PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Primary Structure of Carp Myoglobin in the Context of Molecular Evolution

A. E. Romero-Herrera, N. Lieska, A. E. Friday and K. A. Joysey

Phil. Trans. R. Soc. Lond. B 1982 **297**, 1-25
doi: 10.1098/rstb.1982.0032

Phil. Trans. R. Soc. Lond. B **297**, 1–25 (1982) [1]
Printed in Great Britain

THE PRIMARY STRUCTURE OF CARP MYOGLOBIN IN THE CONTEXT OF MOLECULAR EVOLUTION

By A. E. ROMERO-HERRERA†, N. LIESKA†, A. E. FRIDAY‡ and K. A. JOYSEY‡

†*Department of Anatomy, Wayne State University School of Medicine, Detroit, Michigan* 48201, *U.S.A.*
‡*University Museum of Zoology, Downing Street, Cambridge CB2 3EJ, U.K.*

(*Communicated by H. Lehmann, F.R.S. – Received* 27 *May* 1980 – *Revised* 11 *May* 1981)

CONTENTS

The amino acid sequence of skeletal muscle myoglobin from carp (*Cyprinus carpio*) is presented. Comparisons are made with previously reported myoglobin sequences for several other fish and birds, and many mammals. The functional significance of the amino acid substitutions and 'deletions' in the carp sequence is considered.

The new sequence is used in a re-examination of the evidence for an approximately constant rate of molecular evolution. By using estimates of the dates of divergence of lineages leading to living species and equations put forward by proponents of the 'neutral theory' of biochemical evolution it is demonstrated that similar amounts of change appear to have occurred over periods of time that differ by more than a factor of two.

Vol. 297. B 1084 1 [Published 15 April 1982

1. Introduction

The amino acid sequences of myoglobin from over 60 species have been published to date, mostly derived from mammals. The myoglobins of fish, amphibians and reptiles have so far been very little investigated despite their obvious relevance to the study of vertebrate molecular and organismal evolution.

Recently determined sequences for fish myoglobin and an additional sequence reported here enable new reconstructions of hypothetical ancestral sequences. The new fish sequences also present opportunities for speculation regarding the functional significance of parts of the myoglobin molecule and for further examination of the problem of the rate of molecular evolution.

Very few muscles in non-amniotes possess myoglobin, and even these have very low concentrations as judged by the colour of the tissue. We chose the actinopterygian *Cyprinus carpio* (carp) after it was noted that its body musculature had a narrow bilateral strip containing red pigment. Because it was this animal whose haemoglobin sequence was utilized by Kimura (1968) in formulating his so-called 'neutral theory', we discuss, among other points, the significance that the carp myoglobin sequence has for this hypothesis.

2. Materials and methods

(*a*) *Isolation and purification of myoglobin*

Red skeletal muscle (1.2 kg) was removed from the lateral bands of five carp (*Cyprinus carpio*) weighing a total of 15 kg. The muscle was first ground, then homogenized for 2 min with 1.5 volumes of water containing 2 mM KCN. The homogenate was centrifuged (25 000 *g*, 30 min) and the supernatant adjusted to 55% ammonium sulphate saturation. After 1 h of continuous stirring at 4 °C the material was centrifuged (45 000 *g*) and the supernatant dialysed against 2 mM KCN for 72 h. Following concentration to 100 ml (Amicon PM-10 membrane), 20 ml portions were fractionated at a flow rate of 15 ml/h on columns of agarose–acrylamide (LKB AcA-54, 2.6 cm × 90 cm × 2 in tandem, equilibrated with 50 mM Tris–HCl, pH 8.5, 2 mM KCN). The red material that eluted with an approximate relative molecular mass M_r of 17 000 was concentrated and dialysed against water for 48 h. Removal of the haem group was performed by cold acid–acetone precipitation (1.5% HCl in acetone) followed by four washes with acetone to remove all traces of acid. The precipitated material was dried under a nitrogen current to yield 2.9 g of crude apomyoglobin. Of this, 2.5 g were dissolved in 40 ml of 5 mM Na_2HPO_4, pH 6.35, 1 mM DTT, 8 M urea (starting buffer), and applied to a column of Whatman CM-23 cellulose (2.6 cm × 30 cm). After release of the unbound material the chromatography was developed with a linear gradient composed of 800 ml of the starting buffer and 800 ml of 40 mM Na_2HPO_4, pH 6.35, 1 mM DTT, 8 M urea at a flow rate of 100 ml/h. The purified apomyoglobin fraction was dialysed against water for 72 h and lyophilized to yield 900 mg of protein.

(*b*) *Polyacrylamide gel electrophoresis*

Electrophoresis was performed on 10% acrylamide gels (5 mm × 75 mm) in the presence of 0.1% sodium dodecyl sulphate with use of the continuous buffer system of Porzio & Pearson (1977). Applications consisted of 10–50 μg of protein per gel. Following electrophoresis for 3.5 h at 2 mA per column, the gels were stained with 1% amido black.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

(*c*) *Aminoethylation and cyanogen bromide* (CNBr) *cleavage*

Cysteine residues were converted into aminoethylcysteine by reacting 50 mg of purified apomyoglobin (in 1 M Tris–HCl, pH 9.4, 8 M urea, 1% β-mercaptoethanol) with 0.5 ml of ethyleneimine for 2 h at room temperature (Jones 1964). The modified protein was dialysed against 0.1% ammonium bicarbonate for 72 h and lyophilized.

CNBr cleavage (Gross & Witkop 1962) was performed by dissolving 50 mg of protein in 1 ml of 70% formic acid to which a 100–200 molar excess of CNBr was added (relative to methionine). After standing at room temperature for 16 h the material was diluted with 100 ml of water and lyophilized. Purification of the CNBr peptides was achieved by gel filtration on Sephadex G-75 (2.6 cm × 180 cm column) equilibrated with 1% formic acid. The flow rate was 15 ml/h.

(*d*) *Enzymatic hydrolysis and peptide purification*

Portions of the aminoethylated protein were digested by trypsin and chymotrypsin (Sick *et al.* 1967). The insoluble cores left by these digestions were subsequently hydrolysed by pepsin (Romero-Herrera & Lehmann 1974*a*) and thermolysin (Lorkin *et al.* 1970), respectively. *Staphylococcus aureus* V8 protease digestion of the complete molecule was performed in 50 mM ammonium bicarbonate (pH 7.8) following the method of Houmard & Drapeau (1972). The tryptic peptides containing residues 1–8, 9–27, 42–57, 67–72, 75–93, 101–108 and 114–125 were subsequently digested by thermolysin. All enzymatic peptides were separated by fingerprinting on Whatman paper, by high-voltage electrophoresis at pH 6.5 followed by ascending chromatography at right angles (Romero-Herrera & Lehmann 1974*a*). All neutral and positively charged peptide spots in the fingerprints of the tryptic, peptic and chymotryptic digests routinely were cut from the papers and purified by electrophoresis at pH 3.5 (Romero-Herrera & Lehmann 1974*a*). The negatively charged ones were re-electrophoresed at pH 9.0. Similarly, the neutral regions and all overlapping spots in the fingerprints of specific, redigested peptides were purified by electrophoresis at the appropriate pH.

(*e*) *Amino acid composition of peptides and sequence determination*

Following elution from paper, the peptides were hydrolysed in 6 M HCl at 108 °C for 24–72 h and their amino acid compositions established in an automatic amino acid analyser (Beckman 119 CL) with attached data processor (Beckman 126). In general, residues constituting less than 0.20 mol.% of a peptide were regarded as contaminants and not recorded. Peptides to be sequenced were eluted from the paper with 3% ammonium hydroxide and the solvent removed by rotary evaporation. Dansyl-Edman degradation was performed by the technique of Hartley (1970) and the dansyl derivatives identified by polyamide thin-layer chromatography (Woods & Wang 1967; Crowshaw *et al.* 1967).

Identification of the N-terminal residue was possible by dansylation of the complete molecule (Gray 1972). Establishment of amide and acidic side chain groups was obtained by applying the equation charge $= M_r^{\frac{2}{3}} \times$ mobility (Offord 1966; Dixon 1972) to all negatively and positively charged peptides (see figure 2), while in two instances the assignment was on the basis of the specificity of the V8 enzyme.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

3. Results

Figure 1*a* shows the elution profile resulting from gel filtration of the ammonium sulphate supernatant of the carp muscle on AcA 54. Following haem removal, the protein from peak 3 was fractionated on CM-23 cellulose (figure 1*b*) to yield the purified apomyoblobin (peak 2). The homogeneity of this fraction was attested to by the presence of only one species (of M_r *ca.* 16000) upon polyacrylamide gel electrophoretic analysis in the presence of sodium dodecyl sulphate (figure 1*b*, inset). The fragments generated by cyanogen bromide cleavage of the molecule were separated on Sephadex G-75 (figure 1*c*).

Digestion of the myoglobin by a battery of enzymes produced numerous peptides which were separated by fingerprinting on paper (figure 2) and provided the multiple overlaps necessary for alignment.

The amino acid analyses of the total molecule (table 1) and of the various enzymic peptides have been deposited as a supplement in the archives of the Royal Society and the British Library, Lending Division.† This information and that derived from 77 steps of dansyl-Edman degradation performed on selected peptides allowed determination of the complete amino acid sequence shown in figure 3. The total amino acid composition of the protein as deduced from this sequence was in very close agreement with that determined by acid hydrolysis of the entire molecule (table 1). The molecule contained 146 amino acid residues with a total M_r of 15644, an N-terminus of His and a C-terminus of Gly. The tryptic peptide 127–140 was not present in the soluble fraction because its high content of hydrophobic residues caused it to precipitate from the digestion mixture. However, peptides from this region were recovered in the fingerprints of the chymotryptic, V8 protease, thermolytic and peptic digests.

Glx residues at positions 34 and 49 were concluded to be Glu on the basis of the V8 protease specificity. An interesting finding was that the staphylococcal V8 protease utilized at alkaline pH (Miles Laboratories, lot 0877) hydrolysed not only Glu-X peptide bonds but also Asp–X (positions 76, 117, 129, 134 and 136) and the following peptide bonds: Lys–Phe (41–42), Lys–Ala (57–58), Arg–Leu (100–101), Gly–Ile (44–45), Gly–Ala (60–61), Gly–Phe (143–144), Ala–Thr (61–62) and Met–Asp (128–129).

Small amounts of myoglobin were found to have Glu at positions 31, 36 and 121 (all Gln in the major fraction). This observation was regarded to be the result of artefactual deamidation (e.g. residue 121 in figure 2*e*) because the quantities of these Glu-containing peptides varied substantially between four different tryptic digests of the molecule. Allocation of the two Met positions was possible by analysis of the three CNBr peptides (residues 111–128, 129–146 and 111–146) which eluted together as peak 3 on Sephadex G-75 (figure 1*c*). Amino acid analyses of peaks 1 and 2 from this gel filtration revealed the presence of the intact, unreacted molecule and of the peptide residues 1–110 (with C-terminal homoserine), respectively.

