CMLS, Cell. Mol. Life Sci. 54 (1998) 979–1004
1420-682X/98/090979-26 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Review

Evolution of myoglobin

T. Suzuki[a,*] and K. Imai[b]

[a]Laboratory of Biochemistry, Faculty of Science, Kochi University, Kochi 780-8072 (Japan),
Fax +81 888 44 8356, e-mail: suzuki@cc.kochi-u.ac.jp
[b]Department of Physicochemical Physiology, Medical School, Osaka University, Suita 565-0871 (Japan),
e-mail: kimai@phys1.med.osaka-u.ac.jp

Received 4 March 1998; received after revision 5 May 98; accepted 5 May 1998

Abstract. The distribution, physiological function, amino acid sequence and gene structure of myoglobin and myoglobin-like proteins from various taxa are summarized, and their evolution is discussed. Although it has long been thought that all haemoglobins and myoglobins have evolved from a common ancestral gene, the knowledge presently accumulated about the structures of these proteins suggests that there may be three distinct origins: a 'universal globin', a 'compact globin' and an 'IDO-like globin'.

Key words. Myoglobin; amino acid sequence; gene structure; evolution.

Introduction

It is well known that during evolution three types of respiratory proteins that can bind oxygen reversibly appeared. These are globins (myoglobin, haemoglobin and chlorocruorin), haemocyanin and haemerythrin. Haemocyanin and haemerythrin are distributed among a limited number of species from a restricted number of metazoan phyla such as Arthropod, Mollusca, Sipuncula, Brachiopoda and Annelida, whereas globins are widely distributed from bacteria to vertebrates, although their occurrence is sporadic [1]. An organism usually expresses one of the three oxygen-binding proteins. Coexpression occurs rarely, e.g. in some gastropod molluscs, such as *Aplysia* and *Dolabella*, which express two proteins, haemocyanin in the haemolymph and myoglobin in the buccal mass, and the marine polychaete *Nereis*, which has haemoglobin in coelomic fluid and haemerythrin in body wall muscle. The crystal structures of the three oxygen-binding proteins are quite different, implying that all three evolved from distinctly different ancestors. Globins carry a haem group; protohaeme IX for haemoglobin and myoglobin (fig. 1) and chlorocruorohaem for chlorocruorin. The dioxygen binds to the sixth coordination site of the haem iron. On the other hand, haemocyanin and haemerythrin do not carry a haem group, and the oxygen binds directly to copper or iron atoms coordinated to amino acid side chains of the protein moiety.

Globins are classically characterized by having haem as the prosthetic group, an optical spectrum with two visible bands α (around 580 nm) and β (around 540 nm) in the oxy form and the capability of reversible oxygen binding. Most globins are composed of 140–160 amino acids (M_r 15–18 kDa). Although the amino acid sequence homology between the distantly related globins is low – the amino acid identity between sperm whale and *Aplysia* myoglobins is about 25% – they have a

* Corresponding author.

very similar three-dimensional structure (the so-called myoglobin fold) [2] (fig. 1). An accidental evolutionary event of two or more duplications of a globin gene followed by their fusion sometimes results in a multidomain globin as seen in a blood clam, *Barbatia*, two-domain globin (M_r 32 kDa) [3], and a brine shrimp, *Artemia*, nine-domain globin (M_r 135 kDa) [4]. In addition, two new types of globins, an unusual myoglobin from a mollusc containing one haem per 40-kDa subunit [5] and significantly smaller globins ($\sim$115 residues) from protozoans [6], have been discovered. These globins show no significant amino acid sequence homology with the usual globins composed of 140 to 160 residues.

Myoglobin is a haemoprotein contributing to intracellular oxygen storage and transcellular facilitated diffusion of oxygen [7]. It is found typically in skeletal or cardiac muscle of vertebrates in which continuous oxygen flow is required for high activity of aerobic metabolism. Similar intracellular myoglobin-like haemoproteins, which serve as oxygen stores rather than transport or as an oxygen sensor in some cases, are found in bacteria (usually called haemoglobin), plants (leghaemoglobin or plant haemoglobin), protozoans (haemoglobin or myoglobin by some authors) and invertebrate nerve systems (myoglobin or tissue haemoglobin) [1].

In this review we will describe the distribution, structure, function and evolution of these intracellular, noncirculating oxygen-binding haemoproteins. The term 'myoglobin' will be used for the globins contained in muscles, while the term 'haemoglobin' will be used for all other noncirculating proteins described above, irrespective of their aggregation state. Table 1 lists relatively well-studied haemoproteins selected from the proteins that are described in this review, together with their source, molecular weight, aggregation state and functional properties.

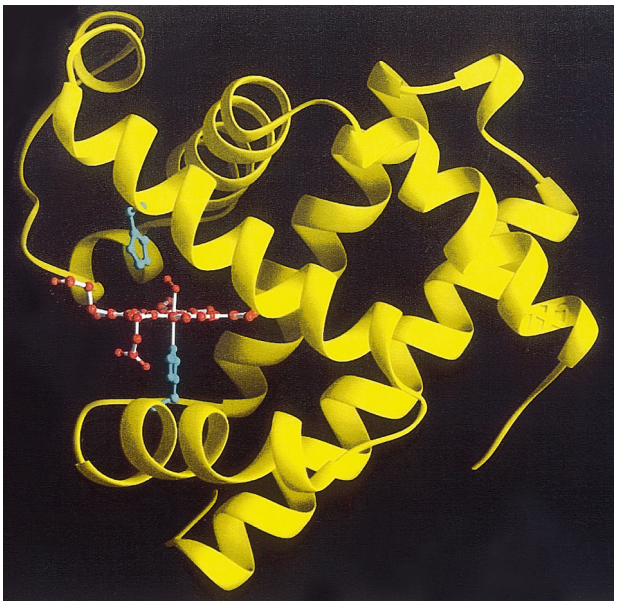

Figure 1. Ribbon diagram of the structure of oxygenated sperm whale myoglobin. The polypeptide chain, shown in yellow, is composed of eight α-helical regions, which are joined via nonhelical segments and folded, forming a compact tertiary structure. This structure is called the 'myoglobin fold'. The haem group, shown in red, is inserted in a pocket lined with hydrophobic amino acid side chains. The dioxygen molecule, also shown in red, is at the centre of the haem. The two residues in bluish-green are the proximal (lower) and distal (upper) histidines. The proximal histidine (His-F8; the eighth residue of helix F), which is covalently bound to the haem iron, is conserved in all the myoglobins and haemoglobins studied so far. The distal histidine (His-E7; the seventh residue of helix E) occurs in many myoglobins and haemoglobins, stabilizing the bound oxygen via hydrogen bonding with it, but it is occasionally replaced by other residues. (By courtesy of Drs. Jeffry C. Nichols, George N. Phillips Jr. and John Olson, Department of Biochemistry and Cell Biology and the W. M. Keck Center for Computational Biology, Rice University, Houston, Texas, USA).

Physiological function

Oxygen-binding properties

The physiological function of myoglobin is based solely on its reversible oxygen-binding capability. The reversible oxygen binding is preserved only when the haem iron is kept in the ferrous state, which is stabilized by hydrophobic environments around the haem group. Oxygen-binding properties are well described by the oxygen dissociation (or oxygen equilibrium) curve, which expresses a relationship between oxygen saturation (SO_2) and partial pressure of oxygen (PO_2) observed in chemical equilibrium. PO_2 can be replaced by a concentration of dissolved oxygen that is proportional to PO_2. For normal aqueous solutions, PO_2 of 1 mmHg (=133 Pa) corresponds to oxygen concentrations of 1.82 μM, 1.66 μM and 1.40 μM at 20, 25 and 37 °C, respectively. Monomeric myoglobin or haemoglobin shows a hyperbolic oxygen dissociation curve as expressed by a plot of SO_2 vs. PO_2. This shape is in distinct contrast with the sigmoid shape of the dissociation curve, typically exhibited by vertebrate tetrameric haemoglobins or annelid polymeric haemoglobins, and in a few cases by invertebrate dimeric myoglobins. The degree of sigmoidicity is measured by the Hill coefficient, n, that is the maximum slope of the Hill plot, $\log [SO_2/(1-SO_2)]$ vs. $\log PO_2$ [8, 9]. For a hyperbolic curve, $n=1$, and n becomes larger as the curve becomes more sigmoid but does not exceed the number of the haem groups per molecule. Oxygen affinity is usually expressed in terms of partial pressure of oxygen at half oxygen saturation, P_{50}. The larger the P_{50}, the lower the

oxygen affinity, and vice versa. The oxygen affinity of most circulating haemoglobins is reduced as blood pH becomes lower in physiological range. This phenomenon is called the '(alkaline) Bohr effect'.

Physiological functions

The physiological functions of myoglobin and noncirculating haemoglobin show diversity that arises from differences in the amino acid sequence of the polypeptide chains. Some of those functions have not yet been completely established. Whether widely accepted or hypothetical, the functions may be enumerated as below (table 1).

1) Short-term storage of oxygen (oxygen buffer)
2) Long-term storage of oxygen
3) Facilitation of intracellular oxygen diffusion
4) Biochemical catalyst
5) Oxygen scavenger
6) Oxygen sensor (and signal transducer)

It is widely accepted that myoglobin serves as an oxygen store during temporary deficits in oxygen supply. In beating heart and exercising skeletal muscle, myoglobin acts as a short-term oxygen store (i.e. an oxygen buffer), tiding the muscle over from one contraction to the next (function 1) [10, 11]. The high concentration of skeletal myoglobin in aquatic mammals was thought to enable a long-term oxygen supply during diving (function 2) [7]. However, this function is of small significance, because the amount of oxygen bound to myoglobin is too little to ensure a long-term supply [12]. A series of experimental and theoretical studies [7, 12–15] provided a line of evidence supporting the validity of function 3, in which myoglobin facilitates intracellular diffusion of oxygen in skeletal muscle cells. This mechanism was based on the idea that the amount of oxygen transported by translational diffusion of myoglobin makes a significant contribution to the total oxygen flux in the cell compared with spontaneous diffusion of free dissolved oxygen. To prove that this mechanism operates in vivo, however, several discrepancies must be resolved [16–19]. Function 4 includes myoglobin-mediated oxidative phosphorylation in cardiac myocytes [20, 21], haemoglobin-mediated sulphide utilization in symbiont-harbouring gill of sulphide-interface clams, terminal oxidase activity of bacterial haemoglobin and so on [20]. The very high oxygen affinities of leghaemoglobin and haemoglobins of nodulating nonlegumes (*Parasponia andersonii* and *Casuarina glauca*) maintain oxygen concentrations at a low level, which is sufficient to enable oxygen-labile nitrogenase to participate in nitrogen fixation in nodules [20, 22]. Thus, these haemoglobins are considered to serve as oxygen scavengers (function 5). Paradoxically, leghaemoglobin maintains a continuous flow of oxygen to bacteroids by facilitated oxygen diffusion to support bacterial oxidative phosphorylation to meet the demand of nitrogenase for adenosine triphosphate (ATP). FixL, a chimeric protein having a haem domain and a kinase domain, occurring in *Rhizobium* or *Bradyrhizobium* binds oxygen reversibly with a very low affinity. The proposed function of this protein (function 6) is to sense oxygen through its haem domain and transduce this signal by controlling the phosphorylation of the transcriptional activator FixJ that in turn induces the expression of nitrogen fixation genes, thereby enabling oxygen-dependent switching of nitrogen fixation [23–25]. Chimeric two-domain O_2-binding proteins are found in other bacteria: the flavohaemoglobins in yeast [26, 27] and the hydrogen bacterium, *Alcaligenes eutrophus* [28], show a diaphorase activity, whereas that of *Escherichia coli* [29] shows a dihydropteridine reductase activity. However, the physiological roles of these proteins in vivo are not clear.

Myoglobin or noncirculating haemoglobin has no cofactor that modifies its oxygen-binding properties, with a few exceptions in plant haemoglobins and invertebrate myoglobins, which show pH dependences (table 1), and this has been one of the most conspicuous features compared with allosteric functional regulations of circulating haemoglobins. However, it was recently found that lactate, a product of glycolysis, causes a marked lowering of oxygen affinity for sperm whale myoglobin and horse heart myoglobin [30]. This phenomenon may have a physiological significance for these myoglobins, just as protons and carbon dioxide do for haemoglobin, and further it may force a change in the traditional thinking that myoglobin is a monomeric nonallosteric protein.

An important physiological function of oxygen-binding proteins that is not the intended subject of this review is transport of oxygen between the sites of gas exchange. The proteins participating in this function include haemoglobin, chlorocruorin, haemerythrin and haemocyanin, which are either encapsulated (i.e. contained in red cells) or directly dissolved in blood plasma, haemolymph or coelomic fluid. The structure and function of these proteins are reviewed in other articles [1, 2, 31–42]. The most striking feature of these proteins, with a few exceptions, compared with myoglobin or noncirculating haemoglobin is cooperative binding of oxygen which arises from interactions among oxygen-binding sites in the same molecule. The n value can be as large as 10–12 for haemocyanins and extracellular haemoglobins, which carry 48–144 oxygen-binding sites per molecule. The sigmoid oxygen dissociation curve undoubtedly realizes efficient oxygen transport. The oxygen affinity shows enormous species-dependent variations, being a manifestation of molecular-level adaptations to environments [33]. The highest affinity known to date is $P_{50} = 0.0015$ mmHg for *Ascaris* perienteric haemoglobin [43], and the lowest is $P_{50} = 420$ mmHg

Table 1. Properties of myoglobins and noncirculating haemoglobins.

Name	Source	MW per haem*	Aggregation state	Function	Oxygenation properties†			Remarks	Ref.
					P_{50}‡	n§	pH-dep.‖		
Bacteria									
Hb	*Vitreoscilla*	16	dimer	O_2 storage and diffusion or terminal oxidase	40	2 (CO binding)			44, 195
FixLT	*Rhizobium meliloti*	43	dimer	O_2 sensor	27	1.0		truncated	23–25
FixL	*Bradyrhizobium japonicum*	56	dimer	O_2 sensor	76	1.0		full-length	24
Cyanoglobin	Cyanobacterium *Nostoc commune*	12.5		O_2 scavenger in nitrogen fixation					46
Unicellular eukaryotes									
FlavoHb	Yeast *Candida norvegensis*	44			0.01	1.0	Absent	chimeric with FAD domain	26, 47, 122, 196
Chloroplast Hb	Green alga *Chlamydomonas eugametos*	17							48
Hb	Protozoa *Paramecium aurelia*	13	monomer		0.6				6, 57
Plants									
LegHb	Legume *Glycine max*	16	monomer	O_2 scavenger in nitrogen fixation	0.04 ~0.06	1.0	present	nodulating; three components	58, 104, 197
Hb	*Parasponia andersonii*	19	dimer	O_2 scavenger in nitrogen fixation	0.049		present	nodulating	61, 62
Hb	*Casuarina glauca*	18	monomer	O_2 scavenger in nitrogen fixation	0.074			nodulating	63
Hb	*Trema tomentosa*	18	monomer	O_2 transfer?				nonnodulating	64
Invertebrates									
Mb	Platyhelminthes *Dicrocoelium deudriticum*	18	monomer		0.15	1.0	present	acid Bohr effect	66, 69
Mb	Nematoda *Ascaris suum*	16	dimer	O_2 transfer?	0.11			from body wall	43, 75
Mb	Annelida *Abarenicola pacifica*	17	monomer		2.8	1.0	absent	from body wall	81
	Travisia foetida	15	dimer	O_2 transfer	0.4	2.0	absent	from body wall	82
	Arenicola marina	18	monomer	O_2 transfer	0.8 ~1.3	1	small	from body wall; two components	79
Nerve Hb	Annelida *Aphrodite aculeata*	16	monomer	?	1.1	1		from nerve	84
Mb	Mollusca (Amphineura) *Sipharochiton pelliserpentis*	17	monomer	O_2 transfer?	0.95	1.03	absent	from radular muscle	88
		17	dimer	O_2 transfer?	3.5	1.35	absent	from radular muscle	88
	Amaurochiton glaucus	17	monomer	O_2 transfer?	0.63	0.85	absent	from radular muscle	88
		17	dimer	O_2 transfer?	2.76	1.37	absent	from radular muscle	88

Table 1. Continued.

