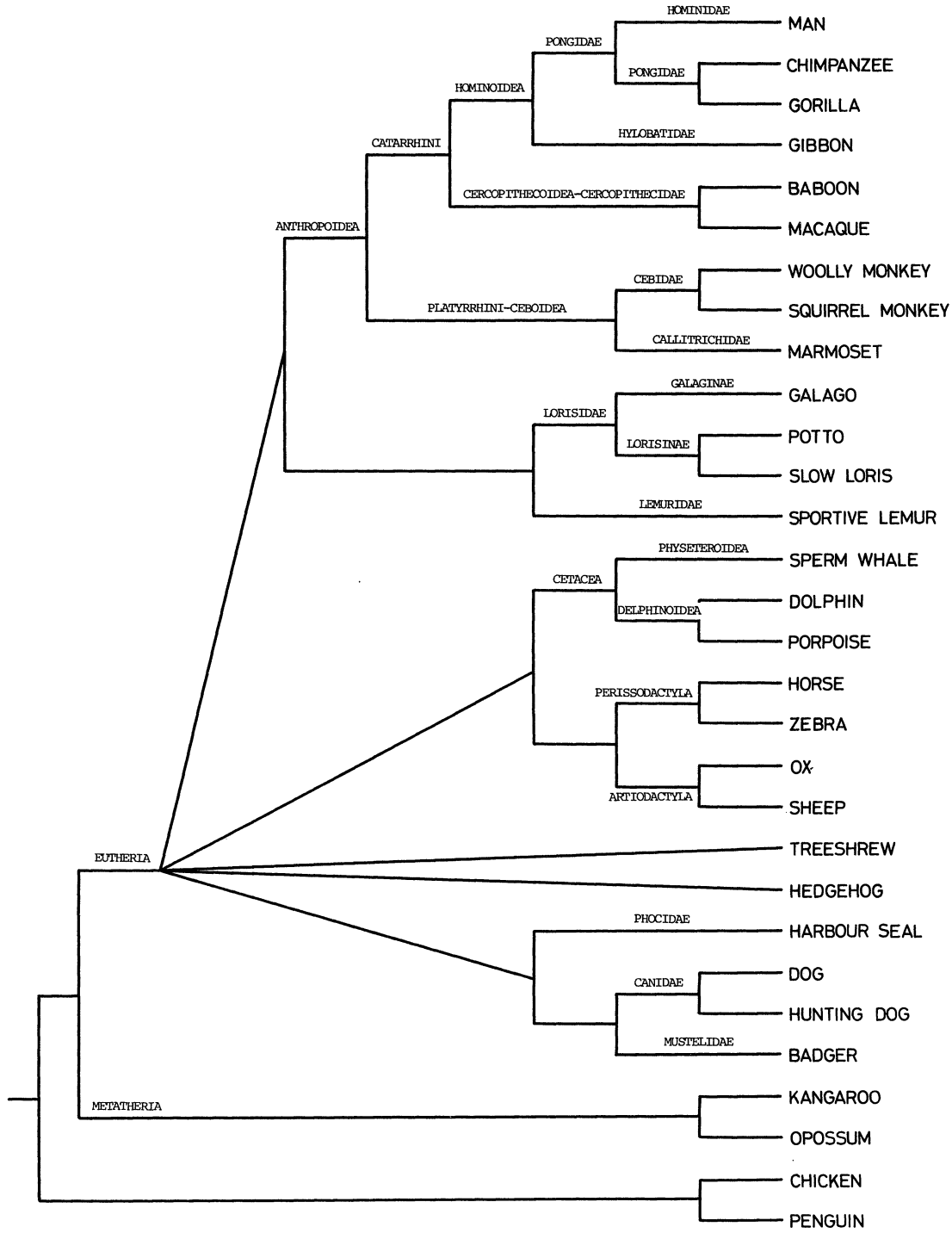


FIGURE 1. Pattern of phylogenetic relationship between 30 living species. The pentachotomy represents the superimposition of four dichotomies, the order of which is not known.

whale, a representative of the superfamily Physeteroidea. The unexpected similarity between the myoglobins of the dolphin and the sperm whale discussed by Romero-Herrera *et al.* (1973) is now known to be a result of the errors reported in the footnote on page 68 of this paper.

Horse and zebra have been linked as representatives of the order Perissodactyla, sheep and ox have been linked as representatives of the order Artiodactyla and these two orders have been linked on the grounds that both probably had their origin within the extinct order Condylarthra. We have tentatively accepted that the order Cetacea arose within the mesonychid condylarthrans (Van Valen 1966; McKenna 1969) and so we have linked them with the ungulates, all being 'condylarthran derivatives'. We are aware that the relationship of these three orders is unclear and that some might favour a link between the Artiodactyla and the Cetacea rather than between the Artiodactyla and the Perissodactyla.

The hedgehog is a representative of the order Insectivora. We have not linked it with the tree shrew because we wish to remain neutral on the contended issue as to whether the tree shrew should be included in the order Insectivora, or the order Primates, or the order Scandentia (Butler 1972).

Among the order Carnivora the two dogs (family Canidae) have been jointly linked with the badger (family Mustelidae), all being Fissipedia. Our placing of the harbour seal reflects the classical division between the suborders Fissipedia and Pinnepedia, although we are aware that some would favour a link between the Phocidae and Mustelidae (Savage 1957; McLaren 1960).

The lines leading to the Primates, condylarthran derivatives, tree shrew, hedgehog and Carnivora are shown as emerging from a pentachotomy. This does not represent our opinion of the phylogenetic pattern but it results from our superimposition of four dichotomies, because there is no consensus of opinion regarding the sequence of branching among early eutherian mammals. Several alternative solutions are discussed later in this section.

Opossum and kangaroo are linked on the assumption that all living marsupials are members of a monophyletic group, and similarly penguin and chicken are linked on the assumption that the class Aves is monophyletic.

The eutherian stem is linked with the marsupial stem on the basis of the evidence that both groups can be derived from a common stem of therian mammals which were still undifferentiated in early Cretaceous times (A. W. Crompton & Z. Kielan-Jaworowska 1975, personal communication). It is generally accepted that mammals originated among synapsid reptiles and that birds arose among diapsid reptiles; the division between these two reptilian groups has been traced back almost to the earliest known reptiles. The link between the mammalian stem and the birds is justified here on the assumption that the amniote egg evolved only once.

Returning now to the pentachotomy, eight different possible solutions are illustrated in figure 2. Possible solution 1 is based on the pattern of kinship which emerged in the course of our earlier work, when fewer myoglobin sequences were available. The Carnivora and the condylarthran derivatives are linked because the harbour seal (representing the Carnivora) was found to share particular residues with some condylarthran derivatives (Romero-Herrera *et al.* 1973). This would be expected if the miacid carnivores were derived from the arctocyonid condylarthrans, and so this pattern deserves further investigation now that additional myoglobin sequences of Carnivora are available. We also found that the tree shrew could be accommodated equally well as a primate or as a non-primate (Lehmann *et al.* 1974). In solution 1 the tree shrew has been placed as a non-primate, linked with the hedgehog, so exploring the

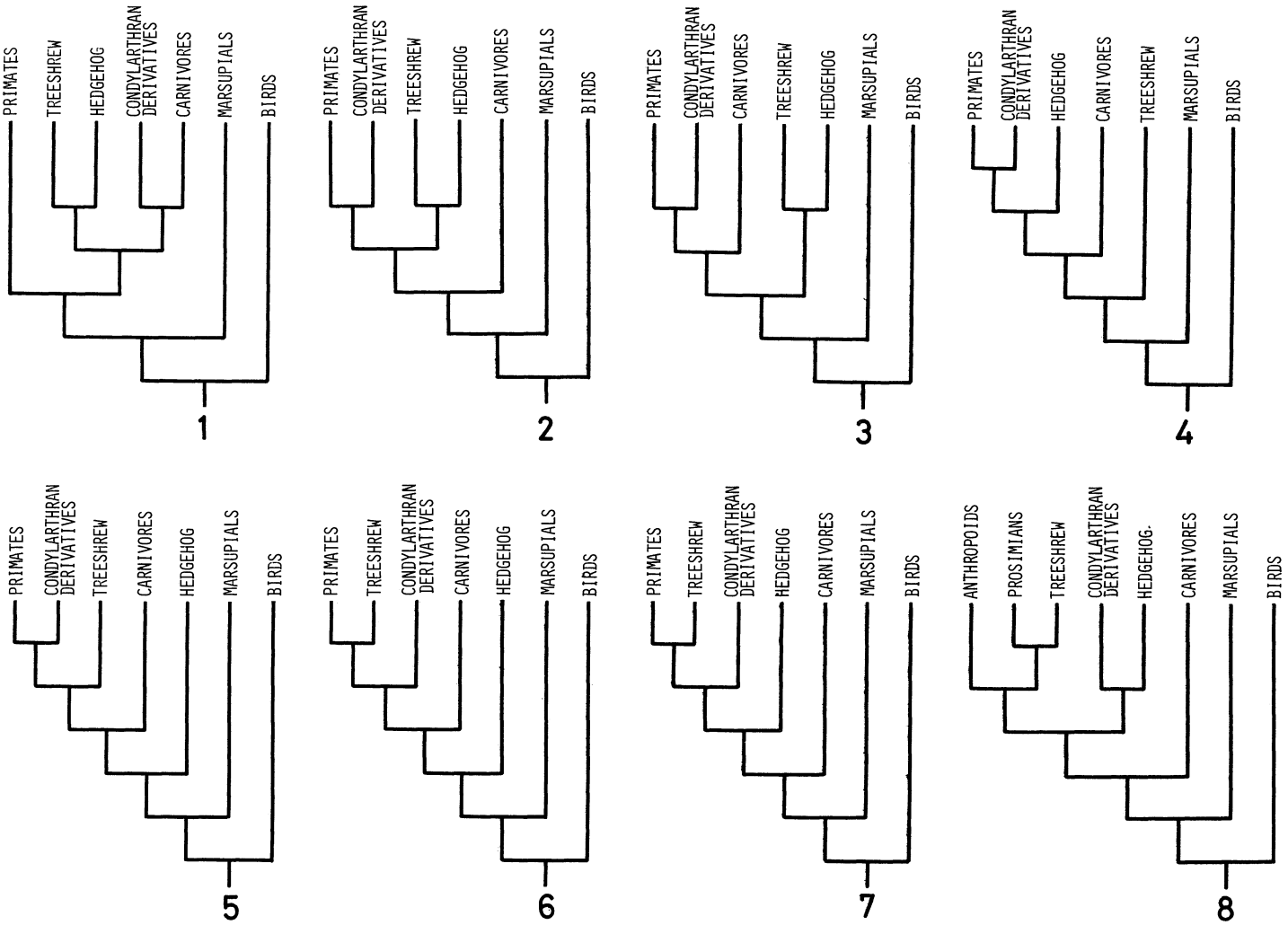


FIGURE 2. Eight alternative possible solutions to the pentachotomy of figure 1. The numbers 1 to 8 are used elsewhere in this paper to refer to cladograms based on each of these patterns.

possibility of placing the tree shrew in the order Insectivora. In summary, solution 1 implies that the Primates branched off a primitive eutherian stock earlier than the Condylarthra, that the Carnivora were derived from the early Condylarthra, and that the primitive stock persisted and gave rise to the modern Insectivora, including the tree shrew.

In all the other possible solutions considered here the condylarthran derivatives are linked more closely to the Primates than to the Carnivora. This reflects the similarity between the teeth of *Protungulatum* and *Purgatorius* and it further implies that the Carnivora were not derived from the arctocyonid condylarthrans. Compared with solution 1, these new relationships can be derived by reversing the Primates and Carnivora, as shown in solution 2, so bringing the condylarthran derivatives closer to the Primates than to the Carnivora, but leaving the tree shrew and hedgehog linked with one another.

In solution 3 the tree shrew and hedgehog are still linked but they are shown as diverging

earlier than the Carnivora, which would result in them being equidistant from Carnivora, condylarthran derivatives and Primates.

For all of the remaining possible solutions the tree shrew and hedgehog lines have been separated. In solution 4 the tree shrew (now as a representative of the order Scandentia) has been left in the same position as in solution 3, implying that this lineage diverged before the differentiation of the ancestors of any of the other eutherian groups represented in this study. The hedgehog has been moved to share a closer relationship with the condylarthran derivatives and the Primates, reflecting the opinion that the *Protungulatum/Purgatorius* molar tooth pattern could have been derived from an erinaceomorph ancestor (W. A. Clemens 1975, personal communication).

By contrast, in solution 5 the hedgehog has been left in the same position as in solution 3, and the tree shrew has been linked to the common ancestor of the condylarthran derivatives and the Primates, which results in the Scandentia being equidistant from the two latter groups.

The three remaining solutions explore the possibility that the tree shrew is more closely related to the Primates than to any other group represented in this study. Solution 6 differs from solution 5 only in this respect, the relative positions of the tree shrew and the condylarthran derivatives having been reversed.

Solution 7 retains the same relations as established in solution 6 for the condylarthran derivatives, the tree shrew and the Primates, but the relative positions of the Carnivora and the hedgehog line have been reversed. This results in a pattern reflecting the possibility that all groups mentioned except the Carnivora might have been derived from an erinaceomorph ancestor, whereas the Carnivora stand apart as the product of an earlier divergence.

The Carnivora retain this position in solution 8, but the hedgehog line is shown as being more closely related to the condylarthran derivatives than to the Primates, so reflecting a possibility at one time suggested by McKenna (1974, personal communication). In addition, the tree shrew has been moved to a position most closely related to the prosimians, thus exploring the possibility of including it within the Primates.

3. Anatomy of alternative cladograms

With the aim of obtaining the most parsimonious acceptable solution of the pentachotomy, eight different cladograms were prepared, in each case minimizing the number of single-point mutations (hits) required to derive the sequences of the living animals.

The sequences were fitted by considering in turn each amino acid site homologous over all species and allocating fixed mutations along the branches of the phylogenetic tree. By using the genetic code and the nucleotide information the number of hits needed to fit the tree was minimized. Implicit in this was the principle that only unavoidable parallel and back mutations were assigned, and single-hit changes were favoured over double hits and these over triple.

In most instances, changes could be allocated with confidence because in most cases the sequences of a number of fairly closely related animals are available, but towards the base of the tree the mutations were more difficult to assign, and considerable uncertainty is attached to any solution of the pentachotomy. A particular difficulty was encountered at the root of the cladogram where the ancestors of birds and mammals diverge, because the assignment of residues to either group has to be arbitrary. One could allocate half of the mutations to each

side, or one could take the extreme case and place all the mutations on one branch; in this context we have adopted the latter course, bearing in mind that any alternative solution would affect the reconstructed ancestral chain in a totally predictable way.

In the case where one must choose between a solution needing a back mutation and, at an identical cost of hits, one requiring two changes in parallel, it has been our policy to construct two complementary cladograms, so that events represented as back mutations in one become parallel changes in the other. In our present state of knowledge we have no reason to prefer one cladogram to the other on any functional grounds.

Cladogram 2 (see figure 3, pullout 1, facing p. 90)

This cladogram is one of the two most parsimonious acceptable solutions and scores a total of 281 hits. There is a net bias towards changes in parallel in preference to back mutations in those cases where such alternatives are available. In this sense, cladogram 5 is complementary to the present cladogram. The various parts of the cladogram are dealt with under taxonomic headings in the detailed description which follows.

Superfamily Hominoidea

Four amino acid residues indicate hominoid common ancestry: 110 Cys, 140 Lys, 144 Ser and 145 Asn. 110 Cys and 145 Asn are peculiar to the hominoids, 144 Ser also appears in the birds, but among mammals it is found only in the Hominoidea. Among the Primates residue 140 Lys occurs only in man, chimpanzee, gorilla and gibbon. Within the Hominoidea, the first dichotomy is between the branch leading to the Hylobatidae (gibbon) which, as far as myoglobin is concerned, does not appear to have introduced any amino acid changes, and the branch leading to the Pongidae, along which residue 23 Gly became fixed. The Pongidae gave rise to the branch leading to Hominidae, and within the Pongidae a dichotomy leads to chimpanzee and gorilla, with the introduction of substitutions 116 His and 22 Ser respectively. The relationship of man, gorilla and chimpanzee is not known; the fossil evidence does not reveal whether the branches leading to chimpanzee and gorilla shared common ancestry after the divergence of the Hominidae or whether one of these species diverged at an earlier time while the other shared a longer period of common ancestry with man. The myoglobins do not resolve this problem because man and chimpanzee share residue 22 Pro, whilst the gorilla has Ser at this position, and man and gorilla share residue 116 Gln, a position occupied by His in the chimpanzee.

The evidence of the known amino acid sequences of other proteins provides no further resolution. The fibrinopeptides A and B of man, chimpanzee and gorilla are the same (Doolittle, Wooding, Lin & Riley 1971). Comparison of the δ-chain of haemoglobin A_2 (Boyer *et al.* 1971) shows that the chimpanzee and the gorilla share residue 126 Val, whereas man has Met at this position. This yields no decisive information because 126 Val is also found in *Hylobates*, *Ateles*, *Saimiri* and *Saguinus* and so it is considered to be the ancestral state for this position.

One might expect that comparison of the adult haemoglobins could help. Man and chimpanzee have identical α- and β-chains (Dayhoff 1972), and differ from the gorilla by one residue in both chains. This could be interpreted as indicating common ancestry for chimpanzee and man after the divergence of the gorilla lineage. However, this situation could also have arisen if a common ancestor of chimpanzee and gorilla diverged from the stem leading to man and the haemoglobin mutations became fixed in the gorilla lineage after this common pongid stock

split further into the branches leading to the living forms. Thus, the haemoglobin evidence does not refute the hypothesis that the phylogenetic distance is the same from man to either of the two pongids. As is well recognized by comparative anatomists, such problems can only be resolved when the ancestral and derived states have been distinguished.

Superfamily Cercopithecoidea

Only two representatives, *Papio anubis* and *Macaca fascicularis*, have been studied, both belonging to the subfamily Cercopithecinae. The common ancestry between baboon and macaque is indicated by the shared residue 106 Leu; elsewhere this residue appears as a parallel mutation only in the squirrel monkey lineage.

The two cercopithecoids differ from each other in only one residue; position 66 is Ala in the baboon and Val in the macaque. In this cladogram 66 Val has been allocated to the macaque branch, because Ala has been chosen for the primate common ancestor at this position (see below). It is of interest that in several of the computed solutions, in which programs designed by Goodman, Moore, Barnabas & Matsuda (1974) were used, these two genera emerged as independent branches from the Catarrhine stem. Exploring such a pattern, with macaque diverging before baboon, an equally parsimonious solution is obtained when 66 Val is considered to be the primate common ancestral residue instead of Ala. This would lead to the assignment of two independent double hits, those of 66 Thr for the sportive lemur and 66 Thr for the squirrel monkey, and one single-point mutation, 66 Ala, in the common ancestor of baboon and the Hominoidea. However, this arrangement results in the change to 106 Leu being considered as three parallel events in the lineages of baboon, macaque and squirrel monkey. In this alternative solution the overall cost within the Primates is 8 mutations, the same cost as that found when Ala is regarded as the primate common ancestral residue for position 66, and 106 Leu is introduced in the cercopithecoid common stem. This latter solution (figure 3) seems to be slightly advantageous because it does not involve a double hit with an unknown intermediate.

Infraorder Catarrhini

Common ancestry between cercopithecoids and hominoids is indicated by the shared residues 86 Ile, 110 Ser and 113 Gln. Within the Primates, residue 86 Ile is also found in the common ancestor of the Lorisidae. Residue 110 is of special interest because in all mammalian myoglobins so far studied, with the exception of those of the hominoids and the cercopithecoids, this position is occupied by Ala. To arrive from alanine to the cysteine of the Hominoidea, more than a single-point mutation in one of the codons for alanine is required. However, GCU or GCC for alanine can be converted by one step into the codon UCU or UCC for serine and a further step can change the serine codon into UGU or UGC for cysteine. Thus the serine in the cercopithecoids provides evidence of the intermediate between alanine in other mammals and the cysteine found in the hominoids. The third residue shared by the Hominoidea and Cercopithecoidea is 113 Gln which, among the Primates, belongs only to these two superfamilies. The sequence of divergence between the Ceboidea, Cercopithecoidea and Hominoidea cannot at present be deduced from the fossil record. The myoglobin information provides its solution to the radiation of the Anthropoidea by forming first a dichotomy between Platyrrhini and Catarrhini and a subsequent divergence of the latter into Cercopithecoidea and Hominoidea in agreement with other molecular and anatomical information.

Infraorder Platyrrhini

Residues 31 Ser, 60 Glu and 117 Lys indicate the shared common ancestry of the Ceboidea because, among the myoglobins investigated, they appeared only in the woolly monkey, squirrel monkey and marmoset. As the myoglobin common ancestor for position 117 is Ser, AGU/C, then the Lys, AAA/G assigned to the ceboid branch is the outcome of two nucleotide substitutions. Three other substitutions also appear to have taken place in the ceboid common stem: 109 Asp, 112 Val and 132 Lys. These three mutations are found within the Primates as parallel changes in the sportive lemur lineage. Two equally parsimonious, alternative phylogenies were found to be possible for the three ceboids, when consideration was given to the residues found at four positions. (In the following discussion residues 114 Ala in the woolly monkey and 106 Leu in the squirrel monkey have been left out because they appear only in these two genera.) The first ceboid phylogeny, as used in figure 3, postulates a common ancestor from which two branches diverge, one leading to the marmoset and the other being the cebid common stem. The latter subsequently forks into the woolly monkey and squirrel monkey lineages. In this interpretation 112 Val is allocated to the ceboid ancestor and this is changed to 112 Ile for the woolly monkey. Two mutations are introduced in the cebid common stem: 13 Ile and 81 Gln. Finally, since Ala is considered to be the primate common ancestral residue for position 66, it becomes necessary to assign in this position Val to the woolly monkey and marmoset, and Thr to the squirrel monkey. This arrangement of residues costs seven single-point mutations.

The alternate possibility is to change the ceboid phylogeny as follows: from the ceboid common ancestor one branch leads to the woolly monkey lineage and another to a common stem for the remaining cebid and the callitrichid. This stem subsequently split into branches leading to the squirrel monkey and marmoset. In such an interpretation 66 Val, instead of Ala, is taken as the primate common ancestral residue; hence, to reach 66 Thr in the squirrel monkey a double hit is required. Residues 13 Ile and 81 Gln are allocated to the ceboid common stem, necessarily leading to the assignment of 13 Val and 81 Gln as back mutations in the marmoset. Lastly, 112 Val is introduced in the common ancestor of squirrel monkey and marmoset. This alternative phylogeny also costs seven hits. Although the first solution is more parsimonious in so far as it does not involve a double hit, the myoglobin evidence does not weigh heavily against the possibility that the callitrichids might have arisen from within the cebids.

Suborder Anthropoidea

Common ancestry for the members of the suborder Anthropoidea is strongly supported by residues 22 Pro, 51 Ser and 142 Met which, among the Primates, are found only in the nine anthropoids so far studied. A fourth residue 23 Ser has been assigned to this common stem because of its presence in the three ceboids, the two cercepithecoids and the gibbon. This Ser was eventually substituted by Gly as a back mutation in the common branch leading to man, chimpanzee and gorilla.

Prosimians

Although it seems likely that the Anthropoidea arose from tarsiiform prosimians, the relationship of tarsiiforms to lemuriform and lorisiform prosimians is obscure. In the present context the term prosimian is used to denote the latter two groups.

After the dichotomy between the anthropoid ancestors and the prosimians the latter fixed residues 34 Thr, 81 Gln and 52 Ala. The first of these is found in all prosimians so far studied, the second in three of them and the third in *Lepilemur* and *Galago*. Residue 52 Ala deserves special comment because the chosen ancestral residue for this position is Glu, whose codons GAA/G can change by a single nucleotide substitution into GCA/G, for Ala, assigned to the prosimian common stem. Another single change in this codon is necessary to reach the Pro, CCA/G found in the Lorisinae.

The common ancestry of the lorisids is indicated by residues 35 Ala and 48 Asn which, among all the myoglobins so far studied, are found only in this family. Residues 21 Leu, 86 Ile and 66 Val are also shared by the three lorisids, the first residue being confined, within the Primates, to this family. Close kinship between potto and slow loris is indicated by the residues 9 Ser and 52 Pro, found so far only in these two lorisines, and by residue 28 Ile also peculiar to them within the order Primates. As can be seen in figure 3, residue 9 Ser, UCA/G, in the lorisine stem is the result of a double hit. This is because the ancestral residue for this position, Leu, has been translated by one of the possible codons CUA/G; such a resolution is unavoidable if parsimony is to be obeyed. The alternative choice of Leu, UUA/G, as the ancestral residue, which can change into Ser, UCA/G, for the potto–loris stem by a single-point mutation, would cost one hit more in the overall codon phylogeny of this position.

Among the myoglobins investigated, three of the eight residues fixed along the *Galago* lineage are found only in this genus: 12 Lys, 125 Thr and 127 Val. Within the primates, positions 27 Asp, 117 Asn and 120 Ser are peculiar to the galago, whereas 13 Ile and 132 Ser are shared as parallel mutations with the two cebids and the slow loris, respectively. In the potto lineage residue 81 His appears as the only back mutation for this position, with the Gln found in the prosimian common ancestor as intermediate. In the present cladogram two mutations were allocated to the slow loris, 109 Gly, unique among all myoglobins so far known, and 132 Ser.

The line leading to the sportive lemur incorporates many substitutions but it does not have any unique residue. All eleven changes are either the result of back mutations or they are found elsewhere as parallel substitutions. Five of these parallel events occur in the lineage leading from the condylarthran stem to horse and zebra, as follows: sportive lemur and the condylarthran common stem share residues 21 Val, 101 Ile, and 116 His; 109 Asp appears also in the perissodactyl-artiodactyl common stem and 66 Thr in the horse–zebra stem. It is for this reason that 'phylogenies' produced by procedures such as the unweighted pair group method and some of the Wagner trees, in which no zoological guidance is given, tend to associate these two groups.

Order Primates

Before discussing the kinship of the order Primates, it is relevant to mention that in several of the results we obtained by the maximum parsimony method of Moore and Goodman, the prosimians were assigned common ancestry with the condylarthran derivatives (cetaceans and ungulates), instead of with the Anthropoidea. A detailed examination of this phenomenon revealed that residues 22 Ala, 23 Gly, 51 Thr and 142 Ile are present in the stems of the condylarthran derivatives and the prosimians, remaining unchanged from the eutherian common ancestral chain, whereas 22 Pro, 23 Ser, 51 Ser and 142 Met have been substituted in the anthropoid stem. In addition, residue 34 Thr is found in the prosimian common ancestor, as

well as in the perissodactyl-artiodactyl common stem. On the other hand, the only substitution present in the common ancestor of the Anthropoidea and the prosimians is residue 66 Ala which, as mentioned above, could alternatively be Val. It is difficult to establish a plausible phylogeny for position 66 because eight different amino acids have been found at this position in the 31 myoglobins so far investigated. At present we are interpreting any apparent diphyletic origin of the Primates as resulting from the parsimony method being influenced by the shared residue 34 Thr, and we have chosen to construct our cladograms on the basis of the phylogeny supported, albeit weakly, by comparative anatomy and the fossil record.

Order Condylarthra

As discussed in §2, the cetaceans, perissodactyls and artiodactyls are regarded as being derivatives of the extinct order Condylarthra. Four amino acid substitutions can be allocated to this common stem. Among all myoglobins so far investigated, 53 Ala is found only in this group. At position 101, Ile is shared by the derivatives of this order, although it appears also in the sportive lemur and the pinnipeds as independent parallel events. Position 116 is occupied by His, which is also present, indicated as parallel changes, in the lineages of the sportive lemur, harbour seal and chimpanzee. The introduction of 21 Val in the common ancestor of cetaceans and ungulates is the most economical choice because it requires only two subsequent substitutions, one in the lineage leading to the porpoise and another in the perissodactyl common stem.

Order Cetacea; suborder Odontoceti: superfamilies Physeteroidea and Delphinoidea

The monophyletic origin of the three species of cetaceans so far studied is supported by the common residues 4 Glu and 118 Arg, which are peculiar to this order. Residues 27 Asp and 140 Lys are almost exclusive to the Cetacea, the exceptions being that the former is found in the kangaroo and galago and the latter constitutes part of the common ancestry of the Hominoidea.

Before our finding errors in the sequence of the dolphin (see footnote on page 68), the Cetacea were used to illustrate the point that alternative solutions, favouring parallel and back mutations respectively, can sometimes be reconstructed without any change in the number of hits involved. This point remains valid despite the errors in the dolphin sequence. In solution A (figure 3), five residues, 1 Val, 12 His, 15 Ala, 28 Ile and 74 Ala, are found as parallel changes in the lineages of the sperm whale and dolphin. Alternatively in solution B (figure 4), residues 1 Val, 12 His, 15 Ala, 28 Ile and 74 Ala have been assigned as part of the cetacean common ancestry, so leading to the introduction of five back mutations in the porpoise lineage. Solution B is interchangeable with solution A in all of the cladograms.

Order Perissodactyla, family Equidae; order Artiodactyla, family Bovidae

As mentioned above, the myoglobin data indicate that the condylarthran stock gave rise to the common stem of the cetaceans and another stem from which both ungulate orders were derived.

Residue 34 Thr is found in the four ungulates so far studied and so it has been assigned as part of their common ancestry. This residue is again found in the prosimian common stem where it is regarded as a parallel substitution. The Asp at position 109 is shared by the four ungulates although it also appears in the cebids, sportive lemur, fissipeds and kangaroo. The

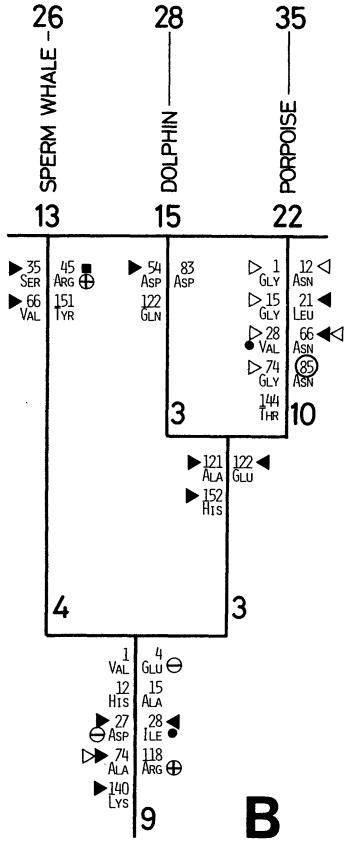


FIGURE 4. An alternative solution to block A of figure 3. Symbols as figure 3. Both local and overall scores (number of hits) are given for each lineage, and these conventions are followed in other figures where appropriate.

third residue allocated to the ungulate common stem is 66 Asn, which is eventually substituted by Thr in the perissodactyl common stem. We have explored the possibility that the Artiodactyla and the Cetacea may be more closely related than the Artiodactyla and the Perissodactyla, and the result is shown in figure 5. In order to arrive at 66 Thr in the perissodactyls two base substitutions are required, but 21 Ile has been eliminated from the same stem. On the other hand the cladistic pattern shown in figure 5 would require the introduction of 34 Thr and 109 Asp as parallel mutations, and so the overall cost of associating the artiodactyls with the cetaceans is two single-point mutations more than that of the classical view.

Ten amino acid substitutions (costing eleven hits) were fixed in the artiodactyl stem before the lines leading to ox and sheep diverged. Four of these, 86 Val, 88 His, 117 Ala and 148 Val are found only in this common stem. Three other residues, 13 Ala, 95 Asn and 121 Ser are each found once elsewhere in the branches leading to opossum, hedgehog and penguin, respectively. Finally, 99 Val, 122 Asn and 132 Ser, although found as parallel changes in other lineages, could be allocated to this ancestral stem with confidence.

The changes at position 117 deserve special comment: when 117 Ser, AGU/C is considered to be the common ancestral residue two single-point mutations are required to obtain the Ala, GCU/C found in the artiodactyl stem. This ancestral codon also necessitates a triple hit in the common stem of the birds, where Gln, CAA/G is found. A further double nucleotide

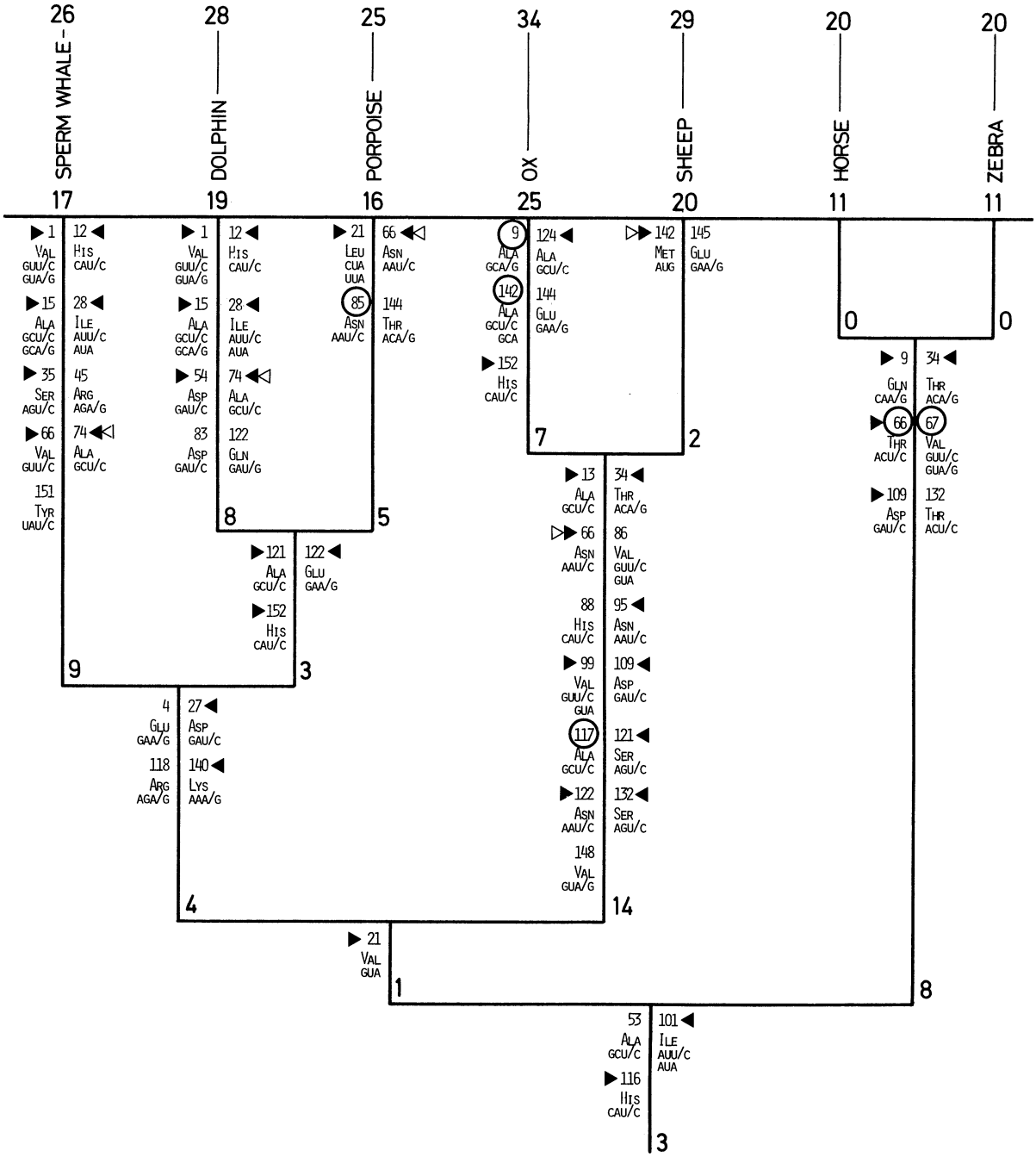


FIGURE 5. A cladogram exploring an alternative pattern of relationships of the Cetacea, Artiodactyla and Perissodactyla.

substitution is required in the ceboid stem to reach Lys, AAA/G. In addition there are single nucleotide substitutions accounting for Asn, AAU/C, found in galago and hunting dog, and Glu, GAA/G, in the chicken. These triple, double and single substitutions make a total of 10 single hits.

Alternatively, if the common ancestral residue is considered to be 117 Ser, UCA/G, only a single hit is necessary to reach the residue 117 Ala, GCA/G, in the artiodactyl stem. This,

together with the assignment of 117 Glu, GAA/G, in the common bird stem, saves two hits when compared with the codon phylogeny above. However, two triple hits are needed to account for the 117 Asn, AAU/C, found independently in the galago and hunting dog lineages, which together with the double substitution Ser, UCA/G, to Lys, AAA/G, found in the ceboids, yields a total of 12 single events.

Residues 9 Ala, 142 Ala and 144 Glu are, among all the myoglobins studied, found only in the lineage of the ox, whereas the other two substitutions occurring in this species, 125 Ala and 152 His, are found elsewhere as parallel mutations, the former in the badger and the latter in the Delphinoidea and harbour seal.

When residue Leu, CUA/G, is regarded as the common ancestral residue at position 9, a double nucleotide substitution is needed to obtain the Ala, GCA/G, found in the ox. The most likely postulated intermediate is Val, GUA/G, and not Pro, CCA/G, because this position occurs in the middle of the A helix (residue 9 corresponds to the helical A7) and Pro would probably provide a less stable configuration of the α helix.

The substitution 142 Ala in the ox has taken place in an internal, haem contact position. The residue at position 142 is either Ile or Met in the rest of the myoglobins, and so a double nucleotide substitution is required to reach Ala. It can be assumed that the intermediate amino acid was Val, rather than Thr, because the former residue has a non-polar side chain. Considering the discrepancy in side chain size between Ile (or Met) and Ala, it is likely that the haem contact function of residue 142 has been lost for the latter. However, this function is probably taken over by the neighbouring Phe 138 and Leu 104, both invariant sites in the known myoglobins.

The lineage leading to the ox scores the highest number of hits since its separation from the metatherian-eutherian common stock, 34 single events, of which seven were fixed along the ox branch since the divergence of Bovinae and Caprinae. In contrast the latter branch accepted two changes: 142 Met and 145 Glu. The first is found only in the sheep amongst the seven condylarthran derivatives so far studied, and is considered a back mutation, whilst the second is peculiar to this species among all myoglobins so far investigated. Horse and zebra have identical myoglobin chains, thus it is virtually certain that all the changes found in these two species were fixed during their common ancestry. Six base substitutions accounting for five detectable amino acid changes were assigned to this stem: 9 Gln, also found in parallel in the common ancestor of the birds; 21 Ile, a back mutation whose intermediate is the Val allocated to the common condylarthran branch; 66 Thr, 67 Val and 132 Thr. The latter two residues are important because, when all myoglobins are taken into account, they are found only in the two representatives of the family Equidae. To get 67 Val from Thr, two nucleotide substitutions are needed, with an intermediate Ala, Met or Ile.

The ancestry of the orders Condylarthra and Primates

This cladogram attributes residues 66 Asp, 113 His and 129 Gly to the common ancestry of condylarthrans and primates. The assignment of 66 Asp to this stem is based on very weak evidence, because it is found only in the dolphin, when the myoglobins of primates (thirteen) and condylarthran derivatives (seven) are considered. The 66 Asp was chosen on the grounds that it is the only amino acid which avoids the generation of double hits in the subsequent radiation, as follows: from Asp, GAU/C, it requires only one substitution to reach Ala, GCU/C, regarded as the primate common ancestral residue at this position; Asp changes to Asn, AAU/C, in the

ungulate common ancestor, and this into Thr, ACU/C, found in the perissodactyl stem. Finally the Val, GUU/C, and Asn, AAU/C, found in the branches of sperm whale and porpoise respectively, demand also one single step each from Asp. The extreme variability at position 66 raises an important issue concerning the reconstruction of past events. No strategy is available which would allow confident qualitative interpolation of further mutations; nevertheless the variability observed makes it unlikely that the parsimony approach results in a good approximation to reality in this case. It is hoped that when more myoglobin sequences are known this uncertainty will be resolved.

Residues 113 His and 129 Gly can be allocated with certainty to the common ancestor of condylarthrans and primates because, with the exception of 113 Gln, a back mutation assigned to the catarrhines, and 129 Ala allocated to the sportive lemur, they are shared by the rest of the descendant genera so far studied.

Order Scandentia and order Insectivora

The phylogenetic positions of the genera *Tupaia* and *Erinaceus* have been difficult to assess because of the imperfection of the fossil record together with the retention of many characters primitive for placental mammals, resulting in a lack of derived character states shared with other groups. It is rather unfortunate that the myoglobins did not help to resolve these problems as one might have hoped. In fact, the reconstructed ancestral myoglobin chain is very similar to that of the tree shrew. In this cladogram (which is one of the two most economical) the common ancestry of the tree shrew and hedgehog is supported by the shared residues 81 Gln and 132 Ser. It is worth noting, at this stage, that the tree shrew has in common with the galago and loris residue 132 Ser, and with these two species plus the sportive lemur residue 81 Gln. Of the three residues attributed to the tree shrew lineage only one, 70 Ser, is unique to this genus among all myoglobins so far studied. The other two changes are at positions 21 Val and 86 Ile. The first is found also in both artiodactyls, two of the cetaceans and the sportive lemur, whereas the second appears in the catarrhines and the lorisids. Thus, at four of the five positions just mentioned the same residues are present in prosimians. On this basis the allocation of tree shrew to the prosimians appears to receive some support, but the weight of this evidence must be evaluated in the light of cladogram 8, discussed below.

After separation from the common stem, the lineage of the hedgehog fixed twelve nucleotide changes which account for eleven detectable amino acid substitutions. (This number contrasts with three nucleotide changes fixed in the tree shrew lineage during the same time.) Four of these amino acid substitutions, 57 Ser, 85 Gln, 87 Ala and 116 Lys are peculiar to the hedgehog among the animals so far studied. The substitution 87 Ala was the result of a double hit on the ancestral myoglobin codon for 87 Lys, with either Thr or Glu as possible intermediates. A difficulty in finding a convincing phylogenetic position for the hedgehog lies in the fact that it shares its other seven substitutions with several species, as can be seen in the following list: 22 Pro is found also in the anthropoids, with the exception of the gorilla; 35 Asp is present in the chicken; 51 Ser is shared by the anthropoids, the carnivores and the marsupials; 66 Thr is found in squirrel monkey, lemur and the perissodactyls; 95 Asn is present in the artiodactyls; 120 Ala is shared by kangaroo and the birds. Lastly, 129 Gly appears in 12 out of 13 primates, all the condylarthran derivatives and the badger.

The ancestry of the eutherians after the divergence of the Carnivora

After the separation of the Carnivora, this cladogram assumes a common stem for the rest of the eutherians whose myoglobins have been studied. The only substitution attributed to this common ancestor was 132 Asn found in the catarrhines, potto and cetaceans. The choice of this amino acid was based on the fact that it serves as intermediate between the Lys of birds, kangaroo and carnivores, and the Ser of hedgehog, tree shrew, artiodactyls, loris and galago, thence avoiding the introduction of double hits. 132 Asn also changes to 132 Thr of the perissodactyls and 132 Lys of the ceboids by a single-point substitution, the latter being a back mutation.

Order Carnivora

The myoglobins of five representatives of this order have been studied and four of them were incorporated when this cladogram was prepared; the fifth, that of the Californian sea lion, will be brought into the discussion later.

The common ancestry of this order is indicated by four residues. Residue 21 Leu is found in all carnivores. Giving priority to single nucleotide substitutions, the presence of 35 Asn, AAU/C in both canids (the domestic dog and the Cape hunting dog) forced us to allocate the 35 Ser, AGU/C found in the harbour seal to the carnivore stem because Gly, GCU/C has been considered to be the common ancestral residue for this position.

For the same reason to reach one of the six codons for 57 Arg (AGA/G, CGU/C, CGA/G) of the harbour seal from one of the four for 57 Ala (GCU/C, GCA/G) of the ancestral myoglobin chain, at least two single steps are required; the logical solution is to use one of the four. Gly codons (GCU/C, GGA/G) found in the three fissipeds as an intermediate. Finally, residue 51 Ser has been assigned to the carnivore stem, although this change can be eliminated if the common ancestral residue for this position is Ser instead of Thr, as shown in some of the subsequent cladograms.

Common ancestry of the badger and the two canids is indicated by residue 109 Asp. This residue occurs in some other myoglobins, but within the carnivores it is found only in the fissipeds. An additional residue which might strengthen this common ancestry is a matter of some uncertainty; the mutation of 122 Asp to Asn could have happened in this stem instead of being two parallel mutations in the lineages of the domestic dog and the badger, but as a consequence of such an arrangement it would be necessary to postulate a back mutation in the branch leading to the hunting dog (see cladogram 5). Eight residues were attributed to the canid common stem; two of them, 35 Asn and 124 His (the latter resulting from a double hit with Arg or Asp as possible intermediates) are peculiar to this family. Residues 41 Asp and 120 Ser are found only in this lineage with the exception of the birds and galago, respectively. The remaining four, 13 Ile, 19 Thr, 51 Thr and 127 Thr, though present elsewhere, also support the canid common ancestry. Three amino acid differences were found between the two dogs. In this cladogram 122 Asn has been attributed to the domestic dog whilst 117 Asn and 128 Glu have been assigned to the hunting dog. Since the separation of the badger lineage from the common ancestor of the Canidae, the mustelid branch has accumulated 10 amino acid substitutions requiring a minimum of 11 single hits. When all known myoglobins are considered residues 82 Gln, 112 Ala and 126 Glu are found only in the badger. The 112 Ala, GCU/C, GCA, is the outcome of at least two single nucleotide substitutions in one of the three possible codons, AUU/C, AUA, for 112 Ile in the reconstructed ancestral myoglobin chain. The

intermediate residue could be either Val or Thr. The change to 35 Gly in the mustelid lineage is somewhat uncertain because within the carnivores there are two possible solutions for the phylogeny of this position. One of them has already been discussed in the context of the common stem of the order Carnivora. The alternative proposes that the codon for 35 Gly of the common ancestral myoglobin changed by a single hit into the codon for 35 Ser in the harbour seal stem, but this would demand a double step to obtain the 35 Asn fixed in the canid lineage. This scheme costs three hits, the same as the one discussed before, but it has the disadvantage that it introduces two nucleotide substitutions, with an unknown intermediate, in one of the codons, and for this reason the other possible solution has been accepted.

When all mammalian myoglobins are taken into account residue 99 Val is found only in the badger and both artiodactyls, and residue 124 Ala is found only in the badger and ox. Within the order Carnivora residues 74 Ala and 129 Gly are peculiar to the badger lineage. On the other hand 101 Ile is also found in the two pinnipeds and 122 Asn is, in addition to the badger, present in the domestic dog.

In the present cladogram the lineage of the harbour seal incorporated 12 amino acid substitutions after this lineage diverged from the carnivore stem. However, because of the introduction of the sea lion, three of these residues have been allocated to the pinniped common ancestor, as can be seen in figure 6. When all myoglobins are considered residue 57 Arg is found only in this suborder, and among the five carnivores residue 113 His is also assigned with confidence to the pinniped common stock. The attribution of 101 Ile to this stem leads to the inevitable incorporation of the same residue in the badger lineage as a parallel event. Among the 29 mammalian genera so far studied, residues 8 His, 56 Arg and 62 Arg are unique to the harbour seal; among the Carnivora residues 54 Asp, 116 His, 121 Ala, 122 Glu and 152 His are unique to this species, whereas residue 19 Thr is also present in the canids.

In our Wagner tree (§4) the harbour seal clustered with the cetaceans rather than with the other carnivores. It may be seen that the harbour seal shares residues 21 Leu with the porpoise; 35 Ser with the sperm whale; 54 Asp with the dolphin; 121 Ala with the dolphin and porpoise; 122 Glu with the porpoise; 152 His with the porpoise and dolphin; 101 Ile and 116 His with the condylarthrans. These similarities could be fortuitous or could perhaps result from convergent evolution due to similarity in mode of life.

As shown in figure 6A, the sea lion branch fixed eight amino acid differences after its separation from the pinniped common stem. Four of them, 22 Val, 66 Lys, 128 His and 147 Arg are peculiar to this species when all known myoglobins are considered. Residues 13 Ile and 35 Gly are found in numerous other species, but they are also present in some of the cetaceans. The first is present in the humpback whale for which only the N-terminal 60 residues have been established (Edman & Begg 1967), and the second in the porpoise and dolphin. Residue 83 Asp is found only in two marine animals: the sea lion and the dolphin. Finally, residue 127 Thr is shared by the sea lion and the canid common ancestor.