Homologous alignment of the carp myoglobin sequence with those representative of mammalian species revealed several interesting features about the molecule: (1) it is four residues shorter at the N-terminus; (2) Trp has been replaced at position A5; (3) Cys is present at position A11; (4) there is no CD8 position; (5) there are no GH5–H1 positions; and (6) two neighbouring Tyr residues are present at positions H22 and H23.

† Copies of the material deposited may be purchased from the British Library, Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K. (reference SUP 10036).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

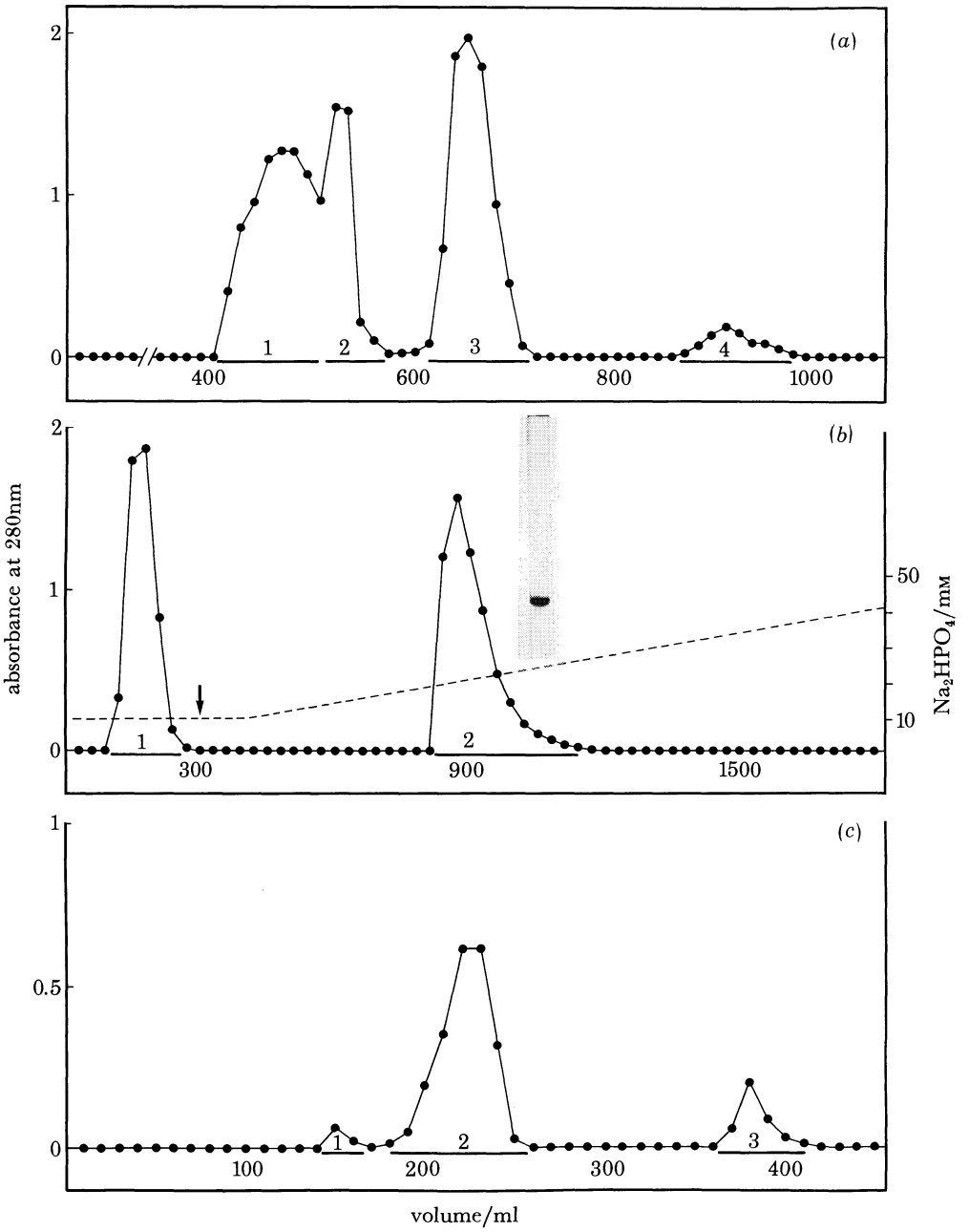


FIGURE 1. (*a*) Elution profile on AcA 54 of the crude muscle extract after ammonium sulphate fractionation. Peaks 1, 2 and 4: proteins other than myoglobin. Peak 3: myoglobin plus contaminating protein. (*b*) Elution profile of apomyoglobin (Aca 54, peak 3) on CM-23 cellulose. The arrow indicates the initiation of the linear gradient. Peak 1: unbound contaminating protein. Peak 2: purified apomyoglobin. Inset: sodium dodecyl sulphate–polyacrylamide gel electropherogram of peak 2 fraction. (*c*) Elution profile of myoglobin CNBr peptides on Sephadex G-75. Peak 1: unreacted material. Peak 2: peptide consisting of residues 1–110. Peak 3: peptides composed of residues 111–146, 111–128 and 129–146.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

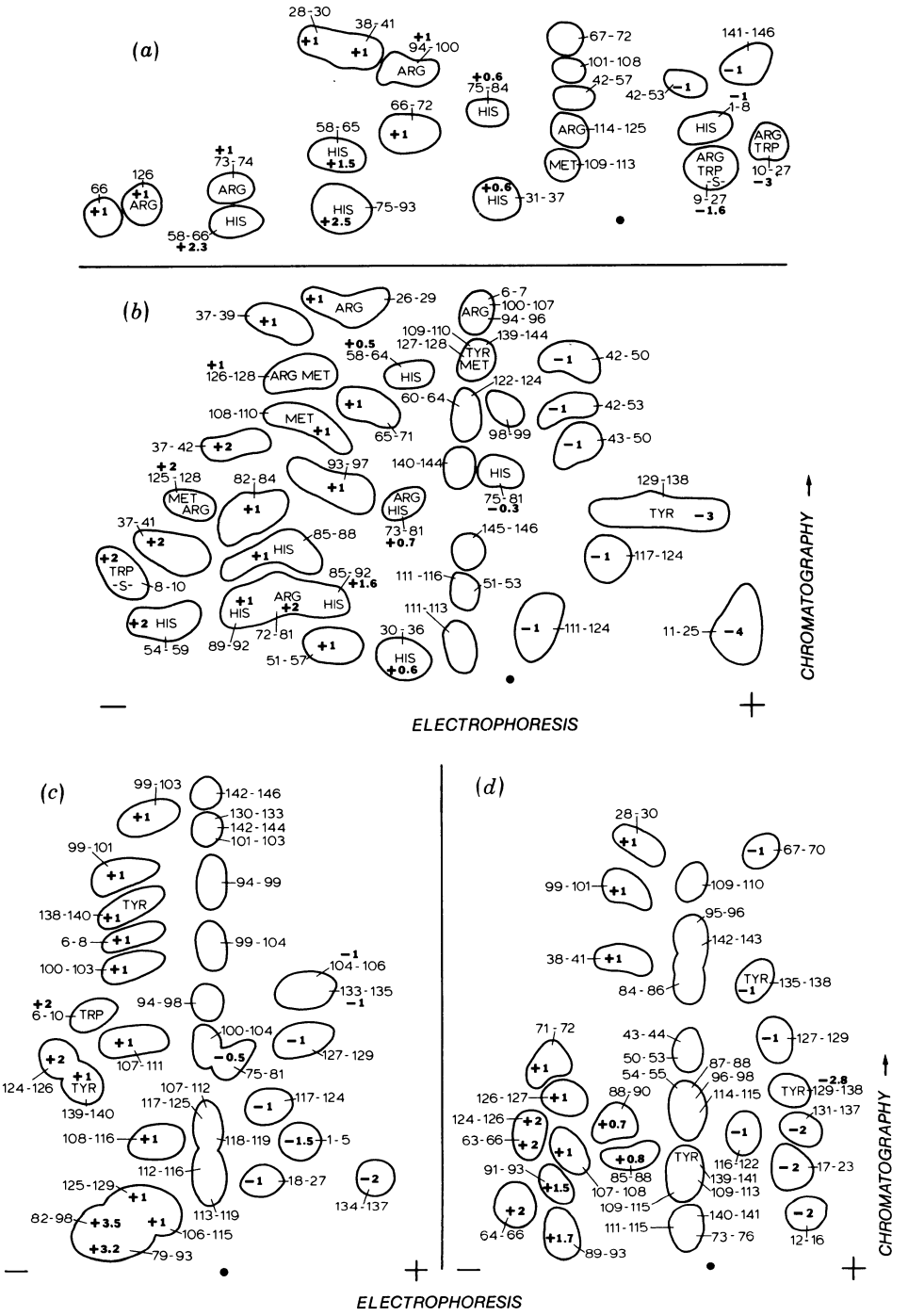


FIGURE 2. Two-dimensional peptide maps of the various enzymic digests of carp myoglobin. Each peptide is identified by its sequential numbers as given in figure 3. The boldfaced numbers indicate the net charges of peptides as calculated by Offord's equation. In (*a*) and (*b*) those peptides positive for specific staining reactions are indicated. (*a*) Soluble tryptic peptides from the aminoethylated globin; (*b*) soluble chymotryptic peptides from aminoethylated globin; (*c*) peptic peptides obtained from the insoluble tryptic core; (*d*) thermolytic peptides obtained from the insoluble chymotryptic core; (*e*) peptides from V8 protease digestion of the total molecule; (*f*) fingerprint of the short CNBr fragments (peak 3, figure 1*c*); (*g–l*) fingerprints of thermolytic peptides derived from specific tryptic peptides as indicated in the left upper margin of each figure. The symbol • represents the point of application of the specimen.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