Name	Source	MW per haem*	Aggregation state	Function	Oxygenation properties†			Remarks	Ref.
					P_{50}‡	*n*§	pH-dep.‖		
Mb	Mollusca (Gastropoda)								
	Buccinum undatum	17	dimer	O_2 transfer?	13	1.4	absent	from radular muscle	102
	Fusitriton oregonensis		dimer	O_2 transfer?	8.5	1.35	absent	from radular muscle	198
	Nassa mutabilis	15	dimer	O_2 transfer?	4.7	1.5	absent	from radular muscle	100
	Cerithidea rhizophorarum	17	dimer	O_2 transfer?	2.7	1.1		from radular muscle	103
	Aplysia kurodai	16	monomer	O_2 transfer?	4.8	1.0		from radular muscle	103
	Dolabella auricularia	16	monomer	O_2 transfer?	1.4	1.0	absent	from triturative stomach	206
IDO-like Mb	Mollusca (Gastropoda)								
	Sulculus diversicolor	41	dimer		3.8	1.02	absent	from radular muscle	169, 199
	Battilus cornutus	40	monomer		3.5	1.04	absent	from radular muscle	208
Mb	Mollusca (Bivalve)								
	Mercenaria mercenaria	17	monomer					from adductor muscle	112
Vertebrates									
Mb	Elasmobranchii								
	Galeorhinus japonicus	16	monomer	O_2 storage and transfer	1.1	1.0	absent	from skeletal muscle	200
	Mustelus japonicus	16	monomer	O_2 storage and transfer	1.2	1.0		from skeletal muscle	199
Mb	Osteichthyes								
	Thunnus thynnus	16	monomer	O_2 storage and transfer	0.9	1.0	absent	from skeletal muscle	201
	Chinnichthyidae (five species)	16	monomer	O_2 storage and transfer				from heart ventricle	116
Mb	Aves								
	Gallus domesticus	18		O_2 storage and transfer	0.5	1.0		from gizzard	202
		18			0.5	1.0		from skeletal muscle	202
Mb	Mammalia								
	Physeter catadon	18	monomer	O_2 storage and transfer	0.51	1.0	absent	from skeletal muscle	Imai and Miyazaki, as quoted in ref. 199
	Elephas maximus	18	monomer	O_2 storage and transfer	0.62	1.19		from skeletal muscle	189
	Homo sapiens	18	monomer	O_2 storage and transfer	0.72	1.03	absent	from skeletal muscle	203

*Molecular weight per haem (kDa). †At 20 to 25 °C and at physiological pH. ‡Partial pressure of oxygen (mmHg). §Hill's coefficient. ‖pH dependence of oxygen affinity (P_{50}).

for *Potamilla* chlorocruorin [207], spanning a 280,000-fold variation. Oxygen transport by oxygen-binding proteins of higher animals are regulated by various cofactors such as H^+, CO_2, organic phosphates, lactate and divalent cations (Ca^{2+} and Mg^{2+}). Human haemoglobin (Hb A) transports part of CO_2 from the peripheral tissues to the lungs by binding it at the amino termini. Hb A further enhances the major part of CO_2 transport (in the form of bicarbonate) by blood plasma through oxygen-linked proton release (the Haldane effect). Hb A also contributes to acid-base balance (pH regulation) of blood. These three physiological functions (O_2 transport, CO_2 transport and pH regulation) are probably and equally expressed by most of vertebrate haemoglobins. In lower animals the importance of haemoglobin in oxygen transport varies from species to species, and oxygen store can be the major role for some species, making a strict division of roles impossible [33, 36].

Distribution and some properties

The distribution of myoglobin in the animal kingdom is more intermittent than that of haemoglobin. Despite of extensive studies of haemoglobins and myoglobins over the last 30 years, myoglobin has been reported in only 7 of 30 phyla in Animalia. This section presents an overview of the distribution and some properties of myoglobin, together with the haemoglobins from prokaryotes, unicellular eukaryotes and plants. Selected proteins are listed in table 1.

Prokaryotes

Webster [44] was the first to discover the presence of haemoglobin in the Gram-negative bacterium *Vitreoscilla*, which was first believed to be a cytochrome *o*. The haemoglobin is a dimer composed of two identical subunits of 16 kDa, has a low affinity for oxygen ($P_{50} = 40$ mmHg) and is expressed in large quantity under hypoxic conditions. Haemoglobin was also isolated from *E. coli* [29]. The *E. coli* haemoglobin is a chimeric protein containing both haem and flavin with molecular mass of 44 kDa, can bind oxygen reversibly and has a dihydropteridine reductase activity.

An oxygen-binding domain was also found in a membrane-bound protein of *Rhizobium meliloti* [45] and in a soluble protein of *Bradyrhizobium japonicum* [24]. These proteins, called 'FixL', are involved in oxygen sensing and signal transduction, and their structure is chimeric: the first domain acts as a transmembrane part (this domain is absent in *Bradyrhizobium* FixL); the second is a haem-containing domain ($P_{50} = 27–76$ mmHg); and the third is a kinase domain [23–25].

Another type of bacterial haemoglobin was discovered from the cyanobacterium *Nostoc commune* [46]. This haemoglobin has a molecular weight of 12.5 kDa, and appears to be involved in some aspects of nitrogen fixation, functioning as a scavenger of oxygen.

Unicellular eukaryotes

Oshino et al. [26, 47] isolated a haemoglobin from the yeast *Candida norvegensis* (previously called '*C. mycoderma*') and characterized it in detail. The haemoglobin consists of a 44-kDa single polypeptide chain, contains both haem and flavin prosthetic groups in the molecule and has a high affinity for oxygen ($P_{50} = 0.01$ mm Hg). Thus *Candida* haemoglobin was considered to have a two-domain structure, like *E. coli* haemoglobin. Such chimeric proteins carrying haem and flavin are called 'flavohaemoglobin'. The flavohaemoglobin was also isolated from the yeast *Saccharomyces cerevisiae* [27]. In contrast to *Candida*, in which haemoglobin is constitutively expressed, *Saccharomyces* haemoglobin requires the disruption of normal mitochondrial electron transport to be expressed.

Two 17–18 kDa haemoglobin-like proteins were isolated from the green alga *Clamydomonas eugametos* [48]. The proteins are predominantly located in the chloroplast, particularly in the pyrenoid and thylakoid regions.

The presence of haemoglobin in protozoans was first observed by Sato and Tamiya [49] for *Paramecium caudatum*. Later, Keilin and Ryley [50] confirmed this by spectroscopic analyses for *Paramecium* and *Tetrahymena*. The proteins were purified from *P. caudatum* [51, 52], *P. primaurelia* [51], *P. tetraurelia* [53] and *T. pyriformis* [54]. The molecular mass for the proteins was estimated to be 13–15 kDa, a 10–20% smaller value than usual myoglobin subunit (16–18 kDa). It is noted that in some species haemoglobin shows a remarkable polymorphism: 5 different components were identified electrophoretically in *P. tetaurelia* [53, 55] and 12 components for *P. aurelia* [56]. The oxygen affinity of *Paramecium aurelia* haemoglobin ($P_{50} = 0.6$ mmHg) is comparable with those of mammalian myoglobins [57].

Plants

Haemoglobin occurs abundantly in the root nodules of many kinds of legumes, such as the soybean *Glycine max* [58] and the tropical legume *Sesbania rostrata* [59]. These haemoglobins, called 'leghaemoglobin' are a monomer with molecular mass of 16–17 kDa and display a notable polymorphism: at least four to seven components are present in *Glycine max*. The oxygen affinity of *Glycine* leghaemoglobin is very high ($P_{50} =$ 0.04–0.06 mmHg), and it is thought that it serves to

protect the oxygen-sensitive bacterial nitrogenase as well as to supply oxygen for most other aerobic metabolism [60].

Haemoglobins also occur in nonlegumes, *Parasponia andersonii* ($P_{50} = 0.049$ mmHg) [61, 62], *Casuarina glauca* ($P_{50} = 0.074$ mmHg) [63], *Trema tomentosa* and *Celtis australis* [64]. *Parasponia* haemoglobin is homodimeric while the others are monomeric.

Invertebrates

Platyhelminthes. Body wall myoglobins have been isolated from the primitive trematodes, *Fasciolopsis buski* [65], *Dicrocoelium dendriticum* [66], *Gastrothylax crumenifer*, *Paramphistomum epiclitum* [67] and *Isoparorchis hypselobagri* [68]. In all cases the major myoglobin is a monomer with molecular mass of 17–18 kDa, but a minor dimer myoglobin also occurs in *Gastrothylax* and *Isoparorchis*. The oxygen affinity for *Dicrocoelium* myoglobins is high ($P_{50} = 0.0160$–0.15 mmHg at 25 °C dependent on pH) [69].

Nemertea. Many nemerteans contain myoglobin in the body wall or haemoglobin in the nerve tissue [70]. Recently, a haemoglobin was isolated and characterized from the neural tissue of *Cerebratulus lacteus* [71]. The polypeptide length (109 residues) is the shortest so far known [72].

Nematoda. A body wall myoglobin was isolated from the parasitic nematode *Ascaris lumbricoides*, and the molecular mass was estimated to be 37 kDa on a gel filtration column [73]. This value is comparable to that (40.6 kDa) of the subunit of homooctameric *Ascaris* perienteric fluid haemoglobin (328 kDa), which shows an extremely high oxygen affinity ($P_{50} = 0.0015$ mmHg at pH 7.0 and 20 °C) [74, 43]. A haem content was estimated rather lower (one mole haem group per 35,000–40,600 g protein) for both of *Ascaris* haemoglobin and myoglobin, suggesting haem is easily lost during purification steps [73].

Recently, a body wall myoglobin was isolated from *Ascaris suum* [75]. The myoglobin is a dimer composed of 16 kDa polypeptide chains. Its oxygen affinity ($P_{50} = 0.11$ mmHg) is higher than that of mammalian myoglobin [43].

Two haemoglobin-like 17–18 kDa isoforms, one localized in the body and the other in the cuticle, were also isolated from *Nippostrongylus brasiliensis* [76]. Both globins have oxygen affinities 100-fold higher than the rodent host's haemoglobins. Interestingly, they are expressed in different stages: the body isoform is first expressed upon invasion of the mammalian host, and the abundant cuticular globin is expressed only by adult nematodes in the gut.

Echiura. The echiuran is the group that is considered to be closely allied to annelids. The presence of myoglobin was first reported by Manwell [77] in the body wall muscle of the echiuroid worm *Arhynchite pugettensis*. The *Urechis caupo* also contains myoglobin in the body wall, as well as coelomic haemoglobin [78]. The myoglobin eluted on a gel filtration column in the position corresponding to a molecular mass of 19 kDa, but the mass was also estimated to be 13.8 kDa by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). *Urechis* myoglobin appears to be polymorphic, since the electrophoresis of the myoglobin under denatured conditions results in multiple protein bands [78].

Annelida. Myoglobins were isolated from the body wall muscles of the marine polychaetes *Arenicola marina* [79], *Glycera robusta* [80], *Abarenicola pacifica* [81] and *Travisia foetida* [82]. Myoglobins from *Arenicola*, *Glycera* and *Abarenicola* are monomers, and the molecular masses were estimated to be 15–16 kDa by gel filtration and SDS-PAGE. *Glycera* myoglobin appears to be homogeneous, while *Arenicola* and *Abarenicola* myoglobins exhibit heterogeneity: two and seven components, respectively, identified by electrophoresis. The P_{50} value for oxygen affinity of *Abarenicola* myoglobin is reported to be 2.8 mmHg at 20 °C, and it lacks Bohr effect (pH dependence of oxygen affinity) between pH 7–8 [81].

Travisia foetida contains three oxygen binding proteins, a vascular giant haemoglobin with a molecular mass of 3400 kDa, a coelomic cell haemoglobin and a body wall myoglobin; the latter two are homogeneous dimers of 15 to 16-kDa polypeptide chains. It should be noted that these dimeric proteins show high cooperativity with the n values (the Hill coefficient) of 1.8 to 2.0. There found a general increase in oxygen affinities (P_{50}) with vascular haemoglobin (2.9 mmHg) < coelomic haemoglobin (1.3) < myoglobin (0.4), and it was suggested that oxygen could be transported from vascular haemoglobin to myoglobin via coelomic cell haemoglobin [77, 82].

Two types of body wall myoglobins, Mb I and Mb II [83], were isolated from the marine polychaete *Arenicola marina*, of which Mb I showed a significantly higher oxygen affinity ($P_{50} = 0.8$ mmHg) than Mb II (1.3 mmHg) [79]. Weber and Pauptit [79] suggest that the differentiation of oxygen affinities in *Arenicola* Mbs I and II presumably coincides with a functional involvement over an increased range of oxygen tensions.

A nerve haemoglobin was isolated from the marine polychaete *Aphrodite aculeata* [84]. This haemoglobin is a monomer consisting of a 16-kDa chain, and the oxygen affinity ($P_{50} = 1.1$ mmHg) is similar to that of nerve haemoglobin from the mollusc *Aplysia californica* (4.0 mmHg).

Mollusca

Amphineura. Amphineura is one of the most primitive groups in the phylum Mollusca and has a fossil record dating back to the upper Cambrian [85]. The radular muscle myoglobins were isolated from the chitons *Acanthopleura granulata* [86], *Katharina tunicata*, *Cryptochiton stelleri*, *Mopalia muscosa* [87], *Sipharochiton pelliserpentis*, *Amaurochiton glaucus* [88, 89] and *Liolophura japonica* [90].

Although *Liolophura* myoglobin is composed of three monomeric fractions with a molecular mass of 17 kDa, all other species have both monomer and dimer. The dimer is disulphide-bonded, which is reduced with 2-mercaptoethanol under mild conditions [87, 88]. In spite of the disulphide-link, *Sipharochiton* and *Amaurochiton* dimers display a cooperative sigmoidal oxygen-binding curve with a Hill interaction coefficient value n of 1.35 to 1.37 and a P_{50} of 2.76 to 3.5 mmHg [88, 89]. On the other hand, the monomer exhibited a hyperbolic binding curve with $n = 0.85–1.03$ and $P_{50} = 0.63–0.95$ mmHg. The occurrence of disulphide-bonded dimer is unique to amphineuran myoglobins. Smith et al. [88] suggested that the simultaneous presence of monomer and dimer myoglobins in the same muscle is related to the fact that chitons, having an intertidal habitat, require oxygen-binding systems optimized for both aquatic and aerobic conditions.

The monomeric myoglobin of *Liolophura japonica* appears to be highly resistant against autoxidation: the autoxidation rate (0.0073 h^{-1}) at pH 7.2 and 25°C is 3–20 times slower than other molluscan myoglobins [90].

Gastropoda. One of the most remarkable occurrences of red muscle in invertebrate animals is found in the buccal mass (radular muscle) and triturative stomach of gastropod molluscs. Myoglobin is abundant in such muscle, and in *Aplysia* buccal mass the concentration is as high as 5–6% on a dry weight basis, two or three times larger than human cardiac and skeletal muscle [86]. Therefore, the gastropod myoglobins have been the most extensively studied among invertebrate myoglobins. Read [86] and Terwilliger and Terwilliger [91] have exhaustively reviewed the physicochemical and physiological properties of molluscan haemoglobins and myoglobins.

By 1966, myoglobins from the buccal mass, radular muscle, pharyngeal muscle and triturative stomach and haemoglobins from nerve had been isolated from 22 species of gastropods: *Patella* sp., *Paludina* sp., *Littorina* sp., *Fusitriton oregonense*, *Buccinum undatum*, *Busycon canaliculatum*, *B. contrarium*, *Bulla gouldiana*, *Aplysia californicus*, *A. depilans*, *A. limacina*, *A. punctata*, *Bursatella* sp., *Tethys carifornicus*, *Physa fontinalis*, *Lymnaea auricularia*, *L. limora*, *L. palustris*, *L. stagnalis*, *Planorbis contortus*, *P. corneus* and *P. vortex* (see Read [86]). More recent studies have added 15 myoglobins from *Nerita peloronta* [92], *Littorina littorina*, *Siphonaria gigas* [93], *Lunatia heros* [94], *Acmaea testudinalis*, *Haliotis kamtschatkana*, *Tegula funebralis*, *Fusitriton oregonensis*, *Thais lamellosa* [95], *Aplysia kurodai* [96], *A. juliana* [97], *Cerithidea rhizophorarum* [98], *Dolabella auriculaia* [99], *Nassa mutabilis* [100] and *Bursatella leachii* [101].

In contrast to monomeric vertebrate myoglobins, some gastropod myoglobins exist as dimers. The dimeric myoglobins from *Buccinum*, *Fusitriton* [102] and *Nassa* [100] show sigmoid oxygen equilibrium curves ($P_{50} =$ 4.7–13 mmHg; $n = 1.3–1.5$). On the other hand, the monomeric myoglobins from *Aplysia* and *Dolabella* show hyperbolic curves ($P_{50} = 1.4–4.8$ mmHg; $n = 1$) [103, 205].