During recent years the classical phylogeny of the Carnivora, in which the sea lion and harbour seal are regarded as derivatives of a pinniped common stem, has been challenged by the view that the families Mustelidae and Phocidae on the one hand are closely related (Savage 1957), and on the other the families Canidae and Otariidae (McLaren 1960). Following this latter pattern we have devised a new cladogram (figure 6B) and explored its implications. It was found that the sea lion and the two dogs shared residues 13 Ile and 127 Thr, whereas the badger and the harbour seal had in common residues 13 Val and 127 Ala. The alternative

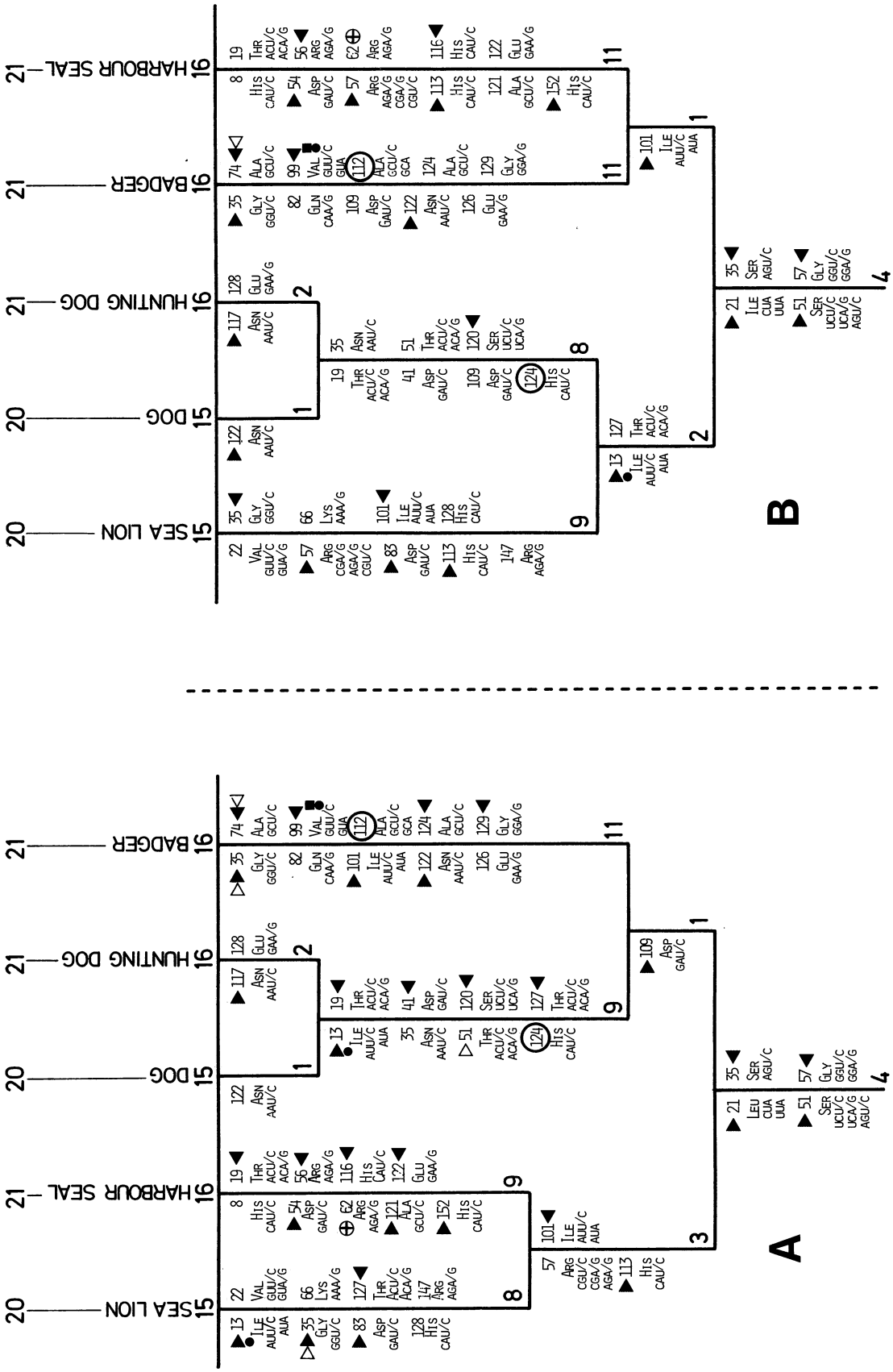


FIGURE 6. Cladograms exploring alternative patterns of relationship within the Carnivora.

set of comparisons (figure 6A) yielded four residues shared by the sea lion and the habour seal, namely 57 Arg, 109 Glu, 113 His and 124 Gly, while the badger and dogs shared residues 109 Asp, 113 Gln and 57 Gly. Thus the comparison of amino acid alignments favoured the classical phylogeny, but such an approach can sometimes be misleading and so it is relevant to examine the possible history of events along particular lines of descent. In the alternative cladogram (figure 6B), the common ancestry of the carnivore stem is supported by the same four residues discussed with reference to the classical scheme. This stem then branches into two lineages, one leading to the mustelids and phocids and the other to the canids and otariids. The former could have fixed residue 101 Ile before diverging into the lineages of the badger and harbour seal, while the latter could have incorporated residues 13 Ile and 127 Thr before dividing into the lineages leading to the dogs and the sea lion. This phylogeny is more economical by one hit because the change to residue 13 Ile occurs only once in the carnivores (Otariidae–Canidae ancestor), whereas the classical interpretation requires the introduction of two parallel events, 13 Ile for the sea lion and 13 Ile for the canid stock. However, when other positions are taken into account the final result is a score of 49 single hits in the order Carnivora for the alternative cladogram (figure 6B) instead of the score of 48 hits for the classical view (figure 6A) because, if the sea lion and the harbour seal did not share an immediate common ancestor, then it is necessary to assign parallel mutations for 57 Arg and 113 His to each of the marine carnivores.

On the principle of parsimony it is evident that the classical division of the Carnivora into the suborders Pinnipedia and Fissipedia is, at present, marginally supported by the myoglobin information.

Subclass Eutheria

Five substitutions have been assigned to the eutherian common stem. Position 115 Leu is shared by all the eutherians so far studied, this residue being Ile in the metatherians and birds. Residue 149 Phe is present only in birds and the kangaroo, whilst in all eutherians the residue at this position is Leu. The opossum also has Leu at this position and has, therefore, been considered to possess this residue as a result of a parallel mutation.

In all eutherians so far studied, with the exception of two cetaceans and the badger, residue 74 is Gly. The position is occupied by Asn in the marsupials, Ala in the chicken and Glu in the penguin. Residue 120 Pro supports the eutherian common ancestry because, discounting the Ser of the galago and the two dogs and the Ala found in the hedgehog (as well as in the birds and kangaroo), it is shared by the rest of the eutherians so far studied. This suggests a second parallel mutation in the eutherian stem and the lineage leading to the opossum, since residue 120 is also Pro in the latter.

Lastly, methionine has been chosen as the myoglobin common ancestral residue at position 142 because it is found in the birds and marsupials, but it changes to Ile in the common stem of the eutherians. During the course of time Met is reintroduced independently in the lineages of the anthropoids and sheep as back mutations, and after a double nucleotide substitution Ala appears at this position in the ox lineage.

Order Marsupialia

The family Macropodidae (Australian) and the family Didelphidae (American) are represented by the red kangaroo and Virginian opossum respectively.

At least three amino acid residues, 74 Asn, 102 Gln and 103 Phe, characterize the common

ancestry of the two metatherians and differentiate them from the other mammals included in this study. The phylogeny of residue 74 Ala can be reconstructed in two different ways depending on the chosen codon for this position in the common ancestral myoglobin. Considering first Ala as being coded by either GCU or GCC, then to reach the Gln in penguin and the Asn in the common marsupial stem will require three and two single steps respectively; alternatively, if the ancestral Ala is coded by either GCA or GCG then the Gln of the penguin and the Asn of the marsupials will demand two and three nucleotide replacements respectively.

The fourth residue allocated to the marsupial common ancestor is 51 Ser. This residue can be eliminated easily without an increase in the total number of hits, if the common ancestral myoglobin residue at this position was Ser instead of Thr, as adopted in cladogram 5.

After the separation of the lineages leading to kangaroo and opossum, the kangaroo line fixed ten amino acid substitutions which are accounted for by 11 detectable nucleotide changes. If the common ancestral residue for position 21 was Ile, then two single changes would have been required to reach 21 Glu, unique to the kangaroo, with either Lys or Val as a potential intermediate. Because of the character of the genetic code the presence of 21 Glu in the kangaroo permits the accurate establishment of the codon for 21 Ile as AUA, for 21 Val as GUA and for 21 Glu as GAA. The first two residues are found in 19 out of the 31 myoglobins studied. Three other residues which are also unique to the kangaroo are 26 Lys, 66 Ile and 140 His. The remaining six residues are 19 Thr in parallel with the harbour seal and the canids; 22 Gly also present in the sportive lemur lineage; 27 Asp which appears also in galago and the cetacean common ancestor; 13 Ile, 109 Asp and 122 Asn which occur in several other lineages as parallel mutations.

During the same period of time the lineage of the opossum incorporated nine amino acid substitutions which require at least eleven nucleotide changes. Considering all known myoglobins, residues 81 Asn, 100 Ser, 125 Gly and 132 Gly are restricted to the opossum. 132 Gly is the result of two nucleotide changes in one of the ancestral codons, AAA or AAG, for lysine. Residue 13 Ala appears independently in the branch of the artiodactyls. The other four residues fixed along the opossum lineage at positions 22, 66, 120 and 149 are discussed below in the context of parallelism.

From the point of view of its myoglobin, the opossum presents a striking example of parallel and convergent evolution. When the amino acid differences between 28 mammalian species were examined, it was noticed that the average value between the opossum and the anthropoids was about 16, whereas a comparison of the opossum with the rest of the mammals gave about 26. To a lesser extent the kangaroo followed the same pattern because the average difference between this marsupial and the anthropoids was about 22, while the average difference between kangaroo and non-anthropoid mammals was about 27. Another indication of parallelism was obtained from the results of tree-forming procedures (discussed in §4), in which the opossum frequently emerged as part of the Ceboidea group. This placement of the opossum sometimes drew the kangaroo into a group with the anthropoids. A detailed investigation in terms of the cladogram showed that the removal of the opossum from its expected zoological position began by its having residues 120 Pro and 149 Leu in parallel with the eutherians. This was compounded by residue 66 Ala shared in parallel with the primate common stem, and by residue 22 Pro shared in parallel with the Anthropoidea. Additional grounds for the translocation of the kangaroo arose from its possession of residues 109 Asp and 13 Ile in parallel with the Ceboidea for the former, and the Cebidae for the latter. These similarities are reinforced by

some residues which are not shown in the lineages of the relevant species on the cladogram because they form part of the ancestral myoglobin chain (i.e. they are shared primitive characters); for example, 21 Ile (anthropoids–opossum); 34 Lys (anthropoids–marsupials); 52 Glu (anthropoids–marsupials); 113 Gln (Catarrhini–marsupials) and 132 Lys (Platyrrhini–kangaroo). We emphasize that although these residues are shared by some other stems scattered over the cladogram they have all been retained as primitive characters in the marsupials and anthropoids.

Class Mammalia; class Aves; orders Sphenisciformes and Galliformes

Before discussing the amino acids which were possibly fixed along the stems of the common mammalian ancestor and the common bird ancestor, we need to mention the weaknesses present at the root of our cladogram. These are produced by the following circumstances. First, only two birds are represented. Secondly, the time of divergence between the branches which eventually gave rise to the birds and mammals has been estimated at 293 Ma, whereas it is at least 52 Ma since the ancestors of the Galliformes and the Sphenisciformes diverged, and a minimum of 79 Ma has elapsed since the divergence of the Metatheria–Eutheria. Thus, for the species involved in this study there has been an estimated time of 241 Ma of avian common ancestry and 214 Ma of mammalian common ancestry. During that time, in this cladogram, we have been able to detect only 15 amino acid differences, which are accounted for by a minimum of 20 nucleotide substitutions. It is reasonable to suppose that some of these amino acids belong to the avian stem whereas others form part of the mammalian common stem, but which belong to each side is not known, and this has important repercussions on the generation of the hypothetical ancestral myoglobin chain. It would be possible, of course, to apportion the number of mutations, allocating seven for the mammals (214 Ma) and eight for the birds (241 Ma), but this provides no solution as to which mutations should be allocated to each side. We have, therefore, adopted the arbitrary method of assigning all to the avian lineage in this cladogram and, as an alternative, assigning all to the mammalian common ancestral stem in cladogram 5. We hope that this problem will be elucidated in the near future by the study of the myoglobin of more birds, and of the prototherians, platypus and echidna.

Nine of the fifteen residues allocated to the avian common ancestor have been found only in the two birds when all 31 known myoglobins are considered: they are 12 Thr, 26 His, 30 Met, 42 Arg and 113 Lys, which could be the outcome of single point substitutions; and 5 Gln, 48 Gly and 116 Ala, which require as yet unknown intermediates between them and their respective ancestral residues. In this cladogram Gln has been assigned as the avian common ancestral residue for position 117; if one considers that Ser, AGU or AGC, is the codon in the ancestral myoglobin chain, then to arrive at Gln, CAA or CAG, at least three single nucleotide changes are necessary. It is obvious that the use of one of the Ser codons, UCA or UCG, would make it possible to reach Gln by a double substitution, but this choice would involve twelve rather than ten hits in the overall phylogeny of position 117. An alternative possible solution is explored in cladogram 5.

The remaining six residues allocated to the avian stem also occur in other lineages: 9 Gln (perissodactyls); 13 Ile (cebids, galago, sea lion, canids and kangaroo); 41 Asp (canids); 54 Asp (dolphin and harbour seal); 57 Gly (fissipeds) and 144 Ser (hominoids). After the divergence of the branches leading to the chicken and the penguin, the former introduced eleven detectable amino acid changes which are accounted for by fifteen single-point mutations. Nine of these

residues are unique to the chicken and can be divided into those which could result from one nucleotide substitution, 85 Asp, 110 Val, 117 Glu, 127 Ser and 140 Asp, and those resulting from at least two nucleotide substitutions: 34 His, 53 Pro, 66 Gln and 75 Gln. The latter one is important because it is Ile in the rest of the myoglobins, and the presence of Gln in this position might be expected to introduce water into the haem pocket. Although the residue at position 75 is regarded as internal it is very close to the surface in the vicinity of the EF corner and it seems most likely either that the Gln side chain is hydrogen bonded (as in the case of the internal and invariant 39 Thr) or that the polar region protrudes towards the surface. Another interesting observation concerning position 75 is that the residue at this position is Ile in the common ancestral myoglobin and, because of the presence of Gln in the chicken, the codon AUA can be established for the former, and the codon CAA for the latter. Finally, residues 35 Asp and 121 Ala occur elsewhere in parallel, the first in the hedgehog and the second in the sportive lemur, the harbour seal and the delphinoid common ancestor.

In this cladogram 21 recognizable amino acid substitutions were assigned to the penguin branch after its separation from the avian common stem, requiring a minimum of 25 nucleotide substitutions. Thirteen of the 21 residues allocated to the penguin branch are found only in this animal: 3 Asn, 13 Met, 19 Ser, 27 Ala, 40 Met, 47 Arg, 53 Glu, 61 Met, 74 Gln, 90 Ser, 92 Thr, 112 Met and 129 Glu. If 74 Ala is coded by either GCU or GCC in the reconstructed ancestral chain, then a triple hit will be required to reach Gln in the penguin and a double hit will be needed to reach Asn in the marsupial stem. On the other hand, if 74 Ala is coded by either GCA or GCG the situation is reversed because only a double hit is required to reach Gln in the penguin, but a triple hit is needed to reach Asn in the marsupial stem. Hence the alternative solutions are equally parsimonious.

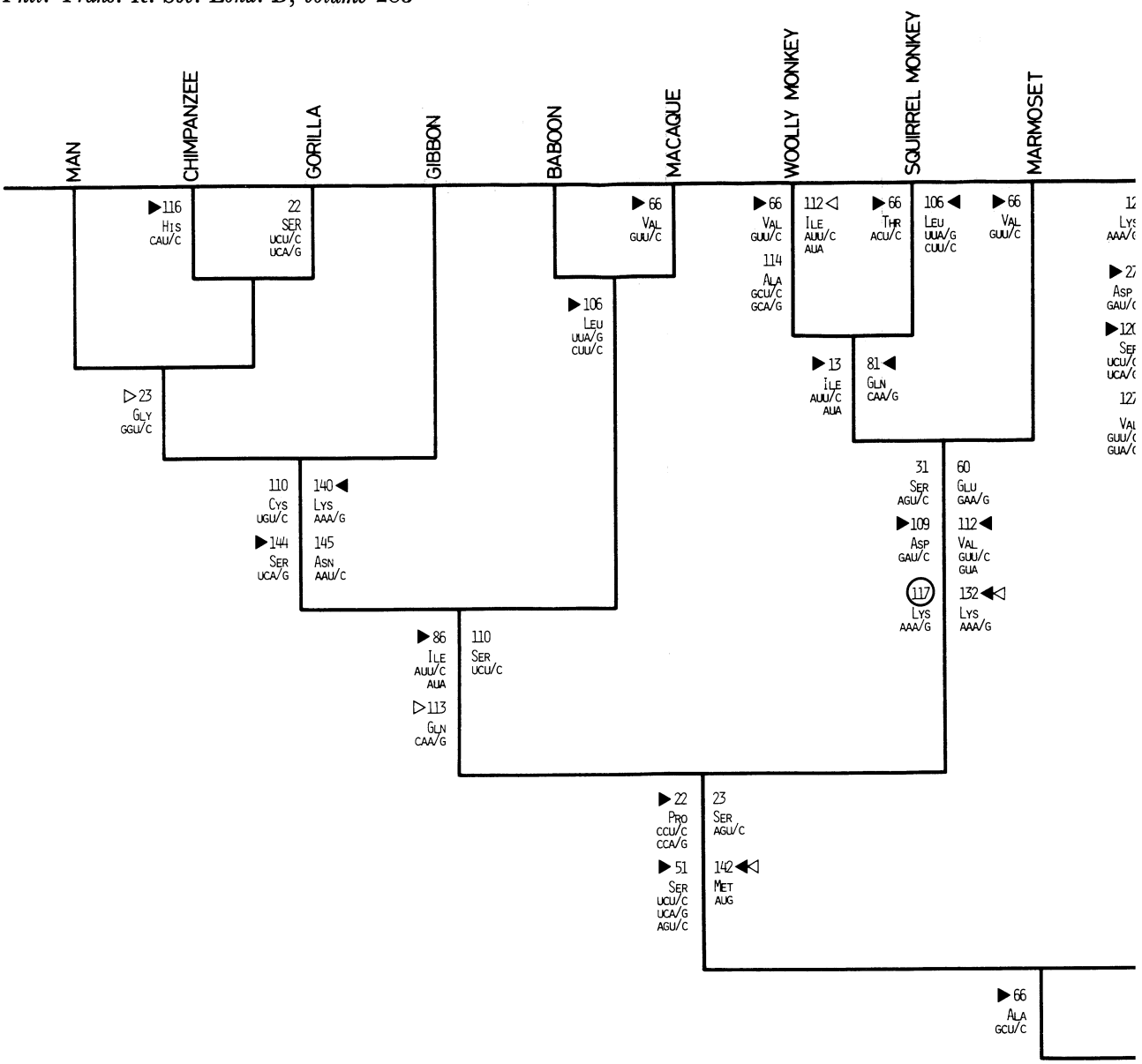
The remaining eight changes in the penguin lineage are found in parallel in various mammals: 21 Leu (lorisids, carnivores and porpoise); 35 Ser (sperm whale and harbour seal); 56 Arg (harbour seal); 99 Val (artiodactyls and badger); 121 Ser (artiodactyls); 122 Asn (artiodactyls, dog, badger and kangaroo); 52 Pro (lorisines) and 66 Val (macaque, woolly monkey, marmoset, lorisids and sperm whale). Double nucleotide changes were necessary to arrive at the last two from the preferred 52 Glu and 66 Asn of the ancestral chain.

Cladogram 5 (see figure 7, pullout 2)

Cladogram 5 and cladogram 2 are equally parsimonious; both require 281 hits. Cladogram 5 differs from cladogram 2 in the following three points:

1. The phylogenetic position of the hedgehog is different. It has been placed as the earliest branch among the eutherian mammals, quite separate from the tree shrew, which remains in the same position as in cladogram 2. This new placing of the hedgehog results in a change in the number of hits incorporated along the tree shrew and hedgehog lineages after the separation between the eutherian and metatherian stems; in cladogram 2 they accumulated 11 and 20 hits respectively, whereas in cladogram 5 the equivalent scores are 14 and 17.

2. It was realized during the construction of cladograms 2 and 5 that a number of mutational events were definitely the result of back mutations, while others were undoubtedly the product of parallel change. However, there were a small number of positions (14) which could be assigned either as parallel mutations or alternatively as back mutations, and so those changes which had been chosen fortuitously as parallel in cladogram 2 were reinterpreted as back mutations in cladogram 5, and vice versa. From the point of view of overall parsimony the two



▶◀ PARALLEL MUTATIONS

▷◁ BACK MUTATIONS

◯ DOUBLE HITS

□ TRIPLE HITS

■ HAEM CONTACTS

● INTERNAL RESIDUES

⊕⊖ CHARGED RESIDUES FORMING SALT BRIDGES

◉ RESIDUES INVOLVED IN HYDROGEN BONDING

1	3	4	5	8	9	12	13	15	19	21	22	23	26	27	28	30	31	34	35	40	41	42	45	4
Gly GGU/C GGA/G	Ser AGU/C	Asp GAU/C	Gly GGA/G	Gln CAA/G	Leu CUA/G	Asn AAU/C	Val GUU/C GUA	Gly GGU/C GGA/G	Ala GCU/C GCA/G	Ile AUA	Ala GCU/C GCA/G	Gly GGU/C	Gln CAA/G	Glu GAA/G	Val GUU/C GUA	Ile AUU/C AUA	Arg AGA/G CGU/C	Lys AAA/G	Gly GGU/C	Leu CUG UUG	Glu GAA/G	Lys AAA/G	Lys AAA/G	Ly AA

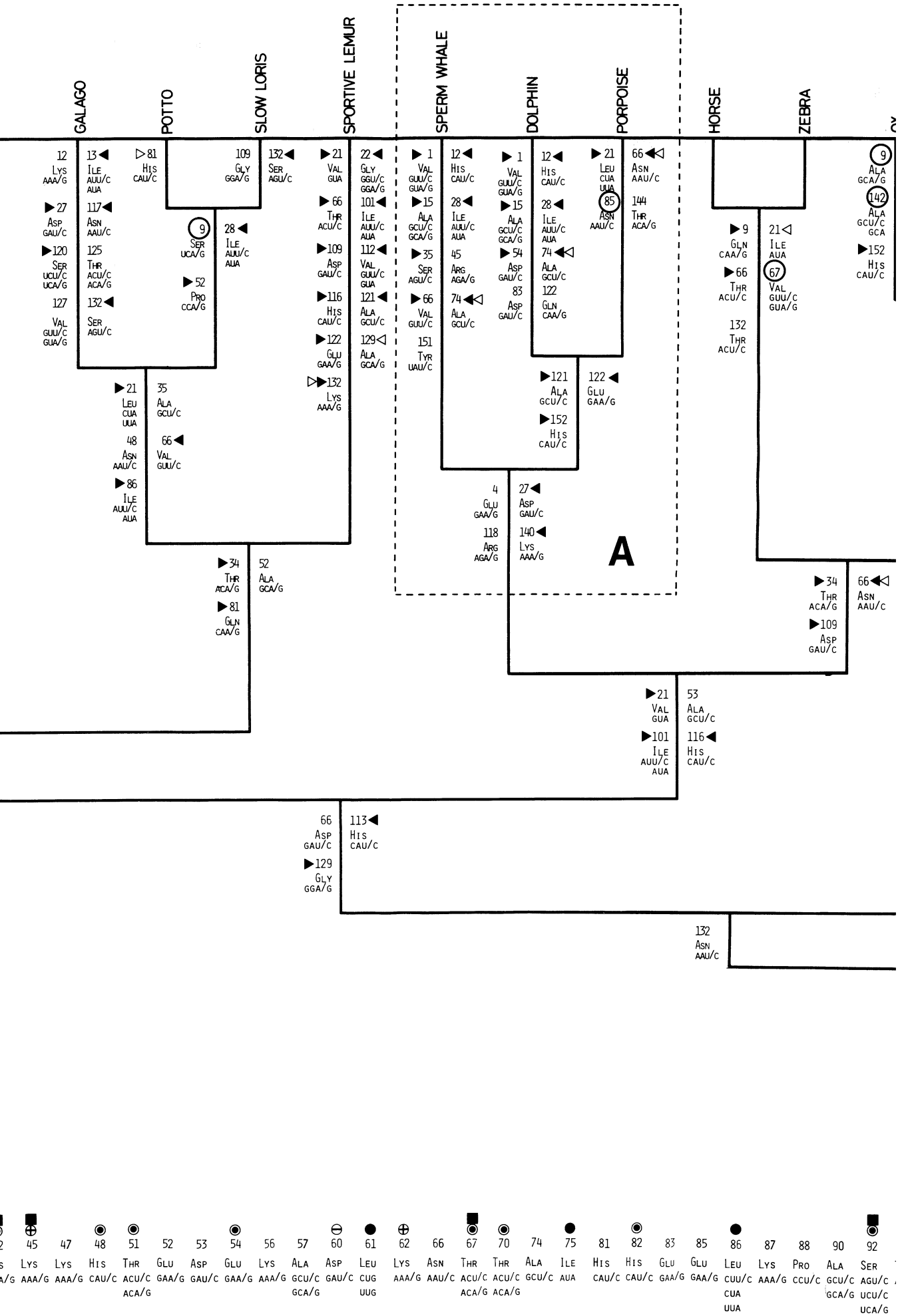
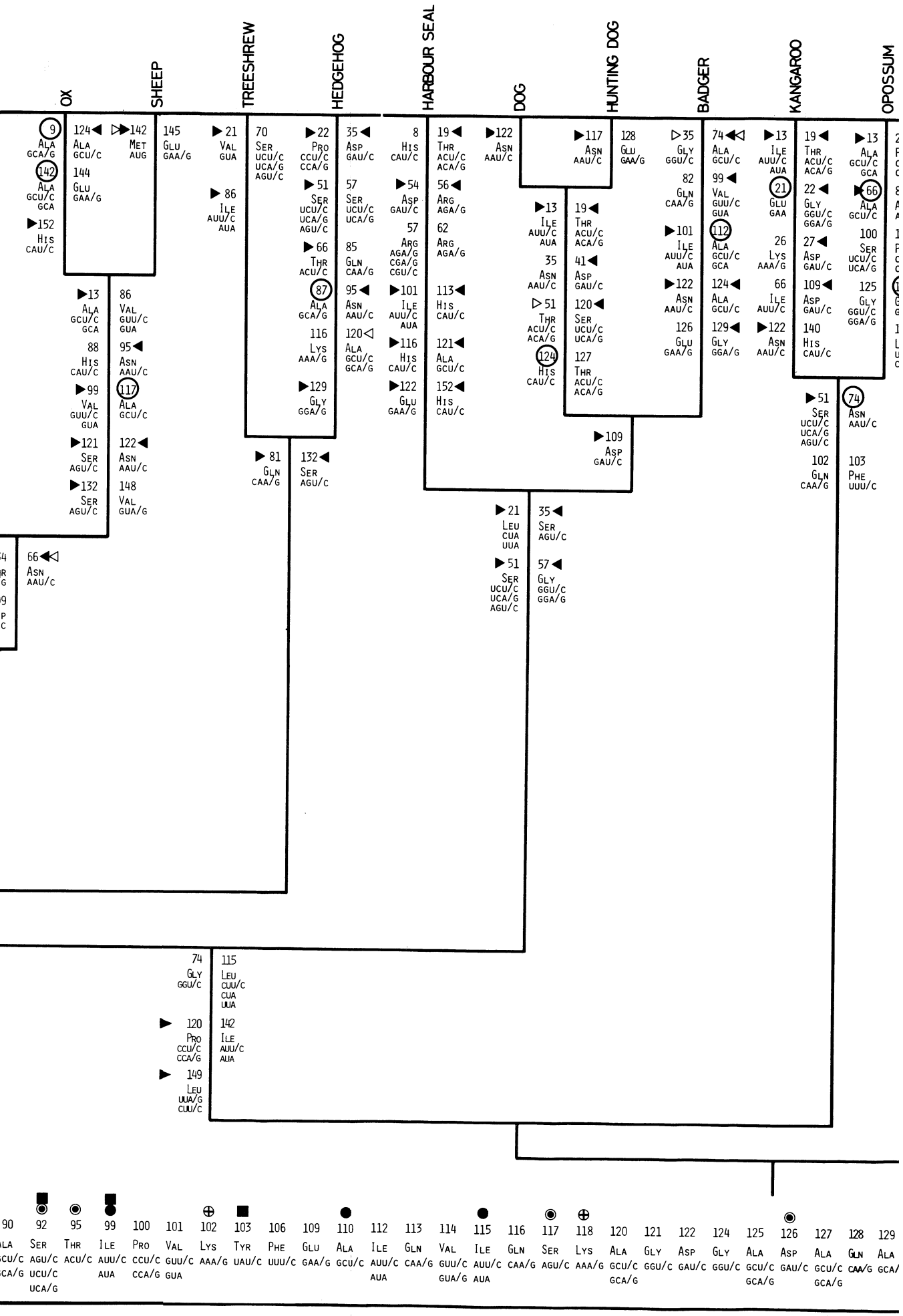
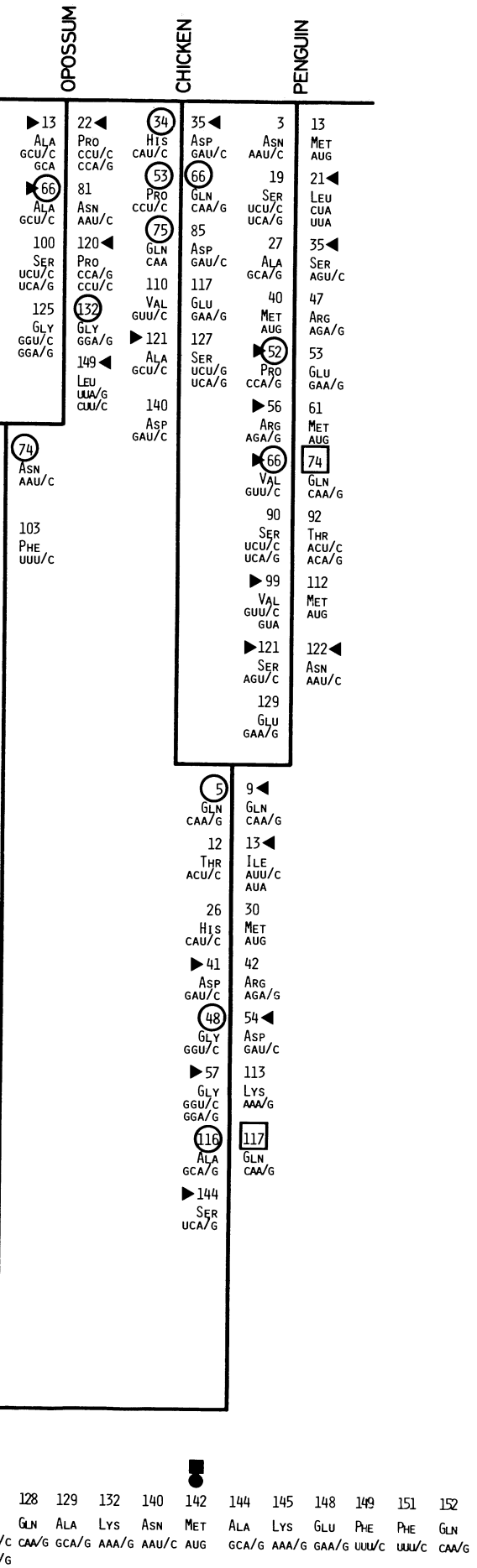
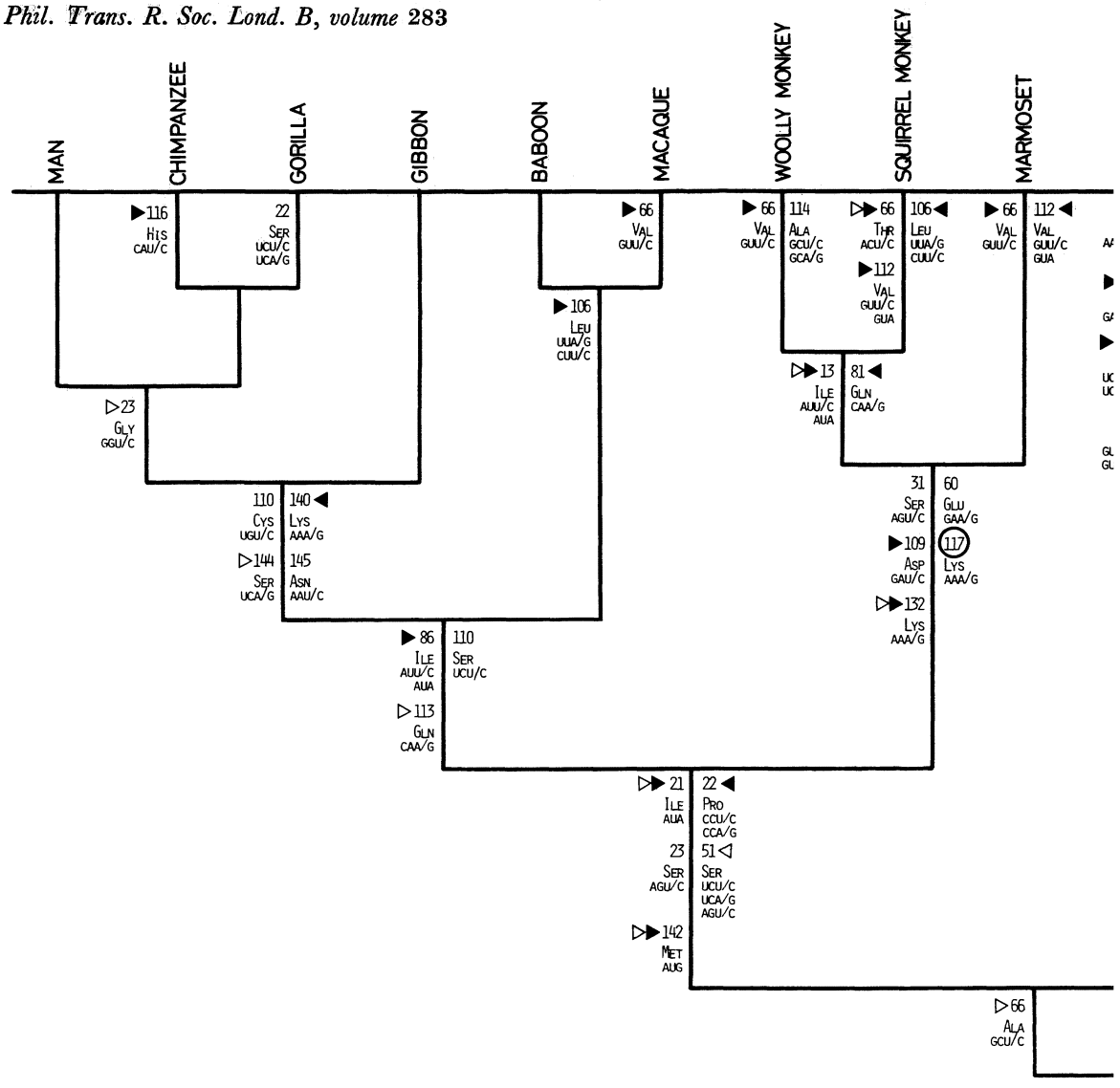


FIGURE 3. Cladogram 2. One of the two most parsimonious acceptable solutions. An alternative solution to block A is



block A is given in figure 4, and described in the text.


OPOSSUM
CHICKEN
PENGUIN
▶13 ALA GCU/C GCA
◀66 ALA GCU/C
100 SER UCU/C UCA/G
125 GLY GGU/C GGA/G
74 ASN AAU/C
103 PHE UUU/C
22◀ PRO CCU/C CCA/G
81 ASN AAU/C
120◀ PRO CCA/G CCU/C
132 GLY GGA/G
149◀ LEU UUA/G CUU/C
34 HIS CAU/C
53 PRO CCU/C
75 GLN CAA
110 VAL GUU/C
▶121 ALA GCU/C
140 ASP GAU/C
35◀ ASP GAU/C
66 GLN CAA/G
85 ASP GAU/C
117 GLU GAA/G
127 SER UCU/G UCA/G
3 ASN AAU/C
19 SER UCU/C UCA/G
27 ALA GCA/G
40 MET AUG
◀52 PRO CCA/G
▶56 ARG AGA/G
◀66 VAL GUU/C
90 SER UCU/C UCA/G
▶99 VAL GUU/C GUA
▶121 SER AGU/C
129 GLU GAA/G
13 MET AUG
21◀ LEU CUA UUA
35◀ SER AGU/C
47 ARG AGA/G
53 GLU GAA/G
61 MET AUG
74 GLN CAA/G
92 THR ACU/C ACA/G
112 MET AUG
122◀ ASN AAU/C
5 GLN CAA/G
12 THR ACU/C
26 HIS CAU/C
▶41 ASP GAU/C
48 GLY GGU/C
▶57 GLY GGU/C GGA/G
116 ALA GCA/G
▶144 SER UCA/G
9◀ GLN CAA/G
13◀ ILE AUU/C AUA
30 MET AUG
42 ARG AGA/G
54◀ ASP GAU/C
113 LYS AAA/G
117 GLN CAA/G
128 GLN CAA/G
129 ALA GCA/G
132 LYS AAA/G
140 ASN AAU/C
142 MET AUG
144 ALA GCA/G
145 LYS AAA/G
148 GLU GAA/G
149 PHE UUU/C
151 PHE UUU/C
152 GLN CAA/G



▶◀ PARALLEL MUTATIONS

▷◁ BACK MUTATIONS

◯ DOUBLE HITS

□ TRIPLE HITS

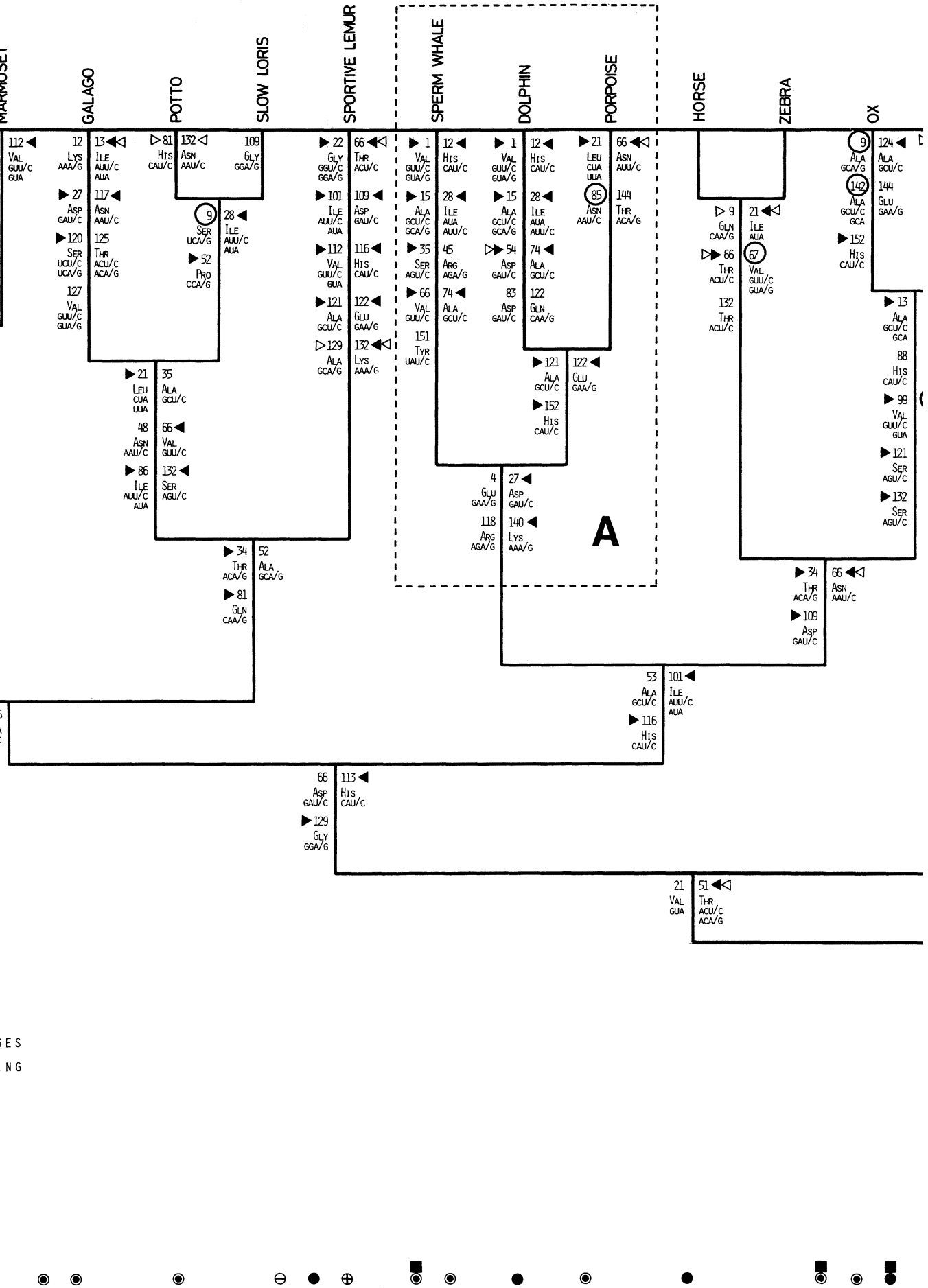
■ HAEM CONTACTS

● INTERNAL RESIDUES

⊕⊖ CHARGED RESIDUES FORMING SALT BRIDGES

◉ RESIDUES INVOLVED IN HYDROGEN BONDING

	◉	⊖					●						◉	⊖	●				◉			■◉	■⊕		◉
1	3	4	5	8	9	12	13	15	19	21	22	23	26	27	28	30	31	34	35	40	41	42	45	47	48
GLY GGU/C GGA/G	SER AGU/C	ASP GAU/C	GLN CAA/G	GLN CAA/G	GLN CAA/G	THR ACU/C	ILE AUU/C AUA	GLY GGU/C GGA/G	ALA GCU/C GCA/G	ILE AUA	ALA GCU/C GCA/G	GLY GGU/C	HIS CAU/C	GLU GAA/G	VAL GUA GUU/C	MET AUG	ARG AGA/G CGU/C	LYS AAA/G	GLY GGU/C	LEU CUG UUG	ASP GAU/C	ARG AGA/G	LYS AAA/G	LYS AAA/G	GLY GGU/C



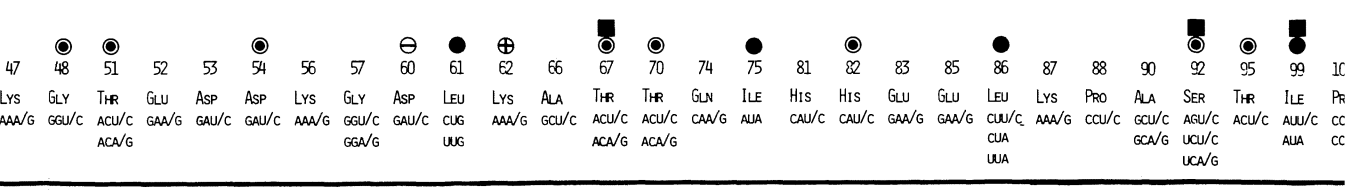
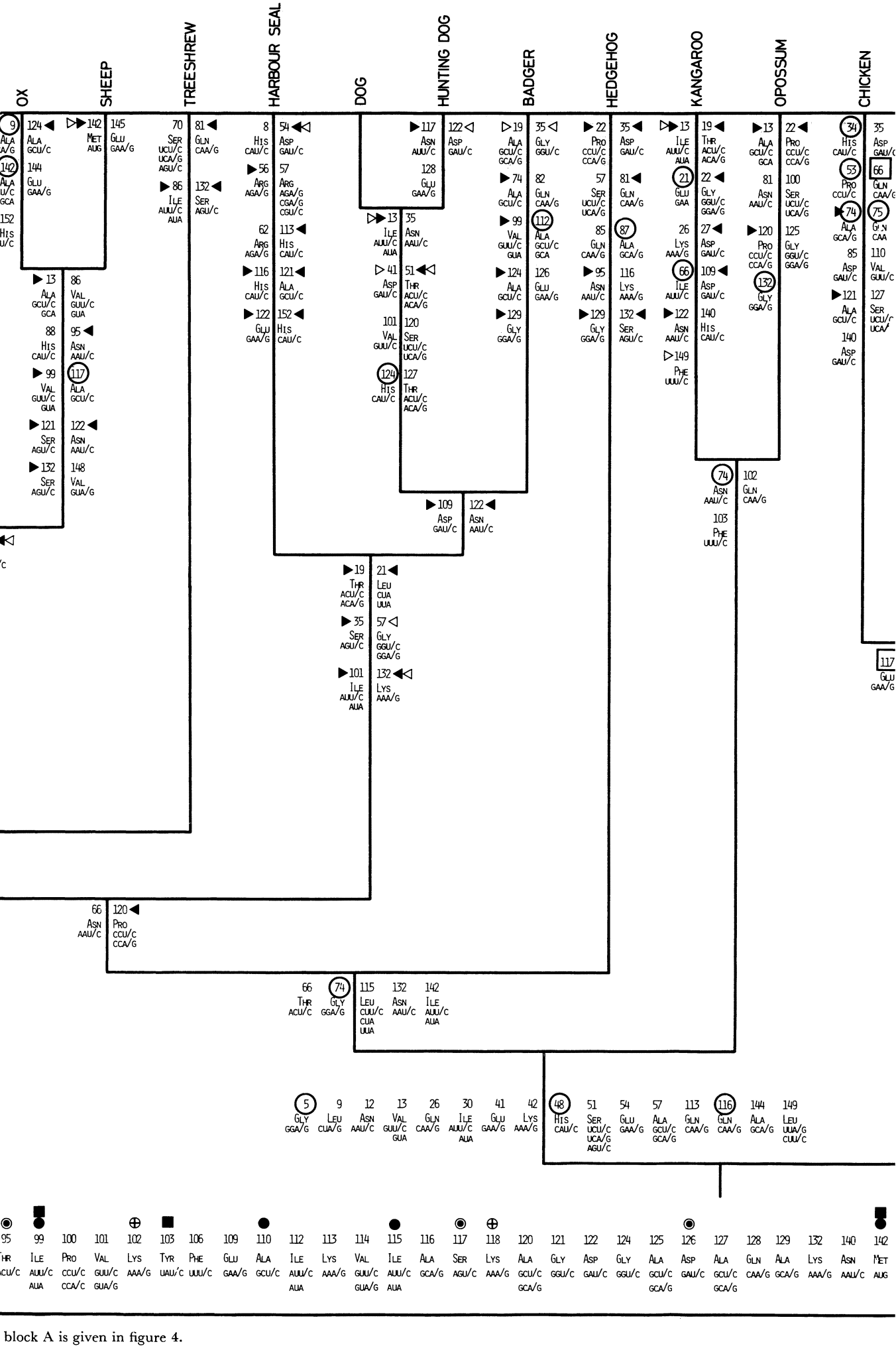
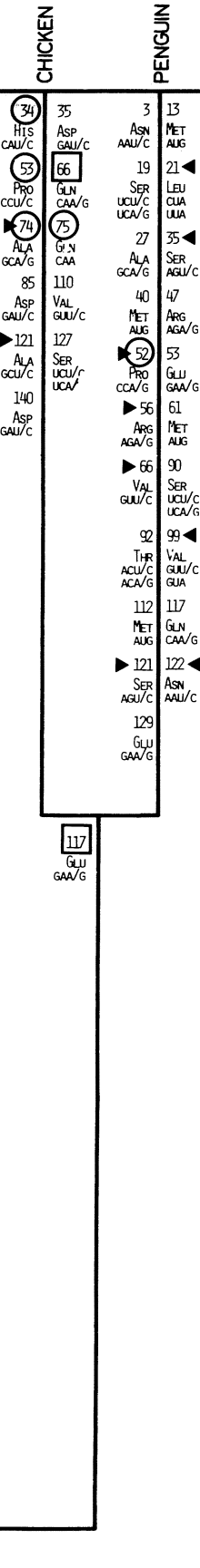


FIGURE 7. Cladogram 5. One of the two most parsimonious acceptable solutions. An alternative solution to block A is



block A is given in figure 4.