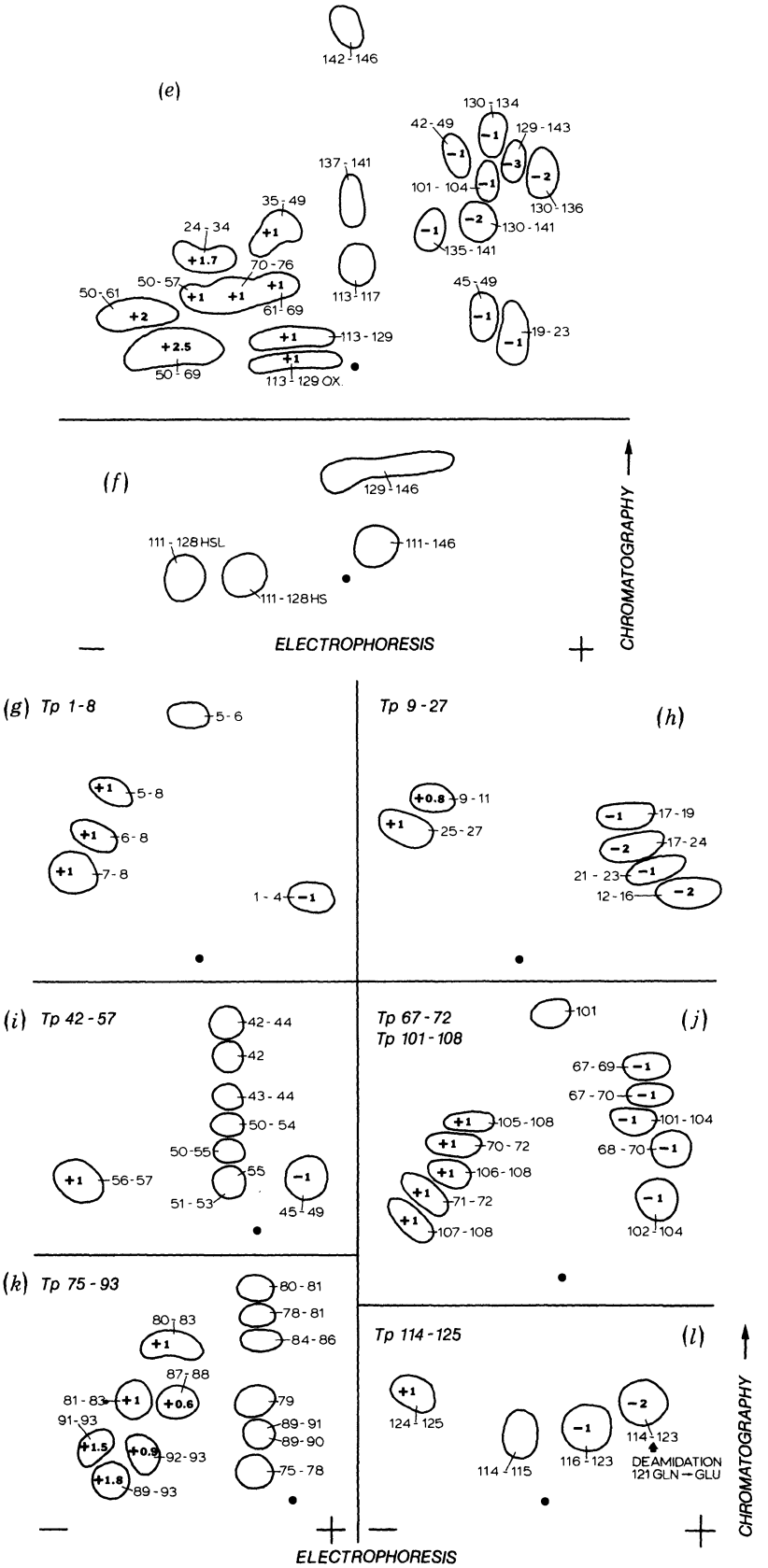


Table 1. Amino acid composition of carp myoglobin

	number of residues per molecule of protein from direct hydrolysis*	deduced from sequence
Cys†	0.98	1
Asp	12.34	12
Thr‡	8.55	9
Ser‡	1.97	2
Glu	13.36	13
Pro	3.15	3
Gly	16.40	16
Ala	19.25	19
Val§	12.21	12
Met‖	1.75	2
Ile§	6.92	7
Leu	17.23	17
Tyr	1.88	2
Phe	6.38	6
His	5.95	6
Lys	13.14	13
Arg	5.14	5
Trp¶	0.85	1
total		146

With reference to hydrolysis data:
* averaged values from 24, 48 and 72 h hydrolysis, except as noted;
† quantitated as cysteic acid following performic acid oxidation;
‡ extrapolated 0 h hydrolysis value;
§ 72 h hydrolysis value;
‖ quantitated as methionine sulphone following performic acid oxidation;
¶ recovered following 24 h mercaptoethane sulphonic acid hydrolysis.

4. Discussion

(*a*) *Comparative morphology of carp myoglobin*

A first step in appreciating the extent of modification that a protein undergoes during its evolution is to align the various amino acid sequences of the molecule from representative species against that sequence whose native conformation and function are best understood. As with comparative anatomy on a gross level this type of exercise on the molecular scale allows one to speculate upon the interrelationship of form and function as expressed through natural selection. The model for such studies with myoglobin is the sperm whale molecule. Figure 4 presents the sequences of the skeletal muscle myoglobins of the sperm whale (Edmundson 1965), carp and the shark *Heterodontus portusjacksoni* (Fisher & Thompson 1979), aligned so as to maximize the homology between molecules. On a purely quantitative level the carp molecule differs at 87 amino acid sites from the sperm whale myoglobin. Also, seven positions, namely 1–4 (NA1–A2), 50 (CD8) and 123–124 (GH5–H1), are absent in comparison with the known mammalian and bird myoglobins. The shark primary structure shows 84 differences from the sperm whale and five fewer positions, while the carp and shark differ at 83 positions, with the carp having two less residues in total.

If one assumes that the generalized tertiary structure of the carp and shark myoglobins resembles that of the sperm whale molecule, whose high resolution (2 Å†) crystallographic

† 1 Å (ångstrom) = 10^{-10} m.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

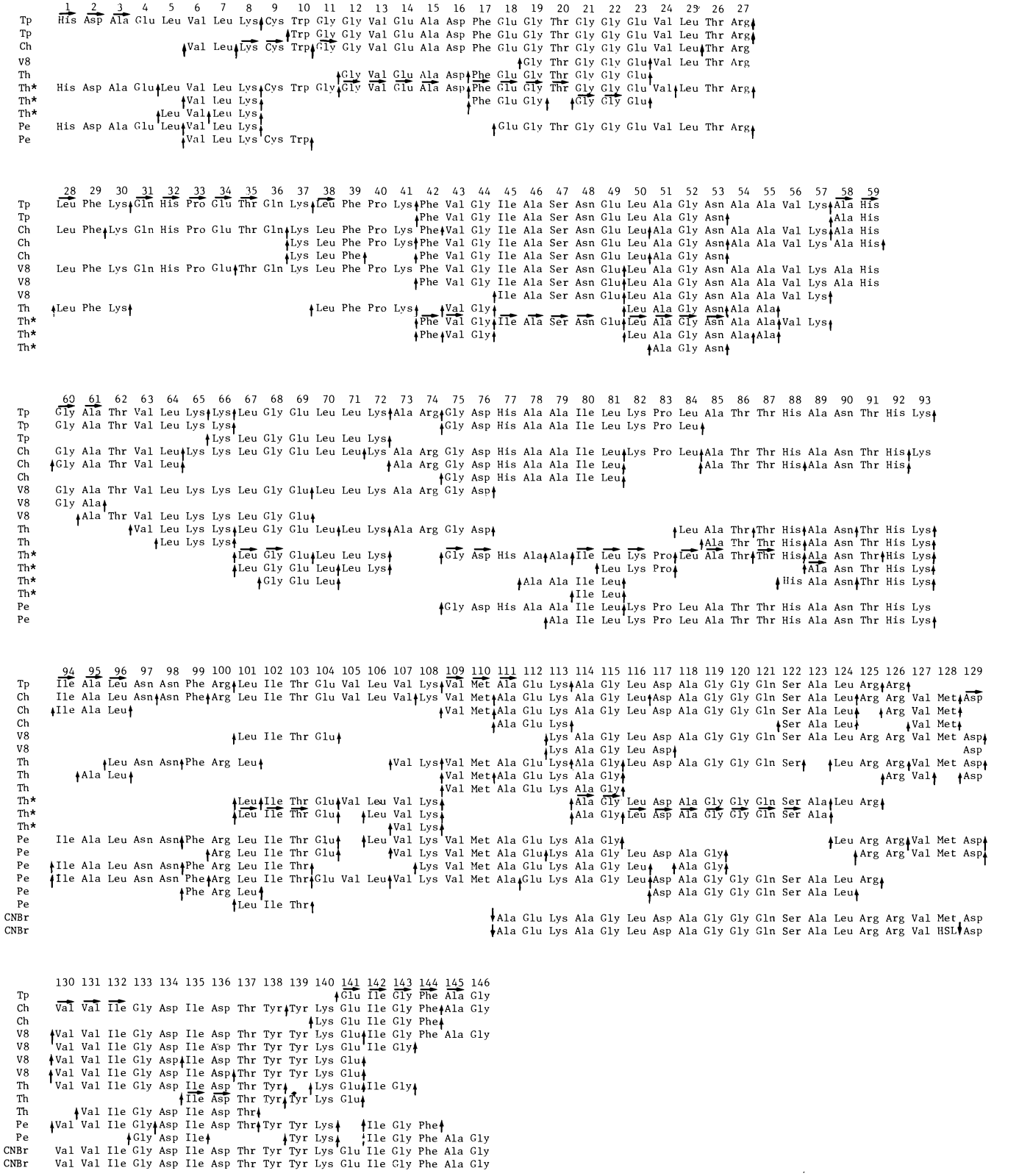


FIGURE 3. The primary structure of carp skeletal muscle myoglobin obtained by aligning overlapping peptides and sequential dansyl-Edman degradation ( ); , site of enzymic hydrolysis. Peptides: Tp, tryptic peptides; V8, V8 proteolytic peptides; Th, thermolytic peptides, Th*, thermolytic peptides derived from tryptic peptides; Ch, chymotryptic peptides; Pe, peptic peptides. CNBr (), cyanogen bromide peptides and points of cleavage.

structure is known (Takano 1977*a*, *b*), the homologous alignment shown in figure 4 can be used to speculate upon the manner in which some of the differences in primary structure may be expressed three-dimensionally. Perspective projections of computer-generated pictures of the sperm whale molecule (based on the coordinates of Takano) were used to aid comparisons. While the overwhelming majority of residues most important for the stability and function of the sperm whale molecule have been conserved in these fish myoglobins, there are some differences of consequence.

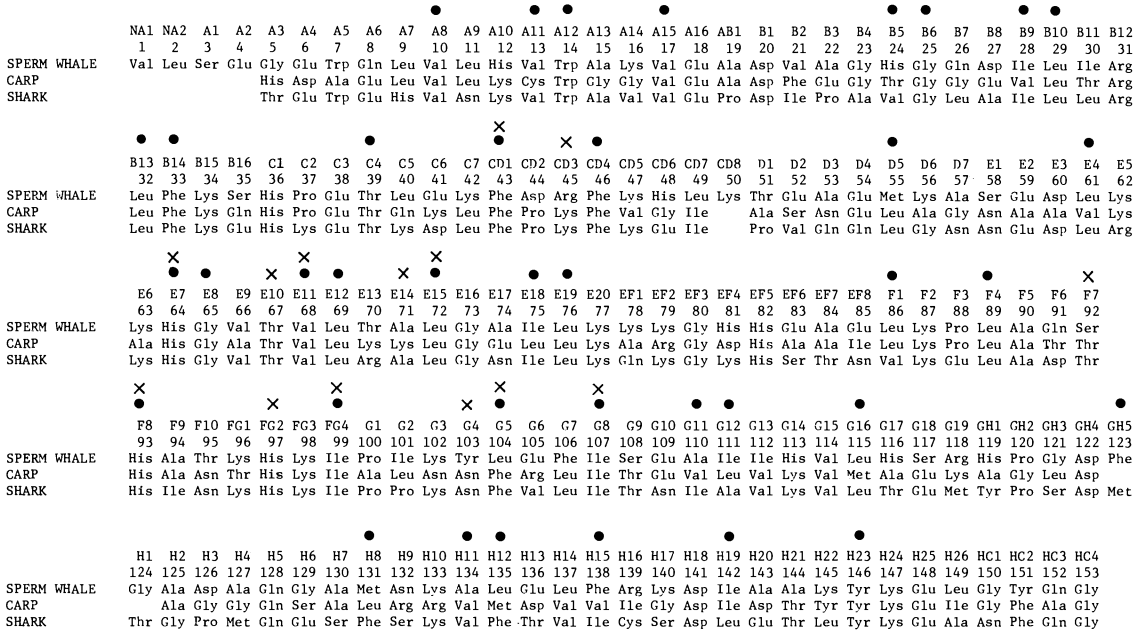


FIGURE 4. Alignment of the myoglobins of sperm whale, carp and shark. Maximum homology suggests that residues at positions 1–4 and 50 in the sperm whale are not a feature of the fish myoglobins. Residues at position 123 and 124 are absent in carp myoglobin when compared to sperm whale and shark myoglobins. Symbols • and × indicate the positions of internal and haem contact residues, respectively.