In the class Prosobranchia, dimeric myoglobin is found in the members of the order Neogastropoda and in a few members of the order Mesogastropoda. On the other hand, monomeric myoglobins are distributed in the members of the subclasses Pulmonata and Opisthobrabchia, and also in the two orders Archaeogastropoda and Mesogastropoda of the class Prosobranchia. This distribution pattern is consistent with the picture drawn by Read [93] that the monomeric form represents an ancestral condition and that the dimeric form displays a more advanced state of myoglobin evolution.

The myoglobin from *Cerithidea rhyzophorarum* undergoes a dimer-monomer conversion coupled with the oxidation of haem iron [103]. Namely, the oxygenated form of *Cerithidea* myoglobin exists as a dimer, but upon oxidation it is converted to hemichrome monomers. This conversion is reversible over a wide range of pH 5 to 11.

The autoxidation rate of the oxy form of *Aplysia* and *Dolabella myoglobins* has been examined over a wide range of pH [99, 105]. Their pH dependences of autoxidation are quite different from that of sperm whale myoglobin [106, 107], and this unusual property was ascribed to the absence of distal (E7) histidine, a functional key residue, in *Aplysia* and *Dolabella* myoglobins.

Abalone. We have recently isolated myoglobin from the abalone (archaeogastropod) *Sulculus diversicolor* [5]. This myoglobin differed from all the myoglobins so far examined in the sense that the polypeptide chain has an unusual molecular mass of 39 kDa, which is about double the size of the usual myoglobin subunit; it contains only one haem per molecule, and the amino acid sequences of several internal peptides shows no significant homology with other molluscan globins [5]. Thus it was first assumed that *Sulculus* myoglobin has a two-domain structure resulting from duplication of globin gene, but one of the two domains does not function as

myoglobin. The complementary DNA (cDNA)-derived amino acid sequence [108] showed that *Sulculus* myoglobin is not a two-domain globin but has a significant homology with human indoleamine dioxygenase (IDO), a tryptophan-degrading enzyme containing haem. As a reaction intermediate, IDO can take an oxygenated form, and its absorption spectrum is very similar to those of haemoglobin and myoglobin [109]. However, the intermediate is extremely unstable, and thus IDO cannot serve as an oxygen carrier. On the other hand, the oxygenated form of *Sulculus* myoglobin is stable enough to function as an oxygen carrier: the autoxidation rate at pH 7.2 and 25 °C (0.030 h^{-1}) was comparable to those of other molluscan myoglobins. So far, we have also isolated IDO-like myoglobins from the gastropod molluscs *Nordotis* [110], *Omphalius pfeifferi* [111], *Battilus* [208], *Chlorostoma argyrostoma* and *Chlorostoma xanthostigma* (T. Suzuki et al., unpublished results for the latter two).

Clams. Myoglobin also occurs in the adductor muscle, heart and foot of a number of bivalves, and its concentration is as high as 2% in some species [86]. *Mercenaria mercenaria* contains two monomer myoglobins (17.4 kDa) in adductor muscle, which are separable by ion-exchange chromatography [112]. Similarly, the beefsteak clam *Saxidomus nuttali* has two distinct myoglobins (pI = 5.8 and 7.2) in adductor muscle [91].

Vertebrates

Myoglobin occurs typically in skeletal or cardiac muscle of mammalians, and it has been isolated from more than 80 species of vertebrates. The P_{50} values for vertebrate myoglobins are 0.5–1.2 mmHg, and the *n* values are 1.0–1.19 (see table 1). A microheterogeneity was demonstrated by electrophoresis and column chromatography for human, horse, tuna and sperm whale myoglobins: for example, at least four components are isolated from *Physeter catodon* (sperm whale) [113], of which only the major myoglobin has been sequenced. Elucidation of the crystal structure of sperm whale myoglobin by Kendrew et al. [114] and subsequent determination of its amino acid sequence by Edmundson [115] were landmarks in structural biology.

Very recently, the presence of myoglobin (but not haemoglobin) was identified in heart ventricle of the ocellated icefish *Chionodraco rastrospinosus* [116]. Sidell et al. [116] also report that of the eight icefish species examined, three do not express myoglobin in heart ventricle, but all species retain the myoglobin gene in their genomes.

Amino acid sequence

Globins (haemoglobins and myoglobins) are typically composed of 145 to 155 amino acid residues, of which only three amino acid residues, CD1-Phe, E7-His and F8-His in haem cavity, are highly conserved. The haem-neighbouring CD1-Phe and iron-binding F8-His are strictly conserved in all globins, whereas the E7-His has been replaced by Gln in a few globins of vertebrate and by Val, Leu, Tyr or Gln in a considerable number of globins of invertebrate. The E7-His is associated with ligand-binding properties, and in an oxygenated form it is capable of forming a hydrogen bond to the bound dioxygen and stabilizing it [117].

Minihaemoglobins from prokaryotes and unicellular eukaryotes

The amino acid sequence of a minihaemoglobin was first reported from *Paramecium caudatum* by Iwaasa et al. [6]. The polypeptide chain has an acetylated N-terminus and consists of 116 residues, much shorter than mammalian myoglobins by 37 residues [118]. It has only two histidine residues at positions 81 and 101 in the sequence (see fig. 2a), and Shikama et al. [118] suggests that these histidines correspond to distal (E7) and proximal (F8) histidines, respectively. The same types of haemoglobins were sequenced from *Tetrahymena pyriformis* [54], *T. thermophila* [119], *Paramecium multimicronucleatum*, *P. triaurelia* and *P. jenningsi* [120].

Two chloroplast haemoglobins from the green alga *Chlamydomonas eugametos* were sequenced via its cDNA [48]. They are composed of 171 and 164 amino acid residues, respectively, of which the N-terminal 38 residues constitute transit peptides of chloroplast precursors. Thus, the mature proteins are composed of 133 and 126 residues, respectively, and show a significant sequence homology (37–42% identity) with protozoan haemoglobins.

A haemoglobin sequence was reported from the cyanobacterium *Nostoc commune* [46]. The polypeptide consists of 118 residues, showing a significant homology (26–30% identity) with those of protozoan haemoglobins. Similar haemoglobin sequences were also reported from another cyanobacterium *Synechococcus* (D90910) and Gram-positive bacterium *Mycobacterium tuberculosis* (MTCY48.23).

Amino acid sequences of minihaemoglobins from *Paramecium*, *Tetrahymena*, *Chlamydomonas*, *Nostoc*, *Synechococcus* and *Mycobacterium* are aligned in figure 2a, and a phylogenetic relationship among the sequences is given in figure 2b.

Bacterial single-domain haemoglobin and flavohaemoglobins in *E. coli* and yeasts

The amino acid sequence of the single-domain haemoglobin from *Vitreoscilla* was the first evidence that the origin of globin genes dates back to a common

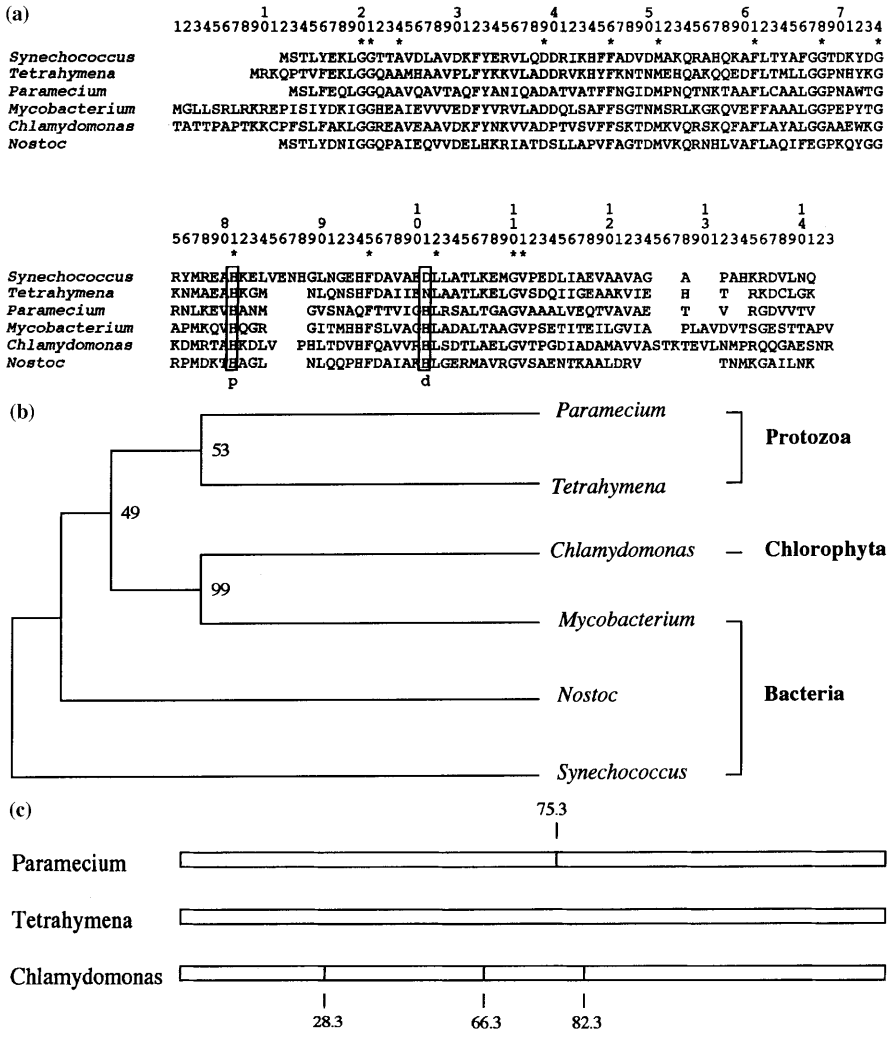

Figure 2. (*a*) Alignment of amino acid sequences of compact globins. Amino acid residue conserved in all sequences is indicated by asterisk. The boxed residues indicate the tentatively assigned haem-binding proximal histidine (p) and distal residue (d). This alignment was obtained using the algorithm of Feng and Doolittle [204]. (*b*) A phylogenetic relationship of compact globins. This tree was obtained using the parsimony method in the PHYLIP package, ver. 3.5c [178]. The numbers at branching points indicate the bootstrap values. (*c*) Gene structure of compact globins. The position of the intron is indicated by a vertical line. For example, the number 75.3 shows that the intron is inserted after the third nucleotide of a codon corresponding to the amino acid sequence position 75 in the alignment in (*a*).

ancestor of prokaryotes and eukaryotes [121]. The *Vitreoscilla* haemoglobin comprises 145 residues and showed a significant homology (24% identity) with plant leghaemoglobins.

The cDNA-derived amino acid sequences for flavohaemoglobins were determined from *E. coli* [29] and yeasts *Saccharomyces cerevisiae* [27] and *Candida norvegensis* [122]. The proteins are composed of approximately 400 amino acid residues and consist of two distinct domains: the N-terminal domain (140 residues) is homologous with globins, and the C-terminal is a flavin-containing reductase domain. *Vitreoscilla* single-domain haemoglobin shows strong homology with the N-terminal globin domains of *E. coli*, *Saccharomyces* and *Candida* (46, 53 and 39% sequence identity, respectively).

Recently, several flavohaemoglobin sequences were determined from the bacteria *Erwinia chrysanthemi*

Figure 3. (*a*) Alignment of amino acid sequences of universal myoglobins. The amino acid residue conserved in all sequences is indicated by an asterisk. The boxed residues indicate the distal-E7 residue (d) and haem-binding proximal-F8 histidine (p). This alignment was obtained using the algorithm of Feng and Doolittle [204].

[123] (81% amino acid sequence identity with *E. coli*-haem domain), *Vibrio parahaemolyticus* (P40609) (79%), *Bacillus subtilis* [124] (45%) and *Alcaligenes eutrophus* [125] (46%).

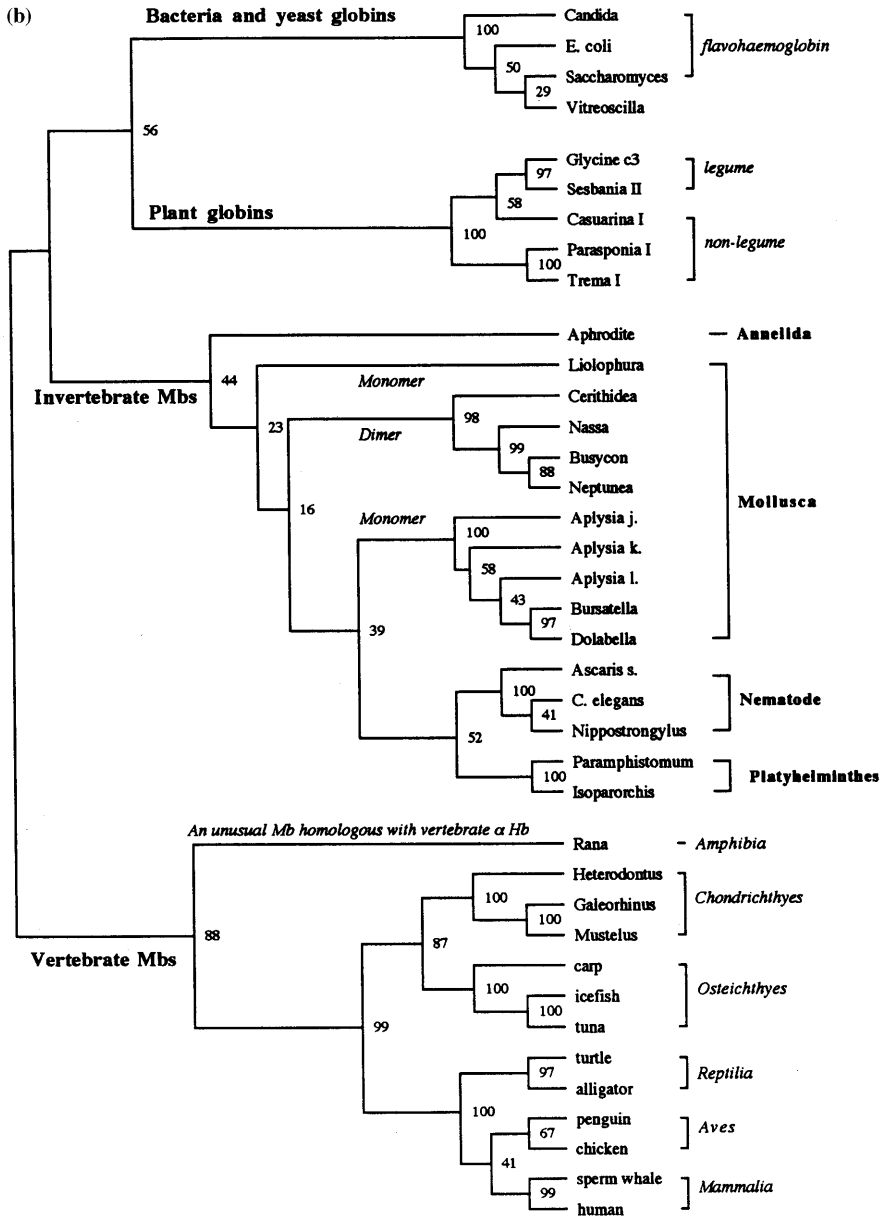

Figure 3. (*b*) The phylogenetic relationship of the universal myoglobins. This tree was obtained using the parsimony method in the PHYLIP package, ver. 3.5c [178]. The numbers at branching points indicate the bootstrap values.

The amino acid sequences of the haem domains from *Vitreoscilla*, *E. coli*., *Candida* and *Saccharomyces* haemoglobins are aligned in figure 3a, and a phyloge netic relationship among the sequences is given in figure 3b.

Bacterial FixL protein with a haem-containing domain

Gilles-Gonzales et al. [23] suggested that the second domain, consisting of 134 amino acid residues, of FixL of *Rhizobium meliloti* corresponds to a haem-containing domain. However, the haem domain showed no significant homology with other globins described in this section. It is possible that this domain is derived from a different origin.