140	142	144	145	148	149	151	152
ASN AAU/C	MET AUG	SER UCA/G	LYS AAA/G	GLU GAA/G	PHE UUU/C	PHE UUU/C	GLN CAA/G

solutions are equivalent because they generate the same number of mutational events, but on the other hand these different interpretations have important repercussions on the number and kind of mutations attributed to each of the branches concerned, as will be described below.

3. In cladogram 2 all of the mutations which occurred in the bird and mammal stems were attributed to the avian side, but in cladogram 5 all have been assigned to the mammals. This redistribution affects the total number of mutations attributed to the lineage leading to each of the species here studied. The transfer of seventeen nucleotide substitutions from the bird stem to the mammal stem also results in changes in the reconstructed ancestral chain, introducing sixteen alternative residues for positions: 5, 9, 12, 13, 26, 30, 41, 42, 48, 54, 57, 66, 74, 113, 116 and 144.

Order Primates; suborder Anthropoidea

Five amino acids were fixed in the anthropoid common ancestor because residue 21 Ile has been added to the four already mentioned in the discussion of cladogram 2.

After the divergence within the anthropoids into catarrhines and platyrrhines, the pattern of mutations of the former remained the same as in cladogram 2, whereas in the phylogeny of the platyrrhines residue 112 Val now appears in parallel in the branches of the squirrel monkey and the marmoset, instead of 112 Val in the ceboid common stem and 112 Ile as a back mutation in the woolly monkey lineage.

Order Primates; suborder Prosimii

The common ancestry of the primates is still supported by residue 66 Ala, but changes at two positions have been introduced among the prosimians. In cladogram 2 the phylogeny of residue 132 was solved by attributing 132 Ser to the branches of the galago and the slow loris, but in the present interpretation 132 Ser has been considered as belonging to the lorisid stem, so leading to the assignment of 132 Asn to the potto as a back mutation. Residue 21 Val has been eliminated from the lineage of the sportive lemur because, in this cladogram, 21 Ile in the common ancestral chain has been replaced by 21 Val in the eutherian stem, and it remains as such for the Madagascan prosimian.

Order Condylarthra, and its derivatives

The number and nature of the amino acid substitutions given in cladogram 2 for the condylarthran derivatives (cetaceans and ungulates) has remained the same in cladogram 5, with the exception that residue 21 Val does not appear in the condylarthran common stem because it has been incorporated at an earlier stage (see figure 7).

The ancestry of the Orders Condylarthra and Primates

Compared to cladogram 2 no modifications have been introduced to this common stem in cladogram 5.

Order Scandentia

In the preceding cladogram the tree shrew had an immediate common ancestry with the hedgehog, which accounted for the shared residues 81 Gln and 132 Ser. In cladogram 5 the hedgehog has been displaced, and so the tree shrew lineage has had to incorporate the fixation of these two residues in addition to 70 Ser and 86 Ile. Furthermore, residue 21 Val is no longer

allocated to the tree shrew lineage because it has already been incorporated, together with 51 Thr, in the Scandentia–Condylarthra–Primate common ancestor, as shown in figure 7.

Order Carnivora

Compared with cladogram 2, several mutations among the carnivores have been redistributed in cladogram 5. Instead of assigning 19 Thr to the harbour seal and the canid stock in parallel, 19 Thr is now incorporated in the carnivore common stem, and it changes into 19 Ala in the badger lineage. Previously residue 101 Ile was assigned to the harbour seal and badger lineages in parallel, whereas in the present cladogram 101 Ile is attributed to the carnivore common stem, and then it (back) mutates into 101 Val for both canids. In cladogram 2, 122 Asn was allocated in parallel to the lineages of domestic dog and badger, whereas in cladogram 5, 122 Asn is incorporated in the fissiped stem and thereafter changes into 122 Asp for the hunting dog.

Finally, the 51 Ser previously allocated to the carnivore common stem has become unnecessary in the present interpretation because of the introduction of 51 Ser in the mammalian common stem. The opposite happens to residue 132 Lys which now appears in the carnivore common stem as a back mutation, because of the introduction of 132 Asn in the eutherian common stem, whereas in cladogram 2, 132 Lys was already present in the reconstructed ancestral chain.

Order Insectivora

In cladogram 2, eleven amino acids were fixed in the hedgehog lineage after its separation from the common ancestor of the Scandentia and Insectivora. In the present cladogram three of these residues have been eliminated from the erinaceoid lineage because 51 Ser and 66 Thr have been incorporated in the mammalian common stem and the eutherian common stem, respectively, and residue 120 Ala is already present in the reconstructed ancestral myoglobin chain. On the other hand, residues 81 Gln and 132 Ser have been removed from the joint Scandentia–Insectivora stem of cladogram 2, and transposed to the lineage of the hedgehog which, in cladogram 5, no longer shares immediate ancestry with the tree shrew.

Subclass Eutheria

Of the five residues assigned to this stem in the preceding cladogram, three have remained the same (74 Gly, 115 Leu and 142 Ile), two have been excluded (120 Pro and 149 Leu) and two new ones have been introduced (66 Thr and 132 Asn) in the present interpretation. The choice of 66 Thr, ACU/C, provides a bridge between the 66 Ala, GCU/C, of the common ancestral chain and the 66 Asn, AAU/C, allocated to the rest of the eutherians after the divergence of the insectivores. Lys is the common ancestral residue for position 132 in both cladograms. In cladogram 2 it changed into 132 Asn in the eutherian lineage after the separation of the carnivores, whilst in cladogram 5, Asn is present at this position in the eutherian stem and it changes to Ser and Lys for the hedgehog and carnivores, respectively, each by a single-point mutation. If the phylogeny of this position were kept as in cladogram 2, it would yield a double nucleotide substitution: ancestral Lys AAA/G—→ Ser AGU/C for the hedgehog, with an unknown intermediate.

Residue 74 Gly of the eutherian common stem deserves comment because, as can be seen in cladogram 2 (figure 3), it was the result of a single-point mutation, whereas in the present interpretation it is the result of a double hit. In the first case the common ancestral residue for

position 74 is Ala, GCU/C, which gave Gln, CAA/G (penguin), Asn, AAU/C (marsupials) and Gly, GGU/C (eutherians) by a triple, a double and a single nucleotide substitution respectively, whereas in the present solution the common ancestral residue has been chosen to be Gln, CAA/G, which yields Ala, GCA/G (chicken), Asn, AAU/C (marsupials) and Gly GGA/G (eutherians) as a series of three double hits; these two different amino acid and codon phylogenies are equally parsimonious.

Order Marsupialia

In cladogram 2, four amino acid substitutions were allocated to the marsupial common stem; in cladogram 5 one of these, 51 Ser, is eliminated because this residue has been incorporated at an earlier stage.

The substitutions attributed to the kangaroo and the opossum lineages have remained essentially the same in both cladograms, with two exceptions. First, in the previous interpretation 149 Leu was allocated to the opossum lineage because Phe was chosen as the common ancestral residue for this position, whereas in cladogram 5, 149 Leu has been assigned to the mammalian common stem, hence 149 Phe is introduced in the kangaroo lineage. Second, for the variable position 66 Ala (instead of Asn) has been selected as the residue in the common ancestral chain, and this leads to the allocation of 66 Val (penguin), 66 Gln (chicken) and 66 Ile (kangaroo) costing single, triple and double nucleotide substitutions, respectively. Because 66 Ala cannot be transformed into 66 Asn by a single-point mutation, then 66 Thr, found in the hedgehog, was chosen as the intermediate to be allocated to the eutherian common stem. We must emphasize, however, that because of the wide range of residues occupying position 66 several alternative phylogenies can be reconstructed, all equally parsimonious to those given in cladograms 2 and 5 at a minimum cost of 20 single hits.

Class Mammalia

In order to solve in the most impartial way the inevitable problem regarding the number and kind of amino acids which might have been incorporated either in the avian common stem or in the mammalian common stem, we assigned all substitutions to the former in cladogram 2 and, wherever possible, to the latter in this cladogram. In consequence, the reconstructed ancestral chain was transformed as shown in figure 7. In cladogram 5 a theoretical maximum of sixteen amino acid substitutions could have occurred along the mammalian common stem. Fourteen of these account for a minimum of seventeen nucleotide changes: 9 Leu, 12 Asn, 13 Val, 26 Gln, 30 Ile, 41 Glu, 42 Lys, 54 Glu, 57 Ala, 113 Gln and 144 Ala (which could be the outcome of single changes in their ancestral triplets) and 5 Gly, 54 Glu and 116 Gln (which require at least double steps). The two remaining substitutions, 51 Ser and 149 Leu, were incorporated in order to demonstrate an alternative phylogeny with the same cost as that given for these two positions in cladogram 2.

Class Aves; orders Sphenisciformes and Galliformes

For reasons explained in relation to cladogram 2, Ser, AGU/C, has been retained at position 117 in the ancestral myoglobin chain. This position is occupied by Gln in the penguin and Glu in the chicken. In order to avoid the introduction of two triple hits (costing six nucleotide substitutions) this codon phylogeny can be reconstructed in two different ways: Ser, AGU/C, changed to Gln, CAA/G, in the avian common stem requiring at least one change in each of

three bases, and subsequently this changed to Glu, GAA/G, in the chicken lineage by a single-point substitution; alternatively, Ser, AGU/C, could be transformed into Glu, GAA/G, in the avian common stem and this to Gln, CAA/G, in the penguin lineage. The two pathways are equally parsimonious. The first possibility was favoured in cladogram 2, and the second in cladogram 5. After the lineages of the two birds had diverged the chicken branch substituted 11 amino acids, requiring a total of 17 nucleotide changes, whereas the 21 amino acid substitutions in the penguin branch require a minimum of 22 base mutations. Residue 74 Ala has been introduced in the chicken branch as the result of a double hit, whereas 74 Gln has been removed from the penguin branch (compare figures 3 and 7), because the chosen residue for this position in the ancestral chain is Gln in the present interpretation, instead of Ala as in cladogram 2.

With Ala at position 66 in this reconstruction of the ancestral chain, the number of hits needed to arrive at the Gln of the chicken has increased from two to three, and for the same reason 66 Val in the penguin has become the outcome of a single substitution.

Cladogram 3 (see figure 8)

This cladogram is the second most parsimonious, scoring 282 hits. In this cladogram the first dichotomy amongst the eutherians is between a lineage common to the tree shrew and the hedgehog and another lineage leading to the rest of the eutherians included in this study.

Compared with cladogram 2, the shift in the position of the tree shrew and hedgehog

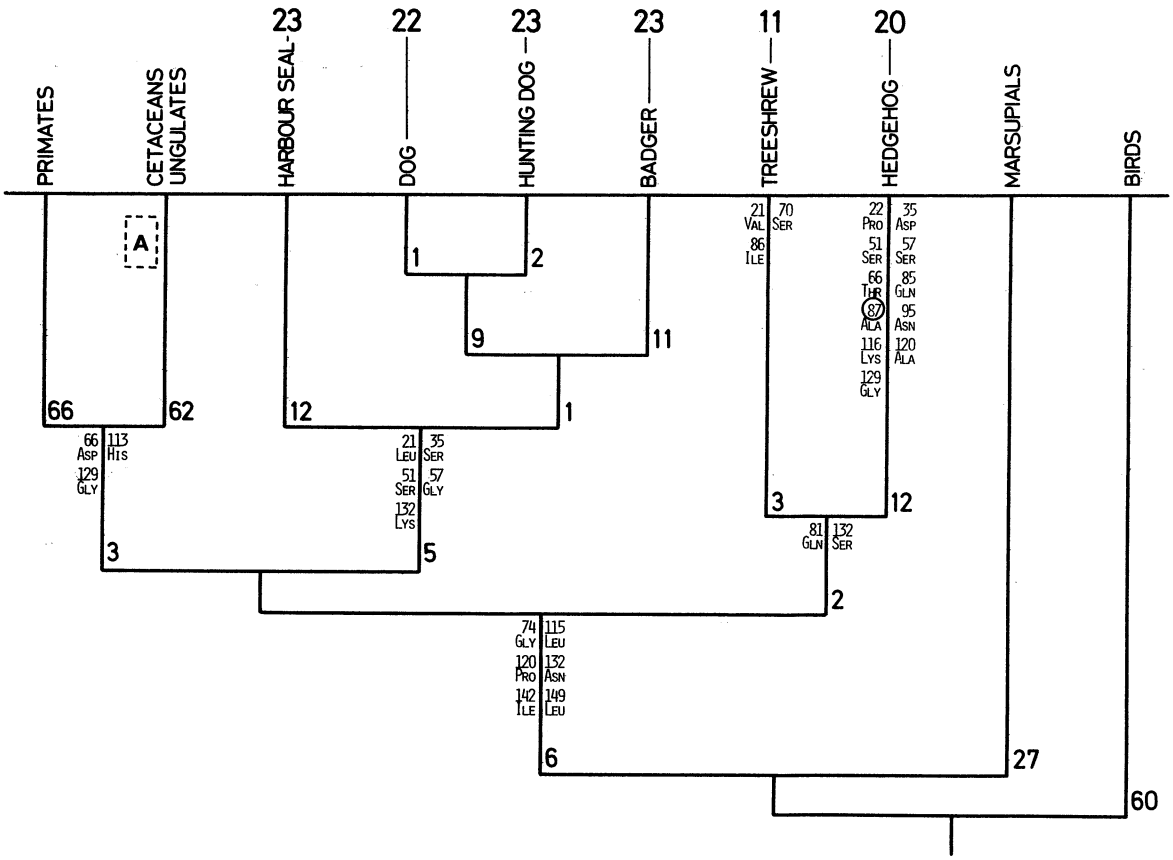


FIGURE 8. Cladogram 3.

increased the overall number of nucleotide substitutions by one, because 132 Lys has to be added to the carnivore common stem, owing to the attribution of 132 Asn to the common eutherian stem.

Cladogram 4 (see figure 9)

In cladogram 4 the sequence of origin of the eutherian lineages is as follows: tree shrew, Carnivora, hedgehog, Condylarthra and Primates. This cladogram has the same distribution of amino acids as that of cladogram 2, except that both 81 Gln and 132 Ser now appear independently along the lineages of tree shrew and hedgehog, because these animals no longer share immediate common ancestry. Therefore, the overall score has been increased to 283 hits.

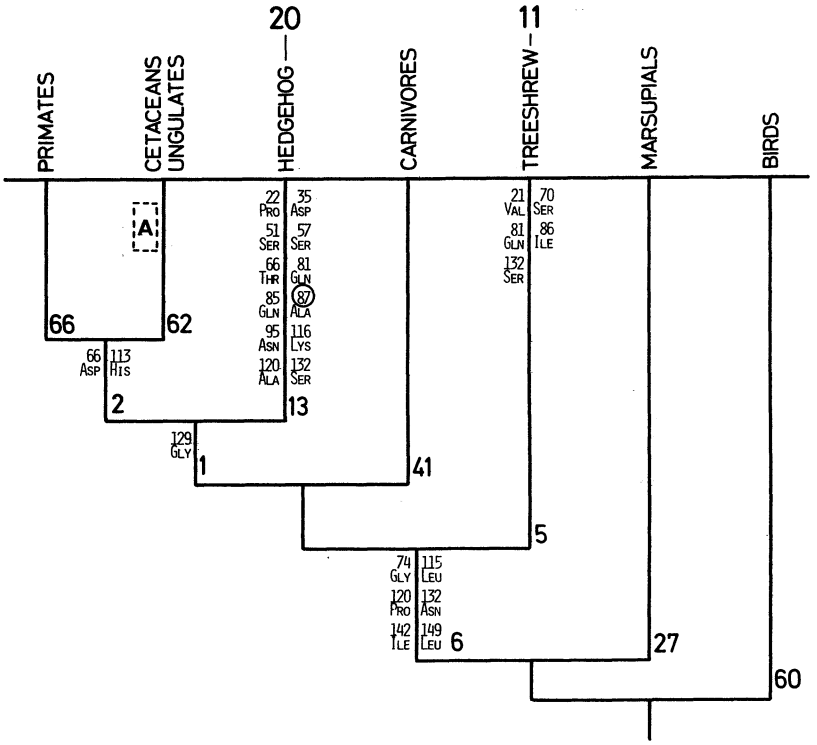


FIGURE 9. Cladogram 4.

Two other positions required modification. The 129 Gly of the hedgehog lineage and the 129 Gly of the common stem of Condylarthra and Primates (in cladogram 2) have been eliminated in the present cladogram because 129 Gly has been allocated to the common stem of the hedgehog, Condylarthra and Primates, thus saving one hit. On the other hand, the reallocation of 132 Asn to the eutherian common stem has led to the introduction of 132 Lys in the carnivore common stem as in cladogram 3 (see figures 8 and 9), so increasing the total cost by one hit.

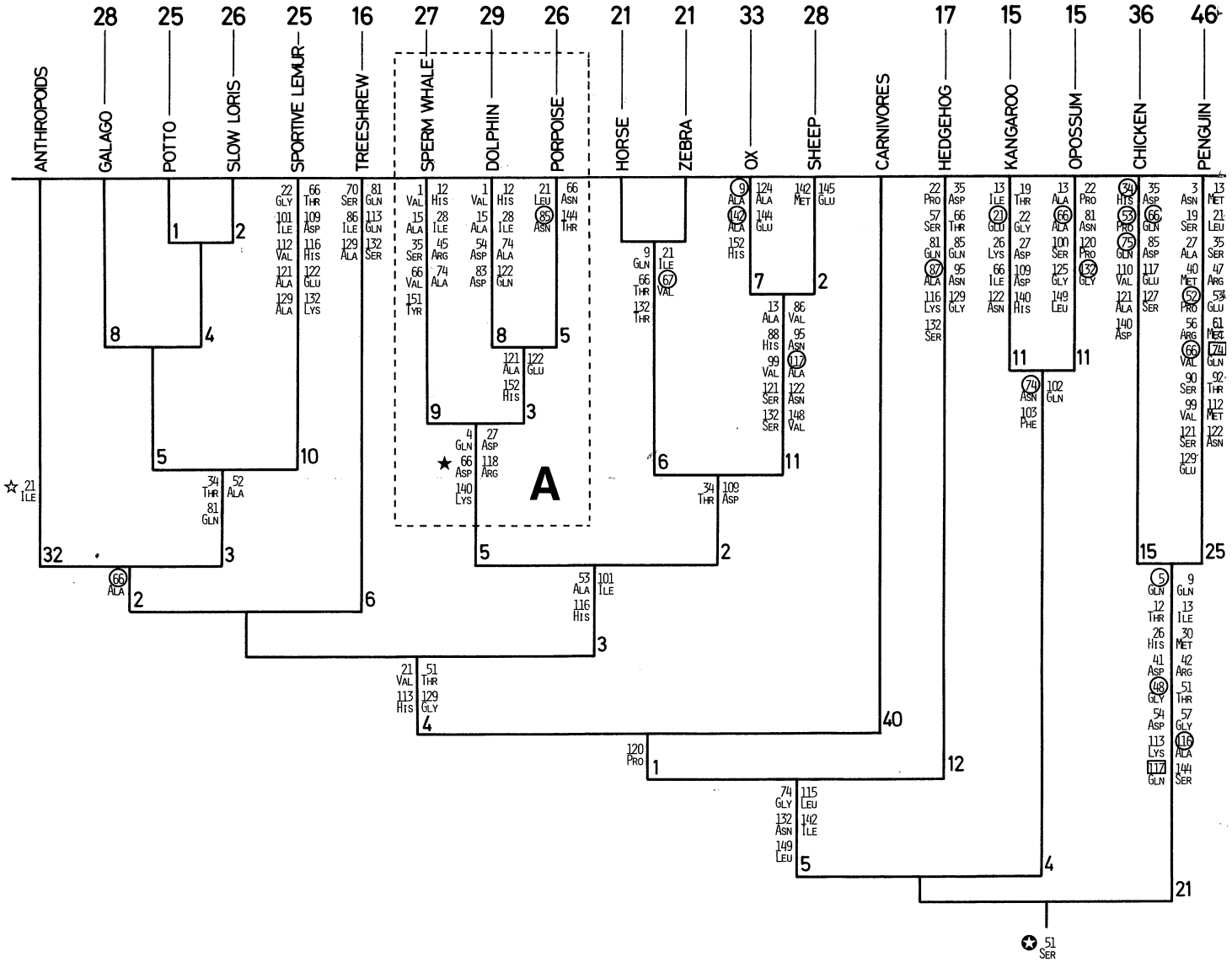


FIGURE 10. Cladogram 6.

Cladogram 6 (see figure 10)

In this cladogram the sequence of origin of the eutherian lineages is as follows: hedgehog, Carnivora, Condylarthra, tree shrew and Primates; the total score is 283 nucleotide changes. In order to obtain this number, several modifications were needed which will be discussed by comparison with cladogram 2.

Residue 21 Val has been removed from the condylarthran stem because it could be incorporated in the common stem of the Condylarthra, tree shrew and Primates; this makes unnecessary its allocation to the tree shrew and sportive lemur lineages, but has necessitated the assignment of 21 Ile to the anthropoid stem, thus saving one hit overall.

In this cladogram Ser has been chosen for position 51 in the reconstructed ancestral chain

instead of the Thr given in the previous four interpretations, because this substitution saves one hit.

In cladogram 2, residue 66 Asn was allocated to the ungulate stem because 66 Asp was fixed along the ancestral stem of Primates and Condylarthran. In the present interpretation 66 Asn of the ungulate stem has been retained from the ancestral chain, and so 66 Asp has been incorporated in the cetacean common stem. As 113 His and 129 Gly are present in the ancestral stem of Primates, tree shrew and Condylarthra, then single-point mutations are necessary to arrive at 113 Gln and 129 Ala in the tree shrew lineage. In comparison with cladogram 2, the back mutation to 120 Ala in the hedgehog lineage has become unnecessary, because the change to 120 Pro can be accommodated after the divergence of this lineage. Finally, the total number of hits accumulated by the carnivores remains at 40, as in cladogram 2, because the change to 51 Ser has been eliminated from their common stem, this residue being derived directly from the ancestral chain, but 132 Lys has been added owing to the assignment of 132 Asn to the eutherian stem.

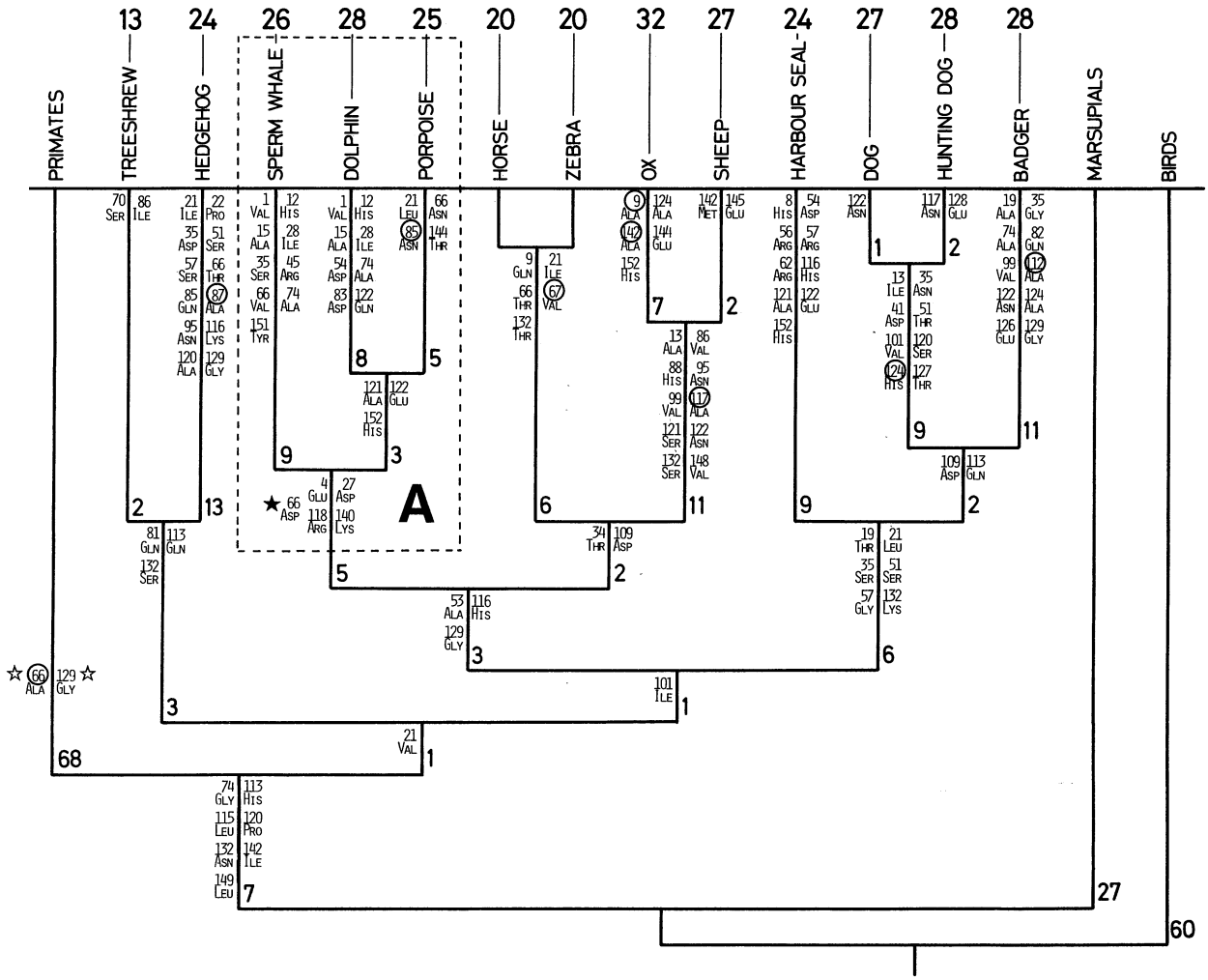


FIGURE 11. Cladogram 1.

Cladogram 1 (see figure 11)

In cladogram 1 the first divergence within the eutherians is between the Primates and the rest. The latter subsequently gave rise to the common stems of the tree shrew and hedgehog on the one hand and to the Carnivora and condylarthran derivatives on the other. This arrangement has led to changes at seven positions compared with cladogram 2, and scores a total of 283 detectable nucleotide substitutions.

Within the Carnivora there are two alternative phylogenies for position 19, which are equally parsimonious; in the first 19 Thr may be introduced as parallel events in the canid common stem and the harbour seal lineage (figure 3), whereas in the second 19 Thr has been assigned to the carnivore common stem and so 19 Ala must be introduced in the badger lineage, as shown in figure 11.

The change to 21 Val has been attributed to the common stem of the tree shrew, hedgehog, condylarthran derivatives and Carnivora, so leading to the elimination of this change from the tree shrew lineage and to the incorporation of 21 Ile in the hedgehog lineage. In this cladogram the 66 Asn of the ancestral myoglobin chain must be transformed into the 66 Ala of the primate stem through an unknown intermediate, requiring a double nucleotide substitution. The assignment of 101 Ile to the common stem of the Carnivora and condylarthran derivatives removes the need for the introduction of this change, as parallel events, in the condylarthran stem and in the lineages of the harbour seal and badger, but requires the introduction of 101 Val in the canid stem.

Two changes have been added to the eutherian stem: 113 His and 132 Asn. The 113 His changes into 113 Gln in the common ancestor of the tree shrew and hedgehog, and also in the fissiped stem; it also removes the change to 113 His in the harbour seal lineage. The presence of 132 Asn in the eutherian stem requires the introduction of 132 Lys in the Carnivora.

Finally, because the Primates and the Condylarthra do not share an immediate common ancestor, residue 129 Gly has to be assigned independently to both stems.

Cladogram 7 (see figure 12)

In this cladogram the sequence of origin of the various eutherian groups is: Carnivora, hedgehog, condylarthran derivatives, tree shrew and Primates. This phylogeny scores a total of 283 hits. In comparison with cladogram 2 residue 21 Ile of the ancestral myoglobin chain has mutated into 21 Val in the common stem of the condylarthrans, tree shrew and primates, so eliminating the need for this change in the sportive lemur and tree shrew lineages; it subsequently changed into 21 Ile in the anthropoids.

In the present interpretation residue 51 Ser has been chosen for the ancestral myoglobin chain, leading to its substitution by 51 Thr in the avian stem and in the ancestral stem of the Primates, tree shrew and Condylarthra; concomitantly this choice has removed the need for the incorporation of 51 Ser in the subsequent ancestry of the marsupials, carnivores and the hedgehog.

Because of the different placing of the tree shrew, residues 66 Asn, 113 Gln and 129 Ala have been incorporated as back mutations in this lineage. In addition, the separation of the tree shrew and hedgehog has necessitated the assignment of 81 Gln and 132 Ser independently to both lineages. Although we are not in a position to choose between cladograms 4, 6, 1, 7 and 8 on the basis of parsimony (each of them scored 283 hits) the placing of the origin of the tree

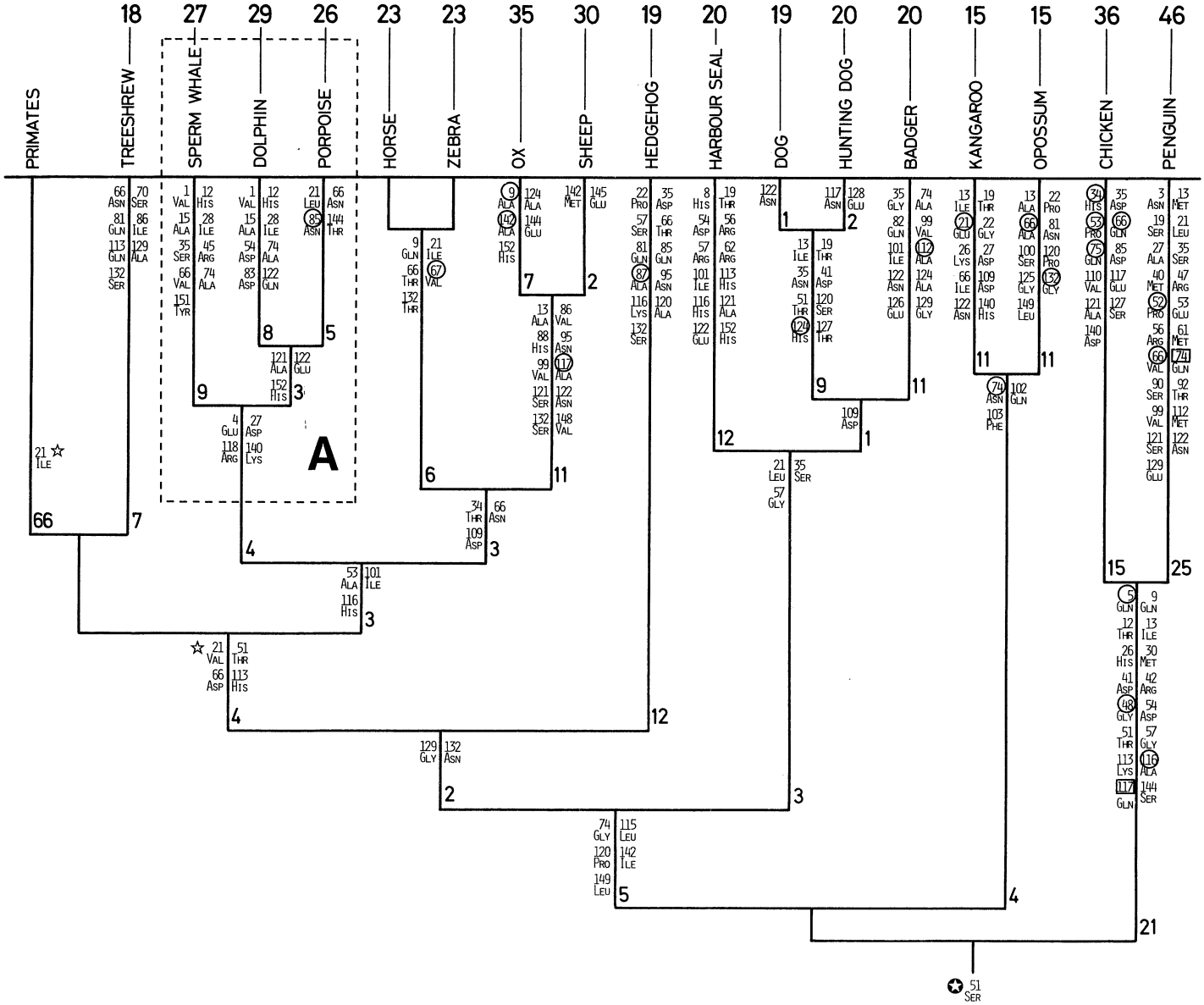


FIGURE 12. Cladogram 7.

shrew lineage in close association with the Primates in this cladogram has involved the introduction of a number of back mutations.

Cladogram 8 (see figure 13)

The phylogenetic relationships shown in this cladogram are the same as those of cladogram 2, with the exceptions that the tree shrew lineage has been moved to share immediate common ancestry with the prosimians, and the hedgehog lineage has been placed to share immediate common ancestry with the Condylarthra. This pattern is supported by the tree shrew sharing residues 21 Val and 129 Ala with the sportive lemur; 51 Thr and 81 Gln with all four prosimians and 86 Ile with the lorisids. Apart from the presence of residues 81 Gln and 86 Ile in the cebids

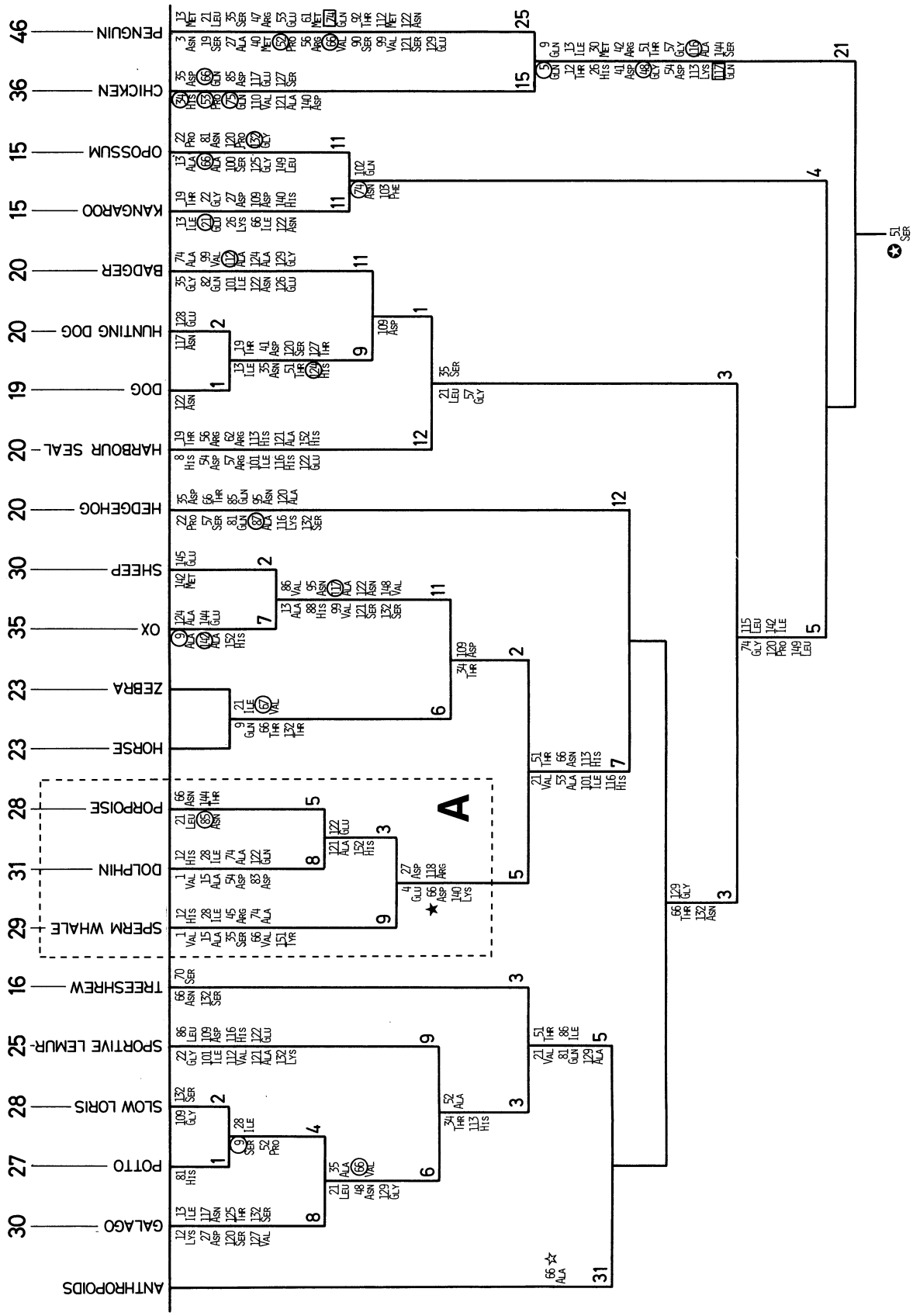


FIGURE 13. Cladogram 8.

and catarrhines, respectively, these five residues are confined to the tree shrew and prosimians amongst the Primates, and so they have been assigned to the common ancestor of the tree shrew and prosimians. In this cladogram, residue 51 Ser has been chosen as part of the ancestral myoglobin chain instead of 51 Thr, because it saves one nucleotide substitution. (In contrast, for position 51 in cladogram 2, Ser and Thr were equally parsimonious.) Although the change to 51 Ser is no longer necessary in the anthropoid stem, 66 Ala has been added, and so the total of 31 hits for the anthropoids remains the same as in cladogram 2.

Because of the new position of the tree shrew, the phylogeny of residues at position 66 has been increased to 22 hits, two more than in cladogram 2. Compared with the latter, 66 Asp in the common stem of Primates and Condylarthra, 66 Thr of the sportive lemur lineage and 66 Asn in the ungulate stem have been eliminated; consequently, in the present cladogram 66 Asn has been allocated in parallel to the condylarthran stem and the tree shrew lineage, while Asp has been incorporated in the cetacean stem. The addition of two nucleotide substitutions in the phylogeny of position 66 is due to the assignment of 66 Thr to the common ancestor of Primates, Condylarthra and the hedgehog which has led to the assignment of 66 Val to the lorisids as the outcome of a double nucleotide substitution. Once again, the placing of the tree shrew in association with the Primates has involved the introduction of a number of back mutations but, in this cladogram, some of these are shared with the prosimians.

Like cladograms 4, 6, 1 and 7, cladogram 8 requires a total of 283 nucleotide substitutions. Thus, of the eight alternative phylogenetic patterns investigated, two require 281 hits, one requires 282 and five require 283 hits. On the evidence at present available none of the solutions has a convincing advantage over the alternatives and the problem posed by the pentachotomy of figure 1 remains unresolved.

4. Analysis of difference matrices

We have employed several techniques in an attempt to evaluate phylogenetic trees without initially restricting branching patterns on zoological grounds. It is of particular interest to compare, in parallel with the full sequence information, trees based on the matrix of amino acid distances (table 4), because the value of a number of phylogenetic methods (such as comparative immunochemical ones) depends upon the production of matrices of distances, without the lengthy use of sequencing.

The original Wagner method (Wagner 1961) has been elaborated and presented in algorithm form by Farris (Kluge & Farris 1969; Farris 1970). The particular algorithm on which our program is based refers to a distance matrix (Farris 1972). The Wagner tree is one type of parsimonious tree in the sense that it seeks to present the data in terms of a smallest tree length (that is, the sum of the individual branch lengths). It is clearly desirable that there be the least difference between the distances in the data matrix and the corresponding distances obtained by summing the branch lengths on the path between the two relevant taxa included in the tree. In the case of sequence data, where the results are not so liable to experimental error as those derived from immunotaxonomic comparisons, it is in addition preferable that no pair of taxa shall be separated by a tree distance less than the entry for them in the data matrix, and the number of extra steps is a minimum estimate of convergence, parallelism and reversal. Farris uses the term 'homoplasy' in this sense (Farris 1967, 1972).

Several procedures (Fitch & Margoliash 1967; Goodman, Barnabas, Matsuda & Moore 1971) have been criticized (Farris 1972) on the grounds that they allow, and have produced,

TABLE 4. MATRIX OF AMINO ACID DIFFERENCES

(Inset: number of comparable sites.)

	MAN	CHIMPANZEE	GORILLA	GIBBON	BABOON	MACAQUE	WOOLLY MONKEY	SQUIRREL MONKEY	MARMOSET	GALAGO	POTTO	SLOW LORIS	SPORTIVE LEMUR	SPERM WHALE	DOLPHIN	PORPOISE	HORSE & ZEBRA	OX	SHEEP	HARBOUR SEAL	DOG	HUNTING DOG	BADGER	HEDGEHOG	TREESHREW	KANGAROO	OPOSSUM	CHICKEN	PENGUIN	LAMPREY	GLYCERA	APLYSIA	
MAN	0	1	1	1	6	7	16	17	14	23	16	19	22	24	26	20	18	29	25	24	22	23	20	15	13	22	15	34	44	105	102	107	MAN
CHIMPANZEE		0	2	2	7	8	17	18	15	24	17	20	21	23	25	19	17	28	24	23	23	24	21	16	14	23	16	34	44	105	102	107	CHIMPANZEE
GORILLA			0	2	7	8	17	18	15	23	16	19	22	24	26	20	18	29	25	24	22	23	20	16	13	22	16	34	44	105	102	107	GORILLA
GIBBON				0	5	6	15	16	13	24	17	20	23	25	27	21	19	30	26	25	23	24	21	17	14	23	16	35	45	105	103	107	GIBBON
BABOON					0	1	13	12	11	22	15	18	21	25	27	22	17	29	25	23	21	22	19	15	12	22	14	36	45	106	103	106	BABOON
MACAQUE						0	12	12	10	21	14	17	21	24	27	22	17	29	25	23	21	22	19	15	12	22	15	36	44	106	104	106	MACAQUE
WOOLLY MONKEY							0	4	4	21	20	19	19	28	31	26	18	29	25	25	21	21	21	18	17	22	18	37	45	107	105	106	WOOLLY MONKEY
SQUIRREL MONKEY								0	4	23	22	21	17	30	32	27	18	30	26	26	21	21	21	18	17	23	19	38	46	107	106	106	SQUIRREL MONKEY
MARMOSET									0	23	18	19	17	26	29	24	16	28	24	23	21	21	18	18	17	22	18	37	43	106	105	106	MARMOSET
GALAGO										0	12	11	21	26	30	25	21	30	27	28	21	20	27	22	16	30	26	37	42	102	105	109	GALAGO
POTTO											0	3	18	20	24	20	14	26	24	22	20	21	21	20	13	28	23	35	40	103	102	108	POTTO
SLOW LORIS												0	17	23	27	23	15	26	24	24	21	22	22	19	12	29	24	38	42	103	103	108	SLOW LORIS
SPORTIVE LEMUR													0	24	23	18	12	23	20	17	20	22	19	20	14	24	23	37	44	102	107	106	SPORTIVE LEMUR
SPERM WHALE														0	9	15	19	31	28	26	29	30	27	27	22	32	30	41	47	103	102	103	SPERM WHALE
DOLPHIN															0	12	21	29	28	23	31	33	27	30	23	32	32	40	48	104	105	102	DOLPHIN
PORPOISE																0	16	23	23	17	24	26	22	24	19	29	28	38	45	105	103	105	PORPOISE
HORSE & ZEBRA																	0	19	17	20	20	21	18	18	15	26	22	35	43	105	103	105	HORSE & ZEBRA
OX																		0	6	28	28	30	24	30	24	35	33	45	47	102	104	109	OX
SHEEP																			0	27	26	28	23	27	21	31	29	43	45	103	104	106	SHEEP
HARBOUR SEAL																				0	19	21	20	23	19	27	26	37	40	104	105	107	HARBOUR SEAL
DOG																					0	3	15	21	15	18	23	32	38	105	105	110	DOG
HUNTING DOG																						0	18	22	16	21	24	32	40	105	105	110	HUNTING DOG
BADGER																							0	21	17	23	24	39	41	104	105	106	BADGER
HEDGEHOG																								0	14	25	19	35	45	107	106	105	HEDGEHOG
TREESHREW																									0	21	16	35	43	103	105	108	TREESHREW
KANGAROO																										0	16	34	40	103	103	104	KANGAROO
OPOSSUM																											0	36	45	104	105	105	OPOSSUM
CHICKEN																												0	28	104	107	108	CHICKEN
PENGUIN																													0	104	104	110	PENGUIN
LAMPREY																														0	97	92	LAMPREY
GLYCERA																															0	106	GLYCERA
APLYSIA																																0	APLYSIA

	THERIANS	CHICKEN	PENGUIN	LAMPREY	GLYCERA	APLYSIA
THERIANS						
CHICKEN	153					
PENGUIN	152	152				
LAMPREY	136	136	135			
GLYCERA	136	136	135	124		
APLYSIA	142	142	141	131	130	

negative branch lengths. Whilst noting the difficulty of interpretation of such values, there can be partial justification where negative branch lengths are allowed in intermediate trees used by iterative procedures which tend finally to obtain parsimonious solutions with few or no negative values. Cook & Hewett-Emmett (1974) provide a short discussion of this issue.

The derivation of the equations employed to form a distance Wagner tree is given in Farris (1972), together with the necessary assumptions in dealing with matrices of amino acid distances. The technique utilizes distance equations which are particularly appropriate to the so-called Manhattan metric, and attempts to minimize the possible over-estimation of tree lengths where the amino acid distance values depart from the theoretical ideal. Essentially similar minimum evolution methods based on Euclidean distance have been used for gene frequency data by Cavalli-Sforza & Edwards (1967) and Thompson (1973), although in this context the procedures have been replaced by maximum-likelihood versions.

An attractive feature of the Wagner method is that it is free of assumptions concerning the homogeneity of rates of divergence, and it is therefore suitable for a preliminary investigation of rate differences amongst lineages. As a consequence, the patristic distance matrix does not necessarily have ultrametric structure (unless the data matrix does) in contrast to many commonly used techniques such as unweighted pair group clustering and single-linkage cluster analysis. If molecular evolution has indeed progressed divergently at constant rate, or even at

approximately constant rate within groups (Jardine, van Rijsbergen & Jardine 1969), then most clustering methods, given dissimilarities between present-day organisms, would generate dendrograms which justifiably could be interpreted as phylogenetic branching patterns. Several authors have suggested distortion statistics which can aid in assessing the departure of data matrices from ultrametric structure (Jardine & Sibson 1971; Estabrook 1972).

For real data it is apparent that, despite much theoretical support for constant rates of molecular evolution (Nei 1975), these should not be assumed. Also, it seems that a very high proportion of change may be in parallel; we have estimated some 50% for the myoglobin information presented in this paper, and Peacock & Boulter (1975), in a critical discussion of the relative merits of matrix and ancestral sequence methods in reconstructing phylogeny, estimate that 33% of all substitutions are non-divergent for higher plant cytochrome c. They emphasize that it is the proportion of such changes that affects accuracy, although on the basis of their computer simulations it appears that a false topology is likely to be chosen only when the total number of such changes is greater than that of shared derived substitutions. This is part of the justification of the maximum parsimony method of Goodman, Moore, Barnabas & Matsuda (1974).

In an experimental situation examined by Moore, Barnabas & Goodman (1973) in which a network was constructed so that very different rates of evolution were imposed, the Wagner technique proceeded immediately to a correct additive solution given the published matrix of distances. The iterative technique of Moore *et al.* uses an unweighted pair group method to approximate an example of an initial dendrogram and, as the authors note, this is inappropriate as a likely phylogeny. For the thirteen mammalian myoglobins used by Goodman *et al.* (1974) the Wagner method again proceeds immediately to an identical topology to that of their figure 3, and the solution produced for the haemoglobin alpha chains of nine animals approached the most parsimonious obtained by Cook & Hewett-Emmett (1974).