An example of compensatory modification which is expressed in the carp molecule is illustrated by a series of reconstructed events which involve the A-helix, the GH-bend and the beginning of the H-helix, secondary to the loss of two residue positions. In the sperm whale an important interaction is formed between A11 Val and the GH-segment and between GH5 Phe and the A-helix. In the three-dimensional model these residues are at the bottom of the hydrophobic cluster around the invariant Trp (A12), thereby helping to close the crevice between the GH and A regions. In reconstructing the molecule's phylogeny it could be assumed that the ancestral state for all vertebrate myoglobins was either GH5 Met, A11 Val (as in the shark) or GH5 Phe, A11 Val or Ile as in most of the known vertebrate myoglobins, with the exception of those of the bony fish. After the divergence of the actinopterygians, the residues at positions GH5 and H1 were deleted. In the ancestral myoglobin chain the sequence GH1, GH2, GH3, GH4,...H2, H3, H4 most probably was: Ala-Gly-Ser-Asp-Ala-Gly-Gly (Ser at position GH3 is here considered as the ancestral state because of its presence in shark and penguin). The resultant shorter GH corner, absence of GH5 with its usually bulky hydrophobic residue and a

missing hydrogen bond between GH1 His and B5 His could have introduced some instability into the molecule. Equilibrium was regained by a series of subsequent mutations: (1) GH3 became Leu, a larger and hydrophobic residue; (2) A11 changed from Val to Cys, which, by virtue of its bulkier and hydrophobic side chain, assumed the function of Phe GH5 in bringing together the GH-segment and A-helix, thus closing the bottom of the hydrophobic cluster around A12 (Takano 1977*a*).

A second quantitative difference concerns residue 50 (CD8, the first position in the non-helical CD segment) and distinguishes the myoglobins of both the carp and the shark from the known myoglobins of tetrapods. Lacking from the fish molecules, this position should be considered an insertion into the ancestral tetrapod myoglobin. In support of this assumption is the fact that the two known invertebrate myoglobins, from *Aplysia limacina* (Tentori *et al.* 1973) and *Busycon canaliculatum* (Bonner & Laursen 1977), also lack this residue. The acquisition of residue CD8 Lys, invariant in birds and mammals, probably conferred additional stability to the molecule in the CD corner, since in the sperm whale this residue is in contact (via a molecule of water) with position C6 (Glu or Asp in all known tetrapod myoglobins).

A final quantitative distinction, the absence of four sequential N-terminal positions in both fish molecules *vis-à-vis* the sperm whale model, contributes to the conclusion that a chain elongation occurred during evolution of the ancestral myoglobin common to birds and mammals. This allowed the formation of a hydrophobic bond between residue NA2 Leu (invariant) and the H-helix. Furthermore, the hydrogen bond created between the hydroxyl group of residue A1 (Ser or Thr) and the main chain imino group of residue A4 appears to assist initiation of the A-helix.

Conservation of the haem contacts has been remarkable in spite of the large overall number of amino acid differences between shark, carp and sperm whale myoglobins and the hundreds of million of years of divergent evolution. Of fourteen short interatomic haem contact residues in the sperm whale (those 4.0 Å distant or less; Takano 1977*a*), nine are identical to those of shark and carp. Of the remaining, three have undergone conservative substitution: CD3 Arg → Lys; F7 Ser → Thr; and G5 Leu → Phe. Alanyl at the external position E14 in the sperm whale and shark has been replaced by Lys in the carp. However, it is possible that in the latter species this haem contact does not exist, and that the two positions are not functionally comparable. Alternatively, the contact may be established through the β-carbon of the Lys side-chain. Lastly, a striking change has taken place at the contact G4 (which is occupied by Asn in the fish myoglobins and by either Phe or Tyr in all other known myoglobins). A detailed discussion of this substitution is given by Fisher & Thompson (1979).

Likewise, the hydrophobic cluster surrounding the distal E7 His is formed by the same residues in the myoglobins of all three species: Leu (B10); Phe (B14); Phe (CD1), whose phenyl ring is oriented parallel to the haem plane; Phe (CD4); Val (E11); and Leu (B13). Interacting with the propionic acid of pyrole III is residue CD3 and with that of pyrole IV is residue FG2. These positions are occupied by Arg and His in the sperm whale, and by Lys and His in carp and shark myoglobins, respectively. The three haem contact residues F4, G5 and H15, which constitute the hydrophobic cluster around the proximal F8 His, similarly show a high degree of conservatism. In the sperm whale these are Leu, Leu and Phe, in the carp Leu, Phe and Val, and in the shark Leu, Phe and Ile, respectively.

The 38 internal positions of sperm whale myoglobin (Kendrew 1962; Watson 1969; Takano 1977*a*) are occupied by hydrophobic residues with the exception of five that are hydrogen-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

bonded or have special functions (B5, C4, E7, F8 and H23). During divergent evolution leading to the present shark, carp and sperm whale, 21 of these positions remained invariant, 16 mutated within a subset of hydrophobic residues (Leu, Val, Ile, Phe, Met, Ala and Cys) and one (B5) was not conserved (His in sperm whale, hydrophilic Thr in the carp and hydrophobic Val in the shark). Of the three internal residues with polar side chains in the sperm whale myoglobin, two are the same as in carp and shark: C4 Thr, a haem contact with its OH group hydrogen-bonded to the main chain carbonyl group of residue C1; and H23 Tyr, whose OH group is hydrogen-bonded to the main chain carbonyl group of residue FG4, thereby stabilizing the C-end of the H-helix. This hydrogen bond is of particular interest in the globin family because it becomes stronger in myoglobin during deoxygenation, whereas in haemoglobin this bond is broken under the same condition. In the sperm whale the third polar internal residue, B5 His, interacts with the imidazole group of GH1 His to provide stability to the GH loop. However, in the carp and shark, B5 is Thr and Val and GH1 is Ala and Tyr, respectively. A salt bridge important in stabilizing the A-helix in sperm whale myoglobin (A4 Glu to H10 Lys) has been highly conserved in both the carp (Asp and Arg) and the shark (Glu and Lys).

One final point of interest concerns observations made in an earlier publication dealing with the myoglobins of 31 tetrapods (Romero-Herrera *et al.* 1978). It was pointed out that the sequence from H10 to H16 constituted an invariant region in these species and that the site generated by the three-dimensional arrangement of positions C3, C6, CD2, CD5, CD8 and FG3 exhibited a high degree of conservatism. It was postulated that this latter region may function as a 'docking-site' for interaction either with a mitochondrial membrane receptor or with the enzyme metmyoglobin reductase. This tendency is not respected in the myoglobin of the fish presently considered: H10–H16 is at variance with the tetrapod sequence and the putative docking site exhibits some rather non-conservative substitutions. With reference to position H16 the presence of Cys in the shark sequence may well be significant when one considers that in all known tetrapod myoglobins Arg occupies this position, while Cys is further present at this site in the major haemoglobin fraction (V) of the lamprey *Petromyzon marinus* (Li & Riggs 1970) and, based upon homologous alignment of the enzymic peptides, in the myoglobin from this same species (Romero-Herrera *et al.* 1979). It is therefore likely that the ancestral globin sequence had Cys at this position.

(*b*) *Reassessment of myoglobin evolution by incorporating Lower Vertebrate sequences*

In addition to the aforementioned shark and carp myoglobin sequences, that of the yellowfin tuna has recently become available to us. (Access to this unpublished sequence was kindly provided by Dr W. Duane Brown and colleagues, University of California at Davis.) This information from non-amniote species provides a unique opportunity to reassess myoglobin evolution because previous evaluations were of necessity based largely on mammalian and bird sequences.

Analysis of sequence differences among myoglobins from extant species (by utilizing a phylogenetic tree strongly supported by other evidence and generating different sets of hypothetical ancestral protein chains) allows one to speculate on which lineages incorporated specific mutational events. Absolute rate calculations, whether determined simply on the basis of the number of differences between the sequences from present-day species or on the basis of reconstructed hypothetical ancestral chains, require one further set of information, namely the dates (millions of years ago) at which the various dichotomies in a cladogram are believed to have occurred.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

This section therefore addresses itself first to the dating of dichotomies, including the rationale in arriving at particular times, then to discussion of possible routes by which the myoglobin sequences of several phylogenetically strategic species arose, then to the rates of molecular change in myoglobin, and finally to the conclusions that analysis of these new sequences brings to bear on the concept of rate constancy and the 'neutral theory' of molecular evolution.

Dates of divergence

To determine dates of divergence we have used criteria similar to those employed by Romero-Herrera *et al.* (1978): on the basis of direct fossil evidence we have sought to establish best estimates of the time at which it can be demonstrated that the ancestors of two living forms had diverged. Because direct evidence demands that the event had already occurred, it follows that our dates are likely to represent minimum values. We believe that in most cases the actual date of divergence was earlier than the date that we have used, but we prefer not to speculate about earlier possible dates in the absence of good evidence. Dates are based on the Geological Society Phanerozoic Time-scale (1964), with modifications suggested in the supplement (1971).