Plant haemoglobins

Symbiosis-dependent legume haemoglobins (leghaemo-

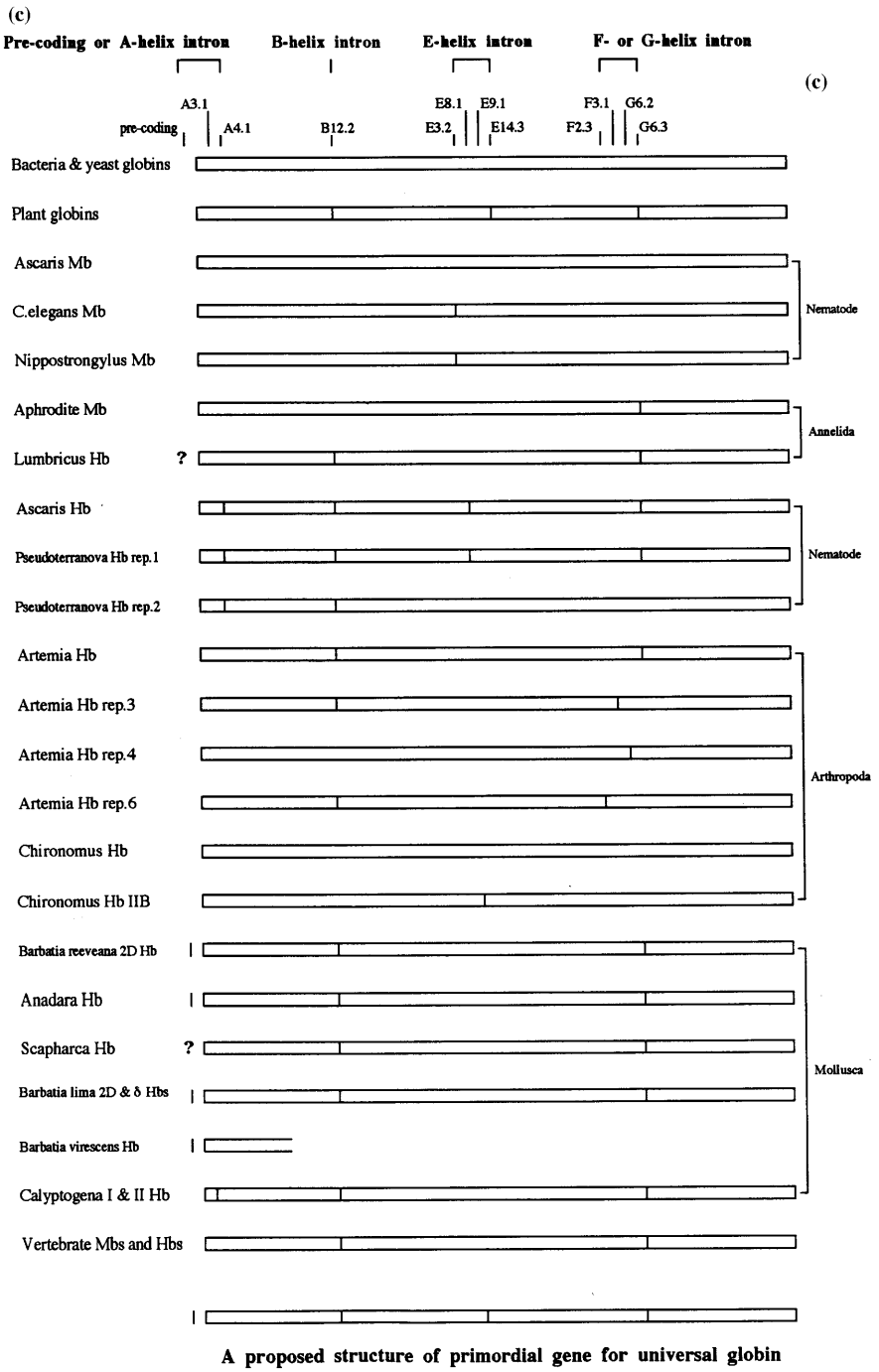

Figure 3. (*c*) Gene structure of the universal globins. The position of the intron is indicated by vertical line. For example, B12.2 shows that the intron is inserted after the second nucleotide of a codon corresponding to the amino acid sequence position B12 in the alignment shown in figure 2a.

globins) were sequenced for soybean (*Glycine max*) components a, c1, c2 and c3, kidney bean a, broad bean I, pea I, yellow lupin I and II, *Sesbania* II, IV, VI and VII, and *Medicago* I and II. Nonlegume haemoglobins were also sequenced for *Parasponia* I and II, *Casuarina* I and II, *Medicago* III and nonnodulating plant *Trema*

I. For references to original sequence papers of plant haemoglobins, see the review by Vinogradov et al. [1].
The amino acid sequences of haemoglobins of *Glycine* c3 [126], *Sesbania* II [127], *Casuarina* I [128], *Parasponia* I [129] and *Trema* I [64] are aligned in figure 3a, and a phylogenetic relationship among the sequences is given in figure 3b.

Invertebrate myoglobins

The amino acid sequences of myoglobins from *Paramphistomum epiclitum* and *Isoparorchis hypselobagri* (trematodes) belonging to the phylum Platyhelminthes, one of the most primitive invertebrates were determined chemically [130]. *Paramphistomum* and *Isoparorchis* myoglobins consist of 147 and 148 amino acid residues, respectively. In both myoglobins, the distal histidine-E7 was surprisingly replaced by a tyrosine residue, which normally causes rapid autoxidation of the haem iron (Fe^{2+} to Fe^{3+}), and thus oxygen cannot bind to the haem, as manifested in human haemoglobin M mutants [131, 132]. In contrast, trematode myoglobins are functional molecules with a high oxygen affinity. Using molecular modelling, Rashid et al. [130] predicted two possible positions for the aromatic ring of Tyr-E7 to explain the high oxygen affinity of trematode myoglobins.
Recently, cDNA-derived amino acid sequences of 154 to 158 residues were determined for myoglobins from nematodes *Ascaris suum* [75], *Nippostrongylus brasiliensis* [76] and *C. elegans* [133]. In *Ascaris* and *C. elegans* myoglobins, a Gln residue occurs at the distal E7 position, whereas in *Nippostrongylus* myoglobin the histidine E7 is replaced by a Leu residue.
A nerve haemoglobin from the polychaete *Aphrodite aculeata* was sequenced by Moens and co-workers [134]. The N-terminal Ala residue is presumably acetylated, and the polypeptide chain comprises 150 amino acids. The nerve haemoglobin is not likely to represent a separate globin family, because its messenger RNA (mRNA) was also detected in muscle, gut and pharynx tissues of *Aphrodite* [134]. In this haemoglobin, the distal E7 histidine is conserved.
Molluscan myoglobins were most extensively sequenced: a monomeric myoglobin from the chiton *Liolophura japonicus* (one of the primitive molluscs) [90], monomeric myoglobins of *Aplysia limacina* [135, 136], *A. kurodai* [96], *A. juliana* [97], *Dolabella* [99], *Bursatella* [101] and *Neptunea* [137], and homodimeric myoglobins from *Busycon* [138], *Cerithidea* [98] and *Nassa* [139]. *Aplysia*, *Dolabella* and *Bursatella* myoglobins contain a single histidine residue which corresponds to the haem-binding proximal histidine, and their distal E7 histidine is replaced by a valine residue. On the other hand, the E7 histidine has been conserved in dimeric myoglobins of *Neptunea*, *Busycon*, *Cerithidea* and *Nassa*.
The amino acid sequences of 16 invertebrate myoglobins are aligned in figure 3a, and a phylogenetic relationship among the sequences is given in figure 3b.
The cDNA-derived amino acid sequence of an unusual 39-kDa myoglobin from the abalone *Sulculus diversicolor* has been determined [108]. The mature protein consists of 376 amino acid residues, and the N-terminus is blocked. The amino acid sequence showed no significant homology with any other globins, but surprisingly showed high homology (35% identity) with human indoleamine dioxygenase (IDO).
The cDNA-derived amino acid sequences of IDO-like myoglobins were determined from another abalone *Nordotis madaka* [110], and turbanshells *Battilus cornutus* [208] and *Omphalius pfeifferi* [111]. The IDO-like myoglobins are aligned with human and mouse IDOs [140, 141], and a homologue in the yeast *Saccharomyces cerevisiae* (Z49578), in figure 4a, and a phylogenetic relationship among the sequences is given in figure 4b.

Vertebrate myoglobins

Myoglobins from many species of vertebrates have been sequenced chemically, and at least 80 amino acid sequences have been deposited in the Swiss-Prot data bank. Those include 5 myoglobins from Chondrichthyes (primitive and modern sharks), 3 from Osteichthyes (carp, yellowfin tuna and icefish), 1 from Amphibia (bullfrog), 4 from Reptilia (turtles, lace monitor and alligator), two from Aves (Emperor penguin and chicken) and 71 from Mammalia [including the primitive Monotremata (Australian echidna and duckbill platypus) and Marsupialia (red kangaroo)].
Of the myoglobins from Reptilia, Aves and Mammalia, all but alligator and penguin myoglobins are composed of 153 amino acid residues. Alligator myoglobin retains an additional Met residue at the N-terminus, which may be derived either from a real insertion of Met or from an initiation Met that is usually cleaved post-translationally, consequently consisting of 154 amino acids [142], while the penguin myoglobin lacks one amino acid at position 71 (E14) and is composed of 152 amino acids [143]. The myoglobins of the three classes have a high amino acid sequence identity (over 60%).
The myoglobins of Chondrichthyes and Osteichthyes are composed of 148 and 146 amino acid residues, respectively, and their N-terminus is acetylated. Deletions of the N-terminal 4 residues and 1 residue at CD7 and in addition, deletion of 2 amino acids at the GH corner in Osteichthyes shorten the polypeptides of Chondrichthyes and Osteichthyes myoglobins by 5 to 7 residues, compared with mammalian myoglobins [144].

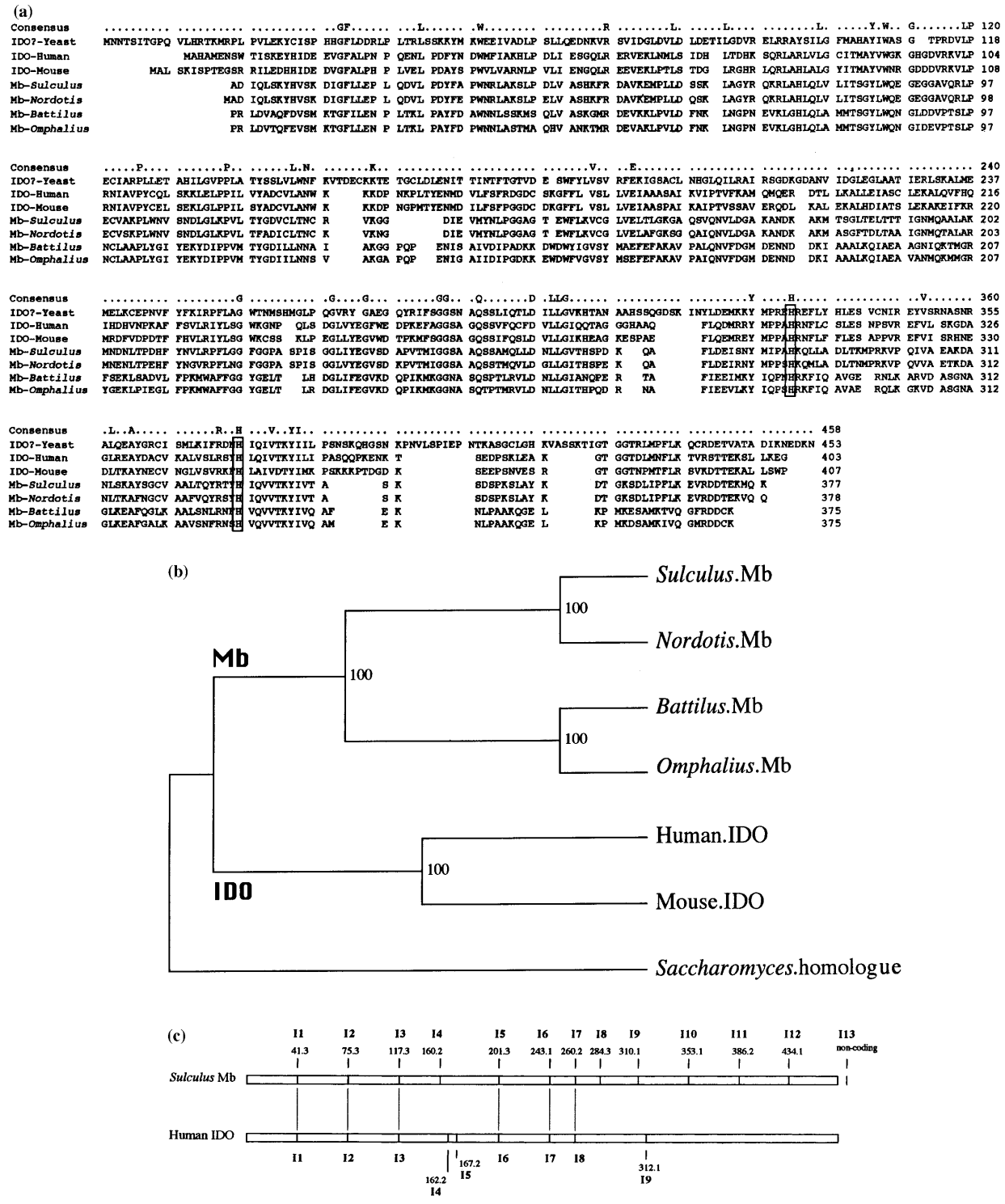

Figure 4. (*a*) Alignment of amino acid sequences of IDO-like myoglobins, IDOs and a homologue in yeast. The boxed residues indicate the tentatively assigned distal and haem-binding proximal histidines. This alignment was obtained using the algorithm of Feng and Doolittle [204]. (*b*) The phylogenetic relationship of IDOs, IDO-like myoglobins and a homologue in yeast. This tree was obtained using the parsimony method in the PHYLIP package, ver. 3.5c [178]. The numbers at branching points indicate the bootstrap values. (*c*) Comparison of gene structure of IDO and IDO-like myoglobin. The position of intron is indicated by a vertical line. For example, the number 41.3 shows that the intron is inserted after the third nucleotide of a codon corresponding to the amino acid sequence position 41 in the alignment shown in figure 3a.

The sequences show significant homology (40–45% identity) with those of mammalian myoglobins.

The myoglobin from the heart muscle of the bull frog *Rana catesbeiana*, consisting of 132 amino acid residues, has been sequened by Maeda and Fitch [145] as a result of extensive efforts to isolate myoglobin from amphibians. The protein was approximately 20 amino acids shorter than mammalian myoglobin.

Amino acid sequences of myoglobins from sperm whale, human, penguin, chicken, map turtle, alligator, bullfrog *Rana*, yellowfin tuna, carp, icefish, modern sharks *Mustelus* and *Galeorhinus* and the primitive shark *Heterodontus* are aligned in figure 3a, and a phylogenetic relationship among the sequences is given in figure 3b.

Gene structure

The first elucidation of globin gene structure was that of the human haemoglobin α and β chains [146, 147]. The genes have two introns conserved exactly at positions B12.2 and G6.3. Go [148] predicted that globin genes should have an additional intron at the central position on the basis of an analysis using distance maps which revealed correlations between DNA exons and protein structural units (modules). This predicted intron was soon after discovered in the leghaemoglobin gene of the soybean *Glycine max*, which consequently has three introns [149]. Thus, until the end of the 1980s, it was widely believed that all of the eukaryotic globin genes had evolved from a primordial ancestor with three introns; the three-intron/four-exon structure had been conserved only in plant globins, whereas animal globins lost one intron located in the central position but retained two remaining introns at strictly conserved positions.

In 1992, Dixon and Pohajdak proposed a new theory that the two-intron/three-exon structure of vertebrate globin genes reflects the ancestral form, and the central intron has been gained independently, following the divergence of plants and animals [150]. This theory is based on their findings that two nematode globin genes from *Pseudoterranova* and *Caenorhabditis* have central introns, but the position and phase of the central intron have not been conserved in the three globin genes *Glycine*, *Pseudoterranova* and *Caenorhabitis*.

Thus the origin of the central intron in globin genes is still controversial. According to the former theory, unconventional position and phase of the central intron are explained by the sliding of the intron [151]. On the other hand, the latter theory claims that the evolution of the intron-exon pattern in globin genes should only be explained in terms of intron loss and independent intron gain [150].

In this section, we summarize the gene structures of myoglobins known so far, together with the genes for haemoglobins, to make clear the process of globin evolution.

Minihaemoglobin

Of the minihaemoglobin genes, eukaryotic genes from the protozoan *Paramecium* and green alga *Chlamydomonas* contain introns, whereas other bacterial genes from *Nostoc*, *Synechococcus* and *Mycobacterium* do not. As shown in figure 2c, *Paramecium* globin genes have a single intron (75.3) in the middle of the protein [152]. On the other hand, the *Chlamydomonas* gene has three introns (28.3, 66.3 and 82.3) [48], none of which corresponds to the intron position of *Paramecium*. Moreover, globin gene from another protozoan, *Tetrahymena*, contains no introns [119]. Since there are no large deletions or insertions in the alignment of miniglobins (see fig. 2a), judgements of intron positions would be correct. At present, we do not know whether an ancestral gene for minihaemoglobin retained at least four introns and some of the introns were lost during evolution, or whether the introns were gained independently.

Vertebrate-like globin

Positions of introns of myoglobins and haemoglobins are summarized in figure 3c.

Bacteria and yeast. Globin or flavohaemoglobin genes for *Vitreoscilla* [153], *Rhizobium* [23], *E. coli* [29] and *Saccharomyces* [27] contain no introns.