It seems, therefore, that as a rapid estimating technique, and as a time-saving starting procedure for an iterative strategy such as the sophisticated Moore and Goodman method, the Wagner technique has much in its favour. It should be noted that when a data matrix contains more than one minimum entry (indicating the initial pair of taxa to be linked), and when the tests for the next taxon to be linked to a developing network result in ties, there may be other (possibly more parsimonious) solutions, and these alternatives should be explored in the Wagner procedure.

The equations derived by Farris (1972) are designed to work with metric data. Difficulties are encountered when transformations are applied to amino acid distances in order to estimate undetected mutations during evolutionary history. We are grateful to Dr R. Holmquist for comments on this problem, and for introducing us to the work of Beyer, Stein, Smith & Ulam (1974). Some degree of inaccuracy may be introduced if a tree is produced on amino acid distances and the leg lengths are subsequently corrected individually. A clear demonstration of the desirability of using a suitably adjusted starting matrix is given by Holmquist (1972*c*), although a rather different empirical strategy has been used with some success in the augmented method of Goodman *et al.* (1974).

Defining a molecular sequence metric is difficult, and measures such as minimum mutation distance, the correction of Zuckerkandl & Pauling (1965) (used also by Dickerson 1971), the formula of Kimura & Ohta (1972) (derived in slightly different form also by Read 1975), and the random evolutionary hit values of Holmquist (Holmquist 1972*a*, *b*; Holmquist, Cantor &

Jukes 1972; Jukes & Holmquist 1972) can all lead to distance measures which do not necessarily obey the triangle inequality. A detailed treatment of this problem is given by Beyer *et al.* (1974), who emphasize that there should be a proper metric between contemporary molecular sequences, that metric properties should be imposed on the construction of inferred phylogenetic trees, and that these constraints should be model-free so that one avoids interpretations based on preconceived evolutionary models.

In the case of the myoglobin sequences used below, where all have the same number of residues, amino acid distance is equivalent to equation (4) of Beyer *et al.* (1974) who call this the Hamming metric. The Wagner method operating on these myoglobin amino acid distances thus fulfils all three of the above criteria, although it will readily be appreciated that the branch lengths of the resulting trees have undergone no correction for undetected mutations, and the longer branch lengths are consequently underestimated relative to the shorter. Beyer *et al.* propose a suitable, more sophisticated metric which we have not yet applied to the myoglobin data.

In this part of our study 28 mammals and the chicken were considered. The latter was included to provide an estimate of the position of the root of the tree and does not itself appear on the diagram. It rapidly became apparent that the opossum was sufficiently discrepant that it warranted special consideration. In the course of development of the network the opossum was assigned to the stem branch of the South American monkeys, and subsequently the kangaroo was joined to it, and the tree was finally rooted on the branch leading to the kangaroo (figure 14*a*).

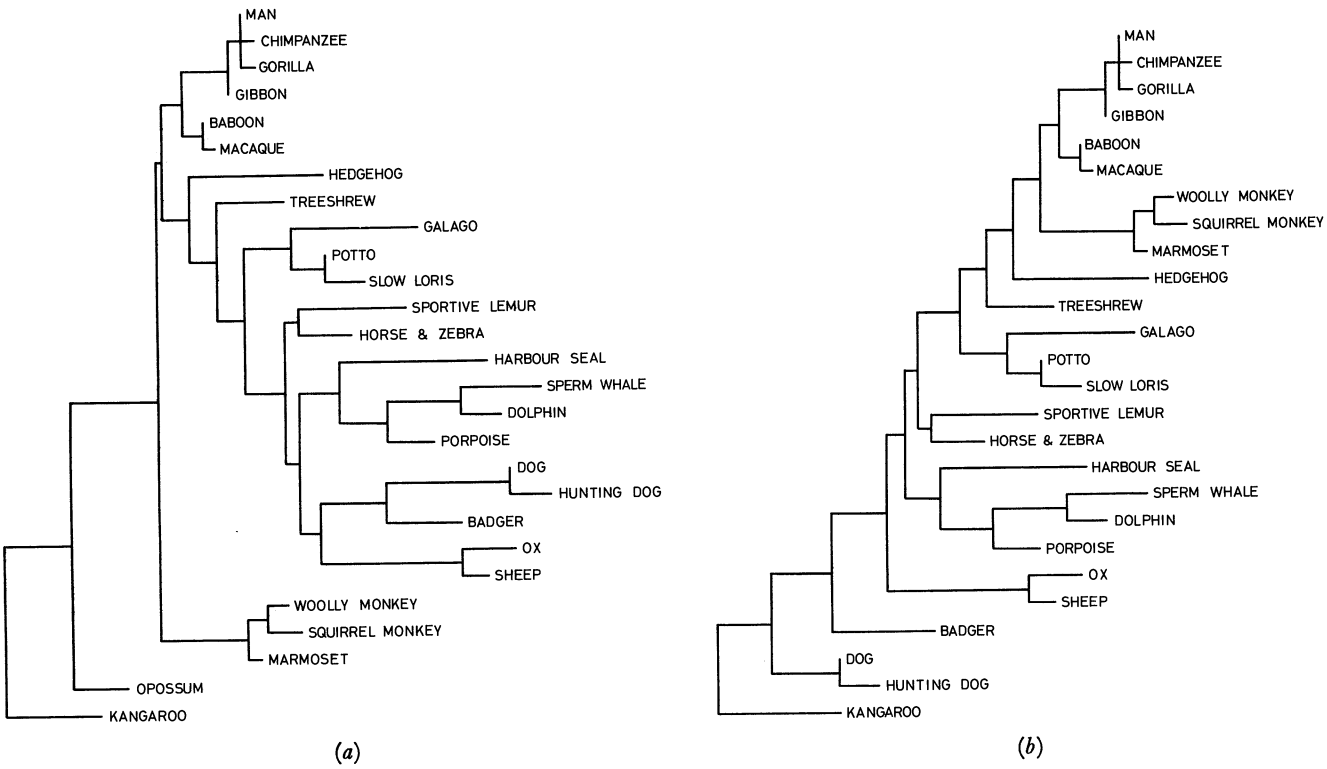


FIGURE 14. Wagner trees based on the matrix of amino acid distances. The branches are drawn to scale. (*a*) Opossum included; (*b*) opossum excluded.

When the opossum was excluded the tree assumed the pattern shown in figure 14*b*. The placings of a number of animals were unexpected on zoological grounds. It is, however, often a property of this type of analysis that if an animal is likely to have an alternative position which is relatively acceptable under the criteria of the tree, it shows significant homoplasy against those animals to which it might be linked. For example, the harbour seal, here linked with the cetaceans, has a moderately high homoplasy value against the other carnivores. Forcing it towards the carnivores would then result in a large homoplasy value between it and the cetaceans. It seems likely that there are functional reasons for the several substitutions acquired in parallel in the myoglobins of these diving mammals.

More surprising is the close placement of the horse and sportive lemur, but we have already discussed the residues common to these two animals (§3).

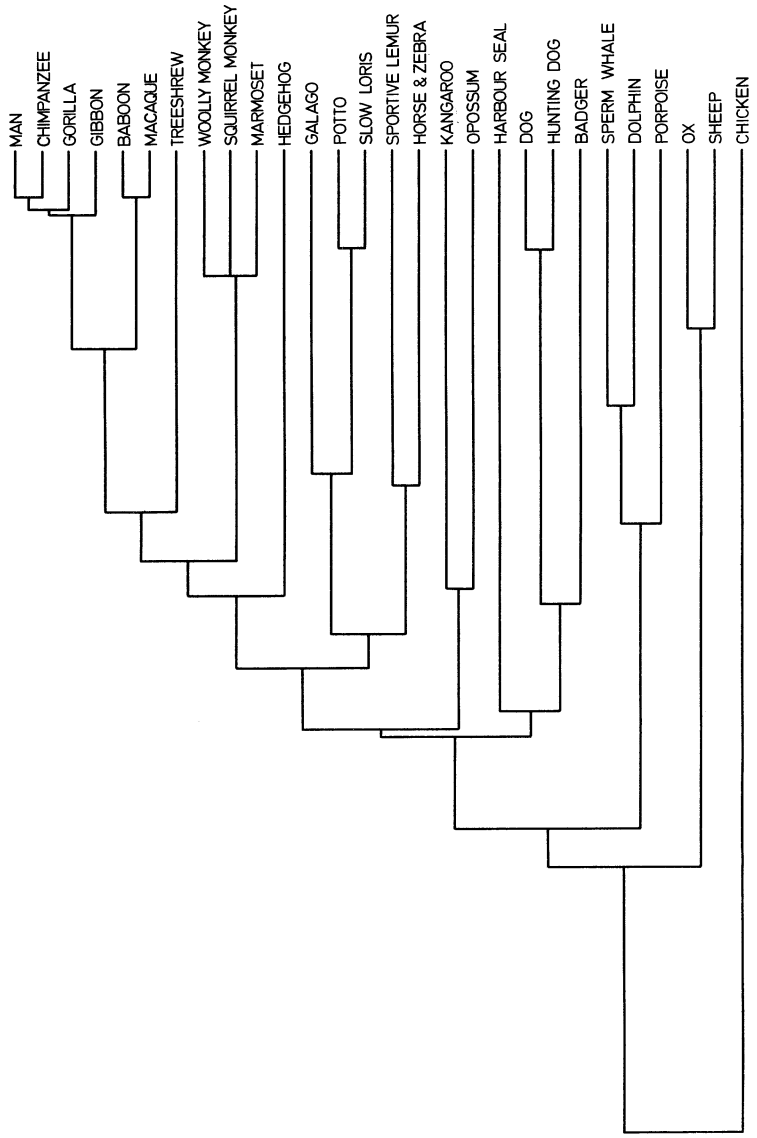


FIGURE 15. An unweighted pair group method analysis based on the matrix of amino acid distances. The branches are drawn to scale.

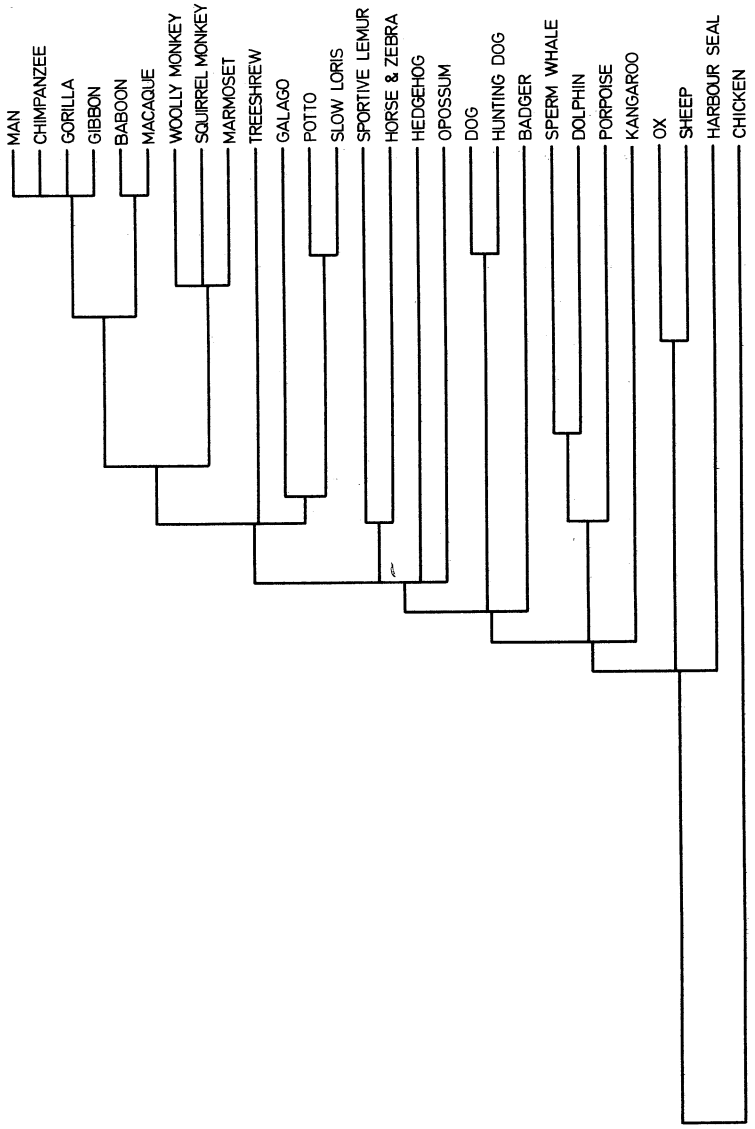


FIGURE 16. A single linkage cluster analysis based on the matrix of amino acid distances. The branches are drawn to scale.

The opossum has a lower amino acid distance from several of the catarrhines than it has from the kangaroo, and the problem it presents is most salutary in this context. Had its zoological affinity been equivocal (as is the case with hedgehog and tree shrew) one might have attempted to read some phylogenetic significance into its original anomalous placement. Because we believe that the opossum is a marsupial doubts do not arise, but it is necessary to bear this in mind when seeking hints concerning the affinities of the hedgehog and tree shrew.

The results of several other widely used clustering procedures are presented in dendrogram form in figures 15, 16 and 17. Farris (1972) is critical of the unweighted pair group method because of its sensitivity to rate differences reflected in the data matrix. An unweighted pair group clustering of the mammals and chicken is shown in figure 15 and it is noteworthy that the two marsupials are well embedded in the dendrogram. From our reconstructed cladograms

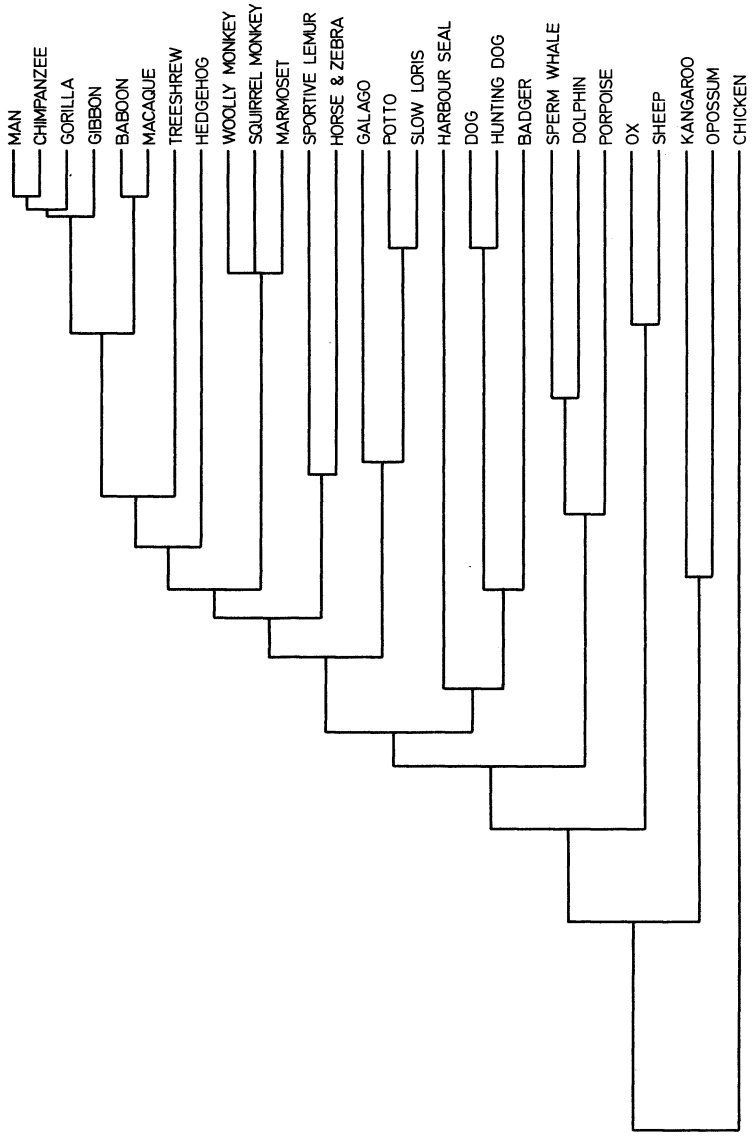


FIGURE 17. Cluster analysis by a method due to McQuitty (see text), based on the matrix of amino acid distances. The branches are drawn to scale.

(§3) the artiodactyls and cetaceans appear to have changed most since the mammalian common ancestor and are positioned in accordance with this.

Figures 16 and 17 show respectively the results of single-linkage cluster analysis and of a procedure due to McQuitty (1966). The former widely separates kangaroo and opossum, while the latter places the horse among the primates.

Certain groupings, such as the cetaceans, the apes, the artiodactyls and the two dogs (although not the carnivores as a whole) remain stable in all the results and can consequently be regarded as reliable (although the repeated clustering of horse with sportive lemur suggests pause for consideration). Zoologically, however, these stable groupings are not informative and, unfortunately, the most interesting issues cannot be regarded as unequivocally resolved.

Some properties of the amino acid distance matrix can be better appreciated in figure 18.

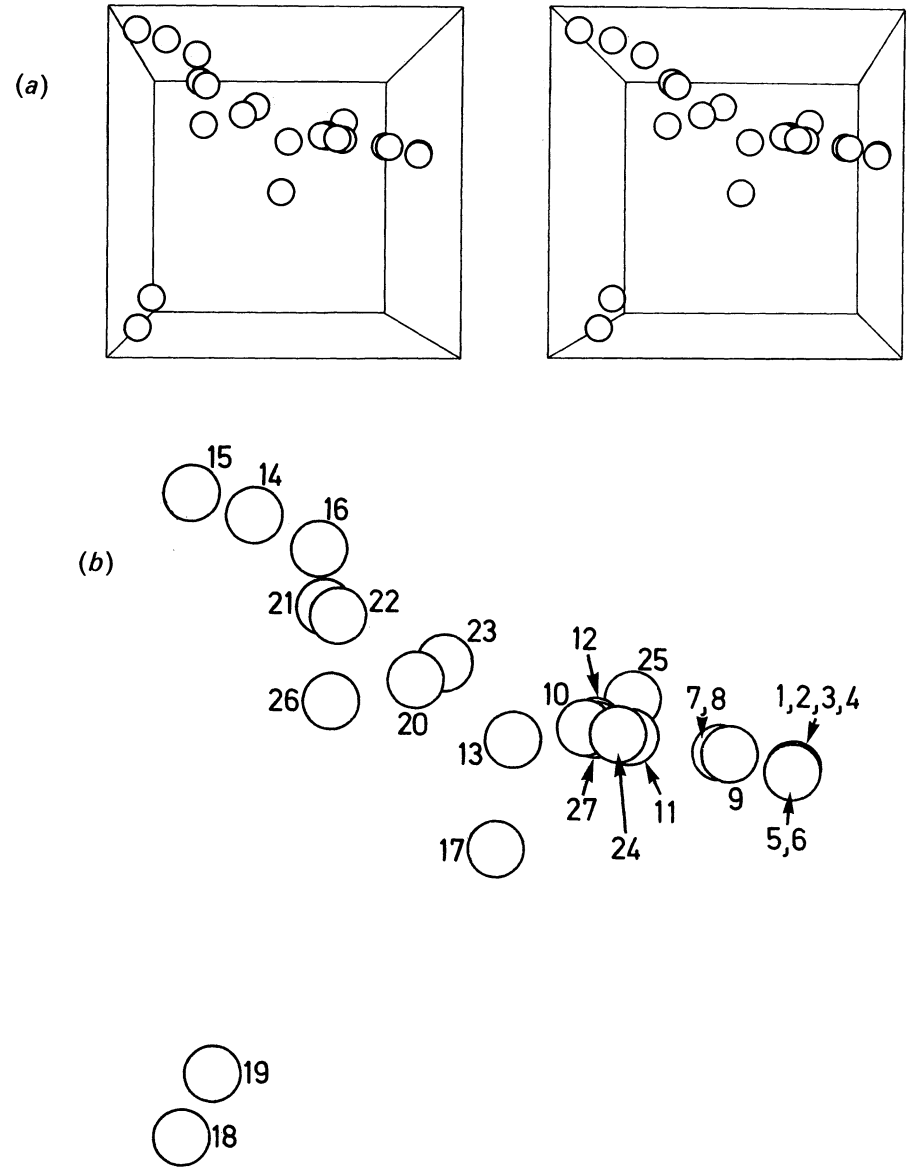


FIGURE 18. (*a*) A stereo-pair derived from the B_k method, combined with a principal-coordinates ordination, based on the matrix of amino acid distances.

(*b*) Key to animals represented in the stereo-pair:

1, man; 2, chimpanzee; 3, gorilla; 4, gibbon; 5, baboon; 6, macaque; 7, woolly monkey; 8, squirrel monkey; 9, marmoset; 10, galago; 11, potto; 12, slow loris; 13, sportive lemur; 14, sperm whale; 15, dolphin; 16, porpoise; 17, horse/zebra; 18, ox; 19, sheep; 20, harbour seal; 21, dog; 22, hunting dog; 23, badger; 24, hedgehog; 25, tree shrew; 26, kangaroo; 27, opossum.

It is too tempting for all but the most cautious of investigators to interpret immediately dendrograms summarizing hierarchical clustering procedures as phylogenetic branching patterns. Jardine & Sibson (1971) have commented on the danger of taking this step without further inspection of the nature of the data. Accordingly, figure 18 and table 5 summarize some results of the B_k method of Jardine & Sibson (1971) combined with an ordination using the principal

TABLE 5. B_k ($k = 3$) METHOD
(Species key in figure 18.)

edge length	species joined	cluster list
1.0	1–2, 5–6, 1–4, 1–3	1–2, 5–6, 1–4, 1–3
2.0	3–4, 2–3, 2–4	1–2–3–4
3.0	11–12, 21–22	11–12, 21–22
4.0	8–9, 7–9, 7–8	7–8–9
5.0	4–5	4–5
6.0	18–19, 4–6, 1–5	18–19, 4–5–6, 1–4–5
7.0	1–6, 2–5, 3–5	1–2–3–4–5–6
9.0	14–15	14–15
10.0	6–9	6–9
11.0	10–12, 5–9	10–12, 5–6–9
12.0	10–11, 6–25, 6–8, 6–7, 5–8, 5–25, 15–16, 12–25, 13–17	10–11–12, 5–6–7–8–9, 5–6–25, 15–16, 12–25, 13–17
13.0	4–9, 3–25, 11–25, 1–25	11–12–25, 1–2–3–4–5–6–7–8–9–25
14.0	24–25, 11–17, 5–27, 13–25, 6–11	24–25, 11–17, 5–27, 13–25, 6–11–25
15.0	1–24, 21–25, 5–11, 5–24, 21–23, 17–25, 12–17, 6–27, 6–24, 14–16, 1–27	21–25, 21–23, 13–17–25, 11–12–17–25, 14–15–16, 1–2–3–4–5–6–7–8–9–11–24–25–27

coordinates technique of Gower (1966). The animals are identified in the figure, but it is worth commenting that the groups with the highest rates of evolution – the artiodactyls and cetaceans – are widely separated in the lower left and upper left corners of the frame respectively. The dendrogram type of presentation (which is not helpful for B_k when k is greater than 1) does not reveal so forcefully that these two groups are not only relatively distant from the other animals, but are also widely separated from one another. In practice most evolutionary biologists would regard it as axiomatic that phylogeny can be accurately presented in tree form; the issue here is how far one can allow a presentation of the structure of the data to influence one's ideas concerning phylogeny.

Through the courtesy of Dr M. Goodman and Dr G. W. Moore, we were given access to their iterative programs for the parsimonious construction of cladograms (Moore *et al.* 1973*a*, *b*; Goodman *et al.* 1974), and a study using this technique on the collected myoglobin sequences will be published elsewhere.

TABLE 6. MUTATIONAL LENGTHS OF VARIOUS TREES

starting topology	initial length	final length
Wagner, figure 14*a*	246	242
Wagner, figure 14*b* plus opossum joined with kangaroo	246	242 (figure 19)
unweighted pair group	256	247
figure 2, cladogram 2	249	247
McQuitty, figure 17	256	249
random	400	

To return to some of the topologies figured here, it is of interest to examine their mutational tree lengths (table 6). Radically different solutions to the pentachotomy of figure 1 make relatively little difference to overall parsimony. This appears to be a feature both of the data

and of the technique, which undervalues those areas of the cladogram which have a lower density of representation.

It is noteworthy that of the topologies listed in table 6 that of figure 14*a*, which is highly improbable zoologically, had one of the lowest mutational tree lengths. This topology led to a solution, at length 242, which entailed some internal rearrangement but was equally improbable zoologically. However, by taking the topology of figure 14*b*, adding opossum to the kangaroo branch, and then submitting the resulting pattern to the iterative procedures, we obtained a further solution at length 242 (figure 19). Rather more of this solution is acceptable from a zoological point of view.

Finally, as a demonstration that quite simple and rapid methods may provide valuable information about sequence distance data, figure 20 shows a minimum spanning tree based on the mammals and the chicken, together with a reconstructed ancestral myoglobin sequence for eutherian mammals based on cladograms 2 and 5 (figures 3 and 7). The result reflects the major groupings noted above and suggests some confidence in the procedures employed in the cladograms.

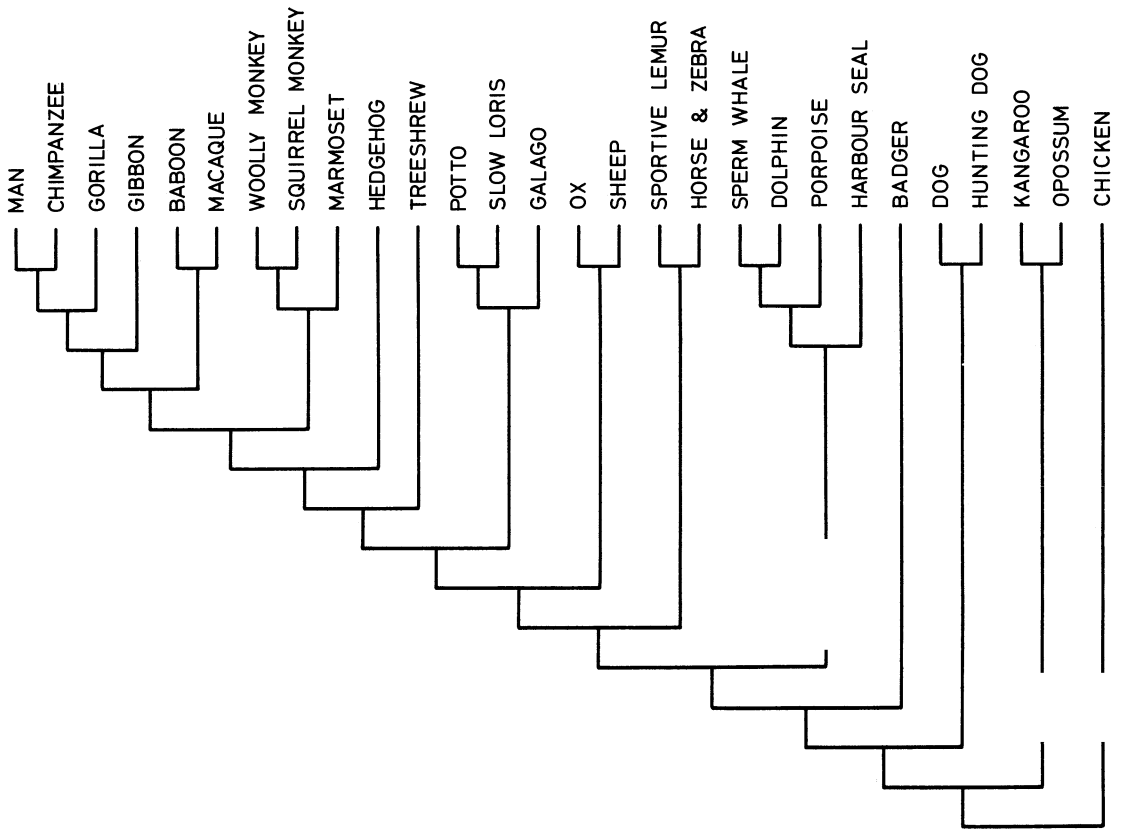


FIGURE 19. A tree derived by the method of Moore and Goodman, from the topology of figure 14*b* (with opossum added to the kangaroo lineage). The resulting topology, not drawn to scale, was one of several with a length of 242 hits, but some placements are unacceptable from a zoological point of view.

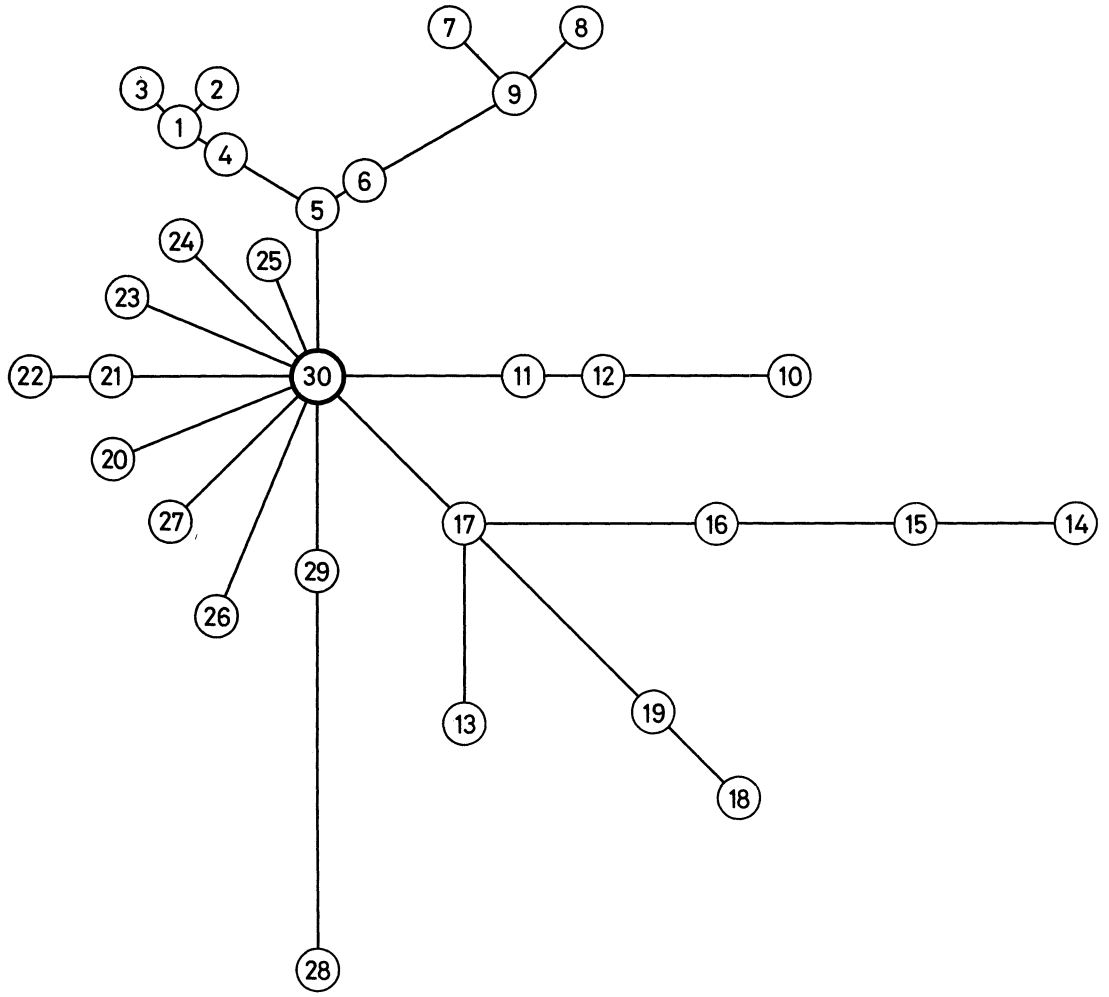


FIGURE 20. A minimum spanning tree based on the matrix of amino acid distances, with the addition of distances derived from comparisons with a reconstructed ancestral eutherian sequence. Numbers as for figure 18 with the addition of 28, chicken; 29, rabbit and 30, reconstructed ancestral sequence. The branches are drawn to scale.

5. FUNCTIONAL MORPHOLOGY OF THE MOLECULE

The availability of amino acid sequences of 31 myoglobins belonging to mammals and birds, and that of the mollusc *Aplysia* (for which an alignment is presented in table 2) provides an opportunity to study the functional relevance of certain positions at which the amino acids appear to have remained unchanged, or have changed conservatively, during about 300 Ma of evolution. Our structural analysis is based on the crystallographic model of sperm whale myoglobin and it is assumed that there is an overall resemblance of the three-dimensional structure in all myoglobins.

Most vertebrate myoglobins so far studied have 153 amino acid residues; the number is 155 in the minor myoglobin fraction found in some South American monkeys (Romero-Herrera & Lehmann 1973*a*), 152 in the penguin (Peiffer 1973) and 145 in the mollusc *Aplysia* (Tentori *et al.* 1973). They can be classified as follows:

1. According to its secondary structural configuration the myoglobin polypeptide chain is divided into eight right handed α helical segments and seven non-helical segments established

by X-ray crystallographic analysis (Kendrew 1962; Perutz *et al.* 1965; Perutz 1965; Watson 1969). The helical segments are designated: A (residues 3–18); B (residues 20–35); C (residues 36–42); D (residues 51–57); E (residues 58–77); F (residues 86–94); G (residues 100–118) and H (residues 124–149). The non-helical segments are: NA (residues 1–2); AB (residue 19); CD residues 43–50); EF (residues 78–85); FG (residues 95–99); GH (residues 119–123) and HC) (residues 150–153). It may be noted that there are no inter-helical sections BC and DE.

Table 7. Variation in bonding structure

SIDE CHAIN HYDROGEN BONDS

	Hydrogen bond No. (Fig. 21)	Sequential number	Helix key	Hydrogen bonded side chains	αCO, αNH or NH IMIDAZOLE	Other amino acids found at the same position
Inter-segmental	1	26	B7	Gln	αCO(56)	Lys(kangaroo), His(birds)
	2	42	C7	Lys	αCO(98)	Arg(birds)
	3	141	H18	Asp	IMIDAZOLE(82)	
	4	146	H23	Tyr	αCO(99)	
	Hydrogen bond No. (Fig. 22)					
Intra-segmental	5	3	A1	Ser	αNH(6)	
	5	6	A4	Glu	αNH(3)	
	6	20	B1	Asp	αNH(23)	
	7	24	B5	His	αCO(20)	
	8	35	B16	Ser	αCO(31)	Asp(hedgehog, chicken), Asn(dogs), Ala(lorisids),
	9	39	C4	Thr	αCO(36)	Gly (other primates, other cetaceans, ungulates, sea lion, badger, treeshrew, marsupials)
	10	48	CD6	His	αCO(45)	Asn(lorisids), Gly(birds)
	11	51	D1	Thr	αNH(54)	Ser(marsupials, hedgehog, carnivores except dogs, anthropoids)
	12	54	D4	Glu	αNH(50)	
	13	58	E1	Ser	αNH(61)	
	14	67	E10	Thr	αCO(63)	Val(perissodactyls)
	15	70	E13	Thr	αCO(66)	Ser(treeshrew)
	16	92	F7	Ser	αCO(89)	Thr(penguin)
	17	93	F8	His	αCO(89)	
	18	95	FG1	Thr	αCO(91)	Asn(hedgehog, artiodactyls)
	19	108	G9	Ser	αCO(104)	
	20	117	G18	Ser	αCO(113)	Lys(new world monkeys), Asn(galago, hunting dog), Glu(chicken) Gln(penguin), Ala(artiodactyls)
		38*	C3	Glu	αNH(38)	
		126*	H3	Asp	αNH(126)	

*Amino acids hydrogen bonded to their own α imino group

SALT BRIDGES

	Sequential number	Helix key			Sequential number	Helix key	Salt bridge variants
Inter-segmental	4	A2	Asp	→ Lys	79	EF2	
	4	A2	Glu	→ Lys	79	EF2	cetaceans
	18	A16	Glu	→ Lys	77	E20	
	20	B1	Asp	→ Lys	118	G19	
	20	B1	Asp	→ Arg	118	G19	cetaceans
	27	B8	Glu	→ Lys	118	G19	
	27	B8	Asp	→ Arg	118	G19	cetaceans
	27	B8	Asp	→ Lys	118	G19	kangaroo, galago
	**27	B8	Ala	Lys	118	G19	penguin
	45	CD3	Lys	→ Asp	60	E3	
	45	CD3	Arg	→ Asp	60	E3	sperm whale
	45	CD3	Lys	→ Glu	60	E3	New World monkeys
Intra-segmental	59	E2	Glu	→ Lys	62	E5	
	59	E2	Glu	→ Arg	62	E5	harbour seal
	102	G3	Lys	→ Glu	105	G6	
	**102	G3	Gln	Glu	105	G6	marsupials

**missing salt bridges

The salt bridges between haem and globin are not included.

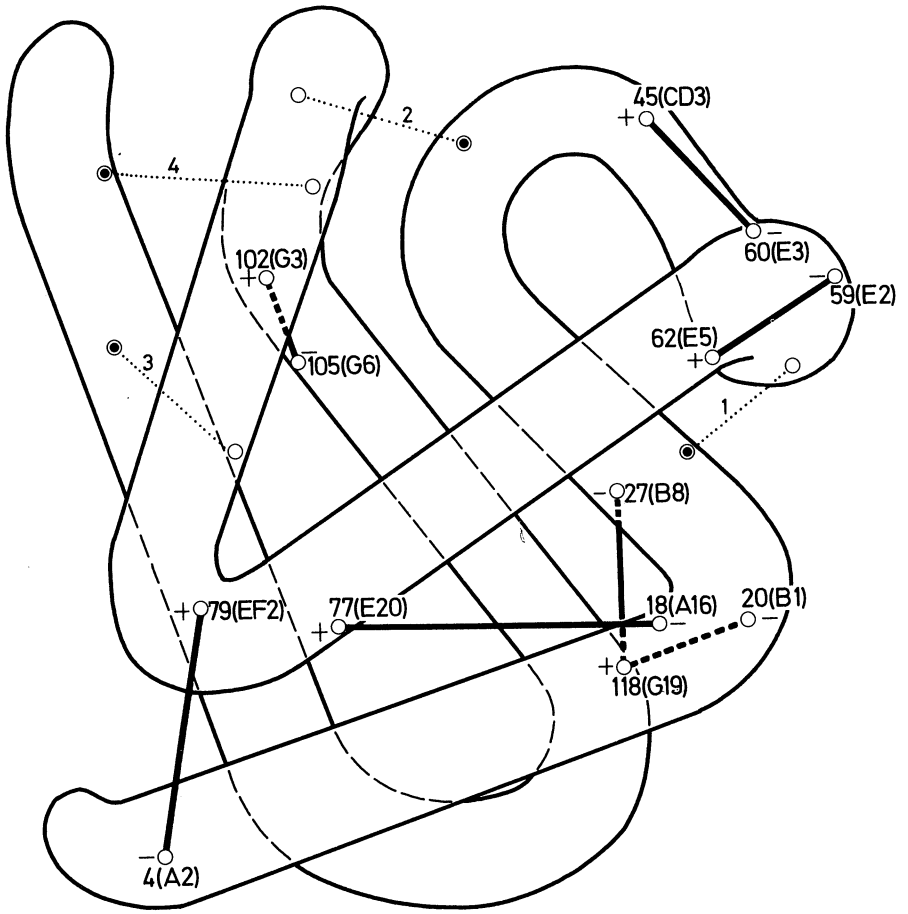


FIGURE 21. An outline drawing of the sperm whale myoglobin molecule showing the location of four inter-segmental hydrogen bonds (numbered 1–4), five inter-segmental and two intra-segmental salt bridges.

2. The myoglobin polypeptide chain is bent in a peculiar fashion to produce a flattened globular molecule of dimensions 2.2 nm × 2.2 nm × 1.25 nm, whose function is to allow the haem group to combine reversibly with molecular oxygen. With the exception of the small region which contains the two propionic acid groups, the haem lies buried in an internal cavity formed by hydrophobic residues ('the haem pocket'). The outside part of the molecule is in contact with the aqueous environment and is mainly composed of hydrophilic residues; 120 residues are external or lining the surface of crevices and 33 residues are internal (Perutz *et al.* 1965).

3. The haem group remains firmly wrapped in its pocket because of numerous atomic contacts with 22 of the surrounding residues. Of these, twelve are internal and interact with the buried part of the prosthetic group by means of numerous van der Waals forces (hydrophobic bonds) (Perutz 1969). The haem iron is covalently linked to a 'proximal' histidine at F8 and it lies opposite to a 'distal' histidine at E7 with the exception of *Aplysia* myoglobin which has only the proximal histidine (Tentori *et al.* 1973). The remaining 8 haem contacts are considered to be external residues which, through their polar and non-polar groups, help to maintain the position of the haem.

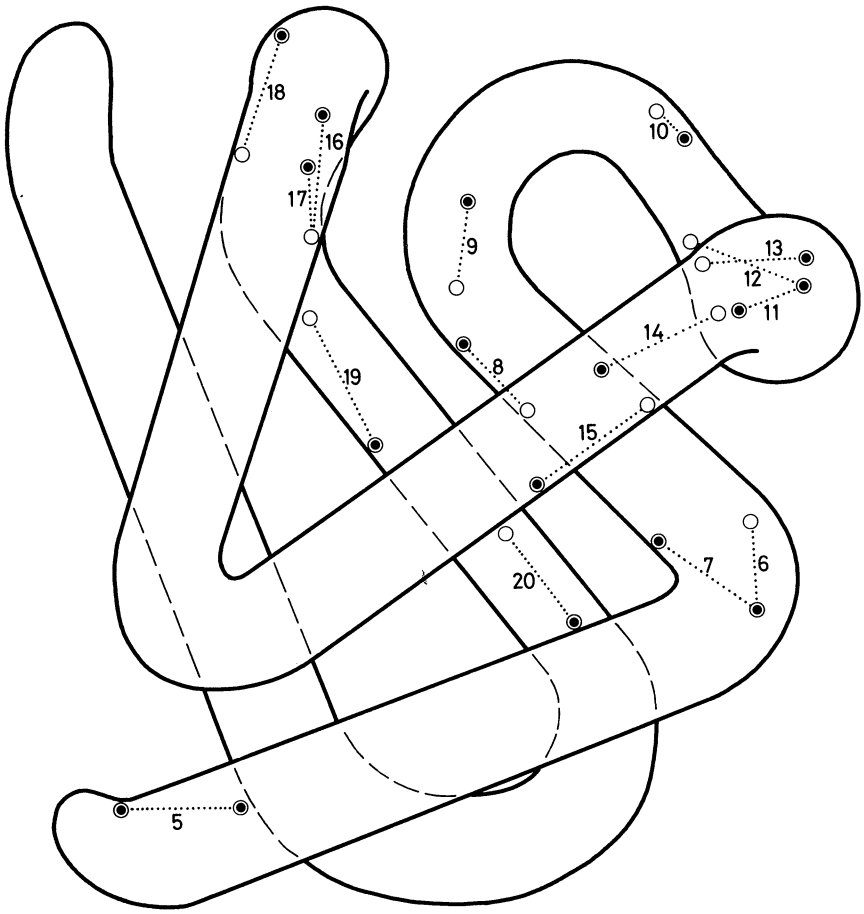


FIGURE 22. An outline drawing of the sperm whale myoglobin molecule showing the location of sixteen intra-segmental hydrogen bonds (numbered 5–20).

4. Four specific interactions maintain the three-dimensional conformation and stability of the myoglobin molecule:

(*a*) The numerous van der Waals forces between the interlocking non-polar side chains of residues within the same segment or belonging to different segments.

(*b*) The multiple main chain imino-carbonyl hydrogen bonds which form and stabilize the α helical segments.

(*c*) Four inter-segmental and 16 intra-segmental hydrogen bonds which have been deduced by X-ray crystallographic studies of sperm whale myoglobin (Watson 1969).

(*d*) The presence of five inter-segmental and two intra-segmental salt bridges (Watson 1969), which possibly play a role in the folding of the molecule, and also help to stabilize it thereafter. It is of special interest that with one exception these salt bridges are missing in the haemoglobin subunits.

The residues involved in the side-chain hydrogen bonds and in the salt bridges are shown in table 7, which also includes alternative substitutions found in some of the 31 myoglobins so far investigated. The locations of these interacting residues are schematically shown in figures 21 and 22.

The two conservative regions

Throughout some 300 Ma of evolution (since the ancestors of the mammals and birds diverged), two regions of the myoglobin molecule have remained remarkably conservative.

There is a flat plane formed by the C helix and the CD inter-helical bend in the vicinity of the haem group (see figure 23). This segment of the molecule contains, on the surface, residues 38 (C3) Glu, 41 (C6) Glu, 44 (CD2) Asp, 47 (CD5) Lys and 50 (CD8) Lys in all the myoglobins so far investigated, with the exception of residue 41 Asp found in the canids and birds, and residue 47 Arg present in the penguin (which are very conservative substitutions). The invariant residue 98 (FG4) Lys, also lies in this plane. It seems unlikely that the conservatism of these residues is due merely to chance and we must therefore assume that preservation of the charged residues in this region is important for the function of the molecule. Wittenberg (1966, 1970) and Scholander (1965) have discussed the possible rôle of myoglobin in facilitating the diffusion of oxygen within the cell. It seems possible that the invariant charged residues on this flat plane are involved in an interaction between myoglobin and another molecule, functioning as a 'docking site' facilitating the release of oxygen in the vicinity of the mitochondria. Another possibility is that this 'docking site', being close to the iron atom, facilitates the interaction with metmyoglobin reductase. In the haemoglobin molecule, C3, C6 and FG4 are $\alpha 1\beta 2$ tetrameric contacts (Perutz, Muirhead, Cox & Goaman 1968) and the preservation of a difference in this part of the myoglobin surface may guarantee its monomeric character.

The other conservative region of the myoglobin molecule, which is completely invariant in the 31 vertebrate species so far investigated, is that formed by residues 133 to 139: Lys–Ala–Leu–Glu–Leu–Phe–Arg (see table 2 and figure 23). It is very difficult to explain why this part of the molecule has not changed during some 300 Ma of evolution. Assuming that this situation has not arisen by chance, several hypothetical explanations deserve discussion.

The seven invariant residues form almost two turns of the α helix and three of them are internal: 134 (H11) Ala, 135 (H12) Leu and 138 (H15) Phe. The Phe 138 is in contact with the haem and the Leu 135 is probably a haem contact. Hence there is a possibility that the conservation of these two particular residues is of importance for steric and functional reasons, and so the fixation of changes may have been constrained by natural selection, but the other invariant residues 133 (H10) Lys, 136 (H13) Glu, 137 (H14) Leu and 139 (H16) Arg are external.