Carp/shark. Among the Gnathostomata (jawed fish) the date of divergence of the Chondrichthyes (cartilaginous fish, so including the sharks) and the Osteichthyes (bony fish, so including both tuna and carp) cannot be considered apart from their possible relationships to two extinct groups of fish, namely the Placodermi and the Acanthodii. This is so because it is widely held, but not universally accepted, that the ancestry of the chondrichthyans may lie among the placoderms and it has been suggested that the acanthodians may be related either to the cartilaginous fish or to the bony fish. It so happens that the fossil record of the acanthodians starts before that of the placoderms and so if it can be shown that the acanthodians shared a common ancestry with one group more recently than with the other, then the presence of acanthodians can be taken to indicate that the chondrichthyans and osteichthyans had already diverged. Various hypotheses of the interrelationships of acanthodians, chondrichthyans and osteichthyans have been reviewed by Miles (1973), who concluded that '*Acanthodes* and primitive osteichthyans exhibit a far reaching mutual resemblance in the skull, which is not shared by chondrichthyans. This resemblance cannot be explained easily by recourse to primitive gnathostome characters, it appears to be phylogenetically significant'. Although Miles found that the evidence was compatible with a relationship between the acanthodians and the osteichthyans, he cautiously suggested that the evidence was not yet sufficient to falsify the hypothesis of a relationship between acanthodians and chondrichthyans. Nevertheless, he took the step of recommending 'that acanthodians and osteichthyans should be classified together as teleostomes'. We have decided to accept his recommendation, which is reinforced by Gardiner's (1973) statement that: 'The osteichthyans, although quite distinct, must have shared a common ancestry with the acanthodians, and are their sister group'. On this basis the date of the earliest acanthodians can be taken as indicative that the ancestors of the chondrichthyans and osteichthyans had already diverged. Romero-Herrera *et al.* (1978) have previously used this same event to date the divergence between penguin and lamprey, because the earliest recorded acanthodian is also the earliest known gnathostome, namely *Nostolepis* from the Llandovery/Wenlock of Czechoslovakia (Miles 1967), which is equivalent to a date of about 420 Ma ago. If it should be felt that the evidence relating to acanthodians and osteichthyans is insufficient, then by similar criteria we would feel that the evidence linking placoderms and chondrichthyans would also have to be regarded as insufficient; under such circumstances the divergence of chondrichthyans and osteichthyans would have to be dated by reference to the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

first undoubted member of either group, and the date would then become coincident with the next point of divergence discussed below.

Carp/tetrapods. If one accepts the conclusions of Miles (1973) and Gardiner (1973) that the Osteichthyes are a monophyletic group, and working on the basis that Actinopterygii (ray-finned fish) were not themselves ancestral to the tetrapods, then the point of divergence required is that between the actinopterygians and whichever other osteichthyan group was ancestral to the tetrapods. It is contended as to whether the tetrapods were derived from the Crossopterygii (tassel-finned fish), the Dipnoi (lung fish), or both. It so happens that the earliest known crossopterygian (*Porolepis*) and dipnoan (*Dipnorhynchus*) are both recorded from the Siegenian division of the Devonian (Andrews 1967), and so it is unnecessary for us to make a choice between conflicting views on tetrapod ancestry. The presence of both of these groups, somewhat before the earliest recorded actinopterygian, indicates that the divergence between both putative ancestors of the tetrapods and the actinopterygians had already occurred 382 Ma before the present; it is pointless to speculate how much earlier the actual divergence might have occurred because no earlier osteichthyans are known.

Mammals/birds. This point of divergence, dated at 293 Ma ago, was discussed by Romero-Herrera *et al*. (1978, p. 138) under the heading *Opossum/chicken*; we are unaware of any additional evidence and our opinion remains unchanged. Nevertheless, our argument was based on the assumption that the Diapsida are a monophyletic group, which must be assumed in the absence of evidence to the contrary. In case this assumption should be incorrect it seems fair to indicate an alternative, completely substantial, minimum date. While it is conceivable that early synapsids might have given rise to early diapsids, and perhaps more than once, it is inconceivable that the therapsids might be ancestral to archosaurs. Hence, the earliest therapsids, recorded from the Permian (Guadelupian), may be used to provide a most conservative minimum date of 250 Ma ago, which has been included in figure 6 in addition to our preferred date of 293 Ma ago. (We have here accepted that birds arose from an archosaur reptile.)

Placentals/marsupials. This divergence was discussed by Romero-Herrera *et al*. (1978, p. 137), under the heading *Badger/kangaroo*. The date of 79 Ma ago was based on the undoubted eutherian affinity of *Zalambdalestes* and *Kennalestes* from the Djadoktha Formation of Bayn Dzak, Mongolia. The age of these deposits remains in doubt; the date of 79 Ma ago reflects a Santonian age, but they may be slightly younger (see footnote on page 252 of Crompton & Kielan-Jaworowska (1978)). It has been represented to us that 79 Ma ago is too conservative as a minimum date and that the divergence probably occurred considerably earlier. We chose 79 Ma ago in conformity with our policy of requiring hard fossil evidence rather than speculation, but we also drew attention to the undescribed forms known as *Prokennalestes* and *Prozalambdalestes*. As these forms remain undescribed we are not in a position to reach an opinion on their affinities but, for the purposes of the present paper, we have decided to adopt the views of Kielan-Jaworowska (1979), who refers to *Prokennalestes* as 'the most primitive of known eutherian mammals'. She also states that, on dental morphological grounds, it is possible 'that all groups of eutherians may be derived from such a creature'. Feeling that we should not venture beyond that, we have here adopted the date of 109 Ma ago, which was discussed in our previous paper but not used. In choosing 109 rather than 79 Ma ago we have detracted from the strength of one of our observations, but we would not wish it to be thought that this observation might be an artefact consequent upon our using the younger date.

Sperm whale/tree shrew. This divergence was discussed by Romero-Herrera *et al*. (1978, p. 136)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

under the heading *Sheep/tree shrew*, which is equivalent to condylarthran/tree shrew. The orders Artiodactyla (including sheep) and Cetacea (including sperm whale) are both believed to be derived from condylarthrans. This is an unsatisfactory date because the phylogenetic position of the tree shrew is not known; it is here assumed that the condylarthran/tree shrew divergence was earlier than the condylarthran/primate divergence. The date of 68 Ma ago represents the latter divergence and therefore provides a minimum date for the former divergence, for which there is no useful fossil evidence.

Kangaroo/opossum. The fossil record does not furnish a useful date of divergence for these species because the earliest mammal fauna recorded in Australia indicates that considerable radiation of Australian marsupials had already occurred. The date at which effective separation of the Australian and South American continents took place might shed some light on this split, but such indirect dating is too speculative to be considered alongside the other dates used in this study.

Chicken/penguin. This divergence was discussed by Romero-Herrera *et al.* (1978, p. 138), when we stated that 52 Ma ago was the best available date and that we were 'very conscious that it might have taken place much earlier'. To emphasize the unsatisfactory nature of this date we draw attention to a speculative phylogenetic chart by Fisher (1967) which indicates a date of divergence between the ancestors of the chicken and penguin at 132 Ma ago. This date is not supported by direct fossil evidence, but the 80 Ma difference between our minimum date and a possible date indicates the latitude available for speculation. Although we have refused to embark on such speculation we must admit, in this case, that ancestral penguins might one day be found anywhere within this range.

Carp/tuna. Among the Actinopterygii the carp and tuna are both included within the Euteleostei, and within this group it is generally accepted that the Ostariophysi (a monophyletic group, including carp, characins and catfish, characterized by the presence of a chain of Weberian ossicles) were derived from among the Protacanthopterygii. Greenwood *et al.* (1966) drew attention to various similarities between the protacanthopterygian order Gonorynchiformes and the Ostariophysi that suggest their derivation from a common stem. This hypothesis was developed further by Rosen & Greenwood (1970), who concluded that the Gonorynchiformes are the sister group of the Ostariophysi, so implying that the two groups shared an immediate common ancestor. In accordance with this view they enlarged the Ostariophysi to include the Gonorynchiformes, so giving the term Ostariophysi a new meaning. The view of Rosen & Greenwood (1970) on the relationship of these two groups has been questioned by Roberts (1973), but appears to have been accepted by Patterson (1975) and Briggs (1979). We are impressed by the evidence that the Gonorynchiformes and the Ostariophysi (old sense) together constitute a monophyletic group. On this basis, the earliest record of any member of either of these groups can be taken to provide evidence that the Ostariophysi (*sensu* Rosen & Greenwood 1970), including the traditional Ostariophysi, had diverged from the rest of the Euteleostei. Patterson (1975) has reviewed the fossil record of these groups; gonorynchiform fish occur in 'Wealden' non-marine deposits in West Africa and in Aptian(?) non-marine deposits in Brazil. These and several other records of gonorynchiforms are earlier than the first records of ostariophysans (old sense), which occur in the Upper Cretaceous of Bolivia. There is considerable uncertainty regarding the stratigraphical correlation of the 'Wealden' non-marine deposits of West Africa; they might be dated anywhere between 110 and 130 Ma ago, and so we have taken 120 Ma ago as the best estimate of this date. In this context we have studied

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Brigg's (1979) review of ostariophysan zoogeography; so far as this group is concerned it well illustrates the highly speculative nature of this line of evidence, which so often detracts from its potential value as a basis for hypotheses regarding dates of divergence.

Pathways of mutation and rates of evolution

Evaluation of the large number of known myoglobin sequences has made apparent two different rates of amino acid substitution: one involving external residues, the other affecting the internal and/or haem contact residues. The greater magnitude of the former over the latter is apparently due to the constraints imposed upon the molecule in conserving the vitally important hydrophobic pocket that maintains haem function. Because of the difficulty in reconstructing mutational pathways when dealing with positions that show extremely high rates of change we have decided to first investigate those 44 amino acid positions defined as internal and/or haem contact residues (Kendrew 1962; Watson 1969; Takano 1977*a*, *b*).

Utilizing the dates of divergence and the myoglobin sequences from a set of species whose phylogenetic relationships with one another are not disputed, we have investigated, by the estimation of ancestral chains, the phenomenon of rate of fixation of mutations in the subset of internal and/or haem contact positions (figure 5*a*). We then tested the effect that a different pattern of mutational events has upon the rates by constructing an alternative allocation of these events (figure 5*b*).

In figure 5*a*, *b* respectively, 77 and 79 amino acid substitutions involving these positions were assigned to the branches of the cladograms. Ten species composed the phylogenetic tree: sperm whale; tree shrew (Romero-Herrera & Lehmann 1974*b*); opossum (Romero-Herrera & Lehmann 1975); kangaroo (Air & Thompson 1971); penguin (Peiffer 1973); chicken (Deconinck *et al.* 1975); yellowfin tuna; carp; shark; and the sea hare, *Aplysia*. The cladogram in figure 5*a* incorporated 19 double mutations and that in figure 5*b* utilized 17 doubles, thereby making the total number of nucleotide substitutions 96 in each.

Due to problems inherent in hypothesizing the ancestral sequence of a protein for any given cladistic group, the allocation of mutations in the present consideration to either the *Aplysia* stem or the vertebrate stem was difficult. We chose to assign all possible changes to the *Aplysia* lineage because this stem served only to root the phylogenetic tree and was thereafter not relevant to the discussion. While the two cladograms are identical in terms of topology and the total number of nucleotide substitutions, they differ at ten positions with respect to the choice of ancestral residues or codons, thereby altering the pathway of mutation along the various lineages. This allows one to test the validity of the rate information derived from the alternative pattern of allocations.

It is apparent from figure 5 that, for the particular set of positions and species considered, the rate encompasses a very wide range of values: from one per 10 Ma (tetrapod stem) to zero per 184 Ma (mammalian stem) in figure 5*a* and from one per 7 Ma (tetrapod stem) to zero per 241 Ma (avian stem) in figure 5*b*.

