Identification, annotation and comparative analysis of cytoglobin and neuroglobin genes from gill transcriptome of three freshwater crayfish

Protick Kumar Bose
Student ID: 150629
Session: 2017-2018

Fisheries and Marine Resource Technology Discipline
Khulna University
Khulna-9208, Bangladesh

December , 2018

Identification, annotation and comparative analysis of cytoglobin and neuroglobin genes from gill transcriptome of three freshwater crayfish

A Project Thesis

By

Protick Kumar Bose

Student ID: 150629

Session: 2017-2018

The project thesis submitted to the Fisheries and Marine Resource Technology Discipline in partial fulfillments of the requirements for the degree of

Bachelor of Science (Hon's)

in

Fisheries

Fisheries and Marine Resource Technology Discipline

Khulna University

Khulna-9208, Bangladesh

December, 2018

Identification, annotation and comparative analysis of cytoglobin and neuroglobin genes from gill transcriptome of three freshwater crayfish

Approved as to the style and content

by

.....

(Signature of the Head)

Dr. Ghausiatur Reza Banu

Professor and Head

Fisheries and Marine Resource Technology Discipline

Khulna University

Khulna-9208, Bangladesh

December, 2018

DECLARATION

Identification, annotation and comparative analysis of cytoglobin and neuroglobin genes from gill transcriptome of three freshwater crayfish

The results presented in the thesis with the above mentioned title are entirely of the candidate's own investigations and no part of the results has been accepted for any degree nor is it being concurrently submitted for any degree.

Protick Kumar Bose

.....

(Signature of Candidate)

Protick Kumar Bose

Student ID: 150629

Session: 2017-2018

.....

(Signature of Supervisor)

Dr. Muhammad Yousuf Ali

Professor

Fisheries and Marine Resource Technology Discipline

Khulna University, Khulna

Fisheries and Marine Resource Technology Discipline

Khulna University, Khulna-9208, Bangladesh

December, 2018

DEDICATED TO

MY BELOVED PARENTS

ACKNOWLEDGEMENT

First of all, my admiration goes to the Almighty for His mercy to enable me to pursue my education in Fisheries and Marine Resource Technology Discipline. It is my great pleasure for being able to complete this thesis for the degree of Bachelor of Science in Fisheries for which I am completely indebted to the Creator of this Universe, The Almighty.

I would like to express my profound appreciation and sincere gratitude to my honorable supervisor, **Dr. Muhammad Yousuf Ali,** Professor, Fisheries and Marine Resource Technology (FMRT) Discipline for his scholastic guidance, constructive criticism, constant encouragement and sustained support that rationalized me to complete this thesis paper successfully.

I would like to express my profound gratitude to Professor **Dr. Ghausiatur Reza Banu**, Head, Fisheries and Marine Resource Technology Discipline, Khulna University for her providence of all necessary facilities.

My heartiest thanks go to all my respected teachers in Fisheries and Marine Resource Technology Discipline, Khulna University and friends whose inspiration was all along with me in the hard time during the course of the present study and for their endless support and encouragement.

Finally, I would like to express my deepest gratitude to my beloved parents for their immeasurable blessing and continuous inspiration.

Protick Kumar Bose
Roll No: 150629
9 December, 2018

Abstract

Neuroglobin and cytoglobin, both proteins bind oxygen reversibly and involve in cellular oxygen homeostasis. We identified and annotated all these genes from gill transcriptome of three commercially important freshwater crayfish, *Cherax quadricarinatus*, *C. destructor* and *C. cainii*. The sequences were made publicly available to NCBI GenBank except *C. quadricarinatus*, But data was submitted in NCBI. Comparative molecular analysis of these genes across the species were also carried out . The cytoglobin obtained from the species *Cherax quadricarinatus*, *C. destructor* and *C. cainii* were named as CqCygb, CdCygb and CcCygb; and neuroglobin named as CqNgb, CdNgb and CcNgb respectively. The deduced proteins encoded by CqCygb, CdCygb and CcCygb contained 190 amino acids; and those encoded by CqNgb, CdNgb and CcNgb were 209 aa. Cygb obtained from our study shared high similarity with Cygbs from *Carcinus maenas* (70% identity), *Penaeus vannamei* (62% identity) and *Hyalella Azteca* (52% identity) and Ngb shared similarity with Ngbs from, *Microplitis demolitor* (63% identity), *Penaeus vannamei* (80% identity). Neuroglobin and cytoglobin are highly conserved, displaying very low levels of non-synonymous nucleotide substitutions. A total of 9 nonsynonymous mutation (amino acid replacement) and 20 synonymous mutations were observed in Cygbs from three crayfish species, while 10 nonsynonymous mutation and 40 synonymous mutation were found in Ngbs across the species. However, all the changes were outside the functional sites. Cygb is more conserve than Ngb. Findings of the study will assist in further molecular analysis and gene expression study.

CONTENT

LIST OF TABLES

LIST OF FIGURES

Chapter One

Introduction:

1.1. Background of the study:

Globins are respiratory proteins that usually bind an oxygen molecule between the iron ion of the porphyrin ring and a histidine of the polypeptide chain. Globins have been found in bacteria, plants, fungi, and animals and play an important role in the respiratory system (Thorsten Burmester et al. 2002). Cytoglobin and neuroglobin are two globin protein that play a vital role in stress management. The stress may come from sudden changes in environmental parameters like O_2, temperature and salinity (LaCount et al. 2000; Thorsten Burmester et al. 2002).

Cytoglobin (Cygb) is a distant evolutionary relative of Mb and is also able to bind oxygen (Trent and Hargrove 2002b). Cygb mRNA is expressed in virtually all mammalian tissues (hence its name). Immunohistochemistry studies show that Cygb is predominantly present in the cytoplasm of fibroblast like cells of connective tissue in very many organs (Schmidt-Kastner et al. 2006) suggesting a physiological function related to the metabolism of the fibro- blast cell lineage. In few other tissues (e.g. brain), Cygb may also be present within the cell nucleus, but its role there is unclear (Schmidt et al. 2003). In order to characterize the modes of gene regulation underlying the cellular functions of Ngb and Cygb, it is important to identify their candidate gene regulatory sequences. Comparative genomic sequence analysis has proven to be a valuable tool in extracting phylogenetically conserved, and thus functionally important, coding and non-coding sequences (Frazer et al. 2003).

Neuroglobin is newly found protein in the globin family which found in the brain of human and rat (Thorsten Burmester, Weich, and Reinhardt 2000). Historically, however, neuroglobin were first observed in invertebrate taxa. Lankester (Gould 1979) had already recorded the brilliant red color of the ganglia of the polychaete annelid *Aphrodite aculeate.* Since then, nHbs have also been found in, or associated with, nerve tissues of several other invertebrate taxa, such as molluscs, arthropods, nemerteans, and nematode species (Dewilde et al. 2006). With few exceptions (e.g. in air-breathing gastropods), the invertebrate Ngbs are expressed at high concentrations (in the mM range) and have therefore been suggested to support nerve function by mediating O_2 storage and/or transport during temporary hypoxia (Wittenberg 1992). For example, in the gastropod mollusc

Aplysia depilans, the oxygenation state of the nHb was correlated with the electrical activity of the neural ganglia, and firing was shown to be proportional to the degree of oxygenation of the globin. Ngb is only distantly related to the Hbs and Mb (ca. 25% amino acid identity), but still features the conserved protein fold which is characteristic for oxygen binding globins (Pesce et al., 2003). Recombinant Among several possible functions of Ngb (Wystub et al. 2003; Pesce et al. 2002), it still seems most probable that Ngb serves to supply oxygen to the metabolically very active and O_2 demanding neuronal and endocrine tissues (Thorsten Burmester, Weich, and Reinhardt 2000). Immuno histochemical data suggest that Ngb protein is differentially expressed in various regions of the mouse brain, suggesting a functional correlation (Thorsten Burmester, Weich, and Reinhardt 2000; Wystub et al. 2003; Hundahl et al. 2008), and the most prominent Ngb expression to date has been found in the extremely oxygen consuming neuroretina. Ngb has been shown to enhance the survival of neurons both in vitro and in vivo in the mouse brain under hypoxic conditions. Ngb was found to be upregulated in hypoxic cell culture (D'Ancona and Fanelli 2009). Other authors, however, did not find any alteration of Ngb mRNA in the mouse brain in vivo under long term hypoxia.

Cygb is a distant evolutionary relative of Mb and is also able to bind oxygen (Trent and Hargrove 2002b). Cygb mRNA is expressed in virtually all mammalian tissues (hence its name). Immuno histochemistry studies show that Cygb is predominantly present in the cytoplasm of fibroblast like cells of connective tissue in very many organs (Schmidt-Kastner et al. 2006) suggesting a physiological function related to the metabolism of the fibroblast cell lineage. In few other tissues (e.g. brain), Cygb may also be present within the cell nucleus, but its role there is unclear (Schmidt et al. 2003). In order to characterize the modes of gene regulation underlying the cellular functions of Ngb and Cygb, it is important to identify their candidate gene regulatory sequences. Comparative genomic sequence analysis has proven to be a valuable tool in extracting phylogenetically conserved, and thus functionally important, coding and non-coding sequences (Frazer et al. 2003).

Red claw (*Cherax quadricarinatus*), Yabby *(C. destructor*) and Marron (*C. cainii*) are the most economically important commercial species endemic to Australia, and naturally distributed in three diverse aquatic environments. They can tolerate a wide range of environmental fluctuations in terms of oxygen, temperature, salinity and pH (McCormack, 2014). Thus, because of the

adaptive physiological behaviors of these animals and their economic significance, they can be used as an important model organism for physiological research in freshwater crustaceans.

Fig 1: Red claw (*Cherax quadricarinatus*)

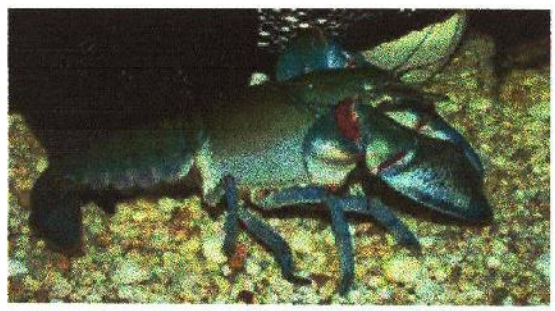

Fig 2: Yabby (*Cherax destructor*)

Fig 3: Maroon (*Cherax cainii*)

Although few previous works have reported some stress-related genes (Ali et al., 2015a; Ali et al., 2015b; Glazer et al., 2013; Pavasovic et al., 2013), there is no information on cytoglobin and neuroglobin in crayfish. Most of the previous studies reported cytoglobin and neuroglobin from vertebrate species (Heaslet and Royer 1999). Taken it in consideration, we identified cytoglobin and neuroglobin genes from gill transcriptomes of *Cherax quadricarinatus*, *C. destructor* and *C. cainii* and made a comparative molecular analysis across the species. The genes were made publicly available through submission in NCBI database (National Centre for Biotechnological Information). Genomic information on cytoglobin and neuroglobin obtained from this study will provide a strong foundation for future gene expression study and comparative genomic research on globin proteins in freshwater crayfish in general, and will assist in understanding their response to oxygen stress in particular.

Most of the study was happened regarding to the vertebrate species. Here we study about the status of globin protein (Ngb, CYgb) in invertebrate especially in crustaceans (Crayfish). We investigate the three species of crustaceans mRNA sequence of globin protein and analyze and compare the sequence with other organisms. We give a estimation about the similarity and conservative nature of globin family.

1.2 Literature Review

1.2.1. Globin Protein:

The number of globin families has grown markedly since the 1970's, when they were limited to the α- and β-globins and Mbs, mostly of vertebrates, and the SHbs (legHbs) of legume plants. The ensuing years brought to light several new globins: SHbs in plants other than legumes, NsHbs in a wide variety of plants, and FHbs, chimeric proteins (~400aa) comprising an N-terminal globin domain and a C-terminal FAD- and NAD-binding domain, related to the ferredoxin NADP+ reductases, found in E. coli and in yeasts. Concurrently, "truncated" globins, sequences shorter than normal (<130aa), were discovered in protozoa, cyanobacteria, a nemertean , and bacteria.(Serge N. Vinogradov et al. 2006)

Globins are heme-containing proteins that bind O_2 and other gaseous ligands between the iron atom at the center of the porphyrin ring and a histidine residue of their polypeptide chain . In addition to supporting aerobic metabolism of cells by providing O_2 supply, globins fulfill a broad range of other functions, including O_2 sensing, detoxification of harmful reactive oxygen species (ROS), the generation of bioactive gas molecules like NO and others (Serge N. Vinogradov et al. 2006) . Thus it is not surprising that the versatile globins are found from bacteria to fungi, protists, plants, and most animal groups. Some years ago, the vertebrate globin gene family was expanded by the discovery of two additional globin types, neuroglobin and cytoglobin (S. N. Vinogradov et al. 2005). Neuroglobin (Ngb) is preferentially expressed in neurons and endocrine cells, and its expression patterns suggest an association with oxidative metabolism and the presence of mitochondria . Cytoglobin (Cygb) is expressed in fibroblast related cell types and distinct neuronal cell populations . The exact physiological function(s) of both proteins are still uncertain, and several, partially contradictory hypotheses have been proposed, including functions in O_2 supply,

ROS detoxification, signal transduction and inhibition of apoptosis . In the biomedical field, Ngb and Cygb have created considerable interest because these proteins appear to convey protection to cells and organs, e.g. after ischemia/reperfusion injury of the brain (Bettina Ebner et al. 2010).

1.2.2. Cytoglobin:

Vertebrates possess multiple respiratory globins that differ in terms of structure, function, and tissue distribution. Three types of globins have been described so far: hemoglobin facilitates the transport of oxygen in the blood, myoglobin serves oxygen transport and storage in the muscle, and neuroglobin has a yet unidentified function in nerve cells. Here we report the identification of a fourth and novel type of globin in mouse, man, and zebrafish. It is expressed in apparently all types of human tissue and therefore has been called cytoglobin (CYGB). Mouse and human CYGBs comprise 190 amino acids; the zebrafish CYGB, 174 amino acids. The human CYGB gene is located on chromosome 17q25. The mammalian genes display a unique exon-intron pattern with an additional exon resulting in a C-terminal extension of the protein, which is absent in the fish CYGB. Phylogenetic analyses suggest that the CYGBs had a common ancestor with vertebrate myoglobins (Thorsten Burmester et al. 2002). cytoglobin is significantly upregulated upon hypoxia and this is dependent on the tissue and severity of hypoxia (Fordel et al. 2004).

1.2.3. Structure of Cytoglobin:

Cygb shares about 30% amino acid sequence identity with Mb, pointing at a shared evolutionary ancestry (Norifumi Kawada et al. 2001),(Thorsten Burmester et al. 2002),(Trent and Hargrove 2002a). Compared to Mb, mammalian Cygb is unusually long, containing 190 amino acids owing to extensions of about 20 amino acids at both, the N- and the C-terminus. Part of the N-terminal extension may be explained by sequence motif duplication, while the C-terminal extension partly derives from a small additional exon, which has been recruited during mammalian evolution and is lacking in fish Cygb sequences . Irrespective of these functionally elusive terminal extensions, which are also observed in some invertebrate globins, Cygb features the sequence hallmarks of a standard globin, e.g., the key residues Phe(CD1), His(E7) and His(F8). The crystal structure (N Kawada et al. 2001) ultimately proves that the Cygb core folds as a classic globin. Unfortunately, however, no interpretable electron density data were obtained for the extended Cygb termini. In

agreement with spectroscopic data (Trent and Hargrove 2002b). The crystal structure proves that Cygb is also a hexa-coordinated globin.Moreover, Cygb displays an unprecedented large apolar protein matrix cavity next to the heme, which is connected to the exterior and may provide a special ligand tunnelling pathway Ligand-binding kinetics of Cygb are – as in the case of Ngb – determined by the comparatively slow phase of displacing the internal His(E7) ligand, before an external gaseous ligand can then rapidly bind to the iron atom (Szymanski et al. 2010). The resultant O_2 affinity of Cygb is also in the range of 1 Torr .As shown for Ngb, Cygb is able to form an internal disulfide bridge. Reduction of this bond lowers O_2 affinity of Cygb only moderately, by about 2-fold .While there is no evidence for intermolecular disulfide bonds, biochemical and crystallographic data suggest that full-length Cygb might act as a homodimeric protein, while the truncated form is monomeric (Hankeln et al. 2005).

1.2.4. Function of Cytoglobin:

The physiological function of CYGBs still has to be investigated. The protein was identified in a proteomics approach by virtue of its heavily upregulated expression in stellate cells of rat liver (and it was therefore dubbed stellate cell activation–associated protein, STAP). The authors could demonstrate that STAP possesses peroxidase activity, and they speculate upon its role as a scavenger of peroxides in fibrotic liver. In fact, it has been reported before that the dehaloperoxidase enzyme of the marine worm Amphritite is phylogenetically related to globins (LaCount et al. 2000) and has retained its ability to bind oxygen . However, without additional data the possible spectrum of physiological role(s) of CYGB within the broad variety of vertebrate tissues can only be hypothesized (Lebioda et al. 1999). CYGB may be involved in intracellular oxygen storage or transfer, have an enzymatic (peroxidase) function, play a role in O2 sensing, or may bind and detoxify NO (Flogel et al. 2001) . Finally, a combination of several of these functions also seems conceivable.

1.2.5. Gene expression patterns of Cytoglobin:

Cytoglobin mRNA was originally detected by Northern blot hybridization in essentially all tissues of the mammalian body, indicating a very widespread expression pattern (Trent and Hargrove 2002a; Thorsten Burmester et al. 2002). At the same time, the rat ortholog of Cygb was independently isolated by a proteomics approach from the fibroblast-related stellate cells of the liver. Subsequently, it was reported that Cygb protein is localized exclusively within the cell nuclei

in a wide variety of tissues . Two recent studies have re-investigated the expression pattern of Cygb (Schmidt et al. 2004; Nakatani et al. 2004). Using independently derived antibodies, both report that Cygb is cytoplasmatically expressed in fibroblasts and fibroblast-related cell types in a broad variety of splanchnic organs like liver, heart, muscle, gut, kidney, lung and pancreas. The earlier publications on the presence of Cygb (synonym STAP) in the fibroblast-like hepatic stellate cells were therefore confirmed. In addition, Cygb expression was also detected in bone osteoblasts and in tracheal chondroblasts, but not in mature osteocytes and chondrocytes (Schmidt et al. 2003). In summary, the data by the majority support the specific expression of Cygb in the cytoplasm of cells that are actively engaged in the production of extracellular matrix components in visceral organs. Adding complexity to the Cygb expression pattern, our results suggest that Cygb is also expressed in specific, but ill-defined neuronal cell populations in the brain, as well as in peripheral nerve cells. Here, Cygb immunostaining yielded signals in both, the cytoplasm and the nucleus, possibly pointing at a specific role of Cygb in nervous tissues (Hitomi Sawai et al. 2003; Thorsten Burmester et al. 2002; Serge N. Vinogradov and Moens 2008)

1.2.6. Neuroglobin

Neuroglobin is expressed in vertebrates brain and belongs to a branch of the globin family that diverged early in evolution (Thorsten Burmester, Weich, and Reinhardt 2000). Sequence conservation suggests a relevant role in the nervous system, with tight structural restraints. Experiments in vivo and in vitro showed increased hypoxic stress damage upon repressing neuroglobin biosynthesis and improved recovery following overexpression. Neuroglobin shows internal heme hexacoordination, which controls oxygen affinity and kinetics. Neuroglobin concentration, oxygen affinity and enhanced auto oxidation question a role in oxygen delivery; thus it was proposed that the neuroprotective effect might be due to radical scavenging or activation of protection mechanisms. Neuroglobin's structure shows a peculiar internal cavity of very large size. Binding of heme ligands is associated to a conformational change involving the heme that "slides" into the pre-existing cavity and makes the sixth coordination position available. These features may pave the way to an understanding of neuroprotection by neuroglobin (Brunori and Vallone 2007). Neuroglobin, globin X, the hemoglobins from Ciona, as well as some invertebrate nerve and intracellular globins, belong to an ancient branch of globins that diverged from other globin types before the split of Protostomia and Deuterostomia . The early evolution of the metazoan is

thought to be associated with the emergence of the nervous system. Thus, the phylogenetically basal position of Ngb in the metazoan tree would suggest that it emerged with the evolution of the nerve system (Dröge et al. 2012).

1.2.7. Structure of Neuroglobin:

Ngb is a substantially divergent member of the globin family, displaying only 20–25% amino acid sequence identity to Mbs and Hbs. Ngb represents a typical Mb-type monomeric globin, which can bind O_2 reversibly (Thorsten Burmester, Weich, and Reinhardt 2000).In spite of its sequence differences, Ngb features the conserved globin fold consisting of the eight a-helices A-H, albeit with some peculiarities which reflect a pronounced adaptive potential of this basic globin structure . The crystal structure of human Ngb (Pesce et al. 2003) has recently been solved, revealing the presence of unusual protein cavities which are not found as such in Hb and Mb and which may influence ligand storage and diffusion paths inside the molecule. The most peculiar structural characteristic of Ngb is the so called hexa-coordinated binding scheme of the heme Fe atom in the ferrous (Fe2+) deoxy and in the ferric (Fe3+) states. The crystallographic data have ultimately confirmed several types of spectroscopic analyses (Uno et al. 2004) , showing that in the absence of external ligands, the histidine at position 7 of the E helix (HisE7) binds the heme iron at its sixth, distal position. Thereby, any external gaseous ligand such as O_2 or CO has to compete with the internal His(E7) ligand for Fe binding. This produces a biphasic ligand-binding kinetics for Ngb: the displacement of the His(E7) is the rate-limiting, slow step, while the inherent affinity of the Fe atom after His(E7) displacement is high and makes the gaseous ligand-binding step fast. Heme hexa-coordination has previously been reported in plant, bacteria and invertebrate globins (Uno et al. 2004), and although this widespread occurrence may suggest a conserved function, its physiological significance is not yet understood. Recent kinetic studies show that even slight variations in pH may cause pronounced changes in the association rates of exogenous ligands in Ngb (Uno et al. 2004) .On the other hand, hexa-coordination in Ngb and other globins may render the process of external ligand binding relatively independent of temperature variations [unpublished data], which might be physiologically relevant under conditions of fluctuating body temperatures, e.g., in poikilothermic animals. Another effect of hexa-coordination may be a protection of the Fe3+ iron atom against various oxidizing molecules, which seems to suppress the formation of cytotoxic ferryl (Fe4+)-heme (Herold et al. 2004) .Despite the complex ligand-

binding scheme, the overall O_2-affinity of mammalian and of fish Ngb is in the range of 1 Torr (Fuchs et al. 2004) which is similar to the O2 affinity of Mb. Recently, it was found that human NGB is able to form an internal disulfide bridge at cysteines CD7 and D5 in vitro which may break up under reducing conditions in the cell, e.g., when NADH+ reduction equivalents accumulate under hypoxia. Reduction of the disulfide bond in turn lowers the O2 affinity of Ngb by a factor of 10, which would lead to a release of O_2 and thus, possibly, an attenuation of hypoxic stress. It is not yet clear if this mechanism is acting in vivo: while fish Ngbs possess two cysteines at roughly equivalent positions, rodent Ngbs lack the CD7. With regard to globin function, it is essential to investigate the possibility of Ngb binding to ligands other than O_2, namely the noxious reactive oxygen and nitrogen species, which accumulate in the cell, e.g., after ischemic insult and subsequent reperfusion of the tissue (Becker 2004). EPR (electron paramagnetic resonance) and kinetic studies revealed that the binding affinity of the reactive nitric oxide (NO) to Ngb-Fe_2+ is relatively low compared to penta-coordinate Hbs and Mbs, which is due to protection of the Fe_2+ by the internal His(E7) ligand. Under an excess of NO, however, Ngb-Fe_2+-NO may readily form and scavenge peroxynitrite, a harmful oxidizing molecule generated under ischemic conditions (Herold et al. 2004).

1.2.8. Function of Neuroglobin:

Putative neuroglobin functions

Until recently, it has generally been accepted that any intracellular globin has a similar function to that of Mb, either storing O_2 for hypoxic phases (e.g. in diving mammals) or facilitating O_2 diffusion from the capillaries to the respiratory chain in the mitochondria. However, within the past ten years it has become evident that certain globins may carry out various other functions. For example, maize globin regenerates NAD+, which is used to promote glycolysis in low oxygen environments **(Sowa et al. 1998).**

Although Ngb is the best-investigated 'novel' globin type, its exact physiological role is still uncertain. Ngb is a highly conserved protein, with an evolutionary rate that is about threefold slower than that of Mb and Hb .Thus Ngb has remained largely unchanged during evolution, pointing to an important role of this protein. Several functions of Ngb have been proposed as:

1. Ngb may exert a Mb-like role, enhancing O_2 supply to the mitochondria of the metabolically active neurons.
2. Ngb may scavenge damaging reactive oxygen or nitrogen species (ROS/RNS), which are generated for example by the respiratory chain.
3. Ngb may detoxify harmful excess of nitric oxide (NO) to nitrate (NO_3–) at normoxia or produce NO for signalling functions from nitrite (NO_2–) at hypoxia for the control of blood pressure.
4. Ngb may be involved in a signal transduction pathway, e.g. by inhibiting the dissociation of GDP from G protein α.
5. Ngb may be part of a redox process that is for example instrumental in preventing apoptosis via reduction of cytochrome c. (T. Burmester and Hankeln 2009).

A role of neuroglobin in O2 supply

A specific function of Ngb in O2 delivery basically requires **(1)** an O_2 affinity that is sufficient to deliver oxygen to the tissue, **(2)** an expression in cells with special needs for O_2, **(3)** an expression level roughly equivalent to that of muscle Mb and **(4)** a system that counterbalances the tendency of Ngb to auto oxidise.

Ngb binds reversibly to O_2 with an affinity that is within the range of a typical Mb [half saturation pressure P50=0.12–0.29kPa (0.9–2.2 Torr)] at 20–37°C . Ngb combines with O_2 at high PO_2 and releases it at low PO_2, fulfilling the basic requirements of a respiratory protein employed in O_2 supply (B. Ebner, Burmester, and Hankeln 2003; Thorsten Burmester et al. 2002)

Hypoxia regulation of Ngb

If Ngb is associated with O_2 consumption, it may be expected that alterations of external O2 supply may also affect Ngb expression. In cell culture systems, Ngb expression may actually be moderately induced by hypoxia (Schmidt-Kastner et al. 2006). In vivo, however, found no difference in Ngb levels in hypoxic and normoxic brains of rodents. various hypoxia regimes never showed increased Ngb mRNA or protein levels in rat brain. Likewise, in situ hybridization, micro-array and quantitative real-time PCR studies showed that ischemia did not induce significant

changes of Ngb. It should be therefore considered unlikely that Ngb expression levels are increased by exposure to hypoxia in rodents (T. Burmester and Hankeln 2009).

1.2.9. Gene expression patterns of Neuroglobin:

Since the initial discovery of Ngb mRNA in the mammalian brain (Thorsten Burmester, Weich, and Reinhardt 2000), several studies have confirmed the widespread expression of both, Ngb mRNA and protein, in nerve cells of the central and peripheral nervous system in mammals (Zhang et al. 2002). Ngb is exclusively ex- pressed in the cytoplasm of neurons, but not in glia, which may be explained by the presence of candidate neuron-restrictive silencer elements in the Ngb gene region . In some studies (Schmidt-Kastner et al. 2006; Herold et al. 2004; Brunori and Vallone 2007), virtually all neurons appear to be Ngb-positive, albeit at regionally substantially different expression intensities. These varying expression levels reconcile these data with other studies that have reported a more focal expression pat- tern of Ngb. Recent mRNA in situ hybridization studies in the zebrafish (*Danio rerio*) revealed a global pattern of Ngb expression in the fish brain (Fuchs et al. 2004). In addition to the central nervous system (CNS) and the peripheral nervous system (PNS), mammalian Ngb is expressed in endocrine tissues such as the adenohypophysis, adrenal gland, testes (Reuss et al. 2002) and the pancreatic islets of Langerhans. Like neurons, these cell-types are known to be metabolically highly active, which appears to be a general feature of Ngb expression sites. While the average Ngb protein content in total mouse brain has been estimated to be rather low, we found that the retina of the mammalian eye is a major site of Ngb expression (Schmidt et al. 2003). Local retinal Ngb concentrations in mouse amount to an estimated 100 lM, which is almost equivalent to Mb content in muscle cells. Ngb is found in the plexiform cell layers and in the inner segments of the photoreceptors, which again coincides with regions of high metabolism and O2 demand. Retinal Ngb expression has recently been confirmed in the zebrafish (Fuchs et al. 2004)

1.2.10. Neuroglobin and Cytoglobin in invertebrate:

There are very few study regarding to the invertebrate. (Bailly and Vinogradov 2008) found that, Bilatarian Sea Urchin and the Radial Starlet Sea Anemone share the almost same coding of the vertebrate animal. But Their function is not so known. A 34 834-bp gene in the genome of the sea urchin Strongylocentrotus purpuratus (Echinodermata) consists of 34 exons and 33 introns, and encodes a 2252aa putative chimeric Hb, comprising an unidentifiable nonglobin N-terminal

domain of ~150aa and 16 globin domains (D1–D16) of ~150aa, except for D2, which lacks helices G and H. The similarity between the globin domains varies from 39 to 51% identity: they are linked tightly to one another, with interdomain segments of 10aa or less. Alignment of the globin domains with other globins shows them to have a D helix and His residues at both the E7 and F8 locations. Intron insertions within the globin domains occur at canonical positions B12.2 and G7.0. Blastp searches with each of the 16 domains showed a strong sequence similarity with vertebrate neuroglobins and globin X, the Cnidarian *Nematostella vectensis*, and with single-domain bacterial globins. A Bayesian analysis of the *S. purpuratus* globins, the 7 single domain globins from the Cnidarian *N. vectensis* and 80 globins from several metazoan groups, indicated that the S. purpuratus and *N. vectensis* globins share a molecular affinity with vertebrate neuroglobins and globin X, while annelid, mollusc, crustacean, lamprey, hagfish and urochordate globins form a clade with vertebrate cytoglobins. The common molecular signatures and gene structures shared between a radial metazoan, *S. purpuratus* and vertebrate globins suggest these proteins could exhibit ancestral metazoan globin properties and predate the Radiata-Bilateria split.

1.2.11. Research gap & Opportunity:

From the above review of the literatures it is evident that there is very few research in the context of globin protein especially Cygb & Ngb on Crustacean. Many Crustacean species are commercially important. As Cherax are commercially important freshwater crayfish, the structure and functions of Cygb and Ngb are essential for future research in the field of gene expression study and comparative molecular analysis. Moreover, it will assist to conduct further molecular research on other commercially important freshwater crustaceans.

1.3. Objectives of the Study:

- To understand the binding site and function of globin protein.
- To compare the Cytoglobin and Neuroglobin of different species.
- To identify and characterize the Cygb and Ngb genes of Crayfish.

Chapter Two

Materials and Method:

2.1. Data Collection

All raw sequence data was obtained from the Sequence Read Archive of NCBI (Ali et al., 2017) and assembled contigs were obtained through personal communication with the author.

2.2. Data assembly and mapping

These reads obtained from the DNA sequencer were assembled with the Trinity short read *de novo* assembler (Version 2014-04-13p1; Haas et al., 2013). Assembled data were used as BLASTx queries against the NCBI protein database using BLAST+ (Version 2.2.29; Camacho et al., 2009). Gene ontology (GO) terms were assigned to annotated transcripts using Blast2Go Pro (Version 3.0; Conesa et al., 2005). Candidate genes were identified based on literature searches, PFAM domains, and GO terms (Eissa and Wang 2016; Kaiser 1975; Ali et al. 2015)

2.3. Identification and annotation of candidate genes

The transcriptome data set was examined to identify transcripts that matched Cygb and Ngb genes and other major genes. The identified gene sequences were aligned with the non-redundant protein database at NCBI (National Centre for Biotechnology Information) in order to compare the similarity to previously identified and annotated genes. Following this validation step, 5 prime and 3 prime untranslated regions (UTRs) as well as open reading frames were determined using ORF Finder interface of NCBI. Open reading frames were translated into proteins. Molecular weight of the translated amino acids were calculated using molecular weight calculator (Manipulation 2000). The putative protein domains, important functional residues and PFAM domains in the predicted amino acid sequences for each candidate gene were determined using SMART database (Simple Modular Architecture Research Tool) (Schultz et al., 1998) and NCBI's Conserved Domain Database (Marchler-Bauer et al., 2010).

2.4. Comparative molecular analysis

The translated amino acid sequences were used as BLASTp queries against the NCBI database. Top BLAST hits with high sequence similarity with the query sequence were downloaded for each gene and aligned with BioEdit software (version 7.2.5; (Hall, 1999)) using a ClustalW alignment platform (Larkin et al., 2007). For phylogenetic analysis of the full-length Cygb and Ngb genes, Neighbour Joining trees were generated based on amino acid data in MEGA (version 6; Tamura et al., 2013) using the Jukes Cantor method with 10000 bootstraps (Kearse et al., 2012). DNA polymorphism at nucleotide and amino acid levels including synonymous and non-synonymous mutations were detected with DnaSP software (Librado and Rozas, 2009).

Chapter Three

Results:

3.1. Identification and Annotation of Cytoglobin and Neuroglobin genes:

Candidate genes were identified based on literature searches and PFAM domains. Multiple units of HSPs were identified and annotated, and the annotated sequence were submitted at NCBI GenBank archive for accession number. Details of the sequences with their accession numbers are listed in Table 1. Cygb and Ngb were unambiguously identified from the gill transcriptome of the three freshwater crayfish *Cherax quadricarinatus*, *C. destructor* and *C. cainii*. Cygb genes obtained from *Cherax quadricarinatus*, *C. destructor* and *C. cainii* were named as CqCygb, CdCygb and CcCygb respectively. The structure of CqCygb (accession AJO70023.1) with their 5′ and 3′-UTR is illustrated in Fig 2. The basic structure of CqCygb, CdCygb and CcCygb were exactly similar with some substitutions at nucleotide level (Table 2 and Appendix). The deduced protein encoded by CqCygb, CdCygb and CcCygb contained 190 amino acids (Table 1) The CdCygb transcript (AJO70023.1) was 1174 bp in length and contained a 332 bp 5′-UTR, 582 bp ORF (including the stop codon), and a 260 bp 3′-UTR (Fig. 4).

The HSP90 transcripts obtained from *Cherax quadricarinatus*, *C. destructor* and *C. cainii* were named as CqNgb, CdNgb and CcNgb respectively. The molecular structure of CdNgb (accession AJO70025.1) with their 5′ and 3′-UTR is illustrated in Fig 5. The structure of CqNgb, CdNgb and CcNgb were almost similar with some changes at nucleotide and amino acid level (Table 2 and Appendix). The length of protein deduced from CqNgb, CdNgb and CcNgb were 209 aa (Table 1). The cytoplasmic Ngb transcript (AJO70025.1) was 1561 bp in length and contained a 326 bp 5′-UTR, 630 bp ORF (including the stop codon), and a 605 bp 3′-UTR (Fig. 5).

```
3     tct tat cat gcc ctc ctc ttg ttt atc ctt aca act gct ggg gcg   47
48    gcg gcg cca tgt cct cat tac cct ctt aag act tag ttg cca ggt   92
93    cac tcg cta tag gga gcc aca gta taa act ggg cgt gag cac ttg  137
138   cct caa gtt agt gat gag tgt gtg gcg gtg aga cgt gga agt agt  182
183   gct agt gag aga ggt gtt act gcg tca gtg ctt cca tag ggt ctt  227
228   cgt tca ttc tct cca tcg cct agc caa gaa cgt gtt ccc tcc acc  272
273   agt taa ctt ccc tag gaa atc ttt cga tac aac aga gtg gtt gag  317
318   GCT GCT GCC AGC ACA ATG GGA GCT GTG TTC AGT GTC TTG TGG AGT  362
                          M   G   A   V   F   S   V   L   W   S
363   TGG GTC TCG CCG AAC TCA AAG GTG GTA CAG GTG AAA GCC ATG ACG  407
11     W   V   S   P   N   S   K   V   V   Q   V   K   A   M   T   25
408   TTC TCA GAC GAG CAG GAC CTG GGA CCT GAG GCA GAC GTG GCC GAC  452
26     F   S   E   E   Q   D   L   G   P   E   A   D   V   A   D   40
453   GCT ACC ACA GGT CTC ACC CTG CGT CAC CGC ACA GCT ATA GTC AGG  497
41     A   T   T   G   L   T   L   R   H   R   T   A   I   V   R   55
498   ACC TGG GAC CTG GTC AGA CCT GAC CTC AAG GAA CAT GGT GTC AAT  542
56     T   W   D   L   V   R   P   D   L   K   E   H   G   V   N   70
543   TTC TTC CTC AGG TTG TTC CAG GAT GCT CCC CTC ATA CAG ACC AGG  587
71     F   F   L   R   L   F   Q   D   A   P   L   I   Q   T   R   85
588   TTC AAG GGC TTC GTT GGC ATG ACG GAG GCT GAA CTC AAG ACC AAC  632
86     F   K   G   F   V   G   M   T   E   A   E   L   K   T   N  100
633   AAG CGC CTT GTC GCC CAC GGC ACC ACG GTA CTG ATG GCA ATA ACG  677
101    K   R   L   V   A   H   G   T   T   V   L   M   A   I   T  115
678   AGC ATG GTG GAC AAC ATA GAT GAT GTG TCT GTG TTA GTG GAG TTG  722
116    S   M   V   D   N   I   D   D   V   S   V   L   V   E   L  130
723   TTG AAG AAC ACT GGT GCC AAC CAC AAC AGT AGA GGC ATA CCC AAG  767
131    L   K   N   T   G   A   N   H   N   S   R   G   I   P   K  145
768   GAC GAC TTT GCG CTG TTG TTC CCG GTG CTC CTG AGC TAC CTG AAG  812
146    D   D   F   A   L   L   F   P   V   L   L   S   Y   L   K  160
813   GAC AAC CTG GGT TCT GCC TGG ACG CCA CTG GCT GAG GAA GGA TGG  857
161    D   N   L   G   S   A   W   T   P   L   A   E   E   G   W  175
858   AAG CAA GCG TGC AAG GTG ATA CAC GCT GTT ATC ATC TCT TCC TAC  902
176    K   Q   A   C   K   V   I   H   A   V   I   I   S   S   Y  190
903   GAC ACC CCC TGA TTC ACC CTT CTC CTG GCT ACA CAT TTC ACG AGT  947
191    D   T   P   *                                              194
948   tta ttg tat cga ata aac aag aaa ttt gtt tgc aag att tat tgg  992
993   tac att ttt tta tat gca ctc acg acc ttg tct tat taa tgg tgt 1037
1038  ctt aag tac gac caa tgt ttt aaa aat tta tgg ctc cag tgt cct 1082
1083  gca att aaa gct aat ttt ata cct tac taa agt aat aaa ctt tag 1127
1128  att att gta cag tat tta aaa aaa aat att gaa tta taa aaa aaa 1172
```

Fig 4: Full length nucleotide sequences along with corresponding deduced amino acid sequences of a CYGBi2 (Cytoglobin, AJO70023) cDNA amplified from the gills Yabby crayfish, C. destructor. Sequences are numbered at both sides of each line. The start codon and stop codon are in bold and shaded. 3′ and 5′ untranslated regions (UTRs) are in small letters and expressed codons (CDS) are in capital letters. One putative AU-rich mRNA instability motif (polyadenylation cleavage signal) (ATTTA) is in bold and shaded.

```
3     gca tcc gtg aac ctc agt ctc gcc tct acc gtg cac gag tgc gtg   47
48    tta ctg ctt cgt gcc tct cac agg cat tag aga aaa agt cat tgt   92
93    gat gtg ctg acg agg tta ctc gtg tgg gaa ttg agc atc gcg gag   137
138   cgg cag atc aat gcc ata tag ctc cgt tgc agc gcg tta tca cta   182
183   taa aat tat taa aga cac tat tgt gtt tta gaa ctc tgt aaa aag   227
228   agc gtc tct gaa aaa ttt att tgg tgt ctt gaa tac aaa gaa aca   272
273   gaa tta aac ttt tta aag tgt taa aca gct taa agc agt taa ata   317
318   taa gat aag ATG GGC TGC CAT TTG ACA AAG GGA AAT AAA AAA AAG   362
                  M   G   C   H   L   T   K   G   N   K   K   K
363   GAA GAG GTT GTA TTG CCA GAG CCT CCT GAA CCA GCG TCA CCA GAC   407
13     E   E   V   V   L   P   E   P   P   E   P   A   S   P   D   27
408   CCT CGT CTG CCT CTC ACT GCA AGA CAA AAG TTC AAC ATC ATC AAA   452
28     P   R   L   P   L   T   A   R   Q   K   F   N   I   I   K   42
453   TCC TGG AAG GGA ATC TCC AGA GCC ATC GAG CCC ACT GGT GTC TAC   497
43     S   W   K   G   I   S   R   A   I   E   P   T   G   V   Y   57
498   ATG TTT GTT AAG TTA TTT GAG AAC AAC GCA GAA TTG ATC AAC TTG   542
58     M   F   V   K   L   F   E   N   N   A   E   L   I   N   L   72
543   TTC ACC AAG TTC AAC CAG CTG CGG ACG AGA GAT GAA CAG GCT GAG   587
73     F   T   K   F   N   Q   L   R   T   R   D   E   Q   A   E   87
588   TCA CTG GAG CTG GCA GAA CAC GCC AAA ACC GTC ATG AAT TCC ATC   632
88     S   L   E   L   A   E   H   A   K   T   V   M   N   S   I   102
633   GAC GAA GGT ATC AAG TCC ATG GAT AAC GTA GAC TAT TTC TTC GAC   677
103    D   E   G   I   K   S   M   D   N   V   D   Y   F   F   D   117
678   ATC CTG CAC CAG ATC GGA GCC TCC CAC CAC AAG ATA CCT GGC TTC   722
118    I   L   H   Q   I   G   A   S   H   H   K   I   P   G   F   132
723   AAG AAG GAG TAC TTT TGG AAG ATC GAG GTT CCC TTC CTG GAA GCT   767
133    K   K   E   Y   F   W   K   I   E   V   P   F   L   E   A   147
768   GTG CGG TTG ACG CTG GAC GAC CGC TAC ACT GAA AAC ATG GAC AAC   812
148    V   R   L   T   L   D   D   R   Y   T   E   N   M   D   N   162
813   ATA TAC AGG ATC ACT ATC AAG TTT CTC CTC GAA ACA GTT GTA GAA   857
163    I   Y   R   I   T   I   K   F   L   L   E   T   V   V   E   177
858   GGT TAC GAG AAG GCA GAA AAA GAC AAA TCC AAC GAA AAT GTG AAG   902
178    G   Y   E   K   A   E   K   D   K   S   N   E   N   V   K   192
903   AAT AAT ATC AAC TCC ACG GAC GGA AAC TGC AGC AAG GGT ATC AAC   947
193    N   N   I   N   S   T   D   G   N   C   S   K   G   I   N   207
948   GGC TCC TGA aga agg cac aaa gaa ttc ctg agg gat gag ttg cgg   992
208    G   S   *                                                  210
993   gaa aaa gct aca ctg cca cta ggg cct aca cca ctt gtt tgt gtt   1037
1038  tgt cca tca cta aga ggt agc tga agt gtg gcc atc gga agc ata   1082
1083  tat gtt ggg gag aca gca tgg cca ttt tgc tcg gat aat gac taa   1127
1128  ctc gct cat ttc tac agg gga taa tac cta att tac tta aca ctt   1172
1173  tgg gaa aca tga cca atc tac tcg tct ctg cca gga ata ata cct   1217
1218  act tca cat aac att gct tgg tta atg aga aat ttg ctc atc tct   1262
1263  gat agg gac ata tta att cgc ctc tac tcc gat aat ggt tta ctt   1307
1308  gct cat tta taa taa ggg taa caa cta agt cac tta ata cta att   1352
1353  ggc taa aga atg agt tgc tga ttt tgg ttg gga tat gac tag act   1397
1398  tta att ctt act atg ttg cac tca tat cta cta ccc tct atg tcg   1442
1443  atc agt aca tgt aaa cta taa agc acc aat tct taa tgg gga aaa   1487
1488  cct att tac gag aga gag caa ctt tct aaa gca aac aaa aca gaa   1532
1533  gta tca cta aga ata cac aca cac aca   1559
```

Fig 5: Full length nucleotide sequences along with corresponding deduced amino acid sequences of a Ngb (Neuroglobin, AJO70025.1) cDNA amplified from the gills Yabby crayfish, *C. destructor*. Sequences are numbered at both sides of each line. The start codon and stop codon are in bold and shaded. 3′ and 5′ untranslated regions (UTRs) are in small letters and expressed codons (CDS) are in capital letters. One putative AU-rich mRNA instability motif (polyadenylation

cleavage signal) (ATTTA) is in bold and shaded.

3.2. Comparative analysis of Cygb and Ngb:

Multiple alignment of the Cherax Cygb with the homologous genes in other crustaceans showed that motifs were highly conserved in all the aligned sequences (Fig. 6). A BLASTp search revealed that the Cygb obtained from our study shared high protein similarity with other Cygbs, including those from *Carcinus maenas* (70% identity), *Penaeus vannamei* (62% identity) and *Hyalella Azteca* (52% identity) (Fig. 6).

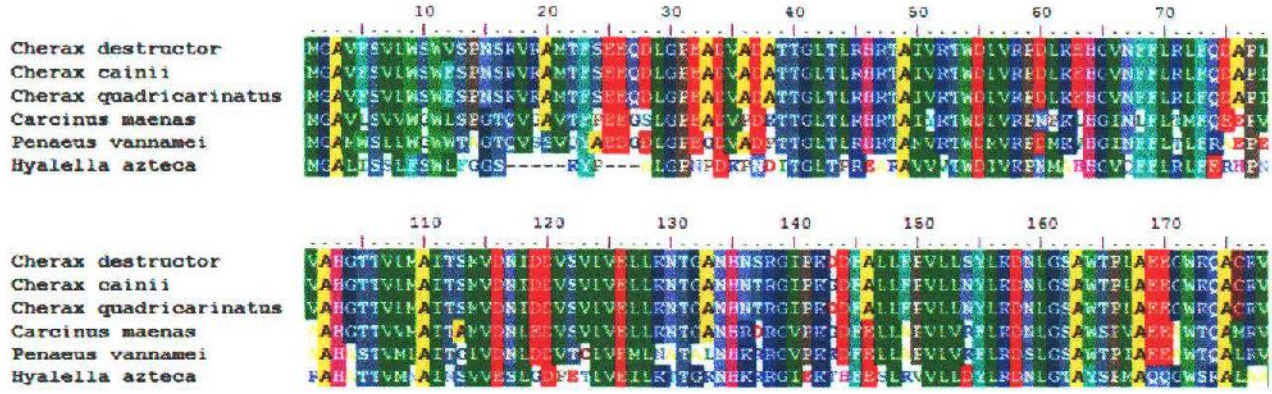

Fig 6: Multiple alignment of translated amino acid sequences of C. destructor (AJO70023.1) and *C. cainii* (AJO70029.1) with Cygb isoforms from some representative crustaceans such as the European green crab *Carcinus maenas* (CBN88274.1), Whiteleg shrimp (ASN74468.1) and Freshwater amphipod *Hyalella azteca* (XP 018008187.1).

BLASTp search determined that the CdNgb was highly conserved in crustaceans and had greatest similarity with the Ngbs isoforms from *Microplitis demolitor* (63% identity), *Penaeus vannamei* (80% identity) (Fig: 7).

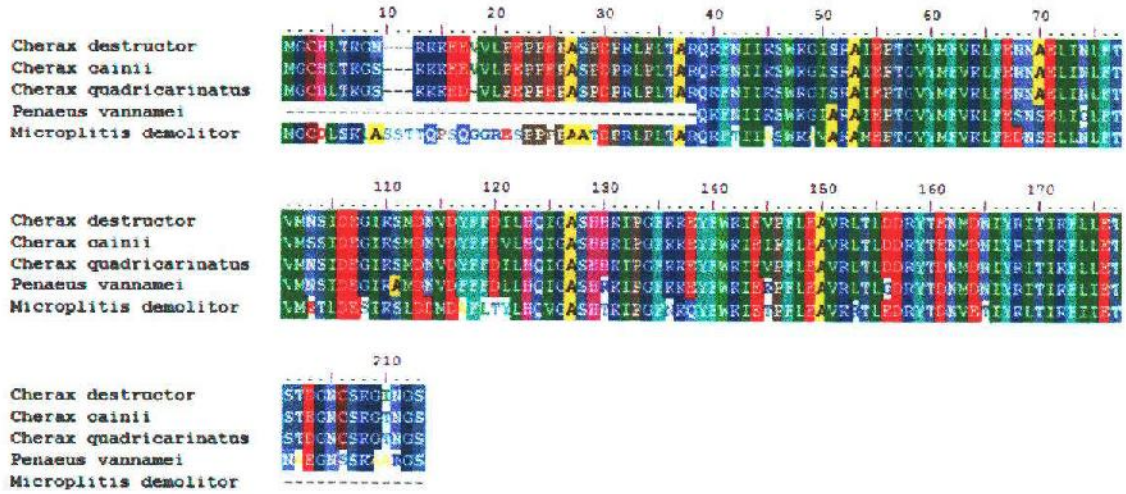

Fig 7: Multiple alignment of translated amino acid sequences of *C. destructor* Ngb (AJO70025) and *C. cainii* (AJO70031) with isoforms from of some representative crustaceans such as the Whiteleg shrimp *Penaeus vannamei* (ASN74469.1), Australian wasp *Microplitis demolitor* (XP_008555957.1).

3.3 Phylogenetic analysis

The phylogenetic tree revealed that Cgbs obtained from several species formed some distinct groups; including Crustacean, Mollusca and Insecta (Fig. 8). All crayfish species, *C. quadricarinatus, C. destructor* and *C. cainii* formed a monophyletic group. CdNgb formed three distinct groups. All crayfish species, *C. quadricarinatus, C. destructor* and *C. cainii* formed a monophyletic group (Fig. 9).

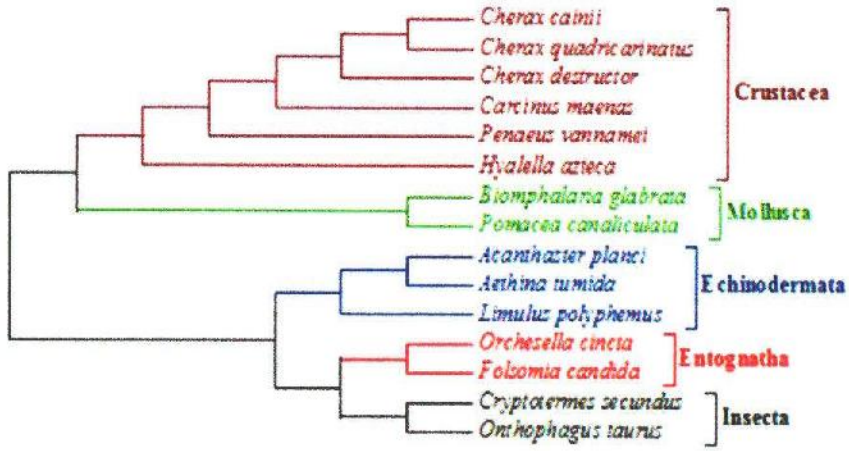

Fig 8 : Phylogenetic relationships between the three Cherax crayfish with other organisms regarding to the Cytoglobin amino acid sequence.

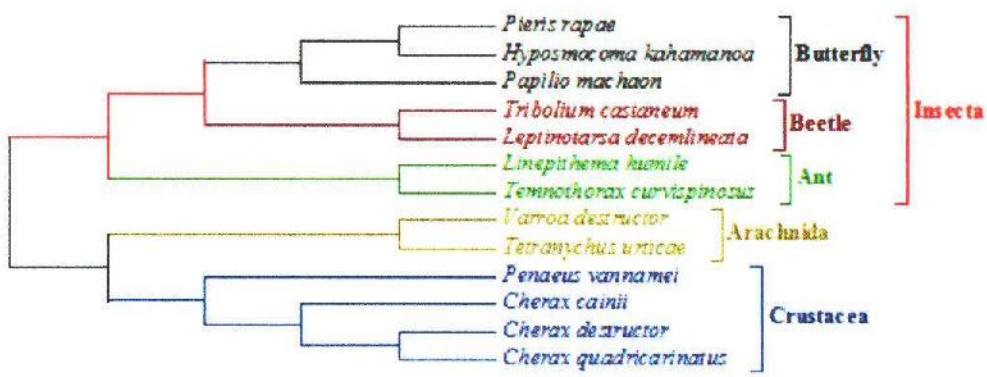

Fig 9: Phylogenetic relationships between the three Cherax crayfish with other organisms regarding to the Neuroglobin amino acid sequence.

Table 1:

List of top candidate genes of Neuroglobin and Cytoglobin identified from the gill transcriptomes of *Cherax cainii, C. destructor* and *C. quadricarinatus.*

Gene	*C. quadricarinatus (Redclaw)*			*C. destructor* (yabby)			*Cherax cainii* (Maroon)		
	Contig Length bp	Protein Length (aa)	Accession #	Contig Length bp	Protein Length (aa)	Access-ion #	Contig Length Bp	Protein length (aa)	Access-ion #
CYgb 2 like isoform 1	1293	190	-	1133	190	AJO70023	1194	190	AJO70029
Ngb Partial	1065	209	-	1561	209	AJO70025	5176	209	AJO70031
Ngb Like	-	-	-	-	-	-	488	45	AJO70033
Globin partial	375	116	-	-	-	-	-	-	-

Table 2. Comparative molecular analyses of cytoglobin and neuroglobin from three crayfish at amino acid/codon level

Features	Cytoglobin	Neuroglobin
Total amino acid/codon in analysis	190	290
Identical sites (monomorphic) in all species	161	240
Total number of mutations (polymorphic sites)	29	50
Number of Nonsynonymous mutation	9	10
Number of Synonymous mutation	20	40
Number of InDel (Insertion-Deletion)	0	1

Chapter Four

Discussion:

Overall this study has demonstrated that most of the candidate genes of Cygb and Ngb are highly conserved in the genus *Cherax* and in other crustacean species in general. While these genes showed little variation across the three target species, no major changes were observed in the functional structure of the genes. The genes encoding Cygb and Ngb were unambiguously identified from three crayfish species and submitted to NCBI GeneBank database. All the submitted sequences contained high similarity to other sequences available in the database based on BLAST searches, which proves that they are correctly annotated. All the annotated sequences were also verified manually by the expert of NCBI as well.

4.1. Cytoglobin:

Cytoglobin from other species have the distinct identical features. Crayfish show same molecular weight of cytoglobin as it is 21.4 kda which is similar of many other species. *C. maenas* cytoglobin is 20.73 kda along with the 189 residues and also 21.22 for P. vannamei along with 190 residues as same as Cherax Crayfish. *Hyalella Azteca* is also similar to the all crustaceans as it is 20.52 kda along with the 184 residues.

. A total of 9 nonsynonymous mutation (amino acid replacement) and 20 synonymous mutations were observed in Cygbs from three crayfish species. There is no insertion or deletion in the sequence of Cygbs.

Cygb belongs of following species have the conserved globin sequence which is identical for globin family. All conserved part starts from 46 amino acid and ends in 153 amino acid without C. destructor and H. Azteca. But all contain 107 amino acid. From the figure (fig: 6) we see conserve globin protein sequence almost same to the all species. Some differences happened in the sequence which is seen in multiple alignment may be due to the insertion or mutation of the gene.

Cytoglobin contain Oxygen transport binding site which is present in all the crustacean species. However the position of binding site is different . But Cherax spp. Contain the same sequence as

well as same structure and same number of amino acid. The binding site is dimer in Cherax spp. However some functional changes occurred due to the change of amino acid sequence as well as structure. Oxygen transport site is conserved for Cherax spp. But a little difference is noticed belongs to the *P. vannamei* and *H. Azteca.*

The *Cherax destructor* contain the phophoglycerate mutase family which is absent to other following species. This family is involving the transfer of phosphor groups between the three carbon atoms of phophoglycerate which is structurally related enzymes that catalyse reactions. However the function of this family is unknown in *Cherax destructor.*

There is no single peptide bond in *Cherax spp* but the crab and shrimp contain single peptide bond.

Some others features are seen in shrimp species which is not found in other respective crustacean. Basic region leucin zipper, H_4, Ribosomal_S15 are the feature of the shrimp species. However There is no prove of their function is found in invertebrate. Those are active in mammals.

Cytoglobin is slightly more conservative than Ngb. Reconstructions of globin phylogeny confirmed that Cygb is distantly related to vertebrate Mbs, (Wittenberg 1992) with which it may have shared a common ancestor before the split of jawless and jawed vertebrates about 450 million years ago (Hankeln et al. 2005). Regarding to the Cygb all classes of organism hold a stable position in evolutionary biology.

4.2. Neuroglobin:

Ngb is a substantially divergent member of the globin family, displaying only 20-25% amino acid sequence identity to Mbs and Hbs. Ngb represents a typical Mb type monomeric globin which can bind O_2 reversibly. Ngb at first found in the nervous system of human but now it is proved that Ngb have the function in invertebrate (B. Ebner, Burmester, and Hankeln 2003). The conserved globin fold of invertebrate show the many similarities. Molecular weight of Ngb in Cherax spp. Is 24 kda along with 209 residues.All other species also almost have same molecular weight from 20-24 kda.

A total 10 nonsynonymous mutation and 40 synonymous mutation were found in Ngbs across the species and only one deletion is found. However all the changes were outside of functional units.

The globin domain of Ngb is conserved for the invertebrate species. All start with the 5'QKFN---

---PFL3`. Globin domain contain 110 amino acid which is the heme binding site for all following species. In Cherax Spp. Domain site starts from 36 amino acid and ends in 145 amino acid. Some differences are shown for other species but the number of amino acid is same and also their function.

Ngb contain metal binding protein site of all following species except *Microplitis demolitor.* However *M. demolitor* contain oxygen transport site which contain 140 amino acid. *Cherax Spp.* bearing metal binding site with the 101 amino acid and amino acid sequence is also differ from other respective species. Ngb of following species have the low complexity region though their amino acid is different except *Cherax spp.*

There is a short conserved region in Ngb which is known as FCH domain *except Peneaus vannamei.* FCH domain contains around 60 amino acid and always N-terminal and is followed by a coiled-coil region. In Cherax spp., there have a DNA primase binding site which is required for functional interaction with DnaB that attracts primase to the replication fork (Letunic, Doerks, and Bork 2015).

Ngb show the very distinct feature towards the evolutionary biology. All of the species from phylogenetic tree are crustacean but regarding to the sequence they differ from one to another. Similarity is present according to the tree. One common ancestor hold the whole subtree of Ngb. Similarity exists around 25% of all of the species (Hankeln et al. 2005). Ngb is representative of an old globin lineage, which already existed before the separation of Protostomia and Deuterostomia more than 600 million years ago. Ngb sequence conservation during mammalian evolution has been especially high, pointing to a strongly selected, important function.

Chapter Five

Conclusion:

Neuroglobin and cytoglobin are mainly stress related protein. In this study we actually only compare the Ngb and Cygb of different species of invertebrate with *Cherax sppp.* These proteins actually conserved protein. We found almost similar sequence of Ngb and Cygb on crustacean though they exist in different corner of the world. Cygb is more conserve than Ngb.

In our study, we clearly identified and characterized all possible units of cytoglobin and neuroglobin gill transcriptome of three freshwater crayfish. The annotated sequence are made publicly available at NCBI database. It will assist in further study in terms of comparative molecular analysis and gene expression study with cytoglobin and neuroglobin.

Chapter Six

References

Ali, Muhammad Yousuf, Ana Pavasovic, Shorash Amin, Peter B. Mather, and Peter J. Prentis. 2015. "Comparative Analysis of Gill Transcriptomes of Two Freshwater Crayfish, Cherax Cainii and C. Destructor." *Marine Genomics* 22: 11–13. https://doi.org/10.1016/j.margen.2015.03.004.

Bailly, Xavier, and Serge N. Vinogradov. 2008. "The Bilatarian Sea Urchin and the Radial Starlet Sea Anemone Globins Share Strong Homologies with Vertebrate Neuroglobins." In *Dioxygen Binding and Sensing Proteins*, 191–201. Milano: Springer Milan. https://doi.org/10.1007/978-88-470-0807-6_16.

Becker, L. 2004. "New Concepts in Reactive Oxygen Species and Cardiovascular Reperfusion Physiology." *Cardiovascular Research* 61 (3): 461–70. https://doi.org/10.1016/j.cardiores.2003.10.025.

Brunori, M., and B. Vallone. 2007. "Neuroglobin, Seven Years After." *Cellular and Molecular Life Sciences* 64 (10): 1259–68. https://doi.org/10.1007/s00018-007-7090-2.

Burmester, T., and T. Hankeln. 2009. "What Is the Function of Neuroglobin?" *Journal of Experimental Biology* 212 (10): 1423–28. https://doi.org/10.1242/jeb.000729.

Burmester, Thorsten, Bettina Ebner, Bettina Weich, and Thomas Hankeln. 2002. "Cytoglobin: A Novel Globin Type Ubiquitously Expressed in Vertebrate Tissues." *Molecular Biology and Evolution* 19 (4): 416–21. https://doi.org/10.1093/oxfordjournals.molbev.a004096.

Burmester, Thorsten, Bettina Weich, and Sigrid Reinhardt. 2000. "A Vertebrate Globin Expressed in the Brain" 407 (September): 1998–2001.

D'Ancona, Piero, and Luca Fanelli. 2009. "Smoothing Estimates for the Schrödinger Equation with Unbounded Potentials." *Journal of Differential Equations* 246 (12): 4552–67. https://doi.org/10.1016/j.jde.2009.03.026.

Dewilde, Sylvia, Bettina Ebner, Evi Vinck, Kambiz Gilany, Thomas Hankeln, Thorsten Burmester, Jill Kreiling, et al. 2006. "The Nerve Hemoglobin of the Bivalve Mollusc

Spisula Solidissima: Molecular Cloning, Ligand Binding Studies, and Phylogenetic Analysis." *Journal of Biological Chemistry* 281 (9): 5364–72. https://doi.org/10.1074/jbc.M509486200.

Dröge, Jasmin, Amit Pande, Ella W. Englander, and Wojciech Makałowski. 2012. "Comparative Genomics of Neuroglobin Reveals Its Early Origins." *PLoS ONE* 7 (10). https://doi.org/10.1371/journal.pone.0047972.

Ebner, B., Thorsten Burmester, and Thomas Hankeln. 2003. "Globin Genes Are Present in Ciona Intestinalis." *Molecular Biology and Evolution* 20 (9): 1521–25. https://doi.org/10.1093/molbev/msg164.

Ebner, Bettina, Georgia Panopoulou, Serge N. Vinogradov, Laurent Kiger, Michael C. Marden, Thorsten Burmester, and Thomas Hankeln. 2010. "The Globin Gene Family of the Cephalochordate Amphioxus: Implications for Chordate Globin Evolution." *BMC Evolutionary Biology* 10 (1). https://doi.org/10.1186/1471-2148-10-370.

Eissa, Nour, and Han Ping Wang. 2016. "Transcriptional Stress Responses to Environmental and Husbandry Stressors in Aquaculture Species." *Reviews in Aquaculture* 8 (1): 61–88. https://doi.org/10.1111/raq.12081.

Eiter, Thomas, Wolfgang Faber, Michael Fink, Gerald Pfeifer, and Stefan Woltran. 2003. "Complexity of Answer Set Checking and Bounded Predicate Arities for Non-Ground Answer Set Programming." *CEUR Workshop Proceedings* 78 (12): 69–83. https://doi.org/10.1093/bioinformatics/bts199.

Flogel, U., M. W. Merx, A. Godecke, U. K. M. Decking, and J. Schrader. 2001. "Myoglobin: A Scavenger of Bioactive NO." *Proceedings of the National Academy of Sciences* 98 (2): 735–40. https://doi.org/10.1073/pnas.98.2.735.

Fordel, E., E. Geuens, S. Dewilde, P. Rottiers, P. Carmeliet, J. Grooten, and L. Moens. 2004. "Cytoglobin Expression Is Upregulated in All Tissues upon Hypoxia: An in Vitro and in Vivo Study by Quantitative Real-Time PCR." *Biochemical and Biophysical Research Communications* 319 (2): 342–48. https://doi.org/10.1016/j.bbrc.2004.05.010.

Frazer, Kelly A, Laura Elnitski, Deanna M Church, Inna Dubchak, and Ross C Hardison. 2003.

"Cross-Species Sequence Comparisons: A Review of Methods and Available Resources." *Genome Research* 13 (1): 1–12. https://doi.org/10.1101/gr.222003.

Fuchs, Christine, Valeska Heib, Laurent Kiger, Mark Haberkamp, Anja Roesner, Marc Schmidt, Djemel Hamdane, Michael C. Marden, Thomas Hankeln, and Thorsten Burmester. 2004. "Zebrafish Reveals Different and Conserved Features of Vertebrate Neuroglobin Gene Structure, Expression Pattern, and Ligand Binding." *Journal of Biological Chemistry* 279 (23): 24116–22. https://doi.org/10.1074/jbc.M402011200.

Gould, S J. 1979. "The Royal Society Is Collaborating with JSTOR to Digitize, Preserve, and Extend Access to Proceedings of the Royal Society of London. Series B, Biological Sciences. {®} Www.Jstor.Org." *Society* 205 (1161): 211–30. https://doi.org/10.1098/rstb.1983.0080.

Hankeln, Thomas, Bettina Ebner, Christine Fuchs, Frank Gerlach, Mark Haberkamp, Tilmann L. Laufs, Anja Roesner, et al. 2005. "Neuroglobin and Cytoglobin in Search of Their Role in the Vertebrate Globin Family." *Journal of Inorganic Biochemistry* 99 (1): 110–19. https://doi.org/10.1016/j.jinorgbio.2004.11.009.

Heaslet, Holly A., and William E. Royer. 1999. "The 2.7 Å Crystal Structure of Deoxygenated Hemoglobin from the Sea Lamprey (Petromyzon Marinus): Structural Basis for a Lowered Oxygen Affinity and Bohr Effect." *Structure* 7 (5): 517–26. https://doi.org/10.1016/S0969-2126(99)80068-9.

Herold, Susanna, Angela Fago, Roy E. Weber, Sylvia Dewilde, and Luc Moens. 2004. "Reactivity Studies of the Fe(III) and Fe(II)NO Forms of Human Neuroglobin Reveal a Potential Role against Oxidative Stress." *Journal of Biological Chemistry* 279 (22): 22841–47. https://doi.org/10.1074/jbc.M313732200.

Hitomi Sawai, ‡,§, ∥ Norifumi Kawada, ⊥ Katsutoshi Yoshizato, @ Hiroshi Nakajima, @ and Shigetoshi Aono, and § Yoshitsugu Shiro*. 2003. "Characterization of the Heme Environmental Structure of Cytoglobin, a Fourth Globin in Humans†." https://doi.org/10.1021/BI027067E.

Hundahl, Christian Ansgar, Gregg C. Allen, Jens Randel Nyengaard, Sylvia Dewilde, Bruce

Douglas Carter, Jesper Kelsen, and Anders Hay-Schmidt. 2008. "Neuroglobin in the Rat Brain: Localization." *Neuroendocrinology* 88 (3): 173–82. https://doi.org/10.1159/000129698.

Kaiser, W. 1975. "Zur Entwicklung Der Anatomischen Und Physiologischen Forschung in Der Dermatologie Des Fruhen 19. Jahrhunderts." *Dermatologische Monatsschrift* 161 (1): 1–10. https://doi.org/10.3389/fphys.2012.00431.

Kawada, N, D B Kristensen, K Asahina, K Nakatani, Y Minamiyama, S Seki, and K Yoshizato. 2001. "Characterization of a Stellate Cell Activation-Associated Protein (STAP) with Peroxidase Activity Found in Rat Hepatic Stellate Cells." *The Journal of Biological Chemistry* 276 (27): 25318–23. https://doi.org/10.1074/jbc.M102630200.

Kawada, Norifumi, Dan Bach Kristensen, Kinji Asahina, Kazuki Nakatani, Yukiko Minamiyama, Shuichi Seki, and Katsutoshi Yoshizato. 2001. "Characterization of a Stellate Cell Activation-Associated Protein (STAP) with Peroxidase Activity Found in Rat Hepatic Stellate Cells." *Journal of Biological Chemistry* 276 (27): 25318–23. https://doi.org/10.1074/jbc.M102630200.

LaCount, M W, E Zhang, Y P Chen, K Han, M M Whitton, D E Lincoln, S A Woodin, and L Lebioda. 2000. "The Crystal Structure and Amino Acid Sequence of Dehaloperoxidase from Amphitrite Ornata Indicate Common Ancestry with Globins." *The Journal of Biological Chemistry* 275 (25): 18712–16. https://doi.org/10.1074/jbc.M001194200.

Lebioda, Lukasz, Michael W. LaCount, Erli Zhang, Yung Pin Chen, Kaiping Han, Margaret M. Whitton, David E. Lincoln, and Sarah A. Woodin. 1999. "Protein Structure: An Enzymatic Globin from a Marine Worm." *Nature* 401 (6752): 445–445. https://doi.org/10.1038/46728.

Letunic, Ivica, Tobias Doerks, and Peer Bork. 2015. "SMART: Recent Updates, New Developments and Status in 2015." *Nucleic Acids Research* 43 (D1): D257–60. https://doi.org/10.1093/nar/gku949.

Manipulation, The Sequence. 2000. "Internet On-Ramp Internet On-Ramp." *Biotechniques* 28 (6). https://doi.org/10.2144/00286ir01.

Nakatani, Kazuki, Hiroaki Okuyama, Yasuyuki Shimahara, Shigeru Saeki, Dong-Ho Kim, Yuji

Nakajima, Shuichi Seki, Norifumi Kawada, and Katsutoshi Yoshizato. 2004. "Cytoglobin/STAP, Its Unique Localization in Splanchnic Fibroblast-like Cells and Function in Organ Fibrogenesis." *Laboratory Investigation* 84 (1): 91–101. https://doi.org/10.1038/labinvest.3700013.

Pesce, Alessandra, Martino Bolognesi, Alessio Bocedi, Paolo Ascenzi, Sylvia Dewilde, Luc Moens, Thomas Hankeln, and Thorsten Burmester. 2002. "3-Embor010" 3 (12): 1–6. papers2://publication/uuid/551CE8BB-2322-44F4-A27F-6DDEFDDDD0DB.

Pesce, Alessandra, Sylvia Dewilde, Marco Nardini, Luc Moens, Paolo Ascenzi, Thomas Hankeln, Thorsten Burmester, and Martino Bolognesi. 2003. "Human Brain Neuroglobin Structure Reveals a Distinct Mode of Controlling Oxygen Affinity." *Structure* 11 (9): 1087–95. https://doi.org/10.1016/S0969-2126(03)00166-7.

Reuss, S, S Saaler-Reinhardt, B Weich, S Wystub, M.H Reuss, T Burmester, and T Hankeln. 2002. "Expression Analysis of Neuroglobin MRNA in Rodent Tissues." *Neuroscience* 115 (3): 645–56. https://doi.org/10.1016/S0306-4522(02)00536-5.

Schmidt-Kastner, Rainald, Mark Haberkamp, Christoph Schmitz, Thomas Hankeln, and Thorsten Burmester. 2006. "Neuroglobin MRNA Expression after Transient Global Brain Ischèmia and Prolonged Hypoxia in Cell Culture." *Brain Research* 1103 (1): 173–80. https://doi.org/10.1016/J.BRAINRES.2006.05.047.

Schmidt, Marc, Frank Gerlach, Aaron Avivi, Tilmann Laufs, Sylvia Wystub, Jeremy C Simpson, Eviatar Nevo, et al. 2004. "Cytoglobin Is a Respiratory Protein in Connective Tissue and Neurons, Which Is Up-Regulated by Hypoxia." *Journal of Biological Chemistry* 279 (9): 8063–69. https://doi.org/10.1074/jbc.M310540200.

Schmidt, Marc, Andreas Giessl, Tilmann Laufs, Thomas Hankeln, Uwe Wolfrum, and Thorsten Burmester. 2003. "How Does the Eye Breathe? Evidence for Neuroglobin-Mediated Oxygen Supply in the Mammalian Retina." *The Journal of Biological Chemistry* 278 (3): 1932–35. https://doi.org/10.1074/jbc.M209909200.

Sowa, Aleksander W, Stephen M G Duff, Phillip A Guy, Robert D Hill, and Bob B Buchanan. 1998. "Altering Hemoglobin Levels Changes Energy Status in Maize Cells under Hypoxia."

Proceedings of the National Academy of Sciences of the United States of America 95 (17): 10317–21. https://doi.org/10.1073/pnas.95.17.10317.

Szymanski, Megan, Ruihua Wang, M. Danielle Fallin, Susan S. Bassett, and Dimitrios Avramopoulos. 2010. "Neuroglobin and Alzheimer's Dementia: Genetic Association and Gene Expression Changes." *Neurobiology of Aging* 31 (11): 1835–42. https://doi.org/10.1016/j.neurobiolaging.2008.10.003.

Trent, James T., and Mark S. Hargrove. 2002a. "A Ubiquitously Expressed Human Hexacoordinate Hemoglobin." *Journal of Biological Chemistry* 277 (22): 19538–45. https://doi.org/10.1074/jbc.M201934200.

Trent, James T, and Mark S Hargrove. 2002b. "A Ubiquitously Expressed Human Hexacoordinate Hemoglobin." *The Journal of Biological Chemistry* 277 (22): 19538–45. https://doi.org/10.1074/jbc.M201934200.

Uno, Tadayuki, Daisuke Ryu, Hiroko Tsutsumi, Yoshikazu Tomisugi, Yoshinobu Ishikawa, Anthony J. Wilkinson, Hideaki Sato, and Takashi Hayashi. 2004. "Residues in the Distal Heme Pocket of Neuroglobin: Implications for the Multiple Ligand Binding Steps." *Journal of Biological Chemistry* 279 (7): 5886–93. https://doi.org/10.1074/jbc.M311748200.

Vinogradov, S. N., D. Hoogewijs, X. Bailly, R. Arredondo-Peter, M. Guertin, J. Gough, S. Dewilde, L. Moens, and J. R. Vanfleteren. 2005. "Three Globin Lineages Belonging to Two Structural Classes in Genomes from the Three Kingdoms of Life." *Proceedings of the National Academy of Sciences* 102 (32): 11385–89. https://doi.org/10.1073/pnas.0502103102.

Vinogradov, Serge N., David Hoogewijs, Xavier Bailly, Raúl Arredondo-Peter, Julian Gough, Sylvia Dewilde, Luc Moens, and Jacques R. Vanfleteren. 2006. "A Phylogenomic Profile of Globins." *BMC Evolutionary Biology* 6: 1–17. https://doi.org/10.1186/1471-2148-6-31.

Vinogradov, Serge N., and Luc Moens. 2008. "Diversity of Globin Function: Enzymatic, Transport, Storage, and Sensing." *Journal of Biological Chemistry* 283 (14): 8773–77. https://doi.org/10.1074/jbc.R700029200.

Wittenberg, J B. 1992. "Functions of Cytoplasmic Hemoglobins and Myohemerythrin." In *Blood*

and Tissue Oxygen Carriers, edited by Charlotte P Mangum, 59–85. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-76418-9_3.

Wystub, Sylvia, Tilmann Laufs, Marc Schmidt, Thorsten Burmester, Ulrike Maas, Sigrid Saaler-Reinhardt, Thomas Hankeln, and Stefan Reuss. 2003. "Localization of Neuroglobin Protein in the Mouse Brain." *Neuroscience Letters* 346 (1–2): 114–16. https://doi.org/10.1016/S0304-3940(03)00563-9.

Zhang, Chenggang, Chunli Wang, Lin Li, Wenlin Xu, Fanwei Meng, Fuchu He, Meiyu Deng, et al. 2002. "Full-Length CDNA Cloning of Human Neuroglobin and Tissue Expression of Rat Neuroglobin." *Biochemical and Biophysical Research Communications* 290 (5): 1411–19. https://doi.org/10.1006/bbrc.2002.6360.

Chapter Seven

Appendix

Gene	*Cherax quadricarinatus*						*Cherax cainii*					
	Contig length bp	Protein length aa	ORF bp	5′ UTR	3′ UTR	Accession #	Contig length bp	Protein length aa	ORF bp	5′ UTR	3′ UTR	Accession #
Cygb	1293	190	503-1075	218	502	-	1214	190	376-948	266	375	AJO70029
Ngb	1065	209	17-644	421	16	-	1607	209	369-998	609	368	AJO70031