Lysine 133 deserves a special comment because it has been substituted by Asn in one of the four human atypical myoglobins so far found (Boyer *et al.* 1963; Boulton *et al.* 1971*a*, *b*, *c*). All four atypical myoglobins were found in the heterozygous state and in patients who had not suffered from muscular disease. The myoglobin mutant 133 (H10) Lys⟶Asn represented only 31% of the total myoglobin (whereas in the other three the mutant form amounted to approximately half). This lysine does not have any obvious function other than interacting with the surrounding water (Kendrew 1962; Perutz *et al.* 1965; Watson 1969); it is furthermore easily alkylated in the native protein (Harris & Hill 1969). In this context it is noteworthy that the human haemoglobin K Woolwich $\alpha_2\beta_2$ 132 (H10) Lys⟶Glu, which does not appear to be unstable (O'Gorman, Lehmann, Alsopp & Sukumaran 1963; Allan, Beale, Irvine & Lehmann 1965), is present at the level of 30% in the peripheral blood without causing abnormality of the red cells (Ringelhann *et al.* 1971). Lys is present at position H10 in almost all myoglobin and haemoglobin chains so far investigated (Dayhoff 1972), exceptions being the β-chain of frog haemoglobin (Chauvet & Acher 1970) and the monomeric component of *Glycera*

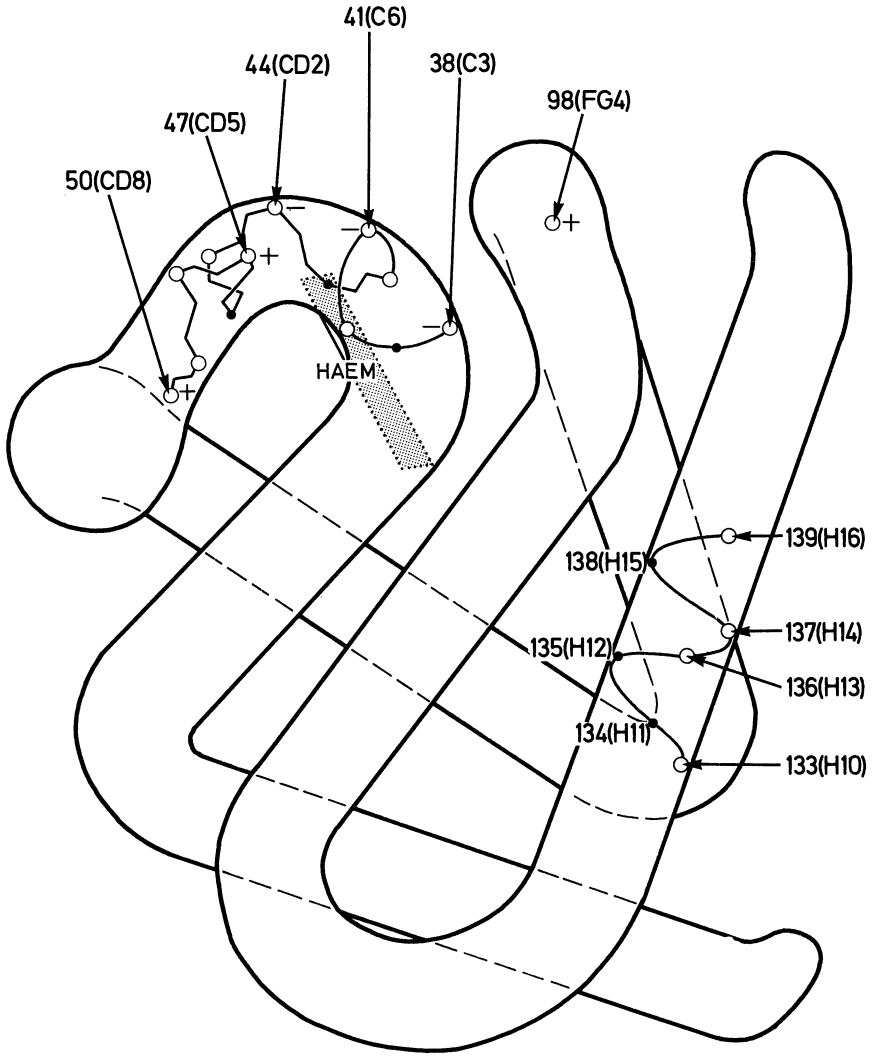


FIGURE 23. An outline drawing of the sperm whale myoglobin molecule showing the location of two conservative regions discussed in the text.

dibranchiata (Imamura *et al.* 1972) where in both instances this residue is Ala. Investigation of the biosynthesis of haemoglobin K Woolwich *in vitro* (Lang, Lehmann & King-Lewis 1974) indicates that its translation time is similar to that found in the normal β-chains. It seems, therefore, that the lower proportion of the K Woolwich β-chain (and presumably of the myoglobin variants which involve residue H10) is perhaps due either to a slower rate of initiation of translation, or to an unstable mRNA, or to some other as yet unknown defect of translation.

It is possible that H10 Lys is concerned with the folding mechanism of the molecule, and when this lysine is substituted by another amino acid the tertiary structure of the molecule cannot be properly completed in time, and the molecule is broken down. This hypothesis is based on the fact that H10 is sterically close to the N-terminus of the molecule. The ϵ-amino group of the Lys, although free as shown by electron density maps, may exert electrostatic attraction on residue A4 Glu during the early stages of folding. In the case of myoglobin this electrostatic attraction may cooperate with that between EF2 Lys and A2 Glu, or Asp, to move

the beginning of the A helix towards the EF corner and the H helix. Once the numerous van der Waals bonds are formed H10 Lys is no longer able to react with A4 Glu because both residues are sterically restricted, and the latter becomes hydrogen bonded to the α-imino group of residue A1.

The segment of seven invariant residues 133–139 invites discussion concerning the possible function of the 21 bases of the mRNA sequence concerned. Owing to the absence of any mutations which eliminate from consideration some of the isosemantic codons, it is not possible to establish the nature of 18 of the 21 bases involved, the exception being the codon for the arginine 139 of human myoglobin. There are two human atypical myoglobins with substitutions in this position: 139 Arg ⟶ Gln (Boulton *et al.* 1971*c*) and 139 Arg ⟶ Trp (Boulton *et al.* 1971*a*). The only one of six codons for arginine which can be converted by a single-point substitution into CAG for glutamine as well as into UGG for tryptophan is the codon CGG.

TABLE 8. ALIGNMENT SHOWING COMPLEMENTARY BASE PAIRS (.)

133	134	135	136	137	138	139
Lys	Ala	Leu	Glu	Leu	Phe	Arg
AAG	GCA	CUC	GAG	CUC	UUU	CGG
.	. .	. .	. . .	. . .	. . .	.
AAC	CUU	GGG	CUC	GAG	AAA	UAC
Gln	Phe	Gly	Leu	Glu	Lys	Tyr
152	151	150	149	148	147	146

We have investigated the feasibility of matching a postulated nucleotide sequence coding for residues 133–139 with another region of the structural mRNA of myoglobin. It is possible to generate a hairpin loop, the complementary sites being aligned as shown in table 8. The significance of such a matching exercise should not be overrated because the genetic code is degenerate and thus affords a wide choice of suitable codons. The 15 complementary base pairs could correspond to those responsible for residues 146–152 in 22 of the 31 myoglobins concerned. By taking into account the nine species which differ, and by assuming single-point substitutions, some of the isosemantic codons could be eliminated from consideration. For example, the sea lion has Arg at position 147, which is occupied by Lys in all other species studied, and so four of the six codons for Arg were eliminated, leaving as likely candidates AGA or AGG. It is of interest that the sheep with 148 Val, the ox with 148 Val and 152 His, the kangaroo and the two birds with 149 Phe, the sperm whale with 151 Phe and 152 His, and the porpoise, dolphin and harbour seal with 152 His, would still be able to pair between 14 and 16 bases out of a total of 21. This hypothetical arrangement leaves six positions (140–145) at which the bases are difficult to pair due to the variability of the residues at these sites in the 31 myoglobins investigated. Whether or not this postulated hairpin loop is part of the myoglobin mRNA must await further investigation, but assuming that it does exist, this secondary structure could perhaps either provide stability to the mRNA or function as a specific recognition site.

The salt bridges

The 13 residues involved in the seven salt bridges appear to have been highly conservative. They probably play an important part in the folding of the myoglobin molecule and in the subsequent stability of its secondary and tertiary structure.

The five inter-segmental salt bridges are so situated that they help the numerous hydrogen bonds and hydrophobic interactions to maintain close contact between parts of the molecule which are distant in terms of the primary structure (see table 7 and figure 21). At position 4 (A2), Asp is present in all known vertebrate myoglobins with the exception of those of the cetaceans, in which this position is occupied by Glu. Either of these negatively charged residues can form a salt bridge with the invariant residue 79 (EF2) Lys, holding together the A helix and the EF corner. The invariant residues 18 (A16) Glu and 77 (E20) Lys form a salt bridge which helps to maintain the proximity of the A and E helices. Residue 20 (B1) Asp appears to form a salt bridge with residue 118 (G19) Arg in the cetaceans; in all other known vertebrate myoglobins the latter position is occupied by Lys, which presumably performs the same function. Residue 118 (G19) Lys or Arg also establishes a salt bridge with residue 27 (B8) Asp, in the cetaceans, galago and kangaroo, or 27 Glu in all other known vertebrate myoglobins except the penguin. Both salt bridges link the relatively distant AB corner and the B helix to the G helix and the GH bend. Residue 45 (CD3) Arg in the sperm whale, or Lys in other vertebrates, forms a salt bridge with residue 60 (E3) Asp, (or Glu in the ceboids) keeping the initial part of the E helix near to the CD segment.

The two intra-segmental salt bridges, which contribute to the stability of the secondary structure of the molecule, are between residue 59 (E2) Glu and residue 62 (E5) Lys (or Arg in the harbour seal), and between residue 102 (G3) Lys and residue 105 (G6) Glu. The latter salt bridge is presumably missing in the marsupials in which residue 102 is Gln.

The hydrogen bonds

The 23 residues whose side chains are hydrogen bonded in sperm whale myoglobin were compared with the corresponding residues in the other vertebrate myoglobins (table 7, figures 21 and 22). Four of these are inter-segmental hydrogen bonds. Residue 26 (B7) is Gln in all known vertebrate myoglobins, with the exceptions of the kangaroo (Lys) and the birds (His). This residue forms a hydrogen bond with the α-carbonyl group of residue 56 (D6) and helps to maintain the closeness of the B and D helices. Lys and His could also form, if sterically permitted, a similar hydrogen bond.

The conservative residue 42 (C7) Lys (Arg in the two birds), forms a hydrogen bond with the α-carbonyl group of residue 98 (FG4). Residue C7 has a triple function; first, it establishes a bridge between the C helix and the relatively distant FG inter-helical segment; secondly, the fully extended long aliphatic chain of Lys (or Arg) closes the uppermost part of a gap which otherwise would permit access of water to the buried part of the prosthetic group; thirdly, the internal part of the hydrophobic side chain of C7 is a haem contact.

The side chain of the invariant residue 141 (H18) Asp is hydrogen bonded to the imidazole group of 82 (EF5) His (Gln in the badger), and helps to maintain proximity between the H helix and the EF inter-helical segment.

The fourth inter-segmental hydrogen bond is between the internal side chain of the invariant residue 146 (H23) Tyr, and the α-carbonyl group of FG5 (Kendrew 1962). In the haemoglobins this latter residue participates in the transition from the deoxy- to the oxy-conformation of the tetramer (Perutz 1969). In the monomeric myoglobin this hydrogen bond is shorter in deoxymyoglobin than in acid metmyoglobin, and so it is stronger in the deoxy-conformation (T. Takano 1975, personal communication).

Turning now to the sixteen intra-segmental hydrogen bonds shown in table 7, it may be

noted that, for example, 35 Ala (lorisids), 67 Val (perissodactyls) and 117 Ala (artiodactyls) are unable to form hydrogen bonds, but these exceptions affect only one position in any particular myoglobin. Some of the residues involved in this kind of bonding in the sperm whale myoglobin differ in other species but, if not sterically restricted, would still be capable of forming hydrogen bonds.

It can be assumed that the adaptive value of the majority of the hydrogen bonds is relatively small when each is considered by itself, but as a group they are undoubtedly important in conferring conformational stability and rigidity to the molecule.

Other important residues

When residue 79 (EF2) Lys forms a salt bridge with residue 4 (A2) Asp, the aliphatic side chain of the former establishes several hydrophobic contacts with the indole ring of residue 7 (A5) Trp, which is invariant in all known vertebrate myoglobins. These bonds help to hold the EF corner to the A helix.

The residue at position 14 (A12) is Trp in all known myoglobins and its bulky side chain is extended towards the internal cavity to form van der Waals contacts with the residues at positions 10 (A8) Val; 13 (A11) Val, Ile, Met or Ala; 17 (A15) Val; 72 (E15) Leu; 76 (E19) Leu; 115 (G16) Ile or Leu; and 131 (H8) Met.

Stability to the AB corner is provided by the hydrogen bond established between the invariant residue 20 (B1) Asp and the main chain imino group of residue 23 (B4). In addition the α-carbonyl group of residue 20 (B1) forms a forked hydrogen bond with the imino group of residue 24 (B5) His and with the imidazole ring of the same invariant residue. The presence of Pro at position 22 (B3) in the anthropoids (with the exception of the gorilla), the opossum and the hedgehog also helps to form the corner because, being an imino acid, it cannot form a hydrogen bond with the α-carbonyl group of residue AB1. The other amino acid residues which are found at this position, Ser, Ala, Gly and Val, are also unable to make this contact because of steric hindrance by the side chain of residue B2.

Between the end of the B helix and the beginning of the E helix, there is a region of 25 residues within which the polypeptide chain changes its direction four times. The core of the resulting structure is formed by the interlocking of eight hydrophobic residues which, with the exception of the highly conservative changes 40 Met and 61 Met in the penguin, have remained invariant in all known vertebrate myoglobins: 29 (B10) Leu, 33 (B14) Phe, 40 (C5) Leu, 43 (CD1) Phe, 46 (CD4) Phe, 49 (CD7) Leu, 55 (D5) Met and 61 (E4) Leu. The stability of the BC corner is provided partly by the invariant 37 (C2) Pro which interrupts the succession of hydrogen bonds along the helix, and partly by the hydrogen bond between the side chain of the invariant 39 (C4) Thr and the α-carbonyl group of residue C1. The corner between the CD inter-helical segment and the D helix is reinforced by the hydrogen bond formed by the α-imino group of residue CD8 and the side chain of residue 54 (D4) which is either Glu or Asp.

In nearly all known vertebrate myoglobins residue 45 (CD4) is Lys; the only exception is Arg, found in the sperm whale. This residue interacts with one of the haem propionic acids and also forms a salt bridge with the charged residue 60 (E3) Asp (or Glu in the ceboids). It seems likely that these interactions confer stability to the molecule. In *Aplysia* myoglobin the condition is reversed: residue CD4 is Asp and residue E3 is Lys. It is not known whether vertebrate and molluscan myoglobins originated independently (and represent an example of convergent evolution) or whether they are derived from a common ancestral myoglobin molecule.

If we assume the latter, it is not known whether the vertebrate, or the molluscan or both myoglobins represent derived states, but if one assumes, for example, that the vertebrate condition represents the persistence of an ancestral state then at least two possibilities for the course of events deserve consideration: (1) residue 60 Asp was first substituted by Lys, leaving the Lys at position 45 interacting with the haem propionic group; when 45 Lys was subsequently substituted by Asp the salt bridge between these two positions was re-established; (2) residue 45 Lys was substituted by Asp leaving the propionic acid unreacted. Subsequently, a compensating second mutation resulted in the substitution of Lys for Asp at position 60. This also would restore the possibility of a salt bridge between these two positions, which would confer stability to the molecule.

With the exception of the Val found in horse and zebra, the known vertebrate myoglobins have Thr at position E10, the side chain of which is hydrogen bonded to the α-carbonyl group of residue 63 (E6). Furthermore, one of the propionic acids (that at position 7 of the haem) is hydrogen bonded to the imidazole ring of residue 97 (FG3) His, which is present in all known vertebrate myoglobins. Because this latter residue is Phe in *Aplysia* the propionic acid at position 7 of the haem is available to interact with the Arg present at E10 in *Aplysia* myoglobin. In this respect it may resemble the haemoglobin β-chain where this propionic acid interacts with E10 Lys.

In the myoglobins residue 104, G5, Leu is invariant, and in the haemoglobins (except *Chironomus*) G5 is Phe (Dayhoff 1972), which helps to maintain the haem plate in the more upright position in deoxyhaemoglobin (Morimoto, Lehmann & Perutz 1971). The abnormal human haemoglobin 'Heathrow' also has Leu at position βG5 (White *et al.* 1973). This variant has an abnormally high oxygen affinity, and it is probable that G5 Leu has a part in conferring that characteristic on myoglobin. Myoglobin, however, being a monomer, would in any case be expected to have a high oxygen affinity.

Residue 110 (G11), which is Ala in most myoglobins (but Ser in cercopithecoids and Cys in hominoids) is at the bottom of the haem pocket, and examination of the sperm whale myoglobin model shows that Ala can be replaced by Ser or Cys without disturbance. Either of these could strengthen the G helix by forming a hydrogen bond with the main chain at G7. C11 is also near the B helix and an increase of the bulk of this side-chain caused by substituting Cys for Ser would strengthen the hydrophobic contact between the G and B helices, and thereby contribute to the stability of the molecule. It may well be that the G11 Cys in the α-chain of mammalian haemoglobin has a similar function.

Three sites, 88 (F3), 100 (G1) and 120 (GH2), are occupied by Pro, in most of the vertebrate myoglobins so far studied. Residues F3 and G1 are at the beginning of the F and G helices; the latter residue also forms part of the corner between the FG inter-helical segment and the G helix. Proline at position GH2 helps to stabilize the sharp GH corner where the main chain bends through 180°. This allows the G and H helices to run parallel in opposite directions like the sides of a ladder whose four rungs are represented by the interlocking hydrophobic bonds of the residues listed in table 9. This whole structure forms part of the bottom of the haem pocket and represents two parallel rods, crossed by the A, E and F helices to form a rigid structure.

Whether the myoglobins of *Aplysia* and the vertebrates were derived from a common ancestor, or whether they arose independently, it is remarkable that the corresponding amino acids involved in holding the haem are either the same or alternatives of a very similar nature. The

TABLE 9. RESIDUES WITH INTERLOCKING HYDROPHOBIC BONDS

123	(GH5)	Phe	invariant;	115 (G16)	Ile or Leu
131	(H8)	Met	invariant;	111 (G12)	Ile invariant
135	(H12)	Leu	invariant;	108 (G9)	Ser invariant
138	(H15)	Phe	invariant;	104 (G5)	Leu invariant

residues involved are CD1 Phe, E15 Leu and F8 His (the same in all known globins); E11 Val (Ile in *Aplysia*); F4 Leu (Phe in *Aplysia*); FG5 Ile or Val (Val in *Aplysia*); G5 Leu (Phe in *Aplysia*); G8 Ile (Val in *Aplysia*); H12 Leu (Phe in *Aplysia*) and H15 Phe (Ile in *Aplysia*). Three substitutions at haem contact sites deserve special comment. (1) Residue E7, usually the distal His, is likely to be Val in *Aplysia*. There are only three haemoglobin chains where E7 is not His: it is Ile in one of the monomeric haemoglobins of *Chironomus thummi thummi* III (Glossman, Horst, Plagens & Braunitzer 1970), Leu in *Glycera dibranchiata* (Imamura *et al.* 1972) and Gln in the α-chain of the opossum (Stenzel 1974). With the exception of substitutions found in the abnormal human haemoglobin chains these are the only replacements of this important residue in all known globins. (2) Residue G4, which is either Tyr or Phe in all known vertebrate myoglobins, is Gln in *Aplysia*. In that respect the *Aplysia* myoglobin resembles the haemoglobins, where this position is usually occupied by the similar residue Asn (Dayhoff 1972). (3) The fact that FG3 is occupied by Phe in *Aplysia* rather than by the His of vertebrate myoglobins has already been mentioned above with reference to the propionic acid at position 7 of the haem. The phenolic ring could well be a conservative substitution for the imidazole ring as a haem contact.

It appears that, during the period of evolution since the haemoglobin and myoglobin molecules diverged, the fixation of mutations at sites important for the common function of the molecules (haem contacts and internal hydrophobic residues) has been constrained. The two molecules resemble one another in their tertiary structure, and yet the myoglobin molecule is very different from the subunits of haemoglobin because the residues on the surface of both molecules have changed progressively to preserve or acquire functions: in the case of the haemoglobin chains to facilitate dimeric and tetrameric combination to produce subunit interaction, the binding of organic phosphates, the Bohr effect (Goodman, Moore & Matsuda 1975) and the capacity of association with the plasma protein haptoglobin; and in the case of myoglobin to preserve its monomeric condition and probably to acquire interactions which have not yet been discovered.

We are grateful to Professor H. C. Watson for reading this section and giving his advice. It would have been difficult for us to attempt the interpretation of molecular structure without the benefit of the model of sperm whale myoglobin (scale 1.25 cm = 0.1 nm) built by Mr S. A. Barker to the specification of Professor Watson and Sir John Kendrew, F.R.S.

6. IMMUNOLOGICAL CONSIDERATIONS

Much information has been accumulated concerning the antigenicity of the globin proteins, and the known tertiary structure of haemoglobin and myoglobin has made both of these proteins highly suitable for the investigation of the structural basis of antigenic activity.

Atassi and his co-workers have carried out extensive studies on myoglobin and have reported that the localization of antigenic reactive regions for sperm whale myoglobin has reached

completion (Atassi 1973). We shall employ the terms 'antigenic reactive region' and 'antigenic reactive site' in the sense in which they are used by Atassi & Saplin (1968). An antigenic reactive region is a part of the primary structure of the molecule involved in a reaction with antibodies; an antigenic reactive site may be made up of two or more antigenic reactive regions brought into close association by the folding of the molecule. The regions involved in a site need not, therefore, be adjacent in the primary structure.

The large body of sequence data now available for the myoglobin of mammals makes it possible to examine the nature and extent of sequence variability in the antigenic reactive regions recognized by Atassi. These regions are listed in table 10, and shown in position on the molecule in figure 24. According to Atassi the sequence 104 to 112 is a doubtful part of reactive region 4, and so it has been omitted from our comparisons.

TABLE 10. ANTIGENIC REACTIVE REGIONS

	15	16	17	18	19	20	21	22	region number
sperm whale	Ala	Lys	Val	Glu	Ala	Asp	Val	Ala	1
rabbit	Gly	Lys	Val	Glu	Ala	Asp	Leu	Ala	
sheep	Gly	Lys	Val	Glu	Ala	Asp	Val	Ala	
	56	57	58	59	60	61	62		
	Lys	Ala	Ser	Glu	Asp	Leu	Lys		2
	Lys	Ala	Ser	Glu	Asp	Leu	Lys		
	Lys	Ala	Ser	Glu	Asp	Leu	Lys		
	94	95	96	97	98	99	100		
	Ala	Thr	Lys	His	Lys	Ile	Pro		3
	Ala	Thr	Lys	His	Lys	Ile	Pro		
	Ala	Asn	Lys	His	Lys	Ile	Pro		
	113	114	115	116	117	118	119	120	
	His	Val	Leu	His	Ser	Arg	His	Pro	4
	His	Val	Leu	His	Ser	Lys	His	Pro	
	His	Val	Leu	His	Ala	Lys	His	Pro	
	145	146	147	148	149	150	151		
	Lys	Tyr	Lys	Glu	Leu	Gly	Tyr		5
	Gln	Tyr	Lys	Glu	Leu	Gly	Phe		
	Gln	Tyr	Lys	Val	Leu	Gly	Phe		

Published studies have used antisera prepared in rabbits and goats, and it became evident that the complete myoglobin sequences for these animals would be of especial interest in an immunological context. During the preparation of this communication we have established the complete myoglobin sequence of the rabbit (table 11). The details of the analysis are given elsewhere (Romero-Herrera, Lehmann & Castillo 1976). The primary structure of goat myoglobin is unfortunately not yet available. We have, therefore, used that of sheep as a substitute. Some justification for this on immunological grounds is available from Atassi, Tarlowski & Paull (1970), where sheep and goat myoglobin are shown to be closely similar although, using absorption techniques, they found that sheep has some immunological similarities with the ox which goat does not share.

The sequences of the five antigenic reactive regions for sperm whale, rabbit and sheep myoglobins are given in table 10. In considering these regions three qualifications are necessary:

(i) the uncertainty concerning the involvement of some other parts of the molecule in antigenic activity;

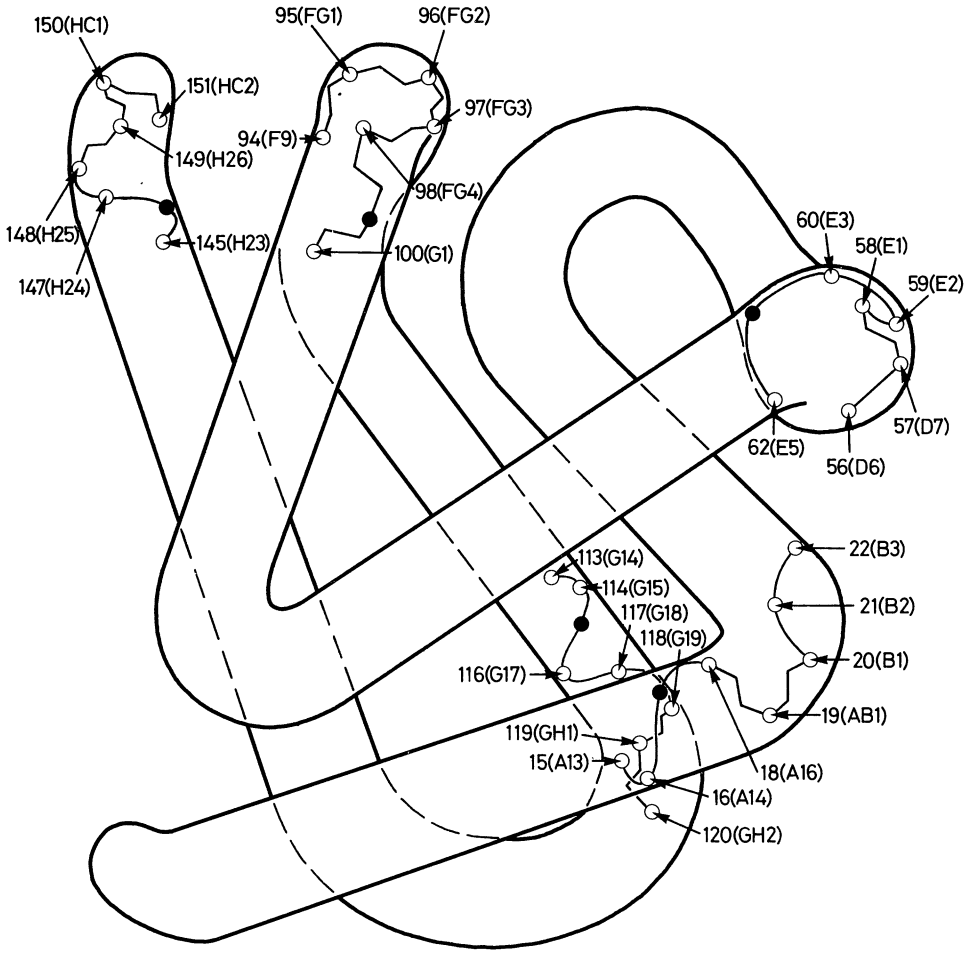


FIGURE 24. An outline drawing of the sperm whale myoglobin molecule showing the location of the five antigenic reactive regions of Atassi (1973). Internal residues are shown by solid symbols.

(ii) some reactive sites formed by the proximity in the tertiary structure of several discrete parts of the primary sequence may not be revealed by the methods of Atassi;

(iii) the substitution of goat by sheep in the sequence information.

The value of immunotaxonomy is usually presented as depending on a consistent relation between sequence difference and immunochemical cross reaction. Kirsch (1969), in modelling the effect of sequence change on serological comparison, made the simplifying assumption that proteins may be reduced, for this purpose, to the antigenic determinants. These sections of the molecule may then be treated in terms of persistence, loss and replacement on a unit basis. Immunochemical distinctions are likely to be sharply made between molecules which differ in residues within reactive regions common to them. Conversely, there may be a profound immunochemical effect of a substitution fixed in parallel in relatively distantly related animals. An example of this is given by Nisonoff, Reichlin & Margoliash (1970) for cytochrome c.

It might, however, be more realistic to envisage that the effects of sequence differences in all parts of the molecule are integrated through the reactive regions and hence are estimated by the immunochemical techniques. Considerable evidence points to the significant effect of conformational change, due to sequence differences, on the immunochemical cross reactivity of

TABLE 11. SEQUENCE OF RABBIT MYOGLOBIN

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Gly	Leu	Ser	Asp	Ala	Glu	Trp	Gln	Leu	Val	Leu	Asn	Val	Trp	Gly	Lys	Val	Glu	Ala	Asp	Leu	Ala	Gly	His	Gly

26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Gln	Glu	Val	Leu	Ile	Arg	Leu	Phe	His	Thr	His	Pro	Glu	Thr	Leu	Glu	Lys	Phe	Asp	Lys	Phe	Lys	His	Leu	Lys

51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
Ser	Glu	Asp	Glu	Met	Lys	Ala	Ser	Glu	Asp	Leu	Lys	Lys	His	Gly	Asn	Thr	Val	Leu	Thr	Ala	Leu	Gly	Ala	Ile

76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Leu	Lys	Lys	Lys	Gly	His	His	Glu	Ala	Glu	Ile	Lys	Pro	Leu	Ala	Gln	Ser	His	Ala	Thr	Lys	His	Lys	Ile	Pro

101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125
Val	Lys	Tyr	Leu	Glu	Phe	Ile	Ser	Glu	Ala	Ile	Ile	His	Val	Leu	His	Ser	Lys	His	Pro	Gly	Asp	Phe	Gly	Ala

126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153
Asp	Ala	Gln	Ala	Ala	Met	Ser	Lys	Ala	Leu	Glu	Leu	Phe	Arg	Asn	Asp	Ile	Ala	Ala	Gln	Tyr	Lys	Glu	Leu	Gly	Phe	Gln	Gly

proteins (Atassi 1973; Teicher, Maron & Arnon 1973). In this event, however, the complexity of the system defies analysis at present, especially because most immunotaxonomists work with large molecules, the structure of which is poorly known.

There has been considerable progress in predicting the secondary structure of proteins from their sequence (Chou & Fasman 1974*a*, *b*). If such predictions were completely successful in differentiating helical and non-helical regions of myoglobin, four of the non-helical regions would overlap with antigenic reactive regions. However, there are two remaining non-helical sections which are not involved in the regions proposed by Atassi, and the region 56–62 is not only helical but varies little, as will be demonstrated below.

The antigenic reactive regions delineated for sperm whale myoglobin occupy particularly accessible parts of the folded molecule (figure 24). It has not yet been demonstrated that other parts of the myoglobin molecule form either alternative or additional reactive regions in other species. However, the fact that three of the five antigenic reactive regions of sperm whale myoglobin are located on corners, a fourth is on the last part of the H helix and continues on to the C-terminal region, and a fifth is highly accessible at the molecule's surface argues strongly that accessibility to antibody molecules is in itself of paramount importance, and that at least some of these regions are likely to be reactive regions in the myoglobins of other species. X-ray crystallographic evidence regarding sperm whale myoglobin and the β-chain of horse haemoglobin does not suggest three-dimensional differences to be sufficiently extensive as to contradict this argument (Perutz 1965; Perutz, Kendrew & Watson 1965). Also, both goat and rabbit appear to make antibodies to the same antigenic reactive regions of sperm whale myoglobin; indeed, the variability in length of the region involved is small ($+0$ or 1 residue on either side of a central section composed of the six residues 16–21 for region 1 according to Koketsu & Atassi (1974)). However, the proteins of even relatively closely related animals do not necessarily cross react with antisera to one another in a fully reciprocal fashion as far as antigenic regions are concerned. Some evidence bearing on this issue is discussed for sperm whale and finback whale myoglobins by Atassi & Saplin (1971), where it is demonstrated that a different combination of reactive regions is involved.

As a first approach to the sequence data all the mammals and the chicken are ranked in order of the amino acid differences between their myoglobin and that of the sperm whale (table 12). Dolphin and porpoise differ least from sperm whale, and chicken is the most distant. Table 13 shows the number of amino acid differences between the same animals and the sperm whale in the five antigenic reactive regions alone. Again, dolphin and porpoise (plus sportive lemur) are closest to the sperm whale, and chicken (plus hunting dog) are the most remote.

TABLE 12. AMINO ACID DIFFERENCES OF VARIOUS SPECIES FROM SPERM WHALE

sperm whale	0	marmoset	26
dolphin	14	galago	26
porpoise	15	harbour seal	26
horse/zebra	19	badger	27
potto	20	hedgehog	27
tree shrew	22	woolly monkey	28
rabbit	22	sheep	28
chimpanzee	23	dog	29
slow loris	23	opossum	30
man	24	hunting dog	30
gorilla	24	squirrel monkey	30
macaque	24	ox	31
sportive lemur	24	kangaroo	32
gibbon	25	chicken	41
baboon	25		

It is a familiar assumption that an animal will not usually make antibodies to parts of a foreign molecule which are identical to the same parts of the homologous molecule in its own body. Thus, although the real situation regarding non-stimulation of antibody formation is often far more complicated, one might expect that when antibodies to sperm whale myoglobin are made in a rabbit, particular significance would attach to those residues within the antigenic reactive regions of sperm whale myoglobin which differ from the homologous positions in rabbit myoglobin. There are five such differences between rabbit and sperm whale: residues 15 (Gly–Ala) and 21 (Leu–Val) in region 1, residue 118 (Lys–Arg) in region 4, residues 145 (Gln–Lys) and 151 (Phe–Tyr) in region 5. We have also sought the presence of these same differences from rabbit in the other animals and performed the same exercise with sheep (substituting for goat) instead of rabbit. This is adopting the point of view that it is the extent to which a heterologous antigen shares these crucial differences with the homologous antigen which would govern the extent of cross reaction, rather than the degree of difference at other positions within the antigenic reactive regions. However, no consistant pattern emerges although it is of interest that where significant cross reactions are reported by Atassi *et al.* (1970) some four to seven of these specific residues are shared by the animals concerned. Conversely, there may be no cross reaction even when four such residues are shared.

There is no evidence for a particular pattern of residues within the antigenic reactive region: these regions are not, as a group, characterized either by marked variability or invariability. However, the region 56–62 (reactive region 2) is of some special interest from the comparative point of view: it is the same in rabbit, sperm whale and sheep. Indeed, the region varies little (except in the harbour seal) within the known mammalian myoglobins, and furthermore, residues 55 and 63 are invariant (table 14). According to Koketsu & Atassi (1974*a*) both rabbit and goat make antisera to this antigenic reactive region in sperm whale myoglobin; a number of speculations may explain why the sperm whale protein provokes antibodies in animals

TABLE 13. AMINO ACID DIFFERENCES OF VARIOUS SPECIES FROM SPERM WHALE IN ANTIGENIC REACTIVE REGIONS

sperm whale	0
dolphin	2
porpoise, sportive lemur	3
horse/zebra	4
rabbit, potto, slow loris, tree shrew	5
galago, chimpanzee, baboon, macaque	6
man, gorilla, gibbon, squirrel monkey, marmoset, opossum, ox	7
harbour seal, badger, woolly monkey, sheep	8
dog, hedgehog, kangaroo	9
hunting dog, chicken	10

TABLE 14. VARIATION IN ANTIGENIC REACTIVE REGION 2

position	variation
55 Met	invariant
56 Lys	Arg in harbour seal
57 Ala	Arg in harbour seal, sea lion; Gly in dog, hunting dog, badger; Ser in hedgehog
58 Ser	invariant
59 Glu	invariant
60 Asp	Glu in S. American monkeys
61 Leu	invariant
62 Lys	Arg in harbour seal
63 Lys	invariant

apparently identical in this region. Once again, sheep may be inadequate as a substitute for goat in this context, and in the case of the rabbit it is possible, but unlikely, that the rabbits used in our sequence studies and the rabbits used by Atassi to produce antisera differed in this region. On the other hand, it is possible that in the region 56–62 there is a determinant which, for the set of animals considered, sustains alterations in its three-dimensional fine structure because of substitution in other parts of the molecule.

It has been suggested that determinants may be of two basic types: 'specific determinants' with which antibodies eventually combine, and 'carrier determinants' which are of importance

during induction of the precursors of antibody-forming cells (Bretscher & Cohn 1968; Erickson & Rosenberg 1973). Erickson & Rosenberg suggest that the structural requirements for the two types may be different, and that non-stimulation of antibody formation is a phenomenon related to the carrier determinants. If this is so, then it complicates further the naive approach forced upon us in trying to correlate sequence difference (or similarity) and immunological cross reactivity. A similarity of carrier determinants between two species could affect the nature and extent of the immune interaction between them. If carrier and specific determinants are not identical, then analysis of the specific determinants alone cannot provide more than an approximation to the true situation. Thus even with the extensive work of Atassi and his co-workers, with the availability of the three-dimensional structure of sperm whale myoglobin and with the complete sequences of the mammalian myoglobins brought together here, firm conclusions are difficult to draw.

We have emphasized above the three-way nature of an immunological comparison involving a heterologous cross reacting antigen and an antibody to the homologous molecule produced in a third species. Nevertheless, there is some evidence to suggest that (at least over a restricted taxonomic range) there is a useful, although approximate, relationship between total sequence difference and immunological reaction for myoglobin. Figure 25 shows the relations between percentage cross reaction measured by quantitative precipitin testing (the homologous reaction being taken as 100%) and total amino acid differences between the pairs of species concerned. The identity of the species is given in the legend, and the points obtained using rabbit and goat antisera are identified on the figure. The assumption has been made following Atassi *et al.* (1970) that values of less than 7% obtained from quantitative precipitin tests do not represent significant cross reactions. Hence point 18 on the baseline represents a minimum estimate for the number of amino acid differences needed to prevent cross reaction in immunological tests using myoglobin. This corresponds to about 12% of the whole molecule. Figure 26 shows the relation between percentage cross reaction and amino acid differences in the antigenic reactive regions alone. The similarity to figure 25 reflects the fact that differences in these reactive regions occur approximately in proportion to those occurring in the molecule as a whole (although this may not remain so when further information becomes available).

With the relatively inconclusive data available from myoglobin, reference should be made to the account of Read (1975) for a discussion of the relative merits and implications of different descriptive equations fitted to the relation between cross reaction and amino acid distance.

It may be asked whether simple amino acid distance is the most informative measure of protein difference which can be investigated against cross reaction values. At first sight it seems unsuitable to attempt to use the sequence information for estimates expressed in terms of nucleotide differences, since antibody activity is clearly directed towards the structure of the completed protein. A more promising approach would perhaps be to score the nature of the differences between sequences, using an index based on the structural and chemical properties of the amino acids involved (Epstein 1967; Sneath 1966; Grantham 1974). However, it has recently been demonstrated (Beyer, Stein, Smith & Ulam 1974) for cytochrome c that, using the index of Sneath (1966), the chemical distance is linear with the genetic coding distance when the procedure of Ohta & Kimura (1971) is used together with the metric proposed by Beyer *et al.* Whether or not this is generally the case remains to be established. Transformation of the myoglobin data using the three indices mentioned did not result in any clarification.

The very approximate nature of simple amino acid difference calculations is shown by the

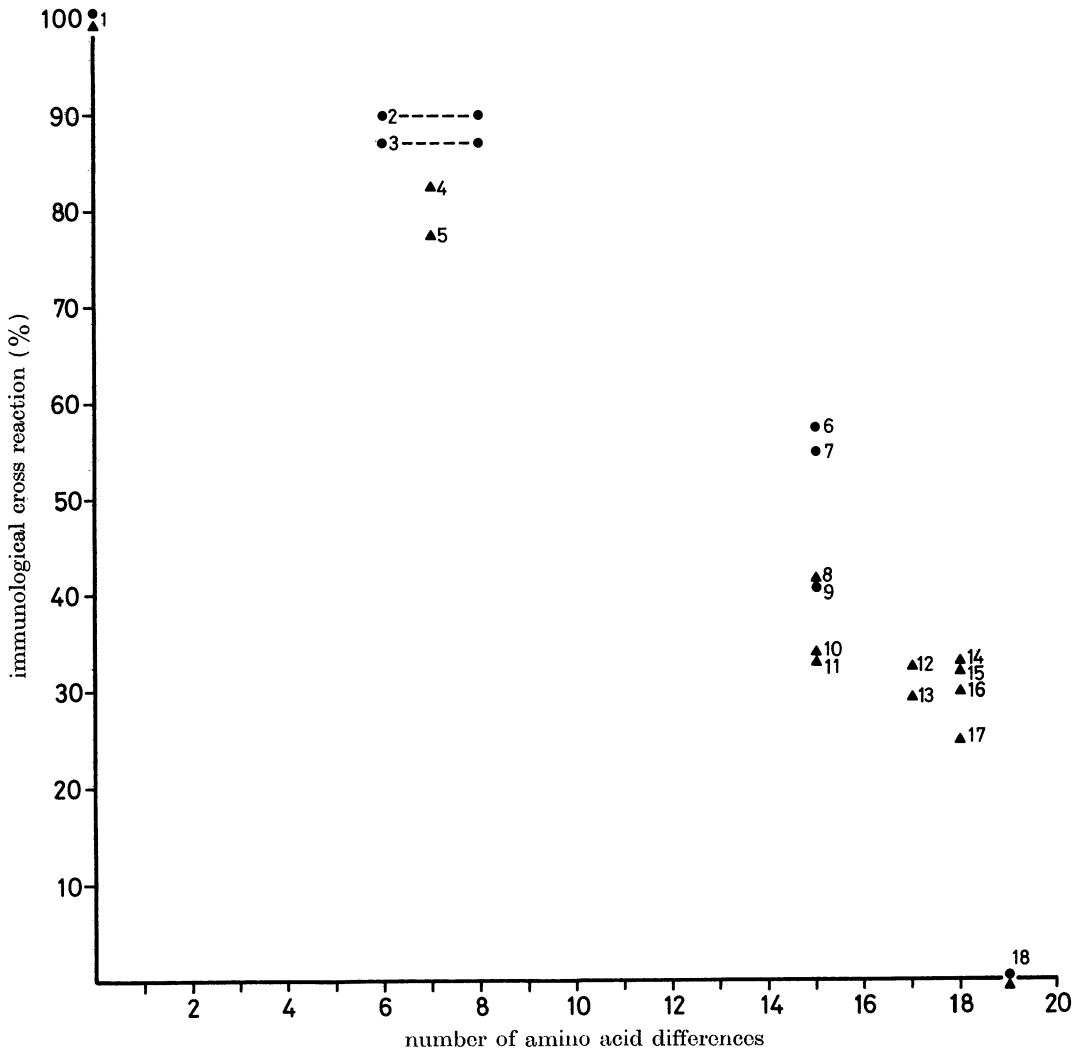


FIGURE 25. Immunological cross reaction, measured by quantitative precipitin testing, plotted against number of amino acid differences between the myoglobin sequences of pairs of species. Homologous reactions are taken as 100% in all cases. Circles represent comparisons made using rabbit antisera, and triangles represent those made using goat antisera. The number on each point indicates the identity of the species pair as follows; cross-reaction data is derived from Atassi *et al.* (1970) except where otherwise indicated. 1. Only two symbols are shown but all homologous reactions were scored as 100%. 2, 3. Anti-ox versus sheep, two different antisera. 4, 5. Anti-man versus macaque, two different antisera. 6, 7. Anti-finback whale versus sperm whale, two different antisera (Atassi & Saplin 1971; the number of amino acid differences has been estimated from table 1 of the same paper). 8, 9, 10. Anti-sperm whale versus finback whale, three different antisera (data as for points 6, 7). 11. Anti-sperm whale versus porpoise (taken from Prager & Wilson 1971). 12, 13. Anti-horse versus macaque, two different antisera. 14, 16. Anti-horse versus man, two different antisera. 15, 17. Anti-man versus horse, two different antisera. 18. Anti-sperm whale versus horse; two different antisera. The symbols coincide. This species pair represents the lowest number of amino acid differences between two animals for which no significant cross reaction was recorded.

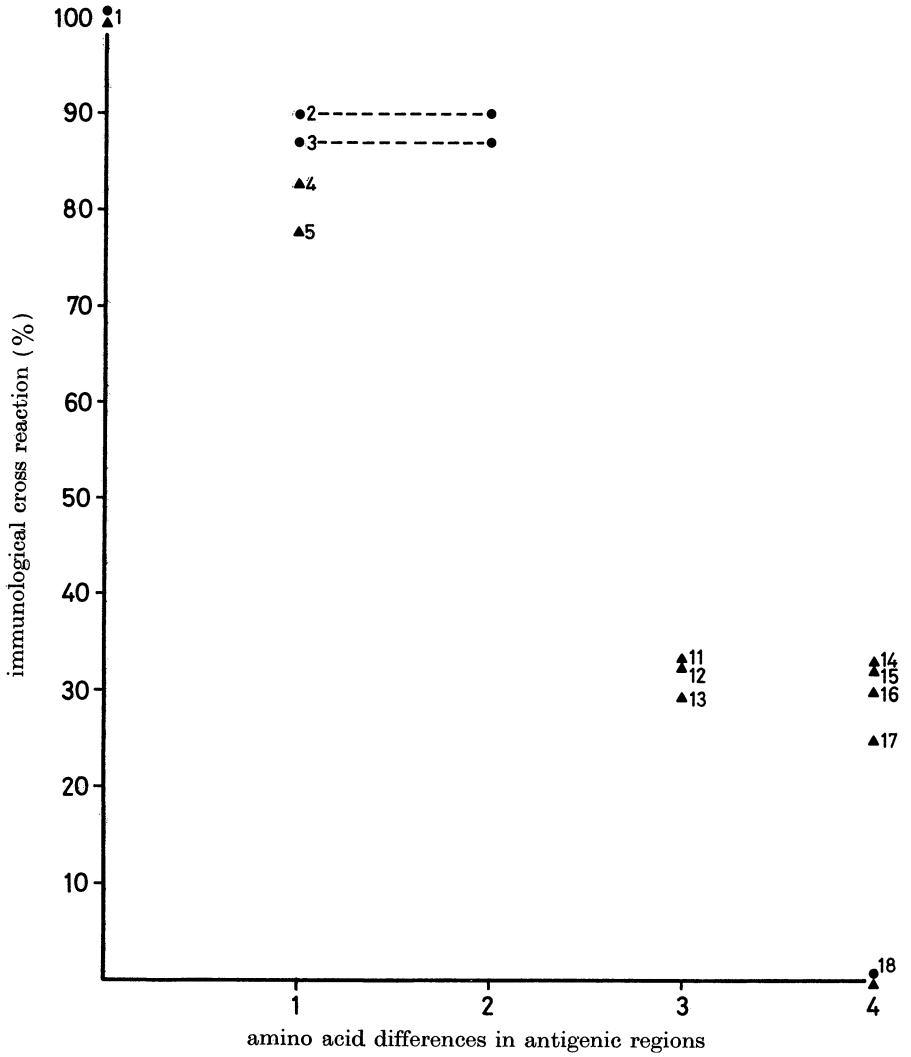


FIGURE 26. As figure 25 but using instead of total amino acid differences between pairs of species the number of differences between them only in the antigenic reactive regions of Atassi (1973). Since a full sequence was not available for finback whale myoglobin points 6,7,8,9 and 10 do not appear on this figure.

following considerations: the myoglobin sequences of horse and man differ by 18 amino acids, while those of horse and sperm whale differ by 19. Yet goat antisera to horse myoglobin react with human myoglobin to the extent of some 30%, while the same antisera do not react at all with the myoglobin of sperm whale (Atassi *et al.* 1970). This is the more striking when one notes that of the five residues by which horse differs from sheep in the antigenic reactive regions (again substituting sheep for goat) man and the sperm whale share 4 each. Also, both man and sperm whale differ from horse in the reactive regions by four residues and so, in terms of simple differences, the two comparisons are almost the same.

Since this text was completed a review by Reichlin (1975) has been noted. This should be consulted for an illuminating treatment of the antigenicity of globin proteins. We are unable, however, to agree with Reichlin's hypothesis that the myoglobin antigenic reactive regions of Atassi (1973) are particularly variable compared to the rest of the molecule. Figure 27 was prepared especially to be comparable with a similar presentation in Reichlin's text.

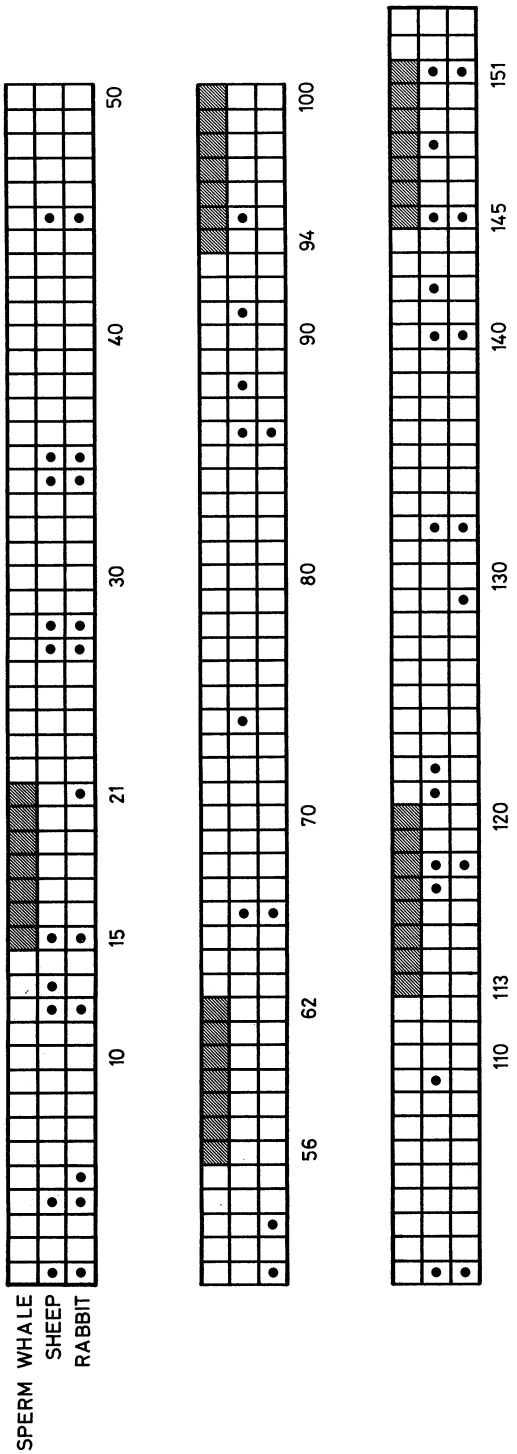


FIGURE 27. An alignment of the myoglobins of sperm whale, sheep and rabbit. The shaded regions in the sperm whale sequence represent the antigenic reactive regions of Atassi (1973), the solid circles for the sheep and rabbit indicate the positions at which they differ from sperm whale. Sequential numbers are given below the sequences.

