Functional Diversification of Multiple Lungfish Globins

Dissertation

with the aim of achieving a
doctoral degree (Dr. rer. nat.)

at the

Faculty of Mathematics,
Informatics and Natural Science

Department of Biology

University of Hamburg

submitted by

Julia Katrin Lüdemann

born in Hamburg

Hamburg, June 2019

Day of oral defense:

27. September 2019

The following reviewers recommended the admission of the dissertation:

Prof. Dr. Susanne Dobler

Prof. Dr. Angela Fago

Everything that human beings or living animals do is done by protein molecules.
And therefore the kind of proteins that one has
and therefore the ability one has
is determined by the genes that one has.

Har Gobind Khorana (1922 – 2011)
Molecular Biologist

Table of Contents

Zusammenfassung 8
Abstract 10
I. Introduction 12
1.1 Hemoglobin *15*
1.2 Myoglobin *15*
1.3 Neuroglobin *17*
1.4 Cytoglobin *19*
1.5 Globin E *20*
1.6 Globin Y *20*
1.7 Globin X *21*
1.8 Androglobin *21*
1.9 Evolution of globins *22*
1.10 Lungfish *25*
1.10.1 Aestivation 27
1.11 Multiple globins in lungfish *28*
1.12 Aim of the thesis *30*
II. Chapter 1: 31
Unusual diversity of myoglobin genes in the lungfish 31
2.1 Abstract *32*
2.2 Introduction *33*
2.3 Results *35*
2.3.1 An unexpected diversity of myoglobins in *P. annectens* 35
2.3.2 Early emergence and fast evolution of lungfish myoglobins 35
2.3.3 Differential, tissue-specific expression of lungfish myoglobins 36
2.3.4 Lungfish myoglobins protect cultured cells from hypoxia and reduce ROS production 38
2.4 Discussion *39*
2.4.1 Ancient and specific amplification of myoglobin genes 39
2.4.2 High myoglobin concentrations may protect the lungfish brain from hypoxia during aestivation 40
2.4.3 Different myoglobins with distinct roles in skeletal muscle and heart 41

2.4.4 Conclusion: Myoglobin diversity in lungfish as an example for the functional diversification of globin genes 43
2.5 Materials and Methods *43*
2.5.1 Transcriptome and database analyses 43
2.5.2 Phylogenetic analyses 44
2.5.3 Tissue preparation, RNA extraction cDNA cloning 45
2.5.4 Quantitative real-time reverse transcription PCR 45
2.5.5 Analyses of myoglobin proteins 46
2.5.6 Cell culture experiments 47
2.6 Acknowledgments *47*
III. Chapter 2: 49
Genetic and functional diversity of the multiple lungfish myoglobins 49
3.1 Abstract *50*
3.2 Introduction *51*
3.3 Material and Methods *53*
3.3.1 Identification and analyses of lungfish Mb cDNA sequences 53
3.3.2 Recombinant expression and purification of P. annectens Mbs 55
3.3.3 O_2 equilibrium measurements 55
3.3.4 Gel-filtration experiments 56
3.3.5 Nitrite reductase activity 56
3.3.6 Dynamic light scattering (DLS) 57
3.3.7 Circular dichroism (CD) spectroscopy 57
3.3.8 Small-angle X-ray scattering (SAXS) 57
3.4 Results *58*
3.4.1 Identification of myoglobin cDNA in four lungfish species 58
3.4.2 O_2 equilibria of *P. annectens* myoglobins 59
3.4.3 Enzymatic activities of P. annectens myoglobins 60
3.4.4 Structural analyses of *P. annectens* myoglobins 61
3.5 Discussion *63*
3.5.1 Emergence of multiple myoglobins as a possible adaptation of lungfish 63
3.5.2 Conserved structure of *P. annectens* myoglobins 65
3.5.3 Functional differentiation of *P. annectens* myoglobins 66
3.5.4 Functional differentiation of lungfish myoglobin in evolution 66

3.6 Acknowledgements *67*
IV. Chapter 3: 68
Globin E is a myoglobin-related, respiratory protein highly expressed in lungfish oocytes 68
4.1 Abstract *69*
4.2 Introduction *69*
4.3 Results *71*
4.3.1 Identification of GbE genes in lungfish species 71
4.3.2 Conservation and lungfish-specific amplification of GbE genes 73
4.3.3 Expression of GbE in the lungfish ovary 74
4.3.4 Localisation of GbE mRNA in previtellogenic oocytes 77
4.3.5 Spectroscopic studies and O2 binding equilibria to GbE1 of *L. paradoxa* 78
4.4 Discussion *79*
4.5 Methods *82*
4.5.1 Lungfish material 82
4.5.2 RNA extraction, library preparation and Illumina sequencing 82
4.5.3 Sequence analyses and phylogenetic inference 83
4.5.4 Quantitative real-time reverse-transcription PCR 84
4.5.5 *In situ* hybridisation. 84
4.5.6 Preparation of recombinant lungfish GbE protein 85
4.5.7 Spectroscopic studies and O_2 binding curves 86
4.5.8 Nitrite reductase activity 87
V. Discussion 88
5.1 Divergence of lungfish genera *88*
5.1.1 Air-breathing is a successful adaptation of fish 90
5.1.2 Aestivation is a special feature of lungfish 91
5.2 Multiple Mb and GbE genes were detected in lungfish transcriptomes *92*
5.3 Mb and GbE form a sister-group *95*
5.4 Mb-Gene expansion protects lungfish brain from hypoxia *96*
5.4.1 Mb probably derived from sub- and/or neofunctionalization 99
5.5 GbE-copies may protect the lungfish eggs *100*
5.5.1 GbE has a possible additive effect 101
5.6 Multiple globin genes are a proof of adaptation *102*
5.6.1 An attempt of explanation: The unexpected globin repertoire in lungfish 105

5.7 Outlook *107*
VI. References 108
VII. Danksagung 120
VIII. Eidesstattliche Versicherung (Declaration on oath) 122
IX. Declaration of own contribution to the manuscripts 123

Zusammenfassung

Globine sind kleine, respiratorische Proteine, die eine wichtige Rolle bei der Sauerstoffversorgung von lebenden Organismen spielen. Zusätzlich können sie auch noch vielfältige weitere Aufgaben übernehmen, wie zum Beispiel als Nitritreduktase fungieren oder die Zellen vor reaktiven Sauerstoffspezies (ROS) schützen. In Vertebraten besteht die Proteinfamilie aus 8 Globintypen: Hämoglobin, Myoglobin, Neuroglobin, Cytoglobin, Globin X, Globin Y, Globin E und Androglobin. Diese 8 verschiedenen Globine zeigen innerhalb der Vertebraten ein divergentes Verteilungsmuster. Einige Globine, wie zum Beispiel Hämoglobin, Neuroglobin oder Cytoglobin, sind in beinahe allen Wirbeltiervertretern vorhanden. Andere Globine, wie zum Beispiel Myoglobin, fehlen in einigen Linien. Im Laufe der Evolution zeigen sich auch multiple Globinkopien in einigen Arten, die weiterer funktioneller Aufklärung bedürfen.

In der vorliegenden Arbeit konnten zum ersten Mal sowohl multiple Myoglobin (Mb) als auch Globin E (GbE) Kopien im Lungenfisch nachgewiesen und diese Diversität näher untersucht werden. Lungenfische sind die nächstlebenden Verwandten von Vertebraten und dabei obligate Luftatmer. Diese beiden Eigenschaften sorgen dafür, dass der Lungenfisch ein ideales Modell darstellt, um evolutionäre Schlüsselmomente zu benennen, die zur Transition des Lebens vom Wasser an Land geführt haben.

In den Afrikanischen Lungenfischen *Protopterus annectens*, *Protopterus dolloi* und *Protopterus aethiopicus* sowie dem Südamerikanischen Lungenfisch *Lepidosiren paradoxa* konnten jeweils fünf bis neun Myoglobin (Mb) Kopien gefunden werden, obwohl Vertebraten normalerweise nur ein Mb besitzen. Diese werden differenziell in den verschiedenen Geweben des Lungenfisches exprimiert. Die höchste Mb Konzentration konnte dabei überraschenderweise im Gehirn und nicht, wie üblich, im Muskelgewebe gefunden werden. Das Herz- und Muskelgewebe exprimiert dabei unterschiedliche Mb Typen. Anhand von rekombinant exprimierten Mbs konnte gezeigt werden, dass diese unterschiedliche O_2 Bindungsverhalten und enzymatische Aktivitäten zeigen. Diese Daten weisen stark darauf hin, dass die Lungenfisch Mbs verschiedene Funktionen ausüben und die *Mb* Gene vermutlich mittels Duplikationen entstanden sind. Da alle Lungenfische eine ähnliche Diversität zeigen, traten die Mb-Amplifikationen wahrscheinlich bereits in der Lungenfisch-Stammlinie auf; vermutlich als Antwort auf die hypoxischen Bedingungen in ihrem Lebensraum.

Ursprünglich wurde angenommen, dass Globin E (GbE) exklusiv im Auge von Vögeln und Reptilien exprimiert wird, wo es eine Mb-ähnliche Funktion hat, um die metabolisch aktive Retina mit O_2 zu versorgen. Im Gegensatz zu dieser Annahme konnten im Südamerikanischen Lungenfisch *Lepidosiren paradoxa* und in den afrikanischen Lungenfischen *Protopterus annectens* und *Protopterus aethiopicus* zwischen fünf und sieben *GbE* Gene gefunden werden, die sehr ähnlich zueinander sind. Überraschenderweise ist GbE im Lungenfisch exklusiv im Ovar exprimiert, wo es sehr hohe mRNA Werte aufweist. Diese Funktionen weisen auf 1) einen additiven Effekt der multiplen GbE Kopien und auf 2) eine Rolle in der Bereitstellung von O_2 im sich entwickelnden Embryo, analog oder komplementär zum embryonischen und fetalen Hämoglobin anderer Vertebraten, hin.

Abstract

Globins are small, respiratory proteins playing an important role in O_2 supply of living organisms. Additionally, they can fulfill several different functions like acting as a nitrite reductase or detoxifying reactive oxygen species (ROS). In vertebrates the globin protein-family consists of eight different globin types: hemoglobin, myoglobin, neuroglobin, cytoglobin, globin X, globin Y, globin E, and androglobin. These eight different globins show a divergent distribution pattern throughout the vertebrates. Some of the globins, for example hemoglobin, neuroglobin or cytoglobin, are present in nearly all vertebrates. Other globins, like myoglobin, are missing in some of the lines. During evolution multiple globin copies also appeared in some species that call for further functional studies.

In this thesis, multiple myoglobin (Mb) and globin E (GbE) copies were detected for the first time in lungfish and their diversity was subject to further investigations. Lungfish are the closest living relatives of terrestrial vertebrates and obligate air-breathers. Both features make them an ideal model to investigate evolutionary key events leading to the transition from water to land.

In the West African lungfish (*Protopterus annectens*), the slender lungfish (*Protopterus dolloi*), the marbled lungfish (*Protopterus aethiopicus*), and the South American lungfish (*Lepidosiren paradoxa*) five to nine copies of Mb were found, whereas normally vertebrates only harbor one Mb. These globins are differentially expressed in various tissues of the lungfish. The highest level of Mb mRNA, which is typically expressed in muscle tissue, was found in the lungfish brain. Heart and muscle express different Mb types. The recombinant Mb proteins show distinct O_2 binding properties and enzymatic activities. These data suggest that the lungfish Mbs carry out distinct functions and that the *Mb* genes evolved through gene duplications. Amplification of the *Mb* genes occurred within lungfish stem lineage, probably in response to a lifestyle in hypoxic waters, as all lungfish show a similar diversity.

GbE expression was initially identified almost exclusively in the eye of birds and reptiles, where it most likely plays a Mb-like role in O_2 supply to the metabolically active retina. In contrast to this assumption, in the South American lungfish, the West African lungfish, and the marbled lungfish (*P. aethiopicus*), between five to seven very similar *GbE* genes were found. Surprisingly, in lungfish GbE is mainly expressed in the ovary, where it reaches very high mRNA levels. These

findings suggest 1) an additive role of the multiple *GbE* gene copies, and 2) that GbE plays a role in the O_2 supply in the developing embryo of lungfish, analogous or complementary to the embryonic and fetal hemoglobins of other vertebrates.

I. Introduction

„Dum spiro, spero“ - While I breathe, I hope

(Marcus Tullius Cicero)

Breathing is the essential task of almost all metazoans to survive (Danovaro et al. 2010). It is obligate that in most living organisms oxygen needs to be transported by the vascular system (Dickerson and Geis 1983). As passive oxygen diffusion is not sufficient enough to provide the metabolic processes with oxygen, respiratory proteins help to store and transport the oxygen (O_2). Hemocyanin is the respiratory protein in the haemolymph of arthropods and molluscs (Burmester 2002, Markl 2013) whereas hemerythrin is present in brachiopods, priapulida, and sipunculids (Bailly et al. 2008).

Globins are the most widespread respiratory proteins in vertebrates, invertebrates, prokaryotes, fungi, archaea, protozoan and plants (Hardison 1996a, Weber and Vinogradov 2001, Vinogradov et al. 2013) that share the characteristic “three-over-three” α-helical structure, consisting of (normally) 8 α-helices (A-H), even though their primary structure is highly diverse. They contain a non-covalently bound ferrous heme that is able to reversibly bind oxygen and other gaseous ligands like carbon monoxide (CO) or nitric oxide (NO) (Perutz 1979). The planar heme-group is localized in a hydrophobic protein pocket built proximally and distally by helices F and E, respectively, with G and H helices each at one site. They share a conserved proximal histidine at position F8 (eighth amino acid of helix F) which is covalently bound to the heme group. In most cases the phenylalanine residue at position CD1 (first amino acid after helix C) is also conserved and gives additional stability (Burmester and Hankeln 2014). The Fe^{2+}-ion can form up to six coordination bonds; four of them bind the tetrapyrrole ring, the fifth binds the proximal histidine at position F8 (HisF8). Oxygen can bind to the sixth coordination bond. The globins differ in their coordination (Fig. I.1). If a globin is hexa-coordinated, it means that the distal HisE7 binds directly to the Fe^{2+}-ion when deoxygenated. In case of oxygenation, the oxygen competes with the distal histidine to bind to the iron. The replacement of the HisE7 is the rate-determining step for binding external ligands (Pesce et al. 2002). With this it comes to conformational changes of the protein which is assumed to be instrumental in signal transduction pathways (de Sanctis et al. 2004). In the penta-coordinated globins, the sixth

binding-site remains empty when deoxygenated which makes it directly accessible to the oxygen to bind the distal site of the Fe^{2+} (Fig. I.1).

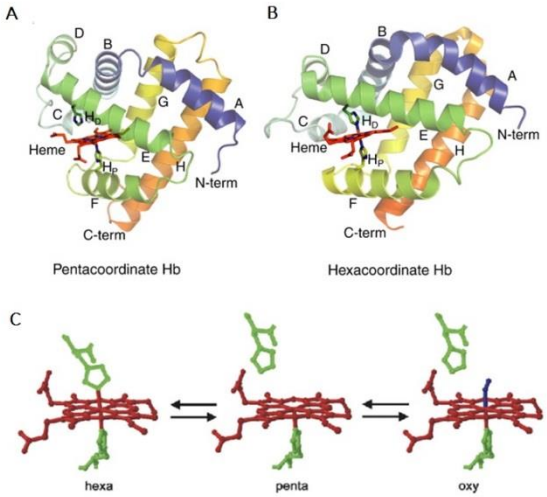


Figure I.1: Three-dimensional structure of penta- and hexa-coordinated globins.

The graphics shows the typical globin 3-over-3 α-helical sandwich structure containing eight α-helices (A-H) with the heme bound. A: The iron of the deoxygenated and penta-coordinated sperm whale Mb is in coordination with the proximal histidine (H_P). B: The iron of the deoxygenated and hexa-coordinated mouse Mb is coordinated by the proximal (H_P) and distal (H_D) histidine. C: A schematic overview of a hexa-coordinated, a penta-coordinated and an oxygenated Fe^{2+}-ions is displayed. Red: heme, green: proximal and distal histidine, blue: oxygen (Modified from (Hankeln et al. 2005, Kakar et al. 2010)).

The globins are expressed in a wide range of tissues and different cell types like muscle and nerve cells or circulating red blood cells associated with a diverse range of quaternary structures (Weber and Vinogradov 2001). This diversity represents a similarly rich diversity of functions including oxygen binding and storage, oxygen sensing, scavenging the reactive oxygen species (ROS) and reactive nitrogen species (RNS), redox signalling, and enzymatic functions (Fig. I.2) (Weber and Vinogradov 2001, Pesce et al. 2002, Fago et al. 2004b, Hankeln et al. 2005, Burmester and Hankeln 2009).

Hemoglobin (Hb) and myoglobin (Mb) are the most prominent members of the globin-superfamily and are probably the best studied proteins in biological science. The globin superfamily is an ideal model to study function and evolution of proteins, genes and gene families (Hardison 1996a, Gillemans et al. 2003, Vinogradov et al. 2007). In the last decades a total of 8 globins were discovered (Fig. I.3). Together with Hb and Mb there are neuroglobin (Ngb) (Burmester et al. 2000), cytoglobin (Cygb) (Kawada et al. 2001, Burmester et al. 2002, Trent and Hargrove 2002), globin X (GbX) (Roesner et al. 2005), globin Y (GbY) (Fuchs et al.

2006), eye-globin or globin E (GbE) (Kugelstadt et al. 2004, Blank et al. 2011a) and androglobin (Adgb) (Hoogewijs et al. 2012). These findings brought an unexpected diversity to the world of globins. Nevertheless, the physiological roles of these globins are still under debate or simply unknown.

Due to gene loss events during evolution every vertebrate lineage has their individual gene repertoire. Hb, Mb, Ngb, Cygb and Adgb are present in nearly all gnathostomata, whereas GbE, GbY and GbX are only present in some of the vertebrates.

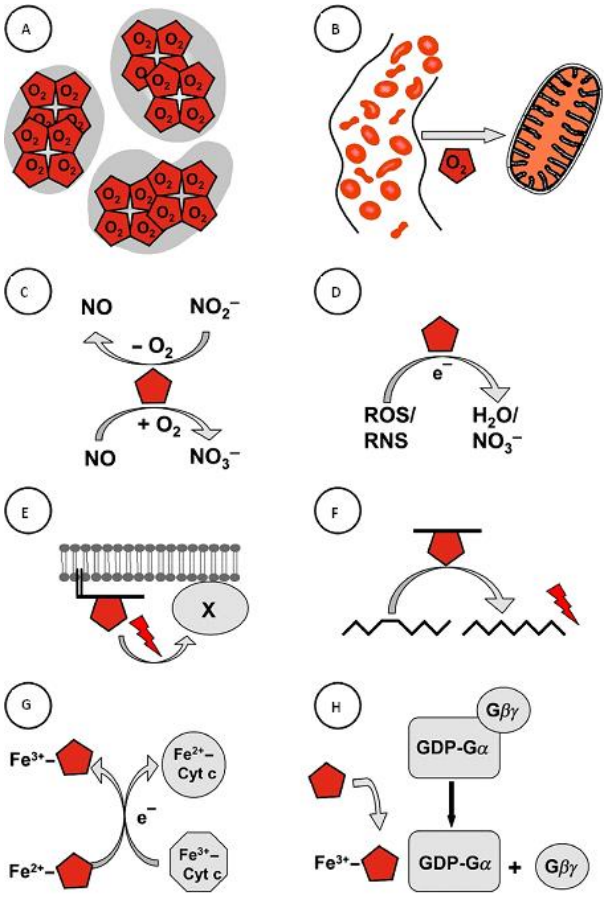


Figure I.2: Possible functions of the vertebrate globins.

A: Hb is able to transport O_2 in the blood; B: Mb, GbE and Ngb (possibly) enhance the intracellular O_2 supply to the mitochondria; C: Hb, Mb, Cygb and Ngb may have enzymatic activity in nitrite metabolism; D: Ngb, Cygb and Mb may detoxify ROS and RNS; E: GbX may be involved in a membrane-mediated signal transduction chain; F: Cygb may help at the fatty acid pathway; G: Ngb may prevent hypoxia-induced apoptosis by cytochrome c reduction; H: Ngb may work as a signal protein inhibiting GDP dissociation from Gα (α-subunit of G proteins) (Modified from (Burmester and Hankeln 2014))

1.1 Hemoglobin

Hb is a tetrameric protein consisting of two α- and two β-subunits in gnathostomata. Their interaction helps to cooperatively bind oxygen (Perutz 1979) and decreases their O_2 affinity, which is significantly lower than that of Mb. Humans harbor six *Hb* genes (α, β, γ, δ, ε, and ζ) which are differentially expressed in the embryo, fetus or adults (Dickerson and Geis 1983, Stamatoyannopoulos 1991). The tetramer of the fetal Hb (2 x α or ζ + 2 x ε) is prenatal replaced by the embryonic Hb (2 x α + 2 x γ). A few years after birth this will be replaced by the adult Hb (2 x α + 2 x β) (Wood 1976, Dickerson and Geis 1983). Depending on the tetrameric Hb structure the O_2 affinity varies: the fetal and embryonic Hb have a higher affinity than the adult Hb, providing O_2 diffusion from maternal to the fetal/embryonic circulatory system in the placenta (Bauer et al. 1975).

To transport the oxygen through the vascular system a large amount of Hb is present in the erythrocytes. This also results in the red colour of vertebrate blood. Hb takes up the oxygen from the respiratory surface and delivers it to the inner organs (Dickerson & Geis 1983). When Hb is deoxygenated it transports CO_2 to release it at the respiratory surface. Additionally, Hb has enzymatic activities. Oxy-Hb is able to convert the signalling molecule nitric oxide (NO) to nitrate whereas deoxy-Hb on the other hand can produce NO from nitrite (Cosby et al. 2003, Shiva et al. 2011). To ensure optimal O_2 supply during ontogenesis the hemoglobin chains of many vertebrates are differentially expressed. This is called hemoglobin switch (Brownlie et al. 2003, Fuchs et al. 2006, Tiedke et al. 2011).

Changes in pH (Bohr effect) and temperature as well as CO_2 or the interaction with organic phosphates (adenosine triphosphate, guanosine triphosphate, inositol pentaphosphate or 2,3-diphosphoglyceric acid) affect the O_2 affinity further (Berg, Tymoczko & Stryer 2006). As far as it is known, Hb is present in all vertebrates except for icefishes (Sidell and O'Brien 2006).

1.2 Myoglobin

Mb was the first protein with resolved structure on the atomic level (Kendrew 1963). In contrast to the heterotetrameric hemoglobin (Hb), myoglobin (Mb) is a monomeric protein containing ~150 amino acids (~17kDa) and is found mainly in the heart and skeletal muscle tissues. In the smooth muscles, endothelial and tumour cells it is expressed to a lower extent (Qiu et al. 1998, Cossins et al. 2009, Gorr et al. 2011).

Mb is responsible for the red colour of the muscles and holds, without any doubt, the key function in facilitating diffusion of O_2 to the respiratory chain in mitochondria. It has by far a higher affinity (P_{50} = ~1 Torr = 0.14 kPa; P_{50} represents the partial pressure necessary for 50% oxygen saturation of globins and is used to measure the oxygen affinity) to O_2 than Hb, which is important for the efficient O_2 extraction out of the blood. In the past 20 years, it has become evident that Mb also has enzymatic functions and plays a role in homeostasis of NO together with detoxification of ROS. Deoxy-Mb is able to reduce nitrite to NO during hypoxia (Shiva et al. 2007) and helps regulating O_2 consumption (Flogel et al. 2010, Helbo et al. 2013, Fago and Jensen 2015). To protect from an excess of ROS during reoxygenation, Mb-derived NO may target the mitochondrial complex I (Hendgen-Cotta et al. 2008). In addition, oxy Mb may convert the bioactive NO into nitrate (NO_3^-) (Flogel et al. 2010) as also known for Hb. The removal of excess NO enhances the mitochondrial respiration (Brunori 2001) as NO works as a reversible inhibitor of cytochrome oxidase (Moncada and Erusalimsky 2002). While reducing NO, oxy-(Fe^{2+}) Mb is oxidized to met-(Fe^{3+}) Mb that needs to be reduced by the cellular Mb reductase (Hendgen-Cotta et al. 2008). It has been suggested that under hypoxia, Mb has the opposite effect whereby deoxy-(Fe^{2+}) Mb functions as nitrite reductase producing NO from nitrite (NO_2^-) (Hendgen-Cotta et al. 2008). The NO inhibits cellular respiration and thereby limits the generation of ROS. It has also been suggested that Mb functions in oxidative defence by removing ROS (George and Irvine 1951, Osawa and Korzekwa 1991, Flögel et al. 2004). Other functions may involve the interaction of Mb with fatty acids (Shih et al. 2014).

In nearly all vertebrates analyzed so far only one myoglobin gene was found. The only exception found so far are the goldfish and the carp which harbor two different *Mb* genes (Mb1 and Mb2) due to a genome duplication event within the cyprinidea lineage. The two genes are furthermore expressed in distinct tissues. While Mb1 is widely expressed throughout various tissues Mb2 is restricted to neuronal tissue (Fraser et al. 2006, Cossins et al. 2009). These two Mbs apparently have distinct primary functions: *In vitro* studies have suggested that Mb1 mainly supplies O_2 and produces NO, whereas Mb2 is more efficient in eliminating H_2O_2 (Helbo et al. 2012). Some vertebrate species lack an *Mb* gene, such as frogs, some ice-fishes, the stickleback and the opossum *Monodelphis domestica* (Maeda and Fitch 1982, Fuchs et al. 2006, Sidell and O'Brien 2006, Hoffmann et al. 2011). These species must have specific mechanisms to compensate for the lack of Mb. In icefish, for example, the low water temperatures in the Antarctic sea may make Mb redundant because the physically dissolved O_2 in the cold water is

sufficient to support the very low metabolic rate of icefish (Sidell and O'Brien 2006). Studies in the frog *Xenopus laevis* suggest that the lack of Mb may be compensated by the expression of an additional monomeric Hb-like chain in the heart, which is not expressed in the blood (Maeda and Fitch 1982), or even by Cygb in muscle cells (Xi et al. 2007).

1.3 Neuroglobin

Ngb was the third discovered member of the globin superfamily and is phylogenetically one of the oldest globins which evolved before the split of deuterostomes and protostomes (Burmester et al. 2000, Droge et al. 2012). Ngb is a hexacoordinated monomer with about 150 amino acids and occurs in most vertebrates in relatively low concentrations of around 1 µM (Burmester et al. 2000). Ngb has a three-fold slower evolutionary rate than Hb or Mb and shares only <21% amino acid sequence similarity with vertebrate Mb and <25% with Hb (Van Acker et al. 2018). In the presence of reducing agents, it can reversibly bind O_2 with a quite high affinity (i.e. low P_{50}) in a range of 0.9-2.2 Torr (0.12-0.29 kPa) which is similar to Mb. Due to the sixth coordinated structure of the heme iron O_2 cannot freely bind, but the occupying His residue can transiently dissociate and allow external ligands to bind (Hankeln et al. 2005). Ngb is widely expressed throughout the central and peripheral nervous system with the highest expression in the mammalian hypothalamus (Fabrizius et al. 2016). Interestingly, it has still relatively low concentrations in the brain when compared to Mb in the muscle. It has been suggested that in the retina a proportion of Ngb protein resides within the mitochondria (Lechauve et al. 2012, Yu et al. 2012), thereby directly interacting with the respiratory chain (Lechauve et al. 2012, Lechauve et al. 2013). This indicates that Ngb is a promising candidate for preventing the degeneration of retinal neurons due to mitochondrial dysfunction (Lechauve et al. 2012 528). On the other hand, there is little doubt that the main fraction of Ngb is in the cytoplasm.

Ngb may be involved in detoxification of NO (Brunori et al. 2005) and may function as a nitrite reductase for the inhibition and protection of the mitochondrial respiration (Tiso et al. 2011). There is also a conclusive evidence that Ngb is associated with the mitochondria and therefore with the oxidative metabolism (Bentmann et al. 2005, Mitz et al. 2009) and the protection of neurons via redox chemistry (Herold et al. 2004). Additionally, it is proposed that Ngb plays an important function in the retina, where it is highly abundant (Yu and Cringle 2001, Lechauve et al. 2013, Schwarze et al. 2015). The most discussed functions of Ngb range from O_2 transport to the mitochondria (although this function needs reducing agents to keep Fe in the ferrous

state, as the Ngb tends to oxidize with oxygen) to the detoxification of ROS through reduction. In addition, Ngb is involved in control of the conversion from NO to NO_3^- under normoxia or production of NO under hypoxia. Furthermore, it may as well function as a signaling protein through interaction with GDP-Gα and inhibition of apoptosis through reduction of cytochrome c (Burmester and Hankeln 2014).

In 2013 Yu et al. showed that Ngb plays an important role in mitochondrial functions such as ATP production, ROS generation, and apoptosis signaling (Yu et al. 2013a). Ngb overexpression enhances the ATP production and preserves the mitochondrial MTT (3-(4, 5-dimethylthiazolyl-2)-2, 5-diphenyltetrazolium bromide) formation, which is an indirect index of mitochondria viability (Liu et al. 2009).

Additionally, Ngb overexpression *in vitro* protects the brain as well as the heart from hypoxia and oxidative stress; in particular it protects the neurons from mitochondrial dysfunction, which is correlated with a wide range of neurological disorders (Khan et al. 2007, Szymanski et al. 2010, Yu et al. 2012, Yu et al. 2013a, Yu et al. 2013b, Fiocchetti et al. 2014). Cell culture experiments in either primary mouse cortical neurons or the neuroblastoma cell-line SH-SY5Y show that Ngb overexpression is needed to rescue the ATP loss during hypoxia (Duong et al. 2009, Liu et al. 2009). This suggests that there is a strong correlation between neuroprotection and overexpression of this globin in hypoxia. The exact mechanism has not yet been properly described.

Ngb is also reported to build a complex together with $G_{\alpha i}$ *in vitro* (Fig I.2). Heterotrimeric guanine-nucleotide-binding proteins (G proteins) serve as molecular switches in signaling pathways playing central roles in a wide variety of biological processes depending on the binding of guanine nucleotides. These heterotrimeric G proteins are composed of three subunits, an α-subunit (Gα) with GTPase activity and a βγ-dimer (Gβγ) (Gilman 1987, Bourne et al. 1991, Simon et al. 1991, Hepler and Gilman 1992, Sprang 1997, Vetter and Wittinghofer 2001, Oldham and Hamm 2006, Oldham and Hamm 2008). The Ngb-$G_{\alpha i}$-complex inhibits the rate of GDP/GTP exchange, which gives Ngb the role of a guanine nucleotide dissociation inhibitor (GDI) (Wakasugi et al. 2003, Wakasugi and Morishima 2005). The interaction of ferric Ngb with the GDP-bound $G_{\alpha i}$ subunit inhibits the G_α activity and prevents the rebinding of $G_{\beta\gamma}$ to G_α. This enhances the $G_{\beta\gamma}$-dependent pathway favoring cell survival via the activation of phosphatidylinositol 3-kinase (PI3K). In the case of oxidative stress, the ferric Ngb undergoes a

structural change and binds to $G_{\alpha i}$ performing its GDI function and preserving the cells from stress-induced death (Watanabe et al. 2012). Under normoxia, ferrous ligand-bound Ngb has no GDI activity *in vitro* (Wakasugi et al. 2003). GDI activity of mammalian Ngb is tightly correlated with its neuroprotective activity (Watanabe and Wakasugi 2008).

Within physiological conditions the functions of Ngb are very modest but under the cellular stress they become strongly upregulated, leading to the conclusion that Ngb is a particularly stable protein (Van Acker et al. 2018).

Ngb-mRNA is found in every vertebrate class (mammals, fish, amphibians, reptiles and birds) so far (Burmester et al. 2000, Awenius et al. 2001, Kugelstadt et al. 2004, Fuchs et al. 2006, Milton et al. 2006). Therefore it is presumed to exist in all vertebrates with a conserved and important function, as the evolutionary rate at the amino acid level is very low (Wystub et al. 2004).

1.4 Cytoglobin

Cygb was discovered and named simultaneously by three research teams in 2001/2002 (Kawada et al. 2001, Burmester et al. 2002, Trent and Hargrove 2002). Due to the ubiquitous expression in mammalian tissues the name 'cytoglobin' pushed through (Burmester et al. 2002). Cygb appears as a dimer (Sugimoto et al. 2004) with an O_2 affinity of around 1 Torr (0.14 kPa) (Trent and Hargrove 2002, Fago et al. 2004a) and occurs in all vertebrates studied so far (Schwarze and Burmester 2013). It consists of 174 to 190 amino acids with a molecular mass of 21 kDa which makes it bigger than the rest of the globin-superfamily. The main expression sites are the fibroblasts and some neurons of the central and peripheral nervous system (Hankeln et al. 2004, Nakatani et al. 2004, Schmidt et al. 2005). It is also present at a lower extent in epithelial and muscle cells, macrophages, and tumor cells (Oleksiewicz et al. 2011, Oleksiewicz et al. 2013). A respiratory function of Cygb is unlikely as there was no correlation of Cygb and O_2 consumption seen (Hankeln et al. 2005, Schmidt et al. 2005). It is assumed that Cygb functions in the NO metabolism, fibrosis, signal transduction, or contributes to the protection from ROS as well as being a tumor suppressor (Hankeln et al. 2005, Schmidt et al. 2005, Xu et al. 2006, Fang et al. 2011, Reeder et al. 2011, Oleksiewicz et al. 2013). In most bony fishes two isoforms of Cygb, *Cygb1* and *Cygb2*, exist (Fuchs et al. 2005, Hoffmann et al. 2011). In zebrafish, *Cygb1* is ubiquitous expressed and *Cygb2* is mainly expressed in neuronal tissues (Tiedke et al. 2011, Corti et al. 2016).

1.5 Globin E

GbE ('eye-globin') was first detected in chicken (Kugelstadt et al. 2004) and later on in other bird genomes (Blank et al. 2011a, Blank et al. 2011b, Storz et al. 2011). Therefore, it was first suggested to be a bird-specific protein. Recent studies observed GbE also in turtles (Schwarze et al. 2015) and coelacanth (*Latimeria chalumnae*) (Schwarze and Burmester 2013). The functional residues like CD1 phenylalanine, the proximal (F8) histidine, and the distal (E7) histidine are conserved in all GbE sequences (Hoffmann et al. 2011). GbE is one of the most highly conserved globins and is characterized by a level of functional constraint similar to that of both Mb and Hb. GbE is penta-coordinated and has a quite low O_2 affinity (P_{50} = 5.8 Torr at 25°C) compared to myoglobin and a slow autoxidation rate ($t_{1/2}$ = ~ 10 h at 25°C) (Blank et al. 2011a). Phylogenetically, it is closely related to myoglobin which suggests a similar function. It is tissue specific and highly expressed in the retina where it is preferentially located in the outer segments of the photoreceptor cells (Blank et al. 2011a, Schwarze et al. 2015). As the avascular retina of birds and turtles has a high metabolic rate, it has been proposed that GbE carries an oxygen storage function in this tissue (Blank et al. 2011a, Hoffmann et al. 2011, Storz et al. 2011, Storz et al. 2013, Schwarze et al. 2015). It originated early in vertebrate evolution as seen by phylogenetic analyses. Within this study, multiple copies of GbE were also found in the lungfish ovary (Ludemann et al. 2019) but not in the retina.

1.6 Globin Y

The structure of GbY looks like a typical penta-coordinated globin, but no *GbY* genes were found in 'higher mammals' (Placentalia and Marsupialia), birds or ray-finned fishes (Actinopterygii) (Burmester and Hankeln 2014). It was first discovered in Xenopus (Fuchs et al. 2006) but later on also found in other vertebrates, for example the platypus (*Ornithorhynchus anatinus*), the lizard (*Anolis carolinensis*), the central bearded dragon (*Pogona vitticeps*), the Australian ghostshark (*Callorhinchus milii)*, the coelacanth, and turtles (Patel et al. 2008, Hoffmann et al. 2010, Patel et al. 2010, Hoffmann et al. 2011, Schwarze 2011, Schwarze and Burmester 2013, Schwarze et al. 2015). Overall, GbY has a relatively low expression but shows the highest mRNA expression in ovary, kidney, eye and lung (Fuchs et al. 2006, Schwarze et al. 2015). During ontogenesis there seem to be timeframes with a high rate of GbE mRNA expression (Fuchs 2007). To date there is no definition of the GbE function or its phylogenetic position. However, synteny studies in the western painted turtle (*Chrysemis picta bellii*) and the Coelecanth

genome show a possible phylogenetic connection of GbY to Hb (Schwarze and Burmester 2013, Schwarze et al. 2015).

1.7 Globin X

GbX together with Ngb belong to the ancient globins. GbX was identified in vertebrates (lampreys, sharks, ray-finned fishes, coelacanth, amphibians and reptiles) but not in birds and mammals (Roesner et al. 2005, Fuchs 2007, Kruszewski 2011, Blank and Burmester 2012, Schwarze and Burmester 2013). It is a hexa-coordinated membrane-bound globin in the central nervous system and it is ~200 amino acids longer than the average globin. This is due to the C- and N-terminal elongations of the protein. In contrast to other globins, the N-terminus shows conserved myristoylation and palmytoylation sites at Gly2 and Cys3 that serve to attach it to the cell membrane (Blank et al. 2011b). This gives a hint to a potential function in protecting the membrane lipids from oxidative stress or to some other membrane-related cellular signaling process (Blank et al. 2011b). The idea of oxidative stress protecting function of GbX at the cell membrane is supported by *in vitro* experiments (Koch and Burmester 2016) that demonstrate that the membrane-binding of GbX is essential for its function.

1.8 Androglobin

Adgb is a multi-domain protein that consists of around 1600 amino acids in humans and mice and is predominantly expressed in testis. It is a 180 kDa chimeric globin containing an N-terminal calpain-like domain, an internal globin domain and an IQ calmodulin-binding motif ([I,L,V]QXXXRXXXX[R,K]) (Bahler and Rhoads 2002, Hoogewijs et al. 2012). It is an ancestral chimeric globin type that is also found in almost all Metazoan lineages and in choanoflagellates. It plays a potential role in reproduction by controlling the sperm maturation as suggested by an increase of the gene expression in fertile males (Hoogewijs et al. 2012). The globin domain of Adgb is split into two parts and permutated (helices C to H followed by helices A and B). Absorbance spectra of recombinantly expressed Adgb indicate a hexa-coordination. Interestingly, Adgb couldn't be found in *C. elegans* or *D. melanogaster*.

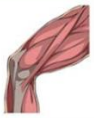
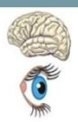
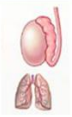
	Hemoglobin	Myoglobin	Globin E	Globin Y	Cytoglobin	Globin X	Neuroglobin	Androglobin
								
Expression sites	red blood cells	skeletal muscle, heart, smooth muscle	eye	kidney, lungs, eye, ovary	fibroblasts, nerve cells	brain, muscle, gut, liver, kidney, gills	neurons, retina, endocrine tissue	testis, lungs
Size [kDa]	~64	~17	~17	~17	~21	~39	~17	~180
Fe-atom coordination	penta	penta	penta	hexa	hexa	hexa	hexa	hexa
O_2 affinity [Torr]	26	1	6-15	?	1	1-13	1-2	?

Figure I.3: Summary of the 8 different vertebrate globins.

The graphic shows the main expression sites, the Fe-atom coordination and the O_2 affinity of the 8 different vertebrate globins. (Modified from (Hankeln et al. 2005, Hoffmann et al. 2010, Laufs 2010, Blank and Burmester 2012, Ebner 2012, Hoogewijs et al. 2012, Fabrizius 2014))

1.9 Evolution of globins

When proteins evolve, their function needs to be maintained in order for the organism to survive. Therefore genome evolution is characterized by gene- and genome-duplications followed by gene loss or retention as well as functional differentiation and physiological adaptations (Ohno 1970, Wolfe and Li 2003). New genes originate by the duplication of genes and genomes, in which the whole genome, individual sections of the chromosomes or single genes/fragments can be duplicated (Fig. I.4) (Davis and Petrov 2005). The resulting new duplicates can have 1) no function (pseudogene), 2) gain a new function (neofunctionalization), 3) subfunctionalize or 4) just lead to more gene product by an additive effect (Ohno 1970, Wolfe 2001). Subfunctionalization means that the original gene has two (or more) functions which are split among the new gene copies. The resulting duplicates perform different functions (Lynch and Force 2000). In contrast to that, neofunctionalization means that one copy keeps the old functions, whereas the other copy gains a new function. Besides the pseudogenization, all the different duplication scenarios result in 2 functional, active genes. The highly essential genes are expected to be retained in phylogenetically diverse taxa and to exhibit especially slow rates of amino acid substitutions relative to paralogous members of the same gene family.

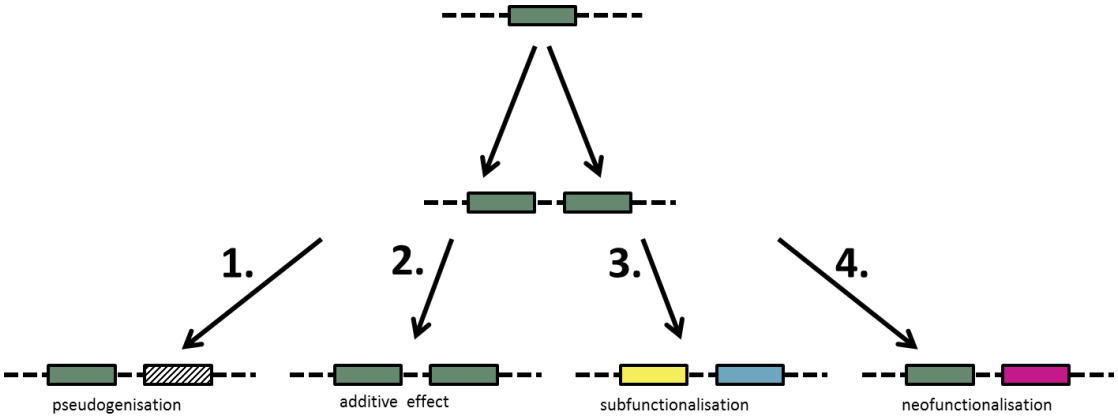


Figure I.4: How do new genes emerge?

The graphic shows the duplication of a genes through gene- or genome-duplications and the possible outcome; 1.:Pseudogenization, which means a loss of one gene copy; 2.: Additive effect with more gene product; 3.:Subfunctionalization resulting in two copies with two different functions of the original gene; 4.: Neofunctionalization which results in one copy with a new function (Ohno 1970, Wolfe 2001)

The majority of vertebrate globins emerged by gene duplication or whole-genome duplication events that occurred in the stem-lineage of vertebrates (Hoffmann et al. 2011, Storz et al. 2011). The two monomeric globins Ngb and GbX are more ancient as they derived from duplications that predate the origin of deuterostomes (Roesner et al. 2005, Ebner et al. 2010). Due to this scientific finding together with their broad phyletic distribution, these two globins are assumed to perform physiological functions related to very fundamental aspects of cellular metabolism.

Synteny and phylogenetic analysis revealed that *Mb* and *GbE* genes represent products of a subsequent tandem gene duplication in the stem lineage of gnathostome vertebrates, whereas the proto-*Mb/GbE* gene and the *Cygb* gene represent paralogous products of a whole-genome duplication event in the stem lineage of vertebrates (Fig. I.4) (Hoffmann et al. 2011). This assumes that *Cygb*, *GbE*, and *Mb* genes were already present in the genome of the gnathostome common ancestor. This also concludes that *GbE* is not the result of a recent bird-specific duplication event but it was rather only retained in birds and got lost in all other gnathostome lineages independently (Hoffmann et al. 2011).

The diverse family of vertebrate globins emerged through small-scale duplications and whole-genome duplication events. The first round of whole-genome duplication produced two paralogons: A proto "*Mb/GbE/Cygb*" and a proto "*α/β-Hb* + *GbY*" paralogon. The subsequent round of whole-genome duplication led to reduplication of the proto "*Mb/GbE/Cygb*"

paralogon which gave rise to the progenitors of *Cygb* and *Mb/GbE* gene lineages. The reduplication of the "*α/β-Hb + GbY*" paralogon gave rise to the progenitor of the α- and β-globin gene families as well as a fourth paralogous globin gene lineage that does not appear to have been retained in any extant vertebrate taxa (Hoffmann et al. 2012). The vertebrate *GbX* and *Ngb* genes diverged early in animal evolution. Their progenitors were also present in the vertebrate common ancestor and were therefore duplicated and reduplicated through two rounds of whole-genome duplication but then reversed to the ancestral single-copy state (Hoffmann et al. 2012).

Key physiological innovations in the vertebrate oxygen transport system reveal a direct link to the vertebrate-specific whole genome duplications. In detail, it shows the functional division of a duplicated gene pair consisting of the oxygen transport protein proto-Hb and the oxygen storage protein proto-Mb, which evolved after the first round of whole-genome duplication (Hoffmann et al. 2012). This first round set the stage for the physiological division of labour in the tissue specific gene expression between the evolutionary precursors of Mb and Hb. The following tandem duplication provides a further functional specialization concerning blood-gas-transport producing α- and β-globin genes. The functional divergence of these proto α- and β-genes leads to the heteromeric quaternary structure of Hb necessary for cooperative oxygen-binding and allosteric regulation of oxygen binding to provide sufficient O_2 to tissues (Fig. I.5) (Hoffmann et al. 2012). Cygb is the third paralog derived through whole-genome duplication with much specialized functions in either gnathostomes or cyclostomes. It is characterized by higher levels of amino acid sequence conservation in comparison with Mb and Hb (Burmester et al. 2002, Hoffmann et al. 2011), suggesting an essential physiological role.

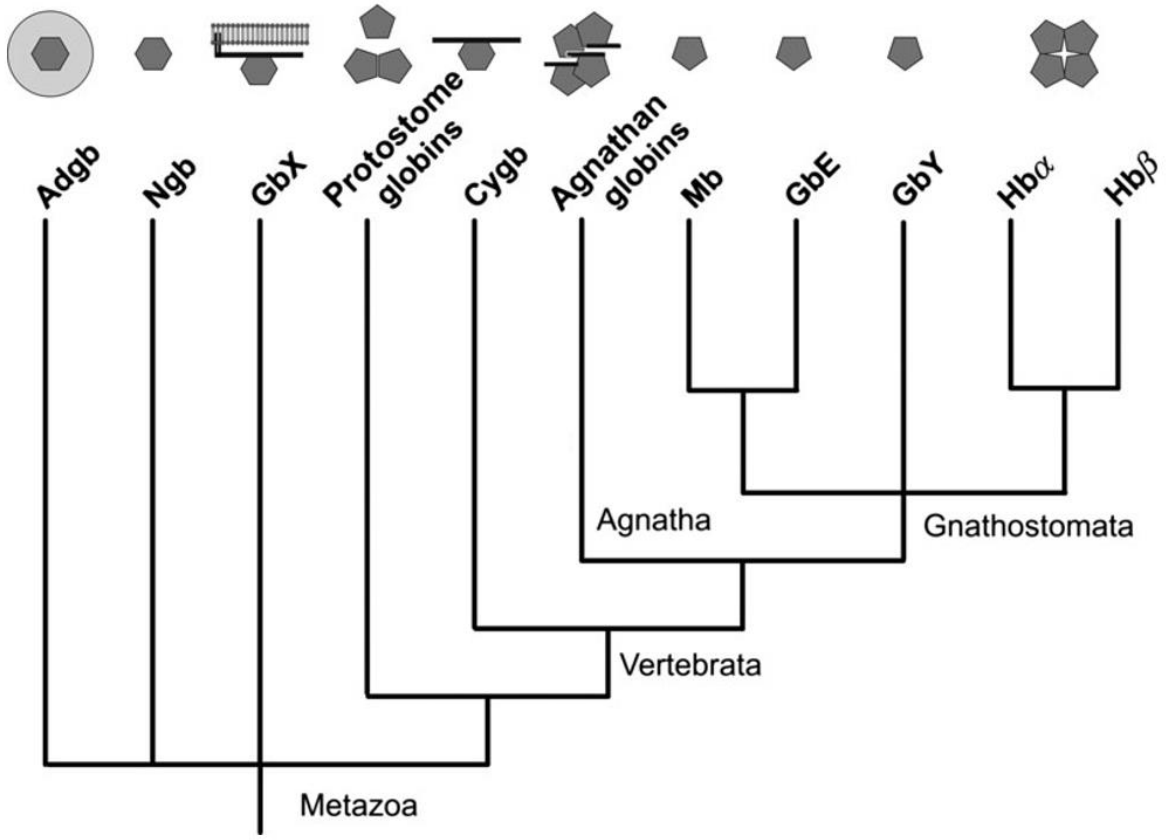


Figure I.5: Hypothesized phylogenic relationship of animal globins.

Modified from (Burmester and Hankeln 2014)

1.10 Lungfish

Lungfish are fresh-water fish that have the ability to take up O_2 through their lungs while also having gills. They belong to the subclass of Dipnoi within the Osteichthyes (bony fishes) and first appeared in the Devonian period around 400 million years ago (Chang and Yu 1984). They represent the closest living relatives of the land-living vertebrates (tetrapods) as seen in phylogenetic analyses (Fig. I.6) (Brinkmann et al. 2004, Amemiya et al. 2013, Biscotti et al. 2016) and build an evolutionary bridge between the tetrapods and the aquatic vertebrates (Amemiya et al. 2013). They harbor an interesting position in the field of evolution of tetrapods and the physiological adaptations accompanying the transition from water to land as they retained their primitive characteristics. Together with the coelacanth they belong to the Sarcopterygii.

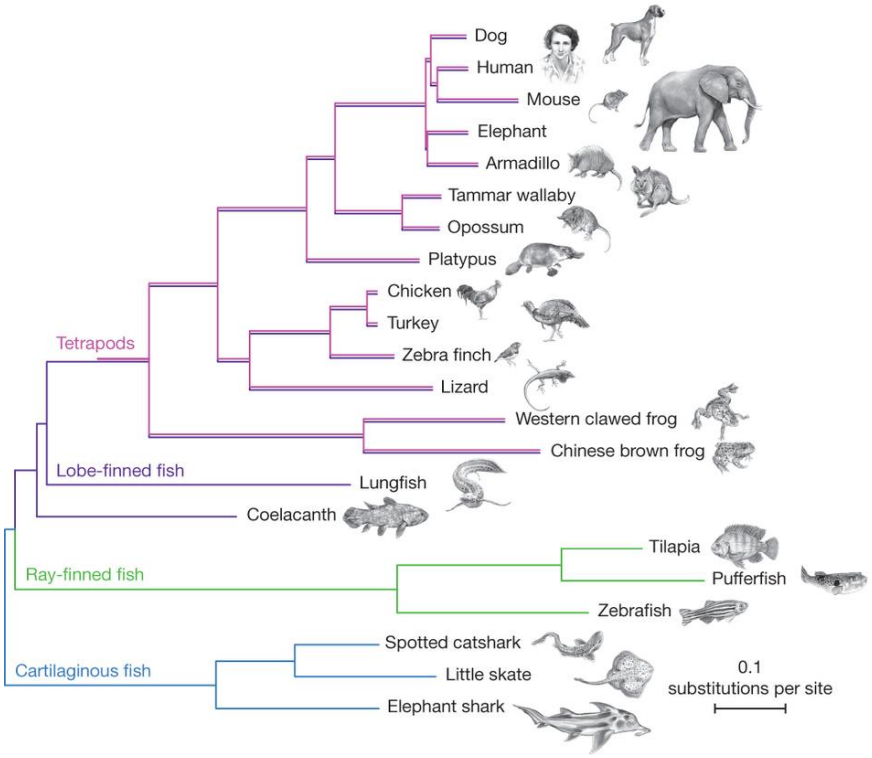


Figure I.6: Phylogenetic relationship of jawed vertebrates shows that the lungfish is the closest living relative to tetrapods. The graphic displays a multiple sequence alignment of 251 genes with a 1:1 ratio of orthologues in 22 different vertebrates used to generate a concatenated matrix of 100,583 unambiguously aligned amino acid positions showing that the lungfish and not the coelacanth is the closest relative of land-living tetrapods. (Amemiya et al. 2013)

There are six extant lungfish species found in fresh-water habitats in Australia, South America and Africa that belong to three different families (Fig. I.7). The Queensland lungfish (*Neoceratodus forsteri)*, the most ancient genus, belongs to the order of Ceratodontiformes and the family of Ceratodontidae and has, in contrast to the other species, just one lung. The other order is Lepidosireniformes that includes the two genera *Lepidosiren* and *Protopterus* (Nelson 2016). The South American lungfish (*Lepidosiren paradoxa)* belongs to the family of Lepidosirenidae. The four different African lungfish all belong to the family of Protopteridae, comprising the Marbled lungfish (*Protopterus aethiopicus*), the Gilled African lungfish (*Protopterus amphibius*), the West African lungfish (*Protopterus annectens*), and the Spotted African lungfish (*Protopterus dolloi*). The different families evolved separately since the beginning of the continental drift that promoted the splitting of big continents (Zaccone 2016).

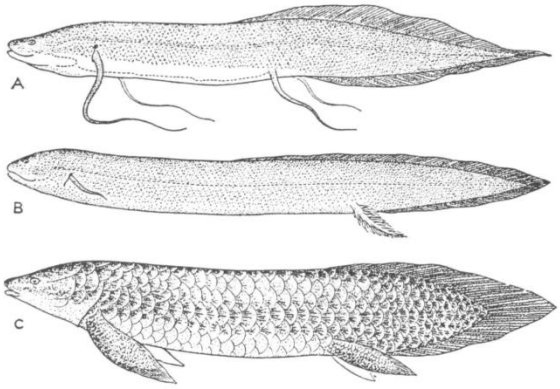


Figure I.7: The three living lungfish genera.

A: African lungfish (*Protopterus annectens*); B: South American lungfish (*Lepidosiren paradoxa*); C: Australian lungfish (*Neoceratodus forsteri*) (Modified from (Norman 1963)).

The African and South American lungfish are obligate air breathers that typically dwell in freshwaters with low dissolved O_2 and high daytime temperatures. They normally have two lungs along with the gills, which is unique throughout the family of fish. They gulp air every 3 to 5 minutes in their natural habitat (Australian lungfish: 30 to 60 minutes) (Hutchins 2003, Zaccone 2016). Around 90% of their O_2 uptake relies on their lungs and they use the gills mainly to excrete 70-80% of their CO_2 and nitrogenous wastes and to regulate the acid-base balance. Only the Australian lungfish needs to respire also through its gills, as it is the most primitive form with only a single lung (Hutchins 2003).

The body of the lungfish is elongated with paired pectoral and pelvic fins. What is special about them, is the presence of the lymphatic system, which is similar to the tetrapods, but not to the other fishes. They are very inactive and mostly stay at the bottom of their water habitat. They can aestivate in a mud cocoon during summer to survive periods of seasonal drought (Ballantyne and Frick 2010). During this period, they reduce their energy metabolism and turn on cytoprotective genes, e.g. to convert ammonia to urea (Hiong et al. 2013).

1.10.1 Aestivation

Aestivation is an adaptive mechanism described for lungfishes, teleost fishes, amphibians, reptiles, birds, and mammals (Gregory 1982, Geiser 1988, Pinder 1992, Graham 1997). It helps the organism to withstand hot and dry environments with low food availability (Pinder 1992, Storey 2000, Secor and Lignot 2010). It is described as “a behavioral strategy accompanied by physiological adjustments” (Seidel 1978) and characterized by three features: 1) inactivity, 2)

fasting, and 3) decrease in basal metabolism (for many species) (Pinder 1992, Guppy and Withers 1999).

Of the five lungfish species that aestivate, the aestivation of *P. annectens* is best described (Greenwood 1987). When the water-level lowers, the lungfish digs a vertical burrow in the mud and positions itself in a U-shaped manner, head and tail facing upwards. With the help of secreted mucous from epithelial mucous glands the lungfish builds a hard cocoon surrounding the body. A small opening is left at the mouth to allow breathing (Greenwood 1987). When the lungfish is dormant it has a decreased gas exchange, cardiac performance, and ammonia production which all together help to reduce their energy expenditure and maintain the homeostasis (Janssens 1968, Delaney et al. 1974, Fishman 1986, Lomholt 1993).

1.11 Multiple globins in lungfish

The genome of the lungfish is the largest known throughout the animal kingdom ranging from ~60 Gbp in *P. annectens* to ~130 Gbp in *P. aethiopicus,* so it is up to 40 times larger than the human genome (3.2 Gbp) (Gregory 2018). During the genome discoveries BLAST searches revealed at least eight distinct *Mb* gene copies in the lungfish, differentially expressed in brain, kidney+gonad and liver+gut. The full-length sequences translate into proteins of 142 to 154 amino acids. All eight Mbs harbor the invariant Phe residue at position CD1 and the Fe^{2+}-binding His at F8 (Koch et al. 2016). Further analysis identified a ninth Mb sequence in *P. annectens* together with multiple Mb copies in the other lungfish as well (Lüdemann et al, 2019, submitted). This finding is unique throughout the animal kingdom as usually species only have one Mb copy. There are a few exceptions, for example some fishes, like the carp and the goldfish, have two Mb copies, others, like frogs, icefishes the stickleback and opossums do not harbor any Mb at all.

Phylogenetic analyses show that the paralogous myoglobins of *P. annectens* developed several million years ago.

Together with the multiple Mb copies four Hb α, five Hb β, one GbX, one GbY, and (preliminary) one GbE in the transcriptomes of the West African lungfish were identified in this study (Koch et al. 2016). No Ngb or Cygb sequence could be detected in the transcriptome.

Additionally, the transcriptomes of the South American lungfish *L. paradoxa* were assembled and checked for the globin repertoire in this species. BLAST searches revealed two Hb α, three

Hb β, one GbX, one GbY, five Mb, and, unexpectedly, six distinct GbE transcripts specifically expressed in the ovary of *L. paradoxa* (Ludemann et al. 2019). Again, no Cygb or Ngb was identified in the transcriptome of *L. paradoxa*, suggesting the absence of these genes in all lungfish genomes.

The transcriptomes of the other lungfish revealed 6 Mbs in *P. dolloi*, 5 Mbs and 7 GbEs in *P. aethiopicus*, and, after re-examination, 5 GbEs in *P. annectens*.

1.12 Aim of the thesis

Globins are a classical example for the evolution of genes. The homologies in their amino acid sequence and the highly conserved tertiary structure, the so-called 'globin-fold', indicate that globins derive from an ancient, common ancestral gene. Proteins with O_2 binding capacities evolved to allow multicellular animals to grow to a large size, because they could not only rely on simple O_2 diffusion through the body surface. It appears that about 500 million years ago a series of gene mutations and duplications occurred to differentiate them. Multiple vertebrate-specific duplication events generated the vertebrate globin gene repertoire with additional globin functions besides the transport of O_2 (Hardison 1998, Alberts et al. 2002).

In the beginning of this PhD project multiple globin copies in the West African lungfish were identified. Therefore, the primary objective of this study was to fully identify the multiple globin genes and to identify how they emerged to better understand the underlying evolution and functional differentiation. Studies on their expression pattern, oxygen-binding kinetics, enzymatic functions, and their three-dimensional structure were employed to solve the emerging issues.

In part one of this thesis the multiple *Mb* genes, their expression, their O_2 binding affinities and their enzymatic functions were analyzed. During the studies on the multiple *Mb* genes new lungfish transcriptomes emerged, which gave more detailed insights into the lungfish globin-family. As a result the second part addresses the surprising finding of multiple GbE copies in the lungfish genome.

This thesis aims to find answers on how the multiple globin genes evolved after multiple gene duplication events and what their underlying function is, if the distinct globins have specific functions other than O_2 supply. This can further help to provide insights into the functional flexibility of the family of globin proteins in evolution and may hint to other functions, also of mammalian globins. The study of respiratory protein may also shed light on some of the evolutionary steps taken in the transition from water to land.

II. Chapter 1:

Unusual diversity of myoglobin genes in the lungfish

Jonas Koch[1], Julia Lüdemann[1], Rieke Spies[1], Marco Last[1], Chris T. Amemiya[2,3] and Thorsten Burmester[1]

[1]Institute of Zoology, University of Hamburg, D-20146 Hamburg, Germany
[2]Molecular Genetics Program, Benaroya Research Institute, Seattle, Washington 98101, USA
[3]Department of Biology, University of Washington, Seattle, Washington 98105, USA

Published in *Molecular Biology and Evolution*, December 2016, Volume 33, Issue 12, Pages 3033-3041

Supplementary Material can be found on the enclosed CD-ROM

2.1 Abstract

Myoglobin is a respiratory protein that serves as a model system in a variety of biological fields. Its main function is to deliver and store O_2 in the heart and skeletal muscles, but myoglobin is also instrumental in homeostasis of nitric oxide (NO) and detoxification of reactive oxygen species (ROS). Almost every vertebrate harbors a single myoglobin gene; only some cyprinid fishes have two recently duplicated myoglobin genes. Here we show that the West African lungfish *Protopterus annectens* has at least seven distinct myoglobin genes (*PanMb1-7*), which diverged early in the evolution of lungfish and showed an enhanced evolutionary rate. These myoglobins are lungfish-specific, and no other globin gene was found amplified. The myoglobins are differentially expressed in various lungfish tissues, and the brain is the main site of myoglobin expression. The typical myoglobin-containing tissues, the skeletal muscle and the heart, have much lower myoglobin mRNA levels. Muscle and heart express distinct myoglobins (*PanMb1* and *PanMb3*, respectively). In cell culture, lungfish myoglobins improved cellular survival under hypoxia albeit with different efficiencies and reduced the production of reactive oxygen species. Only *Mb2* and *Mb6* enhanced the energy status of the cells. The unexpected diversity of myoglobin hints to a functional diversification of this gene: Some myoglobins may have adapted to the O_2 requirements of the specific tissue and help the lungfish to survive hypoxic periods; other myoglobins may have taken over the roles of neuroglobin and cytoglobin, which appear to be missing in the West African lungfish.

2.2 Introduction

Myoglobin (Mb) is a small heme-protein that resides mainly in the striated (skeletal and heart) muscles of vertebrates (Wittenberg and Wittenberg 2003). Mb is probably the best-studied protein in biological sciences and was the first protein whose structure was resolved at the atomic level (Kendrew et al. 1958, Kendrew 1963). Mb is a member of the globin superfamily that also includes hemoglobin and other O_2 binding proteins from vertebrates and invertebrates (Dickerson and Geis 1983, Perutz 1983, Hardison 1996b). For about a century, Hb and Mb had been considered the only globins of vertebrates. However, in recent years six additional globin types have been identified: Neuroglobin (Ngb) (Burmester et al. 2000), cytoglobin (Cygb) (Kawada et al. 2001, Burmester et al. 2002, Trent and Hargrove 2002), globin X (GbX) (Roesner et al. 2005), globin Y (GbY) (Fuchs et al. 2006), eye-globin or globin E (GbE) (Kugelstadt et al. 2004, Blank et al. 2011a), and androglobin (Adgb) (Hoogewijs et al. 2012). These discoveries have added an unprecedented diversity to the vertebrate globin family. While Hb, Mb, Ngb and Cygb are widespread and may be present in almost any gnathostome (jawed) vertebrate (Burmester et al. 2004), GbE, GbX and GbY are restricted to certain taxa, indicating differential loss and retention of globins in vertebrate lineages (Hoffmann et al. 2011, Hoffmann et al. 2012, Storz et al. 2013). The physiological roles of most of these globins are still not well understood (Hankeln et al. 2005, Burmester and Hankeln 2009, Burmester and Hankeln 2014).

High concentrations of Mb are responsible for the red color of the skeletal and heart muscles of most vertebrates. There is no doubt that the main function of Mb is to enhance O_2 supply to the respiratory chain of mitochondria to meet the high energy demand of the myocytes (Merx et al. 2001, Wittenberg and Wittenberg 2003, Helbo et al. 2013). The O_2 affinity of Mb is typically higher than that of Hb, allowing for efficient extraction of O_2 from the blood. Mb may physically carry O_2 to the mitochondria or facilitate O_2 diffusion (Wittenberg and Wittenberg 2003). Surprisingly, *Mb*-knockout mice showed no obvious physiological defects (Garry et al. 1998) but displayed several compensatory mechanisms that enhance O_2 supply (e.g. higher capillary density in the heart) (Gödecke et al. 1999). Recent studies demonstrate a more widespread expression of Mb than initially thought and its presence, for example, in smooth muscle, endothelial and tumor cells, albeit at much lower concentrations than in the heart and the skeletal muscles (Qiu et al. 1998, Cossins et al. 2009, Gorr et al. 2011, Kristiansen et al.

2011). Mb may also have enzymatic functions. Employing Mb-deficient mice, Flögel et al. (2001) demonstrated that Mb is instrumental for decomposition of bioactive nitric oxide (NO) to nitrate (NO_3^-). NO acts as a reversible inhibitor of cytochrome oxidase (Moncada and Erusalimsky 2002), and removal of NO enhances mitochondrial respiration (Brunori 2001). It has been suggested that Mb may also function in oxidative defense by removing reactive oxygen species (ROS) (George and Irvine 1951, Osawa and Korzekwa 1991, Flögel et al. 2004). *In vitro* studies showed that Mb may interact with fatty acids, which could be metabolically relevant under oxygenated conditions and high energy demands (Shih et al. 2014).

Given the variety of Mb functions and its pivotal role in skeletal muscles, it is surprising that some species apparently lack an *Mb* gene. *Mb*-less vertebrates include frogs, some icefishes, the stickleback *Gasterosteus aculeatus* and the opossum *Monodelphis domestica* (Maeda and Fitch 1982, Fuchs et al. 2006, Sidell and O'Brien 2006, Hoffmann et al. 2011). How Mb deficiency is compensated for is not known. Other vertebrate species usually possess only a single *Mb* gene in the genome. The only known exceptions are carp and goldfish, which have Mb isoforms with different tissue expression patterns (Fraser et al. 2006, Roesner et al. 2008). While Mb isoform 2 is specific to brain neurons, the Mb isoform 1 displays a rather widespread expression in muscles, but also various other tissues (Fraser et al. 2006, Cossins et al. 2009). *In vitro* data suggest that Mb1 exerts its standard roles in O_2 supply and NO production while Mb2 is specifically active in eliminating H_2O_2 and could thus function in protecting neurons from ROS in the hypoxia-adapted cyprinids (Helbo et al. 2012).

Here we have investigated the unexpectedly high diversity of the *Mb* genes in the West African lungfish *Protopterus annectens.* Lungfish (Dipnoi) are freshwater fish that share with the land-dwelling tetrapods the ability to breathe air by lungs along with their gills, or have lost their gills completely. Since their discovery in the 19th century, these animals have attracted tremendous scientific interest. Lungfish appeared in the fossil record in the Devonian period (Chang and Yu 1984). Together with the coelacanth (*Latimeria*) and the tetrapods (land-living vertebrates), lungfish belong to the Sarcopterygii. For a long time, the relative relationship among these three taxa was hotly disputed. Recent phylogenomic evidence suggests that the lungfish are the closest living relative of the tetrapods (Brinkmann et al. 2004, Amemiya et al. 2013, Biscotti

et al. 2016). Thus, the lungfish harbors an important position for investigating the evolution of tetrapods and the physiological adaptations accompanying the transition from water to land.

2.3 Results

2.3.1 An unexpected diversity of myoglobins in *P. annectens*

We identified in the transcriptomes of the West African lungfish *P. annectens* four *Hb* α, five *Hb* β, one *GbX*, one *GbY*, and eight distinct *Mb* cDNA sequences (supplementary table SII.1, Supplementary Material online). No *Ngb* or *Cygb* sequences were detected in these data. Because no other known vertebrate showed such high diversity of *Mb*, these genes were subjected to closer investigations. The *Mb* cDNAs were verified by RT-PCR and sequencing. No additional *Mb* transcript was detected during these studies, suggesting that the *Mb* cDNA collection was complete, at least for the adult stage.

The *P. annectens* Mb sequences share 53.9% to 97.4% of the nucleotides and 32.4% to 96.5% of the amino acids (fig. II.1). Two Mb sequences were found highly similar (eleven different nucleotides resulting in five amino acid replacements) and could not be discriminated in qRT-PCR studies. These genes most likely reflect recent duplicates or alleles and have been named *PanMb6a* and *PanMb6b*. The *P. annectens* Mb amino acid sequences (142 to 154 amino acids) show the typical globin fold and possess the conserved globin sites PheCD1 (first amino acid in the interhelical region between helices C and D) and the proximal His at helix position F8 that binds the Fe^{2+}. The distal site at E7, which coordinates the bound O_2, is either occupied by the typical His (PanMb1-5, 7) or by a Gln (PanMb6a+b) (fig. II.1).

2.3.2 Early emergence and fast evolution of lungfish myoglobins

Database searches showed that seven Mb paralogs are apparently restricted to the lungfish. Bayesian phylogenetic analyses of vertebrate globin sequences confirmed this conclusion and showed a single origin of the lungfish Mbs (supplementary table SII.2, supplementary figs. SII.1 and SII.2, Supplementary Material online). The lungfish Mbs showed a notably higher evolution rate (fig. II.2; supplementary Fig. SII.2, Supplementary Material online). Most relative rate tests (Tajima 1993) significantly rejected the null hypothesis of equal rates between the lungfish Mbs and other Mb lineages (supplementary table SII.3, Supplementary Material online). Assuming that lungfishes and tetrapods diverged 411 million years ago (mya) (Hedges et al. 2015), we

calculated for the lungfish Mbs an average evolution rate of 2.77 x 10^{-9} substitutions per amino acid per year (with the LG model of amino acid evolution). Using this divergence time, we calculated for the tetrapod Mbs a rate of 1.24 x 10^{-9} substitutions per amino acid position per year. For the teleost Mbs, we assumed that the Actinopterygii (bony fish) and the Sarcopterygii (lobe-finned fish, including tetrapods) diverged ~430 mya (Hedges et al. 2015) and calculated a rate of 6.11 x 10^{-10} substitutions per amino acid per year. Thus, on average, the lungfish Mb proteins evolved ~2.2 times faster than those of the tetrapods and ~4.5 times faster than those of the teleosts. If we assume a lungfish substitution rate in the lungfish Mb stemline, the first divergence within the lungfish clade occurred ~268 mya; with the assumption of a tetrapod rate, the earliest date of divergence of the lungfish Mbs was only ~94 mya. Estimations of the dN/dS ratios provided no evidence for positive selection within the lungfish Mbs.

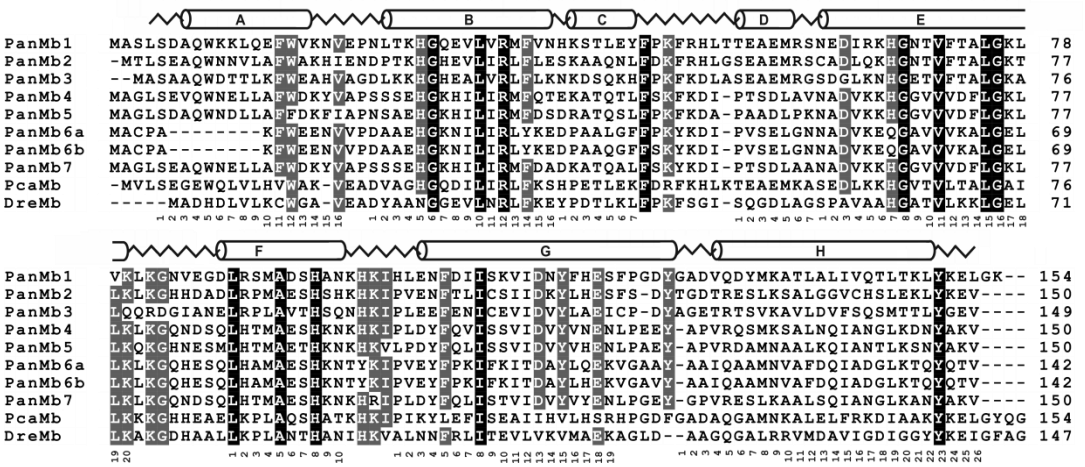


Figure II.1: Multiple sequence alignment of lungfish (Pan = *P. annectens*) Mbs with the Mbs of the sperm whale (PcaMb) and the zebrafish (DreMb).

The secondary structure of sperm whale myoglobin is superimposed in the upper row, with α-helices designated A through H. The consensus numbering of the globin α-helices (Dickerson and Geis 1983) is given below the sequences. Conserved residues are shaded (100% conservation: black; 75%: grey).

2.3.3 Differential, tissue-specific expression of lungfish myoglobins

qRT-PCR expression analyses with *PanMb1* to *7* were performed on a variety of lungfish tissues (fig. II.3). As mentioned, *PanMb6a* and *PanMb6b* could not be discriminated and were treated as a single transcript. Surprisingly, the brain showed the highest level of *Mb* mRNA expression. We found all seven globins expressed in this tissue (supplementary table SII.4, supplementary Fig. SII.4, Supplementary Material online). *PanMb2* was the highest expressed *Mb* in the brain (1.02 x 10^7 copies per µg total RNA), followed by *PanMb3* (9.59 x 10^6 copies/µg RNA) and *PanMb5* (7.50 x 10^6 copies/µg RNA). High levels of *PanMb* mRNA were also found in the eye,

which represents – at least in part – another neuronal tissue. Again, *PanMb2* had the highest copy numbers (3.38 x 10^6 copies/µg RNA), followed by *PanMb3* (2.56 x 10^6 copies/µg RNA). However, compared to the brain, the levels of *PanMb5* were low (2.12 x 10^5 copies/µg RNA).

In the skeletal muscle and in the heart, which are the typical sites of Mb expression in other vertebrates, *Mb* mRNA levels were comparably low. In the muscle, *PanMb1* had the highest copy numbers (2.49 x 10^6 copies/µg RNA), while all other *Mbs* (except *PanMb2*; 3.69 x 10^5 copies/µg RNA) had negligible numbers. In the heart, *PanMb3* had the highest level (6.13 x 10^5 copies/µg RNA), although it did not reach the expression levels as in the brain or the eye. *Mb* expression was also detected in the other tissues, but mRNA levels were low. Notable exceptions were *e.g. PanMb1* and *PanMb2* in the kidney (3.97 x 10^6 and 2.70 x10^5 copies/µg RNA, respectively), *PanMb2* in the gallbladder (5.51 x 10^5 copies/µg RNA), and *PanMb5* in the gonads (2.62 x 10^5 copies/µg RNA). The expression levels of *PanMb4*, *PanMb6* and *PanMb7* were mostly lower than those of the other *Mbs*. The highest level of *PanMb4* was found in the heart (1.93 x 10^5 copies/µg RNA), *PanMb6* peaked in the brain (6.04 x 10^5 copies/µg RNA), and *PanMb7* had its main site of expression in the eye (7.31 x 10^5 copies/µg RNA).

The *Mb* expression pattern was confirmed by the RNA-seq analysis of the three transcriptomes of *P. annectens*, deriving from the brain, a mixture of liver and gut, and a mixture of kidney and gonads (supplementary table SII.5 and Fig. SII.5, Supplementary Material online). Again, *PanMb2* was found the highest expressed *Mb*, which peaks in the brain. We further analyzed the levels of the other globins. Notable amounts were only found for PanGbY in the brain. The expression of the various Hb chains probably reflects the perfusion in the particular organs as the ratios between them were fairly constant. Expression of all four α and five β chains were detected; the α-chains *PanHba3* and *PanHba4* and the β-chains *PanHbb4* and *PanHbb5* reflect the major adult Hb chains.

We additionally identified the PanMb proteins in tissue samples that had been separated by SDS-PAGE (supplementary Fig. SII.7, Supplementary Material online) and 2D gel electrophoresis (supplementary figs. SII.8 and SII.9, Supplementary Material online). Mass spectrometry (supplementary table SII.6, Supplementary Material online) from bands of the SDS-PAGE showed the presence of PanMb1, 2, and 3 in brain, eye and heart. The muscle, liver and lung

samples contained peptides that were assigned to PanMb1 and 2. PanMb5 was only found in the brain. In the 2D gel, mass spectrometry (supplementary table SII.6, Supplementary Material online) detected PanMb1, 2, 3 and 5 in the brain; the muscle sample contained a fragment that either corresponds to PanMb1 or PanMb2. This protein pattern essentially mirrors the variation of the *PanMbs* mRNA levels in different tissues (fig. II.3), showing, for example, the broad expression of *PanMb1, 2* and *3*, and the brain-specificity of *PanMb5*.

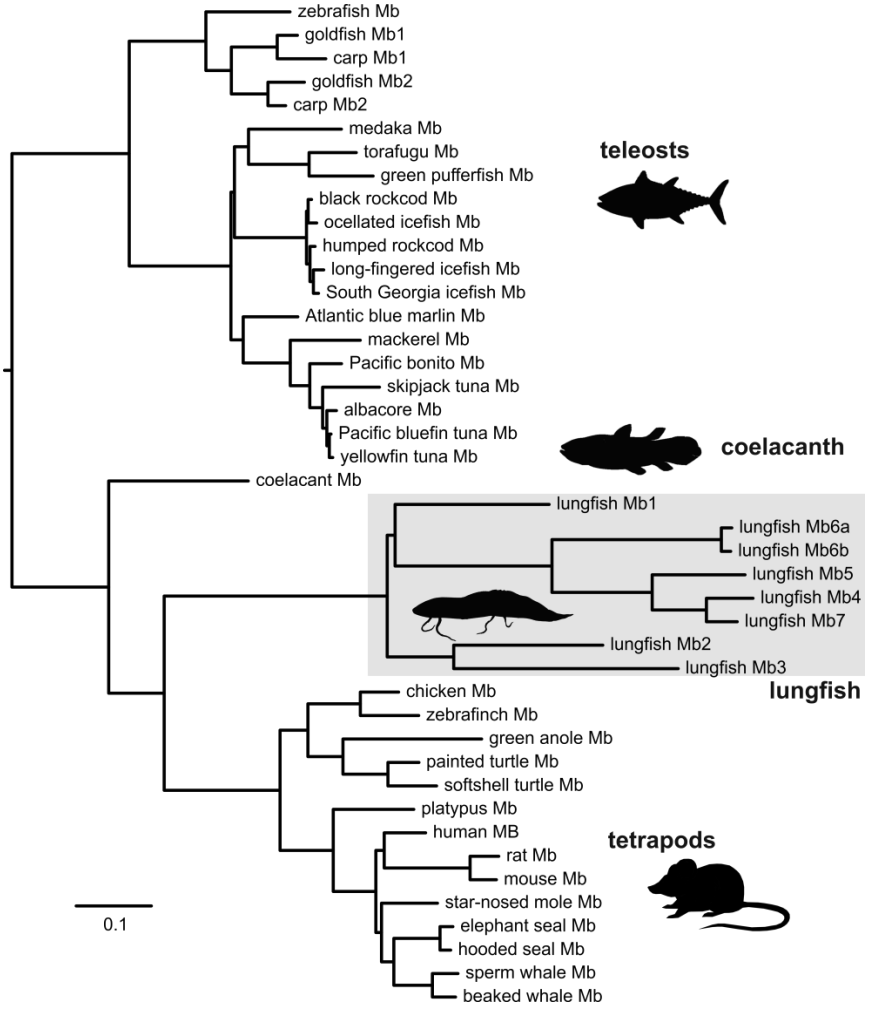


Figure II.2: Phylogenetic tree of vertebrate Mbs.

Tree reconstruction was carried out on the cDNA sequences with PAML assuming the GTR model. The bar represents 0.1 PAM distance. The common names of the species are given; for the accession numbers, see supplementary table SII.2, Supplementary Material online. Note the elevated evolutionary rate of lungfish Mb (shaded in grey).

2.3.4 Lungfish myoglobins protect cultured cells from hypoxia and reduce ROS production

To evaluate the function of the lungfish Mbs at the cellular level, the *PanMb1* – *PanMb6* cDNAs were transfected into the HN33 cell line. *PanMb7* could not be successfully cloned into the pCDNA3.1+ vector and was not considered further. The capability of the lungfish *Mbs* to protect

the cells from hypoxia was assessed by measuring the activity of the mitochondrial dehydrogenases and ATP levels after 24 h at 1% O_2 (fig. II.4; supplementary table SII.7, Supplementary Material online). The expression of Mb increased the activity levels of the mitochondrial dehydrogenases in the HN33 cells compared to the mock control, which only contained the empty vector. Transfection with *PanMb1* and *PanMb2* had the strongest effect, enhancing the cell viability > 2.5-fold. Hypoxia caused significant reduction of the ATP concentration in the mock control by 50%. These values were similar to those observed before (Koch and Burmester 2016). The ATP levels were found reduced in most *PanMb*-transfected cells; only *PanMb3* and *PanMb6* enhanced the ATP status by ~20%.

We evaluated the effect of the lungfish *Mbs* on the production of ROS under hypoxia using the DCFH-DA assay. In the mock control, ROS-production was found enhanced by ~50% compared to the normoxia control (supplementary table SII.7, Supplementary Material online). Most *PanMbs* considerably reduced the production of ROS (fig. II.4). Only for *PanMb4*, the effect was not significant.

2.4 Discussion

2.4.1 Ancient and specific amplification of myoglobin genes

Lungfish have a rich paleontological record, but there are only six extant species (Jørgensen and Joss 2010). Four species occur in Africa (*Protopterus aethiopicus*, *P. amphibius*, *P. annectens* and *P. dolloi*), the two other can be found in Australia (*Neoceratodus forsteri*) and South America (*Lepidosiren paradoxa*), respectively. Notably, lungfish have the largest genomes of animals, with C-value estimates ranging from ~40 to ~130 pg, which is much higher than in humans (3.5 pg) (Gregory 2016). Nevertheless, there is no evidence for that the *P. annectens* genome has experienced recent polyploidization (Vervoort 1980) or the amplification of globin genes other than *Mb*. Thus, the multiplication of the *Mb* genes was specific, and there is no correlation with the large lungfish genome.

The high number of *Mb* genes makes the lungfish unique among the vertebrates and requires physiological explanations. In principle, it is conceivable that the high number of *Mb* genes have additive effects on O_2 transport in muscle tissue. However, there is compelling evidence that the lungfish Mbs have evolved distinct functional properties: Firstly, the Mbs are highly

diverged (up to 67.6% different amino acids; fig. II.1 and fig. II.2) and the distinct *Mb* genes have been retained in evolution since they commenced to diversify 94 to 268 mya. Secondly, the Mb mRNA and proteins are differentially expressed in the various tissues (fig. II.3; (supplementary figs. SII.8 and SII.9, supplementary table SII.6, Supplementary Material online), supporting distinct functions. Thirdly, the *Mbs* have different capabilities to enhance the survival of a neuronal cell line under hypoxia (fig. II.4). The evolutionary mechanisms that underlay the functional diversification of the *Mb* genes must remain uncertain. Probably the diversity of *Mb* genes in the lungfish emerged from the ancestral *Mb* gene by several subfunctionalization and/or neofunctionalization events.

2.4.2 High myoglobin concentrations may protect the lungfish brain from hypoxia during aestivation

African lungfishes are obligate air breathers that typically dwell in freshwaters with low dissolved O_2 and high daytime temperatures. The vertebrate brain is particularly sensitive towards a reduction of O_2 supply. Nevertheless, notable *Mb* expression in the brain – as found in *P. annectens* – is rather an unusual feature, and has previously been – to the best of our knowledge – only observed in the carp (Fraser et al. 2006). The total *PanMb* mRNA levels in the lungfish brain amounts to 2.9 x 10^7 copies per µg total RNA, which is in the same range as Mb in the fish heart (~5 x 10^7 to 5 x 10^8) (Tiedke et al. 2014). This indicates an important function of Mb in the lungfish brain, probably enhancing O_2 supply to the nervous tissues when the ambient O_2 concentration is low.

Lungfish survive periods of seasonal drying aestivating in a mud cocoon (Ballantyne and Frick 2010). During this period, lungfish reduce energy metabolism and turn on cytoprotective gene expression (Hiong et al. 2013). It is conceivable that Mb is a component of these protective mechanisms. For example, the various Mbs may protect the tissues from ROS, as evident from the cell culture experiments (fig. II.4).

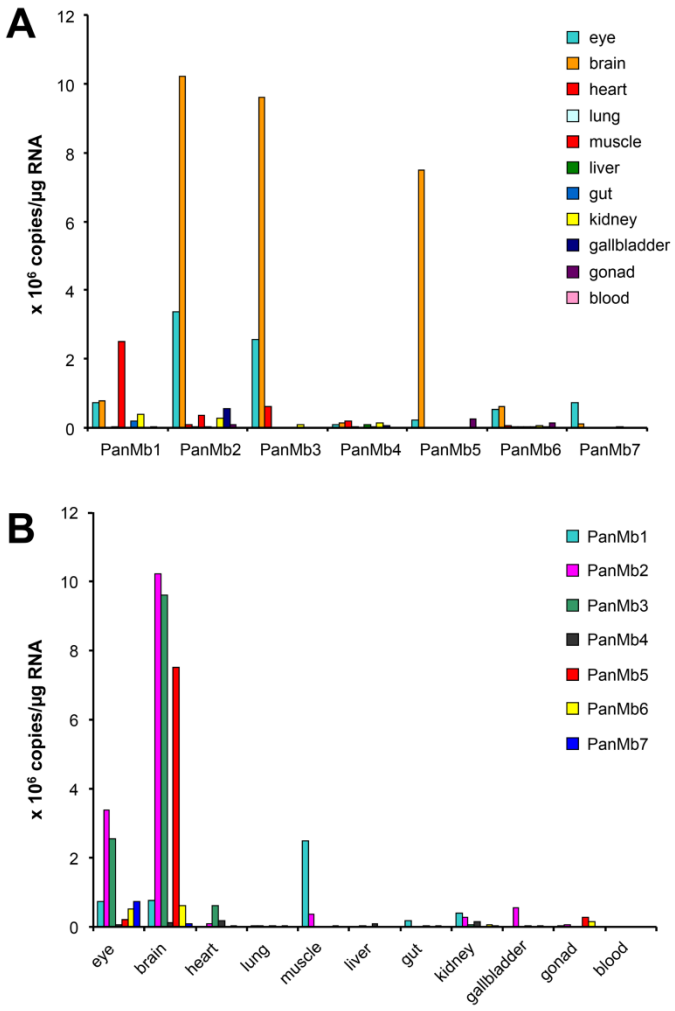


Figure II.3: Expression of *Mb* mRNA in selected *P. annectens* tissue samples, as estimated by qRT-PCR.

(A) Tissue-specificity of *PanMb*; *PanMb2*, *3* and *5* showed the highest expression levels, mainly in the brain. (B) Expression of *PanMb* per tissue. The brain expresses most of the *P. annectens* Mbs. The copy numbers are given in supplementary table SII.4, Supplementary Material online. Log-scale data are presented in log scale in supplementary fig. SII.4, Supplementary Material online. RNA-seq analyses in brain, liver + gut, and kidney + gonad tissues are given in supplementary fig. SII.5, Supplementary Material online.

2.4.3 Different myoglobins with distinct roles in skeletal muscle and heart

Mb is best known for its prominent role in the supply of O_2 to the mitochondria of the skeletal and the heart muscles (Wittenberg and Wittenberg 1989, Wittenberg and Wittenberg 2003). The single *Mb* gene in the lungfish stem-lineage had most likely a similar function. The total *PanMb* copy numbers in the *P. annectens* skeletal muscle (2.9 x 10^6 copies per µg total RNA) and heart (9.4 x 10^5 copies per µg RNA) were low and a tenth or less compared to the brain (fig. II.3; supplementary table SII.4, Supplementary Material online). Thus, in the lungfish, paradoxically the brain is the main site of *Mb* mRNA expression, not the muscles.

Notably, the skeletal muscles and the heart express distinct sets of *PanMbs*. In the skeletal muscle, *PanMb1* is the highest expressed *Mb* gene (2.5 x 10^6 copies per µg RNA), which was not detected in the heart. By contrast, the main *Mb* in the heart is *PanMb3* (6.1 x 10^5 copies per µg RNA), which is only marginally expressed in the skeletal muscle (642 copies per µg RNA). Both tissues require an adequate supply with O_2. The cell culture experiments showed that PanMb1 and PanMb3 have different biochemical properties: PanMb1 increases the activity of mitochondrial dehydrogenases of the cultured cells, which essentially reflects the number of living cells; PanMb3 enhances the ATP level and thus the energy status of the cells (fig. II.4). This may reflect different requirements of the skeletal muscle and the heart.

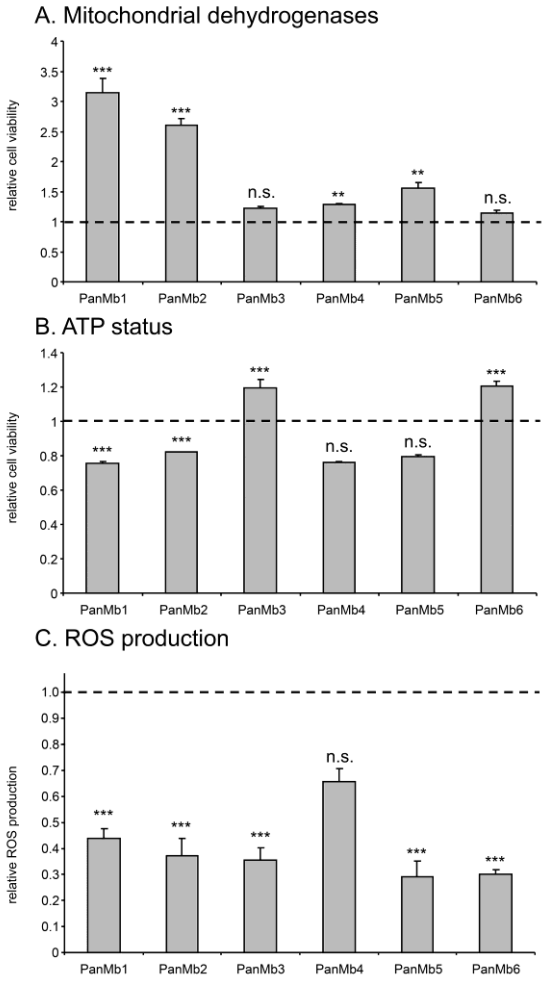


Figure II.4: Effect of lungfish Mb on the viability of HN33 cell viability (A and B) and on ROS production (C) after hypoxia treatment (1% O_2) for 24 h (n=3 each).

The cells were transfected with PanMb1 – 6a. Cell viability was measured by the relative activity of the mitochondrial dehydrogenases (A) and ATP content (B), ROS production with the DCFH-DA test (C). The values were normalized according to the mock controls (vector control; dashed line). The bars show the standard deviations (SD). Asterisks indicate significance levels at $p < 0.01$ (**) and $p < 0.001$ (***).; n.s., not significant ($p < 0.05$ were considered as significant). Asterisks at the bars indicate statistically significant differences between the globin-containing cell line and the mock control.

2.4.4 Conclusion: Myoglobin diversity in lungfish as an example for the functional diversification of globin genes

The presence of seven Mbs in the West African lungfish is unusual and unique among vertebrates. Although additive effects of the expression of multiple *Mb* genes may be important for certain tissues to increase O_2 supply, most likely they carry out distinct functions, as demonstrated for PanMb1 and PanMb3 in the striated muscles. As mentioned, Mb may also decompose NO (Flögel et al. 2001) and ROS (George and Irvine 1951, Osawa and Korzekwa 1991, Flögel et al. 2004), or may transport fatty acids (Shih et al. 2014). It is conceivable that these tasks have been distributed among the various lungfish Mbs.

There is no evidence from the transcriptomes that *P. annectens* harbors *Ngb* or *Cygb* genes. Although Ngb is a highly conserved protein that has an ancient origin that predates the divergence of Protostomia and Deuterostomia (Burmester and Hankeln 2014), it has been lost in certain vertebrate taxa, *i.e.* lampreys (Schwarze et al. 2014) and sharks (Opazo et al. 2015). Cygb has been found in all vertebrates investigated so far; thus, the lungfish may be the first known example the Cygb is absent. Although the functions of Ngb and Cygb are still a matter of debate (Hankeln et al. 2005, Burmester and Hankeln 2009, Burmester and Hankeln 2014), it is conceivable that some *Mb* genes have taken over their roles in the lungfish, thereby explaining, in part, their present diversity.

2.5 Materials and Methods

2.5.1 Transcriptome and database analyses

The globin genes of the West African lungfish *P. annectens* were identified in the transcriptomes from the brain, liver + gut, and kidney + gonads, which have been generated by Illumina sequencing at the Broad Insitute (SRA accession numbers: SRX152529 - SRX152531) (supplementary table SII.1, Supplementary Material online). Assemblies provided by the Broad Insitute and own assemblies generated with the CLC-Genomics Workbench (version 7.5) were used. The coelacanth (Schwarze and Burmester 2013) and turtle (Schwarze et al. 2015) globin sequences were employed to search for globins with BLAST algorithm. If required, the cDNA sequences were assembled by hand or with the aid of the CLC-Genomics Workbench. Protein sequences were derived by translation with the tool provided by the ExPASy Molecular Biology Server (http://www.expasy.org). The *Mb* sequences were verified by RT-PCR and sequencing,

and, in some cases, by RACE (rapid amplification of cDNA ends; see below). The *P. annectens* Mb coding sequences have been deposited at EMBL/GenBank under the accession numbers XXXX – YYYY.

The proportion of transcripts in the Illumina samples was estimated with the RNA-seq tool of CLC Genomics Workbench 7.5. The globin cDNAs were used as reference sequences (mapping options: mismatch cost: 2, insert cost: 3, deletion cost: 3, length fraction: 0.95, similarity fraction: 0.95). Each RNA-Seq dataset was mapped against the reference sequences and normalized according to the transcript length and the size of the dataset. Expression levels were calculated as Reads Per Kilobase of transcript per Million reads (RPKM). Due to the small reference dataset, the highest globin expression level (*PanHba4* in the liver sample for the *Hb* transcripts; *PanMb2* in the brain sample for all other globins) was set to 1 arbitrary unit (AU). The other RPKM values were related to this reference.

2.5.2 Phylogenetic analyses

The *P. annectens* globin sequences were included in a previously used collection of vertebrate globin sequences (Schwarze et al. 2015) (supplementary table SII.2, Supplementary Material online). Multiple sequence alignments of the amino acid sequences were obtained employing MAFFT with the G-INS-i method (Katoh and Toh 2008, Katoh et al. 2009). The appropriate model of amino acid evolution was selected with PROTTEST (Abascal et al. 2005). Phylogenetic analysis was performed with MrBayes 3.2.3 (Huelsenbeck and Ronquist 2001, Ayres et al. 2012) using the LG model (Le and Gascuel 2008), which was implemented into MrBayes (Schwarze and Burmester 2013, Schwarze et al. 2015). Two independent runs of 5,000,000 generations with four simultaneous chains each were performed. The trees were sampled every 1000th generation. The final average standard deviation of split frequencies was < 0.02. The posterior probabilities were estimated after discarding the initial 25% of the trees.

For analysis of Mb evolution, wwe selected the appropriate Mb sequences from a broad range of vertebrate taxa (supplementary table SII.2, Supplementary Material online). The topology within the *P. annnectens* Mbs resulted from the MrBayes tree (supplementary Fig. SII.1, Supplementary Material online), the topology of the Mbs of the other taxa was derived from the known vertebrate phylogeny (Song et al. 2012, Amemiya et al. 2013, Betancur et al. 2013).

The branch lengths were estimated using the programs baseml and codeml from the PAML package (Yang 2007). The GTR model (Yang 1994) was assumed for the nucleotide sequences, the LG model for the protein sequences. Substitution rates were estimated under the assumption that the clades leading to lungfishes and tetrapods split ~411 million years ago (mya) and that Actinopterygii and Sarcopterygii diverged ~430 mya (dates from http://www.timetree.org (Hedges et al. 2015). dN and dS were estimated with codeml from PAML. Relative rate tests (Tajima 1993) were performed with the MEGA package (Tamura et al. 2013).

2.5.3 Tissue preparation, RNA extraction cDNA cloning

Two *Protopterus annectens* specimens were obtained from a pet shop and immediately euthanized in 1 g/l tricaine methanesulfonate. Animals were treated in accordance with the German Animal Welfare Act. Selected tissues (brain, eye, heart, muscle, lung, gut, liver, kidney, gonads, gallbladder and blood) were removed and stored frozen at -80 °C in RNAlater (Qiagen, Hilden, Germany) until further use.

Total RNA from the individual tissues was extracted employing the Crystal RNA Mini Kit (Biolab Products, Gödenstorf, Germany) according to the manufacturer's instructions. The quality and integrity of total RNA were evaluated by formaldehyde agarose gel electrophoresis. RNA concentration was measured with the Nanodrop ND 100 UV-Vis spectrometer (Thermo Scientific, Bonn, Germany). 1 µg total RNA was used for cDNA synthesis employing the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific, Bonn, Germany) following the instructions of the manufacturer. Gene-specific oligonucleotides (supplementary table SII.8, Supplementary Material online) were used for the amplification of the individual *Mb* cDNAs by PCR. The cDNAs were cloned into a pGEM-T (Promega, Mannheim, Germany) or pcDNA3.1+ (Invitrogen, Karlsruhe, Germany) vector and sequenced by a commercial service (GATC, Konstanz, Germany).

2.5.4 Quantitative real-time reverse transcription PCR

Expression levels of *Mb* mRNAs in different lungfish tissues were estimated by quantitative real-time reverse transcription PCR (qRT-PCR). Standard plasmids were generated from the *P. annectens Mb* coding sequences, which had been amplified from the cDNA as described

above. qRT-PCR amplification was performed on an ABI Prism 7500 Real Time PCR System (Applied Biosystems, Darmstadt, Germany). qRT-PCR reactions were performed in triplicates in 20 µl including the Power SYBR-Green PCR Master Mix, 3 µl of cDNA and primers (supplementary table SII.8, Supplementary Material online) (final concentration 0.2 mM). Amplification was carried out using a standard PCR protocol. Dissociation curve analysis was used to validate the specificity of each amplification reaction. Absolute mRNA copies were calculated with the 7500 System Sequence Detection Software 2.0.6 (Applied Biosystems) using the standard curve method by using dilutions (10^7 to 10^2) of the recombinant plasmid.

2.5.5 Analyses of myoglobin proteins

Total protein was isolated from small pieces of heart, muscle, brain, lung, liver and eye. Tissues were homogenized by sonification (Bandelin Sonopuls, Berlin, Germany) in 1 x Complete Protease Inhibitor (Roche, Basel, Switzerland) in 1 x PBS (0.14 M NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, 0,15 mM KH_2PO_4). After centrifugation at maximum speed (20 min, 4° C), the supernatant was collected. Protein concentration was measured with the Pierce™ BCA Protein Assay Kit (Thermo Scientific, Waltham, MA USA) according to the manufacturer's instructions. 150 µg protein from each tissue were separated on a 15% SDS-polyacrylamide gel. The gel was stained with Coomassie brilliant blue.

For 2D gel electrophoresis, 250 µg total protein from brain and muscle were applied. Rehydratisation of the Immobiline™ DryStrips pH 3-10NL, 7 cm (GE Healthcare, Freiburg, Germany) were performed over night at 18°C with rehydratization buffer (8 M urea, 2 M thiourea, 1% Chaps, 19.4 M dithiothreitol, 0.5 % pharmalyte, Bromophenol blue). Isoelectric focusing was performed with an Ettan IPGphor3 IEF System (GE Healthcare) with a 4 step protocol (step 1: 300 V, 1000 Vh, step 2: 1000 V, 300Vh, step 3: 5000 V, 4500 Vh, step 4: 5000 V, 2000 Vh). Afterwards, the IPG strips were equilibrated for 15 min in 50 mM Tris-HCl pH 8.8, 6 M urea, 30% glycerol, 4% sodium dodecyl sulfate, 20 mM dithiothreitol, followed by 15 min in 50 mM Tris-HCl pH 8.8, 6 M urea, 30% glycerol, 4% SDS, 240 mM iodoacetamide. A 12% SDS-polyacrylamide gel was used as the second dimension. Gels were stained with Coomassie brilliant blue.

Putative Mb bands and spots were excised and analyzed by mass spectrometry using a commercial service (Core Facility Mass Spectrometric Proteomics, University Medical Center Hamburg-Eppendorf, Germany).

2.5.6 Cell culture experiments

The lungfish Mb cDNA were cloned into the pcDNA3.1+ vector and transfected into HN33 cells using Nanofection (PAA, Pasching, Austria). Cells transfected with an empty vector served as controls. The cell were grown in Dulbecco's modified Eagle medium (BioWest, Nuaillé, France) containing 10% fetal bovine serum (BioWest, Nuaillé, France), 1% penicillin/streptomycin (BioWest) and 700 µg/ml geneticin (PAA) in a humidified atmosphere at 37° C with 5% CO_2. Gasses were obtained from Linde (Munich, Germany). The success of the transfection and the *PanMb* expression levels were verified by qRT-PCR. Hypoxia experiments were carried out in a CB150 incubator (Binder, Tuttlingen, Germany) with cells (passage number < 35) at 1% O_2, 5% CO_2 and 94% N_2 for 24 h. All experiments were carried out with three technical and three biological replicates with 5 x 10^4 cells per well in a 96-well plate.

Cell viability was assayed by the WST-1 test (Roche, Mannheim, Germany), which measures the activity of the mitochondrial dehydrogenases, and by the CellTiter-Glo® Luminescent Cell Viability Assay (Promega, Mannheim, Germany), which measures the ATP content of the cells. Both tests were carried out according to the manufacturers' instructions in 96 well plates with 5 x 104 cells per well. Absorbance at 450 nm (WST-1) and luminescence (CellTiter-Glo®) were determined with a DTX-880 Multimode Plate Reader (Beckman Coulter, Krefeld). The production of ROS assayed by the 2',7'-dichlorodihydrofluorescein-diacetate (DCFH-DA) method, which measures the ROS-induced oxidation of the non-fluorescent DCFH to the fluorescent DCF (Wang and Joseph 1999). Detection of DCF was carried out with the DTX-880 Multimode Plate Reader using an excitation wavelength of 485 nm and an emission wavelength of 535 nm. Statistical analysis was carried out with IBM SPSS Statistics 20 employing one-way ANOVA with Bonferroni correction. P-values < 0.05 were considered as significant.

2.6 Acknowledgments

This project is supported by the Deutsche Forschungsgemeinschaft (Bu956/19). We thank Ceyda Cubuk (University of Hamburg) for her help with the dissection of the animals, Andrej Fabrizius (University of Hamburg) for his help with the RNA-seq analyses and Pascal Steffen

(Core Facility Mass Spectrometric Proteomics, University Medical Center Hamburg-Eppendorf) for the mass spectrometry.

III. Chapter 2:

Genetic and functional diversity of the multiple lungfish myoglobins

Julia Lüdemann[1], Angela Fago[2], Sven Falke[3], Michelle Wisniewsky[1], Igor Schneider[4], Andrej Fabrizius[1] and Thorsten Burmester[1]

[1]Institute of Zoology, Department of Biology, University of Hamburg, D-20146 Hamburg, Germany

[2]Department of Bioscience, Aarhus University, DK-8000 Aarhus C, Denmark

[3]Institute for Biochemistry and Molecular Biology, Department of Chemistry, University of Hamburg, D-20146 Hamburg, Germany

[4]Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, PA, Brazil

In review in *The FEBS Journal*

Supplementary Material can be found on the enclosed CD-ROM

3.1 Abstract

It is known that the West African lungfish (*Protopterus annectens*) harbours multiple myoglobin genes that are differentially expressed in various tissues and that the myoglobins differ in their abilities to confer tolerance towards hypoxia. Here we show that other lungfish species (*Protopterus dolloi*, *Protopterus aethiopicus*, and *Lepidosiren paradoxa*) display a similar diversity of myoglobin genes and have orthologous myoglobin genes. To investigate the functional diversification of these genes, we studied the structures, O_2 binding properties and nitrite reductase enzymatic activities of recombinantly expressed *P. annectens* myoglobins (PanMbs). Circular dichroism spectroscopy and small-angle X-ray scattering revealed the typical globin-fold in all investigated recombinant myoglobins, indicating a conserved structure. The highest O_2 affinity was measured for PanMb2 (P_{50} = 0.88 Torr at 20 °C), which is mainly expressed in the brain, whereas the muscle-specific PanMb1 has the lowest O_2 affinity (P_{50} = 3.78 Torr at 20 °C), suggesting that tissue-specific O_2 requirements have resulted in the emergence of distinct myoglobin types. Two of the mainly neuronally expressed myoglobins (PanMb3 and PanMb4b) have the highest nitrite reductase rates. These data show different O_2 binding and enzymatic properties of lungfish myoglobins, reflecting multiple subfunctionalisation and neofunctionalisation events that occurred early in the evolution of lungfish. Some myoglobins may have also taken over the functions of neuroglobin and cytoglobin, which are widely expressed in vertebrates but appear to be missing in lungfish.

3.2 Introduction

Lungfish (Dipnoi) are of particular scientific interest because they are considered as "living fossils" and the closest living relatives of tetrapods (Brinkmann et al. 2004, Amemiya et al. 2013, Biscotti et al. 2016). Moreover, lungfish harbour various specific morphological, physiological and molecular adaptations that allow them to tolerate hypoxic environments (Jørgensen and Joss 2010). Recent studies identified eight distinct myoglobin (Mb) genes in the West African lungfish (*Protopterus annectens*) (Koch et al. 2016). This diversity is unusual because most vertebrates harbour a single Mb in the haploid genomes (Burmester and Hankeln 2014).

Mb is a member of the superfamily of globins, which is a classic model system for the study of the evolution and function of proteins and genes (Hardison 1996a, Graur 2000, Gillemans et al. 2003, Vinogradov et al. 2007). Globins are small, globular proteins that are mainly responsible for the binding of oxygen (O_2), but also may have other functions. Globins have a Fe^{2+} ion in the centre of their porphyrin prosthetic group and share the unique structure that typically consists of eight alpha helices. In addition to the well-known Mb and Hb, vertebrates harbour six other globin-types, which commenced to diversify even before the radiation of vertebrates (Burmester and Hankeln 2014). The globin-types are: neuroglobin (Ngb) (Burmester et al. 2000), cytoglobin (Cygb) (Kawada et al. 2001, Burmester et al. 2002, Trent and Hargrove 2002), globin X (GbX), globin Y (GbY), globin E (GbE) (Kugelstadt et al. 2004, Blank et al. 2011a) and androglobin (Adgb) (Hoogewijs et al. 2012). However, the functions of these recently identified globins are mostly not well understood (Hankeln et al. 2005, Burmester and Hankeln 2009, Burmester and Hankeln 2014).

Mb is a monomeric protein consisting of ~150 amino acids (~17 kDa). It was the first protein with a resolved protein structure at an atomic scale (Kendrew et al. 1958, Kendrew 1963). The main function of Mb is the transport and storage of O_2 in striated (skeletal and heart) muscles (Merx et al. 2001, Wittenberg and Wittenberg 2003, Helbo et al. 2013). For the effective extraction of oxygen from the blood the typical Mb has a higher O_2 affinity, i.e. lower P_{50} value (partial pressure at which the protein is 50 % saturated with O_2), than haemoglobin (Hb). Most mammalian and avian Mbs have a P_{50} of approximately 1 Torr at 20 °C (Antonini and Brunori 1971, Enoki et al. 1995, Enoki et al. 2008). In contrast, Mbs of fish have P_{50} values between 1 Torr (e.g. bluefin tuna, blue marlin, and common carp (Madden et al. 2004, Helbo et al. 2012) and 4.9 Torr (rainbow trout) (Pedersen et al. 2010, Helbo and Fago 2011). This variation can be related to structural differences: fish Mbs are usually shorter (~ 146 aa) than mammalian Mbs

and lack the D-helix (Helbo et al. 2013). Remarkably, Mb O_2 affinity, unlike Hb O_2 affinity, does not correlate with the size of the individual (Schmidt-Nielsen 1990).

In the past 20 years, it has become evident that Mb also exhibits enzymatic functions. For example, oxygenated Mb rapidly converts the bioactive nitric oxide (NO) signalling molecule into nitrate (NO3-) under normoxic conditions (Flogel et al. 2010). The removal of NO enhances mitochondrial respiration in normoxia as NO acts as a reversible inhibitor of cytochrome oxidase (Moncada and Erusalimsky 2002). While reducing NO, oxy-Mb is oxidised to met-Mb that needs to be reduced by the cellular Mb reductase (George and Irvine 1951, Osawa and Korzekwa 1991, Flögel et al. 2004). Furthermore, under hypoxia, Mb has the opposite effect whereby deoxy-Mb functions as nitrite reductase producing NO from nitrite (NO_{2^-}) (Hendgen-Cotta et al. 2008), thereby limiting O_2 consumption rates (Flogel et al. 2010, Helbo et al. 2013, Fago 2017). It has also been suggested that Mb functions in oxidative defence by removing reactive oxygen species (ROS) (George and Irvine 1951, Osawa and Korzekwa 1991, Flögel et al. 2004). Additional functions of Mb may involve an interaction with fatty acids (Shih et al. 2014).

Most vertebrates harbour a single *Mb* gene in their haploid genomes (Burmester and Hankeln 2014). Known exceptions are some cyprinid fish, which harbour two distinct *Mb* genes (Fraser et al. 2006, Roesner et al. 2008). While Mb1 is widely expressed in various tissues, the expression of Mb2 is restricted to the brain (Fraser et al. 2006, Cossins et al. 2009). These two Mbs apparently have distinct primary functions: In vitro studies have suggested that Mb1 mainly supplies O_2 and produces NO, whereas Mb2 efficiently eliminates H_2O_2 in the brain (Helbo et al. 2012). Some vertebrate species lack an Mb gene, such as frogs, some ice-fishes, the stickleback and the opossum *Monodelphis domestica* (Maeda and Fitch 1982, Fuchs et al. 2006, Sidell and O'Brien 2006, Hoffmann et al. 2011). These species must have specific mechanisms to compensate for the lack of Mb. Especially in teleost fish the loss of cardiac Mb evolved repeatedly during their evolution. The Mb deficit is in connection with a pale heart colour and affects water- and air-breathing fish species from all salinities and habitats (Macqueen et al. 2014).

Using quantitative qRT-PCR and RNA-Seq, we found that the multiple *Mb* genes of the West African lungfish (PanMbs) are differentially expressed in tissues (Koch et al. 2016). Surprisingly, most of the PanMbs have the highest expression levels in the brain. Further, we found that striated muscles express different sets of PanMbs: PanMb1 was highest in the skeletal muscle

and PanMb4a in the heart. The unexpected diversity of the Mbs in the West African lungfish hinted to a functional diversification by subfunctionalisation and/or neofunctionalisation. Some PanMbs may have taken over the functions of Ngb and Cygb, which could not be detected in the lungfish transcriptomes. To better understand the evolution of *Mb* genes, we have studied their diversification in four lungfish species. Further, the functional differentiation was accessed by oxygen-binding, enzymatic, and structural studies.

3.3 Material and Methods

3.3.1 Identification and analyses of lungfish Mb cDNA sequences

The transcriptomes of the West African lungfish *P. annectens*, the slender lungfish *Protopterus dolloi*, the marbled lungfish *Protopterus aethiopicus*, and the South American lungfish *Lepidosiren paradoxa* were retrieved from the public SRA database at GenBank (for accession numbers, see Table SIII.1). The transcriptomes of each lungfish species were assembled using the CLC Genomics Workbench, version 11.0.1 (Qiagen, Hilden, Germany) using default parameters. Mb cDNA sequences were identified employing BLAST searches, using the Mb cDNA sequences from *P. annectens* (Koch et al. 2016) as queries. When required, the putative Mb sequences were re-assembled from the Illumina reads. RNA-Seq analyses were performed with the CLC Genomics Workbench (mismatch cost: 2, insertion cost: 3, deletion cost: 3, length fraction: 0.95, similarity fraction: 0.95). The mRNA levels of the myoglobins were then calculated as RPKM (reads per kilobase million; Figs. SIII.2 to SIII.5).

The Mb amino acid sequences were included in an alignment of vertebrate Mbs used before (Koch et al. 2016) (Table SIII.2). A multiple sequence alignment was obtained with MAFFT using the L-INS-i method (Katoh and Toh 2008, Katoh et al. 2009). Phylogenetic analysis was performed with MrBayes 3.2.3 (Huelsenbeck and Ronquist 2001, Ayres et al. 2012) using the JTT model (Jones et al. 1992), which was selected by a PROTTEST analysis (Katoh et al. 2005, Katoh et al. 2009). MrBayes was run for 5,000,000 generations with two independent runs and four simultaneous chains. The trees were sampled every 1000th generation, and the posterior probabilities were estimated after discarding the initial 25% of the trees as burnin. The Maximum Likelihood Analysis was carried out using the IQ-TREE web server with the JTT substitution model (Nguyen et al. 2015, Trifinopoulos et al. 2016, Hoang et al. 2018). 1000 bootstrap alignments were run with a maximum of 1000 iterations and a minimum correlation

coefficient of 0.99. The perturbation strength was 0.5 and the IQ-TREE stopping rule 100. The values written on the branches show the ultrafast bootstrap support (Fig. III.1).

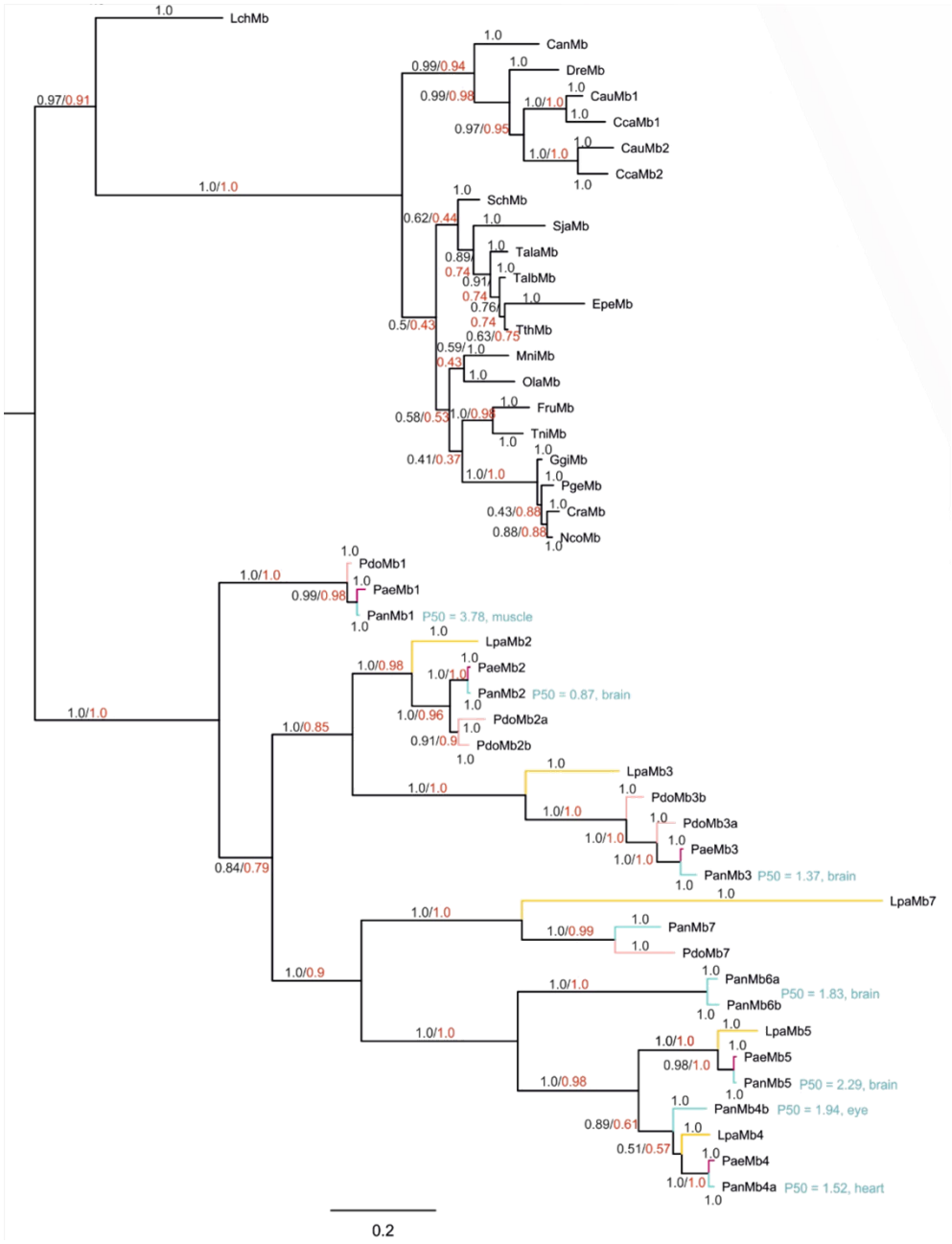


Figure III.1: Bayesian phylogenetic tree of vertebrate Mbs.

Tree reconstruction was carried out with the amino sequences assuming the JTT model of protein evolution (Sick and Gersonde 1969). The bar represents 0.2 substitutions per site. The numbers at the nodes are the Bayesian posterior probabilities (black) and the ultrafast bootstrap support (red). For the abbreviations and accession numbers, see Table SIII.1. For the PanMbs, the P_{50}-values (in Torr) and the main expression sites are given. (blue: *P. annectens* Mbs; pink: *P. aethiopicus* Mbs; light red: *P. dolloi* Mbs; yellow: *L. paradoxa* Mbs)

3.3.2 Recombinant expression and purification of P. annectens Mbs

Tissue preparation, RNA extraction, and cDNA cloning of *P. annectens* Mbs were performed as described before (Koch et al. 2016). The cDNAs were cloned into pET16b using the restriction sites for correct sequence orientation (Novagen - Merck Biosciences, Darmstadt, Germany). The vectors were then transfected into E. coli BL21(DE3) pLysS for recombinant expression. *E. coli* were grown overnight at 37°C in a 5 ml L-Medium (1% bactotryptone, 0.5% yeast extract, 0.5% NaCl, pH 7.5), supplemented with 10 µg/ml ampicillin and 34 µg/ml chloramphenicol. Subsequently, the culture was transferred to 1 l L-medium supplemented with 1 ml of 1M 5-aminolevulinic acid and cells were grown at 37°C until an OD600 between 0.4 and 0.8 was reached. The expression was then induced by 0.4 mM isopropyl-1-thio-β-D-galactopyranoside and the culture was incubated for an additional 3 hours at 37°C. The cell suspension was centrifuged at 4000 × g and 4°C for 45 min and the pellet with the bacterial cells was dissolved in Protease Inhibitor Buffer (50 mM Tris-HCl, 1 mM dithiothreitol, 1 mM Pefabloc, 1 mM $MgCl_2$, 1x CompleteTM proteinase inhibitor mix (Roche Applied Science, Penzberg, Germany), 80 µg DNase, RNase, pH 8.0). After exposing the cells to three freeze-thaw steps in liquid nitrogen and disruption by sonification (10 × 30 s), the cells were incubated at 37°C for 2 hours to digest the DNA and RNA and centrifuged for 1 h at 4500 × g and 4°C. The Mb proteins were purified from the supernatant on a His60-Ni-column (Qiagen) according to the manufacturer's instructions. The histidine-tag was removed by incubation with the Factor Xa protease (20 µg/ml) for 6 h at 37 °C. After inactivation of the protease with 2 µM dansyl-Glu-Gly-Arg-chloromethyl-ketone, the recombinant Mb proteins were transferred to 10 mM HEPES, pH 7.8, 0.5 mM EDTA and stored at 4°C until further analysis.

3.3.3 O_2 equilibrium measurements

A modified diffusion chamber technique was used to determine the O_2 equilibrium curves as described earlier (Sick and Gersonde 1969, Weber 1981, Weber 1992, Madden et al. 2004). Samples of 5 µl (~0,2 mM heme) in 0.1 M HEPES, 0.5 mM EDTA, pH 7.2 at 20 °C were measured in at least duplicates. Before measuring, the ferric lungfish Mbs were reduced to the ferrous form under N_2 for 10 min at room temperature using a met-Hb reductase system (Hayashi et al. 1973, Fago et al. 2004a) consisting of glucose 6-phosphate (677 µM), glucose 6-phosphate dehydrogenase (0.33 µg/ml), NADP (45 µM), ferredoxin (0.17 µM), ferredoxin-NADP reductase

(0.77 µM), catalase (7 nM). All chemicals were supplied by Sigma-Aldrich (Brøndbyvester, Denmark).

The GMS 500 gas mixing system (Loligo System, Denmark) generated a water-saturated gas mixture of ultrapure (>99.998%) N_2 and O_2, which was further used to equilibrate the thin smear of Mb solution by stepwise increasing of the O_2 tension (PO_2). A photomultiplier (model RCA 931-A) and an Eppendorf model 1100 M photometer recorded the absorbance changes continuously at 436 nm to derive saturation at each PO_2 step. The absorption signal was recorded and analysed using the in-house developed software (Specktrosampler, K. Beedholm, Aarhus Univ. Denmark) (Jendroszek et al. 2018). At the beginning and the end of each experiment 0 and 100% O_2 saturation levels were obtained as reference. P_{50} and cooperativity values were calculated from the zero intercept (log(Y/(1-Y)) = 0) and slope of Hill plots, respectively: log(Y/(1-Y)) versus logPO_2, where Y is the fractional saturation. Four to six saturation steps were taken for each curve.

3.3.4 Gel-filtration experiments

The apparent molecular weight of PanMbs was determined by gel-filtration using an Äkta Pure chromatographic system (GE Healthcare, Freiburg, Germany) together with a Superdex 75 10/300 (GE Healthcare) column. A flow rate of 1 ml/min and an elution buffer containing 50 mM potassium phosphate, 0.5 mM EDTA, pH 7.0, 0.15 M NaCl were applied. The absorbance was measured at 280 nm (proteins) and 415 nm (heme-group). Sample solutions (100 µl each) with a concentration of 0.2 mM heme were loaded on the system; human Hb and horse Mb were used as standard at the same concentration and volume as the samples.

3.3.5 Nitrite reductase activity

After complete deoxygenation of the PanMbs (10 µM heme in deoxygenated 50 mM HEPES, pH 7.13) by anaerobically titrating with 1 mM of sodium dithionite in a 1 cm sealed cuvette, sodium nitrite (0.1 mM/cuvette) of freshly made stock solutions was added anaerobically with a Hamilton gastight syringe (Bonaduz, Switzerland) (Fago et al. 2017). The reaction was carried out at 20°C and under pseudo-first-order conditions, i.e. with substrate nitrite in excess of protein. The measurement of the reaction kinetics was started immediately, and absorbance was recorded at different time intervals at 435 nm and 20°C.

3.3.6 Dynamic light scattering (DLS)

DLS measurements were performed to verify the oligomeric state and solution homogeneity of the PanMbs. Sample solutions (15 µl) were exposed to a 660 nm diode laser in a quartz glass cuvette inside a Specrolight 300 instrument (XtalConcepts, Germany) at 20°C for 100 s. The autocorrelation function of scattered light was evaluated using the CONTIN algorithm to determine diffusion constants and finally the hydrodynamic radius (Rh) distribution of the protein solutions.

3.3.7 Circular dichroism (CD) spectroscopy

The secondary structure was investigated by comparing the far-UV CD-spectra of the respective PanMbs. Protein samples were prepared in 10 mM Tris-HCl inside a 1 mm path length glass cuvette, which was positioned inside a Jasco J-815 spectropolarimeter (Jasco, USA). For each sample, ten spectra were recorded and averaged with the subsequent subtraction of the averaged ellipticity of the buffer solution. Experiments were performed at 19 °C with an increment of 0.1 nm per data point and a scanning speed of 100 nm/min. A far-UV CD spectrum of horse Mb (ID: CD0000047000) was obtained from the Protein Circular Dichroism Data Bank (PCDDB) (Whitmore et al. 2017).

3.3.8 Small-angle X-ray scattering (SAXS)

Homogeneous monodisperse solutions of histidine-tagged PanMbs were applied to SAXS experiments to analyse the tertiary structure envelope of selected PanMbs in solution. The experiments were performed at the EMBL beamline P12 (Blanchet et al. 2015) of the PETRA III synchrotron radiation source (DESY, Hamburg, Germany) applying a wavelength of 0.124 nm at a sample to detector distance of 3.1 m. The data were recorded in a momentum transfer range of 0.03-4.80 nm−1 with 20 exposures for 0.05 s per sample and at a temperature of 14 °C. The protein concentration range was 0.4-3.5 mg/ml. The SASFLOW data processing pipeline (Franke et al. 2012), as well as PRIMUS (Petoukhov et al. 2012), were applied for primary data quality verification and reduction, to exclude radiation damage as well as averaging of the scattering amplitudes of the individual X-ray exposures. Subsequently, the averaged scattering intensity of the buffer solution, which was exposed before and after each protein was subtracted and samples were scaled relative to the protein concentration. Radii of gyration (Rg) were derived from both the Guinier plot and the pair-distance distribution function using the PRIMUS user

interface including AUTOGNOM. The pair-distance distribution function was used as an input for further ab-initio modelling. Herein GASBOR (Svergun et al. 2001) was applied to calculate chain-like assemblies of dummy amino acid residues with P1 symmetry to visualise the protein shape.

3.4 Results

3.4.1 Identification of myoglobin cDNA in four lungfish species

Assemblies of the transcriptomes of *P. dolloi*, *P. aethiopicus*, *P. annectens* and *L. paradoxa* (Table SIII.1) revealed multiple Mb copies in all four lungfish species (Table SIII.2). Eight Mb cDNA sequences had been identified before in *P. annectens* (Koch et al. 2016), and a ninth PanMb cDNA sequence was found in the transcriptomes that have recently become available (Table SIII.1). The *Mb* genes were named according to their orthology and phylogenetic position in the tree (Fig. III.1; Fig. SIII.1), which required renumbering of two previously described *P. annectens Mb* genes (PanMb4 => PanMb4a and PanMb7 => PanMb4b). The newly identified PanMb cDNA sequence was named PanMb7. The main expression site of the Mbs in *P. annectens* were the brain, which agrees with the previous results (Koch et al. 2016) (Fig. SIII.2; Table SIII.3).

The transcriptomes of the slender lungfish *P. dolloi* harbours six Mbs (PdoMbs), which were named by orthology to the PanMbs PdoMb1, PdoMb2a, PdoMb2b PdoMb3a, PdoMb3b and PdoMb7. The PdoMbs share between 54.6 and 94% sequence identity on nucleotide level and 36.9 to 93.3% identity on the protein level. The available *P. dolloi* transcriptomes derive from pelvic and pectoral fins, and nasal lymphoid aggregates. The two fin tissues, which presumably mainly consist of bones and muscles, have very similar patterns and show that PdoMb7 is the strongestly expressed Mb in these samples (Fig. SIII.3; Table SIII.4). In the marbled lungfish *P. aethiopicus*, five Mbs were identified (PaeMb1-5) that share 44.9 to 61% of the nucleotides and 29.7 to 66.7% of the amino acids. The expression of the PaeMbs was evaluated by RNA-Seq, which showed the highest mRNA levels for PaeMb2 in the developing jaw (Fig. SIII.4; Table SIII.5). The transcriptomes of *L. paradoxa* revealed five distinct Mb sequences, named LpaMb2, LpaMb3, LpaMb4, LpaMb5, and LpaMb7, respectively. They have a sequence identity of 46.2 to 74.2% on the nucleotide and 28.2 to 72.7% on the amino acid level. Like in *P. annectens*, the highest Mb expression levels were observed in the brain. Also, like PanMb2, LpaMb2 had the highest mRNA level (Fig. SIII.5; Table SIII.6). In none of the lungfish transcriptomes, cDNA

sequences that represent Ngb or Cygb were detected. Comparisons between orthologs showed mostly high conservation (e.g. 80 to 88% identity between *P. annectens* and *L. paradoxa* Mb orthologs), with the exception of the Mb7 proteins (28% identity of PanMb7 and LpaMb7). Phylogenetic analysis shows that all lungfish Mbs form a single clade, which is in sister group position to the tetrapod Mbs (supported with 0.95 Bayesian posterior probability [BPP]; Fig. III.1; Fig. SIII.1). This clade is the sister group of the coelacanth Mb, and the monophyly of the sarcopterygian Mbs is supported with 1.0 BPP. Within the lungfish, distinct Mbs from different species form common clades, giving rise to orthologous *Mb* genes in the different species. Some Mbs are apparently absent in certain species (such as Mb1 in *L. paradoxa*), others as far as the transcriptome data indicates are in-paralog that have duplicated within a species (thus named, e.g., PanMb6a and 6b, or PdoMb3a and 3b). PanMb4b may be an out-paralog that has been lost in other lungfish species, but the support values within this clade are comparably low.

Table III.1: Summary of the results of the PanMbs

	Main expression site	coordination	P_{50} [Torr]	Hill Coeff	Nitrite reductase rate [s^{-1} M^{-1}]
PanMb1	muscle	penta	3.78 ± 0.01	1.63 ± 0.15	26.95 ± 0.005
PanMb2	brain/eye	penta	0.88 ± 0.13	1.35 ± 0.18	13.78 ± 0.0007
PanMb3	brain/eye	penta	1.37 ± 0.04	1.57 ± 0.27	42.13 ± 0.0002
PanMb4a	heart	penta	1.52 ± 0.10	1.59 ± 0.12	22.91 ± 0.0018
PanMb4b	brain/eye	penta	1.94 ± 0.30	1.64	33.63 ± 0.0025
PanMb5	brain/eye	penta	2.29 ± 0.08	1.63 ± 0.06	10.26 ± 0.0032
PanMb6b	brain/eye	penta	1.83 ± 0.07	1.33 ± 0.17	0.00

3.4.2 O_2 equilibria of *P. annectens* myoglobins

The Mbs of *P. annectens*, PanMb1, 2, 3, 4a, 4b, 5, and 6b were successfully recombinantly expressed applying the pET vector system. PanMb6a could not be cloned, possibly due to a low expression level and/or high similarity to PanMb6b. PanMb7 could also not be cloned.

The O_2 equilibrium curves of all seven PanMbs were hyperbolic and showed that all PanMbs reversibly bind oxygen (Fig. III.2). However, the O_2 affinities (P_{50}) and cooperativity values (n) were quite distinct (Table III.1). Within the given parameters, PanMb2 had the highest O_2

affinity (P_{50} = 0.88 ± 0.13 Torr), followed by PanMb3 (P_{50} = 1.37 ± 0.04 Torr), PanMb4a (P_{50} = 1.52 ± 0.10 Torr), PanMb6b (P_{50} = 1.83 ± 0.07 Torr), PanMb4b (P_{50} = 1.94 ± 0.30 Torr), PanMb5 (P_{50} = 2.29 ± 0.08 Torr) and PanMb1 (P_{50} = 3.78 ± 0.01 Torr) with the lowest affinity (Fig. III.2 and Fig. SIII.6). Hill-plots show slight cooperativity (n > 1) for all PanMbs, suggesting an oligomeric assembly (Fig. III.2B): PanMb1 (n = 1.63 ± 0.15), PanMb2 (n = 1.35 ± 0.18), PanMb3 (n = 1.57 ± 0.27), PanMb4a (n = 1.59 ± 0.12), PanMb4b (n = 1.64 ± 0.07), PanMb5 (n = 1.63 ± 0.06), and PanMb6 (n = 1.33 ± 0.17).

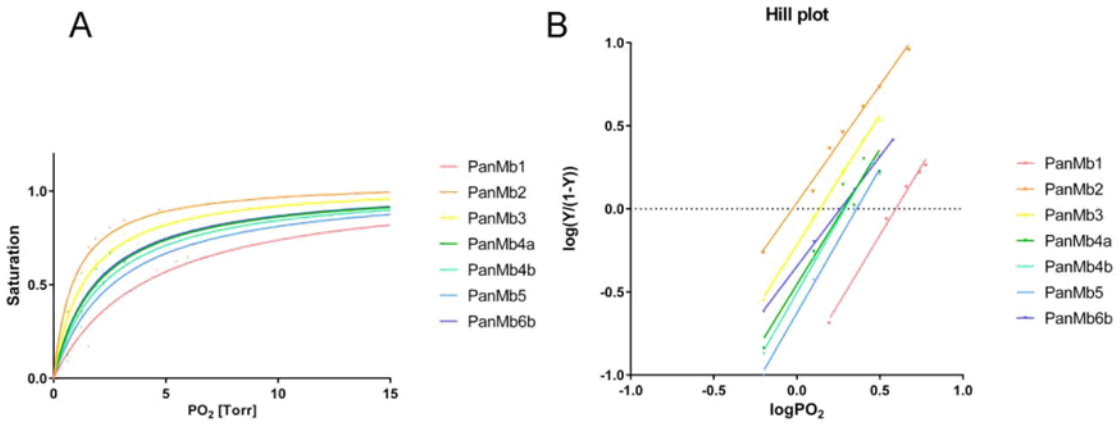


Figure III.2: Oxygen equilibrium curves of recombinant PanMbs, measured at pH 7.2, 20°C.

The sigmoidal curves (A) and linear form of the Hill plot (B) are depicted. Slight cooperativity is evidenced by the slope (n>1) of the Hill plot.

3.4.3 Enzymatic activities of P. annectens myoglobins

The nitrite reductase activity of the deoxy PanMbs was measured after addition of 0.1 mM nitrite. The reactions took place at different observed rates (Fig. III.3, Table III.1, Table SIII.7), with different derived second order rate constants (expressed as M^{-1} s^{-1}). For PanMb3 (42.13 ± 0.0002 M^{-1} s^{-1}) and PanMb4b (33.63 ± 0.0025 M^{-1} s^{-1}) the nitrite reductase activity was high, the nitrite reductase rates of the other Mbs were lower: PanMb1 (26.95 ± 0.005 M-1 s^{-1}), PanMb2 (13.78 ± 0.0007 M^{-1} s^{-1}), PanMb4a (22.91 ± 0.0018 M^{-1} s^{-1}), and PanMb5 (10.26 ± 0.0032 M^{-1} s^{-1}). PanMb6b showed no apparent nitrite reductase activity.

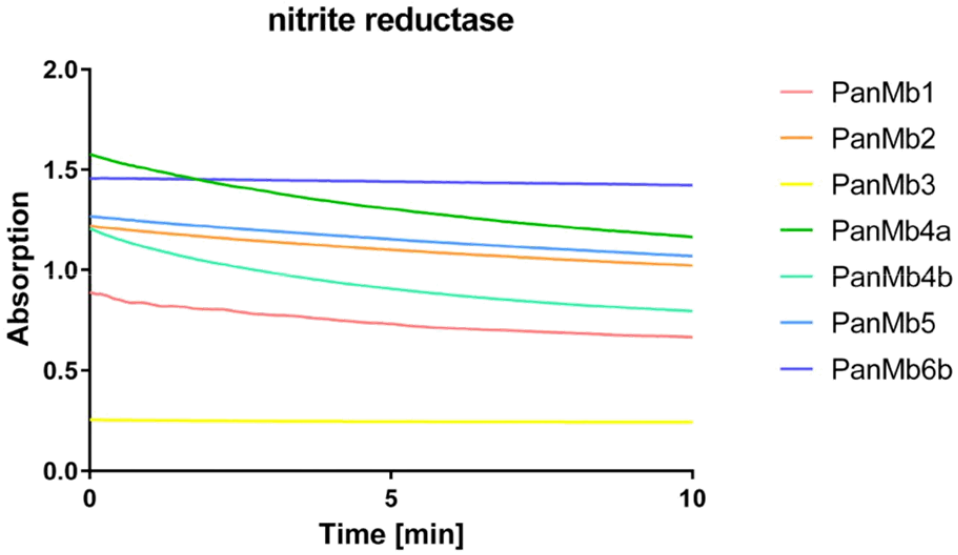


Figure III.3: Nitrite reductase activity of recombinant PanMbs, measured at pH 7.2, 20°C, shows the decrease of the absorbance over time of the deoxy Soret peak.

3.4.4 Structural analyses of *P. annectens* myoglobins

Gel filtration of the oxy-PanMbs showed approximately the same molecular mass for all PanMbs (Fig. SIII.7). These results were confirmed by SDS PAGE (Figure SIII.8). The absorbance peaks of the respective chromatograms also indicate that all PanMbs elute mainly as monomers. An additional small peak in PanMb1, PanMb2 and PanMb6b suggest minor dimeric fractions in these recombinant proteins.

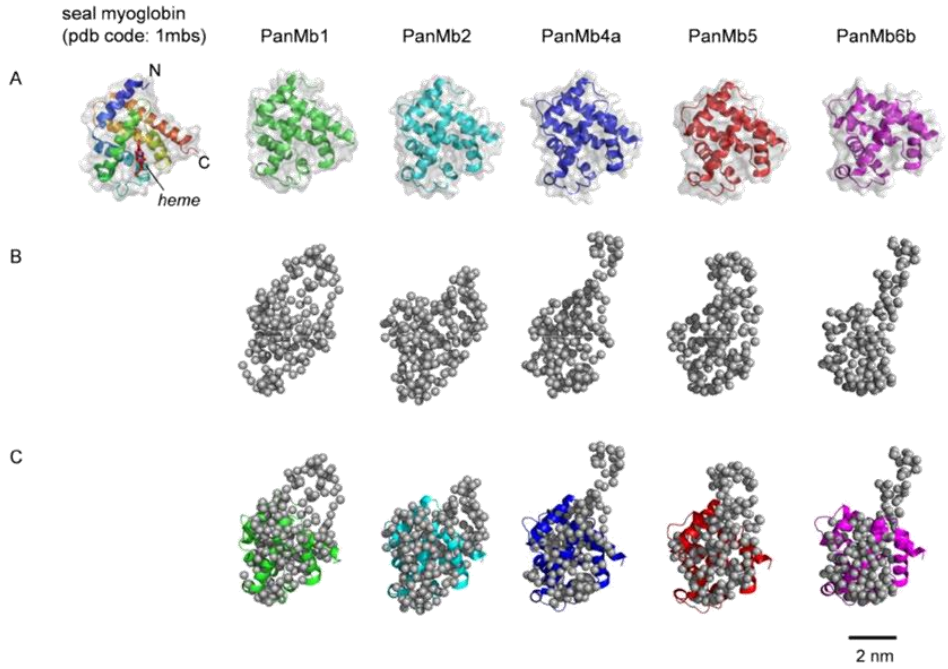


Figure III.4: Superimposition of *in silico* models calculated by Swiss Model and Gasbor *ab initio* models.

A) Cartoon plot of PanMb1-6 *in silico* models in comparison to a myoglobin high-resolution X-ray structure possessing the conserved myoglobin fold. The surface representation is colored light grey. B) *Ab initio* models of C-terminally tagged PanMbs calculated by GASBOR based on the respective P(R) function. A chain-like set of dummy spheres was used to fit the X-ray scattering intensity distribution. The GASBOR χ^2-values of the fit function with the experimental data are 1.11, 1.07, 1.34, 0.99 and 1.03 in the sequence the models are displayed from left to right. A globular core structure of the PanMbs in solution is conserved with a slightly more distinct structure of the respective termini C) Superimposition of both types of models displayed above. The scale bar is 2 nm. The Mb of the harbour seal (pdb code: 1mbs) was used as reference.

DLS experiments showed that all seven analysed PanMbs have approximately the same hydrodynamic radius (Rh) (Table III.3), which is comparable to Rh of human Mb (Fedorov and Denesyuk 1978). SAXS measurements revealed very similar radii of gyration (Rg) and the typical rigid and compact Mb fold for the investigated Mbs with only minor deviation. Ab initio models of the PanMbs based on solution scattering are well superimposable with the respective in silico models calculated via Swiss Model (Biasini et al. 2014, Bienert et al. 2017) using default modelling parameters (Fig. III.4). Further, the respective CD spectra were highly similar and indicated a conserved secondary structure composition compared to each other and to an Mb reference spectrum (around 70% of α-helical structure) (Fig. III.5). The secondary structure composition was also not significantly affected by the purification tag.

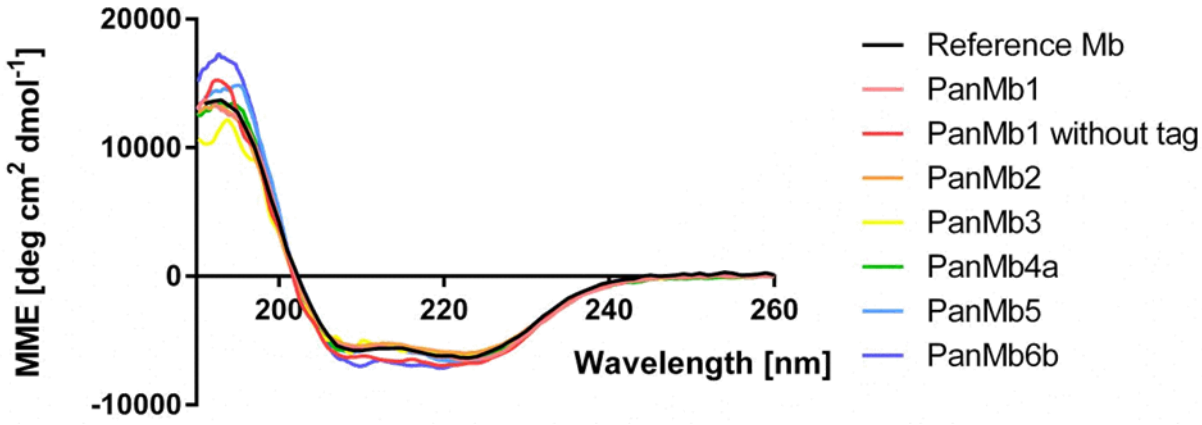


Figure III.5: CD spectra of recombinant PanMbs compared to a horse Mb as reference.

3.5 Discussion

3.5.1 Emergence of multiple myoglobins as a possible adaptation of lungfish

Lungfish (Dipnoi) share with the land-living vertebrates (tetrapods) the ability to breathe air. They first appear in the fossil record in the Devonian ~ 400 million years ago (Clack 2011). There are only six extant lungfish species that live in rivers and (seasonal) freshwater lakes in the tropics (Jørgensen and Joss 2010). Four species of the genus Protopterus live in Africa (*Protopterus* aethiopicus, *P. amphibius*, *P. annectens* and *P. dolloi*), *Lepidosiren paradoxa* in South America and *Neoceratodus forsteri* in Australia. Although lungfish are often considered to be "living fossils"(Lee et al. 2006), they also show some specific adaptations. For example, lungfish can withstand hypoxia and can survive certain periods outside the water. African lungfishes may also aestivate during the dry season in a mud cocoon (Ballantyne and Frick 2010). During this period, lungfishes reduce their energy metabolism, increase urea production to reduce ammonia toxicity and turn on the expression of cytoprotective genes (Chew et al. 2004, Hiong et al. 2013). Lungfish also have the largest animal genomes, with C-values up to 132 pg per haploid genome (Gregory 2017).

As none of the other globins (with the exception of GbE; see (Ludemann et al. 2019)) or, to the best of our knowledge, any other gene family show signs of gene amplification, it is unlikely that the multiplication of the *Mb* genes is caused by hypothetic whole-genome amplifications putatively linked to the large lungfish genomes. Duplicated and even triplicated *Mb* genes have been known from some teleost species (Fraser et al. 2006, Roesner et al. 2008, Cossins et al. 2009, Macqueen et al. 2014, Gallagher and Macqueen 2017). Lungfish are the current record

holders with up to nine distinct Mbs. While the duplicated and triplicated *Mb* genes of teleosts are relatively recent innovations in evolution, the amplification of *Mb* genes in lungfish must have occurred before the genera *Lepidosiren* and *Protopterus* split more than 100 million years ago (Betancur et al. 2015). In the future, the availability of genomic sequences of the Australian lungfish *N. forsteri* may push this date further back to the divergence time of Lepidosireniformes and Ceratodontiformes > 200 million years ago. In general, genomic data would help to confirm some of our assumptions. With the help of genomic analyses we could explicitly attest the losses of Cygb and Ngb in lungfish.

The amplification of the different lungfish Mbs occurred in the lungfish stem lineage (Fig. III.1, Fig. SIII.1), which might have been a response to a lifestyle with low and changing O_2 conditions (Koch et al. 2016) or other physiological challenges. It is noteworthy that most Mb orthologs have been maintained for > 100 million years, suggesting that the lungfish requires specific Mbs rather than just multiple copies of the gene. Only in four cases, we found additional paralogs (PdoMb2a and 2b, PdoMb3a and 3b, PanMb4a and 4b, and PanMb6a and 6b). The *Mb7* gene may have been lost in *P. aethiopicus*, but additional transcriptome or genome sequencing is required for verification. Our results strongly hint to subfunctionalisation or neofunctionalisation of the *Mb* genes during lungfish evolution, a hypothesis that is supported by their tissue-specific expression patterns (Koch et al. 2016), as well as the different kinetic and enzymatic properties of the recombinant proteins (see below).

Table III.2: Amino acid structure of PanMbs at conserved globin positions

	CD1	CD3	E7	E10	E11	F8
PanMb1	Phe	Lys	His	Thr	Val	His
PanMb2	Phe	Lys	His	Thr	Val	His
PanMb3	Phe	Lys	His	Thr	Val	His
PanMb4a	Phe	Lys	His	Val	Val	His
PanMb4b	Phe	Lys	His	Val	Val	His
PanMb5	Phe	Lys	His	Val	Val	His
PanMb6a	Phe	Lys	Gln	Val	Val	His
PanMb6b	Phe	Lys	Gln	Val	Val	His
PanMb7	Phe	Lys	His	Leu	Val	His

3.5.2 Conserved structure of *P. annectens* myoglobins

Despite notable differences of the amino acid sequences (up to 71.1% difference), the secondary structures of the PanMbs are highly similar and show essentially the same high content of α-helices. The minor differences most likely result from, e.g. the N-terminally shortened sequence of PanMb6b or slightly different surface charge distribution. As seen in the DLS measurements, all PanMbs have almost identical hydrodynamic radii, which fit a nearly globular shape of monomeric Mbs in solution (Table III.3). This observation was verified by the SAXS data, which shows similarities to the tertiary structure of sperm whale Mb originally determined by Kendrew et al. (Kendrew et al. 1960). Size-exclusion chromatography experiments confirmed these results and found mainly monomeric PanMb proteins, although for some a minor dimeric fraction was observed, which may explain the slightly cooperative O_2-binding kinetics. However, the SAXS data did not indicate a significant concentration-dependent oligomerisation of the applied monomeric proteins.

Table III.3: Characteristic size parameters of PanMbs, as obtained by DLS and SAXS respectively

Protein	R_h [nm]	R_g [nm]*	D_{max} [nm]	Oligomeric state
PanMb1	2.2 ± 0.3	1.99 ± 0.04 (1.97 ± 0.01)	6.2	Monomer
PanMb2	2.1 ± 0.3	1.83 ± 0.08 (1.85 ± 0.02)	6.4	Monomer
PanMb4a	1.9 ± 0.3	2.13 ± 0.16 (2.13 ± 0.01)	6.7	Monomer
PanMb5	2.0 ± 0.3	1.86 ± 0.10 (1.85 ± 0.02)	5.6	Monomer
PanMb6b	1.9 ± 0.2	2.04 ± 0.20 (2.03 ± 0.01)	6.3	Monomer

*****calculated according to the Guinier approximation; the numbers in parenthesis show R_g values according to the pair-distance distribution function

3.5.3 Functional differentiation of *P. annectens* myoglobins

The O_2-affinities of the PanMbs range between P_{50} values from 0.88 to 3.8 Torr. PanMb1 has the lowest O_2 affinity with P_{50} = 3.8 Torr, similar to the low affinity of some fish Mb, such as mackerel Mb with P_{50} = 3.7 Torr at 25 °C (Madden et al. 2004) and trout Mb with P_{50} = 3.4-4.9 Torr at 25 °C (Pedersen et al. 2010, Helbo and Fago 2011). PanMb1 is mainly expressed in the skeletal muscle and is also by far the most highly expressed Mb of this tissue (Koch et al. 2016). The comparably low affinity suggests that PanMb1 is suitable for efficient delivery of O_2 to the mitochondria. This finding correlates with the high efficiency of PanMb1 to enhance the activity of the mitochondrial dehydrogenases under hypoxia in cell culture (Koch et al. 2016). Higher O2-affinities were observed for the other PanMbs, with the brain-specific PanMb2 having the highest O_2-affinity (Table III.1). The specific O_2-affinities of the different PanMbs probably reflect the individual requirements of the respective tissue. For example, the neurons may require an Mb with higher O_2 affinity than the skeletal muscle.

Deoxygenated globin proteins can reduce nitrite to NO (Tejero and Gladwin 2014, Fago et al. 2018). All PanMbs except PanMb6b display noticeable nitrite reductase activity (Table III.1). It is, therefore, conceivable that the high expression levels of Mb help to protect hypoxic tissues by increasing vasodilation and blood supply. Also, NO may inhibit mitochondrial respiration and limit the generation of ROS (Hendgen-Cotta et al. 2008, Helbo et al. 2013).

The lungfish Mbs support the idea that distinct evolutionary mechanisms can control the heme redox and the non-redox (O_2 binding) activities separately, depending on the need of the respective tissue or cell (Fago et al. 2018). PanMb6a and PanMb6b are the only Mbs of the lungfish with an E7 glutamine substitution in the sequence (Table III.2). The elephant Mb has the same substitution (Romero-Herrera et al. 1981) and reacts with NO 500-1000 times faster than Mbs without substitution. Also, PanMb6b is the only one with no apparent nitrite reductase activity. Taken together the E7 glutamine substitution is unlikely to be the only reason to explain these differences, but in case of the lungfish Mbs, it seems to be responsible for the negligible nitrite reductase enzymatic activity of deoxy PanMb6b.

3.5.4 Functional differentiation of lungfish myoglobin in evolution

Our results strongly suggest that the various lungfish Mbs carry out distinct biological functions. It is noteworthy that the muscle-specific Mb1 is in a sister group position to all other lungfish

Mbs (Fig. III.1, Fig. SIII.1). Thus, the original function of lungfish Mb probably was the supply of O_2 to the muscle tissue, the classic Mb-function found for most vertebrate Mbs. The other lungfish Mbs differentiated from that role after multiple gene duplications. Because the protein sequences of the other lungfish Mbs are highly similar to the sequences of the PanMbs, we suppose them having similar functions in the same tissues. On the one hand, an adaptation to a temporarily hypoxic environment may have led to the broad expression pattern of Mb in various tissues, thereby enhancing O_2 supply or reducing respiration rates via nitrite reduction to NO. All Mbs, in particular, PanMb3 and PanMb4, are highly efficient as nitrate reductases (see Table III.1).

In comparison, vertebrate Mbs have a much lower nitrite reductase activity (Helbo et al. 2013). In addition, it is conceivable that the derived loss of Ngb and Cygb provided the impetus for the functional differentiation of lungfish Mbs, of which some might have taken over the yet still unclear roles of Ngb and Cygb. This hypothesis would also explain the preferential expression of some lungfish Mbs in the brain, which is also the main expression site of Ngb and Cygb (Hankeln et al. 2005, Burmester and Hankeln 2014).

3.6 Acknowledgements

The authors would like to thank Alexey Kikhney (EMBL Hamburg) for the assistance in operating the SAXS beamline P12 at the PETRA III storage ring and Elin E. Petersen (Aarhus) for skilled assistance in the lab. This work is supported by the Deutsche Forschungsgemeinschaft (BU956/19-1), CNPq Universal Program Grant 403248/2016-7 to I.S and Natur og Univers, Det Frie Forskningsråd (grant 4181-00094 to AF).

IV. Chapter 3:

Globin E is a myoglobin-related, respiratory protein highly expressed in lungfish oocytes

Julia Lüdemann[1], Kellen Matos Verissimo[2], Kimberley Dreger[1], Angela Fago[3], Igor Schneider[2] & Thorsten Burmester[1]

[1]Institute of Zoology, University of Hamburg, D-20146, Hamburg, Germany

[2]Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, PA, Brazil

[3]Department of Bioscience, Aarhus University, DK-8000, Aarhus C, Denmark

Published in *Scientific Reports*, January 2019, Volume 9, Issue 1, Article number 280

Supplementary Material can be found on the enclosed CD-ROM

4.1 Abstract

Globins are a classical model system for the studies of protein evolution and function. Recent studies have shown that – besides the well-known haemoglobin and myoglobin – additional globin-types occur in vertebrates that serve different functions. Globin E (GbE) was originally identified as an eye-specific protein of birds that is distantly related to myoglobin. GbE is also present in turtles and the coelacanth but appeared to have been lost in other vertebrates. Here, we show that GbE additionally occurs in lungfish, the closest living relatives of the tetrapods. Each lungfish species harbours multiple (≥5) *GbE* gene copies. Surprisingly, GbE is exclusively and highly expressed in oocytes, with mRNA levels that exceed that of myoglobin in the heart. Thus, GbE is the first known oocyte-specific globin in vertebrates. No GbE transcripts were found in the ovary or egg transcriptomes of other vertebrates, suggesting a lungfish-specific function. Spectroscopic analysis and kinetic studies of recombinant GbE1 of the South American lungfish *Lepidosiren paradoxa* revealed a typical pentacoordinate globin with myoglobin-like O_2-binding kinetics, indicating similar functions. Our findings suggest that the multiple copies of GbE evolved to enhance O_2-supply in the developing embryo of lungfish, analogous to the embryonic and fetal haemoglobins of other vertebrates. In evolution, GbE must have changed its expression site from oocytes to eyes, or vice versa.

4.2 Introduction

A constant supply of oxygen (O_2) is essential for aerobic organisms. The transport and storage of O_2 in vertebrates are mediated by proteins that are members of the globin superfamily (Burmester and Hankeln 2014). Some globins may also have other functions and are, for example, involved in the detoxification of reactive O_2 species (ROS), NO metabolism, or signaling (Weber and Vinogradov 2001, Burmester and Hankeln 2014). The best-known vertebrate globins are haemoglobin (Hb), which is a heterotetramer that transports O_2 in the blood (Dickerson and Geis 1983), and myoglobin (Mb), which is a monomer in the heart and the skeletal muscles, where it facilitates the diffusion of O_2 and enhances O_2 storage (Wittenberg and Wittenberg 2003). Within recent years, six additional globins have been identified in vertebrates (Burmester and Hankeln 2014). The function of neuroglobin (Ngb), which resides mainly in the nervous system(Burmester et al. 2000), is still uncertain (Hankeln et al. 2005, Burmester and Hankeln 2009). There is evidence that Ngb plays a role in oxidative metabolism (Bentmann et al. 2005, Mitz et al. 2009). Cytoglobin (Cygb) is expressed in fibroblast-related cell types and some populations of neurons (Burmester et al. 2002, Nakatani

et al. 2004, Schmidt et al. 2004). Cygb may supply O_2 to specific enzymes and may detoxify ROS (Hankeln et al. 2005). Androglobin (Adgb) expression is restricted to the testis (Hoogewijs et al. 2012). While Hb, Mb, Ngb, Cygb, and Adgb occur in most vertebrates, the occurrence of the globins E, X, and Y (GbE, GbX, and GbY) is restricted to certain taxa. GbX emerged very early in the evolution of Metazoa but is – in vertebrates – only present in non-tetrapods, amphibians and some reptiles (Roesner et al. 2005, Fuchs et al. 2006). The GbX protein is bound to the cell membrane via N-terminal acylation (Blank et al. 2011b, Blank and Burmester 2012), where it may protect the cells from ROS (Koch and Burmester 2016). GbY has an unknown function in some "basal" ray-finned fishes, amphibians, reptiles, and platypus, where it is broadly expressed at low levels (Fuchs et al. 2006, Gallagher and Macqueen 2017).

GbE was initially found in the eye of chicken (Kugelstadt et al. 2004) and was additionally identified in the genomes of other birds (Blank et al. 2011a, Hoffmann et al. 2011, Storz et al. 2011), the coelacanth (Schwarze and Burmester 2013) and turtles (Schwarze et al. 2015). Gene synteny and phylogenetic analyses suggest that Mb is the closest related globin type of GbE, although the divergence of these genes must have occurred before the radiation of the gnathostome classes (Blank et al. 2011a, Hoffmann et al. 2011, Storz et al. 2011, Storz et al. 2013). Immunohistochemistry and quantitative realtime RT-PCR (qRT-PCR) studies showed that GbE is highly and almost exclusively expressed in the eye (thus its name) of chicken and turtles (Schwarze et al. 2015). Estimates of total protein levels were ~10 µM GbE in the chicken retina, which is in the range of Mb in striated muscle cells (Blank et al. 2011a).

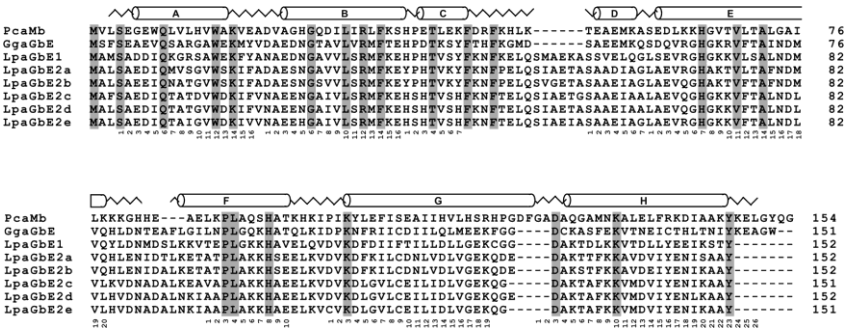


Figure IV.1: Comparison of the GbE and Mb amino acid sequences.

The GbE sequences of the South American lungfish *L. paradoxa* (LpaGbE1–2e), the chicken (GgaGbE) and sperm whale Mb (PcaMb) were aligned. The secondary structure of sperm whale Mb is superimposed in the upper row, with α-helices designated A through H; the globin consensus numbering is given below the sequences. Strictly conserved residues are shaded in grey. The conserved globin residues, i.e. the proximal His F8 and the distal His E7, as well as the Phe CD1 are in boldface.

Together, the available evidence is consistent with GbE having a Mb-like function in O_2 supply to the metabolically highly active avian retina (Blank et al. 2011a).

Lungfish (Dipnoi) have received much scientific interest because of their ability to breathe air, their conserved morphology that remained largely unchanged since the Devoian, and their phylogenetically position as closest living relatives of the tetrapods (Amemiya et al. 2013, Irisarri and Meyer 2016, Irisarri et al. 2017, Takezaki and Nishihara 2017). There are six extant lungfish species that dwell in rivers and (seasonal) freshwater lakes in the tropics (Jørgensen and Joss 2010). Four species of the genus *Protopterus* live in Africa, *Lepidosiren paradoxa* in South America and *Neoceratodus forsteri* in Australia. Nearly all vertebrates have only a single *Mb* gene, which is expressed in the skeletal and heart muscles. In striking contrast, the West African lungfish *P. annectens* harbours at least seven distinct *Mb* genes with tissue-specific expression patterns (Koch et al. 2016). For example, distinct Mb paralogs occur in the heart and skeletal muscle, and highest levels of Mb mRNA were found in the brain. Recombinant paralogous Mb proteins of *P. annectens* display different O_2 binding affinities and enzymatic activities (J. Lüdemann, A. Fago, T. Burmester, unpublished data). The data suggest that the lungfish Mb paralogs carry out distinct functions and that the *Mb* genes evolved by neofunctionalisation and/or subfunctionalisation after multiple gene duplications.

In the present study, we demonstrate the occurrence of multiple *GbE* genes in three lungfish species, which are expressed almost exclusively in the ovary. Together with the biochemical analyses of recombinant lungfish GbE, we propose that GbE may have an Mb-like role in O_2 supply in the ovary of lungfish during development. Our findings add a novel level of complexity to the studies of globin evolution and function.

4.3 Results

4.3.1 Identification of GbE genes in lungfish species

An assembly of transcriptomes of the South American lungfish *L. paradoxa* (Supplemental Information Table IV.1) revealed several globin genes. TBLASTN and BLASTN searches identified two *Hb* α, three *Hb* β, one *GbX*, one *GbY*, five *Mb* cDNA sequences, as well as multiple contigs that had resembled the sauropsid GbE (Supplemental Information Figs. IV.1 and IV.2). No sequences that matched Ngb, Cygb, or Adgb were found in the available transcriptomes of *L.*

paradoxa. The GbE cDNA sequences were verified by backmapping of the reads, reassembled if required, and finally revealed six distinct GbE transcripts, which were named *LpaGbE1* and LpaGbE2a to e. (Fig. IV.1). We should note that the nomenclature is provisional until a full representation of lungfish *GbE* genes is achieved. *LpaGbE1*, *2a*, and *2c* could be further verified by RT-PCR and sequencing. However, some sequences obtained by RT-PCR could not be assembled from the Illumina reads (which derived from a different specimen), indicating either technical issues such as hybrid sequences generated by the PCR step, or biological causes such as different alleles in the population, along with gene conversion or crossing over among the closely related *GbE* genes. Within the coding region of 456 or 459 bp, respectively, the six *LpaGbE* sequences differ between 2.6 and 34.8%. After translation, the differences were 4.7 to 41.4% on the amino acid level. The largest difference was found between *LpaGbE1* and *LpaGbE2b*.

We assembled the publically available transcriptomes of the West African lungfish *P. annectens* (Biscotti et al. 2016) and searched them for *GbE* genes. We found five distinct GbE cDNA sequences, which differ between 1.8 and 27.2% on the nucleotide and between 1.3 and 36.2% on the amino acids level. Because the sequences form two clades in phylogenetic analyses (see below), they were named *PanGbE1a* and *b*, and *PanGbE2a-c*, respectively.

Additionally, we generated 60,785,122 Illumina reads (150 nt, paired-end) from total RNA extracted from the ovary of the marbled lungfish *P. aethiopicus*. The reads were assembled and searched for GbE sequences, of which seven distinct cDNA sequences were identified. The sequences differ 2.4 to 27.7% on the nucleotide and 2.0 to 41.1% on the amino acid level. The sequences were named *PaeGbE1a-c*, and *PaeGbE2a-d*, respectively, based on the position in the phylogenetic tree (see below).

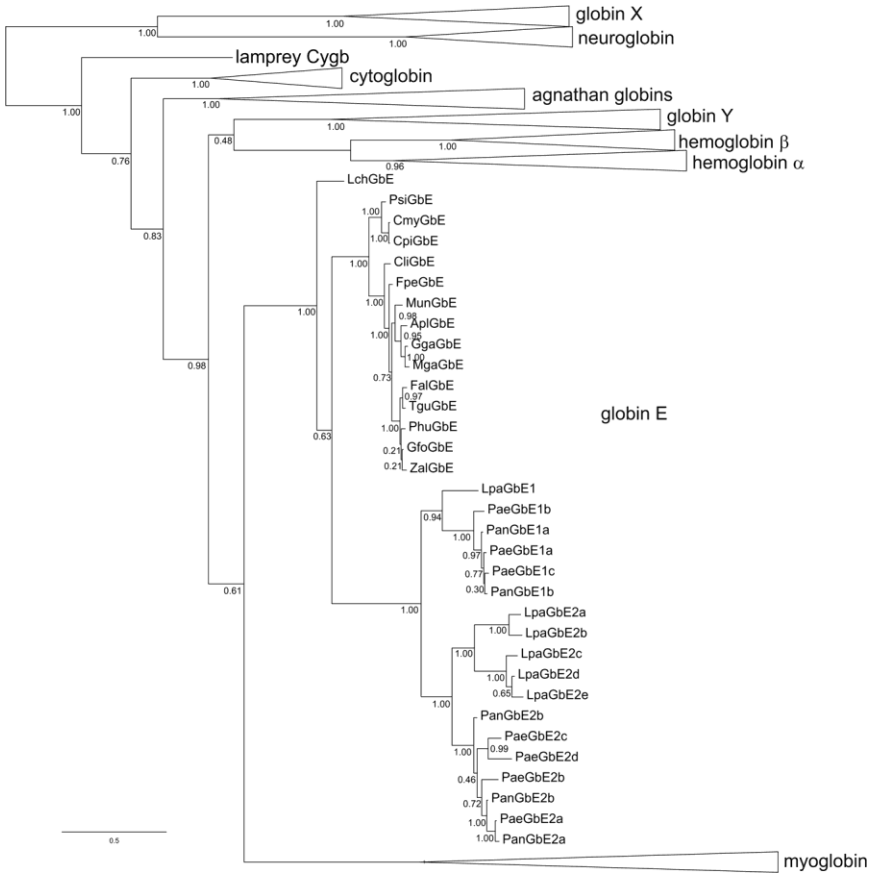


Figure IV.2: Bayesian phylogenetic tree of vertebrate globins.

Tree reconstruction was carried out with the amino acid sequences assuming the LG model. The bar represents 0.1 substitutions per site. The numbers at the nodes are posterior probabilities. The different globin clades – except GbE – have been collapsed; the full tree is given in Supplemental Information Fig. IV.4. The species abbreviations are: Apl, *Anas platyrhynchos*; Cli, *Columba livia*; Cmy, *Chelonia mydas*; Cpi, *Chrysemys picta bellii*; Fal, *Ficedula albicollis*; Fpe, *Falco peregrinus*; Gfo, *Geospiza fortis*; Gga, *Gallus gallus*; Lch, *Latimeria chalumnae*; Lpa, *Lepidosiren paradoxa*; Mga, *Meleagris gallopavo*; Mun, *Melopsittacus undulatus*; Pae, *Protopterus aethiopicus*; Pan, *Protopterus annectens*; Phu, *Pseudopodoces humilis*; Psi, *Pelodiscus sinensis*; Tgu, *Taeniopygia guttata*; Zal, *Zonotrichia albicollis*.

4.3.2 Conservation and lungfish-specific amplification of GbE genes

Our studies resulted in a total of 18 novel *GbE* genes from three lungfish species with 151 or 152 amino acids. An alignment with the known GbE amino acid sequences of birds, turtles and coelacanth showed that all lungfish GbEs carry an insertion of six amino acids in the region between helices C and D, a deletion of four amino acids in the GH interhelical region, and are five amino acids shorter (Fig. IV.1). The maximum divergence within lungfish GbE amino acid sequences was 42.1%. The divergence of lungfish and coelacanth GbE was between 42.1 and 52.8% of the amino acid. Sauropsid and lungfish GbE amino acid sequences differed by up to 56.4%. No other globin is more closely related to the lungfish GbE sequences.

Phylogenetic analyses using GbE and other globin amino acid sequences (Supplemental Information Fig. IV.3) confirmed previous studies, which found a relationship of GbE and Mb, although the support was not particularly high (0.61 posterior probability) (Fig. IV.2; Supplemental Information Fig. IV.4). Within the GbE clade, the overall topology of the tree followed the accepted relationships among gnathostome taxa. The coelacanth GbE formed the sister group of all other GbE sequences. Sauropsid and lungfish GbEs were two separate clades. The lungfish GbEs fell into two clades, of which one was formed by the GbE1 sequences, the other one by the *L. paradoxa* GbE2a-e along with the GbE2 proteins of the genus *Protopterus*.

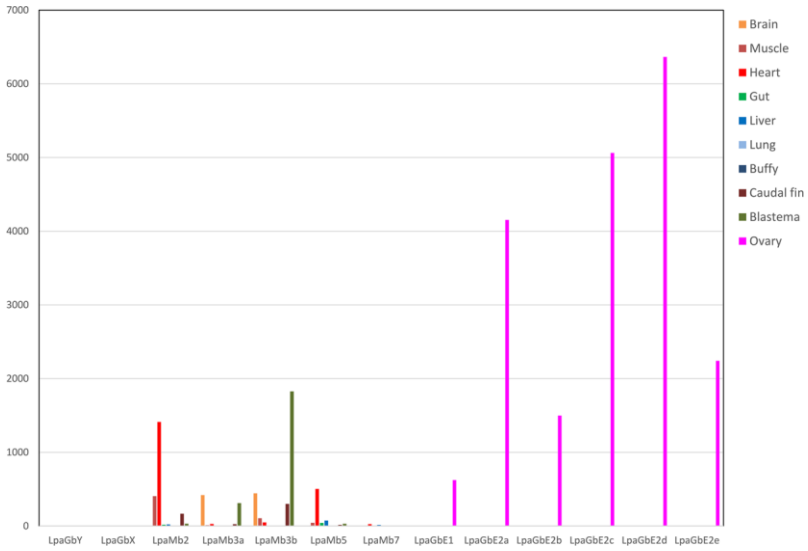


Figure IV.3: Expression of the Mb and GbE genes in selected *L. paradoxa* tissues.

mRNA levels were estimated by RNA-Seq and are displayed as RPKM values. Transcriptome accession numbers are given in Supplemental Information Table IV.1. Note the dominant expression of LpaGbE1-2e in the ovary. The copy numbers are given in Supplemental Information Table IV.3. Log-scale data are presented in Supplemental Information Fig. IV.5. The tissue-specific expression per gene is given in Supplemental Information Figs IV.6 and IV.7.

4.3.3 Expression of GbE in the lungfish ovary

We checked the expression levels of the six *GbE* by RNA-seq employing the transcriptomes of *L. paradoxa* (Fig. IV.3; Supplemental Information Figs. IV.5–IV.7). The five *Mb* genes, as well as *GbX* and *GbY*, were included for comparison. All six *GbE* genes were essentially restricted to the ovary transcriptome. The RPKM (Reads Per Kilobase exon model per Million reads) values were very high and reached 6,364.7 RPKM. In other tissues, the RPKM values were between 0 and 5 (Supplemental Information Table IV.3). The cumulative RPKM value of all six *GbE* genes in the ovaries was 19,942. For comparison: The cumulative RPKM of all five *Mb* genes was 562 in skeletal muscle and 2,017 in the heart (Supplemental Information Table IV.3). The total levels

of Mb mRNA in ovaries amounted to 5.92 RPKM. To validate the ovary-specific expression of *GbE*, quantitative real-time RT-PCR was carried out with *L. paradoxa* ovaries and other selected tissues (Fig. IV.4). Again, *GbE1* mRNA was exclusively detected in the ovaries, but not in other tissues, including the eye.

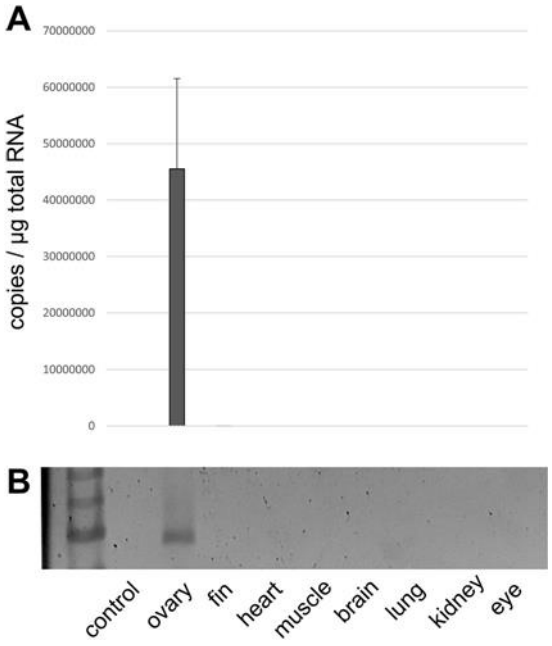


Figure IV.4: Expression of GbE1 in selected *L. paradoxa* tissues.

mRNA levels were determined by qRT-PCR (A) and RT-PCR (B). The standard deviations in (A) derive from three replicates. GbE1 was almost exclusively detected in the ovary.

We further analysed the transcriptomes of the West African lungfish *P. annectens* (Biscotti et al. 2016). These transcriptomes included male and female gonads, along with brain and liver transcriptomes. The five *GbE* genes were found highly expressed in the female gonads, with cumulative RPKM of 31,916 to 37,275 (Supplemental Information Fig. IV.8; Supplemental Information Table IV.4). In the male gonads, expression of GbE was at least 1,800-fold lower with cumulative RPKM of 0.85 and 17.69. In other tissues, cumulative RPKMs of GbE were <4. Except in the gonads, there was no difference in GbE levels between other male and female organs. In *P. annectens*, we also found an Adgb cDNA, which is expressed at low levels mainly in the mature male gonads (Supplemental Information Fig. IV.8). RNA-Seq analyses of the three available transcriptomes of the marbled lungfish *P. aethiopicus* showed the same picture: High expression of *GbE* in the ovary (cumulative RPKM of the seven *GbEs*: 53,661), but only traces of *GbE* in the transcriptomes of the developing jaw/mandible or mixed visceral tissues (Supplemental Information Fig. IV.9; Supplemental Information Table IV.5).

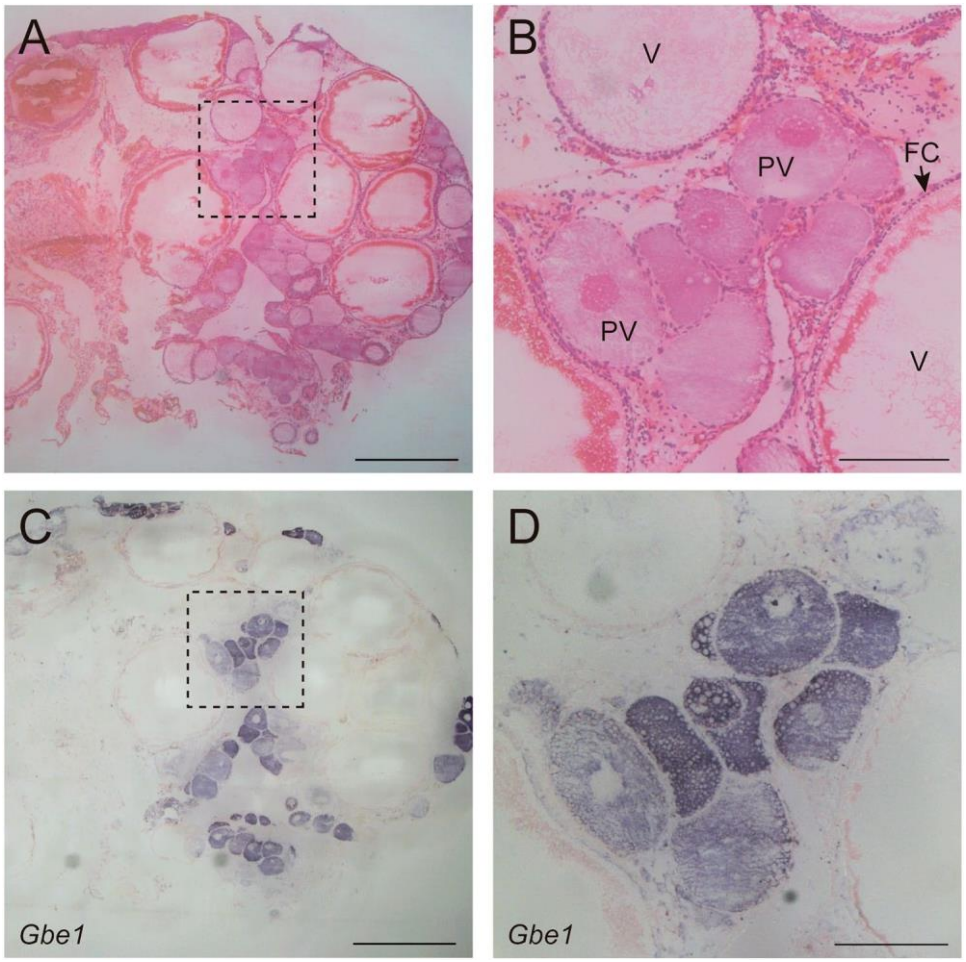


Figure IV.5: GbE1 expression in the ovary of *L. paradoxa*.

(A,B) Haematoxylin and eosin stained in ovary sections, oocytes at various stages of maturation can be seen. (C,D) Antisense GbE1 riboprobe signal was detected in previtellogenic oocytes. (B,D) Represent zoom in of a dashed line box in (A,C) respectively. (B) Arrow indicates follicular cells. Previtellogenic oocytes (PV); Vitellogenic oocyte (V); follicular cell (FC). Scale bar: 2 mm (A,C) and 0.5 mm (B,D). The sense controls are given in Supplemental Information Fig. IV.12.

We further identified GbE proteins in the ovaries of *L. paradoxa* and *P. aethiopicus* by mass spectrometry. After separation of ovary proteins by SDS-PAGE, prominent bands with the expected GbE mass of ~15 kDa were detected (Supplemental Information Fig. IV.10). Mass spectrometry identified in these bands the proteins LpaGbE1-2e in the *L. paradoxa* samples, and PaeGbE1a, b, c, and PaeGbE2 a, b in the *P. aethiopicus* samples, respectively (Supplemental Information Fig. IV.2).

To check whether the oocyte-specific expression was overlooked in previous studies with birds (Kugelstadt et al. 2004, Blank et al. 2011a), we evaluated the ovary- and egg-specific transcriptomes of chicken (*Gallus gallus*) (Supplemental Information Table IV.1). In none of these transcriptomes, *GbE* sequences were found (Supplemental Information Fig. IV.11; Supplemental Information Table IV.6). We randomly checked further ovary transcriptomes of

other vertebrate species, as available at SRA, for putative GbE sequences via BLAST. We included all vertebrate classes, but in none of these datasets, GbE sequences were detected. A comprehensive BLAST search of the available transcriptomes (including the transcriptome shotgun assemblies; TSA) or genomes at Genbank or ENSEMBL identified GbE only of birds, turtles, and the coelacanth.

4.3.4 Localisation of GbE mRNA in previtellogenic oocytes

To determine the spatial expression on *GbE1* of *L. paradoxa* (*LpaGbE1*) we performed *in situ* hybridisations in adult gonads of a female South American lungfish. Hematoxylin-eosin stained sections of the ovary show oocytes in various stages of maturation (Fig. IV.5A,B). Oocyte development in *L. paradoxa* consists of an initial stage of previtellogenic oocytes, characterised by a basophilic cytoplasm. Next continuous yolk deposition results in a rapid increase in cellular volume in vitellogenic oocytes. Flattened follicular cells surround the vitellogenic oocyte, and acidophilus yolk granules cortically localized gradually expand into the cytoplasm. Mature oocytes measure about 2 mm (Chaves 1992). *LpaGbE1* antisense probe signal was observed specifically in the cytoplasm of previtellogenic basophilic oocytes (Fig. IV.5C). *LpaGbE1* expression was not detected in vitellogenic oocytes, follicular cells or other ovarian cells (Fig. IV.5C,D). No signal was detected by the sense control probe (Supplemental Information Fig. IV.12).

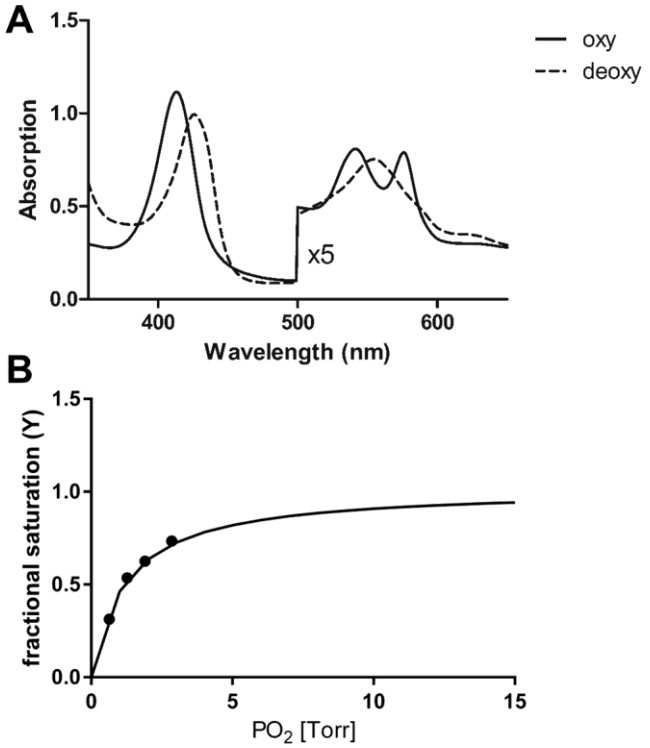


Figure IV.6: Heme coordination and O2 equilibria of recombinant GbE1 of *L. paradoxa*.

(A) Absorbance spectra of purified (solid line) and deoxygenated (dotted line) recombinant GbE1 after addition of dithionite, indicating penta coordinate heme. (B) Representative O2 equilibrium curve, measured at pH 7.2, 20 °C. Fitting of saturation data is indicated (continuous line).

4.3.5 Spectroscopic studies and O2 binding equilibria to GbE1 of *L. paradoxa*

In size-exclusion chromatography, recombinant LpaGbE1 elutes largely as a monomer (data not shown). The absorbance spectrum of purified LpaGbE1 displayed a Soret band at 406 nm, an α band at 533 nm and a β-band at 579 nm (Fig. IV.6A), indicating a mixture between ferric and ferrous oxy forms. For comparison, a pure Mb oxy spectrum displays peaks at 418, 543 and 581 nm (Antonini and Brunori 1970). After reduction with Na-dithionite under nitrogen, the ferrous deoxy-form was obtained, with a large amplitude of the Soret band (427 nm) and a single peak in the visible region (555 nm). The absorption spectrum of deoxy-LpaGbE1 resembled that of Mb and Hb, indicating a pentacoordinate heme. O_2 equilibrium curves (pH 7.2, 20 °C) showed that LpaGbE1 reversibly binds O_2 (Fig. IV.6B), with a P_{50} of 1.2 ± 0.02 torr (0.16 kPa; 1 torr= 0.133 kPa). The O_2 binding curve showed some degree of cooperativity (n= 1.19 ± 0.19), suggesting a dimeric assembly. The autoxidation rate of LpaGbE1 (pH 7.13, 20 °C), measured after removing the met reductase enzymatic system by gel filtration, was low (0.004 ± 0.0003 s−1). The nitrate reductase activity (pH 7.13, 20 °C) of deoxy LpaGbE1 was 18.72 ± 0.0022 s−1M−1.

4.4 Discussion

RNA-seq and qRT-PCR studies showed that GbE is almost exclusively expressed in the lungfish ovary (Figs IV.3 and IV.4). More detailed studies by *in situ* hybridisation found that *LpaGbE1* mRNA is restricted to non-mature previtello- genic oocytes, but was not found in other cells of the ovary, such as follicular cells, and was also not detected in vitellogenic oocytes (Fig. IV.5). The previtellogenic oocytes have not yet commenced accumulating yolk and other material. Thus, *GbE* mRNA is massively deposited in the oocyte and, as suggested by the mass spectrometric data, translated into GbE protein (Supplemental Information Fig. IV.10). Here, GbE may carry out its respiratory function either in the oocyte or may be used to support embryonic development.

After external fertilisation, males provide parental care to the fertilised eggs in underground burrows (Kerr 1900). Stagnant waters where *Lepidosiren* nests have been found are characterised by low oxygen levels, ranging from 0.2 to 1 cm^3 per litre of dissolved O_2 (Carter 1931). Fertilized *Lepidosiren* eggs can reach 7 mm in diameter and, given the low oxygen conditions of the nests, it was unclear how *Lepidosiren* eggs obtained sufficient oxygen supply to sustain development. *Lepidosiren* males develop pelvic fin filaments during the breeding season, which break off and degenerate after larvae hatch and leave the nest (Kerr 1900). Initially, it was suggested that these filaments could contribute to copulation as spawning brushes to spread seminal fluid. However, given the overall similarity between pelvic fin filaments and external gills, it was also proposed that pelvic fin filaments were used to aerate the eggs by releasing oxygen from the male's blood into the water (Cunningham 1929). Recently, however, it was shown that *Lepidosiren* pelvic fin filaments neither have the morphology compatible for oxygen diffusion nor the gene expression profile typical of gill filaments (Lima et al. 2017). Therefore, it is conceivable that GbE may help to extract O_2 from the water to support the development of the embryo, which may be considered analogous to the function of the embryonic Hb in tetrapods (Dickerson and Geis 1983).

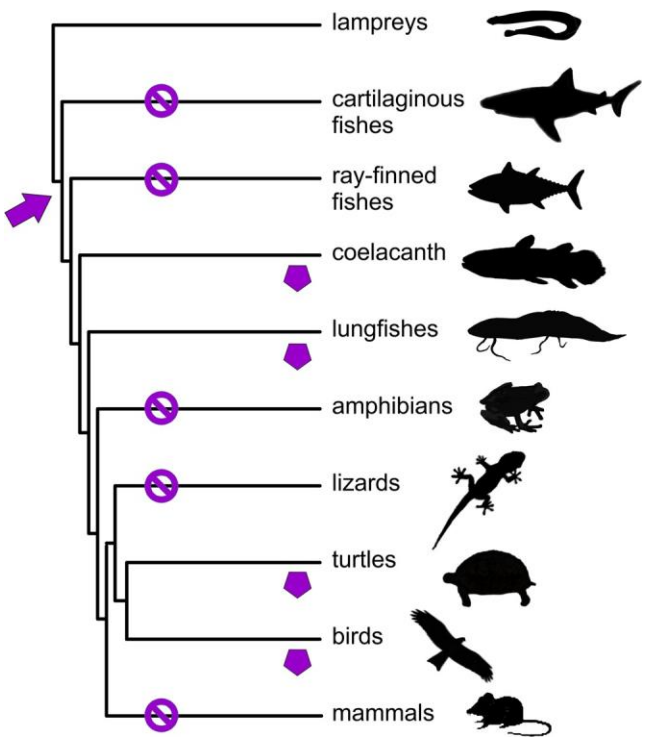


Figure IV.7: Occurrence of GbE in vertebrates.

GbE (purple pentagon) genes are present in the genomes of the coelacanth, lungfishes, turtles, and birds, but has been lost in other vertebrate taxa (indicated by ∅). The arrow indicates the origin of GbE, i.e., its divergence from Mb. Genomes were searched by BLAST at ENSEMBL (https://www.ensembl.org/) and NCBI (https://www.ncbi.nlm.nih.gov/genome/).

In contrast to the lungfish *Mb* genes, which functionally diversified after duplication (Koch et al. 2016), the multiple *GbE* genes most likely encode proteins with similar functions. The presence of multiple palogous *GbE* genes thus enhance *GbE* mRNA levels, as evident by the very high RPKM values between 20,000 and 63,000 (depending on the species). These values exceed the RPKM of Mb in the muscle or heart (Fig. IV.3; Supplemental Information Figs IV.5–IV.7; Supplemental Information Table IV.3) and are even higher than those of *Hb* in blood. Although we do not know the protein levels, the data suggest that GbE has an Mb-like role and contributes to the O_2 supply of the oocytes. This hypothesis is supported by the notable GbE protein bands in the Coomassie-stained gels (Supplemental Information Fig. IV.10), and by the O_2-binding equilibrium curve, with a P_{50} similar to that of a typical vertebrate Mb (Fago and Jensen 2015), as well as by the low autoxidation rate of GbE. In addition, deoxy LpaGbE1 shows nitrite reductase activity similar to that of vertebrate Mbs (Fago and Jensen 2015), indicating that GbE may also contribute to nitrite-dependent NO generation and signaling pathways in the ovary during periods of hypoxia. Other functions of GbE, for example as a vitellogenin-like storage protein, cannot be excluded but are less likely.

The globin family is a classic example of subfunctionalisation and neofunctionalisation of genes after duplication (Ohno 1970). This notion is supported by the existence of eight distinct globin genes in vertebrates, which emerged early in evolution (Burmester and Hankeln 2014). More recent and lineage-specific amplification events of members of the globin family are mirrored by the multiple *Mb* (Koch et al. 2016) and *GbE* genes (this study) of lungfish, which is unusual among vertebrates. The phylogenetic tree (Fig. IV.2) showed that amplification of *GbE* genes commenced already in the lungfish stem lineage (*i.e.*, separation of the GbE1 and GbE2 clades), but then occurred independently within the orders *Lepidosiren* and *Protopterus*. Lungfish are well-known for their gigantic genomes, which may, in theory, also result in the amplification of gene numbers. This explanation is unlikely for the globin genes. We only observed multiple copies of *Mb* and *GbE*, but of no other globin gene. In fact, lungfish have lost *Ngb* and *Cygb* genes. It is possible that their functions have been taken over by *Mb* copies (Koch et al. 2016). The *GbE* genes seem to have an oocyte-specific function in lungfish, which is – to the best of our knowledge – not mirrored by any other globin-type in another vertebrate.

Although additive effects of the expression of multiple *Mb* genes in lungfish may be important to increase O_2 supply, it is more likely that the different Mbs carry out distinct functions. This hypothesis is supported by the distinct expression patterns of the *P. annectens* Mbs, as well as by their different abilities to transfer stress tolerance (Koch et al. 2016) and O_2 binding properties (J. Lüdemann, A. Fago, T. Burmester, unpublished). By contrast, we propose that the multiple *GbE* genes have additive functions. This is evident by their expression in a single tissue, the ovary, as well as by the high similarity of the sequences. While Mb amino acid sequences may differ up to 70%, the maximum divergence of the lungfish GbEs is 40%. We further propose that lungfish GbE has a similar function in the ovary as Mb in the heart and skeletal muscle and may thus provide additional O_2 either by enhancing O_2 storage or by facilitating intracellular O_2 diffusion.

There is little doubt that the globins originated from a single ancestor, and multiple gene duplications have led to their functional diversification (Storz et al. 2011, Storz et al. 2013, Burmester and Hankeln 2014). Phylogenetic analyses suggest that all eight globin types were present in the gnathostome ancestor and that subsequent losses in certain clades have shaped the present globin repertoire of the vertebrate taxa. Globin E has been lost in all vertebrate clades except birds, turtles, lungfish, and coelacanth (Fig. IV.7). Although database searches did

not reveal any other GbE sequences, we cannot rule out that in future additional *GbE* genes will be discovered that are expressed at unexpected sites.

The present data suggest that GbE has a similar role as Mb in cellular O_2 supply. Mb is still considered to exert its function mainly in muscle, but at least in non-tetrapods it displays a more widespread expression pattern in many tissues (Fraser et al. 2006, Cossins et al. 2009, Opazo et al. 2015, Gallagher and Macqueen 2017). GbE and Mb share an ancient common ancestry (Fig. IV.7). It is possible that GbE originated as a variant of Mb that meets the particular needs (for example, regarding O_2 affinity) for a specific tissue. It remains unknown whether the original function was in the oocyte (like in lungfish) or the eye (like in sauropsids). In future, the coelacanth may be helpful to decide on the direction of evolutionary change but none the currently available transcriptomes of this taxon derives from eye or ovary. However, it is evident that either in the lungfish or the sauropsid clade, GbE must have changed its expression site. For unknown reasons, GbE was independently lost in at least five vertebrate clades (Hoffmann et al. 2011) (Fig. IV.7).

4.5 Methods

4.5.1 Lungfish material

A *L. paradoxa* specimen was collected near Breves, Brazil, and euthanised with a lethal dose of tricaine methanesulfonate. This study was approved by IBAMA/SISBIO under license number 47206-1, and experimental procedures and animal care were conducted in accordance to the Ethics Committee for Animal Research at the Universidade Federal do Pará, under the approved protocol number 037-2015.

West African lungfish (*P. annectens*) and marbled lungfish (*P. aethiopicus*) were obtained from a pet shop, euthanised in 1 g/l tricaine methanesulfonate and finally killed by decapitation. Tissues were removed and immediately stored frozen at −80 °C in RNAlater (Qiagen, Hilden, Germany). Animals were treated in accordance with the German Animal Welfare Act.

4.5.2 RNA extraction, library preparation and Illumina sequencing

Total RNA was purified from lung, brain, buffy coat and ovary of South American Lungfish (*L. paradoxa*) and extracted for transcriptome using TRIzol Reagent (Life Technologies, Cat. 15596-026) according to the manufacturer's protocol. RNA samples were further purified using

RNeasy Mini Kit (Qiagen) and treated with DNaseI (Qiagen), according to the manufacturer's protocol. Reference transcriptome and transcript abundance estimation were obtained from each library, sequenced on an Illumina Hiseq platform with 100 bp paired-end reads. Sequencing was carried out by a commercial service from Instituto Nacional do Câncer, Brazil.

RNA was purified from the ovary of the marbled lungfish (*P. aethiopicus*) using the CRYSTAL RNA Mini Kit (biolab products). A library for paired-end sequencing was generated from ~1 µg total RNA. Sequencing of 2 × 150 nt was performed by with the Illumina NextSeq500 technology (StarSEQ, Mainz, Germany).

Additional transcriptomes from various lungfish and other species were retrieved from the public SRA data- base at GenBank (for accession numbers, see Supplemental Information Table IV.1). The transcriptomes from each lungfish species were assembled using either the CLC Genomics Workbench (version 11.0.1) or Trinity v2.6.5. Globin cDNA sequences were identified in the assemblies employing TBLASTN using the coelacanth globins (Schwarze and Burmester 2013) as queries. The final assignment of a globin to a specific clade was done by phylogenetic analyses (see below). The consensus sequences were verified using a backmapping approach using the CLC Genomics Workbench. Broken mate-pairs were employed to identify putative misassemblies. If required, the reads were re-assembled with different parameters, and the procedure was repeated until unambiguous *GbE* cDNA sequences were obtained. Selected GbE sequences were verified by RT-PCR and Sanger sequencing of the cDNA (GATC, Konstanz, Germany). RNA-Seq analyses were performed with the CLC Genomics Workbench. The mRNA levels of the globins were calculated as RPKM.

4.5.3 Sequence analyses and phylogenetic inference

The vertebrate genomes available at ENSEMBL (https://www.ensembl.org/) and NCBI (https://www.ncbi.nlm.nih.gov/genome/) were searched for the presence of *GbE* genes. The lungfish GbE amino sequences were included in an alignment that covers the whole range of vertebrate globins and a broad range of classes (Schwarze et al. 2014, Schwarze et al. 2015, Koch et al. 2016) (Supplemental Information Table IV.2). Adgb was excluded from the phylogenetic analyses because of its permutated globin domain (Hoogewijs et al. 2012). A multiple sequence alignment of the amino acid sequences was obtained with the MAFFT online tool with the L-INS-i method (Katoh and Toh 2008, Katoh et al. 2009). Phylogenetic analysis was

done with MrBayes 3.2.3 (Huelsenbeck and Ronquist 2001, Ayres et al. 2012) using the LG model of amino acid evolution (Le and Gascuel 2008), which was selected using PROTTEST (Darriba et al. 2011), and which was implemented into MrBayes with the general time reversible model as fixed prior and by specifying the aarevmatpr and statefreqpr options (Schwarze and Burmester 2013, Schwarze et al. 2015). The program was run for 5,000,000 generations using the standard option (two independent runs with four simultaneous chains). Trees were sampled every 1000th generation, and the posterior probabilities were estimated after discarding the initial 25% of the trees.

4.5.4 Quantitative real-time reverse-transcription PCR

Reverse transcription was performed with 1 µg of total RNA, oligo-(dT)18 oligonucleotides (10 µM) and 200 U SuperScriptTM II RNase H– Reverse Transcriptase (Invitrogen) according to the manufacturer's protocol. Quantitative real-time RT-PCR (qRT-PCR) experiments were carried out on an ABI 7500 Real-Time PCR system (Applied Biosystems, Darmstadt, Germany) with the "ABI Power SYBR Green Master Mix". The efficiency of the reaction was measured by the slope of a standard curve, deriving from tenfold dilutions of plasmids. Expression data were normalised according to the amount of total RNA. Further analyses were carried out employing the Microsoft Office Excel spreadsheet program.

4.5.5 *In situ* hybridisation.

Ovary tissue was extracted from *L. paradoxa* and embedded in Tissue Tek O.C.T compound (Sakura Finetek) in dry ice, and then stored in −80 °C freezer for subsequent cryosectioning. Then, 20 µm sections were produced on ColorFrost Plus microscope slides (Thermo Fisher Scientific) on a cryostat at −20 °C. Sections were fixed according to a previously established protocol (Nogueira et al. 2016). After drying at room temperature, slides were stored in −80 °C ultrafreezer. Hematoxylin (Sigma-Aldrich) and eosin (Sigma-Aldrich) staining were performed according to standard protocol. A pGEM-T vector (Promega, Mannheim, Germany) containing *GbE1* of *L. paradoxa* served as a template in PCR amplification using forward and reverse M13 primers (reverse: 5′-CAGGAAACAGCTATGAC-3′; forward: 5′-GTAAAACGACGGCCAG-3′). PCR was performed in 50 µl reaction volumes containing 39.2 µl of RNase free water, 1.5 µL of 10 mM $MgCL_2$, 5 µL of 10x buffer, 1 µL of each primer (0.5 M), 1 µL of dNTP mix (10 mM), 0.3 µL of Taq DNA Polymerase, and 1 µL of DNA template. The temperature profile consisted of

preheating at 94 °C for 3 min, 32 cycles of denaturation at 94 °C for 45 s, annealing at 56 °C for 30 s, and extension at 72 °C for 90 s, followed by a final extension step at 72 °C for 10 min. Sense and antisense riboprobes were synthesised using T7 RNA and Sp6 RNA polymerases, respectively, and DIG-labelling mix (Roche). The riboprobe reaction (Life Technologies) was performed in 20 µL reaction volumes containing 0.5 µL of RNase inhibitor, 2 µL of DTT (0.01 M), 2 µL of DIG, 2 µL of 10x reaction buffer, 5 µL of template, 2 µL of Sp6/T7 enzyme mix, and 6.5 µL of nuclease-free water. *In situ* hybridisation was performed according to a previously established protocol (de Lima et al. 2015), using 300 ng/slide of DIG-labelled riboprobe. Slides were imaged using a Nikon SMZ1500 microscope (Nikon Digital Sight DS-Ri1).

Total proteins were isolated from the ovaries of *L. paradoxa* and *P. aethiopicus*. The tissues were homogenised in 10 mM Hepes-buffer with a Minilys Personal Homogeniser (Bertin Instruments, Bretonneux, France) 3 times for 30 s. Then, DNA and RNA were destroyed by sonification (Bandelin Sonopuls, Berlin, Germany). After centrifugation at maximum speed (13,000 × g, 10 min, 4 °C), the supernatant was collected. Total protein from each ovary was separated on a 15% SDS-polyacrylamide gel. The gel was stained with Coomassie brilliant blue. Putative GbE bands were excised and analysed by liquid chromatography-mass spectrometry using a commercial service (Core Facility Mass Spectrometric Proteomics, University Medical Center Hamburg-Eppendorf, Germany).

4.5.6 Preparation of recombinant lungfish GbE protein

The coding sequence of *GbE1* of *L. paradoxa* was amplified by RT-PCR and then cloned into the pET16b expression vector (Novagen - Merck Biosciences, Darmstadt, Germany). Recombinant expression was done in *E. coli* BL21(DE3)pLysS cells (Promega, Mannheim, Germany), which were grown at 37 °C in 5 ml L-medium (1% bactotryptone, 0.5% yeast extract, 0.5% NaCl, pH 7.5) containing 10 µg/ml ampicillin, 34 µg/ml chloramphenicol overnight. The culture was applied to 500 ml L-medium supplemented with 1 mM δ-aminolevulinic acid. The culture was induced at OD_{600}= 0.4 to 0.8 by the addition of isopropyl-1 thio-D-galactopyranoside (final concentration 0.4 mM), and expression was continued at 30 °C overnight. Cells were collected by 45 min centrifugation at 4,000 g and resuspended in 50 mM Tris-HCl, pH 8.0, 1 mM $MgCl_2$, 1 mM dithiothreitol, 10 µg/ml DNase, 5 µg/ml RNase, CompleteTM proteinase inhibitor mix (Roche Applied Science) and Pefabloc (Roth). The cells were broken by three freeze-thaw cycles in liquid nitrogen followed by ultrasonication (10 × 30 s). DNA and RNA were digested for 2 h

at 37 °C. The cell debris was removed by centrifugation for 1 h at 4 °C at 4,500 × g. The recombinant GbE1 was purified from the supernatant using His60-Ni columns (Qiagen) according to the manufacturer's instructions. The His-tag was removed by incubation with the Factor Xa protease (20 µg/ml) for 6 h at 37 °C. After inactivation of the protease with 2 µM dansyl-glu-gly-arg-chloromethyl ketone, the recombinant GbE1 protein was brought to 10 mM Hepes, pH 7.8, 0.5 mM EDTA.

Gel-filtration was carried out with an ÄktaPure chromatographic system (GE Healthcare, Freiburg, Germany) equipped with a Superdex 75 10/300 column (GE Healthcare). The proteins were eluted with 50 mM potassium phosphate, 0.5 mM EDTA, pH 7.0, 0.15 M NaCl at a flow rate of 1 mL/min. The absorbance was measured at 280 and 415 nm. Recombinant GbE1 was applied at a concentration of 0.2 mM heme; Human Hb and horse Mb were employed as references at the same concentration.

4.5.7 Spectroscopic studies and O_2 binding curves

Absorbance spectra of purified GbE1 were taken in the range 350–650 nm. The deoxy-form was obtained by adding sodium dithionite. The determination of O_2 equilibrium curves was done using a modified diffusion chamber technique previously described (Sick and Gersonde 1969, Weber 1981, Weber 1992). Samples were measured in at least duplicates of 5 µL (~0.2 mM heme) in 0.1 M Hepes, 0.5 mM EDTA, pH 7.2 at 20 °C. Ferric GbE1 was reduced for 5 to 10 min in N_2 with the met-Hb reductase system (Hayashi et al. 1973).

In the modified diffusion chamber technique, water-saturated gas mixtures of O_2 and ultrapure (>99.998%) N_2 were generated by GMS 500 gas mixing system (Loligo System, Denmark) and used to equilibrate the thin smear of 5 µL GbE1 sample with stepwise increments of oxygen tension (PO_2). Absorbance traces were sampled at 436 nm by a photomultiplier (model RCA 931-A) and an Eppendorf model 1100 M photometer. The signal was digitalised, and saturation values were obtained using an in-house software (Beedholm 2018). P_{50} and cooperativity values were calculated from the zero intercept and slope of Hill plots, respectively. Each curve consisted of four to six saturation steps.

4.5.8 Nitrite reductase activity

The reaction of recombinant deoxy GbE1 (10 µM heme in deoxygenated 50 mM HEPES, pH 7.13) with nitrite (0.1 mM) was carried out under pseudo-first-order conditions at 20 °C (Fago et al. 2017, Fago et al. 2018). Before adding nitrite, the GbE1 was anaerobically titrated with 1 mM of sodium dithionite in a 1 cm cuvette sealed with a rubber cap. The measurement of the reaction kinetics was started immediately, and absorbance traces were recorded at 435 nm and 20 °C.

V. Discussion

Gene duplications drive the evolution of organisms and make physiological adaptation and adaptive radiation possible. The redundancy after gene duplications allows to have mutations in one gene-copy together with another, functional copy. The mutant copy can fulfil new features compared to the previous one and perform completely different functions to the original gene if it benefits the organism. This results in multiple gene copies or whole families of genes, like the globin-family. The results of this thesis give hints to the unexpected diverse globin repertoire and try to explain the complexity of globin evolution in lungfish.

The number of published transcriptome sequences increases rapidly and gives the possibility to investigate the expressions of genes or gene-families in different species. The new availability of a transcriptome from the West African lungfish (*Protopterus annectens*) was my starting point to examine the globin repertoire of the lungfish as the closest living relative of tetrapods (Brinkmann et al. 2004, Amemiya et al. 2013, Biscotti et al. 2016). At the beginning of this study, at least nine different Mb sequences in the transcriptomes of *P. annectens* were unexpectedly observed (see Chapter 1 and 2). This is very unusual, as most vertebrates harbor only a single Mb gene. The nine newly identified *Mb* genes diverged early in the evolution of lungfish and reflect an enhanced evolutionary rate (see Chapter 1). Further investigations also revealed multiple GbE copies expressed in the ovary of all investigated lungfish (*L. paradoxa*, *P. annectens*, and *P. aethiopicus*). This is the first time observing GbE not being expressed in the eye but in a different tissue (see Chapter 3). No other globin gene except *Mb* and *GbE* was found amplified in the lungfish; rather, *Ngb* and *Cygb* seemed not expressed at all. The unexpectedly high Mb and GbE diversity was a subject to closer investigations.

5.1 Divergence of lungfish genera

The emergence of the multiple *Mb* and *GbE* genes is a possible adjustment that is specific to the lungfish. Lungfish are living fossils that show exceptional adaptation features. They can withstand hypoxia, survive outside of water for a long period of time, and the Lepidosireni-formes can aestivate in a mud cocoon during summer time (see 1.10.1) (Fig.V.1).

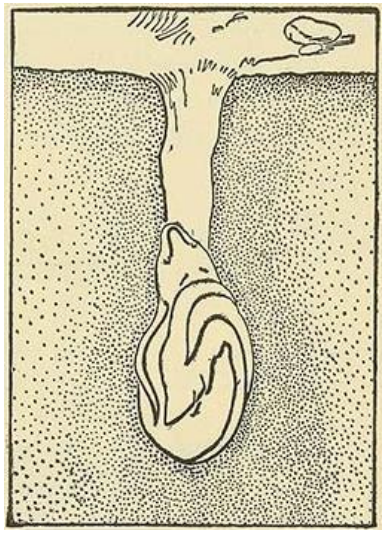


Figure V.1: African lungfish aestivate in a mud cocoon

Modified from (Hilzheimer 1913)

South American (Lepidosirenidae) and African lungfish (Protopteridae), in contrast to the Australian lungfish (Neoceratodontidae), have evolved several metabolic and physiological abilities that resemble amphibians, such as the ability to aestivate in a mud cocoon (some amphibians), their ventilator control, the existence of a real lung, having limb like fins and their ability to breathe air (Wilson 1959, Janssens and Cohen 1966, Heisler et al. 1982, Sanchez 2001, Secor and Lignot 2010). Due to the fact that Lepidosireniformes (*Lepidosiren* and *Protopterus*) are more recently derived than *Neoceratodus*, these advanced adaptations shared with amphibians could also be viewed as examples of convergent evolution. The presence of a true lung in Lepidosireniformes together with their ability to aestivate makes these fish an important biological model to study physiological responses. Besides that, their huge genome size makes this group also interesting from another perspective. The questions arising are: Does the size of the genome matter? And are organisms with a very large genome more specialized in case of adaptation? Or do they have more gene multiplications?

Lungfishes have the largest genomes throughout the animal kingdom with C-values up to 132 pg per haploid genome (*P. aethiopicus*) which makes it around 40-fold the size of the human genome with a C-value of 3.5 (Gregory 2018). To the best of our knowledge, only *Mb* and *GbE* and no other globin gene show signs of amplification in lungfish, hence there is no evidence for polyploidization. First insights into the lungfish genome reveal at least 50 % of the genome as repetitive elements (as communicated by Prof. Dr. Axel Meyer, University of Konstanz, Germany). Therefore, it is unlikely that these globin multiplications are linked to their

large genome size. Ngb and Cygb, as deduced from the RNA-Seq data, got probably lost in the lungfish and it is well possible, that their functions have been taken over by the multiple Mb copies. The lungfish is the first known vertebrate not harboring a Cygb gene. This is unique among vertebrates and requires physiological explanations. But to answer the upcoming questions, the exact functions of Cygb need to be investigated first.

The lungfish has an extensive fossil record which dates back to the Devonian around 380 million years ago, indicating that they lived in both freshwater and shallow sea environments during that time. Lungfish can be considered as living fossil based on the retention of particular anatomical forms and ways of life (Eldredge 1984).

5.1.1 Air-breathing is a successful adaptation of fish

The South American and the African lungfishes differ from other air-breathing fish because of the existence of a real lung and their ability to aestivate (see 5.1.2) (Johansen 1966, Johansen and Lenfant 1967). The only other fish species with a real lung are the *Polypteridae*, but they are not able to aestivate. Typically, aestivation is an adaptation to hypoxic conditions (see 1.10.1). In connection with that, the lungfish are profoundly tolerant of low oxygen supply (Johansen 1966, Johansen 1970, Lahiri et al. 1970). Overall, air-breathing is an ancient vertebrate specialization that appeared in fish long before the evolution of amphibians and the invasion of land by the tetrapods in the Late Silurian Period. Air breathing has persisted throughout the evolutionary history of fishes and is viewed by many as the primary event that led to terrestriality (Graham 1997). To date there are over 70 fish species from twelve families that are able to breathe air while emerging from water. Most of the modern air-breathing fishes live in tropical freshwater regions that are subject to seasonal drying in summer time (Graham 1997). The water heats up and photosynthesis is not sufficient to maintain normoxia conditions in the water which leads to hypoxia or even anoxia in this environment. Additionally, and counterproductively, oxygen consumption and the metabolic rate of the fish go up because of the increased temperature. This common scenario leads to a natural selection pressure and provided a number of successful and different specific developments of air breathing in fishes (Graham 1997).

5.1.2 Aestivation is a special feature of lungfish

The Lepidosireniformes can aestivate, which means, they survive periods of seasonal drying buried in a mud cocoon in the ground (Fig. V.1, see 1.10.1). This ability is not shared with *Neoceratodus*. The Lepidosireniformes can undergo aestivation in nature by entering into a state of corporal torpor at high temperatures. During that time, the lungfish buries itself in a subterranean mud cocoon and can survive without food and water uptake for up to five years (Hiong et al. 2013). Aestivation could also be observed in the laboratory in pure dried mucus cocoons inside of plastic boxes (Chew et al. 2004). During aestivation the lungfish reduces the energy metabolism, turns on cytoprotective genes (such as the *cold inducible RNA-binding protein* and the *glucose regulated protein 58*) and increases its urea production to reduce the ammonia toxicity and to prevent high water losses (Chew et al. 2004, Hiong et al. 2013). Aestivation occurs in three phases: 1) Induction – Environmental cues in the aestivating lungfish are turned into internal signals to induce the necessary changes. It hyperventilates and secrets mucus, to build a dried mucus cocoon within six to eight days. 2) Maintenance – The lungfish is completely encased in the mud cocoon and the food intake and any locomotor activity is stopped. The lungfish needs to prevent cell death; it has to preserve the biological structures and to sustain a slow rate of waste production. 3) Arousal – With the return of the water the lungfish starts arousing from aestivation. The lungfish has to excrete the accumulated waste products and to feed for growth and repair (Chew and Hiong 2014). During aestivation, there are more genes upregulated than downregulated in the lungfish brain compared to the non-aestivating state, indicating a transcriptional and metabolic activity of the brain during aestivation (Hiong et al. 2013). The finding of lungfishes responding to sensory disturbances supports the assumption that the brain is partly active (Delaney et al. 1974). The brain probably maintains certain levels of metabolic and physiological activities (Fig. V.2) and therefore needs to be constantly supplied with oxygen to sustain the basal metabolic rate (Delaney et al. 1974). This assumes a possible function of the different Mbs highly expressed in the lungfish brain could be to protect it from damage. The big deposit of Mb helps to maintain the O_2 supply in this most critical tissue when the O_2 intake by breathing is not sufficient (see 5.4).

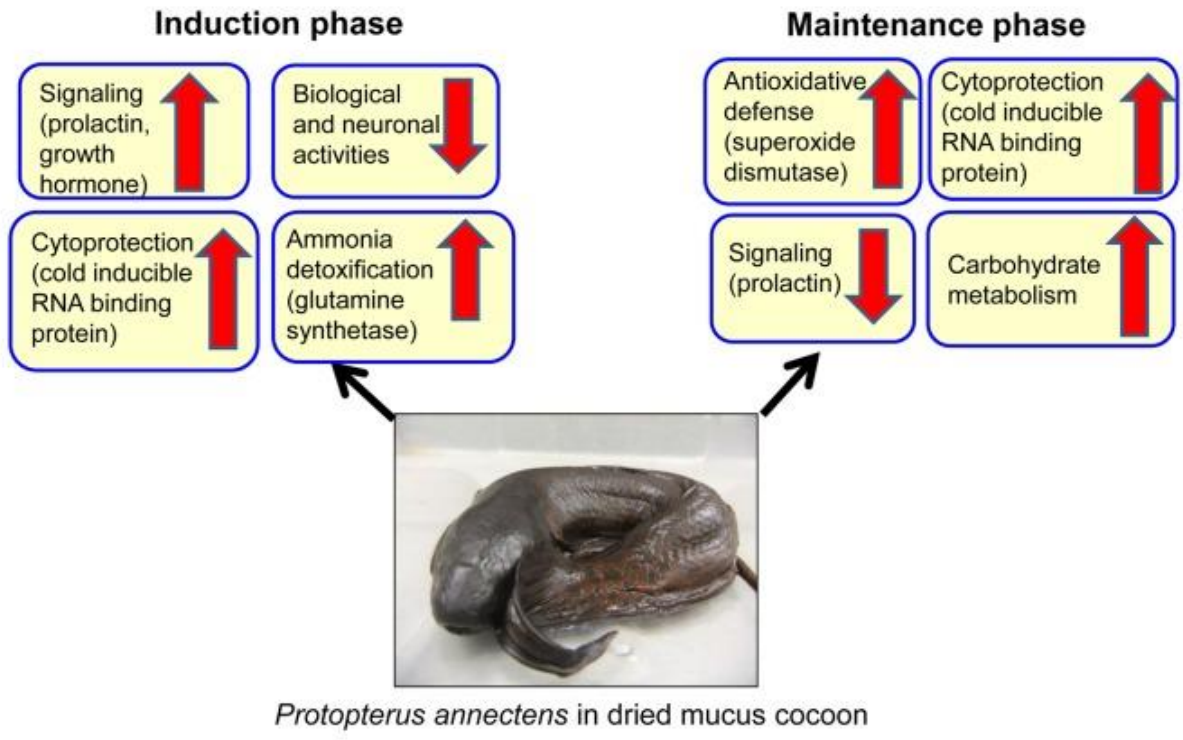


Figure V.2: Changes occurring in the brain of *Protopterus annectens* aestivated in a dried mucus cocoon during aestivation
Changes occurring during the induction and the maintenance phase (Modified from(Chew and Hiong 2014))

5.2 Multiple Mb and GbE genes were detected in lungfish transcriptomes

Knowing that there are multiple Mb copies in the West African lungfish (*P. annectens)*, newly available lungfish transcriptomes were examined in this study, which revealed multiple Mb copies as well. The two African lungfishes *P. dolloi* and *P. aethiopicus*, as well as the South American lungfish *L. paradoxa*, showed a nearly orthologous diversity of *Mb* genes in their transcriptomes. *P. dolloi* harbors six Mbs, *P. aethiopicus* five Mbs and *L. paradoxa* reveals five distinct Mb sequences (see 3.4.1). There must be a reason for the occurrence of multiple *Mb* genes in all of the investigated lungfish species. As the diversification dates back a long time, and is still maintained, all of the different Mbs must have specific functions in the life of lungfishes. In contrast to *Neoceratodus*, the Lepidosireniformes aestivate during the dry summer periods and have to deal with low and changing O_2 conditions in their environment (see 5.1.2). Interestingly, all multiple globin copies discovered so far were found in the Lepidosireniformes, however, a transcriptome of *Neoceratodus* is highly needed to confirm this. To date, there is no available *Neoceratodus* transcriptome in order to assess globin diversity in these species as well.

Searches in different databases revealed that the Mb sequences are restricted to the lungfishes, and phylogenetic analyses from this study showed, that they all derived from the same ancestral copy and form a single clade, which is in a sister group position to the tetrapod Mbs (Fig. V.3).

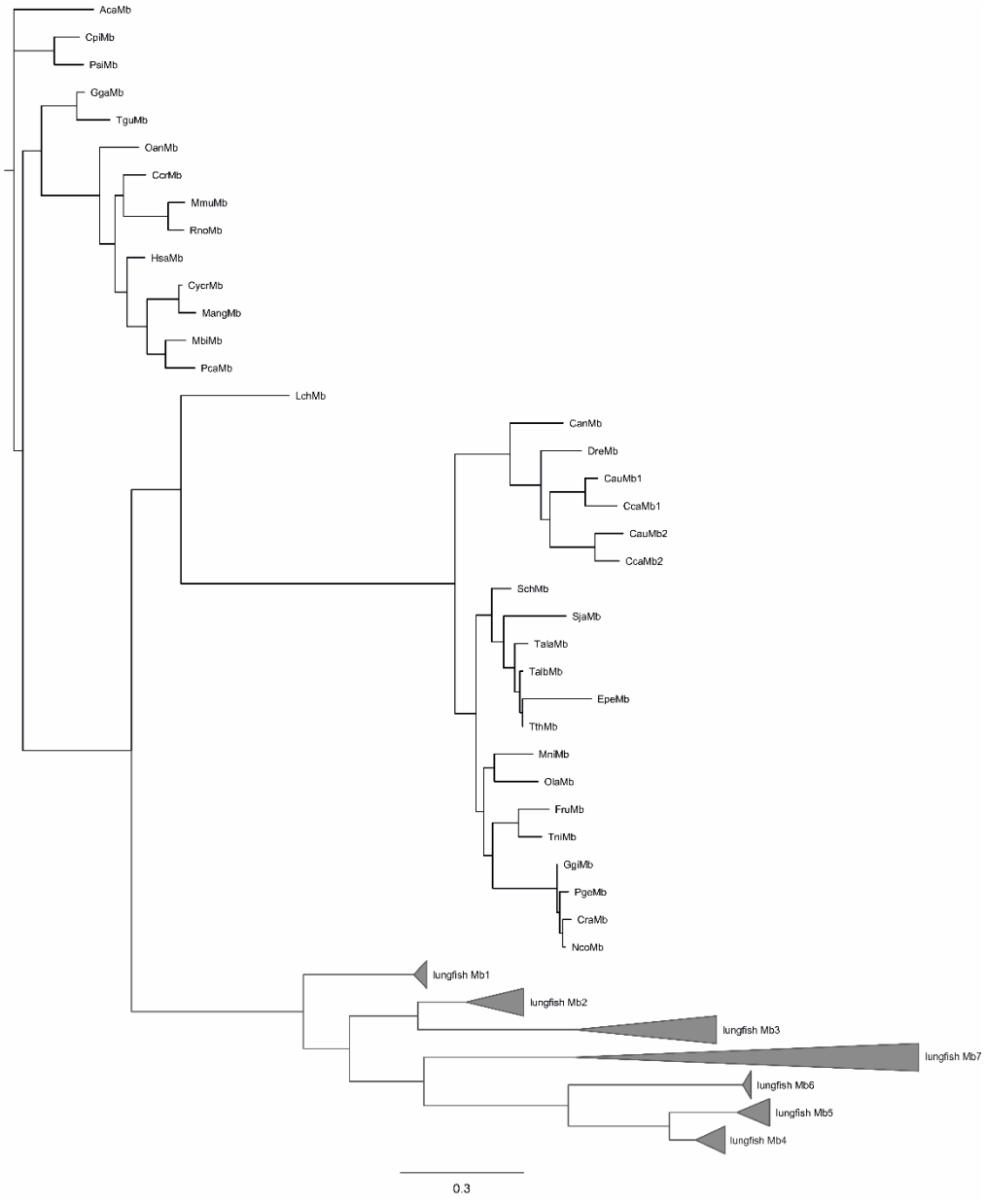


Figure V.3: Phylogenetic Tree of different vertebrate Mbs

Tree reconstruction was carried out with the amino sequences assuming the JTT model of protein evolution (Sick and Gersonde 1969). The bar represents 0.3 substitutions per site. For the abbreviations and accession numbers, see Table SIII.1. The size of each triangle is proportional to the number of myoglobins in the respective group.

As demonstrated in this thesis, expression analysis of various *P. annectens* tissues displayed the distinct Mbs to be differentially expressed with an unexpected main expression in the lungfish brain (see 2.3.3). Followed RNA-Seq analyses on three *P. annectens* transcriptomes and mass

spectrometry from bands of an SDS-PAGE confirmed these results (see 2.3.3). Oxygen binding analyses revealed that all the PanMbs reversibly bind O_2, but have different O_2 affinities and cooperativity values (see 3.4.2, Table III.1). Also, the nitrite reductase activity differed between the PanMbs, leading to the assumption of different functions (see 3.4.3). Nonetheless, all the PanMbs showed approximately the same molecular mass and the same hydrodynamic radius and similar radii of gyration. CD spectra indicated a conserved secondary structure, which fits to the conserved amino acids in their sequence (see 3.4.4). Additionally, cell culture experiments in mouse neuronal cells revealed a protective function of the *P. annectens* Mbs against hypoxia via enhanced cellular survival and reduced ROS production (see 2.3.4).

This unusual diversity of the Mbs in lungfishes hints to a functional diversification of these genes. The results of this thesis reveal that some of the Mbs seem to work as O_2 storage to supply it to the mitochondrial respiratory chain. Other Mbs may be working as ROS scavenger or nitrite reductases under hypoxic conditions. The Mbs show the same, conserved, highly α-helical globin-fold, but display different expression sites and enzymatic properties. Some of the Mbs may have adapted to the tissue-specific O_2 requirements, others may have taken over the functions of Ngb and Cygb, which are missing in lungfish as indicated by RNA-Seq data.

Additionally and unexpectedly, a *de novo* assembly conducted in this study of newly submitted transcriptomes of the South American lungfish *L. paradoxa* revealed multiple contigs that resembled the sauropsid GbE. These sequences all derived exclusively from the ovary tissue, which is very unusual and unexpected as GbE was thought to be a specific ‘eye-globin’ (see 4.3.3). This is the first time that multiple *GbE* genes were identified and also the first time that GbE was detected in another tissue than the eye. Besides GbX in *Danio rerio* (Tiedke et al. 2011), it is the only ovary-specific globin detected so far (Chapter 3).

All other available lungfish transcriptomes from different tissues (see Supplemental Information Table IV.1) were analyzed as well and confirmed the ovary as the exclusive expression site of GbE. Due to the limitation of available ovary transcriptomes, GbE could only be detected in *L. paradoxa*, in *P. annectens* and in *P. aethiopicus* so far, but all specimens displayed multiple GbEs expressed in their ovaries. In *L. paradoxa*, six distinct GbE sequences were found, *P. annectens* revealed five distinct GbE sequences and *P. aethiopicus* revealed seven (see 4.3.3). Phylogenetic analyses done in this thesis show that the lungfish GbEs fall into two clades, of which one is formed by GbE1 sequences, and the other one by GbE2a-e from

L. paradoxa along with GbE2 sequences from the *Protopterus* species (Fig IV.2). The overall expression values of GbE in the ovary are very high, a lot higher than Mb in the muscle and Hb in the blood of *L. paradoxa* (Fig. IV.3; Supplemental Information Figs. IV.5-IV.7; Supplemental Information Table IV.3).

RT-PCR and qRT-PCR analyses validated the ovary-specific expression of LpaGbE1, as it was exclusively detected in the ovary, and not in the eye or any other tissue. GbE proteins could also be described using mass spectrometry. In cooperation with Igor Schneider (Universidade Federal do Pará, Brazil) it was shown in this study, that the LpaGbE1 mRNA is localized in previtellogenic oocytes (see 4.3.4). Oxygen binding studies revealed that LpaGbE1 is able to reversible bind O_2 and, in addition, has a quite high (18.72 s^{-1} M^{-1}) nitrite reductase activity (see 4.3.5). These results again manifest certain similarities with the lungfish Mbs. Both, Mb and GbE, have similarly high nitrite reductase activities and both are able to reversibly bind O_2, suggesting similarities in their function.

5.3 Mb and GbE form a sister-group

Most phylogenetic analyses display GbE and Mb as being closely related groups (Fig.V.4) (Blank et al. 2011a, Hoffmann et al. 2011, Storz et al. 2011, Storz et al. 2013). In previous studies it was shown that there is a conserved synteny in the gene regions of *GbE* and *Mb* of the coelacanth which are linked via the genes NFAM1 and TCF20 (Schwarze and Burmester 2013). Therefore, it is very likely that GbE and Mb evolved from a common ancestor through gene duplication followed by subfunctionalization events. This hints to the already earlier proposed similar functions of these two globin genes (Blank et al. 2011a). It is well possible that GbE originated as a variant of Mb meeting the particular needs of a specific tissue. The results of the present study support this theory, as these two globins reveal similar oxygen binding kinetics and nitrite reductase activities (Table III.1 and see 4.3.5). Additionally, phylogenetic analyses using all known GbE-sequences, including the newly identified lungfish GbE-sequences, together with other globin sequences from lungfish and other vertebrates confirmed the close relationship between GbE and Mb sequences (Fig.IV.2).

Furthermore, it is very striking, that only the *Mb* and *GbE* genes are multiplied in the lungfish. Therefore, one has to ask why there are multiple globin copies in the lungfish, what is their function and how did they emerge?

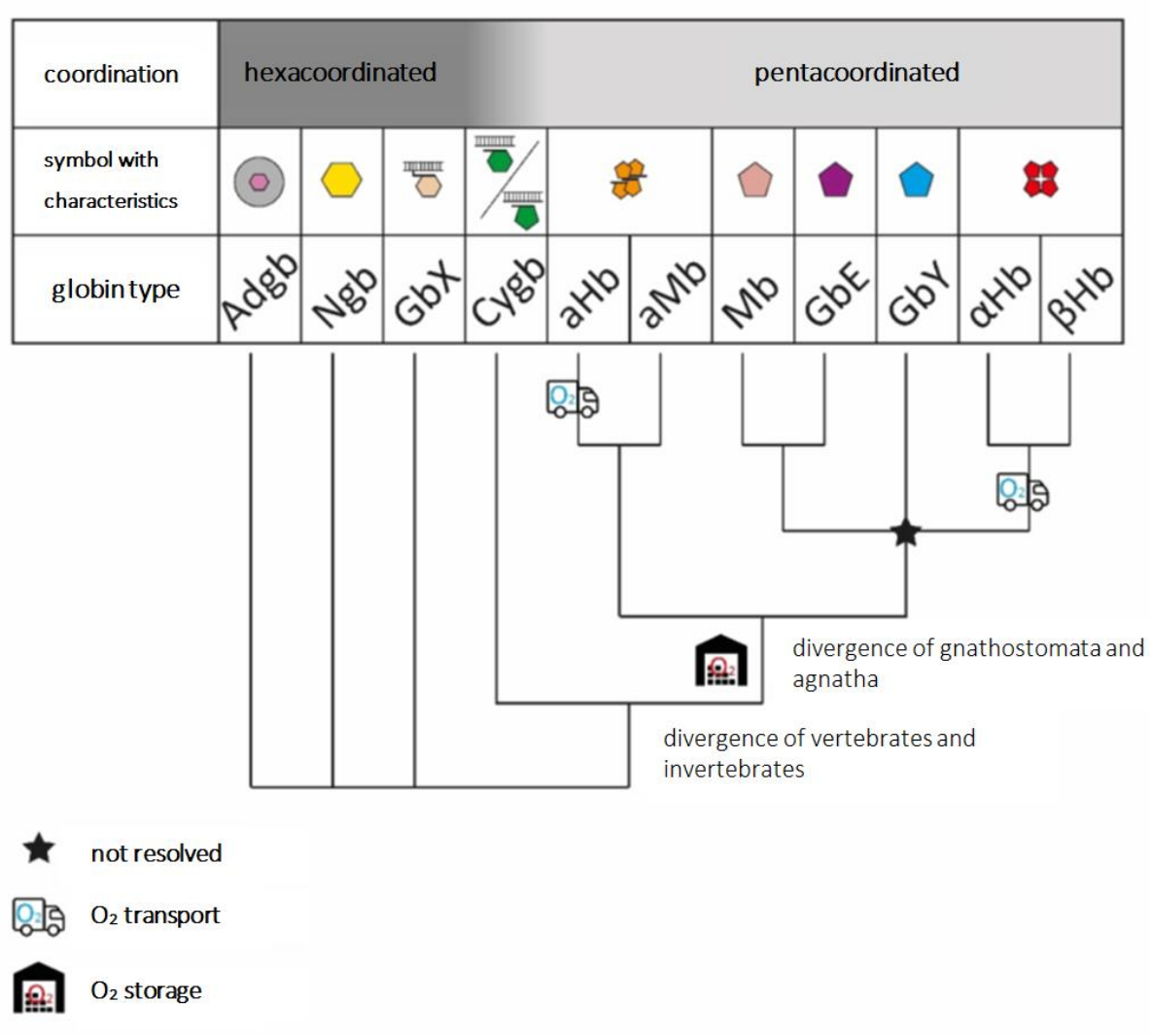


Figure V.4: Schematic phylogeny and characteristics of vertebrate globins

Simplified phylogenetic tree of the vertebrate globins. The hexagon displays hexa-coordination, the pentagon penta-coordination. GbX is membrane-bound, whereas Cygb is able to bind lipids. In *Danio rerio*, there are two Cygb proteins. One of the Cygb is hexa-coordinated, the other Cygb is penta-coordinated (Corti et al. 2016). In all other species investigated so far, Cygb is hexa-coordinated. The position of GbY is not fully solved which is marked with a 'star'. Also shown here is the independent emergence of O_2 transport and O_2 storage. (Modified from (Rohlfing 2016))

5.4 Mb-Gene expansion protects lungfish brain from hypoxia

The multiple lungfish Mbs display a similar diversity and have orthologous myoglobin genes. Most of these orthologs have been maintained for > 100 million years, suggesting that the lungfish requires specific Mbs rather than just multiple copies. Only in 4 cases additional paralogs and only a few losses of a specific Mb, like Mb1 in *L. paradoxa* or Mb 7 in *P. aethiopicus*, were found (see 3.4.1 and Fig. III.1). The different Mbs show strong differences in their amino acid structure together with highly similar secondary structures and essentially the same high content of α-helices. Kinetic and enzymatic function analysis revealed different characteristics, probably resulting in different functions of the various Mbs (see Chapter 2).

They have differences in their O_2 binding capacities and their nitrite reductase activities without a common connection. This also supports the idea of different evolutionary mechanisms that can control the heme redox-like nitrite reductase, and non-redox activities, like O_2 binding, separately, depending on the need of the tissue or the cell (Fago et al. 2018).

There is, without any doubt, a functional differentiation of lungfish Mb in evolution. The PanMbs all appear to have different, specific O_2 affinities meeting the individual requirements in each of the tissues. It appears that the other Mbs differentiated, after multiple gene duplications, to cope with the specific needs of the lungfish. The muscle-specific Mb1 is in a sister group position to all the other lungfish Mbs (Fig. V.4), which leads to the hypothesis that the O_2 supply to the muscles was the original function of Mb and still is the classical Mb function in most vertebrates. Additionally, PanMb1 is shown to have the lowest O_2 affinity of all the PanMbs (see Table III.1). This seems to be suitable for the efficient delivery of O_2 to the mitochondria. Supporting this assumption, PanMb1 enhances the activity of mitochondrial dehydrogenases under hypoxia tested in cell culture (Chapter 1). The total Mb concentration in skeletal and heart muscles of *P. annectens* is at least ten times lower than in the brain (see Chapter 1). The brain-specific PanMb2 has the highest O_2 affinity of 0.88 Torr, indicating the need of Mbs with higher O_2 affinity in neuronal tissue than in muscle tissues (see Table III.1). It is well conceivable that the high Mb concentrations in the neuronal tissues protect the lungfish brain from hypoxia during aestivation as a vertebrate brain is very sensitive towards O_2 reduction (see 5.1.2). This is supported by this thesis results which demonstrate the protective role of Mb from ROS in neuronal cell culture (see Chapter 1), suggesting that the multiple Mbs in the brain probably work as a ROS scavenger as well. Overall, there is a significant and important role of the brain to control a whole-body response during the different phases of aestivation in lungfish and to prevent it from apoptosis (Chew and Hiong 2014). This high Mb expression in the brain has previously only been observed in the carp and the goldfish (Fraser et al. 2006, Roesner et al. 2008). To survive the long periods of aestivation, the lungfish has to shut down most of the metabolic processes, but the brain needs to be active to control the phases of aestivation and to keep the lungfish alive. There is no doubt that the brain of the lungfish is the most critically important tissue during the different phases of aestivation. In the non-aestivating carp, on the other hand, the Mb distribution is significantly different. The Mb1 in the carp has a higher O_2 affinity (1 Torr at 25°C, pH 7.4) than the brain-specific Mb2 with 1.7 Torr under identical conditions (Helbo et al. 2012).

Nevertheless, nearly all PanMbs are highly efficient as nitrite reductases, which results in reduced respiration rate by generating NO under hypoxia. Additionally, this may help to protect hypoxic tissues by increase in vasodilation and the blood supply. It is known, that nitrite and Mb interact in a dose-dependent manner to increase vasodilation under hypoxia. Even modest hypoxia leads to a significant nitrite-dependent vessel relaxation (Totzeck et al. 2014). NO may furthermore inhibit the mitochondrial respiration, contribute to metabolic depression in aestivation and limit the generation of ROS. The Mbs of other vertebrates have much lower nitrite reductase activity, suggesting that this enzymatic activity is important specifically for the lungfish. Only PanMb6a/b has no nitrite reductase (see Table III.1). The amino acid sequence points out that PanMb6a/b has an E7 histidine to glutamine substitution. This substitution is also present in elephant Mb (Romero-Herrera et al. 1981), which reacts with NO 400 time faster than sperm-whale Mb (Sharma et al. 1987). This E7 substitution seems to be also responsible for the negligible nitrite reductase activity of PanMb6a/b, as it is, at least to our knowledge, the only apparent difference to the other PanMbs. Even though it cannot be the only reason, as elephant Mb does not miss the nitrite reductase activity.

Similar to the lungfish, deep diving mammals also use myoglobin as an adaptation to their life under extreme conditions. Some of them (the sperm whale) can dive up to two hours without breathing. Therefore, these animals need high internal oxygen storage to sustain diving. The Mb concentration in the muscles of diving mammals is ten times higher than, for example, in humans (Mirceta et al. 2013). This means that the oxygen storage capacity of the muscles is also increased by ten-fold helping the organism to maintain their physiological functions (Mirceta et al. 2013).

The derived loss of Ngb and Cygb in lungfish may have promoted the functional differentiation of the lungfish Mbs. Interestingly, it seems that the lungfish is the first known vertebrate missing Cygb. The exact functions of Ngb are not fully established by now, but it is well possible that, for some reason, Ngb alone would not be sufficient to protect the lungfish brain from low oxygen conditions and aestivation. It is possible that the different Mbs can achieve this protection better than the single Ngb could, maybe by dividing the different protection mechanisms among each other to fine tune the needs for specific functions in different tissues. This may explain why there is no Ngb, but a variety of Mb expressed in the lungfish brain. Another possible explanation could be that Ngb in lungfishes got lost before Mb took over the

functions of Ngb. Maybe the critical effects resulting from the loss of Ngb led to the emergence of the multiple Mb copies with a main expression in the lungfish brain. In support of this argument the muscle specific Mb1 constitutes a sister group to all the other Mbs and probably displays the original function of Mb. It is conceivable, that through the loss of Ngb the original Mb1 differentiated and the other Mbs compensate for its function in the brain.

The amplification of the different Mbs might have been a response to the changing O_2 conditions of the environment of the lungfish (see Chapter 2) or to other physiological reasons. The different Mb copies have been maintained for over 100 million years suggesting a specific function for all of them. It probably helps the lungfish to cope with the low O_2 environment and to survive these hypoxic periods better, even though it can receive oxygen by breathing air. Additionally, it is possible, that the lungfish, compared to other fish species, can be better at adapting to fluctuating oxygen ranges by acquiring oxygen from water or air. This is also consistent with the latest findings in the *Clarias batrachus* genome. The walking catfish (*C. batrachus*) just like the lungfish is frequently exposed to low-oxygen habitats and harbors multiple Mb copies, mainly expressed in the brain (Li et al. 2018) (see 5.6). Both fishes can aestivate in a mud cocoon during the hot summer periods and they both have Mb predominantly expressed in their brain. This again supports the assumption, that the high Mb concentration in the brain is needed for the fish to survive the periods of seasonal drying in their habitat and a low O_2 environment without severe brain damage. This also hints to the assumption that this adaptive mechanism evolved in the Sarcopterygii independently from the Actinopterygii.

5.4.1 Mb probably derived from sub- and/or neofunctionalization

The longer gene duplication events date back, the higher is the diversity of the gene copies. The current study shows that the different lungfish Mb amino acid sequences reveal an identity of only 36 – 78%. The relatively high substitution rate is comparable with the amino acid substitution of the two Mb-isoforms of the carp and the goldfish (Helbo et al. 2012). It could be possible, that the different Mbs have additive effects on O_2 transport in muscle tissue but all of the Mbs display distinct functional properties. They are highly diverged and retained in evolution. Additionally, the Mb mRNAs as well as the proteins are expressed in various tissues of the lungfish and show different capabilities to enhance the viability of a neuronal cell line under hypoxia. Taken together, the results presented in this thesis suggest that lungfish *Mb*

genes, in contrast to *GbE* genes, evolved through sub- and/or neofunctionalization after gene duplication events. As the different *Mb* genes are widely and specifically expressed in the lungfish tissues, it leads to the assumption of diverse functions of the single Mbs. It can be confirmed by the results achieved in this study, that the different Mbs have specific main expression sites, different oxygen-binding kinetics and enzymatic activities. All these results support the theory of sub- and/or neofunctionalization events after gene duplication from an ancestral *Mb* gene.

5.5 GbE-copies may protect the lungfish eggs

In earlier studies, only a single GbE expression was found in the sauropsids eye (Kugelstadt et al. 2004, Blank et al. 2011a, Hoffmann et al. 2011, Storz et al. 2011, Schwarze and Burmester 2013, Schwarze et al. 2015). Therefore, the lungfish is the first species found not only to harbor GbE in the ovary, but also to possess multiple copies of this gene.

GbE was first identified in the eye of the chicken *Gallus gallus*, later on also in turtles and the coelacanth. This study identified for the first time GbE, which is not expressed in the eye, but in the ovary of all tested lungfish using RNA-Seq and qRT-PCR studies. The findings are also supported by *in situ* hybridization and recombinantly expressed LpaGbE1. The GbE mRNA-levels thereby exceed the values from Mb in heart and skeletal muscle and Hb in blood. Additionally, I was able to discover multiple copies of GbE for the first time, as the lungfish expresses at least five different GbEs. The P_{50} of GbE is with 1.2 Torr similar to typical Mb, it has a low autoxidation-rate and also the deoxy-LpaGbE1 has a similar nitrite reductase rate to Mb. It can be assumed that GbE also contributes to the nitrite dependent NO generation and signaling pathways in the ovary during periods of hypoxia. *In situ* hybridization shows that GbE is massively deposited in the non-mature previtellogenic oocyte but not in any other cell of the ovary and that it is translated into GbE protein. Oxygen is a very important commodity to breed. In lungfish, the males provide the parental care for the embryo in underground burrows with low oxygen levels (Kerr 1900). The GbE seems to have a Mb-like role by contributing to the O_2 supply of oocytes. It provides additional O_2 either by enhancing the O_2 storage or by facilitating the intracellular O_2 diffusion. Therefore, it is assumed that GbE helps to extract the O_2 from the water to support the development of the embryos. This function is analogous to the function of embryonic Hb in tetrapods (Dickerson, Greis, 1983).

Additionally, the high expression level of GbE in the ovaries of the lungfish may protect the previtellogenic eggs during aestivation to ensure the reproduction after aestivation. GbE is, in contrast to Mb, not expressed in the lungfish brain but specifically in the ovary. During aestivation, the lungfish has to withstand highly hypoxic conditions, where it has to make sure, that the eggs are not getting damaged. Similar to the high Mb concentrations in the brain, the lungfish has high GbE concentrations in the oocytes. As Mb and GbE have similar functions, it is conceivable that they also fulfil similar tasks in the lungfish to provide protection through maintained oxygen supply, detoxification of ROS and to increase the vasodilatation and the blood supply through nitrite reductases. This oocyte-specific function of GbE in lungfish is, to the best of our knowledge, not mirrored by any other globin-type. But there must be a reason for it.

The *GbE* genes commenced already in the lungfish stem-lineage but then they proliferated independently in the orders *Lepidosiren* and *Protopterus*. This again displays the importance of the *GbE* genes in the oocytes of lungfish.

Currently it is unclear whether the original function of GbEs was in the oocytes, like in lungfish, or in the eye, as seen in sauropsids. But there is no doubt that GbE must have changed its expression site, either in the lungfish or in the sauropsid clade and got lost in at least five vertebrate clades (Hoffmann et al, 2011). To solve this issue the transcriptomes of the ovary and the eye of the other extant sarcopterygian fish, the coelacanth, or its tissue are needed to perform qRT-PCR analyses.

5.5.1 GbE has a possible additive effect

The amino acid sequences of the different GbEs are very similar (maximum difference of 40%) and they are all exclusively expressed in the ovary, favoring for an additive effect of these genes. The very high RPKM values of GbE, which are between 20,000 and 63,000 depending on the species, lead to the assumption, that the presence of multiple paralogous GbE copies enhances the GbE mRNA levels. This, in turn, assumes that the GbE proteins, other than the multiple Mb copies, seem to encode similar functions. These results suggest a similarity in function of the Mb in the heart and skeletal muscle that leads to the enhanced supply with additional oxygen by either increase in the O_2 storage or by facilitating the intracellular O_2 diffusion. In summary this gives hint to an additive function of GbE in lungfish oocytes.

5.6 Multiple globin genes are a proof of adaptation

The globin family is a classic example to study evolution. Duplications are a way of adding new genetic information to a genome and are the drivers of evolution. This is also the explanation for whole families of related genes, such as the globin family. The vertebrate globin repertoire evolved through repeated rounds of gene duplication and divergence. The evolution of this family shows subfunctionalization as well as neofunctionalization after gene duplication events. These mechanisms are supported by the existence of eight distinct globin genes in vertebrates, which emerged very early in evolution.

There is no doubt that all globins derived from a single ancestor with multiple gene duplications leading to their functional diversification. The emergence of multicellular organisms during evolution, which resulted in a larger body size and a higher complexity of organisms, made simple diffusion inefficient for living organisms. They needed to evolve the mechanism to improve O_2 uptake. All of the eight different globins were present in the gnathostome ancestor, but subsequent losses shaped the present globin repertoire of the vertebrates. There are only a few vertebrates known to harbor all of the eight globin types in their genome: the coelacanth (*Latimeria chalumnae*), the chinese softshell turtle (*Pelodiscus sinensis*) and the western painted turtle (*Chrysemis picta bellii*) (Schwarze and Burmester 2013, Schwarze et al. 2014). It is undisputed that Adgb, GbX, and Ngb are phylogenetically the oldest of the globins and emerged before the split of the protostomia and the deuterostomia as these globins exist in vertebrates as well as in invertebrates (Roesner et al. 2005, Blank and Burmester 2012, Droge et al. 2012, Hoffmann et al. 2012, Hoogewijs et al. 2012, Storz et al. 2013). Nevertheless, till now it is not resolved, which one was the most ancestral globin (Burmester and Hankeln 2014). All of the original globins are hexa-coordinated pointing out the possible initial coordination of the heme Fe-ion in vertebrates and invertebrates (Burmester and Hankeln 2014). GbX additionally has a membrane-binding site, an N-terminal acylation mediating the membrane binding. This function dates back very long and is found in plants, brown algae, and fungi (Blank and Burmester 2012). Evaluation of ROS and hypoxia protection in a neuronal mouse cell line revealed a significantly more effective protective role of membrane bound GbX than mutated, soluble GbX (Koch and Burmester 2016). Thus, the protection from ROS-stress is most likely the original function of globins before they acquired a respiratory role and the membrane-binding, which is displayed by the original structure (Blank and Burmester 2012, Koch and Burmester 2016).

Even though most vertebrates only harbor one copy of Mb, Cygb, Ngb or GbX, multiple or duplicated globin genes can be found throughout the animal kingdom. The last common ancestor of the vertebrates probably had 4 copies of GbX, which were reduced to one copy during evolution in most of the vertebrates. Nevertheless, most of the Actinopterygii harbor two GbX copies (Opazo et al. 2015, Gallagher and Macqueen 2017), besides the rainbow trout (*Oncorhynchus mykiss*) and the Atlantic salmon (*Salmo salar*) which harbor 3 different GbX copies (GbX2a1, GbX2a2, and GbX2b) (Gallagher and Macqueen 2017). Additionally, phylogenetic analysis revealed two Ngb paralogs in salmonids, coding for full-length proteins (Gallagher and Macqueen 2017). Most bony fish harbor two different *Cygb* genes, probably due to the third whole-genome duplication event in their stem lineage (Fuchs et al. 2005, Hoffmann et al. 2011). The different *Cygb* genes may give insights in the evolution from hexa- to penta-coordinated globins, as Cygb1 in zebrafish surprisingly shows a penta-coordination whereas Cygb2, as an ortholog to mammalian Cygb, is hexa-coordinated (Fig. V.4) (Corti et al. 2016). *Cygb1* is ubiquitously expressed and shows a 150-fold higher expression rate in the adult zebrafish compared to *Cygb2* (Tiedke et al. 2011). *Cygb2* on the other hand is 300-fold higher expressed in neuronal tissues than *Cygb1* (Fuchs et al. 2005). This example shows how globin duplicates can have different expression sites and therefore also different functions. This correlates with the finding of multiple *Mb* genes in the lungfish. Certainly, the lungfish Mbs are all penta-coordinated, but still they are expressed in different tissues. Additionally, they all show disparate functional characteristics in case of for example O_2 binding and nitrite reductase function. The multiple *GbE* genes, in contrast, are only expressed in one tissue and probably display similar functions which are added up to cover the special needs of the ovary.

Other examples are the common carp (*Cyprinus carpio*) and the gold fish (*Carassius auratus*), which harbor two Mb copies each, Mb1 and Mb2, instead of just only one (Fraser et al. 2006, Roesner et al. 2008). This is probably also due to whole-genome duplications in the stem lineage of Cyprinidae around 12 million years ago (Larhammar and Risinger 1994, David et al. 2003, Zhao et al. 2014). The two Mb copies developed different functions as Mb1 is mainly expressed in the muscle, liver, kidney, and gills (Fraser et al. 2006, Cossins et al. 2009). The oxygen affinity and nitrite reductase activity are similar to vertebrate Mb and it probably functions as oxygen storage and in NO reductase regulation during hypoxia. Mb2, on the other hand, is specifically expressed in the brain and may protect the brain from ROS by detoxifying H_2O_2 (Fraser et al. 2006, Roesner et al. 2008, Cossins et al. 2009, Helbo et al. 2012). This Mb

duplication event occurred in an ancestor of both the schizothoracine and Cyprinidae fish, but it seems that the Mb2 paralog was subsequently lost in the schizothoracine fish (Gallagher and Macqueen 2017, Qi et al. 2018). The duplication of Mb and the subsequent evolutionary adjustments concerning some of the functions and expression patterns leads to an improved tolerance against hypoxia in fish (Zhao et al. 2014). This shows the advantage of multiple globin copies in organisms. These findings are, to some extent, also reflected by the results of the multiple Mb copies in lungfish. Lungfish Mb1 is mainly expressed in the muscle tissues and reflects the common vertebrate Mb, which is normally expressed in muscle and heart tissue. The other lungfish Mbs are mainly expressed in the brain, like Mb2 in the carp and the goldfish, which probably helps to detoxify ROS and/or to supply the lungfish brain with O_2 during aestivation.

Recent findings from Gallagher and Macqueen display triplicated *Mb* genes in the arowana lineage Osteoglossinae: the silver arowana (*Osteoglossum bicirrhosum*) and the Asian arowana (*Scleropages formosus*), which are both not able to breathe air. Both of them harbor Mbα, Mbβ, and Mbγ (Gallagher and Macqueen 2017). Phylogenetic analysis show that Mb duplications happened in the last common ancestor of the Cypriniformes. They probably arose by tandem duplication and diverged under positive selection. It needs to be established whether they also exist in Arapaiminae, the other subfamily of Osteoglossidae.

The walking catfish (*Clarias batrachus*) is a bony fish, but in contrast to the lungfish it belongs to the Actinopterygii whereas the lungfish belongs to the Sarcopterygii. Similar to the lungfish, the walking catfish is able to breathe air but without having a lung. Instead the air-breathing organ consists of suprabranchial chambers, gill fans and arborescent organs (Munshi 1961, Lewis 1979). It can live in various low-oxygen habitats and buries itself in the mud during summer periods (Das 1928, Islam et al. 2007, Saha et al. 2007). The genome size of the walking catfish is with a C-value of 1.2 quite small. In comparison, the human genome has a C-value of 3.5 and the African lungfish *P. aethiopicus* a C-value of 132.8 (Gregory 2018). In spite of their relatively small genome size, genomic analyses revealed 15 myoglobin gene copies in the *C. batrachus* genome which may be located on different chromosomes (Li et al. 2018). This again suggests, that the massive genome size of the lungfish is not causative of the high Mb diversity in this species. The different Mbs from the walking catfish are all expressed, but to a different extent and in different tissues. Overall, the brain has the highest myoglobin expression

level (Li et al. 2018), which is consistent with my results of the lungfish myoglobins. These latest findings show high similarity between the Mb distribution of lungfish and walking catfish and include other important similarities. They both can aestivate during the hot summer periods and have the highest Mb concentration in the brain. O_2 is the critical factor for the brain to fulfil the required tasks and it seems that the high Mb expression helps the brain to survive by detoxifying ROS and supplying sufficient amounts of oxygen.

Additionally, multiple copies of the thought-to-be eye specific globin GbE in lungfish ovaries were detected. This is the first time multiple GbE copies were detected. So far there is no other species with multiplied globin genes in the oocytes, but there must be a reason, as they remained multiplied in lungfish throughout evolution. It can be assumed, that there is a need for the lungfish, to protect the previtellogenic eggs during aestivation to sustain reproduction (see 5.5).

5.6.1 An attempt of explanation: The unexpected globin repertoire in lungfish

The work of this thesis demonstrates that multiple GbE and Mb copies were found in all the Lepidosireniformes which results in the conclusion that this amplification evolved before the split of the genera *Lepidosiren* and *Protopterus* more than 100 million years ago. The Mb orthologs also remained functional for over 100 million years. But then the multiplication of *GbE* genes occurred independently in the orders *Lepidosiren* and *Protopterus*. The arising question is, how do the multiple globin copies emerge and what is their function?

The coelacanth and the turtle are the only known vertebrates harboring all eight globins (Schwarze and Burmester 2013). The lungfish on the other hand has probably lost Cygb and Ngb (and Adgb in some species), and is therefore the only known vertebrate without the *Cygb* gene. The exact functions of these two globins are still a matter of debate but it is known, that they are both expressed in the brain. It is very likely, that the highly expressed Mbs in the brain have taken over the functions of Ngb and Cygb in the lungfish. GbE is independently lost in nearly all vertebrates besides birds, turtles, the coelacanth, and the lungfish.

The lungfish aestivates during the hot summer periods, which means hypoxic conditions for the lungfish. The multiple Mbs may help the lungfish to protect the brain from oxidative stress, especially during the dry season. GbE on the other hand may protect the previtellogenic eggs as it is only expressed at this stage. It may help extract the O_2 from the water to help the

embryos to survive and/or protects the eggs during aestivation, as it supplies them the oxygen needed.

5.7 Outlook

As shown in this thesis, most of the lungfish species harbor multiple Mb and GbE copies. In the future it would be necessary, to get an Australian lungfish (*Neoceratodus forsteri*) genome to confirm, if there are also multiple Mb and GbE copies. As the *N. forsteri* does not have an actual lung and does not aestivate in summertime, it is well likely, that they do not have multiple Mb or GbE copies. In contrast to the Lepidosireniformes they do not need this adaptation to a temporarily hypoxic environment. The Australian lungfish needs to live in permanent water, little mud, and vegetation. They are found in deep water in winter and during the day and in shallow water during the spawning season and at night (Kemp 1986). The Australian lungfish can grow up to a size of 2 m and can weigh up to 48 kg. It has a single lung and is a facultative air breather, which means, it uses the lung only for aerobic respiration when it is more animated and needs additional oxygen (Kemp 1986).

If the Australian lungfish also harbors the multiple globin genes, it would be interesting to have a look at the genomic sequences to find out, if the divergence dates even further back in evolution than thought before. Also, it would be very interesting to see, if these multiple globin genes have similar distribution and expression pattern in the different tissues, or if there are significant differences, as the Australian lungfish does not aestivate.

Additionally, the genome of the coelacanth would be very helpful as well to decide on the direction of evolutionary change of GbE. To date, none of the available transcriptomes derives either from the eye or oocytes, but it is clear, that GbE must have changed expression site at one stage.

In general, it would be important to have more genomes, or at least transcriptomes of eye and ovary tissues, of other bony-fish, to compare with the globin expression in lungfish tissues. Obtaining the lungfish genomes would also be important to find out about the existence of Adgb in lungfish. So far, I only found fragments of Adgb in *P. annectens*, but there is no clear evidence regarding the general existence of Adgb in lungfish.

The exact functions of Ngb and Cygb remain to be investigated. It is important not only for a better understanding of these globins, but also to identify, if the multiple lungfish Mbs are able to account for their multiple functions.

VI. References

Abascal, F., Zardoya, R. and Posada, D. (2005). "ProtTest: selection of best-fit models of protein evolution." Bioinformatics **21**(9): 2104-2105.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002). Molecular Biology of the Cell. New York, Garland Science.

Amemiya, C. T., Alfoldi, J., Lee, A. P., Fan, S., Philippe, H., Maccallum, I., . . . Lindblad-Toh, K. (2013). "The African coelacanth genome provides insights into tetrapod evolution." Nature **496**(7445): 311-316.

Antonini, E. and Brunori, M. (1970). "Hemoglobin." Annu Rev Biochem **39**: 977-1042.

Antonini, E. and Brunori, M. (1971). Hemoglobin and Myoglobin in Their Reactions with Ligands. North-Holland, Amsterdam, Frontiers of Biology.

Awenius, C., Hankeln, T. and Burmester, T. (2001). "Neuroglobins from the zebrafish Danio rerio and the pufferfish Tetraodon nigroviridis." Biochem Biophys Res Commun **287**(2): 418-421.

Ayres, D. L., Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O., . . . Suchard, M. A. (2012). "BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics." Systematic Biology **61**(1): 170-173.

Bahler, M. and Rhoads, A. (2002). "Calmodulin signaling via the IQ motif." FEBS Lett **513**(1): 107-113.

Bailly, X., Vanin, S., Chabasse, C., Mizuguchi, K. and Vinogradov, S. N. (2008). "A phylogenomic profile of hemerythrins, the nonheme diiron binding respiratory proteins." BMC Evol Biol **8**: 244.

Ballantyne, J. S. and Frick, N. T. (2010). Lungfish metabolism. The biology of lungfishes. J. M. Jørgensen and J. Joss, New Hampshire: Science Publishers**:** 305–340.

Bauer, C., Tamm, R., Petschow, D., Bartels, R. and Bartels, H. (1975). "Oxygen affinity and allosteric effects of embryonic mouse haemolglobins." Nature **257**(5524): 333-334.

Beedholm, K. (2018). Spektrosampler. Aarhus University, Denmark.

Bentmann, A., Schmidt, M., Reuss, S., Wolfrum, U., Hankeln, T. and Burmester, T. (2005). "Divergent distribution in vascular and avascular mammalian retinae links neuroglobin to cellular respiration." J Biol Chem **280**(21): 20660-20665.

Betancur, R. R., Broughton, R., Wiley, E., Carpenter, K., López, J., Li, C., . . . Ortí, G. (2013). "The Tree of Life and a New Classification of Bony Fishes." PLOS Currents Tree of Life **ecurrents.tol.53ba26640df0ccaee75bb165c8c26288**.

Betancur, R. R., Orti, G. and Pyron, R. A. (2015). "Fossil-based comparative analyses reveal ancient marine ancestry erased by extinction in ray-finned fishes." Ecol Lett **18**(5): 441-450.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., . . . Schwede, T. (2014). "SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information." Nucleic Acids Res **42**(Web Server issue): W252-258.

Bienert, S., Waterhouse, A., de Beer, T. A., Tauriello, G., Studer, G., Bordoli, L. and Schwede, T. (2017). "The SWISS-MODEL Repository-new features and functionality." Nucleic Acids Res **45**(D1): D313-D319.

Biscotti, M. A., Gerdol, M., Canapa, A., Forconi, M., Olmo, E., Pallavicini, A., . . . Schartl, M. (2016). "The Lungfish Transcriptome: A Glimpse into Molecular Evolution Events at the Transition from Water to Land." Sci Rep **6**: 21571.

Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Jeffries, C. M., . . . Svergun, D. I. (2015). "Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY)." J Appl Crystallogr **48**(Pt 2): 431-443.

Blank, M. and Burmester, T. (2012). "Widespread occurrence of N-terminal acylation in animal globins and possible origin of respiratory globins from a membrane-bound ancestor." Mol Biol Evol **29**(11): 3553-3561.

Blank, M., Kiger, L., Thielebein, A., Gerlach, F., Hankeln, T., Marden, M. C. and Burmester, T. (2011a). "Oxygen supply from the bird's eye perspective: globin E is a respiratory protein in the chicken retina." J Biol Chem **286**(30): 26507-26515.

Blank, M., Wollberg, J., Gerlach, F., Reimann, K., Roesner, A., Hankeln, T., . . . Burmester, T. (2011b). "A membrane-bound vertebrate globin." PLoS One **6**(9): e25292.

Bourne, H. R., Sanders, D. A. and McCormick, F. (1991). "The GTPase superfamily: conserved structure and molecular mechanism." Nature **349**(6305): 117-127.

Brinkmann, H., Venkatesh, B., Brenner, S. and Meyer, A. (2004). "Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates." Proc Natl Acad Sci U S A **101**(14): 4900-4905.

Brownlie, A., Hersey, C., Oates, A. C., Paw, B. H., Falick, A. M., Witkowska, H. E., . . . Zon, L. (2003). "Characterization of embryonic globin genes of the zebrafish." Dev Biol **255**(1): 48-61.

Brunori, M. (2001). "Nitric oxide moves myoglobin centre stage." Trends Biochem Sci **26**(4): 209-210.

Brunori, M., Giuffre, A., Nienhaus, K., Nienhaus, G. U., Scandurra, F. M. and Vallone, B. (2005). "Neuroglobin, nitric oxide, and oxygen: functional pathways and conformational changes." Proc Natl Acad Sci U S A **102**(24): 8483-8488.

Burmester, T. (2002). "Origin and evolution of arthropod hemocyanins and related proteins." J Comp Physiol B **172**(2): 95-107.

Burmester, T., Ebner, B., Weich, B. and Hankeln, T. (2002). "Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues." Mol Biol Evol **19**(4): 416-421.

Burmester, T., Haberkamp, M., Mitz, S. A., Roesner, A., Schmidt, M., Ebner, B., . . . Hankeln, T. (2004). "Neuroglobin and cytoglobin: genes, proteins and evolution." IUBMB Life **56**(11-12): 703-707.

Burmester, T. and Hankeln, T. (2009). "What is the function of neuroglobin?" J Exp Biol **212**(Pt 10): 1423-1428.

Burmester, T. and Hankeln, T. (2014). "Function and evolution of vertebrate globins." Acta Physiol (Oxf) **211**(3): 501-514.

Burmester, T., Weich, B., Reinhardt, S. and Hankeln, T. (2000). "A vertebrate globin expressed in the brain." Nature **407**(6803): 520-523.

Carter, G. S. B., L. C. (1931). "Reports of an Expedition to Brazil and Paraguay in 1926–7, supported by the Trustees of the Percy Sladen Memorial Fund and by the Executive Committee of the Carnegie Trust for the Universities of Scotland. The Fauna of the Swamps of the Paraguayan Chaco in relation to its Environment. —II. Respiratory Adaptations in the Fishes." Zool. J. Linn. Soc. **37**: 327-368.

Chang, M.-M. and Yu, X. (1984). "Structure and phylogenetic significance of Diabolichthys speratus gen. et sp. nov., a new dipnoan-like form from the Lower Devonian of eastern Yunnan, China." Proceedings of The Linnean Society of New South Wales **107**(3): 14.

Chaves, P. T. C. (1992). "Some aspects of the oogenesis in the South american lungfish, Lepidosiren paradoxa Fitzinger (Dipnoi)." Rev. Bras. Zool. **9**: 93-98.

Chew, S. F., Chan, N. K., Loong, A. M., Hiong, K. C., Tam, W. L. and Ip, Y. K. (2004). "Nitrogen metabolism in the African lungfish (Protopterus dolloi) aestivating in a mucus cocoon on land." J Exp Biol **207**(Pt 5): 777-786.

Chew, S. F. and Hiong, K. (2014). "Aestivation and brain of the African lungfish Protopterus annectens." Temperature (Austin) **1**(2): 82-83.

Clack, J. A. S., L.; Long, J. A. (2011). The fossil record of lungfishes. The Biology of lungfishes (Jørgensen, J. M. & Joss, J., eds). Enfield, USA, Science Publishers.

Corti, P., Ieraci, M. and Tejero, J. (2016). "Characterization of zebrafish neuroglobin and cytoglobins 1 and 2: Zebrafish cytoglobins provide insights into the transition from six-coordinate to five-coordinate globins." Nitric Oxide **53**: 22-34.

Cosby, K., Partovi, K. S., Crawford, J. H., Patel, R. P., Reiter, C. D., Martyr, S., . . . Gladwin, M. T. (2003). "Nitrite reduction to nitric oxide by deoxyhemoglobin vasodilates the human circulation." Nat Med **9**(12): 1498-1505.

Cossins, A. R., Williams, D. R., Foulkes, N. S., Berenbrink, M. and Kipar, A. (2009). "Diverse cell-specific expression of myoglobin isoforms in brain, kidney, gill and liver of the hypoxia-tolerant carp and zebrafish." J Exp Biol **212**(Pt 5): 627-638.

Cunningham, J. T. (1929). "The vascular filaments on the pelvic limbs of Lepidosiren, their function and evolutionary significance." Proc. R. Soc. Lond. **105**: 484-493.

Danovaro, R., Dell'Anno, A., Pusceddu, A., Gambi, C., Heiner, I. and Kristensen, R. M. (2010). "The first metazoa living in permanently anoxic conditions." BMC Biol **8**: 30.

Darriba, D., Taboada, G. L., Doallo, R. and Posada, D. (2011). "ProtTest 3: fast selection of best-fit models of protein evolution." Bioinformatics **27**(8): 1164-1165.

Das, B. K. (1928). "The bionomics of certain air-breathing fishes of India, together withan account of the development of their air-breathing organs." Philosophical Transactions of the Royal Society of London **216**: 183-219.

David, L., Blum, S., Feldman, M. W., Lavi, U. and Hillel, J. (2003). "Recent duplication of the common carp (Cyprinus carpio L.) genome as revealed by analyses of microsatellite loci." Mol Biol Evol **20**(9): 1425-1434.

Davis, J. C. and Petrov, D. A. (2005). "Do disparate mechanisms of duplication add similar genes to the genome?" Trends Genet **21**(10): 548-551.

de Lima, J. L., Soares, F. A., Remedios, A. C., Thom, G., Wirthlin, M., Aleixo, A., . . . Schneider, P. N. (2015). "A putative RA-like region in the brain of the scale-backed antbird, Willisornis poecilinotus (Furnariides, Suboscines, Passeriformes, Thamnophilidae)." Genet Mol Biol **38**(3): 249-254.

de Sanctis, D., Pesce, A., Nardini, M., Bolognesi, M., Bocedi, A. and Ascenzi, P. (2004). "Structure-function relationships in the growing hexa-coordinate hemoglobin sub-family." IUBMB Life **56**(11-12): 643-651.

Delaney, R. G., Lahiri, S. and Fishman, A. P. (1974). "Aestivation of the African lungfish Protopterus aethiopicus: cardiovascular and respiratory functions." J Exp Biol **61**(1): 111-128.

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

Droge, J., Pande, A., Englander, E. W. and Makalowski, W. (2012). "Comparative genomics of neuroglobin reveals its early origins." PLoS One **7**(10): e47972.

Duong, T. T., Witting, P. K., Antao, S. T., Parry, S. N., Kennerson, M., Lai, B., . . . Harris, H. H. (2009). "Multiple protective activities of neuroglobin in cultured neuronal cells exposed to hypoxia re-oxygenation injury." J Neurochem **108**(5): 1143-1154.

Ebner, B. (2012). Phylogenetische & Molekulare Analyse des Globin-Repertoires der Metazoa, Johannes Gutenberg-Universität.

Ebner, B., Panopoulou, G., Vinogradov, S. N., Kiger, L., Marden, M. C., Burmester, T. and Hankeln, T. (2010). "The globin gene family of the cephalochordate amphioxus: implications for chordate globin evolution." BMC Evol Biol **10**: 370.

Eldredge, N. S., S. M. (1984). Living fossils. New York, Springer-Verlag.

Enoki, Y., Matsumura, K., Ohga, Y., Kohzuki, H. and Hattori, M. (1995). "Oxygen affinities (P50) 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 B Biochem Mol Biol **110**(1): 193-199.

Enoki, Y., Ohga, Y., Ishidate, H. and Morimoto, T. (2008). "Primary structure of myoglobins from 31 species of birds." Comp Biochem Physiol B Biochem Mol Biol **149**(1): 11-21.

Fabrizius, A. (2014). Untersuchungen zur Lokalisation und Expression des respiratorischen Proteins Neuroglobin bei Säugetieren, Johannes Gutenberg-Universität.

Fabrizius, A., Andre, D., Laufs, T., Bicker, A., Reuss, S., Porto, E., . . . Hankeln, T. (2016). "Critical re-evaluation of neuroglobin expression reveals conserved patterns among mammals." Neuroscience **337**: 339-354.

Fago, A. (2017). "Functional roles of globin proteins in hypoxia-tolerant ectothermic vertebrates." J Appl Physiol (1985) **123**(4): 926-934.

Fago, A., Hundahl, C., Dewilde, S., Gilany, K., Moens, L. and Weber, R. E. (2004a). "Allosteric regulation and temperature dependence of oxygen binding in human neuroglobin and cytoglobin. Molecular mechanisms and physiological significance." J Biol Chem **279**(43): 44417-44426.

Fago, A., Hundahl, C., Malte, H. and Weber, R. E. (2004b). "Functional properties of neuroglobin and cytoglobin. Insights into the ancestral physiological roles of globins." IUBMB Life **56**(11-12): 689-696.

Fago, A. and Jensen, F. B. (2015). "Hypoxia tolerance, nitric oxide, and nitrite: lessons from extreme animals." Physiology (Bethesda) **30**(2): 116-126.

Fago, A., Parraga, D. G., Petersen, E. E., Kristensen, N., Giouri, L. and Jensen, F. B. (2017). "A comparison of blood nitric oxide metabolites and hemoglobin functional properties among diving mammals." Comp Biochem Physiol A Mol Integr Physiol **205**: 35-40.

Fago, A., Rohlfing, K., Petersen, E. E., Jendroszek, A. and Burmester, T. (2018). "Functional diversification of sea lamprey globins in evolution and development." Biochim Biophys Acta **1866**(2): 283-291.

Fang, J., Ma, I. and Allalunis-Turner, J. (2011). "Knockdown of cytoglobin expression sensitizes human glioma cells to radiation and oxidative stress." Radiat Res **176**(2): 198-207.

Fedorov, B. A. and Denesyuk, A. I. (1978). "Sperm whale myoglobin structure in solution differs from its structure in crystal by a shift of the 'hairpin' GH." FEBS Lett **88**(1): 114-117.

Fiocchetti, M., Nuzzo, M. T., Totta, P., Acconcia, F., Ascenzi, P. and Marino, M. (2014). "Neuroglobin, a pro-survival player in estrogen receptor alpha-positive cancer cells." Cell Death Dis **5**: e1449.

Fishman, A. P. P., A. I.; Delaney, R. G.; Galante, R. J. (1986). "A. P. Fishman et al., "Estivation in Protopterus," Journal of Morpholology 190 Supplement 1, (1986): 237–248." Journal of Morphology **190**(S1): 237-248.

Flogel, U., Fago, A. and Rassaf, T. (2010). "Keeping the heart in balance: the functional interactions of myoglobin with nitrogen oxides." J Exp Biol **213**(Pt 16): 2726-2733.

Flögel, U., Gödecke, A., Klotz, L. O. and Schrader, J. (2004). "Role of myoglobin in the antioxidant defense of the heart." FASEB Journal **18**(10): 1156-1158.

Flögel, U., Merx, M. W., Gödecke, A., Decking, U. K. and Schrader, J. (2001). "Myoglobin: A scavenger of bioactive NO." Proceedings of the National Academy of Sciences of the United States of America **98**(2): 735-740.

Franke, D., Kikhney, A. G. and Svergun, D. I. (2012). "Automated acquisition and analysis of small angle X-ray scattering data." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **689**: 8.

Fraser, J., de Mello, L. V., Ward, D., Rees, H. H., Williams, D. R., Fang, Y., . . . Cossins, A. R. (2006). "Hypoxia-inducible myoglobin expression in nonmuscle tissues." Proc Natl Acad Sci U S A **103**(8): 2977-2981.

Fuchs, C. (2007). Das Globingen-Repertoire von Amphibien und Teleostiern: Molekulare Evolution and Analysen zur Genexpression. PhD, Johannes-Gutenberg University.

Fuchs, C., Burmester, T. and Hankeln, T. (2006). "The amphibian globin gene repertoire as revealed by the Xenopus genome." Cytogenet Genome Res **112**(3-4): 296-306.

Fuchs, C., Luckhardt, A., Gerlach, F., Burmester, T. and Hankeln, T. (2005). "Duplicated cytoglobin genes in teleost fishes." Biochem Biophys Res Commun **337**(1): 216-223.

Gallagher, M. D. and Macqueen, D. J. (2017). "Evolution and Expression of Tissue Globins in Ray-Finned Fishes." Genome Biol Evol **9**(1): 32-47.

Garry, D., Ordway, G., Lorenz, J., Radford, N., Chin, E., Grange, R., . . . Williams, R. (1998). "Mice without myoglobin." Nature **395**(6705): 905-908.

Geiser, F. (1988). "Reduction of metabolism during hibernation and daily torpor in mammals and birds: temperature effect or physiological inhibition?" J Comp Physiol B **158**(1): 25-37.

George, P. and Irvine, D. H. (1951). "Reaction of metmyoglobin with hydrogen peroxide." Nature **168**: 164–165.

Gillemans, N., McMorrow, T., Tewari, R., Wai, A. W., Burgtorf, C., Drabek, D., . . . Philipsen, S. (2003). "Functional and comparative analysis of globin loci in pufferfish and humans." Blood **101**(7): 2842-2849.

Gilman, A. G. (1987). "G proteins: transducers of receptor-generated signals." Annu Rev Biochem **56**: 615-649.

Gödecke, A., Flögel, U., Zanger, K., Ding, Z., Hirchenhain, J., Decking, U. K. and Schrader, J. (1999). "Disruption of myoglobin in mice induces multiple compensatory mechanisms." Proceedings of the National Academy of Sciences of the United States of America **96**(18): 10495-10500.

Gorr, T. A., Wichmann, D., Pilarsky, C., Theurillat, J. P., Fabrizius, A., Laufs, T., . . . Kristiansen, G. (2011). "Old proteins - new locations: myoglobin, haemoglobin, neuroglobin and cytoglobin in solid tumours and cancer cells." Acta Physiol (Oxf) **202**(3): 563-581.

Graham, J. B. (1997). Air-Breathing Fishes: Evolution, Diversity, and Adaptation. San Diego, USA, Academic Press.

Graur, D. L., Wen-Hsiung Li (2000). Fundamentals of Molecular Evolution, Sinauer Associates, Incorporated.

Greenwood, P. H. (1987). "The natural history of African lungfishes." Journal of morphology **190**(S1): 163-179.

Gregory, P. T. (1982). Reptilian hibernation. New York, Academic Press.

Gregory, T. R. (2016). "Animal Genome Size Database." http://www.genomesize.com.

Gregory, T. R. (2017). "Animal Genome Size Database." from http://www.genomesize.com

Gregory, T. R. (2018). "Animal Genome Size Database." from http://www.genomesize.com

Guppy, M. and Withers, P. (1999). "Metabolic depression in animals: physiological perspectives and biochemical generalizations." Biol Rev Camb Philos Soc **74**(1): 1-40.

Hankeln, T., Ebner, B., Fuchs, C., Gerlach, F., Haberkamp, M., Laufs, T. L., . . . Burmester, T. (2005). "Neuroglobin and cytoglobin in search of their role in the vertebrate globin family." J Inorg Biochem **99**(1): 110-119.

Hankeln, T., Wystub, S., Laufs, T., Schmidt, M., Gerlach, F., Saaler-Reinhardt, S., . . . Burmester, T. (2004). "The cellular and subcellular localization of neuroglobin and cytoglobin -- a clue to their function?" IUBMB Life **56**(11-12): 671-679.

Hardison, R. (1998). "Hemoglobins from bacteria to man: evolution of different patterns of gene expression." J Exp Biol **201**(Pt 8): 1099-1117.

Hardison, R. C. (1996a). "A brief history of hemoglobins: plant, animal, protist, and bacteria." Proc Natl Acad Sci U S A **93**(12): 5675-5679.

Hardison, R. C. (1996b). "A brief history of hemoglobins: Plant, animal, protist, and bacteria." Proceedings of the National Academy of Sciences of the United States of America **93**(12): 5675-5679.

Hayashi, A., Suzuki, T. and Shin, M. (1973). "An enzymic reduction system for metmyoglobin and methemoglobin, and its application to functional studies of oxygen carriers." Biochim Biophys Acta **310**(2): 309-316.

Hedges, S. B., Marin, J., Suleski, M., Paymer, M. and Kumar, S. (2015). "Tree of life reveals clock-like speciation and diversification." Mol Biol Evol **32**(4): 835-845.

Heisler, N., Forcht, G., Ultsch, G. R. and Anderson, J. F. (1982). "Acid-base regulation in response to environmental hypercapnia in two aquatic salamanders, Siren lacertina and Amphiuma means." Respir Physiol **49**(2): 141-158.

Helbo, S., Dewilde, S., Williams, D. R., Berghmans, H., Berenbrink, M., Cossins, A. R. and Fago, A. (2012). "Functional differentiation of myoglobin isoforms in hypoxia-tolerant carp indicates tissue-specific protective roles." Am J Physiol Regul Integr Comp Physiol **302**(6): R693-701.

Helbo, S. and Fago, A. (2011). "Allosteric modulation by S-nitrosation in the low-O(2) affinity myoglobin from rainbow trout." Am J Physiol Regul Integr Comp Physiol **300**(1): R101-108.

Helbo, S., Weber, R. E. and Fago, A. (2013). "Expression patterns and adaptive functional diversity of vertebrate myoglobins." Biochim Biophys Acta **1834**(9): 1832-1839.

Hendgen-Cotta, U. B., Merx, M. W., Shiva, S., Schmitz, J., Becher, S., Klare, J. P., . . . Rassaf, T. (2008). "Nitrite reductase activity of myoglobin regulates respiration and cellular viability in myocardial ischemia-reperfusion injury." Proc Natl Acad Sci U S A **105**(29): 10256-10261.

Hepler, J. R. and Gilman, A. G. (1992). "G proteins." Trends Biochem Sci **17**(10): 383-387.

Herold, S., Fago, A., Weber, R. E., Dewilde, S. and Moens, L. (2004). "Reactivity studies of the Fe(III) and Fe(II)NO forms of human neuroglobin reveal a potential role against oxidative stress." J Biol Chem **279**(22): 22841-22847.

Hilzheimer, D. M. (1913). Handbuch der Biologie der Wirbeltiere, Enke.

Hiong, K. C., Ip, Y. K., Wong, W. P. and Chew, S. F. (2013). "Differential gene expression in the brain of the African lungfish, Protopterus annectens, after six days or six months of aestivation in air." PLoS One **8**(8): e71205.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. and Vinh, L. S. (2018). "UFBoot2: Improving the Ultrafast Bootstrap Approximation." Mol Biol Evol **35**(2): 518-522.

Hoffmann, F. G., Opazo, J. C. and Storz, J. F. (2011). "Differential loss and retention of cytoglobin, myoglobin, and globin-E during the radiation of vertebrates." Genome Biol Evol **3**: 588-600.

Hoffmann, F. G., Opazo, J. C. and Storz, J. F. (2012). "Whole-genome duplications spurred the functional diversification of the globin gene superfamily in vertebrates." Mol Biol Evol **29**(1): 303-312.

Hoffmann, F. G., Storz, J. F., Gorr, T. A. and Opazo, J. C. (2010). "Lineage-specific patterns of functional diversification in the alpha- and beta-globin gene families of tetrapod vertebrates." Mol Biol Evol **27**(5): 1126-1138.

Hoogewijs, D., Ebner, B., Germani, F., Hoffmann, F. G., Fabrizius, A., Moens, L., . . . Hankeln, T. (2012). "Androglobin: a chimeric globin in metazoans that is preferentially expressed in Mammalian testes." Mol Biol Evol **29**(4): 1105-1114.

Huelsenbeck, J. P. and Ronquist, F. (2001). "MRBAYES: Bayesian inference of phylogenetic trees." Bioinformatics **17**(8): 754-755.

Hutchins, M. (2003). Grzimek's Animal Life Encyclopedia: Fishes, Thomson-Gale.

Irisarri, I., Baurain, D., Brinkmann, H., Delsuc, F., Sire, J. Y., Kupfer, A., . . . Philippe, H. (2017). "Phylotranscriptomic consolidation of the jawed vertebrate timetree." Nat Ecol Evol **1**(9): 1370-1378.

Irisarri, I. and Meyer, A. (2016). "The Identification of the Closest Living Relative(s) of Tetrapods: Phylogenomic Lessons for Resolving Short Ancient Internodes." Syst Biol **65**(6): 1057-1075.

Islam, M. N., Islam, M. S. and Alam, M. S. (2007). "Genetic structure of different populations of walking catfish (Clarias batrachus L.) in Bangladesh." Biochem Genet **45**(9-10): 647-662.

Janssens, P. A. and Cohen, P. P. (1966). "Ornithine-Urea Cycle Enzymes in the African Lungfish, Protopterus aethiopicus." Science **152**(3720): 358-359.

Janssens, P. A. C. P. P. (1968). "Nitrogen Metabolism in the African Lungfish." Comp. Biochem. Physiol. **24**: 879-886.

Jendroszek, A., Malte, H., Overgaard, C. B., Beedholm, K., Natarajan, C., Weber, R. E., . . . Fago, A. (2018). "Allosteric mechanisms underlying the adaptive increase in hemoglobin-oxygen affinity of the bar-headed goose." J Exp Biol **221**(Pt 18).

Johansen, K. (1966). "Air breathing in the teleost Symbranchus marmoratus." Comp Biochem Physiol **18**(2): 383-395.

Johansen, K. (1970). "Cardiorespiratory adaptations in the transition from water breathing to air breathing. Introduction." Fed Proc **29**(3): 1118-1119.

Johansen, K. and Lenfant, C. (1967). "Respiratory function in the South American lungfish, Lepidosiren paradoxa (Fitz)." J Exp Biol **46**(2): 205-218.

Jones, D. T., Taylor, W. R. and Thornton, J. M. (1992). "The rapid generation of mutation data matrices from protein sequences." Comput Appl Biosci **8**(3): 275-282.

Jørgensen, J. M. and Joss, J. (2010). The Biology of Lungfishes. Enfield, New Hampshire, Science Publishers.

Kakar, S., Hoffman, F. G., Storz, J. F., Fabian, M. and Hargrove, M. S. (2010). "Structure and reactivity of hexacoordinate hemoglobins." Biophys Chem **152**(1-3): 1-14.

Katoh, K., Asimenos, G. and Toh, H. (2009). "Multiple alignment of DNA sequences with MAFFT." Methods Mol Biol **537**: 39-64.

Katoh, K., Kuma, K., Toh, H. and Miyata, T. (2005). "MAFFT version 5: improvement in accuracy of multiple sequence alignment." Nucleic Acids Res **33**(2): 511-518.

Katoh, K. and Toh, H. (2008). "Recent developments in the MAFFT multiple sequence alignment program." Brief Bioinform **9**(4): 286-298.

Kawada, N., Kristensen, D. B., Asahina, K., Nakatani, K., Minamiyama, Y., Seki, S. and Yoshizato, K. (2001). "Characterization of a stellate cell activation-associated protein (STAP) with peroxidase activity found in rat hepatic stellate cells." J Biol Chem **276**(27): 25318-25323.

Kemp, A. (1986). "The biology of the australian lungfish, Neoceratodus forsteri (krefft 1870) " Journal of morphology **190**(S1): 181-198.

Kendrew, J. C. (1963). "Myoglobin and the structure of proteins." Science **139**(3561): 1259-1266.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. and Phillips, D. C. (1958). "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis." Nature **181**(4610): 662-666.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. and Shore, V. C. (1960). "Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution." Nature **185**(4711): 422-427.

Kerr, J. G. V. (1900). "The external features in the development of Lepidosiren paradoxa, fitz." Proc. R. Soc. Lond. **192**(178-184).

Khan, A. A., Sun, Y., Jin, K., Mao, X. O., Chen, S., Ellerby, L. M. and Greenberg, D. A. (2007). "A neuroglobin-overexpressing transgenic mouse." Gene **398**(1-2): 172-176.

Koch, J. and Burmester, T. (2016). "Membrane-bound globin X protects the cell from reactive oxygen species." Biochem Biophys Res Commun **469**(2): 275-280.

Koch, J., Ludemann, J., Spies, R., Last, M., Amemiya, C. T. and Burmester, T. (2016). "Unusual Diversity of Myoglobin Genes in the Lungfish." Mol Biol Evol **33**(12): 3033-3041.

Kristiansen, G., Hu, J., Wichmann, D., Stiehl, D. P., Rose, M., Gerhardt, J., . . . Gorr, T. A. (2011). "Endogenous myoglobin in breast cancer is hypoxia-inducible by alternative transcription and functions to impair mitochondrial activity: a role in tumor suppression?" Journal of Biological Chemistry **286**(50): 43417-43428.

Kruszewski, K. (2011). Untersuchungen zur Expression von Globinen in Neunaugen. Master, University of Hamburg.

Kugelstadt, D., Haberkamp, M., Hankeln, T. and Burmester, T. (2004). "Neuroglobin, cytoglobin, and a novel, eye-specific globin from chicken." Biochem Biophys Res Commun **325**(3): 719-725.

Lahiri, S., Szidon, J. P. and Fishman, A. P. (1970). "Potential respiratory and circulatory adjustments to hypoxia in the African lungfish." Fed Proc **29**(3): 1141-1148.

Larhammar, D. and Risinger, C. (1994). "Molecular genetic aspects of tetraploidy in the common carp Cyprinus carpio." Mol Phylogenet Evol **3**(1): 59-68.

Laufs, T. L. (2010). Untersuchungen zur Expression und Lokalisation der respiratrischen Proteine Neuroglobin (Ngb) und Cytoglobin (Cygb) in Säugern, Johannes Gutenberg-Universität.

Le, S. Q. and Gascuel, O. (2008). "An improved general amino acid replacement matrix." Molecular Biology and Evolution **25**(7): 1307-1320.

Lechauve, C., Augustin, S., Cwerman-Thibault, H., Bouaita, A., Forster, V., Celier, C., . . . Corral-Debrinski, M. (2012). "Neuroglobin involvement in respiratory chain function and retinal ganglion cell integrity." Biochim Biophys Acta **1823**(12): 2261-2273.

Lechauve, C., Augustin, S., Roussel, D., Sahel, J. A. and Corral-Debrinski, M. (2013). "Neuroglobin involvement in visual pathways through the optic nerve." Biochim Biophys Acta **1834**(9): 1772-1778.

Lee, J., Alrubaian, J. and Dores, R. M. (2006). "Are lungfish living fossils? Observation on the evolution of the opioid/orphanin gene family." Gen Comp Endocrinol **148**(3): 306-314.

Lewis, S. V. (1979). "The morphology of the accessory air-breathing organs of the catfish, Clarias batrachus: A SEM Study " Journal of Fish Biology **14**(2): 187-191.

Li, N., Bao, L., Zhou, T., Yuan, Z., Liu, S., Dunham, R., . . . Liu, Z. (2018). "Genome sequence of walking catfish (Clarias batrachus) provides insights into terrestrial adaptation." BMC Genomics **19**(1): 952.

Lima, S. Q., Costa, C. M., Amemiya, C. T. and Schneider, I. (2017). "Morphological And Molecular Analyses of an Anatomical Novelty: The Pelvic Fin Filaments of the South American Lungfish." J Exp Zool B Mol Dev Evol **328**(1-2): 97-105.

Liu, J., Yu, Z., Guo, S., Lee, S. R., Xing, C., Zhang, C., . . . Wang, X. (2009). "Effects of neuroglobin overexpression on mitochondrial function and oxidative stress following hypoxia/reoxygenation in cultured neurons." J Neurosci Res **87**(1): 164-170.

Lomholt, J. P. (1993). Breathing in the aestivating African lungfish, Protopterus amphibius. Delhi, Narendra Publishing House.

Ludemann, J., Verissimo, K. M., Dreger, K., Fago, A., Schneider, I. and Burmester, T. (2019). "Globin E is a myoglobin-related, respiratory protein highly expressed in lungfish oocytes." Sci Rep **9**(1): 280.

Lynch, M. and Force, A. (2000). "The probability of duplicate gene preservation by subfunctionalization." Genetics **154**(1): 459-473.

Macqueen, D. J., Garcia de la Serrana, D. and Johnston, I. A. (2014). "Cardiac myoglobin deficit has evolved repeatedly in teleost fishes." Biol Lett **10**(6).

Madden, P. W., Babcock, M. J., Vayda, M. E. and Cashon, R. E. (2004). "Structural and kinetic characterization of myoglobins from eurythermal and stenothermal fish species." Comp Biochem Physiol B Biochem Mol Biol **137**(3): 341-350.

Maeda, N. and Fitch, W. M. (1982). "Isolation and amino acid sequence of a monomeric hemoglobin in heart muscle of the bullfrog, Rana catesbeiana." J Biol Chem **257**(6): 2806-2815.

Markl, J. (2013). "Evolution of molluscan hemocyanin structures." Biochim Biophys Acta **1834**(9): 1840-1852.

Merx, M. W., Flogel, U., Stumpe, T., Godecke, A., Decking, U. K. and Schrader, J. (2001). "Myoglobin facilitates oxygen diffusion." FASEB J **15**(6): 1077-1079.

Milton, S. L., Nayak, G., Lutz, P. L. and Prentice, H. M. (2006). "Gene transcription of neuroglobin is upregulated by hypoxia and anoxia in the brain of the anoxia-tolerant turtle Trachemys scripta." J Biomed Sci **13**(4): 509-514.

Mirceta, S., Signore, A. V., Burns, J. M., Cossins, A. R., Campbell, K. L. and Berenbrink, M. (2013). "Evolution of mammalian diving capacity traced by myoglobin net surface charge." Science **340**(6138): 1234192.

Mitz, S. A., Reuss, S., Folkow, L. P., Blix, A. S., Ramirez, J. M., Hankeln, T. and Burmester, T. (2009). "When the brain goes diving: glial oxidative metabolism may confer hypoxia tolerance to the seal brain." Neuroscience **163**(2): 552-560.

Moncada, S. and Erusalimsky, J. D. (2002). "Does nitric oxide modulate mitochondrial energy generation and apoptosis?" Nat Rev Mol Cell Biol **3**(3): 214-220.

Munshi, J. S. (1961). "The accessory respiratory organs of Clarias batrachus (Linn.)." J Morphol **109**: 115-139.

Nakatani, K., Okuyama, H., Shimahara, Y., Saeki, S., Kim, D. H., Nakajima, Y., . . . Yoshizato, K. (2004). "Cytoglobin/STAP, its unique localization in splanchnic fibroblast-like cells and function in organ fibrogenesis." Lab Invest **84**(1): 91-101.

Nelson, J. S. G., Terry C.; Wilson, Mark V. H. (2016). Fishes of the World. Hoboken, New Jersey, USA, Wiley.

Nguyen, L. T., Schmidt, H. A., von Haeseler, A. and Minh, B. Q. (2015). "IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies." Mol Biol Evol **32**(1): 268-274.

Nogueira, A. F., Costa, C. M., Lorena, J., Moreira, R. N., Frota-Lima, G. N., Furtado, C., . . . Schneider, I. (2016). "Tetrapod limb and sarcopterygian fin regeneration share a core genetic programme." Nat Commun **7**: 13364.

Norman, J. R. G., Peter Humphry (1963). A History of Fishes, Hill and Wang.

Ohno, S. (1970). Evolution by Gene Duplication. New York, Springer-Verlag.

Oldham, W. M. and Hamm, H. E. (2006). "Structural basis of function in heterotrimeric G proteins." Q Rev Biophys **39**(2): 117-166.

Oldham, W. M. and Hamm, H. E. (2008). "Heterotrimeric G protein activation by G-protein-coupled receptors." Nat Rev Mol Cell Biol **9**(1): 60-71.

Oleksiewicz, U., Liloglou, T., Field, J. K. and Xinarianos, G. (2011). "Cytoglobin: biochemical, functional and clinical perspective of the newest member of the globin family." Cell Mol Life Sci **68**(23): 3869-3883.

Oleksiewicz, U., Liloglou, T., Tasopoulou, K. M., Daskoulidou, N., Bryan, J., Gosney, J. R., . . . Xinarianos, G. (2013). "Cytoglobin has bimodal: tumour suppressor and oncogene functions in lung cancer cell lines." Hum Mol Genet **22**(16): 3207-3217.

Opazo, J. C., Lee, A. P., Hoffmann, F. G., Toloza-Villalobos, J., Burmester, T., Venkatesh, B. and Storz, J. F. (2015). "Ancient Duplications and Expression Divergence in the Globin Gene Superfamily of Vertebrates: Insights from the Elephant Shark Genome and Transcriptome." Mol Biol Evol **32**(7): 1684-1694.

Osawa, Y. and Korzekwa, K. (1991). "Oxidative modification by low levels of HOOH can transform myoglobin to an oxidase." Proc Natl Acad Sci U S A **88**(16): 7081-7085.

Patel, V. S., Cooper, S. J., Deakin, J. E., Fulton, B., Graves, T., Warren, W. C., . . . Graves, J. A. (2008). "Platypus globin genes and flanking loci suggest a new insertional model for beta-globin evolution in birds and mammals." BMC Biol **6**: 34.

Patel, V. S., Ezaz, T., Deakin, J. E. and Graves, J. A. (2010). "Globin gene structure in a reptile supports the transpositional model for amniote alpha- and beta-globin gene evolution." Chromosome Res **18**(8): 897-907.

Pedersen, C. L., Faggiano, S., Helbo, S., Gesser, H. and Fago, A. (2010). "Roles of nitric oxide, nitrite and myoglobin on myocardial efficiency in trout (Oncorhynchus mykiss) and goldfish (Carassius auratus): implications for hypoxia tolerance." J Exp Biol **213**(Pt 16): 2755-2762.

Perutz, M. F. (1979). "Regulation of oxygen affinity of hemoglobin: influence of structure of the globin on the heme iron." Annu Rev Biochem **48**: 327-386.

Perutz, M. F. (1983). "Species adaptation in a protein molecule." Molecular Biology and Evolution **1**(1): 1-28.

Pesce, A., Bolognesi, M., Bocedi, A., Ascenzi, P., Dewilde, S., Moens, L., . . . Burmester, T. (2002). "Neuroglobin and cytoglobin. Fresh blood for the vertebrate globin family." EMBO Rep **3**(12): 1146-1151.

Pinder, A. W. S., K. B.; Ultsch, G. R. (1992). Estivation and Hibernation. . Chicago, University of Chicago Press.

Qi, D., Chao, Y., Zhao, Y., Xia, M. and Wu, R. (2018). "Molecular evolution of myoglobin in the Tibetan Plateau endemic schizothoracine fish (Cyprinidae, Teleostei) and tissue-specific expression changes under hypoxia." Fish Physiol Biochem **44**(2): 557-571.

Qiu, Y., Sutton, L. and Riggs, A. F. (1998). "Identification of myoglobin in human smooth muscle." J Biol Chem **273**(36): 23426-23432.

Reeder, B. J., Svistunenko, D. A. and Wilson, M. T. (2011). "Lipid binding to cytoglobin leads to a change in haem co-ordination: a role for cytoglobin in lipid signalling of oxidative stress." Biochem J **434**(3): 483-492.

Roesner, A., Fuchs, C., Hankeln, T. and Burmester, T. (2005). "A globin gene of ancient evolutionary origin in lower vertebrates: evidence for two distinct globin families in animals." Mol Biol Evol **22**(1): 12-20.

Roesner, A., Mitz, S. A., Hankeln, T. and Burmester, T. (2008). "Globins and hypoxia adaptation in the goldfish, Carassius auratus." FEBS J **275**(14): 3633-3643.

Rohlfing, K. (2016). Evolution der Globindiversität der Vertebraten. PhD, University of Hamburg.

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**(3): 140-147.

Saha, N., Datta, S., Kharbuli, Z. Y., Biswas, K. and Bhattacharjee, A. (2007). "Air-breathing catfish, Clarias batrachus upregulates glutamine synthetase and carbamyl phosphate synthetase III during exposure to high external ammonia." Comp Biochem Physiol B Biochem Mol Biol **147**(3): 520-530.

Sanchez, A. P. G., M. L. (2001). "Effects of environmental hypercapnia on pulmonary ventilation of the South American lungfish." Journal of Fish Biology **58**: 1181-1189.

Schmidt-Nielsen, K. (1990). Animal Physiology: Adaptation and Environment. Cambridge, Cambridge University Press.

Schmidt, M., Gerlach, F., Avivi, A., Laufs, T., Wystub, S., Simpson, J. C., . . . Burmester, T. (2004). "Cytoglobin is a respiratory protein in connective tissue and neurons, which is up-regulated by hypoxia." J Biol Chem **279**(9): 8063-8069.

Schmidt, M., Laufs, T., Reuss, S., Hankeln, T. and Burmester, T. (2005). "Divergent distribution of cytoglobin and neuroglobin in the murine eye." Neurosci Lett **374**(3): 207-211.

Schwarze, K. (2011). Die Organisation der Globingene des Meerneunauges (Petromyzon marinus) und des Rotkehlanolis (Anolis carolinensis). PhD, University of Hamburg.

Schwarze, K. and Burmester, T. (2013). "Conservation of globin genes in the "living fossil" Latimeria chalumnae and reconstruction of the evolution of the vertebrate globin family." Biochim Biophys Acta **1834**(9): 1801-1812.

Schwarze, K., Campbell, K. L., Hankeln, T., Storz, J. F., Hoffmann, F. G. and Burmester, T. (2014). "The globin gene repertoire of lampreys: convergent evolution of hemoglobin and myoglobin in jawed and jawless vertebrates." Mol Biol Evol **31**(10): 2708-2721.

Schwarze, K., Singh, A. and Burmester, T. (2015). "The Full Globin Repertoire of Turtles Provides Insights into Vertebrate Globin Evolution and Functions." Genome Biol Evol **7**(7): 1896-1913.

Secor, S. M. and Lignot, J. H. (2010). "Morphological plasticity of vertebrate aestivation." Prog Mol Subcell Biol **49**: 183-208.

Seidel, M. E. (1978). "Terrestrial dormancy in the turtle Kinosternon flavescens: Respiratory metabolism and dehydration." Comparative Biochemistry and Physiology Part A: Physiology **61**(1): 1-4.

Sharma, V. S., Traylor, T. G., Gardiner, R. and Mizukami, H. (1987). "Reaction of nitric oxide with heme proteins and model compounds of hemoglobin." Biochemistry **26**(13): 3837-3843.

Shih, L., Chung, Y., Sriram, R. and Jue, T. (2014). "Palmitate interaction with physiological states of myoglobin." Biochim Biophys Acta **1840**(1): 656-666.

Shiva, S., Huang, Z., Grubina, R., Sun, J., Ringwood, L. A., MacArthur, P. H., . . . Gladwin, M. T. (2007). "Deoxymyoglobin is a nitrite reductase that generates nitric oxide and regulates mitochondrial respiration." Circ Res **100**(5): 654-661.

Shiva, S., Rassaf, T., Patel, R. P. and Gladwin, M. T. (2011). "The detection of the nitrite reductase and NO-generating properties of haemoglobin by mitochondrial inhibition." Cardiovasc Res **89**(3): 566-573.

Sick, H. and Gersonde, K. (1969). "Method for continuous registration of O2-binding curves of hemoproteins by means of a diffusion chamber." Anal Biochem **32**(3): 362-376.

Sidell, B. D. and O'Brien, K. M. (2006). "When bad things happen to good fish: the loss of hemoglobin and myoglobin expression in Antarctic icefishes." J Exp Biol **209**(Pt 10): 1791-1802.

Simon, M. I., Strathmann, M. P. and Gautam, N. (1991). "Diversity of G proteins in signal transduction." Science **252**(5007): 802-808.

Song, S., Liu, L., Edwards, S. V. and Wu, S. (2012). "Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model." Proc Natl Acad Sci U S A **109**(37): 14942-14947.

Sprang, S. R. (1997). "G protein mechanisms: insights from structural analysis." Annu Rev Biochem **66**: 639-678.

Stamatoyannopoulos, G. (1991). "Human hemoglobin switching." Science **252**(5004): 383.

Storey, K. B. (2000). "Turning Down The Fires Of Life: Metabolic Regulation Of Hibernation And Estivation." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **126, Supplement 1**: S90.

Storz, J. F., Opazo, J. C. and Hoffmann, F. G. (2011). "Phylogenetic diversification of the globin gene superfamily in chordates." IUBMB Life **63**(5): 313-322.

Storz, J. F., Opazo, J. C. and Hoffmann, F. G. (2013). "Gene duplication, genome duplication, and the functional diversification of vertebrate globins." Mol Phylogenet Evol **66**(2): 469-478.

Sugimoto, H., Makino, M., Sawai, H., Kawada, N., Yoshizato, K. and Shiro, Y. (2004). "Structural basis of human cytoglobin for ligand binding." J Mol Biol **339**(4): 873-885.

Svergun, D. I., Petoukhov, M. V. and Koch, M. H. (2001). "Determination of domain structure of proteins from X-ray solution scattering." Biophys J **80**(6): 2946-2953.

Szymanski, M., Wang, R., Fallin, M. D., Bassett, S. S. and Avramopoulos, D. (2010). "Neuroglobin and Alzheimer's dementia: genetic association and gene expression changes." Neurobiol Aging **31**(11): 1835-1842.

Tajima, F. (1993). "Simple methods for testing molecular clock hypothesis." Genetics **135**: 599-607.

Takezaki, N. and Nishihara, H. (2017). "Support for Lungfish as the Closest Relative of Tetrapods by Using Slowly Evolving Ray-Finned Fish as the Outgroup." Genome Biol Evol **9**(1): 93-101.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0." Molecular Biology and Evolution **30**: 2725-2729.

Tejero, J. and Gladwin, M. T. (2014). "The globin superfamily: functions in nitric oxide formation and decay." Biol Chem **395**(6): 631-639.

Tiedke, J., Gerlach, F., Mitz, S. A., Hankeln, T. and Burmester, T. (2011). "Ontogeny of globin expression in zebrafish (Danio rerio)." J Comp Physiol B **181**(8): 1011-1021.

Tiedke, J., Thiel, R. and Burmester, T. (2014). "Molecular response of estuarine fish to hypoxia: a comparative study with ruffe and flounder from field and laboratory." PLoS One **9**(3): e90778.

Tiso, M., Tejero, J., Basu, S., Azarov, I., Wang, X., Simplaceanu, V., . . . Gladwin, M. T. (2011). "Human neuroglobin functions as a redox-regulated nitrite reductase." J Biol Chem **286**(20): 18277-18289.

Totzeck, M., Hendgen-Cotta, U. B., Kelm, M. and Rassaf, T. (2014). "Crosstalk between nitrite, myoglobin and reactive oxygen species to regulate vasodilation under hypoxia." PLoS One **9**(8): e105951.

Trent, J. T., 3rd and Hargrove, M. S. (2002). "A ubiquitously expressed human hexacoordinate hemoglobin." J Biol Chem **277**(22): 19538-19545.

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A. and Minh, B. Q. (2016). "W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis." Nucleic Acids Res **44**(W1): W232-235.

Van Acker, Z. P., Luyckx, E. and Dewilde, S. (2018). "Neuroglobin Expression in the Brain: a Story of Tissue Homeostasis Preservation." Mol Neurobiol.

Vervoort, A. (1980). "Tetraploidy in Protopterus (Dipnoi)." Experientia **36**: 294-296.

Vetter, I. R. and Wittinghofer, A. (2001). "The guanine nucleotide-binding switch in three dimensions." Science **294**(5545): 1299-1304.

Vinogradov, S. N., Bailly, X., Smith, D. R., Tinajero-Trejo, M., Poole, R. K. and Hoogewijs, D. (2013). "Microbial eukaryote globins." Adv Microb Physiol **63**: 391-446.

Vinogradov, S. N., Hoogewijs, D., Bailly, X., Mizuguchi, K., Dewilde, S., Moens, L. and Vanfleteren, J. R. (2007). "A model of globin evolution." Gene **398**(1-2): 132-142.

Wakasugi, K. and Morishima, I. (2005). "Identification of residues in human neuroglobin crucial for Guanine nucleotide dissociation inhibitor activity." Biochemistry **44**(8): 2943-2948.

Wakasugi, K., Nakano, T. and Morishima, I. (2003). "Oxidized human neuroglobin acts as a heterotrimeric Galpha protein guanine nucleotide dissociation inhibitor." J Biol Chem **278**(38): 36505-36512.

Wang, H. and Joseph, J. A. (1999). "Quantifying cellular oxidative stress by dichlorofluorescein assay using microplate reader." Free Radical Biology and Medicine **27**(5-6): 612-616.

Watanabe, S., Takahashi, N., Uchida, H. and Wakasugi, K. (2012). "Human neuroglobin functions as an oxidative stress-responsive sensor for neuroprotection." J Biol Chem **287**(36): 30128-30138.

Watanabe, S. and Wakasugi, K. (2008). "Neuroprotective function of human neuroglobin is correlated with its guanine nucleotide dissociation inhibitor activity." Biochem Biophys Res Commun **369**(2): 695-700.

Weber, R. E. (1981). "Cationic control of O2 affinity in lugworm erythrocruorin." Nature **292**: 386.

Weber, R. E. (1992). "Use of ionic and zwitterionic (Tris/BisTris and HEPES) buffers in studies on hemoglobin function." J Appl Physiol (1985) **72**(4): 1611-1615.

Weber, R. E. and Vinogradov, S. N. (2001). "Nonvertebrate hemoglobins: functions and molecular adaptations." Physiol Rev **81**(2): 569-628.

Whitmore, L., Miles, A. J., Mavridis, L., Janes, R. W. and Wallace, B. A. (2017). "PCDDB: new developments at the Protein Circular Dichroism Data Bank." Nucleic Acids Res **45**(D1): D303-D307.

Wilson, D. M. (1959). "Function of Giant Mauthner's Neurons in the Lungfish." Science **129**(3352): 841-842.

Wittenberg, B. A. and Wittenberg, J. B. (1989). "Transport of oxygen in muscle." Annual Review of Physiology **51**: 857-878.

Wittenberg, J. B. and Wittenberg, B. A. (2003). "Myoglobin function reassessed." J Exp Biol **206**(Pt 12): 2011-2020.

Wolfe, K. H. (2001). "Yesterday's polyploids and the mystery of diploidization." Nat Rev Genet **2**(5): 333-341.

Wolfe, K. H. and Li, W. H. (2003). "Molecular evolution meets the genomics revolution." Nat Genet **33 Suppl**: 255-265.

Wood, W. G. (1976). "Haemoglobin synthesis during human fetal development." Br Med Bull **32**(3): 282-287.

Wystub, S., Ebner, B., Fuchs, C., Weich, B., Burmester, T. and Hankeln, T. (2004). "Interspecies comparison of neuroglobin, cytoglobin and myoglobin: sequence evolution and candidate regulatory elements." Cytogenet Genome Res **105**(1): 65-78.

Xi, Y., Obara, M., Ishida, Y., Ikeda, S. and Yoshizato, K. (2007). "Gene expression and tissue distribution of cytoglobin and myoglobin in the Amphibia and Reptilia: possible compensation of myoglobin with cytoglobin in skeletal muscle cells of anurans that lack the myoglobin gene." Gene **398**(1-2): 94-102.

Xu, R., Harrison, P. M., Chen, M., Li, L., Tsui, T. Y., Fung, P. C., . . . Farzaneh, F. (2006). "Cytoglobin overexpression protects against damage-induced fibrosis." Mol Ther **13**(6): 1093-1100.

Yang, Z. (1994). "Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods." J Mol Evol **39**(3): 306-314.

Yang, Z. (2007). "PAML 4: phylogenetic analysis by maximum likelihood." Molecular Biology and Evolution **24**(8): 1586-1591.

Yu, D. Y. and Cringle, S. J. (2001). "Oxygen distribution and consumption within the retina in vascularised and avascular retinas and in animal models of retinal disease." Prog Retin Eye Res **20**(2): 175-208.

Yu, Z., Liu, N., Li, Y., Xu, J. and Wang, X. (2013a). "Neuroglobin overexpression inhibits oxygen-glucose deprivation-induced mitochondrial permeability transition pore opening in primary cultured mouse cortical neurons." Neurobiol Dis **56**: 95-103.

Yu, Z., Poppe, J. L. and Wang, X. (2013b). "Mitochondrial mechanisms of neuroglobin's neuroprotection." Oxid Med Cell Longev **2013**: 756989.

Yu, Z., Xu, J., Liu, N., Wang, Y., Li, X., Pallast, S., . . . Wang, X. (2012). "Mitochondrial distribution of neuroglobin and its response to oxygen-glucose deprivation in primary-cultured mouse cortical neurons." Neuroscience **218**: 235-242.

Zaccone, G. D., K.; Hedrick, Michael S.; Fernandes, J.M.O.; Icardo, Josè M. (2016). Phylogeny, Anatomy and Physiology of Ancient Fishes. Boca Raton, CRC Press.

Zhao, Z. X., Xu, P., Cao, D. C., Kuang, Y. Y., Deng, H. X., Zhang, Y., . . . Sun, X. W. (2014). "Duplication and differentiation of common carp (Cyprinus carpio) myoglobin genes revealed by BAC analysis." Gene **548**(2): 210-216.

VII. Danksagung

Zu allererst gilt mein Dank Herrn Prof. Dr. Thorsten Burmester, der mir dieses interessante Forschungsthema zur Verfügung gestellt hat und mich während der ganzen Zeit sehr gut betreut hat. Ich danke dir von Herzen und wünsche dir weiterhin eine gute Genesung.

Nicht minder weniger möchte ich mich bei Frau Prof. Dr. Susanne Dobler bedanken. Vielen, vielen Dank, dass du so kurzfristig die Betreuung und alle weiteren, damit verbundenen Aufgaben übernommen hast und mir bei allen Fragen mit Rat und Tat zur Seite gestanden hast. Das hat mir wirklich sehr geholfen.

Additionally, I would like to thank Prof. Dr. Angela Fago from Aarhus University. Thank you so much for all your help, support and the professional suggestions on my thesis. Furthermore, I would like to thank you and Elin for the help on my experiments, your hospitality and the possibility, to perform some of my experiments in your lab.

Ich danke meiner Masterstudentin Kimberley Dreger und unseren TA's, insbesondere Michelle Kruse, für die Hilfe bei der Durchführung einiger Versuche. Ihr habt mir mit eurer Unterstützung und euren tollen Ergebnissen sehr geholfen und dafür bin ich euch sehr dankbar. Vielen Dank für eure Mühe und euer Interesse an meinem Forschungsthema.

Ich danke zusätzlich Dr. Andrej Fabrizius für die stets kompetente und unterstützende Hilfe, das geduldige Gegenlesen meiner Arbeit und jeder weiteren Unterstützung, die ich erfahren habe. Und auch Danke an Dr. Cornelia Geßner fürs Zuhören, die vielen Kaffees und die intensiven Gespräche (nicht nur beruflicher Art).

Ein besonderer Dank gilt außerdem allen meinen Doktoranden-Kollegen: Kathrin Helfenrath und Elsa Ziegler, mit denen ich das Glück hatte, von Beginn an zusammenzuarbeiten, zusammen Konferenzen, Weiterbildungen und Sommerfeste zu besuchen, die immer mitgefiebert haben und mit denen ich viel Spaß hatte. Alena Krüger, mit der ich seit 3 Jahren das Büro teile, das ein oder andere (nicht nur berufliche) Gespräch geführt habe und die die besten Partys organisiert. Danke dir auch dafür. Paressa Papadopoulou-Wörner, Annette Schlosser und Gerrit Martens: Danke fürs Anteilnehmen, fürs Ohr leihen und die netten Gespräche beim Mittagessen. Katja Reimann, Michelle Kruse und Maria Machola: Danke für die

stets kompetente Hilfe, die Ideen, wenn ich selbst nicht weiterwusste, und die nette Atmosphäre.

Also big thanks to you, Dr. Nikolay Gresko, for carefully reading the manuscript and checking on my English. You are the best roommate I ever had.

Der größte Dank gebührt meiner Familie. Vielen Dank Mama und Papa, dass ihr mir das Studium sowohl finanziell als auch emotional ermöglicht habt und mich auch sonst in allem unterstützt und mich in meinen Entscheidungen bestärkt habt. Ihr habt mir immer vermittelt, dass ich alles schaffen kann, wenn ich es möchte und auch Zweifel ok sind. Ich danke auch meiner Schwester Katja, die trotz Zeitstress und ‚wenig' biologischem Hintergrund diese Arbeit gegengelesen hat und zur Not immer Kuchen parat hatte. Zum Schluss Danke ich dir Tim, für alles. Ich freue mich darauf, bald Frau Orth sein zu dürfen!

VIII. Eidesstattliche Versicherung (Declaration on oath)

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, Juni 2019

_____

Julia Katrin Lüdemann

IX. Declaration of own contribution to the manuscripts

Chapter 1: Koch, J., Ludemann, J., Spies, R., Last, M., Amemiya, C. T. and Burmester, T. (2016). "Unusual Diversity of Myoglobin Genes in the Lungfish." Mol Biol Evol 33(12): 3033-3041.

Thorsten Burmester designed the research, Jonas Koch, Rieke Spies, Marco Last and I performed the research. I did the tissue preparation, RNA extraction, cDNA cloning and RT-qPCR with one of the two lungfish specimen and the analysis of the myoglobin proteins. Jonas Koch, Thorsten Burmester and I analyzed the data and wrote the paper.

Chapter 2: Ludemann, J., Fago, A., Falke, S., Wisniewsky, M., Schneider, I., Fabrizius, A. and Burmester, T. (2019). "Genetic and functional diversity of the multiple lungfish myoglobins." in press

Thorsten Burmester designed the research and I performed all the experiments with the help of Angela Fago, Sven Falke and Michelle Wisniewsky and analyzed them with the help of Sven Falke, Igor Schneider, Angela Fago and Thorsten Burmester. I then wrote the manuscript together with Thorsten Burmester.

Chapter 3: Ludemann, J., Verissimo, K. M., Dreger, K., Fago, A., Schneider, I. and Burmester, T. (2019). "Globin E is a myoglobin-related, respiratory protein highly expressed in lungfish oocytes." Sci Rep 9(1): 280.

T. Burmester designed the research and I performed most of the experiments besides the *in situ* hybridization, which was done by Kellen Matos Verissimo. The RNA extraction, library preparation, Illumina sequencing and qRT-PCR from *P. aethiopicus* tissues were done by Kimberley Dreger within the scope of a master thesis which I supervised. Afterwards I analyzed the data together with Igor Schneider, Angela Fago and Thorsten Burmester. Thorsten Burmester wrote the manuscript and I revised it together with Igor Schneider and Angela Fago.