In our previous publication, in which 31 myoglobins were considered (Romero-Herrera *et al.* 1978) we found a remarkably low rate of evolution on the stems leading to the mammals and birds subsequent to their divergence approximately 293 Ma ago. Due to the lack of myoglobin sequences representing the lower vertebrates, lamprey (*Petromyzon marinus*) haemoglobin rather than myoglobin was utilized in comparisons for the more ancient lineages.

In the present analysis the introduction of the fish myoglobins has permitted an evaluation

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

of the rates of evolution between 420 Ma ago (divergence of the shark) and 382 Ma ago (divergence of the actinopterygians) and 293 Ma ago (divergence of the lineages leading to modern birds and mammals). The rate of myoglobin evolution involving the internal and/or haem contact residues was found to be relatively high in the 89 Ma period between 382 Ma and 293 Ma ago: the cladogram in figure 5*a* shows an average rate of one fixed mutation per *ca.* 10 Ma along this lineage, whereas the cladogram in figure 5*b* reveals one fixed mutation per *ca.* 7 Ma. This high-rate period includes the time during which the evolving tetrapods were invading land. It is here suggested that this might also have coincided with the appearance of myoglobin in muscles other than just the heart and/or a few muscles of sustained propulsion; this appearance may be related to the postural requirements imposed upon skeletal muscle in counteracting the increased effects of gravity on land. Following this high rate period there was a lower average rate of fixation of mutations, for this subset of residues, during the 293 Ma since the divergence of bird and mammal ancestors: one fixed mutation per 88 Ma (cladogram, figure 5*a*), or one fixed mutation per 65 Ma (cladogram, figure 5*b*) for the six modern species considered.

Such observations are consistent with the hypothesis that high replacement rates at such positions may be correlated with the development of new functional roles which, once acquired, tend to be maintained by selection. Furthermore, although it appears that the present-day species have fixed an approximately equal number of mutations when the total time interval from a very distant common ancestor is considered, the rates are seen to fluctuate widely when the component time periods along the respective lineages from that ancient dichotomy are considered.

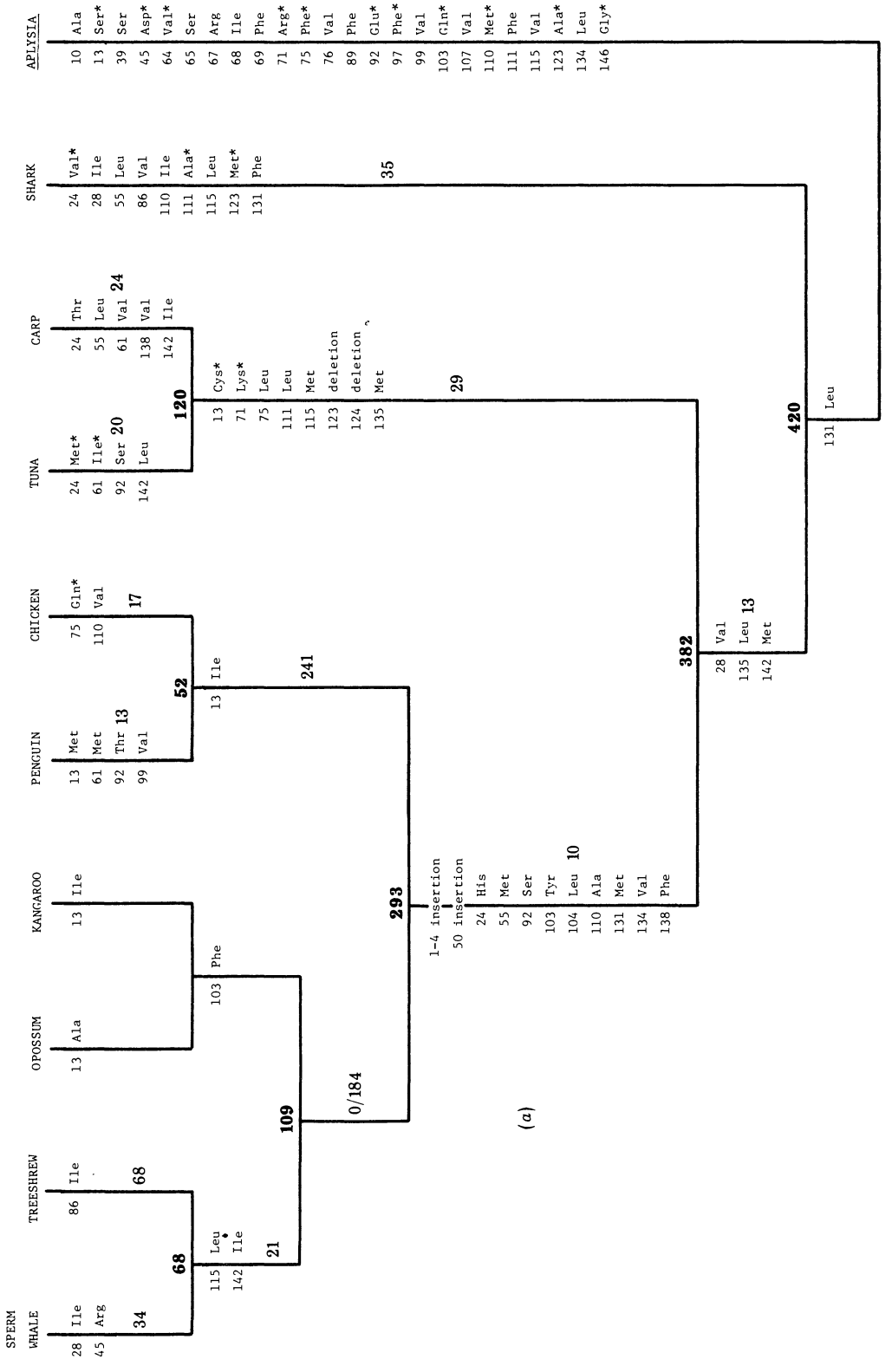
Testing Kimura's concept of rate constancy

Returning to a consideration of the whole molecule, any discussion of protein evolution with its necessary emphasis on the differences between amino acid sequences naturally raises the question of adaptive significance of the changes implied. A carefully reasoned but controversial theory advanced by Kimura holds that the majority of the mutational events fixed into a protein's primary structure are adaptively neither more nor less advantageous than the original (see Kimura (1968, 1979) and Kimura & Ohta (1971) for complete descriptions). This view

FIGURE 5. (*a*) Cladogram depicting one possible pathway of mutations that occurred during the evolution of myoglobin along the different branches leading to present-day species with respect to the 44 internal and/or haem contact residues. The introduction of *Aplysia* serves only for rooting the tree, and thus all possible mutations in the branches of the first dichotomy have been allocated to this mollusc. Asterisks (*) indicate those mutations that required two base substitutions. The insertions at positions 1–4 and 50 in the common ancestor of birds and mammals and the deletion at position 124 in the actinopterygian stem have not been considered because they are external positions. The boldface numbers indicate minimum dates of divergence (see text), but branch lengths are not proportional to time. The numbers to the right of each branch represent the *average* rate of fixation of mutations in terms of millions of years per detected base substitution.

(*b*) An alternative cladogram to that in (*a*). At ten positions (13, 28, 61, 75, 92, 110, 111, 123, 131 and 142) a different ancestral residue or codon has been chosen (see bottom sequence). Consequently, the allocation of mutations along the different stems and the resultant average rates differ from the first cladogram. The designations for this cladogram are the same as given for figure 5*a*. To avoid a triple hit at 75 Gln on the chicken stem, 75 Ile of the vertebrate common stem (AUU/AUC) has been changed into 75 Ile (AUA) on the tetrapod stem. Similarly, to avoid a double hit at 142 Ile on the carp stem, 142 Leu (UUG/CUG) of the common ancestor has been changed into 142 Leu (UUA/CUA) along the actinopterygian common stem.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



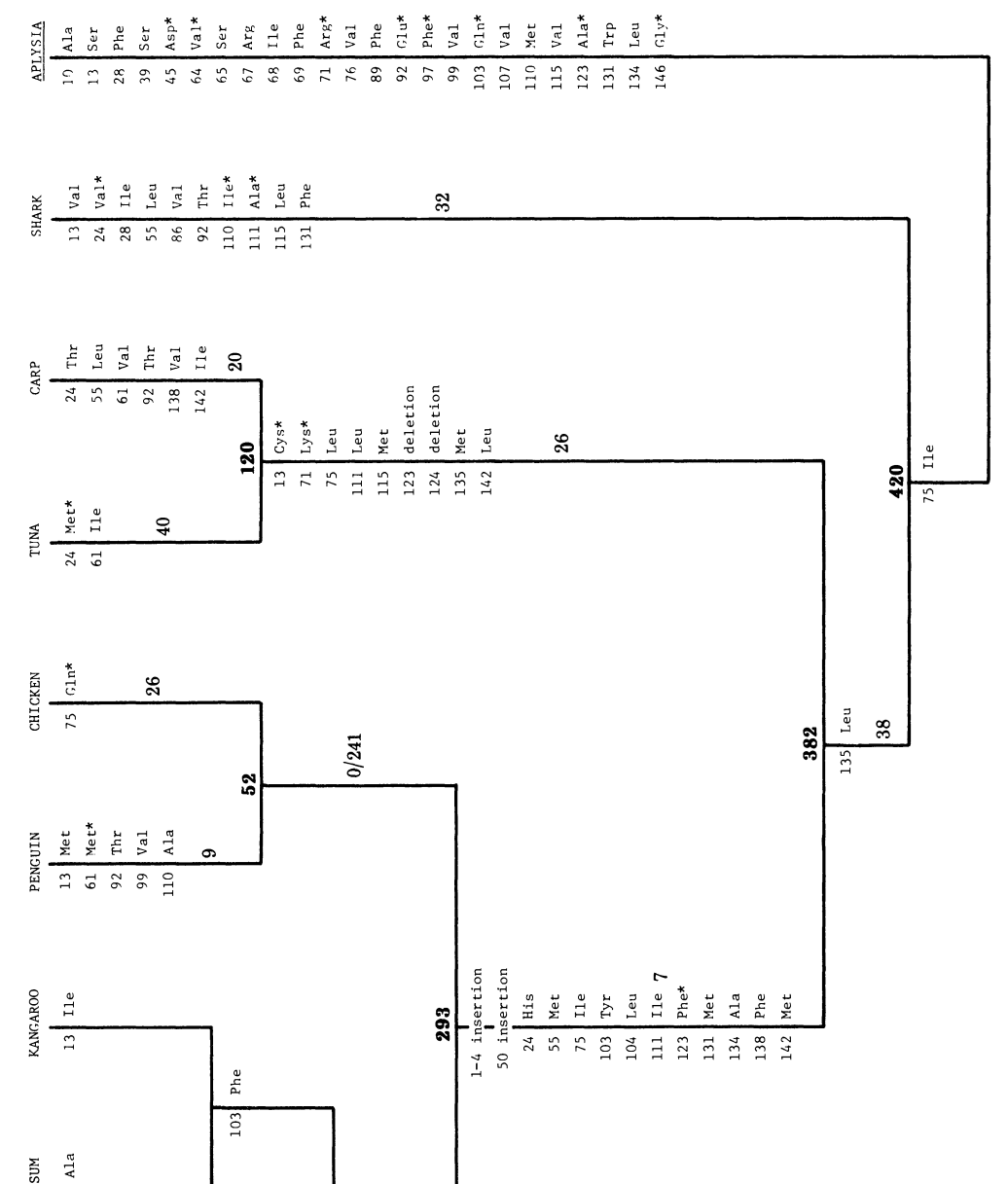


FIGURE 5. For description see p. 17.

has come to be known as the 'neutral theory' of molecular evolution and is based on three fundamental interpretations of molecular studies: (1) the rates of molecular evolution are too high to be accommodated by conventional estimates of the 'cost of selection' (Haldane 1957); (2) where rates of substitution are high, the changes observed in molecules are apparently random (within the constraints of the genetic code); and (3) the rates of evolution for a single molecule measured over a variety of species are constant within the statistical limits expected for a stochastic process. While these points are more difficult to explain if the majority of fixed mutations are not selectively neutral (or slightly deleterious), several investigators (Smith 1968; Sved *et al.* 1967; Lewontin 1974; Holmquist 1979) have addressed themselves to Kimura's arguments concerning the 'cost of selection' and the randomness of amino acid substitution. We shall confine our comments here to the third point, that of constant rate.