7. Rates of evolution

Estimates of the rate of molecular evolution are of interest from two complementary points of view. The first concerns the nature of evolutionary change at the molecular level. Since several influential papers in which it was argued that the major contribution to such change was from the fixation of mutations which were neutral or nearly neutral as far as selection was concerned (Kimura 1968; King & Jukes 1969; Ohta 1974), the extent of the fixation of neutral versus non-neutral mutations has been widely debated. The conclusions of Kimura depended partly on his claim that, considering the cost of selection, the observed amounts of change at the molecular level over periods whose duration appeared well established were too great to be accounted for if one assumed selective processes to be operating on the majority of mutations.

The second area of interest concerns the availability of a molecular clock, on the assumption that there is some measurable parameter in molecular evolution which (when optimistically adjusted to take account of unrecorded events) has changed at an approximately constant rate. In establishing the timing of phylogenetic events a molecular clock would clearly be of considerable value. A more technical issue concerns the sensitivity to differences in rate of some of the procedures employed to estimate phylogeny (see §4).

Pooled evidence from several molecules combined with a dated phylogenetic tree has been used to suggest that molecular information may be sufficiently reliable for dating purposes (Langley & Fitch 1974), but because the underlying process is of a stochastic rather than deterministic nature misleading differences might occur over relatively short periods of time. With myoglobin the dating of events in the evolution of the hominoids, for example, might be expected to suffer from sampling errors in the clock. On the other hand, if a molecule has become 'saturated' when one organism is compared with others there may have been manifold substitutions at any given site, and estimation of the true number of events is difficult. This is the case for the myoglobin of *Aplysia* compared with the myoglobin of any of the birds and mammals. For estimating dates, therefore, it is clearly important to choose molecules which have accepted change at an appropriate average rate.

It should be emphasized that differing rates of evolution do not preclude the reconstruction of phylogeny (§4), and that in all phylogenetic reconstructions the nature of the change, divergent versus convergent and parallel, introduces a further fundamental difficulty. Conversely, detecting differences in rates of evolution of the same molecule between different lineages need not involve estimates of time provided that one can construct with confidence a phylogenetic tree and then compare amounts of change in lineages arising from the same point of branching.

Dates of divergence

In order to investigate the absolute rate of substitutions in the lineages leading to the present species it is necessary to establish best estimates of the dates of divergence of those lineages. Because the actual pattern of phylogeny is not known several alternative phylogenetic patterns have been discussed in earlier sections. These alternative patterns reflect different interpretations of the nature and sequence of the dichotomies. A particular point of dichotomy in one phylogenetic pattern may not have a homologue in an alternative pattern, and so the date of divergence between two living forms sometimes depends upon the chosen pattern of phylogeny. We have chosen cladogram 2 (figure 3) as the basis for our further investigation of

rates, this being that one of the two most parsimonious solutions which happens to favour parallel mutations, as described in §3.

As in our previous work (Romero-Herrera *et al.* 1973) we have used the 'Geological Society Phanerozoic time scale' (1964), with modifications suggested in the Supplement (1971), for most of the geological succession, but we have adopted Berggren & Van Couvering's scale (1974) for post-Eocene time. In practice the adoption of the new scale has had little effect because already many of the dates were based on radiometric brackets and had been placed on the absolute time scale regardless of their stratigraphical correlation.

We recognize, of course, that all attempts to date the divergence between two lineages are to some extent unreliable by reason of the imperfection of the geological record, of the problems inherent in stratigraphical correlation and of the uncertainties which bedevil all dating techniques. We have sought to establish, on the basis of fossil evidence, the best estimate of the point in time at which it can be demonstrated that the ancestors of two living forms had diverged. We recognize that such direct evidence indicates that the event had already occurred. In consequence, our dates of divergence are likely to be minimum estimates, but we have rejected speculations regarding earlier dates of dichotomy which require assumptions about fossils not yet found, because such speculation is open-ended.

Various problems have arisen in attempting to assign dates to the points of dichotomy present in cladogram 2, and so each point will be considered separately, defined by reference to pairs of living species.

Man/chimpanzee, equivalent to Hominidae/Pongidae. Assuming that the Hominidae is a monophyletic group which arose from within the Pongidae, then divergence can be established on the basis of the earliest known hominid. *Ramapithecus* has been widely accepted as a hominid and the most accurately dated record is from Fort Ternan, Kenya (Andrews 1971), with a radiometric age of 14 Ma (Bishop, Miller & Fitch 1969). In our earlier work we accepted the then current consensus of opinion among primatologists that *Ramapithecus* provides the most probable point of dichotomy, although we expressed doubts as to whether sufficient was known about this animal to be sure of its hominid affinity (Romero-Herrera *et al.* 1973). *Ramapithecus* does not have the rounded dental arcade postulated in earlier reconstructions (Walker & Andrews 1973), and the pendulum of opinion has, to some extent, swung away from its hominid affinity (Delson & Andrews 1975). The earliest undoubted hominid is *Australopithecus* from Lothagam 1 Bed, Kenya, with an estimated date of 5 Ma (Patterson, Behrensmeyer & Sill 1970), and we have adopted this date for the present study.

Chimpanzee/gorilla. The relationship between these two species and man has been discussed in earlier sections of this paper. The phylogenetic pattern shown in cladogram 2 reflects the present classification; it assumes that the divergence between the chimpanzee and gorilla lineages post-dates the divergence of the hominids from the pongids, and this was neither confirmed nor denied by the evidence of myoglobin. Some authors prefer to place chimpanzee and gorilla in the single genus *Pan*, whereas others, including ourselves, prefer to regard them as belonging to distinct genera. Adopting the former view would provide the only known example of a difference in the myoglobin sequence between two species belonging to the same genus, but little weight can be attached to this observation because only one pair of congeneric species (horse and zebra) has been investigated so far. Both viewpoints are compatible with cladogram 2. At present there is no fossil material which can be regarded without doubt as ancestral to one of these species and not to the other, and so no date is provided for this dichotomy.

Gorilla/gibbon, equivalent to Pongidae/Hylobatidae. Within the Hominoidea the relationship of these two families is obscure, and so it is necessary to establish the earliest time at which representatives of both groups were present. The suggestion that *Aeolopithecus* (Yale quarry I, Fayum, Egypt) has hylobatid affinities (Simons 1965), in contrast to the pongid affinities of the better known *Aegyptopithecus* (same site), is not generally accepted. *Limnopithecus legetet*, at one time regarded as a hylobatid, is now believed to be more closely related to the pongid *Dryopithecus*, but '*Limnopithecus*' *macinnesi*, soon to be assigned to a new genus, is still regarded as a probable hylobatid (Delson & Andrews 1975). The coexistence of this form with *Dryopithecus* (*Proconsul*) at Mfwangano Island, Kenya provides a best estimate of 20 Ma, based on radiometric dating.

Gibbon/baboon, equivalent to Hominoidea/Cercopithecoidea. Within the Catarrhini the relationship of the hominoids and cercopithecoids is obscure and so it is necessary to establish the presence of both groups in order to be confident that divergence had already occurred. We have hitherto accepted the parapithecine *Apidium* as a cercopithecoid (Simons 1972) and regarded its presence at Yale Quarry G, Fayum, alongside the hominoid *Propliopithecus*, as indicating the most probable point. But it is not universally accepted that *Apidium* and *Parapithecus* are cercopithecoid, and the weight of opinion has moved away from their assignment to this group. Undoubted cercopithecoids are known from the Miocene of East Africa and according to Simons (1972) the oldest known is *Victoriapithecus macinnesi* recorded from Rusinga Island, Kenya (von Koenigswald 1969), dated at about 18 Ma. We understand from Dr P. J. Andrews (personal communication) that, judging from its preservation, all the *Victoriapithecus* material is likely to have come from Maboko Island, Kenya, which on faunal grounds is probably of comparable age to Fort Ternan at 14 Ma. Although some unnamed cercopithecoid material has been found at Napak, Uganda (*ca.* 19 Ma), all of the fossil cercopithecoids recorded from East Africa are younger than the date of the subsequent divergence between hylobatids and pongids (20 Ma). As we are seeking an earlier date for the divergence between hominoids and cercopithecoids, then no reliable date can be provided at present.

Baboon/macaque. This dichotomy within the cercopithecoids represents the derivation of the baboons from the less specialized stock which includes macaques. *Dolichopithecus*, *Procynocephalus* and *Libypithecus* have been interpreted as terrestrially adapted macaques rather than baboons (Simons 1972), and the gelada *Theropithecus* (= *Simopithecus*) is first known at about the same time as *Parapapio*, which is regarded as ancestral to the modern baboon *Papio* (Freedman 1957). The earliest record of *Parapapio* (Kanapoi Beds, Turkana, Kenya) is dated at about 4 Ma (Patterson *et al.* 1970).

Macaque/woolly monkey, equivalent to Catarrhini/Platyrrhini. Within the Anthropoidea the relationship between these two major groups is obscure; it seems reasonable to work on the assumption that those catarrhines which retain only two premolars are unlikely to have given rise to platyrrhines, which characteristically have three premolars. Hence the earliest record of an undoubted catarrhine with only two premolars can be taken as evidence that divergence had occurred. *Amphipithecus* (Upper Eocene, Burma) might be a catarrhine, but it has three premolars (Simons 1972). *Oligopithecus* (Yale quarry E, Fayum, Egypt) has only two premolars, is probably catarrhine and has been interpreted as a hominoid (Simons & Pilbeam 1972), but has been the subject of considerable differences of opinion, summarized by Simons (1971). *Propliopithecus* (Yale quarry G, Fayum) also has only two premolars and is generally accepted both as a catarrhine and as a hominoid. The dates usually attributed to the various fossiliferous horizons in the Fayum sequence are somewhat arbitrary estimates projected from the

radiometric date on an overlying basalt (Tattersall 1970). The type of *Oligopithecus* was found about 44 m below the level at which a tooth of *Propliopithecus* has been found *in situ*, but the source of the type of *Propliopithecus* is not known. The often quoted age of 30 Ma used here could be applied equally well to either horizon.

Woolly monkey/squirrel monkey. Among the Cebidae none of the few known fossils gives any indication of a possible time of divergence between the woolly monkey and squirrel monkey.

Squirrel monkey/marmoset, equivalent to Cebidae/Callitrichidae. The relationship between these two groups of Ceboidea (New World monkeys) is obscure. Absence of the third molar in living callitrichids merely suggests that they could have been derived from an ancestral form with three molars, characteristic of cebids. Clearly the presence of three molars does not debar an animal from callitrichid ancestry and it has been suggested that *Dolichocebus* is a callitrichid and *Homunculus* a cebid (Romer 1966); conversely Simons (1972) and Hershkovitz (1974) do not accept the former as a callitrichid. In the absence of any other fossil callitrichid no date can be provided for this dichotomy.

Marmoset/galago, equivalent to Anthropoidea/Prosimii. Among the Primates, assuming that the suborder Anthropoidea arose from within the suborder Prosimii, then at first sight this dichotomy can be fixed at the earliest record of the anthropoids. On the other hand, it is widely accepted that the Anthropoidea arose from the infraorder Tarsiiformes within the Prosimii. As it is most unlikely that Tarsiiformes gave rise either to the infraorder Lemuriformes or to the infraorder Lorisiformes (which includes *Galago*) then the dichotomy might be represented by the earliest known member of the Tarsiiformes (excluding the Paromomyidae, which some authorities now place in the infraorder Plesiadapiformes) (Simons 1972). *Berruvius* (Upper Palaeocene, France) and *Navajovius* (Upper Palaeocene, Colorado) are possible candidates, but Szalay (1972) regards both as paromomyids. Several anaptomorphid tarsioids (including *Tetonius*) were present in the Lower Eocene (Simons 1972), and we have taken 52 Ma as the best available date, although aware that the ancestry of the anthropoids remains open to dispute.

Galago/potto, equivalent to Galaginae/Lorisinae. The earliest fossil lorisids are *Mioeuoticus*, *Progalago* and *Komba*, now known from several Miocene sites in East Africa (Walker 1970, 1974). In his most recent review of the material Walker (1974) has assigned *Mioeuoticus* to the lorisines and regards *Progalago* and *Komba* as galagines. All three genera are present at Rusinga, Kenya and Napak, Uganda, radiometrically dated at about 18 and 19 Ma respectively (Van Couvering & Miller 1969; Bishop, Miller & Fitch 1969).

Potto/slow loris. There is no fossil material relevant to the divergence between the ancestors of these two modern lorisines and so no date is available.

Slow loris/sportive lemur, equivalent to Lorisidae/Lemuridae. There is general agreement that the Lorisiformes were derived from early members of the Lemuriformes and their divergence cannot be established until such time as both groups were present. Lemuriformes are represented in the Eocene, but Lorisiformes are unknown before *Mioeuoticus*, *Progalago* and *Komba* (Miocene, East Africa). As described above, these genera provide evidence for the divergence between lorisine and galagine lorisids, which must postdate the dichotomy now being considered. Hence, in the absence of any earlier fossil lorisids no date can be provided.

Sportive lemur/sperm whale, equivalent to Primates/Condylarthra, which is based on the assumption that the order Cetacea was derived from within the extinct order Condylarthra. *Purgatorius* (known by many teeth in the Lower Palaeocene, Montana and one tooth from the

Maestrichtian, Montana) has been claimed as the earliest known Primate (Van Valen & Sloan 1965; Szalay 1969). This form hardly differs from several early eutherians, including arctocyonid Condylarthra (Szalay 1968; Simons 1972), and doubt has been cast on its assignment to the Primates (Simpson 1971). There is a body of opinion that the Cetacea may have arisen from mesonychid Condylarthra (McKenna 1969; Van Valen 1966), and it is generally accepted that the order Condylarthra was ancestral to the Artiodactyla and Perissodactyla. The earliest known condylarthran is *Protungulatum* (Maestrichtian, Montana) (Sloan & Van Valen 1965). If we accept *Purgatorius* as a primate and *Protungulatum* as a condylarthran, both first known at the same time, then a date of 68 Ma is the best estimate for this point of divergence. If *Purgatorius* should prove not to be a Primate, then many Middle Palaeocene primate genera would provide dates at about 60 Ma (Simons 1972).

Sperm whale/dolphin, equivalent to Physeteroidea/Delphinoidea. Among the toothed whales (Odontoceti) the earliest indication of divergence between the more specialized sperm whales (Physeteroidea) and the less specialized dolphins (Delphinoidea) is provided by *Physeter* itself (Lower Miocene, Europe) and by *Apenophyster*, *Diaphorocetus* and *Idiorophus* (Lower Miocene, South America) (Simpson 1945).

Dolphin/porpoise, equivalent to Delphinidae/Phocaenidae. Among the Delphinoidea the divergence between these two groups cannot be established until such time as both groups were present. Several dolphins are recorded from the Lower Miocene but the earliest known porpoises are *Palaeophocaena* and *Protophocaena* (Upper Miocene, Europe) (Kellogg 1928; Slijper 1936, Simpson 1945). Before our finding errors in the sequence of the dolphin (see footnote on page 68) it appeared that, in terms of the myoglobin sequence there was a greater similarity between sperm whale and dolphin than between dolphin and porpoise. This greater similarity had been interpreted either as a product of remarkable parallel evolution or as an indication that the divergence between dolphin and porpoise might predate that between sperm whale and dolphin. In view of this uncertainty we regarded this point as unreliable and did not make use of the available date.

Porpoise/horse, equivalent to Cetacea/Perissodactyla, which again assumes that the Cetacea were derived from the Condylarthra. If it is accepted that the Cetacea may have arisen from the mesonychid condylarthrans then it could be argued that the differentiation of the mesonychids from the arctocyonids represents the point of divergence of Cetacea. On the other hand, it is not inconceivable that the Cetacea arose from arctocyonids through some other intermediate group still unknown. Hence this point of dichotomy must be based on the earliest recorded cetacean. The single bone of *Anglocetus* (Lower Eocene, England) described by Tarlo (1964) may be chelonian rather than cetacean (Halstead, personal communication), but even this date (52 Ma) cannot be used because the subsequent divergence between the lineages leading to zebra and ox has been dated at 54 Ma.

Horse/zebra. Azzaroli (1966) recognized four subgenera within the genus *Equus*, one of these including the horse *Equus* (*Equus*) *caballus* and another including the zebra *Equus* (*Hippotigris*) *burchelli*. Quinn (1955) regarded *Equus* and *Hippotigris* as being distinct in the Middle Miocene of North America, but this claim was based on dental features not accepted by Azzaroli (1966). Nevertheless, the later author agreed that the evolutionary lines leading to the living species seem to have begun to differentiate before the Pliocene migration from North America. According to Cruickshank (1974), Azzaroli now believes that the split between the *Equus* (*Equus*) group and *Equus shoshonensis*, the common ancestor of the other three subgenera, took place

in pre-Blancan times in North America. Based on this indirect information we have tentatively accepted a date of divergence at 4 Ma.

Zebra/ox, equivalent to Perissodactyla/Artiodactyla. It is generally accepted that each of these orders arose separately from within the Condylarthra, and so the first undoubted record of either group can be used as evidence of divergence. The perissodactyl *Hyracotherium* (Late Palaeocene/Eocene transition, Wyoming) (Jepson & Woodburne 1969) just antedates several artiodactyls, including *Diacodexis* in the lowest Eocene (Hooijer 1967; Van Valen 1971), so providing a date of 54 Ma.

Ox/sheep, equivalent to Bovinae/Caprinae. Within the family Bovidae the ox and sheep are included in different subfamilies. We have accepted Pilgrim's opinion (1939, 1947) that the boselaphines are ancestral to the bovines, but unlikely to have given rise to the caprines. Dr A. W. Gentry has kindly informed us that it is unlikely that ox and sheep would have a common ancestry less than 18 Ma ago, and impossible that they should have done so less than 14 Ma ago. The boselaphine *Eotragus* is recorded from Artenay, France (Ginsburg & Heintz 1968), dated at about 17 Ma (Van Couvering 1972) and at Gebel Zelten, Libya (Hamilton 1973) which is regarded as equivalent to early Burdigalian (Savage & Hamilton 1973), corresponding to about 18–19 Ma (Berggren & Van Couvering 1974). The earliest definite caprine is *Oioceros*, different species of which are known from Tung Gur, Mongolia and Fort Ternan, Kenya (Gentry, personal communication), the latter site being dated at 14 Ma. As we see no reason to doubt Pilgrim's judgement that the boselaphines are unlikely to be ancestral to caprines, we have adopted the date of 18 Ma.

Sheep/tree shrew, equivalent to Primate–condylarthran stem/Tree shrew–hedgehog stem. There are no undoubted fossil tupaiids, and the Adapisoricidae, which are regarded as the earliest erinaceomorphs (Butler 1972), are first known from the Middle Palaeocene of Montana (Saban 1958), dated at about 59 Ma. This available date is too late to be relevant because the divergence between Primates and Condylarthra, which is shown as subsequent to the dichotomy under consideration, has already been dated at 68 Ma.

Tree shrew/hedgehog. The date given above for the earliest erinaceomorphs is relevant here because they are regarded as being unsuitable as ancestors to the tupaiids (Butler 1972). Whatever the relation between the tupaiids and the erinaceomorphs might be, we hold the view that they diverged not less than 59 Ma ago.

Hedgehog/harbour seal, equivalent to origin of Carnivora. The relationships of the Carnivora are not known and, although it is possible that the miacid carnivores were derived from the arctocyonid condylarthrans, the phylogenetic pattern in cladogram 2 is one which implies that the Carnivora were a distinct group before the origin of the Condylarthra, already dated at 68 Ma. The earliest known Carnivora are the miacids *Protictis* and *Ictidopappus* (Middle Palaeocene, North America), although a possible *Ictidopappus* tooth has been reported from the Lower Palaeocene (MacIntyre 1966). Nevertheless, as these records at about 60 Ma are younger than a subsequent dichotomy dated at 68 Ma, no reliable date can be given to this point of divergence.

Harbour seal/dog, equivalent to Pinnipedia/Fissipedia. As discussed in an earlier section, it is not known whether the Pinnipedia diverged from the Fissipedia before the radiation of the latter group, or whether the Pinnipedia might be a polyphyletic group derived from different families within the Fissipedia (McLaren 1960). In particular it has been suggested that the Phocidae might have affinities with the lutrine Mustelidae (Savage 1957), but in the phylogenetic pattern, adopted in cladogram 2, the divergence of the Phocidae is shown as antedating

the divergence between Canidae and Mustelidae. As no fossil seals are recorded before the early Miocene (Savage, personal communication) whereas the canids and mustelids had diverged by late Eocene, no reliable date can be provided for this point on the cladogram.

Dog/hunting dog. These two dogs have been placed in separate genera, *Canis* and *Lycaon* respectively, which are usually referred to different subfamilies, the Caninae and Cuoninae (= Simocyoninae), within the Canidae (Simpson 1945; Piveteau 1961). If it were assumed that these subfamilies were monophyletic then the divergence between the two genera would have to be placed in the Oligocene, but in a recent review of the Canidae (Clutton-Brock, Corbet & Hills 1976) these subfamilies are no longer recognized as worthwhile groups. The abandonment of the subfamilies requires us to make some other assumption about the possible relation between these two genera. We have chosen to assume only that they are sufficiently different that *Canis* is unlikely to have given rise to *Lycaon*, or vice versa, and on this basis the earliest undoubted record of either genus can be taken as evidence that divergence had occurred. *Lycaon* is known only from the Pleistocene, whereas *Canis* has been recorded from the Hemphillian–Blancan boundary (dated at 4 Ma) in North America where, it has been claimed, the genus evolved from smaller Hemphillian canines which had been a separate lineage since the Miocene (Williams 1969). On the other hand, according to Piveteau (1961) it is claimed that a true *Canis* was already present at Roussillon, France, for which Berggren & Van Couvering (1974) have suggested a date of about 5 Ma.

Hunting dog/badger, equivalent to Canidae/Mustelidae. These two families are included within the superfamily Canoidea, and although they were both derived from an arctoid stock it is generally agreed that neither family is likely to be derived from the other. On this basis, the first undoubted record of either family can be taken as evidence of divergence. The oldest known members of the Mustelidae are recorded from the Lower Oligocene, but several genera of Canidae, including *Procynodictis* and *Cynodictis*, have long been known from the Upper Eocene of North America and Europe (in the Uintan and Bartonian Stages, respectively) (Piveteau 1961), and so it is possible to assign a date of about 42 Ma.

Badger/kangaroo, equivalent to Eutheria/Metatheria. It is generally accepted that the placentals and marsupials diverged from a common therian stem, but it is controversial whether this stem was metatherian or eutherian or neither. The affinities of *Pappotherium* and *Holoclemensia* (Albian, Texas), claimed as eutherian and metatherian respectively (Slaughter 1971), are open to question (Clemens 1971; Simpson 1971). *Zalambdalestes* and *Kennalestes* (Bayn Dzak, Mongolia) are undoubted eutherians and unsuitable as ancestors to Metatheria. Kielan-Jaworowska (1968) suggested a possible Coniacian/Santonian age for the Bayn Dzak fauna, and McKenna (1969) preferred a Cenomanian age, but Crompton & Kielan-Jaworowska (personal communication) now favour a Santonian age, on which we have based our date of 79 Ma. A record of Santonian marsupials in North America (Graham & Ride 1967) is not currently accepted (Clemens 1971) but marsupials and placentals do coexist in the Lower Campanian of Alberta (Fox 1970), and so a date of 75 Ma would become relevant if Bayn Dzak should turn out to be even younger. At the other extreme we have taken note of *Prokennalestes* and *Prozalambdalestes* recorded from the Khovboor Beds of the Gobi Desert, Mongolia (Beliajeva, Trofimov & Reshetov 1974), but not yet described. Their chosen names imply eutherian affinity and if this should be confirmed, and linked with evidence that they are unsuitable as marsupial ancestors, then their estimated Aptian age would become relevant, being equivalent to a date of about 109 Ma.

Kangaroo/opossum, equivalent to Macropodidae/Didelphidae. During recent years there have

been several suggestions that the order Marsupialia deserved to be split into several orders, but regardless of the rank or classification chosen, these two marsupials are very distant from one another. The divergence between them is that between American and Australian marsupials which, on zoogeographic grounds, is presumed to have taken place when Australia lost any effective land connection to South America. Various dates have been proposed according to different tectonic models, but at the present stage we prefer not to introduce indirect estimates of dates of divergence. Turning to the fossil record of marsupials in Australia it is found that the earliest faunas include representatives of all the modern Australian families, and so they are evidently too late to provide a reliable date for the dichotomy under consideration.

Opossum/chicken, equivalent to mammals/birds. It is generally accepted that the mammals are descended from synapsid reptiles and that the birds are derived from the archosaurian diapsid reptiles. Although it is conceivable that early synapsids might have given rise to diapsids, the converse is unacceptable. On this basis, and assuming the Diapsida to be a monophyletic group, the earliest undoubted record of a diapsid reptile can be taken as evidence that divergence between the ancestors of birds and mammals had occurred. We understand from Robert Reiss (personal communication) that his reinvestigation of *Petrolacosaurus*, from the Upper Pennsylvanian of Kansas (Peabody 1952), has established that it is diapsid. The horizon at which it was found has been correlated with the middle Stephanian, and is equivalent to a date of 293 Ma. In adopting this date, we are aware that the archosaur radiation was essentially Triassic and that the earliest known bird is Jurassic (Kimmeridgian, equivalent to about 148 Ma), but we are also impressed by the radiation of synapsids in the early Permian, giving rise to mammal-like reptiles which cannot possibly have been ancestral to archosaurs.

Chicken/penguin, equivalent to Galliformes/Sphenisciformes. The chicken and penguin belong to different orders of birds and their relationship is a matter for conjecture. Nevertheless, representatives of each of these orders have been recorded from the Lower Eocene (Fisher 1967), and even assuming that the Galliformes might not be a monophyletic group it seems unlikely that any Galliformes were derived from Eocene penguins. Although we have adopted 52 Ma as the best available date for this point of divergence, we are very conscious that it might have taken place much earlier.

The following points of divergence are not represented in cladogram 2, but the dates need to be established for the subsequent discussion of rates.

Penguin/lamprey, equivalent to Gnathostomata/Agnatha. Assuming that the Gnathostomata are a monophyletic group then this divergence can be equated with the earliest evidence of a gnathostome, provided by the acanthodian *Nostolepis* (Llandovery/Wenlock, Czechoslovakia) (Miles 1967), which is equivalent to a date of about 420 Ma.

Lamprey/Aplysia, equivalent to Chordata/Mollusca. We have adopted the classical view that the phyla Mollusca and Polychaeta are more closely related to one another than is either to the Chordata. On this basis the hypothetical common stem to the Mollusca and Polychaeta must have diverged from the Chordate stem at some time earlier than the divergence between the Mollusca and Polychaeta. As this latter divergence has been dated at not less than 585 Ma (see below), we have arbitrarily added 5 Ma in order to provide a date of divergence between Chordata and Mollusca (and between Chordata and Polychaeta) at not less than 590 Ma.

Aplysia/Glycera, equivalent to Mollusca/Polychaeta. Although Mollusca and Polychaeta may well have arisen from a common stem, the Mollusca are regarded as too specialized to have been ancestral to the Polychaeta, and so the earliest undoubted record of a mollusc can

be taken as evidence that the two phyla had diverged. A molluscan fauna, which antedates the earliest undoubted polychaetes, has recently been reported from the lower Cambrian of Siberia (Matthews & Missarzhevsky 1975), equivalent to a date of 585 Ma.

Estimation of rates

In this section we use the estimates of time derived from the dated points of branching together with the evidence of our cladograms to investigate rates of evolution. We also employ various procedures which attempt to correct the observed values of amino acid differences for undetectable changes, and which transform them into estimates of the number of base substitutions. In order to express the amount of change in the myoglobin molecule we have used five methods which are to some extent dependent upon one another.

1. Based on the number of base substitutions reconstructed in cladogram 2 (one of the two most parsimonious, figure 3). Information from the equally parsimonious complementary cladogram 5 (figure 7) has also been utilized where appropriate.

2. Based on the number of amino acid differences between pairs of aligned sequences. These are given in a matrix (table 4).

3. Based on the values in 2 above transformed to take account of multiple changes at a site by the method used in Dickerson (1971).

4. Based on an estimated number of base substitutions, derived from 2 above by equations suggested by Kimura & Ohta (1972), and related to approach 5.

5. Based on an estimated number of base substitutions ('random evolutionary hits') derived from a comparison of homologous positions between pairs of myoglobin sequences using the approach of Holmquist (1972*a*, *b*) and based on an extended algorithm published by Jukes & Holmquist (1972). The relationship between time and the parameter T_2 (the number of variable sites in the molecule), suggested by these authors as a suitable parameter for a molecular clock, is also shown graphically.

The results of each of these approaches will be considered in turn.

Amount of change based on the number of base substitutions reconstructed in the cladograms

Figures 28 and 29 show the two most parsimonious cladograms (2 and 5), each of which scored 281 hits. The number of base substitutions reconstructed within each section of a lineage is indicated by a figure at the bottom right of each branch. The total number of base substitutions accumulated within each lineage leading to the living species appears at the top. These numbers are similar to one another because the different lineages share a common ancestry to some degree, and so the mutations fixed along the lower part of the cladogram are necessarily components of the higher part.

As has been discussed already in §3, the assignment of mutations to either side of the earliest dichotomy (between the ancestors of birds and mammals) presents special difficulties. As an arbitrary measure we allocated all to the avian ancestral stem in cladogram 2 and all to the mammalian ancestral stem in cladogram 5. The mutations involved in this arbitrary allocation have been eliminated from the totals shown at the top of figures 28 and 29, by considering only those mutations subsequent to the earliest dichotomy within the birds and mammals respectively, so making possible a realistic comparison between the two most parsimonious solutions. The minor differences between the totals shown in figures 28 and 29 also reflect the other basic differences between cladograms 2 and 5, the hedgehog lineage being differently placed, and wher-

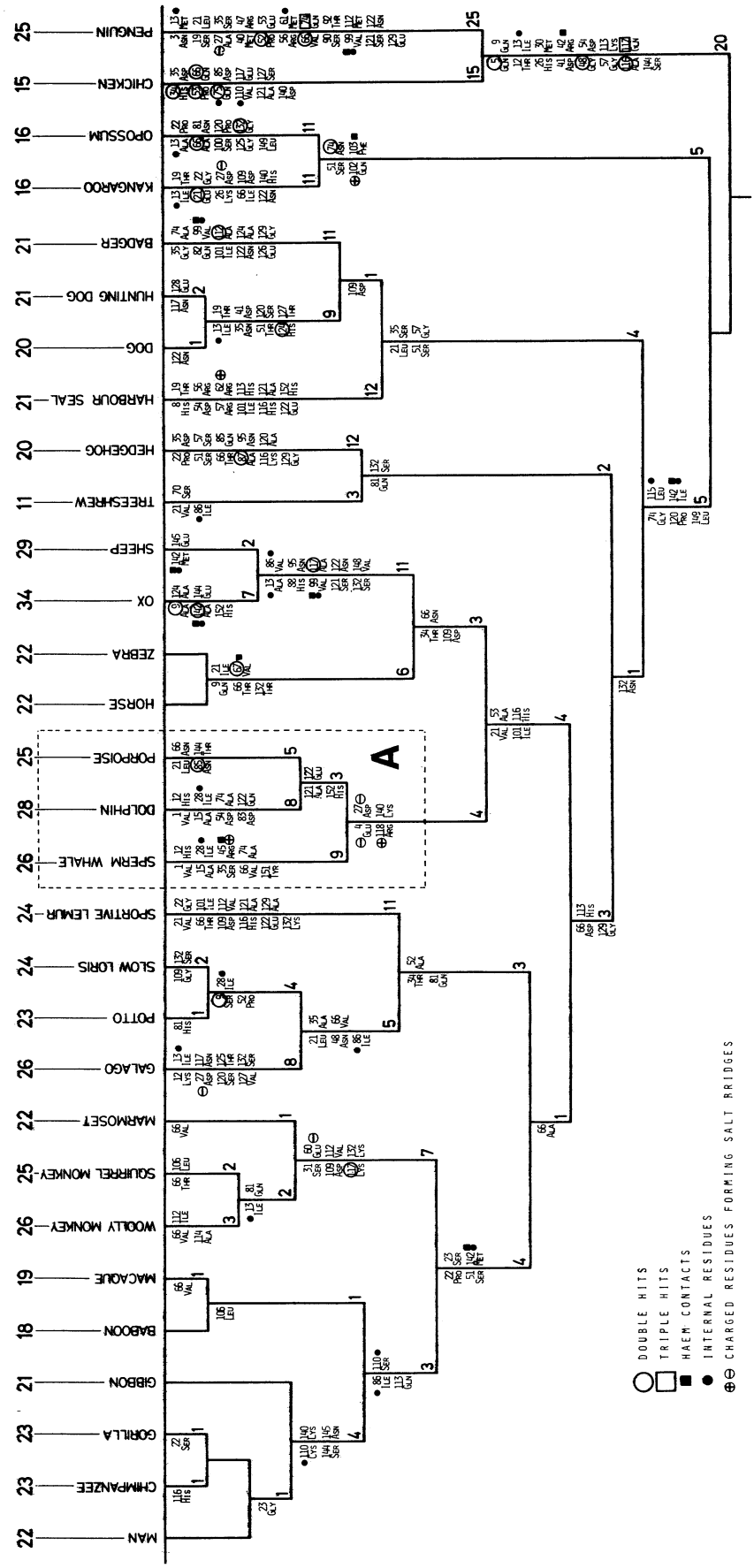


FIGURE 28. Cladogram 2 (see figure 3) with the number of base substitutions reconstructed within each section of a lineage indicated at the bottom right of each. The total number of base substitutions accumulated within each lineage leading to the living species is shown at the top, with the exception of the birds (see text). Alternative solution A has been adopted for the cetaceans.

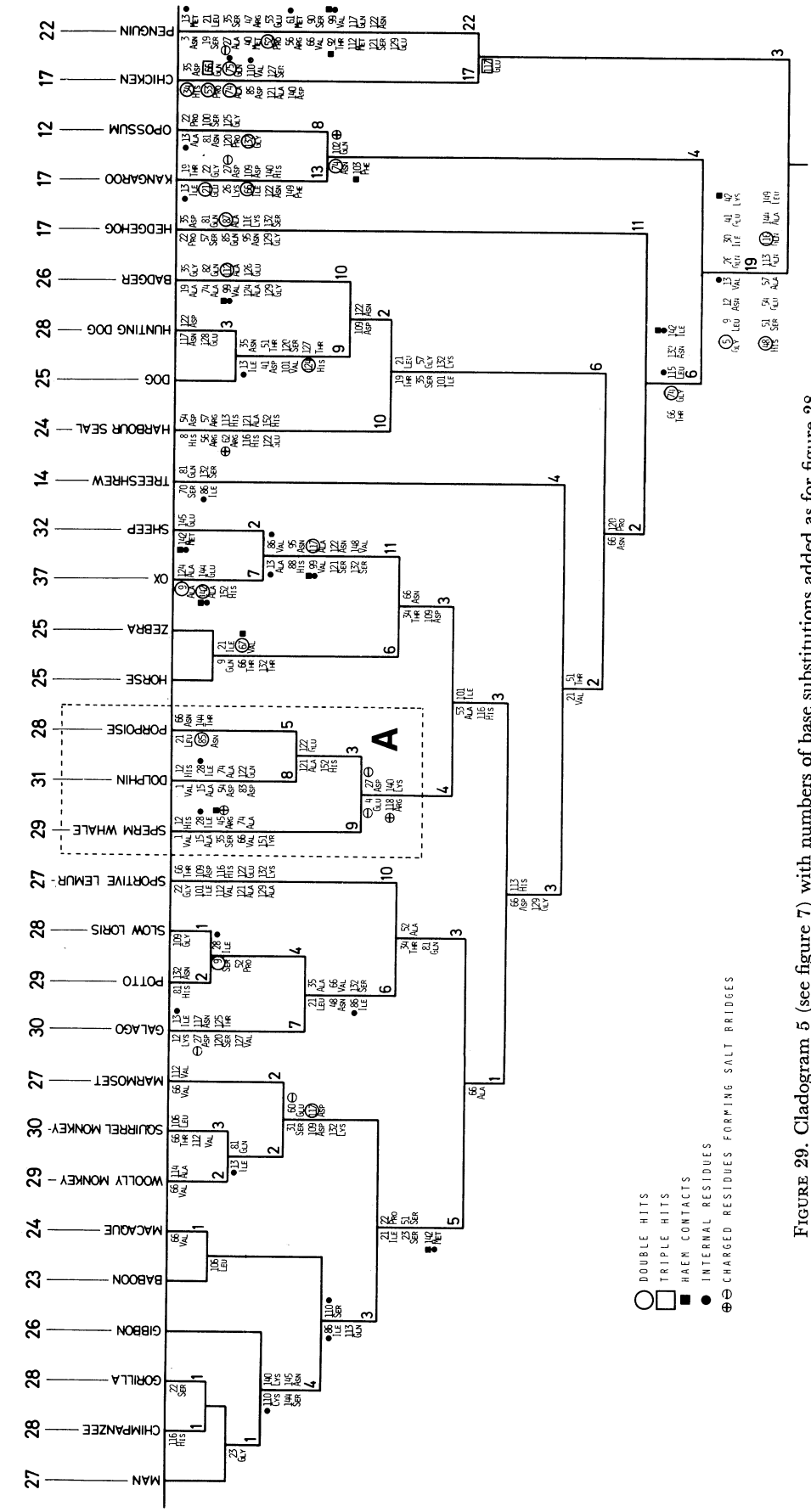


FIGURE 29. Cladogram 5 (see figure 7) with numbers of base substitutions added as for figure 28. Alternative solution A for the cetaceans is shown.

ever possible the choice between parallel and back mutations having been reversed between the two solutions.

With the use of the data derived from figure 28, and the dates of divergence discussed earlier in this section, it is possible to arrive at figure 30. This shows, wherever possible, the best estimate of the date of divergence above each point of dichotomy, and on the right of each branch the time span between successive dichotomies. The number of mutations reconstructed along each lineage between successive points of divergence is included on the left of each vertical line.

The horizontal dotted lines represent dichotomies for which no reliable date was available, and these have not been used in our calculations.

The dichotomies involving the Agnatha, Polychaeta and Mollusca are not included in figure 30 because the first two are represented by haemoglobins, and because the differences between the sequences are so large that there would be considerable uncertainty in reconstructing the course of possible mutational events. Figure 30 shows that the slowest rate of fixation of mutations is present in the lineage leading to the gibbon, which has not fixed any mutation during the last 20 Ma. The fastest rate lies between 79 and 68 Ma; 9 mutations were fixed during this period of 11 Ma, i.e. a rate of one mutation in 1.22 Ma.

Other rates which deserve special comment include several found in the Anthropoidea. Between 30 and 20 Ma ago 7 mutations were fixed along the lineage leading to the Hominoidea, but within the Hominoidea the slow rate already mentioned for the gibbon lineage is almost matched by that of the lineage leading to man (one mutation in 20 Ma). It should be noted that this observation is unaffected by the time of divergence of hominids from pongids. Furthermore, if *Ramapithecus* is a hominid (at about 14 Ma), then the lineage leading to man would have fixed no mutations during the last 14 Ma, and the lineages leading to chimpanzee and gorilla would each have fixed only one mutation during the same period. Hence, whatever position may be adopted regarding hominid phylogeny and dates of divergence all four hominoid lineages so far studied appear to have accepted little change in their myoglobin during the past 20 Ma. This is consistent with the comments made by Goodman, Barnabas, Matsuda & Moore (1971) on the rates of molecular evolution among higher primates. A similar observation applies to the two cercopithecoid myoglobins so far studied, but neither ceboid nor prosimian myoglobins share this feature.

Turning now to the condylarthran derivatives, it appears that the lineages leading to all three cetacean species have shown above average rates of fixation during the past 18 Ma, but it must be remembered that this date of divergence may represent an underestimate of the time involved. Subsequent to the divergence of perissodactyls and artiodactyls, the lineage leading to horse and zebra has fixed 6 mutations in 54 Ma (one in 8.33 Ma), whereas the lineages leading to ox and sheep have fixed 18 and 13 mutations during the same period, i.e. one in 3 Ma and one in 4.15 Ma respectively. On the other hand, after the divergence of the ox and sheep lineages at about 18 Ma ago, 7 mutations have been fixed in the former and only 2 in the latter. This represents rates of one in 2.57 Ma and one in 9 Ma, respectively.

The phylogenetic relationship of the tree shrew and hedgehog is, of course, open to considerable doubt, but the difference between 3 and 12 fixed mutations respectively during the past 59 Ma suggests that the tree shrew lineage has had a relatively slow rate of evolution. Regardless of its point of divergence from the hedgehog lineage, the line leading to the tree shrew has fixed 11 mutations during the past 79 Ma, i.e. one in 7.18 Ma. The low total for the tree shrew stands out in figure 28, and even adopting the alternative pattern shown in figure

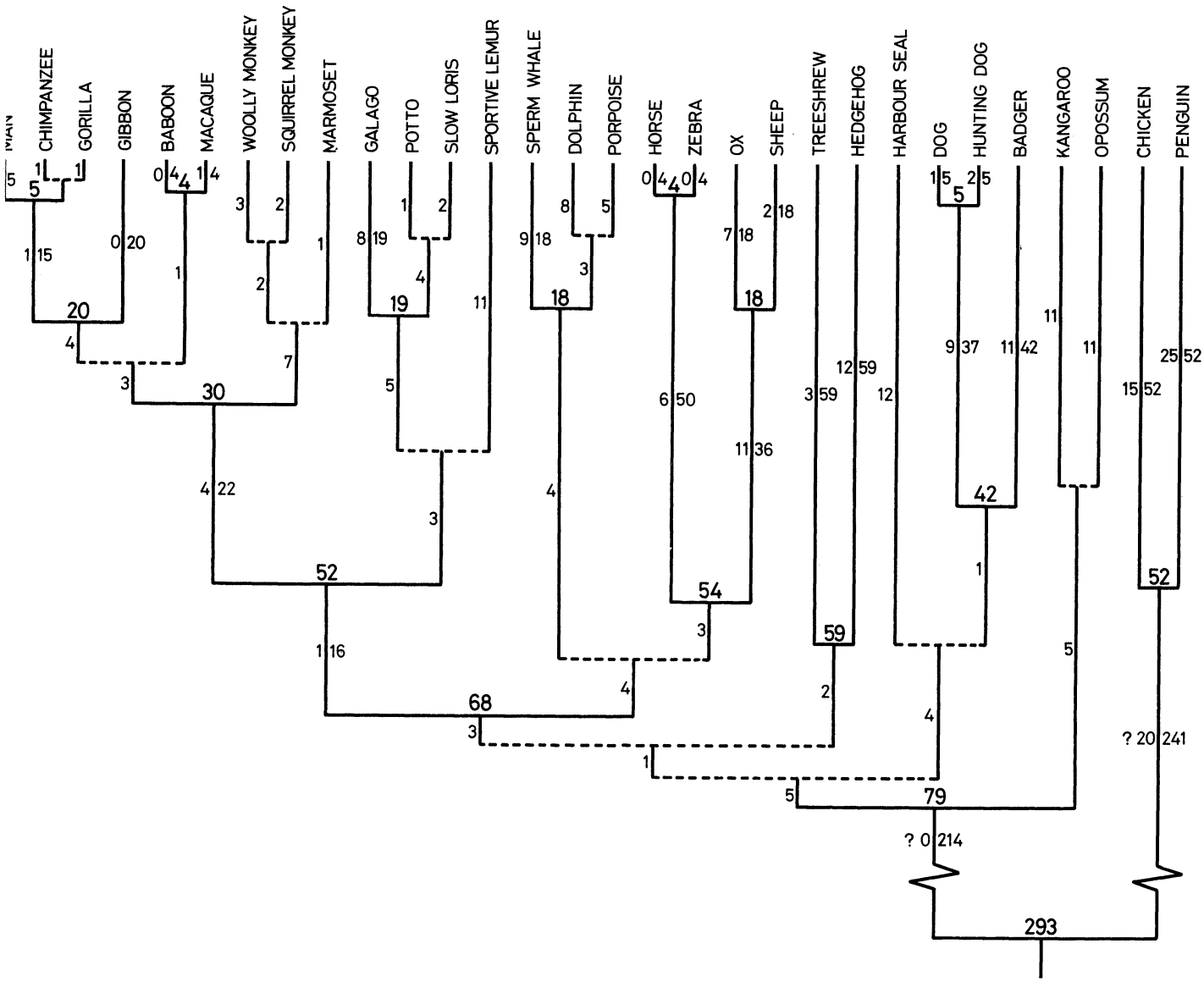


FIGURE 30. A cladogram based on figure 28 showing the date of divergence (where available) above each dichotomy, and to the right of each branch the time span between successive dichotomies. The number of mutations reconstructed along each lineage between successive dichotomies is shown on the left of each branch. Horizontal dotted lines represent dichotomies for which no reliable date was available.

29, the total of 14 mutations is still in the lower part of the range. The low overall rate in figure 28 becomes more impressive when one considers that the early part of the ancestry of the tree shrew includes 6 out of 9 mutations already noted as occurring between 79 and 68 Ma. Similarly in figure 29, the tree shrew shares 10 of 13 mutations during the same 11 Ma period (a rate of one in 0.85 Ma), but the tree shrew itself still emerges with a total of only 14 mutations in 79 Ma (one in 5.64 Ma). This is because only 4 mutations were fixed subsequent to the divergence of its ancestor from the primate/condylarthran stem (some time before 68 Ma ago).

To return to figure 30, both of the marsupials appear to have fixed 16 mutations in 79 Ma (one in 4.93 Ma). However, in the alternative cladogram (figure 29) a total of 12 mutations

is accumulated by the opossum since the dichotomy at 79 Ma (i.e. one in 6.58 Ma) and so the tree shrew lineage no longer has the distinction of having the slowest rate of fixation of mutations.

The rates for birds can only be considered during the last 52 Ma, because of the uncertainty concerning the allocation of mutations either to the bird or to the mammal stem after the initial point of dichotomy. However, the nature of the mutations in the bird stem prior to this date necessarily affects the allocation of mutations within the chicken and penguin lineages. The differences between the common ancestral myoglobin chains reconstructed in cladograms 2 and 5 also affect the reconstruction of the course of events after the 52 Ma point. Because of these two sources of difference the totals of 15 and 25 for chicken and penguin respectively in cladogram 2 become 17 and 22 respectively in cladogram 5. Even taking the lowest available total for the penguin the fixation of 22 mutations in 52 Ma represents a rate of one in 2.36 Ma, which is faster than that of the chicken and of similar magnitude to the faster rates among mammals.

To turn now to the ancestral stems of the mammals and birds subsequent to their divergence about 293 Ma ago, even in the extreme case of assigning all 20 hits to the mammal stem, the rate of fixation of mutations would be remarkably low, being 20 mutations in 214 Ma (i.e. one in 10.7 Ma). A similar observation would apply if all mutations were arbitrarily assigned to the birds. It appears that the myoglobins in the lineages ancestral to both groups fixed only a small number of amino substitutions throughout the late Pennsylvanian, Permian, Triassic, Jurassic and part of the Cretaceous, a span exceeding 200 Ma. The majority of these few changes were of a conservative nature and all but one (residue 13) external in location. It should be remembered, however, that during such a long period of evolution undetectable back mutations and undetectable nucleotide changes yielding the same amino acid (isosemantic mutations) are likely to have occurred.