Plants. All of the plant haemoglobin genes from *Glycine*, *Sesbania*, *Medicago*, *Trema*, *Parasponia* and *Casuarina* have three introns at positions B12.2, E14.3 and G6.3 (fig. 3c) [126, 154, 155].

Invertebrates. There is considerable variation in the intron positions of invertebrate globin genes (see fig. 3c). To date, gene structures for four invertebrate myoglobins, three from nematodes and one from an annelid have been determined. The myoglobin gene of the nematode *Ascaris suum* contained no introns [75]. On the other hand, the gene for another nematode myoglobin, *Caenorhabditis elegans*, contains a single intron at position E3.2, which is located close to but is not in the same position as the central intron (E14.3) of plant globin genes [133]. Two myoglobin genes from the nematode *Nippostrongylus brasiliensis* also have a single intron located at E3.2. Blaxter et al. [76] suggest that the exact position of introns in *Nippostrongylus* and *Caenorhabditis* genes indicate that it arose from an independent insertion event in the Strongylida-Rhabditia line. A nerve haemoglobin gene from the polychaete *Aphrodite aculeata* has been determined recently [134]. It contains a single intron at position G6.3, and the intron corresponds exactly to one of the two introns conserved in all vertebrate globin genes.

Invertebrate haemoglobin genes also show a variation in intron positions (see fig. 3c). The gene for the earthworm *Lumbricus terrestris* has two introns at positions B12.2 and G.6.3 [156]. However, the inconsistency of the 5′ noncoding sequences between cDNA and genomic DNA of *Lumbricus* haemoglobin is indicative of the presence of an additional intron in the preceding region. The genes for extracellular two-domain haemoglobins from the nematodes *Ascaris suum* and *Pseudoterranova decipiens* are of special interest. In both repeats of *Ascaris* genes, there are four introns at positions A4.1, B12.2, E8.1 and G6.3 [157, 205]. On the other hand, repeat 1 of *Pseudoterranova* gene has four introns at positions A4.1, B12.2, E8.1 and G6.3, like *Ascaris* repeats, but repeat 2 is lacking two introns at positions E8.1 and G6.3. The intron in the A-helix (A4.1) of repeat 1 is located very close to the position between the secretory leader sequence and mature protein, and this type of intron is widely observed in genes coding eukaryotic extracellular proteins [158]. The A4.1 intron in repeat 2 is considered to be a so-called bridge intron which separates the two repeats [3] rather than as an intradomain intron.

The exon-intron organization for a nine-domain haemoglobin gene from the brine shrimp *Artemia* is noted [159]. In the nine repeats, introns are found in five different positions (B12.2, F2.3, F3.1, G6.2 and G6.3). Jellie et al. [159] proposed that the unconventionally located introns (F2.3, F3.1 and G6.2) in repeats 3, 4 and 6 have moved from the conventionally located intron G6.3, based on the observation that the repeats with unconventional introns do not have the G6.3 intron. On the other hand, the haemoglobin genes IIβ and IX from another arthropod *Chironomus thummi thummi* have a single intron at position E9.1, in contrast to most other *Chronomus* haemoglobin genes that have no intron [160].

The blood clam *Barbatia reeveana* expresses two types of haemoglobins, a tetramer of α and β chains and a polymer consisting of two-domain chain, in erythrocytes [161]. The gene structure for the two-domain haemoglobin has been determined [3]. It contained the precoding and bridge introns, in addition to two conventional introns at positions B12.2 and G6.3. The precoding intron is located one nucleotide upstream from the initiation codon ATG. The gene structure having two introns at B12.2 and G6.3 for a minor tetrameric haemoglobin from the blood clam *Anadara trapezia* has also been determined [162]. Although Totchen et al. [162] do not mention it, we can find an indication of a precoding intron in the same position as in *Barbatia* gene. The gene structures of the coding regions for tetrameric haemoglobin from the blood clam *Scapharca inaequivalvis* were determined recently [163]. Both genes have two introns at positions B12.2 and G6.3.

One of the authors sequenced the haemoglobin genes of another blood clam *Barbatia lima*, a closely related species to *Barbatia reeveana* (T. Suzuki, unpublished results). *B. lima* expresses three types of intracellular haemoglobins, a homodimer of chain δ, a tetramer of chains α and β, and a polymer consisting of a two-domain chain and the δ chain [164]. This is in sharp contrast to the haemoglobin composition (tetramer and polymer) of *B. reeveana*. The *B. lima* δ chain, which is not expressed in *B. reeveana*, is the ancestral single-domain chain for the two-domain chain [164]. The gene for the δ chain contained three introns, a precoding intron and two conventional introns B12.2 and G6.3, while the structure for the two-domain chain was identical with that for the *B. reeveana* two-domain chain. We have also confirmed the presence of a precoding intron in the gene for an intracellular, heterodimeric haemoglobin from the related clam *Barbatia virescens*. Thus, it was suggested that the ancestral gene for intracellular clam haemoglobins should have the precoding intron, in addition to B12.2 and G6.3 introns. Furthermore, two genes of homodimeric haemoglobins from the deep-sea clam *Calyptogena soyoae* [165] were sequenced (T. Suzuki and S. Ohta, unpublished results). Surprisingly, it possessed no precoding intron, but alternatively an additional intron in the A helix (A3.1), together with the two conserved introns (B12.2 and G6.3). This observation strongly suggests that the shift of the intron from precoding to A-helix occurred in the haemoglobin genes of *Calyptogena*.

Vertebrate. Gene structures for typical vertebrate myoglobins from gray seal [166], human [167], mouse [168] and *Chionodraco rastrospinosus* (icefish) (U71059) have been determined. All genes had two introns at conventional positions B12.2 and G6.3 (fig. 3c). A pseudo-gene for myoglobin from *Chaenocephalus aceratus* is also reported (U71153).

IDO-like myoglobin

The gene for the 39-kDa myoglobin from the abalone *Sulculus* was amplified by polymerase chain reaction (PCR), and its structure determined [169]. The gene has a 14-exon/13-intron structure, which is quite different from that of usual globin gene but is homologous with that (10-exon/9-intron) of the human indoleamine dioxygenase gene [170]: the splice junctions of 6 introns were conserved exactly between the two genes (see fig. 4c), suggesting that these introns have been conserved at least 600 million years [169].

Three distinct roots for haem-containing oxygen-binding proteins

It has long been thought that all haemoglobins and myoglobins have evolved from a common ancestral gene.

However, the discovery of IDO-like myoglobin from the abalone *Sulculus diversicolor* clearly indicated that there was an alternative origin for myoglobin evolution. Furthermore accumulation of knowledge about amino acid sequences and gene structures of myoglobin or myoglobin-like proteins from various taxa now makes it possible to separate them into three distinct categories, universal globin, compact globin and IDO-like globin (fig. 5). This idea was first proposed by Go and co-workers [171]. A similar picture is seen in the evolution of haemocyanins: molluscan and arthropod haemocyanins have evolved from quite different ancestors, tyrosinase and insect storage proteins, respectively [31].

Universal globin

Most of the myoglobins and haemoglobins, distributed from bacteria to mammals, fall under the 'universal globin', for which we propose to use the term 'globin superfamily'. The mature universal globin is usually composed of 145 to 155 amino acid residues with exceptions: the shortest globin (132 residues) is the cardiac myoglobin from *Rana* [145], and the longest (163 residues) is the haemoglobin α chain from the blood clam *Barbatia* [164]. To accommodate all the globin sequences so far known (about 700) using computer-assisted alignment, 183 positions are needed, of which 84 are common to all, including the absolutely conserved CD1-Phe and F8-His [172]. The mean number of amino acid substitutions per position ranges from 8 to 13 for all globins, and 5 to 9 even for internal positions, indicating that the globin sequences are remarkably deviated [172]. However, the crystal structures of universal globins have the same 'myoglobin fold' (fig. 1) [2], supporting their common evolutionary origin.

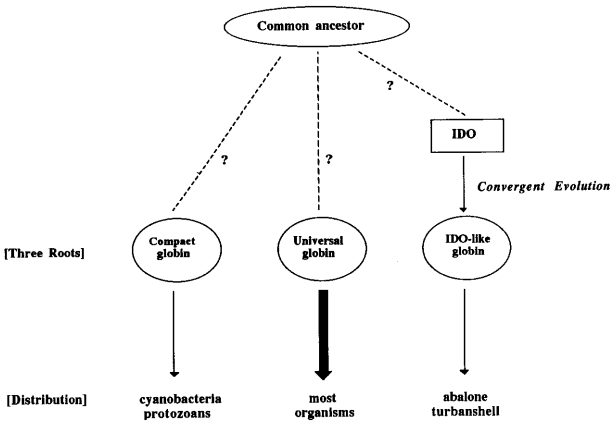

Figure 5. Three roots for haem-binding oxygen-binding proteins.

Compact globin

The minihaemoglobins usually consist of about 120 amino acid residues, and this feature discriminates them from other globins. These globins are called 'cyanoglobin' [46], 'truncated globin' [1] or 'contracted globin' [118], but we propose a new name 'compact globin' for these globins. The origin of compact globins is still controversial. Takagi [173], suggested that *Paramecium* and *Tetrahymena* myoglobins might be evolved from a different ancestral gene, not from a common globin gene, since they show no significant homology with any other globins so far known. On the contrary, Moens et al. [174] concluded that these proteins fit all the essential determinants of the myoglobin fold, and the scores for their V template [175] and new NV template [174] are still comparable to those for other nonvertebrate globins.

A unique globin folding was suggested for *Paramecium* compact haemoglobin by kinetic analyses of autoxidation rate. An amino acid sequence alignment of *Paramecium* haemoglobin with vertebrate and invertebrate myoglobins suggests that the distal E7-His is replaced by Gln in *Paramecium* haemoglobin [1, 48, 174]. The presence or absence of the distal E7-His has a substantial effect on the autoxidation of myoglobin. In fact, the pH dependences for autoxidation of *Aplysia* myoglobin with E7-Val [105], *Dolabella* myoglobin with E7-Val [99] and *Galeorhinus* myoglobin with E7-Gln [176] are quite different from that of sperm whale myoglobin possessing the distal His. However, the pH dependence of *Paramecium* haemoglobin was quite similar to that of sperm whale myoglobin [177]. In addition, the spontaneous formation of hemichrome in *Paramecium* methaemoglobin would support the presence of histidine residue as the sixth ligand of haem iron [177]. On the basis of these facts, we assume that *Paramecium* haemoglobin would have a globin-folding structure distinctly different from the usual myoglobin fold.

Alignment of amino acid sequences of compact globins (fig. 2a) indicates that the one histidine residue conserved at position 81 in all sequences is the most probable candidate for the haem-binding proximal histidine. If the *Paramecium* globin retains the histidine residue near the sixth coordination site of haem iron as suggested by Shikama and co-workers [118, 177], the so-called distal histidine is assigned to His at position 101. Thus, the topology of the distal and proximal residues in compact globins are assumed to be inverse compared with that of universal globins (compare figs 2a and 3a).

Furthermore, it should be pointed out that prokaryotes possess both a compact globin (in *Nostoc*, *Synechococcus* and *Mycobacterium*) and a typical universal globin (in *Vitreoscilla* and *E. coli*). This fact itself implies that there were two primordial genes for oxygen-binding haemoproteins.

Determination of the crystal structure of compact globins is needed to resolve this problem. We conclude provisionally, that current evidence suggests that compact globins have a distinct origin.

IDO-like globin

The cDNA-derived amino acid sequence and the gene of myoglobin from the abalone *Sulculus* clearly indicate that this myoglobin was derived from a copy of the gene for IDO [108, 169]. No notable homology was noted among the sequences of *Sulculus* myoglobin and other universal or compact globins. Moreover, the presence of a homologue of IDO or IDO-like myoglobin gene in the yeast *Saccharomyces* strongly suggests an ancient origin of IDO-related genes [199]. Thus, IDO-like myoglobin represents an alternative path for myoglobin evolution.

In the sequence alignment of IDO and IDO-like myoglobins in figure 4a, two histidine residues at positions 335 and 380 (boxed) are conserved in all sequences, one of which appears most likely to be functioning as the iron-bound proximal histidine. Since they are also conserved in the *Saccharomyces* sequence, the *Saccharomyces* homologue may function either as an IDO or IDO-like myoglobin. If histidine-380 is assigned to the iron-bound proximal histidine, then histidine-335 may be assigned to the distal histidine by analogy to universal myoglobins and haemoglobins. An analysis of pH dependence of autoxidation rate for *Battilus* IDO-like oxymyoglobin suggests that an amino acid residue with the pK_a value of 5.4, which may be assigned to histidine, is actually involved in the reaction [208].

Evolution

At present, we have no evidence to make clear the relationship among the three distinct haem-containing oxygen-binding proteins. Although we have to wait for the crystal structures of a compact globin and an IDO-like globin to be solved, it might be possible to assume that at least a portion of these three proteins evolved from a common ancester (fig. 5).

Universal globin

A phylogenetic tree constructed using a protein parsimony algorithm [178] clearly separated all the universal globins into four distinct clusters: bacteria and yeast globins, plant globins, invertebrate myoglobins and vertebrate myoglobins (fig. 3b).

The sequences of bacterial *Vitreoscilla* haemoglobin and *E. coli* flavohaemoglobin have an unusually high homology (31–53% identity) with those of the flavohaemoglobins from the eukaryotic unicellular yeasts, *Saccharomyces* and *Candida*, and they form a separate group with a bootstrap value of 100% (fig. 3b). Since bacteria and eukaryotes diverged at least 1800 million years ago [179], the evolutionary rate (pauling: amino acid displacement rate/one amino acid site/1000 million years) for the globins would be estimated roughly to be 0.33 at maximum. This rate is much slower than that of invertebrate and vertebrate myoglobins. Zhu and Riggs [27] suggest that this high homology is derived from the special functional roles for flavohaemoglobins, unrelated to haemoglobin or myoglobin functions. They proposed that the haem domain of flavohaemoglobins may act as a molecular switch to control the flavo domain through conformational changes associated with oxygen binding, but not as an oxygen carrier. Thus it appears that the globin domains of flavohaemoglobins have evolved independently of other universal haemoglobins and myoglobins. Alternatively, Moens et al. [174] attributed the unusually high homology between the prokaryote and eukaryote globins to horizontal globin gene transfer.

Runnegar [180] pointed out that globins, cytochromes *b*5 and *b*2 and chicken sulphite oxidase point to a common ancestry. It was proposed that all globins evolved from a family of ancestral haemoprotins with MW of 17 kDa; this family had a globin fold and functioned as redox proteins [174]. In our definition, these globins are categorized under universal globin. Moens et al. [174] hypothesized that redox haemoproteins existed during the early Proterozoic when dioxygen had not yet accumulated in the atmosphere; oxygen-reactive globins likely evolved as atmospheric oxygen appeared and increased after the radiation of photosynthetic bacteria, approximately 2000 million years ago; high oxygen affinity may have evolved repeatedly due to a selective pressure towards enhanced oxygen affinity that was required for oxygen binding and scavenging. Such functional shifts could arise from minor structural changes in the vicinity of the haem group without any alteration in the protein fold as manifested in recombinant myoglobins [181].

The distal E7 residue is occupied by a Gln residue in all the four bacteria and yeast globins (fig. 3a). It is likely that the E7 Gln is required for their unique function and that it represents the ancestral residue in universal globins.

Plant globins share a high sequence homology (42–93% identity) with each other and form a separate cluster with a bootstrap value of 100% (fig. 3b). All the globins conserve the distal E7 histidine (fig. 3a).

The invertebrate myoglobin cluster contains 16 sequences from four phyla, Annelida, Mollusca, Nematode and Platyhelminthes (fig. 3b). The fossil record indicates that metazoans have become abundant during the Edi-

acarian period of the late Precambrian (670 to 550 million years ago) [182, 183]. The bootstrap values of higher branching points within the cluster is very low (16–44%), suggesting that invertebrate myoglobin sequences are markedly diverse: the sequence identity between *Aphrodite* nerve haemo-globin and *Isoparorchis* body wall myoglobin is only 11%. This might reflect the diverse functional properties of invertebrate myoglobins (or haemoglobins), adapted in various living conditions or tissue specificity during evolution. Consistent with that idea, one of the functional key residues, distal-E7 His, is frequently replaced by Val, Tyr, Leu or Gln in invertebrate myoglobins, and such replacement would alter the functional properties dramatically.

In the vertebrate myoglobin cluster (fig. 3b), the sequence of the amphibian *Rana* is distantly related to other vertebrate myoglobins, with only 19–33% sequence identity. This comes from the fact that *Rana* myoglobin is unusually homologous with α chains of vertebrate haemoglobins [145]. Maeda and Fitch [145] assumed that in *Rana* heart muscle a normal myoglobin was lost but, instead, a haemoglobin α chain homologue has evolved. Except for *Rana*, vertebrate myoglobins share a high homology (more than 34% sequence identity), and the branching pattern is essentially consistent with the classical taxonomy. Since the myoglobins of Osteichthes diverged about 400 million years ago [184], the evolutionary rate (in pauling units) for vertebrate myoglobins would be estimated roughly to be 1.4 at most.

Although the presence of a histidine residue at the E7 position is conventional in most vertebrate myoglobins and haemoglobins (fig. 3a), we note that an exceptional amino acid replacement (His to Gln) at E7 position occurrs in the myoglobins from three modern sharks *Mustelus antarcticus* [185], *Galeorhinus australis* [186] and *G. japonicus* [187] and in Indian and African elephants [188, 189]. In vertebrate haemoglobins, the same replacement is also found in hagfish monomer haemoglobin [190] and the α chain of opossum haemoglobin [191]. The functional properties of myoglobins with Gln at E7 are comparable to those of myoglobins with His-E7: Gln-E7 is also able to form a hydrogen bond with the bound dioxygen, like His-E7 [189, 192, 193].

How can we assume the exon/intron organization of a primordial gene for universal globin? It was believed for a long time that an ancestral globin gene had three-intron/four-exon structure, and an exon corresponds to a minimal structural unit (module) of globin [148]. This idea was reinforced by a detailed study on the pattern of introns in a nine-domain globin gene of *Artemia* [159]. Jellie et al. [159] proposed a concept of general positional stability of the three introns in ancestral globin gene, but also suggested that sliding or deletion of introns occurs after inheritance from an ancient gene. On the other hand, Dixon and Pohajdak [150] claim that a primordial globin gene had two introns just like the vertebrate globin gene, and the unconventional central intron was gained independently in various lines of evolution. Ascoli and co-workers support this idea [163].

Two or three introns in a primordial globin gene? On the assumption that in some cases introns can slide or be deleted, the comparison of gene structures of all universal globins in figure 3c appears to support the three-intron structure, because most of the introns in the coding region are classified into three locations, the absolutely conserved B-helix intron (B12.2), the central E-helix intron and the F- or G-helix intron, and we do not find any inserted introns outside the three locations. However, several questions remains to be unsolved. Why is the position of the central intron in E-helix the most variable? The E-helix is the functional key region in myoglobins and haemoglobins and possesses two conserved residues, His-E7 and Val-E11, to control O_2 affinity. On the other hand, why was the B12.2 intron conserved absolutely in most of the globin genes?

Furthermore, we note the significance of the precoding or A3.1 and A4.1 introns in globin genes, which were found in globin genes from several invertebrate phyla, Nematode (*Ascaris* and *Pseudoterranova*), Mollusca (*Barbatia reeveana*, *B. lima*, *B. virescens*, *Anadara* and *Calyptogena*) and presumably Annelida (*Lumbricus*). In the *Barbatia* two-domain globin gene, the precoding intron plays a crucial role in the fusion of the duplicated single-domain globin genes [3]. In nematode extracellular globin genes, it separates the secretory leader peptide and the mature protein [150]. Thus it is natural to consider that in the primordial globin gene, the precoding or A-helix intron was an important element in promoting various types of globin evolution. We propose here that a primordial globin gene contained a total of four introns, including the precoding one (fig. 3c).

Compact globin

Compact globins are distributed in protozoans, *Chlamydomonas* and bacteria (fig. 2b), and they show a high homology (more than 37% sequence identity). The evolutionary rate (pauling) was roughly estimated to be 0.28 at maximum, which is comparable to that for flavohaemoglobins but is significantly slower than that of other universal globins. Based on this unusually high homology among the compact globins, Moens et al. [174] posit horizontal compact globin gene transfer from an ancestor common to protozoans and *Clamydomonas* to a bacterial ancestor.

IDO-like globin

The distribution of IDO-like myoglobin was unexpectedly wide, but it is not found in any other animals except gastropod molluscs.

A phylogenetic tree constructed from the amino acid sequences of IDOs and IDO-like myoglobins shows that the IDO sequences were clearly separated from IDO-like myoglobin sequences, and the sequence of the yeast homologue is outside of the two major clusters (fig. 4b).

When did the IDO-like myoglobin arise? The chitons, one of the most primitive molluscs, have the universal 16-kDa myoglobin, as do the more advanced gastropod molluscs *Aplysia* and *Dolabella*. Thus the IDO-like myoglobin must have arisen exclusively along the lineage of *Sulculus* in molluscan evolution.

Molecular phylogeny using the amino acid sequences of arginine kinases from *Nordotis*, *Battilus*, *Liolophula* (chiton) and *Penaeus* (shrimp) indicates that the abalone *Nordotis* and the turban shell *Battilus* diverged about 220 million years ago [194]. *Nordotis* and *Battilus* express IDO-like myoglobins, but not universal globins. Hence, the IDO-like myoglobin must already have been present in the common ancestor of *Nordotis* and *Battilus*. The period of 220 million years results in a 65% amino acid sequence difference in the two IDO-like myoglobins, and this evolutionary rate (2.4 pauling) seems to be about two times faster than in universal haemoglobins and myoglobins. This is consistent with the observation that vertebrate IDOs evolve faster than haemoglobins [110].

Previously we proposed that the gene for a 'normal' myoglobin was altered or lost in *Sulculus* and that a modified indoleamine dioxygenase, which should have derived from a duplicated gene for the true IDO, evolved as a substitute [108]. To obtain a functional analogue of myoglobin, the enzymatic activity of IDO was probably altered to favour the reversible binding of oxygen, by introducing point mutations in or near the haem cavity. In agreement with this proposal, no IDO enzymatic activity was found in *Sulculus* myoglobin [108]. Thus, *Sulculus* myoglobin represents a typical case of functional convergence, and suggests that molecular evolution of proteins is adaptable. We have not yet determined whether a separate indoleamine dioxygenase gene or an active IDO is present in *Sulculus*.

We compared the amino acid sequence corresponding to each exon of *Sulculus* myoglobin with those of ordinary globin sequences so far known to find if there was any evolutionary relationship between them. The search on exons 10 and 11 of *Sulculus* myoglobin was especially important because they contain the conserved histidine residues, one of which is the most probable candidate for the iron-bound proximal histidine. However, we found no indication of a significant evolutionary relationship between the IDO-like myoglobin and universal myoglobin and haemoglobin.

Very recently, we found that the 16-kDa myoglobin of the gastropod mollusc *Theliostyla albicilla* is present mostly in an oxidized met-form in vivo (T. Suzuki et al., unpublished results). The met-form cannot bind oxygen and is physiologically inactive. Therefore, *Theliostyla* myoglobin might correspond to the 'altered' myoglobin postulated earlier, but *Theliostyla* does not have IDO-like myoglobin. Alternatively, it might lack the myoglobin reduction system. The complete amino acid sequence of *Theliostyla* myoglobin shows that it has an unusually long N-terminal extension (seven residues), compared with other molluscan myoglobins, but it retains the three functional key residues, CD1-Phe, E7-His and F8-His (T. Suzuki, H. Takao and T. Takagi, unpublished results).

Acknowledgement. We thank Dr. S. N. Vinogradov of Wayne State University for reading the manuscript and giving us valuable suggestions.

1 Vinogradov S. N., Walz D. A., Pohajdak B., Moens L., Kapp O., Suzuki T. et al. (1993) Adventitious variability? The amino acid sequences of nonvertebrate globins. Comp. Biochem. Physiol. **106B:** 1–26
2 Bolognesi M., Bordo D., Rizzi M., Tarricone C. and Ascenzi P. (1997) Nonvertebrate hemoglobins: structural bases for reactivity. Prog. Biophys. Mol. Biol. **68:** 29–68
3 Naito Y., Riggs C. K., Vandergon T. L. and Riggs A. F. (1991) Origin of a 'bridge' intron in the gene for a two-domain globin. Proc. Natl. Acad. Sci. USA **88:** 6672–6676
4 Manning A. M., Trotman C. N. A. and Tate W. P. (1990) Evolution of a polymeric globin in the brine shrimp *Artemia*. Nature **348:** 653–656
5 Suzuki T. and Furukohri T. (1989) The ear-shell (*Sulculus diversicolor aquatilis*) myoglobin is composed of an unusual 39 kDa polypeptide chain. Experientia **45:** 998–1002
6 Iwaasa H., Takagi T. and Shikama K. (1989) Protozoan myoglobin from *Paramecium caudatum*. J. Mol. Biol. **208:** 355–358
7 Wittenberg J. B. (1970) Myoglobin-facilitated oxygen diffusion: role of myoglobin in oxygen entry into muscle. Physiol. Rev. **50:** 559–636
8 Imai K. (1981) Analysis of ligand binding equilibria. Methods Ezymol. **76:** 470–486
9 Imai K. (1982) Allosteric Effects in Haemoglobin, Cambridge University Press, London
10 Millikan G. A. (1937) Experiments on muscle haemoglobin in vivo; the instantaneous measurement of muscle metabolism. Proc. R. Soc. (Lond.) **B123:** 218–241
11 Millikan G. A. (1939) Muscle hemoglobin. Physiol. Rev. **19:** 503–523
12 Wyman J. (1966) Facilitated diffusion and the possible role of myoglobin as a transport mechanism. J. Biol. Chem. **241:** 115–121
13 Wittenberg J. B. (1959) Oxygen transport – a new function proposed for myoglobin. Biol. Bull. **117:** 402–403
14 Scholander P. F. (1960) Oxygen transport through hemoglobin solutions. Science **131:** 585–590
15 Wittenberg B. A. and Wittenberg J. B. (1989) Transport of oxygen in muscle. Annu. Rev. Physiol. **51:** 857–878
16 Jones D. P. and Kennedy F. G. (1982) Intracellular O_2 gradients in cardiac myocytes. Lack of a role for myoglobin in facilitation of intracellular O_2 diffusion. Biochem. Biophys. Res. Commun. **105:** 419–424
17 Jones D. P. and Kennedy F. G. (1986) Analysis of intracellular oxygenation of isolated adult cardiac myocytes. Am. J. Physiol. **250:** C384–C390
18 Jurgens K. D., Peters T. and Gros G. (1994) Diffusivity of myoglobin in intact skeletal muscle cells. Proc. Natl. Acad. Sci. USA **91:** 3829–3833
19 Paradopoulos S., Jurgens K. D. and Gros G. (1995) Diffusion of myoglobin in skeletal muscle cells – dependence on fibre typem contraction and temperature. Pflugers Arch. **430:** 519–525

20 Wittenberg J. B. and Wittenberg B. A. (1990) Mechanisms of cytoplasmic haemoglobin and myoglobin function. Annu. Rev. Biophys. Biophys. Chem. **19:** 217–241
21 Doeller J. E. and Wittenberg B. A. (1991) Myoglobin function and energy metabolism of isolated cardiac myocytes: effect of sodium nitrite. Am. J. Physiol. **261:** H53–H62
22 Powell R. and Gannon F. (1988) The leghaemoglobins. Bioessays **9:** 117–121
23 Gilles-Gonzalez M. A., Ditta G. S. and Helinski D. R. (1991) A haemoprotein with kinase activity encoded by the oxygen sensor of *Rhizobium meliloti*. Nature **350:** 170–172
24 Gilles-Gonzalez M. A., Gonzalez G., Perutz M. F., Kiger L., Marden M. C. and Poyart C. (1994) Heme-based sensors, exemplified by the kinase FixL, are a new class of heme protein with distinctive ligand binding and autoxidation. Biochemistry **33:** 8067–8073
25 Winkler W. C., Gonzalez G., Wittenberg J. B., Hille R., Dakappagari N., Jacob A. et al. (1996) Nonsteric factors dominate binding of nitric oxide, azide, imidazole, cyanide and fluoride to the Rhizobial heme-based oxygen sensor FixL. Chem. Biol. **3:** 841–850
26 Oshino R., Asakura T., Takio K., Oshino N., Chance B. and Hagihara B. (1973) Purification and molecular properties of yeast haemoglobin. Eur. J. Biochem. **39:** 581–590
27 Zhu H. and Riggs A. F. (1992) Yeast flavohaemoglobin is an ancient protein related to globins and a reductase family. Proc. Natl. Acad. Sci. USA **89:** 5015–5019
28 Probst I., Wolf G. and Schlegel H. G. (1979) An oxygen-binding flavohemoprotein from *Alcaligenes eutrophus*. Biochim. Biophys. Acta **576:** 471–478
29 Vasudevan S. G., Armarego W. L. F., Shaw D. C., Lilley P. E., Dixon N. E. and Poole R. K. (1991) Isolation and nucleotide sequence of the *hmp* gene that encodes a haemoglobin-like protein in *Escherichia coli* K-12. Mol. Gen. Genet. **226:** 49–58
30 Giardina B., Ascenzi P., Clementi M. E., De Sanctis G., Rizzi M. and Coletta M. (1996) Functional modulation my lactate of myoglobin. J. Biol. Chem. **271:** 16999–17001
31 van Holde K. E. and Miller K. I. (1995) Hemocyanins. Adv. Protein Chem. **47:** 1–81
32 Chung M. C. M. and Ellerton H. D. (1979) The physico-chemical and functional properties of extracellular respiratory haemoglobins and chlorocruorins. Prog. Biophys. Mol. Biol. **35:** 53–102
33 Weber R. E. (1980) Functions of invertebrate haemoglobins with special reference to adaptations to environmental hypoxia. Am. Zool. **20:** 79–101
34 van Holde K. E. and Miller K. I. (1982) Haemocyanins. Q. Rev. Biophys. **15:** 1–129
35 Vinogradov S. N. (1985) The structure of invertebrate extracellular haemoglobins (erythrocruorins and chlorocruorins). Comp. Biochem. Physiol. **82B:** 1–15
36 Mangum C. P. (1985) Oxygen transport in invertebrates. Am. J. Physiol. **248:** R505–R514
37 Bunn H. F. and Forget B. G. (1986) Hemoglobin: Molecular, Genetic and Clinical Aspects, Saunders, Philadelphia
38 Gotoh T. and Suzuki T. (1990) Molecular assembly and evolution of multi-subunit extracellular annelid haemoglobins. Zool. Sci. **7:** 1–16
39 Kurtz D. M. (1992) Molecular structure/function relationships of hemerythrins. Advan. Comp. Environ. Physiol. **13:** 151–171
40 Mangum C. P. (1992) Physiological function of the hemerythrins. Adv. Comp. Environ. Physiol. **13:** 173–192
41 Terwilliger N. B. (1992) Molecular structure of the extracellular heme proteins. Adv. Comp. Environ. Physiol. **13:** 193–230
42 Perutz M. F., Wilkinson A. J., Paoli M. and Dodson G. G. (1998) The stereochemical mechanism of the cooperative effects in hemoglobin revisited. Annu. Rev. Biophys. Biomol. Struct. **27:** 1–34
43 Gibson Q. H. and Smith M. H. (1965) Rates of reaction of *Ascaris* haemoglobins with ligands. Proc. R. Soc. (Lond.) **B163:** 206–214
44 Webster D. A. (1987) Structure and function of bacterial haemoglobin and related proteins. Adv. Inorg. Biochem. **7:** 245–262
45 Ditta G., Virts E., Palomeres A. and Kim C. H. (1987) The *niA* gene of *Rhizobium* is oxygen-regulated. J. Bacteriol. **169:** 3217–3223
46 Potts M., Angeloni S. V., Ebel R. E. and Bassam D. (1992) Myoglobin in a cyanobacterium. Science **256:** 1690–1692
47 Oshino R., Oshino N., Chance B. and Hagihara B. (1973) Studies on yeast haemoglobin. The properties of yeast haemoglobin and its physiological function in the cell. Eur. J. Biochem. **35:** 23–33
48 Couture M., Chamberland H., St.-Pierre B., Lafontaine J. and Guertin M. (1994) Nuclear genes encoding chloroplast haemoglobins in the unicellular green alga *Chlamydomonas eugametos*. Mol. Gen. Genet. **243:** 185–197
49 Sato T. and Tamiya H. (1937) Über die Atmungsfarbstoffe von *Paramecium*. Cytologia (Fujii Jubilee volume): 1133–1138
50 Keilin D. and Ryley J. F. (1953) Haemoglobin in protozoa. Nature **172:** 451
51 Irie T. and Usuki I. (1980) Disparity of native oxyhaemoglobin components isolated from *Paramecium caudatum* and *Paramecium primaurelia*. Comp. Biochem. Physiol. **67B:** 549–554
52 Steers E. Jr., Barnett A. and Lee C. E. (1981) Isolation and characterization of the haemoglobin from *Paramecium caudatum*. Comp. Biochem. Physiol. **70B:** 185–191
53 Steers E. Jr. and Davis R. H. Jr. (1979) Purification and characterization of the myoglobins of *Paramecium tetraurelia*. Comp. Biochem. Physiol. **62B:** 393–402
54 Iwaasa H., Takagi T. and Shikama K. (1990) Protozoan haemoglobin from *Tetrahymena pyriformis*. J. Biol. Chem. **265:** 8603–8609
55 Davis R. H. Jr. and Steers E. Jr. (1976) Myoglobin from the ciliate protozoan *Paramecium aurelia*. Comp. Biochem. Physiol. **54B:** 141–143
56 Usuki I. and Irie T. (1983) Inter- and intrasyngenic variation of *Paramecium* hemoglobin-I. *Paramecium aurelia* complex. Comp. Biochem. Physiol. **75B:** 415–420, 23–33
57 Smith M. H., George P. and Preer J. R. (1962) Preliminary observations on isolated *Paramecium hemoglobin*. Arch. Biochem. Biophys. **99:** 313–318
58 Fuchsmann W. H. and Appleby C. A. (1979) Separation and determination of the relative concentrations of the homogeneous components of soybean leghaemoglobin by isoelectric focusing. Biochim. Biophys. Acta **579:** 314–324
59 Bogusz D., Kortt A. A. and Appleby C. A. (1987) *Sesbania rostrata* root and stem nodele leghaemoglobins: purification and relationship among the seven major components. Arch. Biochem. Biophys. **254:** 263–271
60 Appleby C. A. (1974) Leghaemoglobin. In: The Biology of Nitrogen Fixiation, pp. 521–554, Quispel A. (ed.), North-Holland, Amsterdam
61 Appleby C. A., Tjepkema J. D. and Trinick M. J. (1983) Hemoglobin in a non-leguminous plant, *Parasponia*: possible genetic origin and function in nitrogen fixation. Science **220:** 951–953
62 Wittenberg J. B., Wittenberg B. A., Gibson Q. H., Trinick K. J. and Appleby C. A. (1986) The kinetics of the reactions of *Parasponia andersonii* haemoglobin with oxygen, carbon monoxide, and nitric oxide. J. Biol. Chem. **261:** 13624–13631
63 Fleming A. I., Wittenberg J. B., Wittenberg B. A., Dudman W. F. and Appleby C. A. (1987) The purification, characterization and ligand-binding kinetics of haemoglobins from root nodules of the non-leguminous *Casuarina glauca-Frankia* symbiosis. Biochim. Biophys. Acta **911:** 209–220
64 Bogusz D., Appleby C. A., Landsmann J., Dennis E. S., Trinick M. J. and Peacock W. J. (1988) Functioning haemoglobin genes in non-nodulating plants. Nature **331:** 178–180
65 Cain G. D. (1969) Purification and properties of hemoglobin from *Fasciolopsis buski*. J. Parasitol. **55:** 311–320

66 Tuchsmid P. E., Kunz P. A. and Wilson K. J. (1978) Isolation and characterization of the haemoglobin from the lanceolate fluke *Dicrocoelium dendriticum*. Eur. J. Biochem. **88:** 387–394

67 Haque M., Rashid K. A., Stern M. S., Sharma P. K., Siddiqi A. H., Vinogradov S. N. et al. (1992) Comparison of the haemoglobins of the platyhelminths *Gastrothylax crumenifer* and *Paramphistomum epiclitum* (Trematoda: Paramphistomatidae). Comp. Biochem. Physiol. **101B:** 673–676

68 Rashid K. A., Haque M., Siddiqi A. H., Stern M. S., Sharma P. K., Vinogradov S. N. et al. (1993) Purification and properties of the haemoglobins of the platyhelminth *Isoparorchis hypselobagri* (Trematoda: Isoparorchidae) and its host *Wallagu attu* (catfish). Comp. Biochem. Physiol. **106B:** 993–998

69 Smit J. D., Sick H., Peterhans A. and Gersonde K. (1986) Acid Bohr effect of a monomeric haemoglobin from *Dicrocoelium dendriticum*. Mechanism of the allosteric conformation transition. Eur. J. Biochem. **155:** 231–237

70 Moore J. and Gibson J. (1985) The evolution and comparative physiology of terrestrial and freshwater nemerteans. Biol. Rev. **60:** 257–312

71 Vandergon T. L., Riggs C. L. and Riggs A. F. (1992) Hemoglobin of neural tissue from *Cerebratulus lacteus*: characterization of protein and gene. International Congress on Invertebrate Dioxygen Carriers, 12–17 April, Lunteren, The Netherlands

72 Vandergon T. L., Riggs C. L., Gorr T. A., Colacino J. M. and Riggs A. F. (1998) The mini-hemoglobins in neural and body wall tissue of the nemertean worm, *Cerebratulus lacteus*. J. Biol. Chem. **273:** 16998–17011

73 Okazaki T., Wittenberg B. A., Briehl R. W. and Wittenberg J. B. (1967) The haemoglobin of *Ascaris* body walls. Biochim. Biophys. Acta **140:** 258–265

74 Okazaki T. and Wittenberg J. B. (1965) The haemoglobin of *Ascaris* perienteric fluid. III. Equilibria with oxygen and carbon monoxide. Biochim. Biophys. Acta **111:** 503–511

75 Blaxter M. L., Vanfleteren J. R., Xia J. and Moens L. (1994) Structural characterization of an *Ascaris* myoglobin. J. Biol. Chem. **269:** 30181–30186

76 Blaxter M. L., Ingram L. and Tweedie S. (1994) Sequence, expression and evolution of the globins of the parasitic nematode *Nippostrongylus brasiliensis*. Mol. Biochem. Parasitol. **68:** 1–14

77 Manwell C. (1960) Histological specificity of respiratory pigments. I. Comparisons of the coelom and muscle haemoglobins of the polychaete worm *Travisia pupa* and the echiuroid worm *Arhynchite pugettensis*. Comp. Biochem. Physiol. **1:** 267–276

78 Hall R. E., Terwilliger R. C. and Terwilliger R. N. (1981) Hemoglobins and myoglobin of the echiuran *Urechis caupo* (Fisher and MacGinitie). Comp. Biochem. Physiol. **70B:** 353–357

79 Weber R. E. and Pauptit E. (1972) Molecular and functional heterogeneity in myoglobin from the polychaete *Arenicola marina* L. Archs. Biochem. Biophys. **148:** 322–324

80 Terwilliger R. C., Garlick R. L. and Terwilliger N. B. (1976) Hemoglobins of *Glycera robusta*: structures of coelomic cell haemoglobin and body wall myoglobin. Comp. Biochem. Physiol. **54B:** 149–153

81 Garlick R. L. and Terwilliger R. C. (1977) Structure and oxygen equilibrium of haemoglobin and myoglobin from the pacific lugworm *Abarenicola pacifica*. Comp. Biochem. Physiol. **57B:** 177–184

82 Terwilliger R. C., Garlick R. L. and Terwilliger N. B. (1980) Characterization of the haemoglobins and myoglobin of *Travisia foetida*. Comp. Biochem. Physiol. **66B:** 261–266

83 Kreutzer U. and Jue T. (1996) ^{1}H-NMR signal of *Arenicola marina* myoglobin in vivo as an index of tissue oxygenation. Eur. J. Biochem. **235:** 622–628

84 Wittenberg B. A., Briehl R. W. and Wittenberg J. B. (1965) Haemoglobins of invertebrate tissues. Nerve haemoglobins of *Aphrodite*, *Aplysia* and *Halosydna*. Biochem. J. **96:** 363–371

85 Barnes R. D. (1980) Invertebrate Zoology, 4th ed., Saunders, Philadelphia

86 Read K. R. H. (1966) Molluscan haemoglobin and myoglobin. In: Physiology of Mollusca, Vol. 2, pp. 209–232, Wilbur K. M. and Yonge C. M. (eds), Academic Press, New York

87 Terwilliger R. C and Read K. R. H. (1970) The radular muscle myoglobins of the amphineuran molluscs, *Katharina tunicata* Wood, *Cryptochiton stelleri* Middendorf and *Mopalia muscosa* Gould. Int. J. Biochem. **1:** 281–291

88 Smith S. E., Brittain T. and Wells R. M. G. (1988) A kinetic and equilibrium study of ligand binding to the monomeric and dimeric haem-containing globins of two chitons. Biochem. J. **252:** 673–678

89 Brittain T. (1990) Co-operative functioning of the dimeric haemoglobin obtained from the radular muscle of the amphineurian mollusc *Amaurochiton glaucus*. Comp. Biochem. Physiol. **96B:** 291–295

90 Suzuki T., Furukohri T. and Okamoto S. (1993) Amino acid sequence of myoglobin from the chiton *Liolophura japonica* and a phylogenetic tree for molluscan globins. J. Protein Chem. **12:** 45–50

91 Terwilliger R. C. and Terwilliger N. B. (1985) Molluscan hemoglobins. Comp. Biochem. Physiol. **81B:** 255–261

92 Koppenheffer T. L. and Read K. R. H. (1968) Radular muscle myoglobin of the gastropod mollusc *Nerita peloronta* L. Comp. Biochem. Physiol. **26:** 753–756

93 Read K. R. H. (1968) The myoglobins of the gastropod molluscs *Busycon contrarium* Conrad, *Lunatia heros* Say, *Littorina littorea* L. and *Siphonaria gigas* Sowerby. Comp. Biochem. Physiol. **25:** 81–94

94 Koppenheffer T. L., Terwilliger R. C. and Read K. R. H. (1971) Myoglobin of the gastropod mollusc *Lunatia heros* Say. Int. J. Biochem. **2:** 111–116

95 Terwilliger R. C. and Read K. R. H. (1970) The radular muscle myoglobins of the gastropod molluscs, *Acmaea tetudinalis scutum* Eschscholtz, *Haliotis kamtschatkana* Jones, *Tegula funebralis* Adams, *Fusitriton oregonensis* Redfield, and *Thais lamellosa* Gmelin. Int. J. Biochem. **1:** 445–450

96 Suzuki T., Takagi T. and Shikama K. (1981) Amino acid sequence of myoglobin from *Aplysia kurodai*. Biochim. Biophys. Acta **669:** 79–83

97 Takagi T., Iida S., Matsuoka A. and Shikama K. (1983) *Aplysia* myoglobins with an unusual amino acid sequence. J. Mol. Biol. **180:** 1179–1184

98 Takagi T., Tobita M. and Shikama K. (1983) Amino acid sequence of dimer myoglobin from *Cerithidea rhizophorarum*. Biochim. Biophys. Acta **745:** 32–36

99 Suzuki T. (1986) Amino acid sequence of myoglobin from the mollusc *Dolabella auricularia*. J. Biol. Chem. **261:** 3692–3699

100 Geraci G., Sada A. and Cirotto C. (1977) Cooperative, low-molecular-weight dimeric myoglobins from the radular muscle of the gastropod mollusc *Nassa mutabilis* L. Eur. J. Biochem. **77:** 555–560

101 Suzuki T. and Furukohri T. (1990) Amino acid sequence of myoglobin from the mollusc *Bursatella leachii*. J. Protein Chem. **9:** 69–73

102 Terwilliger R. C. and Read K. R. H. (1971) Oxygen equilibrium studies of the radular muscle myoglobins of the gastropod molluscs, *Buccinum undatum L.* and *Busycon canaliculatum* L. Int. J. Biochem. **2:** 253–261

103 Matsuoka A., Ohie Y., Imai K. and Shikama K. (1996) The dimer-monomer conversion of *Cerithidea* myoglobin coupled with the heme iron oxidation. Comp. Biochem. Physiol. **115B:** 483–492

104 Wittenberg J. B., Appleby C. A. and Wittenberg B. A. (1972) The kinetics of the reactions of leghaemoglobin with oxygen and carbon monoxide. J. Biol. Chem. **247:** 527–531

105 Shikama K. and Katagiri T. (1984) *Aplysia* oxymyoglobin with an unusual stability property. J. Mol. Biol. **174:** 697–704

106 Suzuki T. and Shikama K. (1983) Stability properties of sperm whale oxymyoglobin. Arch. Biochem. Biophys. **224:** 695–699

107 Shikama K. and Matsuoka A. (1986) *Aplysia* oxymyoglobin with an unusual stability property: kinetic analysis of the pH dependence. Biochemistry **25:** 3898–3903

108 Suzuki T. and Takagi T. (1992) A myoglobin evolved from indoleamine-2,3-dioxygenase. J. Mol. Biol. **228:** 698–700

109 Hirata F., Ohnishi T. and Hayaishi O. (1977) Indoleamine 2,3-dioxygenase. Characterization and properties of enzyme. O_2-complex. J. Biol. Chem. **252:** 4637–4642

110 Suzuki T. (1994) Abalone myoglobins evolved from indoleamine dioxygenase. The cDNA-derived amino acid sequence of myoglobin from *Nordotis madaka*. J. Protein Chem. **13:** 9–13

111 Kawamichi H. and Suzuki T. (1998) The cDNA-derived amino acid sequence of indoleamine dioxygenase-like myoglobin from the gastropod mollusc *Omphalius pfeifferi*. J. Protein Chem., in press

112 Koppenheffer T. L. and Read K. R. H. (1969) The adductor muscle myoglobins of the bivalve mollusc *Mercenaria mercenaria* L. Comp. Biochem. Physiol. **31:** 829–832

113 Hardman K. D., Eylar E. H., Ray D. K., Banaszak L. J. and Gurd F. R. N. (1966) Isolation of sperm whale myoglobin by low temperature fractionation with ethanol and metallic ions. J. Biol. Chem. **241:** 432–442

114 Kendrew J. C., Bodo G., Dintzis H. M., Parrish R. G., Wyckoff H. and Philips D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. Nature **181:** 662–666

115 Edmundson A. B. (1968) Amino-acid sequence of sperm whale myoglobin. Nature **205:** 883–887

116 Sidell B. D., Vayda M. E., Small D. J., Moylan T. J., Londraville R. L., Yuan M. L. et al. (1997) Variable expression of myoglobin among the haemoglobinless Antarctic icefishes. Proc. Natl. Acad. Sci. USA **94:** 3420–3424

117 Phillips S. E. V. and Schoenborn B. P. (1981) Neutron diffraction reveals oxygen-histidine hydrogen bond in oxymyoglobin. Nature **292:** 81–82

118 Shikama K., Matsuoka A. and Iwaasa H. (1995) The unique structures of protozoan myoglobin and yeast haemoglobin: an evolutionary diversity. Int. J. Biochem. Cell. Biol. **27:** 1107–1115

119 Takagi T., Iwaasa H., Yuasa H., Shikama K., Takemasa T. and Watanabe Y. (1993) Primary structure of *Tetrahymena* hemoglobin. Biochim. Biophys. Acta **1173:** 75–78

120 Yamauchi K., Tada H. and Usuki I. (1995) Structure and evolution of *Paramecium* hemoglobin genes. Biochim. Biophys. Acta **1264:** 53–62

121 Wakabayashi S., Matsubara H. and Webster D. A. (1986) Primary sequence of a dimeric bacterial haemoglobin from Vitreoscilla. Nature **322:** 481–483.

122 Iwaasa H., Takagi T. and Shikama K. (1992) Amino acid sequence of yeast hemoglobin: a two-domain structure. J. Mol. Biol. **227:** 948–954

123 Favey S., Labesse G., Vouille V. and Boccara M. (1995) The flavohemoprotein HmpX: a new determinant of *Erwinia chrysanthemi* strain 3937. Microbiology **1411:** 863–871

124 LaCelle M., Kumano M., Kurita K., Yamane K., Zuber P. and Nakano M. M. (1996) Oxygen-controlled regulation of the flavohemoglobin gene in *Bacillus subtilis*. J. Bacteriol. **178:** 3803–3808

125 Cramm R., Siddiqui R. A. and Friedrich B. (1994) Primary sequence and evidence for a physiological function of the flavohemoprotein of *Alcaligenes eutrophus*. J. Biol. Chem. **269:** 7349–7354

126 Hyldig-Nielsen J. J., Jenssen E. O., Paludan K., Wiborg O., Garret R., Jorgensen P. et al. (1982) The primary structure of two leghaemoglobin genes from soybean. Nucleic Acids Res. **10:** 689–701

127 Kortt A. A., Strike P. M., Bogusz D. and Appleby C. A. (1987) Amino acid sequence of haemoglobin II from *Sesbania rostrata* stem nodules. Biochem. Int. **15:** 509–516

128 Kortt A. A., Inglis A. S., Fleming A. I. and Appleby C. A. (1988) Amino acid sequence of hemoglobin I from root nodules of the non-leguminous *Casuarina glauca-Frankia* symbiosis. FEBS Lett. **231:** 341–346

129 Kortt A. A., Trinick M. J. and Appleby C. A. (1988) Amino acid sequence of haemoglobins I and II from root nodules of the non-leguminous *Parasponia rigida-Rhizobium* symbiosis and a correction of the sequence of Hb I for *P. andersonii*. Eur. J. Biochem. **175:** 141–149

130 Rashid K. A., Van Hauwaert M.-L., Haque M., Siddiqi A. H., Lasters I., De Maeyer M. et al. (1997) Trematode myoglobins, functional molecule with a distal tyrosine. J. Biol. Chem. **272:** 2992–2999.

131 Hayashi A., Suzuki T., Shimizu A. and Yamamura Y. (1968) Properties of haemoglobin M. Unequivalent nature of the a and b subunits in the haemoglobin molecule. Biochim. Biophys. Acta **168:** 262–273

132 Dickerson R. E. and Geis E. (1983) Hemoglobin: Structure, Function, Evolution and Pathology, Benjamin Cummings, Menlo Park, CA

133 Kloek A. P., Sherman D. R. and Goldberg D. E. (1993) Novel gene structure and evolutionary context of *Caenorhabditis elegans* globin. Gene **129:** 215–221

134 Dewilde S., Blaxter M., Van Hauwaert M.-L., Vanfleteren J., Esmans E. L., Marden M. et al. (1996) Globin and globin gene structure of the nerve myoglobin of *Aphrodite aculeata*. J. Biol. Chem. **271:** 19865–19870

135 Tentori L., Vivaldi G., Carta S., Marinucci M., Massa A., Antonini E. et al. (1973) The amino acid sequence of myoglobin from the mollusc *Aplysia limacina*. Int. J. Peptide Protein Res. **5:** 182–200

136 Cutruzzola F., Allocatelli C. T., Brancaccio A. and Brunori M. (1996) *Aplysia limacina* myoglobin cDNA cloning: an alternative mechanism of oxygen stabilization as studied by active-site mutagenesis. Biochem. J. **314:** 83–90

137 Yamaguchi M. and Takagi T. (1995) Search for IDO-like myoglobins in molluscs. [Abstract in Japanese]. Seikagaku **67:** 855

138 Bonner A. S. and Laursen R. A. (1977) The amino acid sequence of a dimeric myoglobin from the gastropod mollusc *Busycon canaliculatum*. FEBS Lett. **73:** 201–203

139 Parente A., Verde C., Malorni A., Montecucchi P. C., Aniello F. and Geraci G. (1993) The amino acid sequence of the dimeric myoglobin from the clam *Nassa mutabilis*. Biochim. Biophys. Acta **1162:** 1–9

140 Dai W. and Gupta S. L. (1990) Molecular cloning, sequencing and expression of human interferon-γ-inducible indoleamine 2,3-dioxygenase cDNA. Biochem. Biophys. Res. Commun. **168:** 1–8

141 Habara-Ohkubo A., Takikawa O. and Yoshida R. (1991) Cloning and expression of a cDNA encoding mouse indoleamine 2,3-dioxygenase. Gene **105:** 221–227

142 Dene H., Sazy J., Goodman M. and Romero-Herrera A. E. (1980) The amino acid sequence of alligator (*Alligator mississippiensis*) myoglobin. Phylogenetic implications. Biochim. Biophys. Acta **624:** 397–408

143 Peiffer S., Deconinck M., Paul C., Depreter J., Schnek A. G. and Leonis J. (1973) Penguin (*Aptenodytes forsteri*) myoglobin. A 70 residue N-terminal sequence. FEBS Lett. **37:** 295–297

144 Watts D. A., Rice R. H. and Brown W. D. (1980) The primary structure of myoglobin from yellowfin tuna (*Thunnus albacares*). J. Biol. Chem. **255:** 10916–10924

145 Maeda N. and Fitch W. M. (1982) Isolation and amino acid sequence of a monomeric haemoglobin in heart muscle of the bullfrog, *Rana catesbeiana*. J. Biol. Chem. **257:** 2806–2815

146 Liebhaber S. A., Goossen M. L. and Kan Y. W. (1980) Cloning and complete nucleotide sequence of human 5′-α-globin gene. Proc. Natl. Acad. Sci. USA **77:** 7054–7058

147 Lawn R. M., Efstratiadis A., O'Connell C. and Maniatis T. (1980) The nucleotide sequence of human β-globin gene. Cell **21:** 647–651

148 Go M. (1981) Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature **291:** 90–92

149 Blake C. C. F. (1981) Exons and the structure, function and evolution of haemoglobin. Nature **291:** 616

150 Dixon B. and Pohajdak B. (1992) Did the ancestral globin gene of plants and animals contain only two introns? Trends Biochem. Sci. **17:** 486–488

151 Moens L., Vanfleteren J., De Baere I., Jellie A. M., Tate W. and Trotman C. N. A. (1993) Commentary Response. FEBS Lett. **320:** 284–287

152 Yamauchi K., Ochiai T. and Usuki I. (1992) The unique structure of the *Paramecium caudatum* hemoglobin gene: the presence of one intron in the middle of the coding region. Biochim. Biophys. Acta **1171:** 81–87

153 Dikshit K. L. and Webster D. A. (1988) Cloning, characterization and expression of the bacterial globin gene from *Vitreoscilla* in *E. coli*. Gene **70:** 377–386

154 Christensen T., Dennis E. S., Peacock J. W., Landsmann J. and Marcker K. A. (1991) Hemoglobin genes in non-legumes: cloning and characterization of a *Casuarina glauca* haemoglobin gene. Plant. Mol. Biol. **16:** 339–344

155 Hardison R. C. (1996) A brief history of haemoglobins: plant, animal, protist and bacteria. Proc. Natl. Acad. Sci. USA **93:** 5675–5679

156 Jhiang S. M. and Riggs A. F. (1989) The structure of the gene encoding chain c of the haemoglobin of the earthworm, *Lumbricus terrestris*. J. Biol. Chem. **264:** 19003–19008

157 Dixon B., Walker B., Kimmins W. and Pohajdak B. (1992) A nemotode hemoglobin gene contains an intron previously thought to be unique to plants. J. Mol. Evol. **35:** 131–136

158 Breathnach R. and Chambon P. A. (1981) Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50:** 349–383

159 Jellie A. M., Tate W. P. and Trotman C. N. A. (1996) Evolutionary history of introns in a multidomain globin gene. J. Mol. Evol. **42:** 641–647

160 Kao W. Y., Trewitt P. M. and Bergtrom G. (1994) Intron-containing globin genes in the insect *Chironomus thummi*. J. Mol. Evol. **38:** 241–249

161 Grinich N. P. and Terwilliger R. C. (1980) The quaternary structure of an unusual high molecular weight intracellular haemoglobin from the bivalve mollusc *Barbatia reeveana*. Biochem. J. **189:** 1–8

162 Titchen D. A., Glenn W. K., Nassif N., Thompson A. R. and Thompson E. O. P. (1991) A minor globin gene of the bivalve mollusc *Anadara trapezia*. Biochim. Biophys. Acta **1089:** 61–67

163 Piro M. C., Gambacurta A. and Ascoli F. (1996) *Scapharca inaequivalvis* tetrameric haemoglobin A and B chains: cDNA sequencing and genomic organization. J. Mol. Evol. **43:** 594–601

164 Suzuki T., Kawasaki Y., Arita T. and Nakamura A. (1996) Two-domain haemoglobin of the blood clam *Barbatia lima* resulted from the recent gene duplication of the single-domain δ chain. Biochem. J. **313:** 561–566

165 Suzuki T., Takagi T. and Ohta S. (1989) Primary structure of a dimeric hemoglobin from the deep-sea cold seep clam *Calyptogena soyoae*. Biochem. J. **260:** 177–182

166 Blanchetot A., Wilson V., Wood D. and Jeffreys A. J. (1983) The seal myoglobin gene: an unusually long globin gene. Nature **301:** 732–734

167 Akaboshi E. (1985) Cloning of the human myoglobin gene. Gene **33:** 241–249

168 Blanchetot A., Price M. and Jeffreys A. J. (1986) The mouse myoglobin gene. Characterization and sequence comparison with other mammalian myoglobin genes. Eur. J. Biochem. **159:** 469–474

169 Suzuki T., Yuasa H. and Imai K. (1996) Convergent evolution. The gene structure of *Sulculus* 41 kDa myoglobin is homologous with that of human indoleamine dioxygenase. Biochim. Biophys. Acta **1308:** 41–48

170 Kadoya A., Tone S., Maeda H., Minatogawa Y. and Kido R. (1992) Gene structure of human indoleamine 2,3-dioxygenase. Biochem. Biophys. Res. Commun. **189:** 530–536

171 Fukami-Kobayashi K., Mizutani M. and Go M. (1995) Correlation between module boundaries and intron positions in haemoglobins from various taxa. In: Tracing Biological Evolution in Protein and Gene Structures, pp. 271–282, Go M. and Schimmel P. (eds), Elsevier Science, The Netherlands

172 Kapp O. H., Moens L., Vanfleteren J., Trotman C. N. A., Suzuki T. and Vinogradov S. N. (1995) Alignment of 700 globin sequences: extent of amino acid substitution and its correlation with variation in volume. Protein Sci. **4:** 2179–2190

173 Takagi T. (1993) Hemoglobins from single-celled organisms. Curr. Opin. Struct. Biol. **3:** 413–418

174 Moens L., Vanfleteren J., Van der Peer Y., Peeters K., Kapp O., Czeluzniak J. et al. (1996) Globins in nonvertebrate species: dispersal by horizontal gene transfer and evolution of the structure-function relationships. Mol. Biol. Evol. **13:** 324–333

175 Bashford D., Chothia C. and Lesk A. A. (1987) Determinants of a protein fold: unique features of the globin amino acid sequences. J. Mol. Biol. **196:** 199–216

176 Suzuki T. (1987) Autoxidation of oxymyoglobin with the distal(E7) glutamine. Biochim. Biophys. Acta **914:** 170–176

177 Tsubamoto Y., Matsuoka A., Yusa K. and Shikama K. (1990) Protozoan myoglobin from *Paramecium caudatum*. Its autoxidation reaction and hemichrome formation. Eur. J. Biochem. **193:** 55–59

178 Felsenstein J. (1993) PHYLIP (Phylogeny Inference Package), version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, WA, USA

179 Doolittle R. F., Feng D. F., Johnson M. S. and McClure M. A. (1986) Relationships of human protein sequences to those of other organisms. Cold Spring Harbor Symp. Quant. Biol. **51:** 447–455

180 Runnegar B. (1984) Derivation of the globins from type *b* cytochromes. J. Mol. Evol. **21:** 33–41

181 Springer B. A., Sligar S. G., Olson J. S. and Phillips G. N. (1994) Mechanisms of ligand recognition in myoglobin. Chem. Rev. **94:** 699–714

182 Cloud P. and Glassner M. F. (1982) The Ediacaran period and system: Metazoa inherit the earth. Science **217:** 783–788

183 Goodman M., Pedwaydon J., Czelusniak J., Suzuki T., Gotoh T., Moens L. et al. (1988) An evolutionary tree for invertebrate globin sequences. J. Mol. Evol. **27:** 236–249

184 Goodman M., Moore G. W. and Matsuda G. (1975) Darwinian evolution in the genealogy of hemoglobin. Nature **253:** 604–608

185 Fisher W. H., Koureas D. D. and Thompson E. O. P. (1980) Myoglobins of cartilaginous fishes. II. Isolation and amino acid sequence of myoglobin of the shark *Mustelus antercticus*. Aust. J. Biol. Sci. **33:** 153–167

186 Fisher W. H., Koureas D. D. and Thompson E. O. P. (1981) Myoglobins of cartilaginous fishes. III. Amino acid sequence of myoglobin of the shark *Galeorhinus australis*. Aust. J. Biol. Sci. **34:** 5–10

187 Suzuki T., Suzuki T.-N. and Yata T. (1985) Shark myoglobin. II. Isolation, characterization and amino acid sequence of myoglobin from *Galeorhinus japonicus*. Aust. J. Biol. Sci. **38:** 347–354

188 Dene H., Goodman M. and Romero-Herrera A. E. (1980) The amino acid sequence of elephant (*Elephas maximus*) myoglobin and the phylogeny of Proboscidea. Proc. R. Soc. Lond. **B207:** 111–127

189 Romero-Herrera A. E., Goodman M., Dene H., Bartnicki D. E. and Mizukami H. (1981) An exceptional amino acid replacement on the distal side of the iron atom in proboscidean myoglobin. J. Mol. Evol. **17:** 140–147

190 Liljeqvist G., Braunitzer G. and Paleus S. (1979) The amino acid sequence of hemoglobin III from Myxine glutinosa L. A new heme complex: E7 glutamine, E11 isoleucine. Hoppe-Seyler's Z. Physiol. Chem. **360:** 125–135

191 Stenzel P., Brimhall B., Jones R. T., Black J. A., McLachlan A. and Gibson D. (1979) Opossum hemoglobin. J. Biol. Chem. **254:** 2071–2076

192 Nagai K., Luisi B., Shih D., Miyazaki G., Imai K., Poyart C. et al. (1987) Distal residues in the oxygen binding site of haemoglobin studied by protein engineering. Nature **329:** 858–860

193 Zhao X., Vyas K., Nguyen B. D., Rajarathnam K., La Mar G. N., Li T. et al. (1995) A double mutant of sperm whale myoglobin mimics the structure and function of elephant myoglobin. J. Biol. Chem. **270:** 20763–20774

194 Suzuki T., Ban T. and Furukohri T. (1997) Evolution of Phosphagen Kinase V. cDNA-derived amino acid sequences of two molluscan arginine kinases from the chiton *Liolophura japonica* and the turbanshell *Battilus cornutus*. Biochim. Biophys. Acta **1340:** 1–6
195 Tyree B. and Webster D. A. (1978) The binding of cyanide and carbon monoxide to cytochrome *o* purified from *Vitreoscilla*. J. Biol. Chem. **253:** 6988–6991
196 Oshino R., Oshino N. and Chance B. (1971) The oxygen equilibrium of yeast hemoglobin. FEBS Lett. **19:** 96–100
197 Appleby C. A. (1962) The oxygen equilibrium of leghemoglobin. Biochim. Biophys. Acta **60:** 226–235
198 Manwell C. (1960) Heme-heme interactions in the oxygen equilibrium of some invertebrate myoglobins. Arch. Biochem. Biophys. **89:** 194–201
199 Suzuki T. and Imai K. (1997) Comparative studies of the indoleamine dioxygenase-like myoglobin from the abalone *Sulculus diversicolor*. Comp. Biochem. Physiol. **117B:** 599–604
200 Yamamoto Y., Iwafune K., Chujo R., Inoue Y., Imai K. and Suzuki T. (1992) ^{1}H-NMR comparative study of the active site in deoxy shark (*Galeorhinus japonicus*), horse and sperm whale myoglobins. J. Biochem. **112:** 414–420
201 Rossi-Fanelli A., Antonini E. and Giuffre R. (1960) Oxygen equilibrium of myoglobin from *Thunnus thynnus*. Nature **186:** 896–897
202 Enoki Y., Matsumura K., Ohga Y., Kohzuki H. and Hattori M. (1995) Oxygen affinities (P_{50}) of myoglobins from four vertebrate species (*Canis familiaris*, *Rattus norvegicus*, *Mus musculus* and *Gallus domesticus*) as determined by a kinetic and an equilibrium method. Comp. Biochem. Physiol. **110B:** 193–199
203 Rossi-Fanelli A. and Antonini E. (1958) Studies on the oxygen and carbon monoxide equilibria of human myoglobin. Arch. Biochem. Biophys. **77:** 478–492
204 Feng D. F. and Doolittle R. F. (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic tree. J. Mol. Evol. **25:** 351–360
205 Sherman D. R., Kloek A. P., Krishnan B. R., Guinn B. and Goldberg D. E. (1992) *Ascaris* haemoglobin gene: plant-like structure reflects the ancestral globin gene. Proc. Natl. Acad. Sci. USA **89:** 11696–11700
206 Yamamoto Y., Iwafune K., Chujo R., Inoue Y., Imai K. and Suzuki T. (1992) Molecular mechanism for ligand stabilization in the mollusc myoglobin possessing the distal Val residue. J. Mol. Biol. **228:** 343–346
207 Imai K. and Yoshikawa S. (1985) Oxygen-binding characteristics of *Potamilla* chlorocruorin. Eur. J. Biochem. **147:** 453–463
208 Suzuki T., Kawamichi H. and Imai K. (1998) Amino acid sequence, spectral, oxygen binding and autoxidation properties of indoleamine dioxygenase like-myoglobin from the gastropod mollusc *Turbo cornutus*. J. Protein. Chem. in press