Despite the fact that this postulate is only a part of the 'neutral theory', it is an important one. Under a selective model, which is usually regarded as predicting fluctuating rates of change, one can envisage circumstances at times leading to approximate constancy of rate (Van Valen 1973; Blundell & Wood 1975; Hartl & Dykhuizen 1979). However, it is more difficult to see how a neutralist hypothesis could readily accommodate non-constancy of rate. Kimura justifiably claims the vulnerability of the 'neutral theory' as a point in its favour: it is in principle falsifiable, whereas testing the selective model is logically difficult. Furthermore, it should be emphasized that Kimura (1979) appears to claim constancy not only for the rate of amino acid substitution (the process that he has always emphasized in his rate studies), but also for the rate of nucleotide substitution (on the basis of work by Langley & Fitch (1974)). In view of what is known about the relation between the nucleotide and amino acid levels, it is difficult to reconcile these claims, a point already made by Holmquist (1972) with respect to the observations of Wilson & Sarich (1969).

The primary structures of the proteins of living species are the result of an evolutionary process in which the organisms share varying degrees of common ancestry with one another. The extent of shared ancestry clearly depends on the particular pattern of evolutionary history, and we take it as axiomatic that there is a single true pattern of kinships. Use of protein sequences to establish the true pattern of evolutionary relationships is now recognized as a much more difficult exercise than was formerly believed (Romero-Herrera *et al.* 1978; Romero-Herrera *et al.* 1979). Nevertheless, it has become evident from this approach that calculations of mutational rates by simple comparisons of protein sequences may not reveal actual rate fluctuations because of the unavoidable incorporation of identical mutational events (due to common ancestry) on the shared parts of an evolutionary tree. This becomes a problem especially when pairwise comparisons between many distant species on a cladogram are made. In these circumstances the observations are not independent and the effect will be to homogenize the data. Use of the time intervals represented by the individual branch lengths of a cladogram having dated points of dichotomy overcomes this objection (Uzzell & Corbin 1972; Romero-Herrera *et al.* 1973, 1978), but the events assigned to the branches depend on the particular pattern of relationships chosen.

Stebbins & Lewontin (1973) and Romero-Herrera *et al.* (1973) both suggested that, in comparing rates calculated over long periods of time, Kimura was mistaking an average for a constant. Local constancy of rate over small sections of an evolutionary tree is in fact sufficient for some tactics of data analysis and tree construction, and Jardine *et al.* (1969) in advocating such methods have argued that the phenomenon of local constancy is biologically plausible. In

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

this context it is significant that protein evolution seems to exhibit the reverse, i.e. exceedingly different local rates, but rather similar values in comparisons between distantly related lineages (Romero-Herrera *et al.* 1978). Thus, while the relevant values are dependent upon the pattern of evolutionary history reconstructed, this observation in general supports the criticism that Kimura's 'constancy' is achieved, at least in part, by the averaging of locally fluctuating rates of change. Kimura (1979) has rebutted the criticism on the grounds that there is no reason to expect such a high degree of similarity between the average rates from different lineages unless the rate were in fact constant for the particular protein throughout the entire evolutionary tree under consideration.

TABLE 2. PAIRWISE COMPARISON OF MYOGLOBIN SEQUENCES BY KIMURA'S TECHNIQUE

	N	$\bar{x}$	s.d.	p_d	K_{aa}	s.e.	T	$10^{10}k_{aa}$
mammals/birds	90	41.40	4.53	0.271	0.316	0.0492	293	5.38
tuna/carp	1	41	—	0.281	0.330	0.0517	120	13.74
tetrapods/actinopterygians	94	83.18	3.03	0.570	0.843	0.0952	382	11.04
all others/chondrichthyans	49	84.67	1.88	0.572	0.849	0.0950	420	10.11

Symbols and abbreviations used in this table: N, number of sequence cross comparisons; $\bar{x}$, mean amino acid differences; s.d., standard deviation on amino acid differences; p_d, from Kimura & Ohta (1971); K_{aa}, from Kimura & Ohta (1971); s.e., standard error on K_{aa}; T, divergence time; k_{aa}, as Kimura & Ohta (1971), final rate change per amino acid site per year.

Wilson *et al.* (1977) have commented on examples of apparently non-constant rates of molecular evolution and Lessios (1979) has used electrophoretic data from Panamanian sea urchins to claim some 20-fold difference in amounts of molecular change over similar periods of time. This finding, however, cannot so directly be used in disputing constancy because Lessios's data are not strictly comparable with those of Kimura.

While the body of evidence presented thus far appears to refute Kimura's contention of rate constancy, one final exercise seems most illustrative: determining rates by applying Kimura's own computational methods to the known vertebrate myoglobin sequences. The presence in this set of the carp, tuna and shark data reinforces the definitive nature of any such calculation because of the breadth of vertebrate phylogeny and divergence time thereby represented. Although a variety of statistical transformations exist for treating raw amino acid difference data, we restrict ourselves here to the simple Poisson correction of Zuckerkandl & Pauling (1965) because it was the one used by Kimura & Ohta (1971) in their own calculations.

All vertebrate species for which myoglobin sequences are known can be assigned to one or other of the groups listed in table 2. The evolutionary interrelationships of these groups are well established. Therefore, the criticism that we have biased our calculations by using a questionable set of opinions concerning the relationships of these groups cannot easily be entertained. The dates of divergence for the four dichotomies represented by these groups have been previously discussed. Following Kimura's approach, the myoglobin sequences were aligned and pairwise comparisons were made in terms of the number of amino acid differences. These values were divided by the number of comparable positions to give the proportion p_d of Kimura & Ohta (1971).

Corrected values (K_{aa}) were obtained from the equation $K_{aa} = -\log_e(1-p_d)$ (Kimura & Ohta 1971, p. 18, eqn 2).

Standard errors for the K_{aa} values are given by $\sigma_K = p_d/(1-p_d)\,N_{aa}$ (Kimura & Ohta 1971, p. 19, eqn 3), where N_{aa} is the number of comparable positions.

The final values of rates of amino acid substitution (k_{aa}) were obtained from $k_{aa} = \frac{1}{2}K_{aa}/T$ (Kimura & Ohta 1971, p. 19, eqn 4), where T is the relevant time of divergence.

The results of these calculations are presented in table 2. By plotting K_{aa} values against corresponding times of divergence, the same information can be graphically represented (figure 6). Since each point is provided with a vertical bar representing two standard errors either side of the mean, the 'discrepancy' in rate between the bird–mammal comparison and

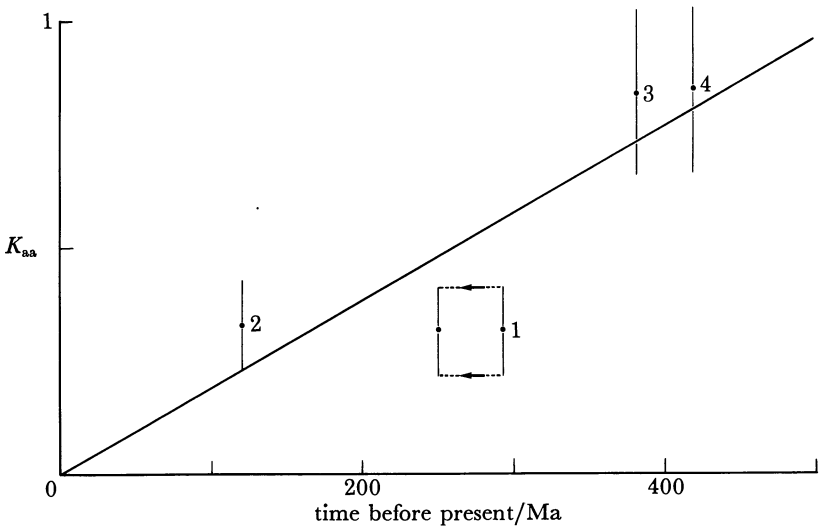


FIGURE 6. The values of K_{aa} (Kimura & Ohta 1971) from table 2 plotted against corresponding estimates of dates of divergence. A least squares fitted line is drawn through the origin. Vertical bars give two standard errors either side of the mean (see table 2). The points represent comparisons between: (1) mammals and birds; (2) tuna and carp; (3) tetrapods and actinopterygians; (4) all others and chondrichthyans. The dotted lines relative to point 1 indicate the displacement when the most conservative date (250 Ma age) is used for this dichotomy.

the other three groups is especially pronounced. We have commented on the low rate of fixation of mutations during the early ancestry of both birds and mammals in this paper and in a previous publication (Romero-Herrera *et al.* 1978). The availability of a comparison between two bony fish reinforces this point because the fish species differ by about the same number of amino acid differences as the average between mammals and birds, yet the dating of the cladogram indicates that very different spans of time are believed to have been involved. To date the mammal/bird divergence on the basis of the amount of molecular change compared with rates averaged from the rest of the cladogram would yield a questionable value. If the Poisson correction is undercompensating for undetected events, this applies equally to the almost identical values of amino acid differences between the mammals and birds on one hand and the two bony fish on the other.

An analysis of the average K_{aa} distances between the groups under consideration using the distance Wagner technique of Farris (1972) is shown in figure 7. Table 3 gives the intergroup distances used as the input data and the extra steps inserted by the program in building the network. The input data strongly determine the network pattern and there is no reason to

suspect that the low rate of change recorded between mammals and birds is largely because of the nature of the bird sequences, of which there are only two compared to over 40 for mammals. The network distances reinforce the conclusion that after the divergence of the ancestors of mammals and birds both lineages underwent relatively few changes.