From cladogram 2 (figures 28 and 30), the number of reconstructed base substitutions has been plotted cumulatively against time (figure 31). The slope of each section indicates the best available measure of the rate of fixation of mutations between two dated points. It should be noted that figure 31 differs slightly in construction from comparable figures in two of our earlier publications (Romero-Herrera *et al.* 1973; Lehmann *et al.* 1974). We then encountered a difficulty arising from the resolution of trichotomies in the phylogenetic chart, in so far as the available date could apply to only one of the two successive dichotomies. In the present study, no satisfactory date is available for several of the dichotomies shown in figure 30 and so, in order to obtain the maximum information about rates, we have bypassed these points and joined the adjacent dated points. This has created pentachotomies and trichotomies which bear no relation to the pentachotomy shown in figure 1, and which have no phylogenetic significance and no implication concerning speciation. In figure 31 it should be noted that, whereas the mammalian lineages radiate from a point at 79 Ma, the divergence of the two bird lineages occurs at 52 Ma.

It may be seen that the rate of fixation of mutations fluctuates along different lines of descent and that periods of high rate may be followed by low-rate episodes, and vice versa. These fluctuations could result in a few cases from the chosen interpretation of the pathway of mutations or they could reflect inaccuracies in the times of divergence, or they could be real.

On the first point it is evident that the allocation of mutations in any cladogram is controlled by the chosen phylogenetic pattern. The chosen interpretation will affect the number of

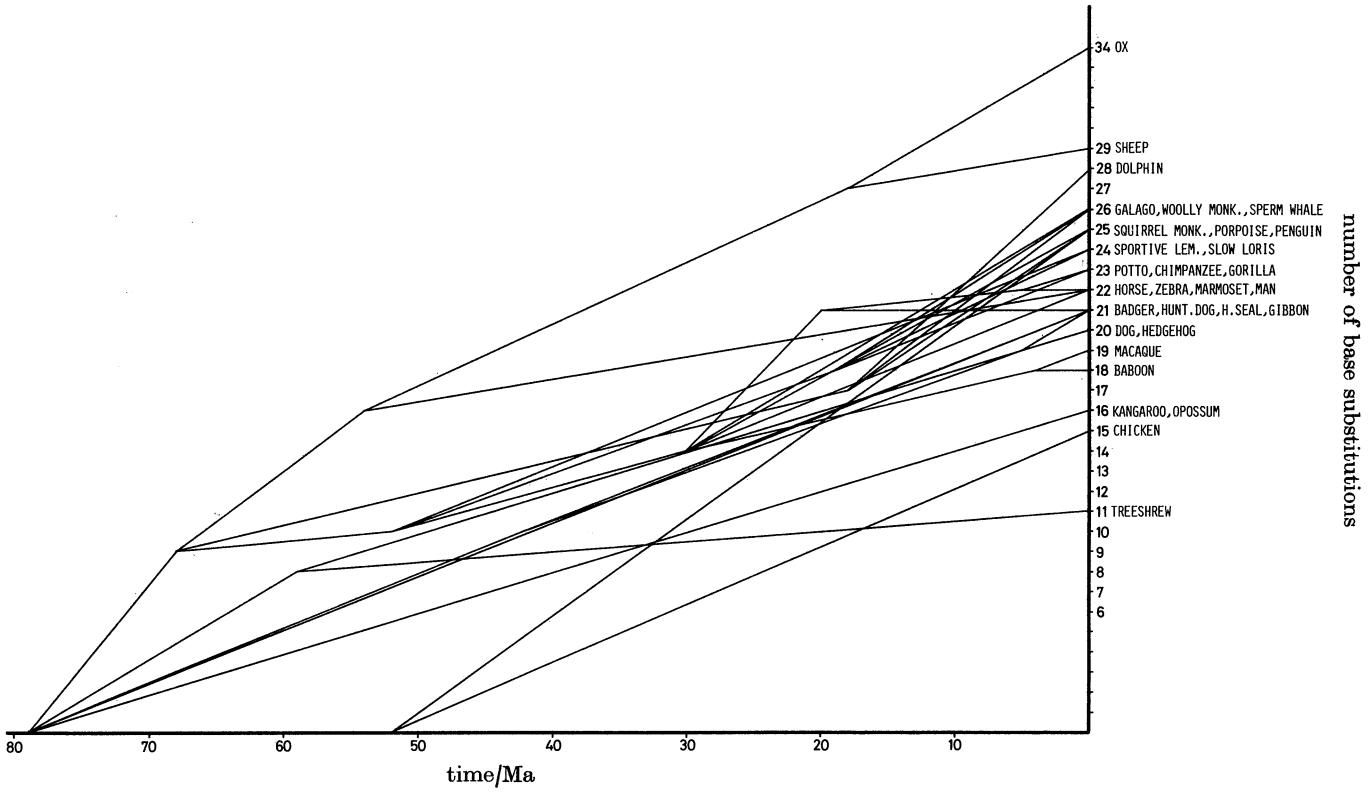


FIGURE 31. Number of reconstructed base substitutions for cladogram 2 (see figures 28 and 30) plotted cumulatively against time. The polychotomies are explained in the text.

mutations assigned along the lineage leading to any given species, and also affect the reconstructed rate. We have chosen to conduct our investigation of rates of fixation of mutations on one of the two most parsimonious cladograms. It is also evident that, within any given cladogram, complementary solutions favouring either back or parallel mutations at the same overall cost will change the allocation of mutations and so cause minor differences in rate. Similarly, when two alternative solutions in a particular cladogram are equally parsimonious, the residue chosen for a given position in the ancestor will affect the rates in some descendant lineages.

On the second point, the dates of divergence provided and discussed earlier in this section are a matter of opinion, and others might prefer alternative interpretations and dates to those given here. Rather than make guesstimates we have chosen not to speculate beyond the fossil evidence, and so some of the dates of divergence provided here are likely to be younger than the actual event. In this context it must be emphasized that the choice of an alternative date for any given dichotomy, say the point of divergence of sheep and ox, will not affect the fact that the two lineages arising from this point have fixed a different number of mutations from one another during the same period of time. It must also be appreciated that adjusting any given date of divergence with a view to smoothing the fluctuations in one lineage is liable to introduce more marked discrepancies elsewhere. For example, if the date of eutherian/metatherian divergence were judged to be earlier than here accepted (79 Ma), this would reduce the rate of evolution during the fast episode present in early eutherian history (79–68 Ma), but it would have the concomitant effect of lengthening the span of time during which the opossum and kangaroo

accumulated only 16 base substitutions. When two lineages arising from the same point differ in their rate of fixation of mutations it is self-evident that both cannot have the same rate as their common stem and so, apart from differences in rate between lineages there are also fluctuations in rate within lineages. However, differences in rate and fluctuations in rate would be expected to occur as the product of a stochastic process, and it is relevant to try to determine whether these differences are of such magnitude as to suggest non-random effects.

Over relatively short periods of time one would expect low numbers of mutations to be subject to high sampling error, but over a period of 79 Ma such fluctuations would tend to be smoothed out, and so the overall rate for each lineage forms the best basis for comparison, even though such a procedure might be masking the presence of real fluctuations along that lineage. These different overall rates are reflected in the total number of fixed mutations shown for each species (figure 31), the highest value being 34 for the lineage leading to the ox, and 11 for the tree shrew lineage in the same period of time. The average value for all the mammal lineages is 22.4 fixed mutations, which is equivalent to an average overall rate of one fixed mutation in 3.53 Ma. The highest overall rate (for the ox lineage) is one fixed mutation in 2.32 Ma, whereas the lowest rate (for the tree shrew lineage) is one fixed mutation in 7.18 Ma. The total number of mutations fixed in the different lineages during the past 79 Ma, with the exception of the ox and tree shrew, all fall within two standard deviations of the mean value, and these two exceptions fall within three standard deviations. We are, of course, aware that similar totals are to be expected for those species which have diverged relatively recently because a large part of their ancestry has followed an identical course.

Figure 32 is a plot of the number of reconstructed base substitutions along individual branches of cladogram 2 against the corresponding estimate of time. Thus, none of the points represent cumulative totals and this avoids the reinforcement of higher values on the graph by lower values which have been added together. It should be remembered that the values are derived from a cladogram, the pattern of which was determined prior to the assignment of base substitutions. Clearly, the number of base substitutions is correlated with time. For short periods of time, sampling error of the occurrence of mutational events is likely to be a significant influence. However, the estimates of the number of base substitutions over longer periods of time also show a considerable range of variation. Of particular interest are the values of 25 base substitutions in 52 Ma and 3 base substitutions in a comparable period of time of 59 Ma. Figure 33 is also based on cladogram 2, but in this case the higher values include earlier ancestry and are therefore additive. Once again there is a wide variation in the number of base substitutions at any point in time.

Without commitment as to the nature of the relation we have fitted a straight line so that our results can be compared with those of others. In the case of figure 32, a regression line was fitted through the origin by the least mean squares method ($y = 0.24\,x$), and the coefficient of correlation is 0.76. Similarly for figure 33, $y = 0.26\,x$ and the coefficient of correlation is 0.89; the higher degree of correlation is to be expected because the lower values are included within the higher values. Such a graph is broadly comparable to that of Langley & Fitch (1974) although their treatment is based on several different molecules and the values for each point in time are averaged to give a single value. In this context it is worth noting that Dickerson (1971) found an approximately linear relation between time and the number of amino acid differences (corrected for multiple changes at the same position). Others have accepted this result, but have been perplexed concerning the underlying mechanism because no plausible

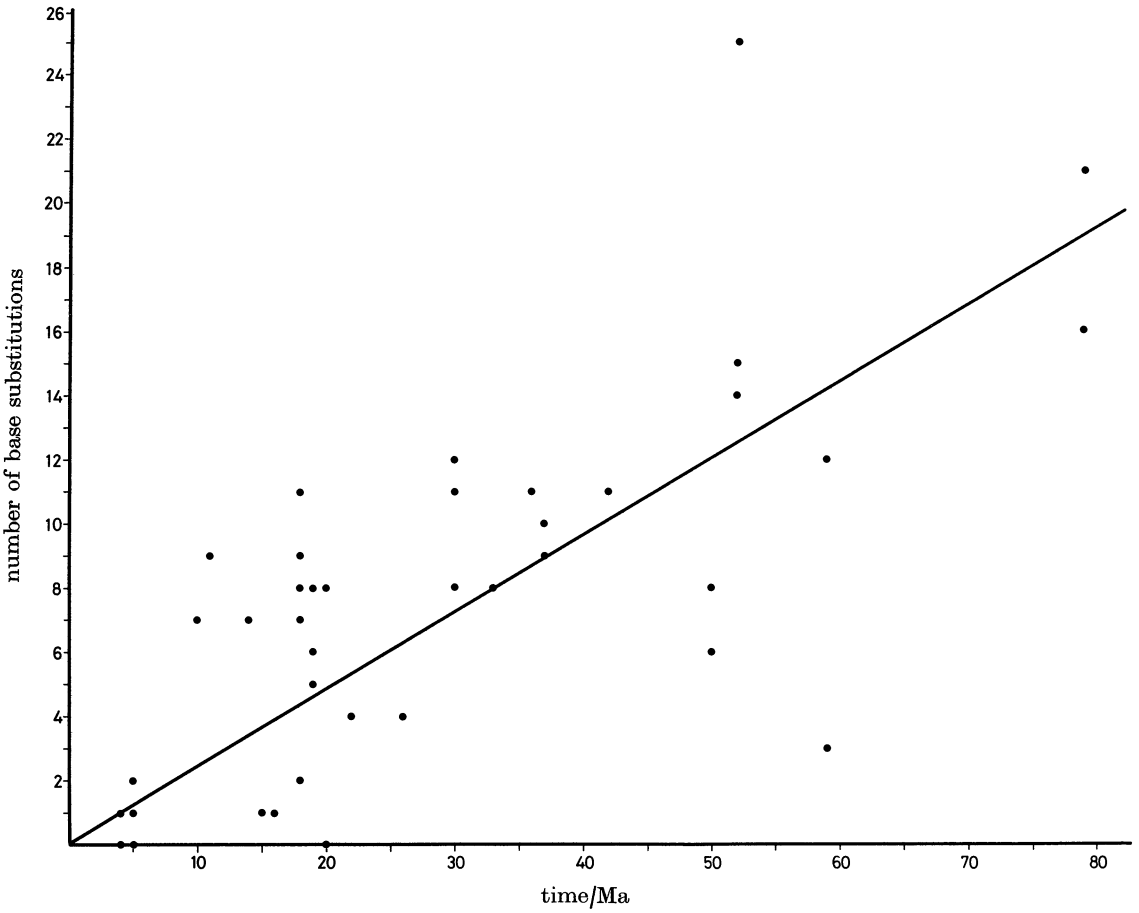


FIGURE 32. Number of reconstructed base substitutions for cladogram 2 (see figures 28 and 30) plotted against time. Each section of the lineages has been plotted separately so that no point represents a cumulative total.

process at the nucleotide level would be expected to provide such a linear relation (Holmquist 1972). Fitch & Markowitz (1970) have suggested that the expected decrease in amino acid differences might have been counterbalanced by the new availability of sites for change.

To return to figures 32 and 33, if a straight line is fitted to the scatter of points in each case then both methods of presentation indicate an average rate of fixation of one mutation in approximately 4 Ma. On the other hand, whether or not any simple relation exists, and if it does, whether it is best expressed by a straight line or by a curve, we feel that the most important point emerging from these two graphs is the wide range of variation in the estimates of the number of base substitutions at any given period of time. Whatever the cause of these differences may be, their presence must detract from the value of any hypothetical myoglobin clock for estimating dates of divergence. Bearing in mind that, even over a relatively long period of time (79 Ma), the fastest rate of fixation of mutations is three times the slowest rate, then we are inclined to discard the clock as unreliable.

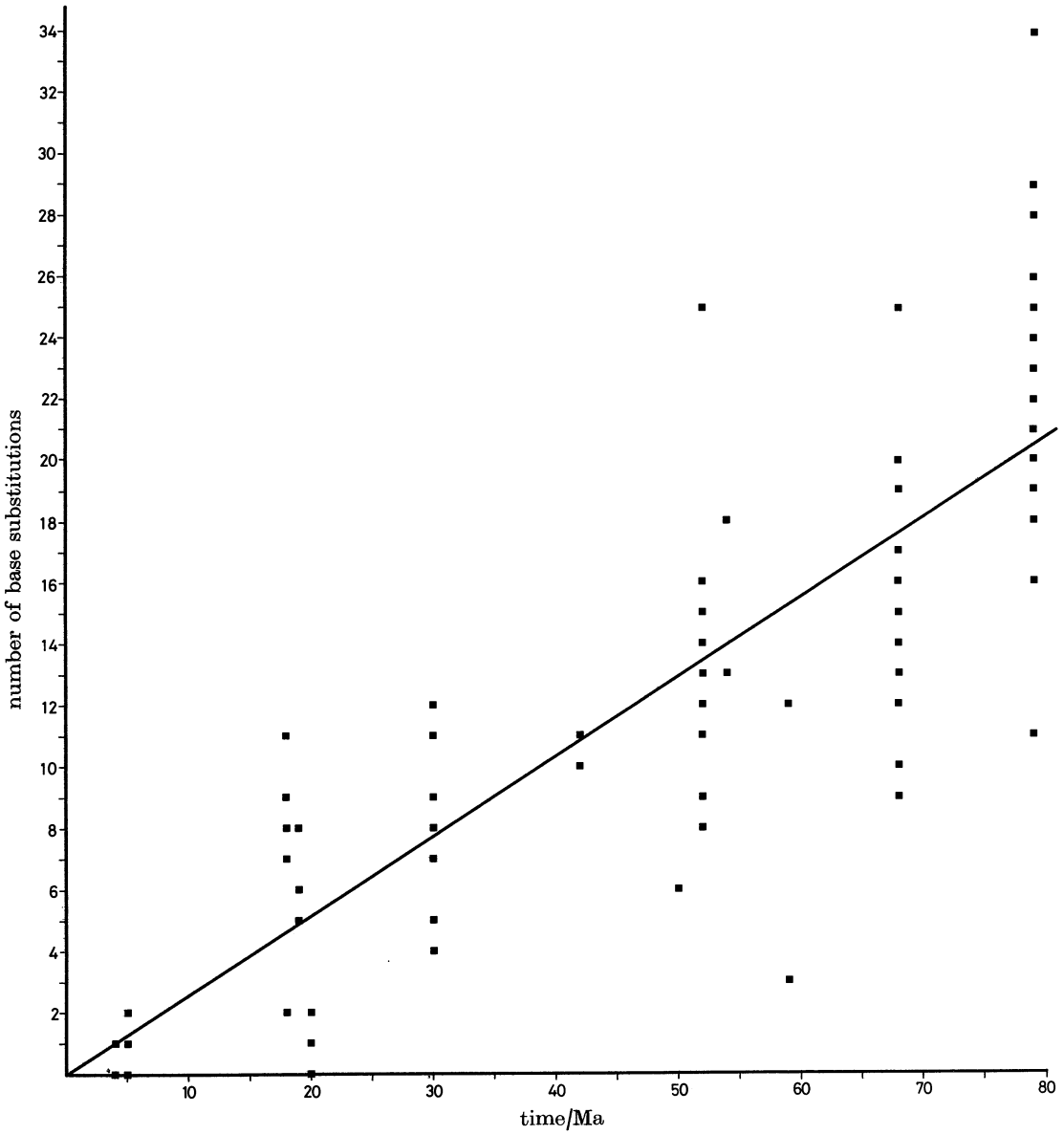


FIGURE 33. As figure 32, but smaller values have been added to produce some composite larger values.

Amount of change based on the matrix of amino acid differences, and two transformations of these values

Figure 34 shows the ranges of amino acid differences between pairs of aligned sequences (table 4) plotted against time of divergence. All comparisons have been included for each of the dated points of figure 30, and the mean values of amino acid differences are indicated.

It should be noted that, although the molecule sequenced for *Aplysia* is a myoglobin, lamprey and *Glycera* are represented by haemoglobins. With this difficulty in mind, figure 34 reflects the point made in connection with table 4 concerning the eventual saturation of the molecule by manifold changes. Due to multiple changes at the same site, back mutations and other hidden events, one might expect that the true number of amino acid substitutions would be

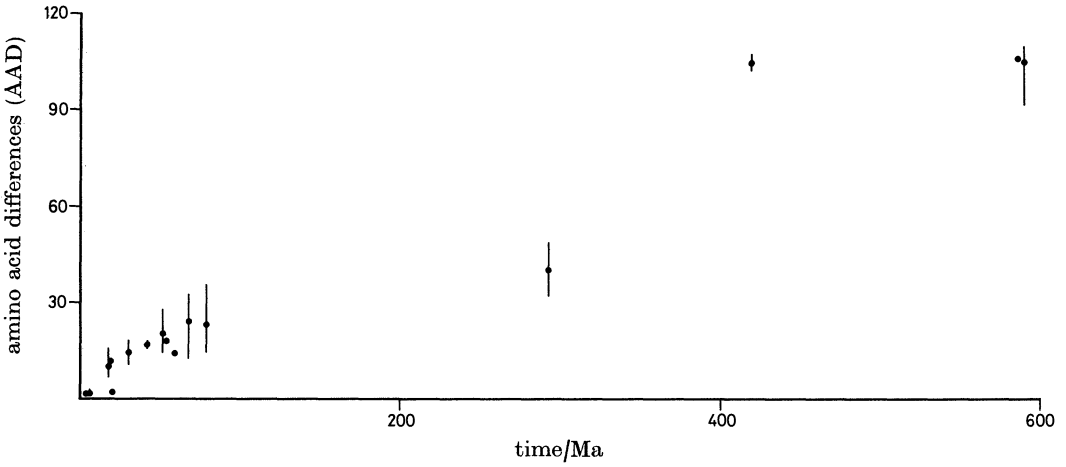


FIGURE 34. Amino acid distances between pairs of aligned sequences plotted against times of divergence.

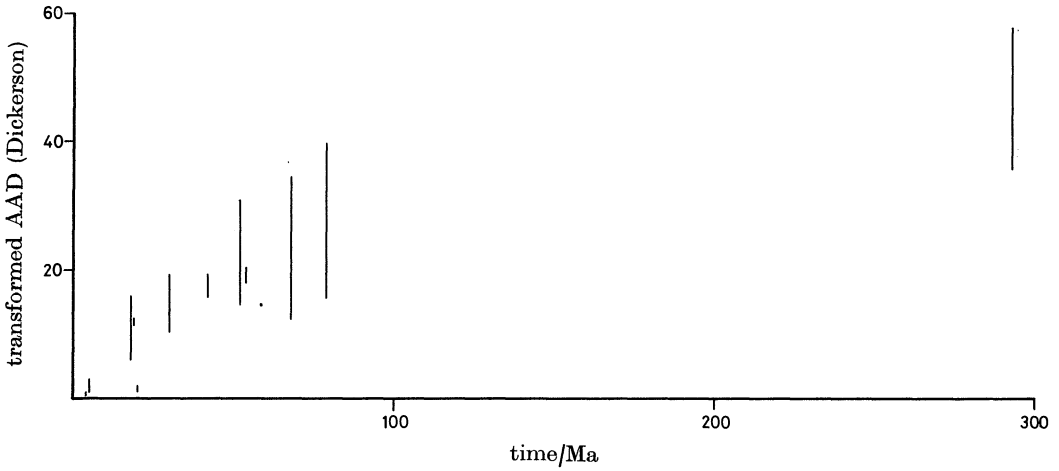


FIGURE 35. Transformed amino acid distances plotted against time up to 300 Ma. Ranges only are shown and are directly comparable with corresponding bars of figure 34. The transformation is that used by Dickerson (1971).

underestimated by the observed differences, and so the relationship might be expected to fall off with time (in the absence of compensating processes such as an increase in the number of sites available for change). Nevertheless, there is some suggestion that the amount of change apparent between birds and mammals (293 Ma) is smaller than might be expected. On the other hand it is likely that the apparent saturation of the molecule at 420 Ma is misleading because this date represents the divergence between lamprey and the other vertebrates rather than the date of divergence between myoglobin and haemoglobin. For the latter event Dickerson & Geis (1969) have used a date of about 930 Ma.

In an attempt to compensate for multiple changes at a site we have employed the transformation used by Dickerson (1971), and the results, covering the period up to 300 Ma, are shown in figure 35. This figure is, therefore, comparable to that of Dickerson & Geis (1969), which predicts a value for mammals compared with birds of about 39 *observed* amino acid differences on

average. The actual figure from our sequence comparisons is an average of 40.1, in rather striking agreement. A line from the origin through the mean of the range at 293 Ma would, however, be a very poor fit to the remaining data, although the span of time considered here is short enough to expect substantial departures from a line fitted in the manner of Dickerson, who was primarily considering the prediction of ancient evolutionary events. It should also be emphasized that Dickerson's (1971) expanded treatment of his data makes it clear that, with the exception of the myoglobin of the sperm whale, his globin line is based on haemoglobins.

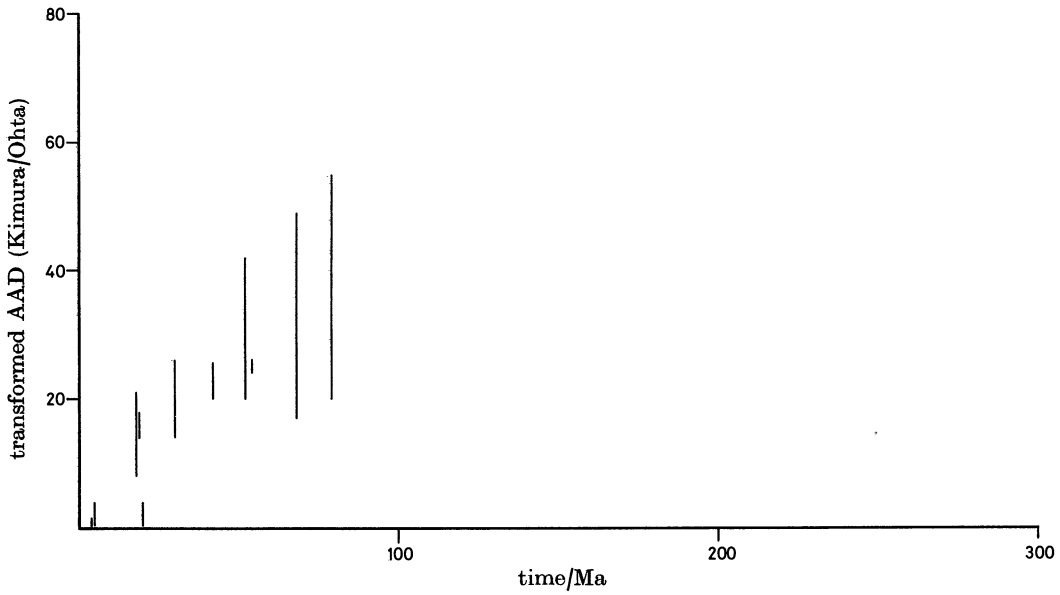


FIGURE 36. As figure 35 but using a transformation derived from Kimura & Ohta (1972).

Figure 36 covers the same dated points, but in this case the transformed values of amino acid differences are dreived from those of table 4 by the equations of Kimura & Ohta (1972), and represent the idealized stochastic model of Holmquist (1972*a*, *b*).

The transformations shown in figures 35 and 36 clearly increase the amounts of change used as distance measures, but do not otherwise substantially alter the appearance of the data.

Amount of change based on the extended stochastic model of Holmquist

When the aligned sequences (table 2) are used, and account taken of the nature of each substitution between pairs of myoglobins (single-, double- or triple-hit changes as a minimum) the full method of Holmquist (1972*a*, *b*; Holmquist, Cantor & Jukes 1972; Jukes & Holmquist 1972) may be applied. Deviations from the expected ratio of single- to double- to triple-hit changes according to the model may have profound effects on the derived values. Figure 37 illustrates this effectively for the chicken/penguin value (represented by a triangle on the figure). It may be seen that rates during the past 79 Ma still appear to have been greater than the mean rate measured over 293 Ma.

Jukes & Holmquist (1972) suggested that, whereas amino acid differences, minimum base distances and even random evolutionary hits were not suitable parameters for a molecular 'clock' their parameter T_2 might be. T_2 is the number of variable sites in a pair-wise compari-

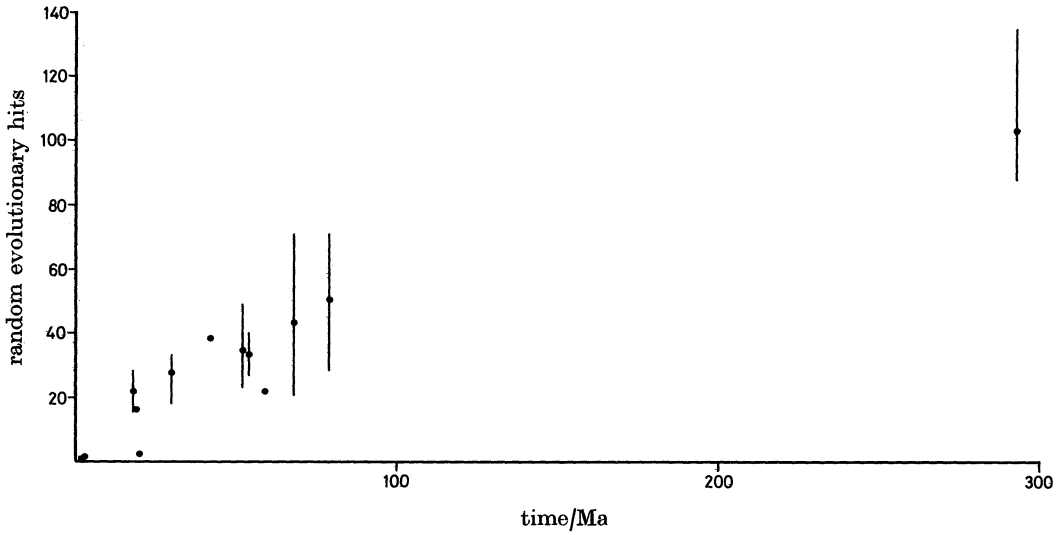


FIGURE 37. Random evolutionary hit values of Holmquist (described in the text) plotted against time. Mean values are indicated by solid circles and the solid triangle represents the value derived from the chicken/penguin comparison.

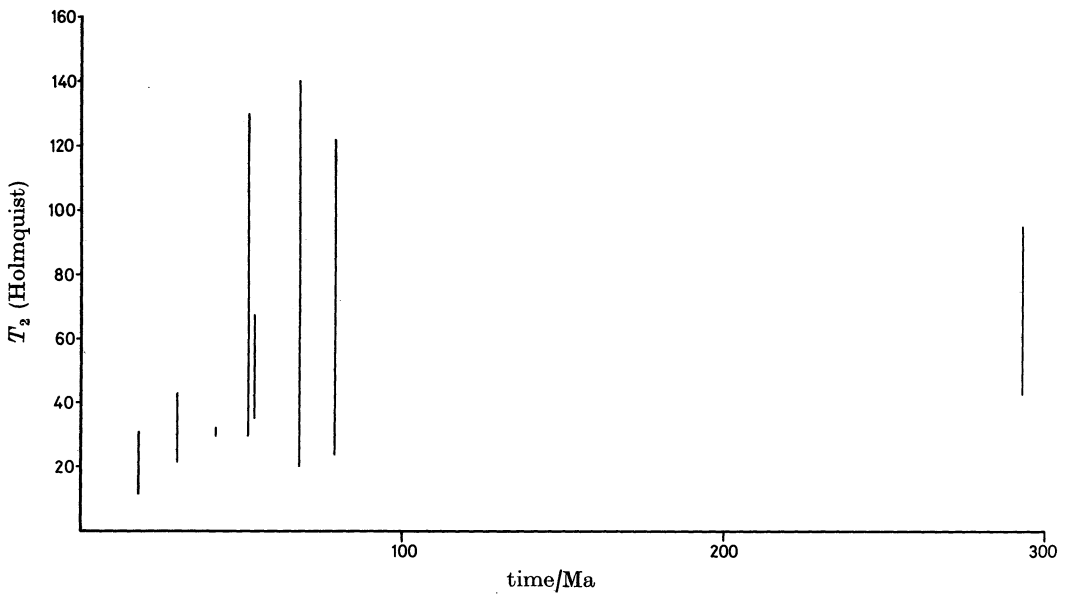


FIGURE 38. T_2 values of Jukes & Holmquist (1972) plotted against time. Ranges only are shown. Comparisons which would result in statistically invalid estimates have been omitted (see Holmquist *et al.* 1976 for details of such cases).

son of sequences and is obtained by application of the algorithm in the appendix to the paper by Jukes & Holmquist (1972). They further suggest that the parameter μ_2 (hits per variable codon) is species-dependent (or lineage-dependent), whereas T_2 is possibly a better indicator of the timing of phylogenetic events – at least for cytochrome c. T_2 is shown plotted against the times of divergence in figure 38. Our results indicate a strong element of dependence for both μ_2 and T_2 on the identity of the animals considered. Often, however, there is a reciprocal

relationship between μ_2 and T_2 resulting in similar values as the product (random evolutionary hits) of a low μ_2 and high T_2, and vice versa. For a comparison of sperm whale against primates and horse against primates, both carried out from the 68 Ma point, the values of T_2 for sperm whale are generally higher than for horse, although the ranges overlap. However, the values of μ_2 for the sperm whale are relatively low and those for the horse relatively high, resulting in closely similar numbers of random evolutionary hits for both animals in comparison with the thirteen primates. We note that the calculations are extremely sensitive to small differences in the ratio of single- to double- (and triple-) hit substitutions. Indeed, this is part of the reasoning behind the procedures. As a consequence, from a single comparison-species, apparently closely related animals may have quite different μ_2 and T_2 values. From horse, the baboon has $\mu_2 = 0.40$ and $T_2 = 67.8$. Macaque, however, has $\mu_2 = 0.83$ and $T_2 = 38.7$. The corresponding REH values are therefore very similar: 27.1 and 32.1. The differences are accounted for by the nature of the residues at the hypervariable position 66 in the three animals. Baboon has Ala, macaque Val and the horse Thr. Val–Thr needs at least two hits, whereas Ala–Thr needs only one. Hence, macaque has one extra double-hit substitution than baboon has in comparison with the horse, and this results in a radically different μ_2 and T_2.

We are, therefore, inclined to reject the possible value of T_2 as proposed by Jukes & Holmquist, at least in the case of myoglobin where T_2 appears to be a poor indicator of relative times of divergence, and to have an unacceptably large range for comparisons from a point in time. However, if the situation encountered above is due to stochastic sampling error and is not a reflection of real restriction in the number of variable sites it is likely to be less influential in the calculations when the method is applied over longer periods of time (as was the case with the original example of Jukes & Holmquist). The value of their proposals needs to be investigated over a larger range of more diverse data, and we should emphasize that the stochastic model of Holmquist and his co-workers has been conceptually most helpful to us in thinking about the causes and effects of molecular evolution. Further comments concerning the stochastic model and the evolution of myoglobin have recently appeared (Moore *et al.* 1976; Holmquist *et al.* 1976) and we are indebted to Dr Holmquist for his communications concerning some details of interpretation.

8. Discussion

In §7 we noted that despite a number of procedures which attempted to correct observed values for unobserved events at the nucleotide level, the average rate of evolution of both mammals and birds during the past 79 Ma appears to have been greater than that during the previous 214 Ma. Using an alternative correction for unobserved nucleotide substitutions on information derived from seven molecules, and also using a different set of dates spanning 125 Ma, Fitch & Langley (1977) have claimed a linear relation between elapsed time and nucleotide substitutions. Applying their dates to our data would increase the difference between these conflicting results.

Considerable attention has been directed towards the claim that molecular evolution may be more readily understood in terms of the fixation of predominantly neutral, or nearly neutral, mutations (Kimura 1968; Ohta & Kimura 1971; Kimura & Ohta 1972; Ohta 1974). However, it cannot be too clearly emphasized that the original argument concerning the predominance of neutral change turns partly on the amount of change observed in molecules when this is

interpreted within a classical genetic framework (taking account also of the genetic load observed in present-day populations), partly on the regularity of this change, and also partly on its nature compared to expected events on a random basis (Ohta & Kimura 1971).

There are usually parts of a molecule which can be demonstrated to be invariant or very conservative because they have a known (or suspected) chemical function. For the remaining variable residues there are conflicting hypotheses concerning the nature of change, such as that of Jukes & Holmquist (1972), noted in the previous section, where a Poisson parameter is applied over a restricted part of the molecule (claimed to increase in size with time). However, Fitch (1971) viewed cytochrome c as having some 4–10 residues, not necessarily adjacent, each of which is able to accept substitutions at any one time, although the composition of this suite of residues may vary as previously labile residues are hit and change their identity.

Allowing that there may be neutral change, how does one establish its basic rate? Ohta (1974) has implied that the fibrinopeptides A and B, and proinsulin, may be accepting change at about this rate. It seems likely that the well-known but relatively little understood 'hot spots' of, for example, the T4 phage (Benzer 1955) may result from special structural features of nucleic acids (Whitehouse 1973). This suggests the possibility that there could be a 'fast' neutrality and a 'slow' (or perhaps more correctly 'normal') neutrality.

Since there is no agreement about the observable effects of predominantly neutral change, it is not sufficient to point out that change may be consistent with a random model (with fixation of neutral mutations by genetic drift) when it may also be consistent with rapid selective processes (Smith 1968). Before one can separate the relative contribution by neutral and selectively significant mutations, one has to take into account molecular function. A promising start has been made in this direction by Goodman, Moore & Matsuda (1975).

Until further progress has been made in understanding molecular function, the likelihood of nearly neutral mutations is difficult to evaluate. Insufficient is known at present about the energetics of protein synthesis to fully assess the effects of nucleotide substitutions at this level. At a higher level of organization, substitutions in the external positions of the haemoglobin molecule may well change respiratory function less than the internal substitutions (Perutz & Lehmann 1968), yet they are more likely to affect the immune interactions of the molecule.

We are left with the conclusion that both 'neutralists' and 'selectionists' would agree that invariant sites represent selective constraints upon possible change, but that a 'neutralist' would regard a substantial rate of change in the variable regions of the molecule as indicating the absence of selective constraints, whereas a 'selectionist' would interpret such a rate as indicating adaptive evolution. Thus it is difficult, at present, to determine whether the apparent increase in the rate of change in both mammals and birds should be interpreted as an increase or a decrease in the importance of the physiological role of myoglobin. Nevertheless, this same apparent change in the rate of nucleotide substitution through time detracts from arguments for neutrality based on regularity of change.

In §7 we have drawn attention to the wide range in rates of evolution between different lineages. Ignoring fluctuations in rate, we demonstrated that even over a long period (79 Ma), the number of base substitutions accepted by the lineage leading to the ox is about three times that in the lineage leading to the tree shrew. Under the present circumstances, the relation between change and time must be assessed empirically and evaluated in such terms as the construction of phylogenies, at least in comparison with the order of events and, less satisfactorily at present, their absolute time.

Over the period of time of about 293 Ma, it is evident that the correlation of various distance measures with time is not sufficiently reliable to establish unknown dates of divergence with any degree of confidence, or even to establish the relative order of divergence of the major groups. Even if the data are sufficiently constant for the 'neutralist' geneticist, they do not satisfy the needs of the phylogeneticist for dating dichotomies.

In attempting to date phylogeny from amount of change, the opossum again provides a striking anomaly when the evidence from myoglobin and the haemoglobins is compared (Stenzel 1974).

TABLE 15. ANALYSIS OF KNOWN VARIATION

	no.	%	expected	observed	observed / expected
haem contact residues	22	14.38	40	12	0.30
internal residues	21	13.72	38	21	0.53
external residues	110	71.90	200	245	1.23
total	153	100	278	278	
external residues involved in salt bridges	13	11.82	24	10	0.42
other external residues	97	88.18	176	190	1.08
total	110	100	200	200	

The figures of table 15 strongly suggest that change in myoglobin has not been at random over the molecule. This is not surprising, given the known functional constraints and the constraints of the genetic code which impose a limited suite of changes on any residue taken as starting point. It is to be expected that both of these phenomena will contribute to the likelihood of parallel change in different lineages. The analysis of cladogram 2 (figure 3) reveals that parallel evolution is commonplace at the amino acid level. The nature of these parallel changes is analysed in table 16. Of the 278 reconstructed mutations, 139 changes (50 %) are found to be in parallel. Parallel events occurred at 38 out of the 83 positions which have accepted change. The constraints of the genetic code are relevant to the incidence of back mutations which are also shown in table 16. At four positions the back mutation was also a parallel event.

It should be recalled that in constructing the equally parsimonious cladogram 5 (figure 7), whenever one of the parallel mutations in cladogram 2 could be alternatively interpreted in terms of back mutations, then the complementary solution was adopted. Accordingly, the total number of back mutations is increased from 20 to 34, but as indicated in table 17 this makes little difference to the frequency of parallel events. In this case, 124 out of 278 mutations are in parallel (44.6 %). (Footnote on page 68 accounts for anomalous total.)

We have used the term 'parallel' with respect to events at the same site in different lineages, but such a restricted use could obscure important functional parallelism, where identical events might occur at different sites and confer similar properties on the molecule. In the terminology of morphological evolution such changes might be described as analogous rather than homologous. It is of interest to note (see table 3) that the myoglobins of diving mammals (cetaceans and pinnipeds) and penguin contain 3–5 residues of arginine, whereas those of land mammals contain only 1 or 2 residues of this amino acid. The higher number of arginine in the diving

TABLE 16. PARALLEL AND BACK MUTATIONS IN CLADOGRAM 2

(Symbols as in figure 3.)

PARALLEL MUTATIONS

Sequential number	9	13	13	19	21	21	22	22	27	28	34	35	35	41	51	52	54	56	57	66	66	66	66	74	81
Helix Key	A7	A11	A11	AB1	B2	B2	B3	B3	B8	B9	B15	B16	B16	C6	D1	D2	D4	D6	D7	E9	E9	E9	E9	E17	EF4
Ancestral amino acids	Leu	Val	Val	Ala	Ile	Ile	Ala	Ala	Glu	Val	Lys	Gly	Gly	Glu	Thr	Glu	Glu	Lys	Ala	Asn	Asn	Asn	Asn	Ala	His
Parallel mutation	Gln	Ile	Ala	Thr	Leu	Val	Gly	Pro	Asp	Ile	Thr	Ser	Asp	Asp	Ser	Pro	Asp	Arg	Gly	Val	Thr	Ala	Asn	Ala	Gln
No. of parallel mutations	2	5	2	3	4	3	2	3	3	2	2	3	2	2	4	2	3	2	2	6	4	2	2	2	3
Coefficient of difference	1.00	0.05	0.40	0.41	0	0.03	0.10	0.41	0.03	0.05	0.21	0.32	0.54	0.03	0.03	1.00	0.03	0.05	0.10	1.00	0.20	0.61	0	0	0.20
Other amino acids found	Ser	Ala	Ile	Ser	Val	Leu	Pro	Gly	Ala		His	Asp	Ser			Ala			Arg	Thr	Val	Thr	Thr	Asn	Asn
at the same position	Ala	Met	Met		Glu	Glu	Ser	Ser				Ala	Ala						Ser	Ala	Ala	Val	Val	Gly	
							Val	Val				Asn	Asn							Asp	Asp	Asp	Asp		
																				Lys	Lys	Lys	Lys		
																				Ile	Ile	Ile	Ile		
																				Gln	Gln	Gln	Gln		
																							Ala		

Sequential number	86	95	99	101	106	109	112	113	116	117	120	120	121	121	122	122	124	129	132	132	140	142	144	149	152
Helix Key	F1	FG1	FG5	G2	G7	G10	G13	G14	G17	G18	GH2	GH2	GH3	GH3	GH4	GH4	H1	H6	H9	H9	H17	H19	H21	H26	HC3
Ancestral amino acids	Leu	Thr	Ile	Val	Phe	Glu	Ile	Gln	Gln	Ser	Ala	Ala	Gly	Gly	Asp	Asp	Gly	Ala	Lys	Lys	Asn	Met	Ala	Phe	Gln
Parallel mutation	Ile	Asn	Val	Ile	Leu	Asp	Val	His	His	Asn	Pro	Ser	Ser	Ala	Glu	Asn	Ala	Gly	Ser	Lys	Lys	Met	Ser	Leu	His
No. of parallel mutations	3	2	3	4	2	5	2	2	4	2	2	2	2	4	3	5	2	3	4	2	2	2	2	2	3
Coefficient of difference	0	0.20	0.03	0.05	0.08	0.03	0.03	0.20	0.20	0.20	0.41	0.40	0.32	0.10	0.03	0	0.10	0.10	0.22	0	0.03	0	0.40	0.08	0.20
Other amino acids found	Val					Gly	Ala	Lys	Lys	Lys	Ser	Pro	Ala	Ser	Asn	Glu	His	Glu	Asn	Ser	His	Ala	Thr		
at the same position							Met		Ala	Ala					Gln	Gln			Thr	Thr	Asp	Ile	Glu		
										Glu									Gly	Gly					
										Gln										Asn					

BACK MUTATIONS

Sequential number	21	23	35	51	66	74	81	112	113	120	121	122	129	132	142	152
Helix Key	B2	B4	B16	D1	E9	E17	EF4	G13	G14	GH2	GH3	GH4	H6	H9	H19	HC3
Intermediate amino acids	Val	Ser	Ser	Ser	Asp	Gly	Gln	Val	His	Pro	Ala	Glu	Gly	Asn	Ile	His
Back mutation	Ile	Gly	Gly	Thr	Asn	Ala	His	Ile	Gln	Ala	Gly	Asp	Ala	Lys	Met	Gln
No. of back mutation(s)	1	1	1	1	2	2	1	1	1	1	1	1	1	2	2	1
Coefficient of difference	0.05	0.30	0.30	0.03	0	0.10	0.20	0.05	0.20	0.40	0.10	0.03	0.10	0.03	0.05	0.20
Other amino acids found	Leu		Ala		Val	Asn	Asn	Ala	Lys	Ser	Ser	Gln	Glu	Ser	Ala	
at the same position	Glu		Asn		Ala	Gln		Met				Asn		Thr		
			Asp		Thr									Gly		
					Ile											
					Gln											

species might be due to chance. However, it may be that an increase in arginine at the surface of the molecule confers some specific adaptation in those species which depend heavily upon myoglobin as a long-term oxygen store. In most cases arginine substitutes for lysine, the exception being arginine for alanine at position 57 in the pinnipeds. The main difference between arginine and lysine concerns the pK of the guanidinium group of the first and the ϵ-amino group of the second. Consequently, if this increase in arginine has an adaptive role, this may be related to the differences in the isoelectric point of the protein, which would confer some advantage on the myoglobin, enabling it to be stored at high concentration in a milieu temporarily deprived of oxygen. This situation is remarkable in the light of Ohta & Kimura's (1971) observation that changes from other residues to arginine are less frequent than would be expected on a random basis.

Returning to the consideration of parallel events at the same site, another observation derived from the reconstruction of our cladograms is that there has been parallel acquisition of

TABLE 17. PARALLEL AND BACK MUTATIONS IN CLADOGRAM 5
(Symbols as in figure 7.)

PARALLEL MUTATIONS

Sequential number	13	13	19	21	21	22	22	27	28	34	35	35	51	52	54	56	66	66	66	74	81	86	95
Helix Key	A11	A11	AB1	B2	B2	B3	B3	B8	B9	B15	B16	B16	D1	D2	D4	D6	E9	E9	E9	E17	EF4	F1	FG1
Ancestral amino acids	Ile	Ile	Ala	Ile	Ile	Ala	Ala	Glu	Val	Lys	Gly	Gly	Thr	Glu	Asp	Lys	Ala	Ala	Ala	Gln	His	Leu	Thr
Parallel mutations	Ile	Ala	Thr	Leu	Ile	Pro	Gly	Asp	Ile	Thr	Ser	Asp	Thr	Pro	Asp	Arg	Val	Thr	Asn	Ala	Gln	Ile	Asn
No. of parallel mutations	4	2	2	4	2	3	2	3	2	2	3	2	2	2	2	2	6	3	2	3	4	3	2
Coefficient of difference	0	0.43	0.41	0	0	0.41	0.10	0.03	0.05	0.21	0.32	0.54	0	1.00	0	0.05	0.41	0.41	0.62	0.61	0.20	0	0.20
Other amino acids found	Val	Val	Ser	Val	Leu	Ser	Pro	Ala		His	Ala	Ser	Ser	Ala	Glu		Thr	Val	Val	Gly	Asn	Val	
at the same position	Ala	Met		Glu	Val	Gly	Ser				Asn	Ala					Asn	Asn	Thr	Asn			
	Met				Glu						Asp	Asn					Asp	Asp	Asp				
																	Ile	Ile	Ile				

Sequential number	99	101	106	109	112	113	116	117	120	120	121	121	122	122	124	129	132	132	140	142	149	152
Helix Key	FG5	G2	G7	G10	G13	G14	G17	G18	GH2	GH2	GH3	GH3	GH4	GH4	H1	H6	H9	H9	H17	H19	H26	HC3
Ancestral amino acids	Ile	Val	Phe	Glu	Ile	Lys	Ala	Ser	Ala	Ala	Gly	Gly	Asp	Asp	Gly	Ala	Lys	Lys	Asn	Met	Phe	Gln
Parallel mutations	Val	Ile	Leu	Asp	Val	His	His	Asn	Pro	Ser	Ala	Ser	Asn	Glu	Ala	Gly	Lys	Ser	Lys	Met	Leu	His
No. of parallel mutations	3	3	2	5	3	2	4	2	2	2	4	2	4	3	2	3	3	4	2	2	2	3
Coefficient of difference	0.03	0.05	0.08	0.03	0.03	0.20	0.47	0.20	0.41	0.40	0.10	0.32	0	0.03	0.10	0.10	0	0.22	0.03	0	0.08	0.20	
Other amino acids found				Gly	Ala	Gln	Gln	Lys	Ser	Pro	Ser	Ala	Glu	Asn	His	His	Ser	Asn	His	Ile			
at the same position					Met		Lys	Ala									Asn	Gly	Asp	Ala			
								Glu									Gly						
								Gln															

BACK MUTATIONS

Sequential number	9	13	19	21	23	35	41	51	51	54	57	66	66	66	66	81	101	113	122	129	132	142	144
Helix Key	A7	A11	AB1	B2	B4	B16	C6	D1	D1	D4	D7	E9	E9	E9	E9	EF4	G2	G14	CH4	H6	H9	H19	H21
Intermediate amino acids	Leu	Val	Thr	Val	Ser	Ser	Glu	Ser	Thr	Glu	Ala	Asp	Asn	Asp	Ala	Gln	Ile	His	Asn	Gly	Asn	Ile	Ala
Back mutation	Gln	Ile	Ala	Ile	Gly	Gly	Asp	Thr	Ser	Asp	Gly	Asn	Thr	Ala	Thr	His	Val	Gln	Asp	Ala	Lys	Met	Ser
No. of Back mutations	1	4	1	2	1	1	1	2	1	2	1	2	1	1	2	1	1	1	1	1	3	2	1
Coefficient of difference	1.00	0.05	0.40	0.05	0.30	0.30	0.03	0.03	0.03	0.03	0.10	0	0.20	0.61	0.41	0.20	0.03	0.20	0	0.10	0.03	0.05	0.40
Other amino acids found	Ala	Ala	Ser	Leu		Ala					Ser	Val	Val	Val	Val	Asn			Gln	Glu	Gly	Ala	Thr
at same position	Ser	Met		Glu		Asn					Arg	Ala	Ala	Ala	Asp				Glu		Ser		Glu
	Leu					Asp						Ser	Ser	Ser	Ser						Thr		
												Thr	Asp	Thr	Asn								
												Ile	Ile	Ile	Ile								
												Gln	Gln	Gln	Gln								

some residues between the cetaceans and pinnipeds: 54 Asp (harbour seal and dolphin); 83 Asp (sea lion and dolphin); 121 Ala (harbour seal, dolphin and porpoise); 122 Glu (harbour seal and dolphin); 152 His (harbour seal, dolphin and porpoise). Whether or not the increase in arginine in the myoglobin of the diving animals and the residues acquired in parallel between cetaceans and pinnipeds are significant adaptations to the aquatic environment awaits further investigation of the solubility, stability and oxygen-binding properties of the myoglobin molecule.

The observations in §5 indicate that there are a limited number of ways of remaining a functional myoglobin molecule. Functional morphology is paramount whether we are considering the anatomical or the molecular level. The limitations of mechanics at the anatomical level, and chemical limitations at the molecular level, place constraints on the pattern of evolution. The laws of physics and chemistry have to be obeyed and so we find that, superimposed on divergent change, there is a high degree of parallelism due partly to these constraints and surely partly to adaptive evolution. Regardless of the causes of parallel evolution we must emphasize that this phenomenon has contributed to unexpected similarities between myoglobins, such as those of horse and sportive lemur, and those of opossum and the anthropoids.

Such similarities have contributed to the difficulties of phylogenetic reconstruction in a manner already familiar to comparative anatomists.

We are grateful to many colleagues who have given help and advice during the course of this work. Those concerned with the earlier stages were acknowledged in our previous publications, and those who contributed exclusively to this report are mentioned where appropriate in the text. Part of this work was supported by the Muscular Dystrophy Group of Great Britain.

References

Air, G. M. & Thompson, E. O. P. 1971 Studies on marsupial proteins. IV. Amino acid sequence of myoglobin from the red kangaroo *Megaleia rufa*. *Aust. J. biol. Sci.* **24**, 75–95.

Allan, N., Beale, D., Irvine, D. & Lehmann, H. 1965 Three haemoglobins K: Woolwich, an abnormal, Cameroon and Ibadan, two unusual variants of human haemoglobin A. *Nature, Lond.* **208**, 658–661.

Andrews, P. 1971 *Ramapithecus wickeri* Mandible from Fort Ternan, Kenya. *Nature, Lond.* **231**, 192–194.

Åstrand, I., Åstrand, P., Christensen, E. H. & Hedman, R. 1960*a* Intermittent muscular work. *Acta physiol. scand.* **48**, 448–453.

Åstrand, I., Åstrand, P., Christensen, E. H. & Hedman, R. 1960*b* Myohemoglobin as an oxygen-store in man. *Acta physiol. scand.* **48**, 454–460.

Atassi, M. Z. 1973 Antigenic structures of proteins inferred from myoglobin as the first protein whose antigenic structure has reached completion. Pp. 118–135 in *Specific receptors of antibodies, antigens and cells*. (3rd Int. Convoc. Immunol. Buffalo, N.Y. 1972.) Basel: Karger.

Atassi, M. Z. & Saplin, B. J. 1968 Immunochemistry of sperm whale myoglobin. I. Interaction of some tryptic peptides and of peptides containing all the respective regions of the antigen. *Biochemistry, N.Y.* **7**, 688–698.

Atassi, M. Z. & Saplin, B. J. 1971 Immunochemistry of sperm whale myoglobin. X. Regions responsible for immunochemical cross-reaction with finback whale myoglobin. Some general conclusions concerning immunochemical cross-reaction of proteins. *Biochemistry, N.Y.* **10**, 4740–4747.

Atassi, M. Z., Tarlowski, D. P. & Paull, J. H. 1970 Immunochemistry of sperm whale myoglobin. VII. Correlation of immunological cross-reaction of eight myoglobins with structural similarity and its dependence on conformation. *Biochim. biophys. Acta* **221**, 623–635.

Azzaroli, A. 1966 Pleistocene and living horses of the Old World. *Paleontogr. ital.* **61**, 1–15.

Beliajeva, E. I., Trofimov, B. A. & Reshetov, V. Ju. 1974 In *Fauna and biostratigraphy of Mesozoic and Cainozoic of Mongolia. Paleontological Expedition* (ed. N. N. Kramarenko). *Trudy mosk. vyp.* **1**, 19–45.

Benzer, S. 1955 Fine structure of a genetic region in bacteriophage. *Proc. natn. Acad. Sci. U.S.A.* **41**, 344–354.

Berggren, W. A. & Van Couvering, J. A. 1974 The late Neogene. Biostratigraphy, geochronology and paleoclimatology of the last 15 million years in marine and continental sequences. *Paleogeogr., Paleoclimat., Paleoecol.* **16**, 1–216.

Beyer, W. A., Stein, M. L., Smith, T. F. & Ulam, S. M. 1974 A molecular sequence metric and evolutionary trees. *Math. biosci.* **19**, 9–25.

Bishop, W. W., Miller, J. A. & Fitch, F. J. 1969 New potassium–argon age determinations relevant to the Miocene fossil mammal sequence in East Africa. *Am. J. Sci.* **267**, 669–699.

Boulton, F. E., Huntsman, R. G., Lorkin, P. A. & Lehmann, H. 1969 Abnormal human myoglobin: 53 (D4) glutamic acid-lysine. *Nature, Lond.* **223**, 832–833.

Boulton, F. E., Huntsman, R. G., Romero-Herrera, A. E., Lorkin, P. A. & Lehmann, H. 1971*a* The third variant of human myoglobin showing an unusual amino acid substitution: 138 (H16) arginine-tryptophan. *Biochim. biophys. Acta* **229**, 716–719.

Boulton, F. E., Huntsman, R. G., Romero-Herrera, A. E., Lorkin, P. A. & Lehmann, H. 1971*b* A human myoglobin variant 133 (H10) Lysine-asparagine. *Biochim. biophys. Acta* **229**, 871–876.

Boulton, F. E., Huntsman. R. G., Yawson, G. I., Romero-Herrera, A. E., Lorkin, P. A. & Lehmann, H. 1971*c* The second variant of human myoglobin; 138 (H16) arginine-glutamine. *Br. J. Haemat.* **20**, 69–74.

Boyer, S. H., Crosby, E. F., Noyes, A. N., Fuller, G. F., Leslie, S. E., Donaldson, L. J., Vrablik, G. R., Schaefer, E. W. & Thurmon, T. F. 1971 Primate hemoglobins: Some sequences and some proposals concerning the character of evolution and mutation. *Biochem. Genet.* **5**, 405–448.

Boyer, S. H., Fainer, D. C. & Naughton, M. A. 1963 Myoglobin: inherited structural variation in man. *Science, N.Y.* **140**, 1228–1231.

Bradshaw, R. A. & Gurd, F. R. N. 1969 Comparison of myoglobins from harbor seal, porpoise and sperm whale. V. The complete amino acid sequences of harbor seal and porpoise myoglobins. *J. biol. Chem.* **244**, 2167–2181.

Bretscher, P. A. & Cohn, M. 1968 Minimal model for the mechanism of antibody induction and paralysis by antigen. *Nature, Lond.* **220**, 444–448.

Butler, P. M. 1972 The problem of Insectivore classification. Pp. 253–265 in *Studies in vertebrate evolution* (eds K. A. Joysey & T. S. Kemp). Edinburgh: Oliver & Boyd.

Cavalli-Sforza, L. L. & Edwards, A. W. F. 1967 Phylogenetic analysis: models and estimating procedures. *Evolution, Lancaster, Pa.* **21**, 550–570.

Chauvet, J. P. & Acher, R. 1970 The β-chain of frog hemoglobin (*Rana esculenta*): the complete amino acid sequence. *FEBS letters* **10**, 136–138.

Chou, P. Y. & Fasman, G. D. 1974*a* Conformational parameters for amino acids in helical β-sheet and random coil regions calculated from proteins. *Biochemistry, N.Y.* **13**, 211–221.

Chou, P. Y. & Fasman, G. D. 1974*b* Prediction of protein conformation. *Biochemistry, N.Y.* **13**, 222–245.

Clemens, W. A., Jr. 1971 Mammalian evolution in the Cretaceous. Pp. 165–180 in *Early mammals* (eds D. M. Kermack & K. A. Kermack). *Zool. J. Linn. Soc.* **50**, Suppl. 1.

Clutton-Brock, J., Corbet, G. B. & Hills, M. 1976 A review of the Family Canidae, with a classification by numerical methods. *Bull. Br. Mus. nat. Hist.* (*Zool.*) **29**, 1–199.

Cook, C. N. & Hewett-Emmett, D. 1974 The uses of protein sequence data in systematics. Pp. 937–958 in *Prosimian Biology* (eds R. D. Martin, G. A. Doyle & A. C. Walker). London: Duckworth.

Cruickshank, A. R. I. 1974 News and views: 21st Symposium on Vertebrate Palaeontology and Comparative Anatomy. *South Afr. J. Sci.* **70**, 100–102.

Darbre, P. D., Romero-Herrera, A. E. & Lehmann, H. 1975 Comparison of the myoglobin of the zebra (*Equus burchelli*) with that of the horse (*Equus caballus*). *Biochim. biophys. Acta* **393**, 201–204.

Dautrevaux, M., Boulanger, Y., Han, K. & Biserte, G. 1969 Structure covalente de la myoglobine de cheval. *Eur. J. Biochem.* **11**, 267–277.

Dayhoff, M. O. 1972 *Atlas of protein sequence and structure.* Silver Spring, Maryland: National Biomedical Research Foundation.

Deconinck, M., Peiffer, S., Depreter, J., Paul, C., Schnek, A. G. & Leonis, J. 1975 The primary sequence of chicken myoglobin (*Gallus gallus*). *Biochim. biophys. Acta* **386**, 567–575.

Delson, E. & Andrews, P. 1975 Evolution and interrelationships of the catarrhine primates. Pp. 405–446 in *Phylogeny of the Primates* (eds W. P. Luckett & F. S. Szalay). New York: Plenum.

Dickerson, R. E. & Geis, I. 1969 *The structure and action of proteins.* Menlo Park, California: Benjamin.

Dickerson, R. E. 1971 The structure of cytochrome c and the rates of molecular evolution. *J. molec. Evol.* **1**, 26–45.

Doolittle, R. F., Wooding, G. L., Lin, Y. & Riley, M. 1971 Hominoid evolution as judged by fibrinopeptide structure. *J. molec. Evol.* **1**, 74–83.

Ducastaing, S. 1973 Structure primaire de la Myoglobine de Blaireau. Thèse Doct. Sciences, Université de Bordeaux.

Dumur, V., Dautrevaux, M. & Han, K. 1976 The covalent structure of dog myoglobin. *Biochim. biophys. Acta* **420**, 376–386 & 427, 759–761.

Edman, P. & Begg, G. 1967 A protein sequenator. *Eur. J. Biochem.* **1**, 80–96.

Edmundson, A. B. 1965 Amino-acid sequence of sperm whale myoglobin. *Nature, Lond.* **205**, 883–887.

Englander, S. W. & Staley, R. 1969 Measurement of the free and the H-bonded amides of myoglobin. *J. molec. Biol.* **45**, 277–295.

Epstein, C. J. 1967 Non-randomness of amino-acid changes in the evolution of homologous proteins. *Nature, Lond.* **215**, 355–359.

Erickson, R. P. & Rosenberg, L. T. 1973 The role of tolerance in the lack of cross-reactivity between homologous (structurally related) proteins: cross-reactions between α-lactalbumins and between lysozymes recognized by antibodies from intolerant rabbits. *Immunochemistry* **10**, 467–470.

Estabrook, G. F. 1972 Theoretical concepts in systematics and evolutionary studies. *Prog. theor. Biol.* **2**, 23–86.

Farris, J. S. 1967 The meaning of relationship and taxonomic procedure. *Syst. Zool.* **16**, 44–51.

Farris, J. S. 1970 Methods for computing Wagner trees. *Syst. Zool.* **19**, 38–92.

Farris, J. S. 1972 Estimating phylogenetic trees from distance matrices. *Am. Nat.* **106**, 645–668.

Fisher, J. 1967 Aves. Pp. 733–762 in *The fossil record* (eds W. B. Harland *et al.*). London: Geological Society.

Fisher, W. K. & Thompson, E. O. P. 1976 Studies on monotreme proteins. VII. Amino acid sequence of myoglobin from the platypus, *Ornithorhynchus anatinus*. *Aust. J. biol. Sci.* **29**, 57–72.

Fitch, W. M. 1971 Rate of change of concomitantly variable codons. *J. molec. Evol.* **1**, 84–96.

Fitch, W. M. & Langley, C. H. 1977 Evolutionary rates in proteins. In *Molecular anthropology* (eds M. Goodman & R. E. Tashian), pp. 197–219. New York: Wenner-Gren.

Fitch, W. M. & Margoliash, E. 1967 Construction of phylogenetic trees. *Science, N.Y.* **155**, 276–284.

Fitch, W. M. & Markowitz, E. 1970 An improved method for determining codon variability in a gene and its application to the rate of fixations of mutations in evolution. *Biochem. Genet.* **4**, 579–593.

Fox, R. C. 1970 Eutherian mammal from the early Campanian (Late Cretaceous) of Alberta, Canada. *Nature, Lond.* **227**, 630–631.

Freedman, L. 1957 The fossil Cercopithecoidea of South Africa. *Ann. Trans. Mus.* **23**, 121–262.

Geological Society Phanerozoic Time-scale 1964 *Q. J. geol. Soc. Lond.* **120**, (Suppl.), 260–262.

Geological Society Phanerozoic Time-scale – a supplement 1971 Special Publication No. 5. London: Geological Society.

Ginsburg, L. & Heintz, E. 1968 La plus ancienne Antilope d'Europe, *Eotragus artensis* du Burdigalien d'Artenay. *Bull. Mus. natn. Hist. nat., Paris* **40**, 837–842.

Glossman, H., Horst, J., Plagens, U. & Braunitzer, G. 1970 Hämoglobine XVIII. Isoelektrisches Fokussieren von Hämoglobinen (Erythrocrurorinen) der Insekten (*Chironomus thummi thummi*, Diptera). *Hoppe-Seyler's Z. physiol. Chem.* **351**, 342–384.

Goodman, M., Barnabas, J., Matsuda, G. & Moore, G. W. 1971 Molecular evolution in the descent of man. *Nature, Lond.* **233**, 604–613.

Goodman, M., Moore, G. W., Barnabas, J. & Matsuda, G. 1974 The phylogeny of human globin genes investigated by the maximum parsimony method. *J. molec. Evol.* **3**, 1–48.

Goodman, M., Moore, G. W. & Matsuda, G. 1975 Darwinian evolution in the genealogy of haemoglobin. *Nature, Lond.* **253**, 603–608.

Gower, J. C. 1966 Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338.

Graham, S. J. & Ride, W. D. R. 1967 Infraclass Metatheria. Pp. 767–768 in *The fossil record* (eds W. B. Harland *et al.*). London: Geological Society.

Grantham, R. 1974 Amino acid difference formula to help to explain protein evolution. *Science, N.Y.* **185**, 862–864.

Hamilton, W. R. 1973 The Lower Miocene ruminants of Gebel Zelten, Libya. *Bull. Br. Mus. nat. His.* (*Geol.*) **21**, 73–150.

Han, K., Dautrevaux, M., Chaila, X. & Biserte, G. 1970 The covalent structure of beef heart myoglobin. *Eur. J. Biochem.* **16**, 465–471.

Han, K., Tetaert, D., Moschetto, Y., Dautrevaux, M. & Kopeyan, C. 1972 The covalent structure of sheep heart myoglobin. *Eur. J. Biochem.* **27**, 585–592.

Harris, C. M. & Hill, R. L. 1969 The carboxymethylation of human met-myoglobin. *J. biol. Chem.* **244**, 2195–2203.

Hemmingsen, E. A. 1963 Enhancement of oxygen transport by myoglobin. *Comp. Biochem. Physiol.* **10**, 239–244.

Hemmingsen, E. A. 1965 Accelerated transfer of oxygen through solutions of heme pigments. *Acta physiol. scand.* **64** (*Suppl.* 246), 1–53.

Hershkovitz, P. 1974 A new genus of Late Oligocene monkey (Cebidae, Platyrrhini) with notes on postorbital closure and Platyrrhine evolution. *Folia primatol.* **21**, 1–35.

Hill, R. 1936 Oxygen dissociation curves of muscle haemoglobin. *Proc. R. Soc.* B **120**, 472–483.

Holmquist, R. 1972*a* Theoretical foundations for a quantitative approach to paleogenetics. Part I: DNA. *J. molec. Evol.* **1**, 115–133.

Holmquist, R. 1972*b* Theoretical foundations for a quantitative approach to paleogenetics. Part II: Proteins. *J. molec. Evol.* **1**, 134–149.

Holmquist, R. 1972*c* Empirical support for a stochastic model of evolution. *J. molec. Evol.* **1**, 211–222.

Holmquist, R., Cantor, C. & Jukes, T. H. 1972 Improved procedures for comparing homologous sequences in molecules of proteins and nucleic acids. *J. molec. Biol.* **64**, 145–161.

Holmquist, R. & Jukes, T. H. 1972 Sequence data: magnitude and implications of some ambiguities. *J. molec. Evol.* **2**, 10–16.

Holmquist, R., Jukes, T. H., Moise, H., Goodman, M. & Moore, G. W. 1976 The evolution of the globin family genes: concordance of stochastic and augmented maximum parsimony genetic distances for α hemoglobin, β hemoglobin and myoglobin phylogenies. *J. molec. Biol.* **105**, 39–74.

Hooijer, D. A. 1967 Superorders Mesaxonia and Paraxonia. Pp. 780–781 in *The fossil record* (ed W. B. Harland *et al.*). London: Geological Society.

Imamura, T., Baldwin, T. O. & Riggs, A. 1972 The amino acid sequence of the monomeric hemoglobin component from the bloodworm, *Glycera dibranchiata*. *J. biol. Chem.* **247**, 2785–2797.

Irving, L. 1939 Respiration in diving mammals. *Physiol. Rev.* **19**, 112–134.

Irving, L. 1964 Comparative anatomy and physiology of gas transport mechanism. Pp. 177–212 in *Handbook of Physiology: Respiration* (section 3, 1). Washington, D.C.: American Physiological Society.

Jardine, N. & Sibson, R. 1971 *Mathematical taxonomy*. London: Wiley & Sons.

Jardine, N., van Rijsbergen, C. J. & Jardine, C. J. 1969 Evolutionary rates and the inference of evolutionary tree forms. *Nature, Lond.* **224**, 185.

Jepsen, G. L. & Woodburne, M. O. 1969 Paleocene Hyracothere from Polecat Bench Formation, Wyoming. *Science, N.Y.* **164**, 543–547.

Jukes, T. H. & Holmquist, R. 1972 Estimation of evolutionary changes in certain homologous polypeptide chains. *J. molec. Biol.* **64**, 163–179.

Karadjova, M., Nedrov, P., Bakardjieva, A. & Genov, N. 1970 Differences in amino acid sequence between dolphin and sperm whale myoglobins. *Biochim. biophys. Acta* **221**, 136–139.

Kellogg, R. 1928 The history of whales – their adaptation to life in the water. *Q. Rev. Biol.* **3**, 29–76, 174–208.

Kendrew, J. C. 1962 Side-chain interactions in myoglobin. *Brookhaven Symp. Biol.* **15**, 216–228.

Kielan-Jaworowska, Z. 1968 Results of the Polish-Mongolian Palaeontological Expeditions. Part I. *Paleont. pol.* **19**, 171–191.

Kimura, M. 1968 Evolutionary rate at the molecular level. *Nature, Lond.* **217**, 624–626.
Kimura, M. & Ohta, T. 1972 On the stochastic model for estimation of mutational distance between homologous proteins. *J. molec. Evol.* **2**, 87–90.
King, J. L. & Jukes, T. H. 1969 Non-Darwinian evolution. *Science, N.Y.* **164**, 788–798.
Kirsch, J. A. W. 1969 Serological data and phylogenetic inference: the problem of rates of change. *Syst. Zool.* **18**, 296–311.
Kluge, A. G. & Farris, J. S. 1969 Quantitative phyletics and the evolution of anurans. *Syst. Zool.* **18**, 1–32.
Kluh, I. & Bakardjieva 1971 Primary structure of N-terminal part of molecule of dolphin myoglobin. *FEBS letters* **17**, 31–34.
Koketsu, J. & Atassi, M. Z. 1974*a* Immunochemistry of sperm-whale myoglobin. XVI. Accurate delineation of the single reactive region in sequence 1–55 by immunochemical studies of synthetic peptides. Some conclusions concerning antigenic structures of proteins. *Immunochemistry* **11**, 1–12.
Koketsu, J. & Atassi, M. Z. 1974*b* Immunochemistry of sperm whale myoglobin. XIX. Accurate delination the single reactive region in sequence 54–85 by immunochemical study of synthetic peptides. *Biochim. biophys. Acta* **342**, 21–29.
Kooyman, G. L. 1969 The Weddell seal. *Scient. Am.* **221**, 100–106.
Lang, A., Lehmann, H. & King-Lewis, P. A. 1974 Hb K Woolwich, the cause of a thalassaemia. *Nature, Lond.* **249**, 467–469.
Langley, C. H. & Fitch, W. M. 1974 An examination of the constancy of the rate of molecular evolution. *J. molec. Evol.* **3**, 161–177.
Lehmann, H. & Huntsman, R. G. 1974 *Man's haemoglobins.* Amsterdam: North-Holland Publishing Co.
Lehmann, H., Romero-Herrera, A. E., Joysey, K. A. & Friday, A. E. 1974 Comparative structure of myoglobin: Primates and tree-shrew. *Ann. N.Y. Acad. Sci.* **241**, 380–391.
Li, S. L. & Riggs, A. 1970 The amino acid sequence of hemoglobin V from the lamprey, *Petromyzon marinus. J. biol. Chem.* **245**, 6149–6169.
MacIntyre, G. T. 1966 The Miacidae (Mammalia, Carnivora) Part I. The systematics of *Ictidopappus* and *Protictis. Bull. Am. Mus. nat. Hist.* **131**, 115–210.
Matthews, S. C. & Missarzhevsky, V. V. 1975 Small shelly fossils of late Precambrian and early Cambrian age: a review of recent work. *Q. J. geol. Soc. Lond.* **131**, 289–304.
McKenna, M. C. 1969 The origin and early differentiation of Therian mammals. *Ann. N.Y. Acad. Sci.* **167**, 217–240.
McLaren, I. A. 1960 Are the Pinnipedia biphyletic? *Syst. Zool.* **9**, 18–28.
McQuitty, L. L. 1966 Similarity analysis by reciprocal pairs for discrete and continuous data. *Educ. psychol. Measur.* **26**, 825–831.
Miles, R. S. 1967 Class Acanthodii. Pp. 637–640 in *The fossil record* (ed W. B. Harland *et al.*). London: Geological Society.
Millikan, G. A. 1939 Muscle haemoglobin. *Physiol. Rev.* **19**, 503–523.
Moore, G. W., Barnabas, J. & Goodman, M. 1973 An iterative approach from the standpoint of the additive hypothesis to the dendrogram problem posed by molecular data sets. *J. theor. Biol.* **38**, 423–457.
Moore, G. W., Goodman, M. & Barnabas, J. 1973 A method for constructing maximum parsimony ancestral amino acid sequences on a given network. *J. theor. Biol.* **38**, 459–485.
Moore, G. W., Goodman, M., Callahan, C., Holmquist, R. & Moise, H. 1976 Stochastic *versus* augmented parsimony method for estimating superimposed mutations in the divergent evolution of protein sequences. Methods tested on cytochrome c amino acid sequences. *J. molec. Biol.* **105**, 15–37.
Morimoto, H., Lehmann, H. & Perutz, M. F. 1971 Molecular pathology of human haemoglobin: sterochemical interpretation of abnormal oxygen affinities. *Nature, Lond.* **232**, 408–413.
Nei, M. 1975 *Molecular population genetics and evolution.* Amsterdam: North-Holland Publishing Co.
Nisonoff, A., Reichlin, M. & Margoliash, E. 1970 Immunological activity of cytochrome c. II. Localization of a major antigenic determinant of human cytochrome c. *J. biol. Chem.* **245**, 940–946.
O'Gorman, P., Lehmann, H., Alsopp, K. M. & Sukumaran, P. K. 1963 Sickle-cell haemoglobin K disease. *Br. med. J.* ii, 1381–1382.
Ohta, T. 1974 Mutational pressure as the main cause of molecular evolution and polymorphism. *Nature, Lond.* **252**, 351–354.
Ohta, T. & Kimura, M. 1971 Amino acid composition of proteins as a product of molecular evolution. *Science, N.Y.* **174**, 150–153.
Patterson, B., Behrensmeyer, A. K. & Sill, W. D. 1970 Geology and faunal correlations of a new Pliocene locality in north-western Kenya. *Nature, Lond.* **226**, 918–921.
Peabody, F. E. 1952 *Petrolacosaurus kansensis* Lane, a Pennsylvanian reptile from Kansas. *Paleont. Contr. Univ. Kans. Vertebrate* 1–41.
Peacocke, D. & Boulter, D. 1975 Use of amino acid sequence data in phylogeny and evaluation of methods using computer simulation. *J. molec. Biol.* **95**, 513–527.
Peiffer, S. 1973 *La myoglobine du manchot empereur (Aptenodytes forsteri).* Thèse, Université libre de Bruxelles, Faculté des Sciences.

Perutz, M. F. 1965 Structure and function of haemoglobin. I. A tentative atomic model of horse oxyhaemoglobin. *J. molec. Biol.* **13**, 646–668.

Perutz, M. F. 1969 X-ray analysis, structure and function of enzymes. *Eur. J. Biochem.* **8**, 455–466.

Perutz, M. F., Kendrew, J. C. & Watson, H. C. 1965 Structure and function of haemoglobin. II. Some relations between polypeptide chain configuration and amino acid sequence. *J. molec. Biol.* **13**, 669–678.

Perutz, M. F. & Lehmann, H. 1968 Molecular pathology of human haemoglobin. *Nature, Lond.* **219**, 902–909.

Perutz, M. F., Muirhead, H., Cox, J. M. & Goaman, L. C. G. 1968 Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model. *Nature, Lond.* **219**, 131–139.

Pilgrim, G. E. 1939 The fossil Bovidae of India. *Mem. geol. Surv. India Palaeont. Indica (n.s.)* **26**, 1–356.

Pilgrim, G. E. 1947 The evolution of the buffaloes, oxen, sheep and goats. *J. Linn. Soc.* **41**, 272–286.

Piveteau, J. 1961 *Traité de Paléontologie* (ed J. Piveteau), vol. 6. Paris: Masson.

Prager, E. M. & Wilson, A. C. 1971 The dependence of immunological cross-reactivity upon sequence resemblance among lysozymes. II. Comparison of precipitin and microcomplement fixation results. *J. biol. Chem.* **246**, 7010–7017.

Quinn, J. H. 1955 Miocene Equidae of Texas Gulf Coastal Plain. *Univ. Texas Publs.* **5516**, 1–102.

Read, D. W. 1975 Primate phylogeny, neutral mutations, and 'molecular clocks'. *Syst. Zool.* **24**, 209–221.

Reichlin, M. 1975 Amino acid substitution and the antigenicity of globular proteins. *Adv. Immun.* **20**, 71–123.

Ringelhann, B., Konotey-Ahulu, F. I. D., Talapatra, N. C., Nkrumah, F. K., Wiltshire, G. B. & Lehmann, H. 1971 Haemoglobin K Woolwich ($\alpha_2\beta_2$ 132 lysine-glutamine) in Ghana. *Acta haemat.* **45**, 250–258.

Romer, A. S. 1966 *Vertebrate paleontology* (third edition). Chicago: Univ. of Chicago Press.

Romero-Herrera, A. E., Darbre, P. D. & Lehmann, H. 1976 The myoglobin of the Cape hunting dog (*Lycaon pictus*). *Biochim. biophys. Acta* **420**, 350–357.

Romero-Herrera, A. E. & Lehmann, H. 1971 The myoglobin of primates. I. *Hylobates agilis* (gibbon). *Biochim. biophys. Acta* **251**, 482–488.

Romero-Herrera, A. E. & Lehmann, H. 1972*a* The myoglobin of primates. II. *Pan troglodytes* (chimpanzee). *Biochim. biophys. Acta* **278**, 62–67.

Romero-Herrera, A. E. & Lehmann, H. 1972*b* The myoglobin of primates. III. Cercopithecidae (Old World monkeys): *Papio anubis* (olive baboon) and *Macaca fascicularis* (= *irus*, crab-eating monkey). *Biochim. biophys. Acta* **278**, 465–481.

Romero-Herrera, A. E. & Lehmann, H. 1973*a* The myoglobin of primates. IV. New World monkeys: Cebidae: (1) *Saimiri sciureus* (squirrel monkey); (2) *Lagothrix lagothricha* (Humboldt's woolly monkey). Callitrichidae: *Callithrix jacchus* (common marmoset). *Biochim. biophys. Acta* **317**, 65–84.

Romero-Herrera, A. E. & Lehmann, H. 1973*b* The myoglobin of primates. V. Prosimians: *Galago crassicaudatus* (thick-tailed galago) and *Lepilemur mustelinus* (sportive lemur). *Biochim. biophys. Acta* **322**, 10–22.

Romero-Herrera, A. E. & Lehmann, H. 1974*a* The amino acid sequence of human myoglobin and its minor fractions. *Proc. R. Soc.* B **186**, 249–279.

Romero-Herrera, A. E. & Lehmann, H. 1974*b* The myoglobin of primates. VI. *Tupaia glis belangeri* (common tree shrew). *Biochim. biophys. Acta* **359**, 236–241.

Romero-Herrera, A. E. & Lehmann, H. 1975*a* The myoglobin of primates. VII. *Perodicticus potto edwardsi* (potto). *Biochim. biophys. Acta* **393**, 205–214.

Romero-Herrera, A. E. & Lehmann, H. 1975*b* The primary structure of the myoglobin of *Didelphis marsupialis* (Virginia opossum). *Biochim. biophys. Acta* **400**, 387–398.

Romero-Herrera, A. E., Lehmann, H. & Castillo, O. 1976 The myoglobin of primates. IX. *Nycticebus coucang* (slow loris). *Biochim. biophys. Acta* **420**, 350–357.

Romero-Herrera, A. E., Lehmann, H. & Castillo, O. 1976 The primary structure of the myoglobin of rabbit (*Oryctolagus cuniculus*). *Biochim. biophys. Acta* **439**, 51–54.

Romero-Herrera, A. E., Lehmann, H. & Fakes, W. 1975 The primary structure of the myoglobin of the insectivore *Erinaceus europaeus* (common European hedgehog). *Biochim. biophys. Acta* **379**, 13–21.

Romero-Herrera, A. E., Lehmann, H. & Fossey, D. 1975 The myoglobin of primates. VIII. *Gorilla gorilla beringei* (Eastern Highland gorilla). *Biochim. biophys. Acta* **393**, 383–388.

Romero-Herrera, A. E., Lehmann, H., Joysey, K. A. & Friday, A. E. 1973 Molecular evolution of myoglobin and the fossil record: a phylogenetic synthesis. *Nature, Lond.* **246**, 389–395.

Saban, R. 1958 Insectivora. Pp. 822–909 in *Traité de Paléontologie.* **6**, 2 (éd J. Piveteau). Paris: Masson.

Savage, R. J. G. 1957 The anatomy of *Potamotherium*, an Oligocene lutrine. *Proc. zool. Soc. Lond.* **129**, 151–244.

Savage, R. J. G. & Hamilton, W. R. 1973 Introduction to the Miocene mammal faunas of Gebel Zelten, Libya. *Bull. Br. Mus. nat. Hist. (Geol.)* **22**, 515–527.

Scholander, P. F. 1960 Oxygen transport through hemoglobin solutions. *Science, N.Y.* **131**, 585–590.

Scholander, P. F. 1961 Physiological adaptation to diving in animals and man. *Harvey Lect.* **57**, 93–110.

Scholander, P. F. 1964 Animals in aquatic environments: diving mammals and birds. Pp. 729–739 in *Handbook of Physiology. Adaptation to the environment* (section 4). Washington, D.C.: American Physiological Society.

Scholander, P. F. 1965 Tension gradients accompanying accelerated oxygen transport in a membrane. *Science, N.Y.* **149**, 876–877.

Scholander, P. F., Irving, L. & Grinnell, S. W. 1942 Aerobic and anaerobic changes in seal muscles during diving. *J. biol. Chem.* **142**, 431–440.

Simons, E. L. 1965 New fossil apes from Egypt and the initial differentiation of Hominoidea. *Nature, Lond.* **205**, 135–139.

Simons, E. L. 1971 Relationships of *Amphipithecus* and *Oligopithecus*. *Nature, Lond.* **232**, 489–491.

Simons, E. L. 1972 *Primate evolution: an introduction to man's place in nature.* New York: Macmillan.

Simons, E. L. & Pilbeam, D. R. 1972 Hominoid paleoprimatology. Pp. 36–62 in *The functional and evolutionary biology of Primates* (ed R. Tuttle). Chicago: Aldine, Atherton.

Simpson, G. G. 1945 The principles of classification and a classification of the mammals. *Bull. Am. Mus. nat. Hist.* **85**, 1–350.

Simpson, G. G. 1971 Concluding remarks: Mesozoic mammals revisited. Pp. 181–198 in *Early mammals* (eds D. M. Kermack & K. A. Kermack). *Zool. J. Linn. Soc.* **50**, Suppl. 1.

Slaughter, B. H. 1971 Mid-Cretaceous (Albian) therians of the Butler Farm local fauna, Texas. Pp. 131–143 in *Early mammals* (eds D. M. Kermack & K. A. Kermack). *Zool. J. Linn. Soc.* **50**, Suppl. 1.

Slijper, E. J. 1936 Die Cetaceen-vergleichendanatomisch und systematisch. *Capita zool.* **7**, 1–590.

Sloan, R. E. & Van Valen, R. 1965 Cretaceous mammals from Montana. *Science, N.Y.* **148**, 220–227.

Smith, J. Maynard 1968 'Haldane's dilemma' and the rate of evolution. *Nature, Lond.* **219**, 1114–1116.

Sneath, P. H. A. 1966 Relations between chemical structure and biological activity in peptides. *J. theor. Biol.* **12**, 157–195.

Stenzel, P. 1974 Opossum Hb chain sequence and neutral mutation theory. *Nature, Lond.* **252**, 62–63.

Szalay, F. S. 1968 The beginnings of Primates. *Evolution, Lancaster, Pa.* **22**, 19–36.

Szalay, F. S. 1969 Mixodectidae, Microsyopidae, and the Insectivore – Primate transition. *Bull. Am. Mus. nat. Hist.* **140**, 197–330.

Szalay, F. S. 1972 Paleobiology of the earliest Primates. Pp. 3–35 in *The functional and evolutionary biology of Primates* (ed R. Tuttle). Chicago: Aldine, Atherton.

Tarlo, L. B. H. 1964 A primitive whale from the London Clay of the Isle of Sheppey. *Proc. Geol. Ass.* **74**, 319–323.

Tattersall, I. 1970 *Man's ancestors.* London: John Murray.

Teicher, E., Maron, E. & Arnon, R. 1973 The role of specific amino acid residues in the antigenic reactivity of the loop peptide of lysozyme. *Immunochemistry* **10**, 265–271.

Tentori, L., Vivaldi, G., Carta, S., Marinucci, M., Massa, A., Antonini, E. & Brunori, M. 1973 The amino acid sequence of myoglobin from the mollusc *Aplysia limacina*. *Int. J. Peptide Protein Res.* **5**, 187–200.

Tetaert, D., Han, K., Plancot, M. T., Dautrevaux, M., Ducastaing, S., Hombrados, I. & Neuzil, E. 1974 The primary sequence of badger myoglobin. *Biochim. biophys. Acta* **351**, 317–324.

Thompson, E. A. 1973 The method of minimum evolution. *Ann. hum. Genet.* **36**, 333–340.

Theorell, H. 1934 Kristallinisches myoglobin. V. Die saverstoffbindungskurve des myoglobins. *Biochem. Z.* **268**, 73–82.

Van Couvering, J. A. 1972 Radiometric calibration of the European Neogene. Pp. 247–271 in *Calibration of hominoid evolution* (eds W. W. Bishop & J. A. Miller). New York: Wenner-Gren.

Van Couvering, J. A. & Miller, J. A. 1969 Miocene stratigraphy and age determinations, Rusinga Island, Kenya. *Nature, Lond.* **221**, 628–632.

Van Valen, L. 1966 Deltatheridia, a new order of mammals. *Bull. Am. Mus. nat. Hist.* **132**, 1–126.

Van Valen, L. 1971 Toward the origin of artiodactyls. *Evolution, Lancaster, Pa.* **25**, 523–529.

Van Valen, L. & Sloan, R. E. 1965 The earliest primates. *Science, N.Y.* **150**, 743–745.

Vigna, R. A., Gurd, L. J. & Gurd, F. R. N. 1974 Californian sea lion myoglobin. Complete covalent structure of the polypeptide chain. *J. biol. Chem.* **249**, 4144–4148.

Von Koenigswald, G. H. R. 1969 Miocene Cercopithecoidea and Oreopithecoidea from Miocene of East Africa. *Foss. Vert. Afr.* **1**, 39–52.

Vötsch, W. & Anderer, F. A. 1972 The N-terminal amino acid sequence of sheep heart myoglobin. *Z. Naturforsch.* **27b**, 157–159.

Vötsch, W. & Anderer, F. A. 1975 The amino acid sequence of myoglobin from skeletal muscles of red deer (*Cervus elaphus*). *J. molec. Evol.* **5**, 315–326.

Wagner, W. H., Jr. 1961 Problems in the classification of ferns. Pp. 841–844 in *Recent advances in botany.* Toronto: Univ. Toronto Press.

Walker, A. 1970 Postcranial remains of the Miocene Lorisidae of East Africa. *Am. J. phys. Anthrop.* **33**, 249–261.

Walker, A. C. 1974 A review of the Miocene Lorisidae of East Africa. Pp. 435–447 in *Prosimian biology* (eds R. D. Martin, G. A. Doyle & A. C. Walker). London: Duckworth.

Walker, A. & Andrews, P. 1973 Reconstruction of the Dental Arcades of *Ramapithecus wickeri*. *Nature, Lond.* **244**, 313–314.

Watson, H. C. 1969 The stereochemistry of the protein myoglobin. *Prog. Stereochem.* **4**, 299–333.

Whitehouse, H. L. K. 1973 *Towards an understanding of the mechanism of heredity* (third edition). London: Arnold.

Williams, C. T. 1969 Evolution of the Caninae and Borophaginae (Canidae). *Geol. Soc. Am. Abst.* (Pt. **3**), p. 74.

Wittenberg, J. B. 1959 Oxygen transport: a new function proposed for myoglobin. *Biol. Bull. mar. biol. Lab., Woods Hole* **117**, 402–403.

Wittenberg, J. B. 1966 The molecular mechanism of hemoglobin-facilitated oxygen diffusion. *J. biol. Chem.* **241**, 104–114.

Wittenberg, J. B. 1970 Myoglobin-facilitated oxygen diffusion: role of myoglobin in oxygen entry into muscle. *Physiol. Rev.* **50**, 559–635.

Wyman, J. 1966 Facilitated diffusion and the possible role of myoglobin as a transport mechanism. *J. biol. Chem.* **241**, 115–121.

Zuckerkandl, E. & Pauling, L. 1965 Evolutionary divergence and convergence in proteins. Pp. 97–166 in *Evolving genes and proteins* (eds V. Bryson & H. J. Vogel). New York: Academic Press.

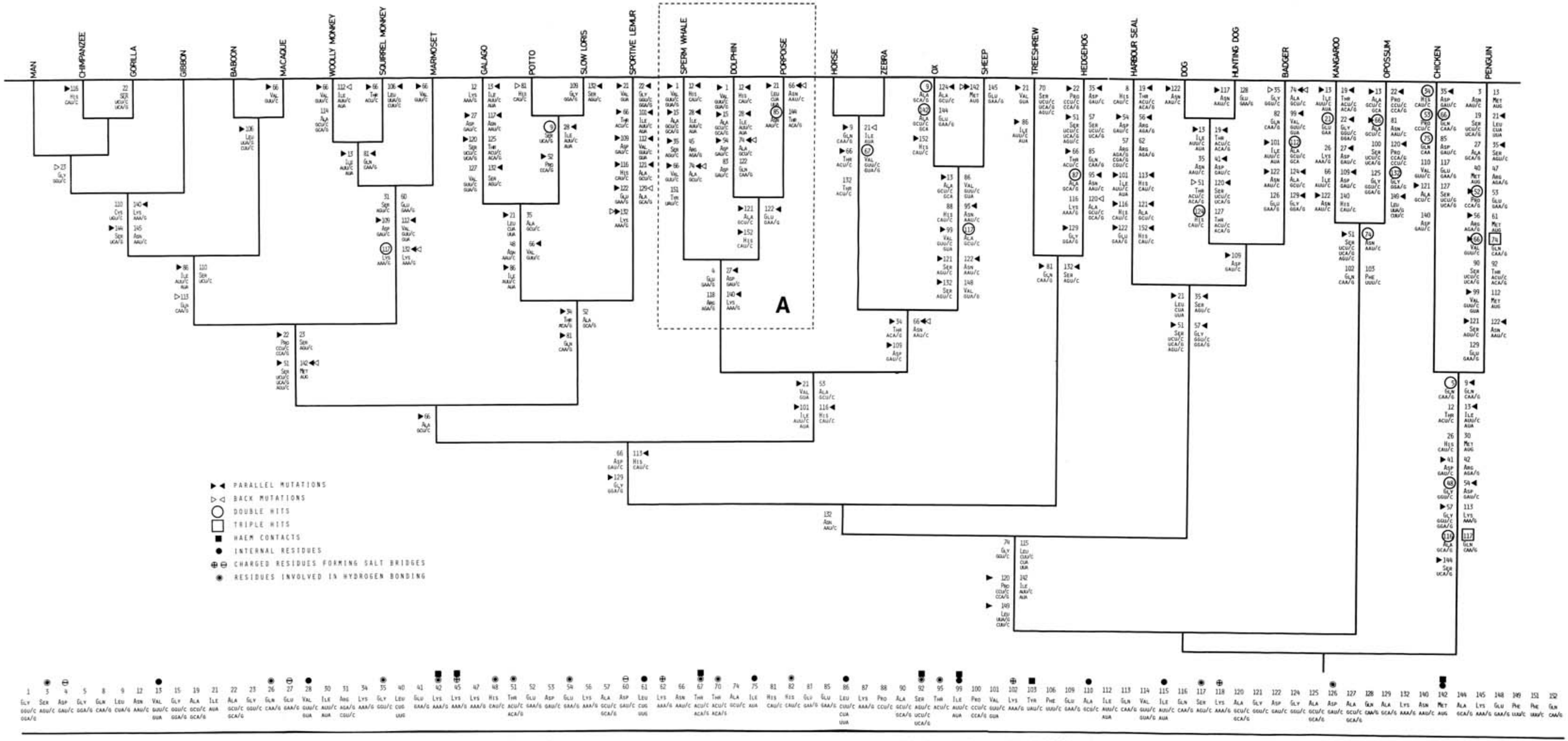


FIGURE 3. Cladogram 2. One of the two most parsimonious acceptable solutions. An alternative solution to block A is given in figure 4, and described in the text.

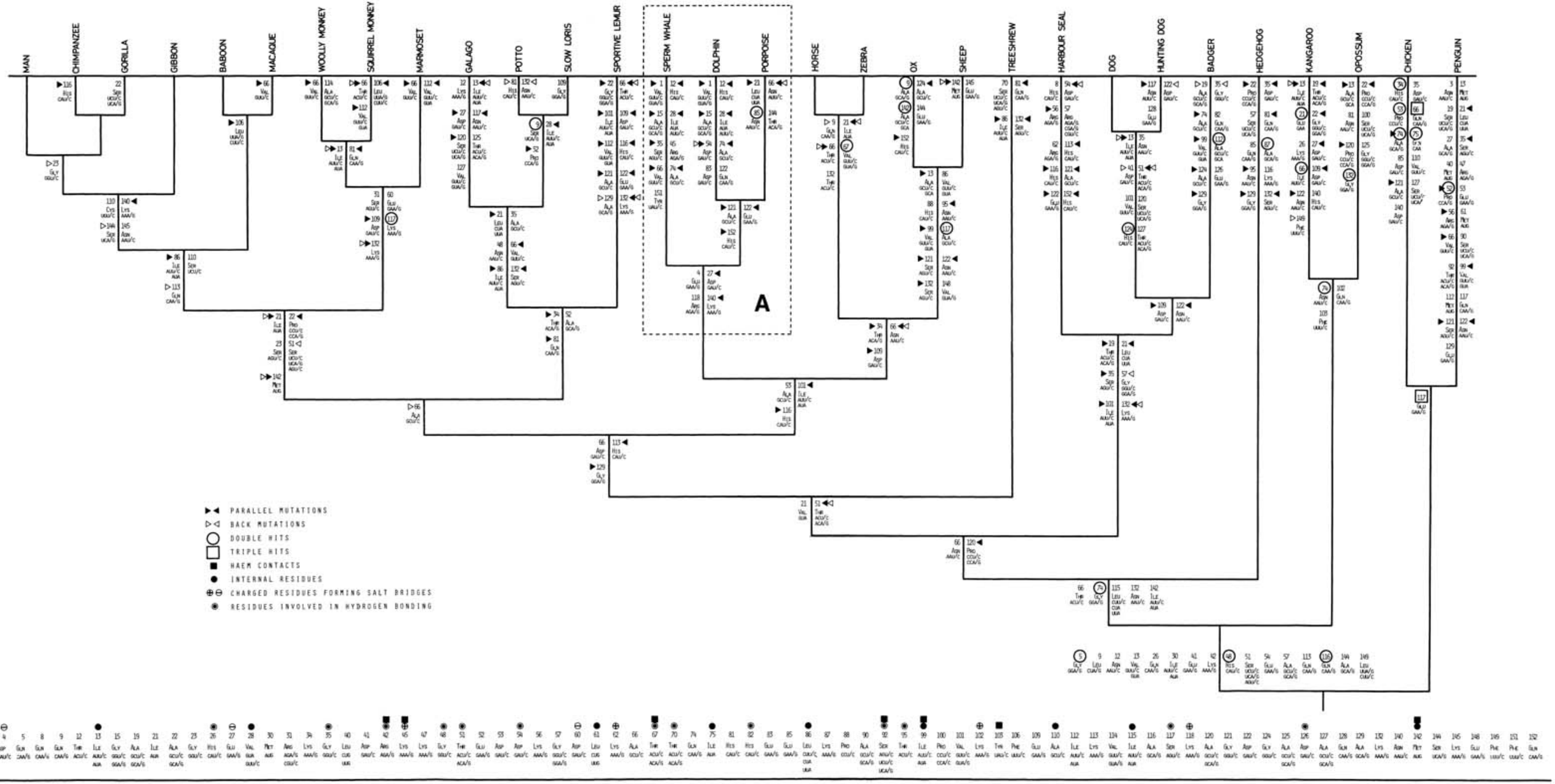


FIGURE 7. Cladogram 5. One of the two most parsimonious acceptable solutions. An alternative solution to block A is given in figure 4.