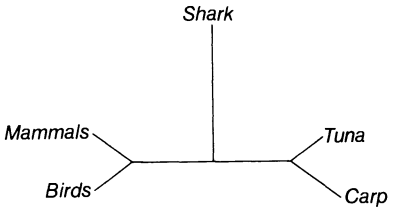


FIGURE 7. A Wagner network based on the input matrix shown in table 3. The branch lengths are drawn to scale.

TABLE 3. INPUT MATRIX FOR WAGNER DISTANCE PROGRAM

	mammals	birds	tuna	carp	shark
mammals		0	0	0	0.024
birds	0.316		0	0.045	0
tuna	0.807	0.802		0	0
carp	0.883	0.833	0.330		0.076
shark	0.851	0.870	0.840	0.840	

Below diagonal: K_{aa} values (Kimura & Ohta 1971).
Above diagonal: number of extra steps introduced between pairs of animal taxa by Wagner distance program.

In conclusion, despite our belief that rate determination utilizing average rates over long periods of time may not be the best way to investigate this aspect of molecular evolution, it is revealing that if one uses this approach the results nonetheless appear to refute a hypothesis of constant rate of change.

We are grateful to Dr K. S. Bannister for helpful discussion of ostariophysan zoogeography. Support for this project was provided by grants DEB 7619924 and PCM 7717644 from the National Science Foundation, Washington, D.C. (U.S.A.).

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.

Andrews, S. M. 1967 Class Crossopterygii and class Dipnoi. In *The fossil record* (ed. W. B. Harland *et al.*), pp. 641–644. London: Geological Society.

Blundell, T. L. & Wood, S. P. 1975 Is the evolution of insulin Darwinian or due to selectively neutral mutation? *Nature, Lond.* **257**, 197–203.

Bonner, A. G. & Laursen, R. A. 1977 The amino acid sequence of a dimeric myoglobin from the gastropod mollusc, *Busycon canaliculatum*. *FEBS Lett.* **73**, 201–203.

Briggs, J. C. 1979 Ostariophysan zoogeography: an alternative hypothesis. *Copeia*, pp. 111–118.

Crompton, A. W. & Kielan-Jaworowska, Z. 1978 Molar structure and occlusion in Cretaceous therian mammals. In *Development, function and evolution of teeth* (ed. P. M. Butler & K. A. Joysey), pp. 249–287. London: Academic Press.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Crowshaw, K., Jessup, S. J. & Ramwell, P. W. 1967 Thin-layer chromatography of 1-dimethylaminonaphthalene-5-sulphonyl derivatives of amino acids present in superfusates of cat cerebral cortex. *Biochem. J.* **103**, 79–85.

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.

Dixon, H. B. F. 1972 A reaction of glucose with peptides. *Biochem. J.* **129**, 203–208.

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

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

Fisher, J. 1967 Fossil birds and their adaptive radiation. In *The fossil record* (ed. W. B. Harland *et al.*), pp. 133–154. London: Geological Society.

Fisher, W. K. & Thompson, E. O. P. 1979 Myoglobin of the shark *Heterodontus portusjacksoni*: isolation and amino acid sequence. *Aust. J. biol. Sci.* **32**, 277–294.

Gardiner, B. G. 1973 Interrelationships of teleostomes. In *Interrelationships of fishes* (ed. P. H. Greenwood, R. S. Miles & C. Patterson), pp. 105–135. London and New York: Academic Press.

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

Geological Society Phanerozoic Time-scale – a supplement 1971 *Spec. Publs geol. Soc. Lond.*, no. 5.

Gray, W. R. 1972 End-group analysis using dansyl-chloride. *Methods Enzymol.* **25**, 121–138.

Greenwood, P. H., Rosen, D. E., Weitzman, S. H. & Myers, G. S. 1966 Phyletic studies of teleostean fishes, with a provisional classification of living forms. *Bull. Am. Mus. nat. Hist.* **131**, 339–456.

Gross, E. & Witkop, B. 1962 Nonenzymatic cleavage of peptide bonds: the methionine residues in bovine pancreatic ribonuclease. *J. biol. Chem.* **237**, 1856–1860.

Haldane, J. B. S. 1957 The cost of natural selection. *J. Genet.* **55**, 511–524.

Hartl, D. & Dykhuizen, D. 1979 A selectively driven molecular clock. *Nature, Lond.* **281**, 230–231.

Hartley, B. S. 1970 Strategy and tactics in protein chemistry. *Biochem. J.* **119**, 805–822.

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

Holmquist, R. 1979 β-Galactosidase and selective neutrality. *Science, N.Y.* **203**, 1012–1014.

Houmard, J. & Drapeau, G. R. 1972 Staphylococcal protease: a proteolytic enzyme specific for glutamoyl bonds. *Proc. natn Acad. Sci. U.S.A.* **69**, 3506–3509.

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

Jones, R. T. 1964 Structural studies of aminoethylated hemoglobins by automatic peptide chromatography. *Cold Spring Harb. Symp. quant. Biol.* **29**, 297–308.

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

Kielan-Jaworowska, Z. 1979 Results of the Polish–Mongolian Palaeontological Expeditions. Part VIII. *Paleont. Ziem Pol.* **38**, 5–41.

Kimura, M. 1968 Evolutionary rate at the molecular level. *Nature, Lond.* **217**, 624–626.

Kimura, M. 1979 The neutral theory of molecular evolution. *Scient. Am.* **241**, 94–104.

Kimura, M. & Ohta, T. 1971 *Theoretical aspects of population genetics*. Princeton: University Press.

Langley, C. H. & Fitch, W. M. 1974 An examination of the constancy of the rate of molecular evolution. *J. molec. Evol.* **3**, 161–177.

Lessios, H. A. 1979 Use of Panamanian sea urchins to test the molecular clock. *Nature, Lond.* **280**, 599–601.

Lewontin, R. C. 1974 *The genetic basis of evolutionary change*. Columbia: University Press.

Li, S. L. & Riggs, A. 1970 The amino acid sequence of hemoglobin V from the lamprey *Petromyzon marinus*. *J. biol. Chem.* **245**, 6149–6169.

Lorkin, P. A., Charlesworth, D., Lehmann, H., Rahbar, S., Tuchinda, S. & Lie-Injo Luang Eng 1970 Two haemoglobins Q, α74 (EF3) and α75 (EF4) aspartic acid ⟶ histidine. *Br. J. Haemat.* **19**, 117–125.

Miles, R. S. 1967 Class Acanthodii. In *The fossil record* (ed. W. B. Harland *et al.*), pp. 637–640. London: Geological Society.

Miles, R. S. 1973 Relationships of acanthodians. In *Interrelationships of fishes* (ed. P. H. Greenwood, R. S. Miles & C. Patterson), pp. 63–103. London and New York: Academic Press.

Offord, R. E. 1966 Electrophoretic mobilities of peptides on paper and their use in determination of amide groups. *Nature, Lond.* **211**, 591–593.

Patterson, C. 1975 The distribution of Mesozoic freshwater fishes. *Mém. Mus. natn. Hist. nat., Paris* A **88**, 156–174.

Peiffer, S. 1973 La myoglobine du manchot empereur (*Aptenodytes forsteri*). Doctoral thesis, Université libre de Bruxelles, Faculté des Sciences.

Porzio, M. A. & Pearson, A. M. 1977 Improved resolution of myofibril proteins with sodium dodecyl sulfate–polyacrylamide gel electrophoresis. *Biochem. biophys. Acta* **490**, 27–34.

Roberts, T. R. 1973 Interrelationships of ostariophysans. In *Interrelationships of fishes* (ed. P. H. Greenwood, R. S. Miles & C. Patterson), pp. 373–395. London and New York: Academic Press.

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

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

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Romero-Herrera, A. E. & Lehmann, H. 1975 The primary structure of the myoglobin of *Didelphis marsupialis* (Virginia opossum). *Biochim. biophys. Acta* **400**, 387–398.
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.
Romero-Herrera, A. E., Lehmann, H., Joysey, K. A. & Friday, A. E. 1978 On the evolution of myoglobin. *Phil. Trans. R. Soc. Lond.* B **283**, 61–163.
Romero-Herrera, A. E., Lieska, N., Goodman, M. & Simons, E. L. 1979 The use of amino acid sequence analysis in assessing evolution. *Biochimie* **61**, 767–779.
Romero-Herrera, A. E., Lieska, N. & Nasser, S. 1979 Characterization of the myoglobin of the lamprey *Petromyzon marinus*. *J. molec. Evol.* **14**, 259–266.
Rosen, D. E. & Greenwood, P. H. 1970 Origin of the Weberian apparatus and the relationships of the ostariophysan and gonorynchiform fishes. *Am. Mus. Novit.* **2428**, 1–25.
Sick, K., Beale, D., Irvine, D., Lehmann, H., Goodall, P. T. & MacDougall, S. 1967 Haemoglobin G Copenhagen and haemoglobin J Cambridge. Two new β chain variants of haemoglobin A. *Biochim. biophys. Acta* **140**, 231–242.
Smith, J. Maynard 1968 'Haldane's dilemma' and the rate of evolution. *Nature, Lond.* **219**, 1114–1116.
Stebbins, G. L. & Lewontin, R. C. 1973 Comparative evolution at the levels of molecules, organisms and populations. *Proc. 6th Berkeley Symp. math. stat. Prob.*, vol. 6 (ed. L. LeCam, J. Neyman & E. L. Scott), pp. 23–42. Berkeley: University of California Press.
Sved, J. A., Reed, T. E. & Bodmer, W. F. 1967 The number of balanced polymorphisms that can be maintained in a natural population. *Genetics, Princeton* **55**, 469–481.
Takano, T. 1977*a* Structure of myoglobin refined at 2.0 Å resolution. I. Crystallographic refinement of metmyoglobin from sperm whale. *J. molec. Biol.* **110**, 537–568.
Takano, T. 1977*b* Structure of myoglobin refined at 2.0 Å resolution. II. Structure of deoxymyoglobin from sperm whale. *J. molec. Biol.* **110**, 569–584.
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.
Uzzell, T. & Corbin, K. W. 1972 Evolutionary rates in cistrons specifying mammalian hemoglobin α- and β-chains: phenetic versus patristic measurements. *Am. Nat.* **106**, 555–573.
Van Valen, L. 1973 A new evolutionary law. *Evolut. Theory* **1**, 1–30.
Watson, H. C. 1969 The stereochemistry of the protein myoglobin. *Prog. Stereochem.* **4**, 299–333.
Wilson, A. C. & Sarich, V. M. 1969 A molecular time scale for human evolution. *Proc. natn. Acad. Sci. U.S.A.* **63**, 1088–1093.
Wilson, A. C., Carlson, S. S. & White, T. J. 1977 Biochemical evolution. *A. Rev. Biochem.* **46**, 573–639.
Woods, K. R. & Wang, K. T. 1967 Separation of dansyl amino acids by polyamide layer chromatography. *Biochim. biophys. Acta* **133**, 369–370.
Zuckerkandl, E. & Pauling, L. 1965 Evolutionary divergence and convergence in proteins. In *Evolving genes and proteins* (ed. V. Bryson & H. J. Vogel), pp. 97–166. New York: Academic Press.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES