



UNIVERSITY OF ESSEX  
SCHOOL OF LIFE SCIENCES

A thesis for the degree of Master of Science (by Dissertation)

**Molecular assembly and function of Androglobin, a novel  
chimeric globin.**

By Sareeta Bagri

Supervisors: Dr Brandon Reeder and Professor Christopher Reynolds

Date: November 2020

## Abstract

Androglobin (Adgb) is a multi-domain protein that contains a calpain-like and a circularly permuted globin-like domain. Adgb is the fifth mammalian globin to be discovered and is predominately expressed in the testes. Whilst Adgb is probably the least studied and understood globin member this project aims to provide further insight into the haem domain and overcome issues associated with generating the full-length protein by recombinant techniques. Comparing 362 Adgb sequences using multiple sequence alignments, the critical amino acids of the haem domain were identified as the proximal His F8, distal Gln E7, and CD1 Tyr residues. Consistent with the literature, Adgb does not have a close phylogenetic relationship with the other mammalian globin members. Despite being circularly permuted, the 3D model of the haem domain was capable of producing a globin-like fold, even though the haem domain does not resemble the customary secondary structure order, as observed in the other globin members. The 3D haem domain model suggests a pentacoordinate haem geometry. The purified haem domain protein partially supports this, revealing mixed penta/hexa-coordinated properties. Using 3D modelling it was predicted that Adgb contains two surface-exposed Cys in close proximity to form a disulfide bond in the CD region (in a similar position to that found in Ngb), which was confirmed by dithiodipyridine experiment. Adgb binds to NO as a pentacoordinate haem iron species, which is highly unusual for globins and similar to that observed for NO binding to cytochrome *c'*. Adgb exhibits a higher NiR activity than other globin members although, similar to that of Globin X. Therefore, it can be proposed that Adgb may have a specialised function in nitrite reduction that could be related to spermatogenesis. The ligation of the full Adgb gene into the pET28a vector and truncation of the full gene was accomplished by molecular cloning techniques something that has not been previously achieved however, challenges were still encountered during protein expression. Overall this study provides a deeper insight into the structure and biochemical properties of the haem domain of Adgb, an unusual globin.

## **Acknowledgements**

I would like to thank my primary supervisor Dr Brandon Reeder for his continuous support, guidance, and feedback throughout my Master's project and thesis write-up. I would like to say thank you to my secondary supervisor Professor Christopher Reynolds for his advice during board meetings and progress reports which helped me to deepen my understanding in my research topic. In addition, I would like to show my gratitude to Dr Elizabeth Welborn, who has been a very supportive, patient, and a reliable technician. She has helped me on endless occasions and has taught me laboratory skills that I can apply in the future therefore, I am very appreciative of all her help. Lastly, I would like to thank my friends and family for their unconditional love and encouragement throughout this year.

## **Declaration**

I declare that the work throughout this project and within this thesis is my own work unless stated otherwise in this document.

## List of Abbreviations

AAs - Amino acids

AD - Alzheimer's disease

Adgb - Androglobin

ALA - 5-aminolaevulinic acid

APS - Ammonium persulfate

Asn - Asparagine

Ca - Calcium

cGMP - Cyclic guanosine monophosphate

CO - Carbon monoxide

Cygb - Cytoglobin

Cys - cysteine

DMSO - Dimethyl sulfoxide

DNA - Deoxyribonucleic acid.

DTT - Dithiothreitol-Clelands reagent

ER - Endoplasmic reticulum

FHbs - Flavohaemoglobins

GbE - Globin E

GbX - Globin X

GbY - Globin Y

GCSs - Globin coupled sensors

Gln - Glutamine

Hb - Haemoglobin

Hb $\alpha$  - Haemoglobin alpha subunit

HbF - Foetal Hb

Hb $\beta$  - Haemoglobin beta subunit

HCl - Hydrochloric acid

HD - Haem domain

His - Histidine

HPLC - High performance liquid chromatography

I-TASSER - Iterative Threading ASSEmbly Refinement

iPSCs - Induced pluripotent stem cells

IPTG - Isopropyl beta-D-1 thiogalactopyranoside

LB - Luria-Bertani

Leu - Leucine

Loss of heterozygosity - LOH

Mb - Myoglobin

MSC - Microbiology safety cabinet

MWC model - Monod-Wyman Changeux

N<sub>2</sub>O - Nitrous oxide

NaOH - Sodium hydroxide

Ngb - Neuroglobin

Nitrite - NO<sub>2</sub>

Nitrite reductase - NiR

NLS - Nuclear localisation signal

NO - Nitric oxide

NOD- Nitrite Dioxygenase

O/N - Overnight

ORF - Open reading frame

PCR - Polymerase chain reaction

Pfu - *Pyrococcus furiosus*

Phe - Phenylalanine

Phyre2 - Protein Homology/AnalogY Recognition Engine

ROS - Reactive oxygen species

SDS - Sodium dodecyl sulfate

Ser - Serine

sGC - Soluble guanylate cyclase

TCEP - Tris(2- carboxyethyl)phosphine

TEMED - N,N,N',N'-Tetramethyl ethylenediamine

Tyr - Tyrosine

Val - Valine

WT - Wild type

## Contents

|                                                                              |           |
|------------------------------------------------------------------------------|-----------|
| <b>List of Tables.....</b>                                                   | <b>11</b> |
| <b>List of Figures.....</b>                                                  | <b>12</b> |
| <b>Chapter 1.Introduction.....</b>                                           | <b>14</b> |
| 1.1 Globins.....                                                             | 14        |
| 1.1.1 The globin fold.....                                                   | 16        |
| 1.1.2 The haem Pocket and co-ordination of globins.....                      | 17        |
| 1.2 Cellular role and reactions of globins.....                              | 20        |
| 1.2.1 Storage and Transport.....                                             | 20        |
| 1.3 Androglobin.....                                                         | 21        |
| 1.3.1 The Structure and Evolution of Androglobin.....                        | 21        |
| 1.3.2 The haem domain of Androglobin.....                                    | 22        |
| 1.3.3 The calpain domain.....                                                | 25        |
| 1.3.4 Other domains of Adgb.....                                             | 25        |
| 1.4 Allosteric regulation of globins.....                                    | 27        |
| 1.5 The role of globins in Nitric oxide metabolism.....                      | 30        |
| 1.6 Thiol groups in globins.....                                             | 32        |
| 1.7 Biomedical implications of globins.....                                  | 35        |
| 1.8 Functions of Androglobin.....                                            | 38        |
| 1.8.1 Androglobin's role in Spermatogenesis.....                             | 38        |
| 1.9 Aims and Objectives.....                                                 | 41        |
| <b>Chapter 2. Materials and Methods.....</b>                                 | <b>43</b> |
| 2.1 Materials.....                                                           | 43        |
| 2.2 Molecular cloning.....                                                   | 43        |
| 2.2.1 Primer Design.....                                                     | 43        |
| 2.2.2 Polymerase chain reaction.....                                         | 45        |
| 2.2.3 Transformation of plasmid DNA into XL1B or BL21 DE3 E. coli cells..... | 48        |

|                                                                                                                                                      |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 2.2.4 Inoculation of liquid culture.....                                                                                                             | 48        |
| 2.2.5 DNA miniprep.....                                                                                                                              | 49        |
| 2.2.6 Determination of DNA concentration.....                                                                                                        | 49        |
| 2.2.7 Restriction enzyme digests of DNA.....                                                                                                         | 49        |
| 2.2.8 DNA ligations.....                                                                                                                             | 49        |
| 2.2.9 Agarose gel electrophoresis.....                                                                                                               | 52        |
| 2.2.10 Gel extraction.....                                                                                                                           | 52        |
| 2.2.11 Sanger sequencing.....                                                                                                                        | 52        |
| 2.3 Protein synthesis.....                                                                                                                           | 53        |
| 2.3.1 Androglobin haem domain protein expression and purification.....                                                                               | 53        |
| 2.3.2 SDS PAGE analysis of Androglobin protein.....                                                                                                  | 54        |
| 2.4 Determination of Androglobin haem domain disulfide bonds.....                                                                                    | 56        |
| 2.5 Nitrite reductase and Nitric oxide binding analysis.....                                                                                         | 56        |
| 2.6 Sequence Homology Analysis.....                                                                                                                  | 57        |
| 2.7 Molecular modelling.....                                                                                                                         | 57        |
| <b>Chapter 3. Bioinformatic analysis and 3D modelling of Androglobin haem domain.....</b>                                                            | <b>59</b> |
| 3.1 Sequence homology analysis of key residues in the haem domain of Androglobin.....                                                                | 59        |
| 3.2 Phylogenetic analysis of Androglobin with other globin members.....                                                                              | 70        |
| 3.3 A 3D model prediction of Androglobin haem domain.....                                                                                            | 72        |
| <b>Chapter 4. Structural and biochemical characterisation of Androglobin haem domain...80</b>                                                        |           |
| 4.1 Optical characteristics of Androglobin haem domain and distal mutants.....                                                                       | 80        |
| 4.2 Determination of the cysteine state in the haem domain of Androglobin.....                                                                       | 86        |
| 4.3 Nitric oxide binding analysis of Androglobin haem domain and distal mutants.....                                                                 | 92        |
| 4.4 Nitrite Reductase Activity of the haem domain of Androglobin and distal mutants.....                                                             | 95        |
| 4.5 Cloning the full Androglobin gene from a eukaryotic expression system vector (pcDNA 3.1 +) into an <i>E.coli</i> expression vector (pET28a)..... | 104       |
| 4.6 Truncation of the pET28a-full Adgb Plasmid.....                                                                                                  | 113       |

|                                                                                                  |                |
|--------------------------------------------------------------------------------------------------|----------------|
| <b>Chapter 5. Discussion.....</b>                                                                | <b>117</b>     |
| 5.1 Sequence homology of amino acids in Androglobin haem domain.....                             | 119            |
| 5.2 Phylogenetic distribution of mammalian globins.....                                          | 122            |
| 5.3 Molecular modelling of Androglobin haem domain.....                                          | 123            |
| 5.4 Absorbance properties of Androglobin and distal mutants.....                                 | 125            |
| 5.5 Measurement of surface-exposed Cysteines in Androglobin haem domain.....                     | 126            |
| 5.6 NO binding and nitrite reductase activity of Androglobin haem domain and distal mutants..... | 127            |
| 5.7 Expression and truncation of the full Androglobin protein in the pET28a vector.....          | 133            |
| <br><b>Chapter 5. Conclusion.....</b>                                                            | <br><b>136</b> |

## List of Tables

| Chapter 2                                                                                                                                      |    |
|------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Table 2.2.1.1. Properties of forward (F) and reverse (R) primers for the truncation of the full Androglobin gene sequence by PCR.              | 44 |
| Table 2.2.2.1. The PCR conditions for each stage of the PCR reaction 1-4.                                                                      | 46 |
| Table 2.2.2.2. The volumes, concentrations and, reagents for the restriction enzyme digest of the 50 $\mu$ l PCR reaction mixtures using DpnI. | 47 |
| Table 2.2.8.1. Volume and Concentrations of the reagents included in the ligation reactions.                                                   | 51 |
| Table 2.3.2.1. Reagents used in the preparation of gel and buffers for SDS-PAGE gel electrophoresis.                                           | 55 |
| Table 2.7.1. Amino acid sequences for the globin domain of Androglobin for 3D modelling.                                                       | 58 |

## List of Figures

| Chapter 1                                                                                                                                                                    |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1.1.1. A Venn diagram portraying the three globin lineages Fhbs, GCss and, 2/2 Hbs (truncated haemoglobins) in the three kingdoms of life.                            | 15 |
| Figure 1.1.2.1. The structure haem group and the haem pocket.                                                                                                                | 18 |
| Figure 1.3.2.1. A schematic representation of the domain structure of human Adgb (A).                                                                                        | 24 |
| Figure 1.4.1. Oxygen saturation curves of haemoglobin and myoglobin.                                                                                                         | 29 |
| Figure 1.6.1. A 3D diagram of the crystallographic structure for human Ngb.                                                                                                  | 34 |
| Figure 1.7.1. Adgb knockdown inhibits cell proliferation of glioma cell lines U251 and U87.                                                                                  | 37 |
| Figure 1.8.1.1. A schematic overview of major events in spermatogenesis.                                                                                                     | 42 |
| Chapter 3                                                                                                                                                                    |    |
| Figure 3.1.1. Multiple sequence alignment of the Androglobin haem domain.                                                                                                    | 60 |
| Figure 3.1.2. Single data series representing the sequence homology of the three key amino acids in Adgb.                                                                    | 62 |
| Figure 3.1.3. Amino acid charts for the key amino acids in the haem domain of Androglobin.                                                                                   | 63 |
| Figure 3.1.4. Sequence homology analysis of Tyr975 in Androglobin across 362 species.                                                                                        | 66 |
| Figure 3.1.5. Percentage frequency of animals containing a Phe instead of Tyr at position 975 in the Essex model and Cys instead of Phe at position 770 in the Zurich model. | 68 |
| Figure 3.2.1. Cladogram for the evolution of human globins.                                                                                                                  | 71 |
| Figure 3.3.1. 3D model of the Androglobin-haem domain, including the IQ-binding motif.                                                                                       | 73 |
| Figure 3.3.2. A 3D model of Androglobin haem pocket and the potential disulfide bond formed in the CD region of the Essex model.                                             | 76 |
| Figure 3.3.3. 3D model of the Essex and Zurich models compared to Neuroglobin.                                                                                               | 78 |
| Chapter 4                                                                                                                                                                    |    |
| Figure 4.1.1. UV-Vis spectra of Androglobin-haem domain (wildtype).                                                                                                          | 81 |
| Figure 4.1.2. UV-Vis spectra of Q791H Adgb distal mutant.                                                                                                                    | 82 |

|                                                                                                                                                                                                                                        |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 4.1.3. UV-Vis spectra of Q791G Adgb distal mutant.                                                                                                                                                                              | 83  |
| Figure 4.2.1. The determination of the number of cysteines exposed on the surface of the haem domain in androglobin.                                                                                                                   | 87  |
| Figure 4.2.2. The determination of the number of cysteines exposed on the surface of the Haem domain in androglobin.                                                                                                                   | 89  |
| Figure 4.3.1. The absorbance spectra for ferrous state of Adgb-HD (WT), Q791H Adgb, and Q791G Adgb when NO is bound.                                                                                                                   | 93  |
| Figure 4.4.1. Optical changes resulting from NiR activity in the deoxyferrous-haem domain of Adgb-HD and distal mutants.                                                                                                               | 97  |
| Figure 4.4.2. The time course for nitrite reductase activity for Adgb-HD (blue), Q791H Adgb (orange) and, Q791G Adgb (grey) distal mutants.                                                                                            | 99  |
| Figure 4.4.3. The pseudo-first-order rate for Adgb-HD (blue), Q791H Adgb (orange) and, Q791G Adgb (grey) NiR activity and for the formation of ferrous-NO.                                                                             | 101 |
| Figure 4.5.1. Schematic representation of the ligation of the pET28a vector and the full Adgb gene insert.                                                                                                                             | 105 |
| Figure 4.5.2. Agarose gel electrophoresis for the pET28a vector digested with BamHI and NotI restriction enzymes.                                                                                                                      | 106 |
| Figure 4.5.3. Agarose gel electrophoresis for the pcDNA 3.1 His-tagged full Adgb plasmid digested with BamHI, NotI, and AseI restriction enzymes.                                                                                      | 108 |
| Figure 4.5.4. Agarose gel electrophoresis for the pET28a-full Adgb recombinant plasmid digested with BamHI.                                                                                                                            | 110 |
| Figure 4.5.5. SDS-PAGE gel electrophoresis for the expression analysis of pET28a-full Androglobin protein                                                                                                                              | 112 |
| Figure 4.6.1. Agarose gel electrophoresis for the truncation of the pET28a-full Adgb by PCR.                                                                                                                                           | 114 |
| Figure 4.6.2 SDS-PAGE gel electrophoresis for the expression analysis of truncated pET28 full Adgb protein.                                                                                                                            | 116 |
| Chapter 5                                                                                                                                                                                                                              |     |
| Figure 5.6.1. Schematic diagram of the mechanism for the formation of the proximal penta-coordinate-NO haem in cytochrome c' of <i>Alcaligenes xylosoxidans</i> (contains a distal Leu residue instead of the QIn residue of Adgb-HD). | 130 |

## Chapter 1. Introduction

### 1.1 Globins

Globins are ubiquitous across the three main kingdoms of life (Archaea, Bacteria and, Eukarya) (Vinogradov et al., 2006). They are classified into three globin lineages which were constructed by phylogenetic sequence analysis of genomes from the three kingdoms of life. One lineage comprises of FHbs (flavo-haemoglobins) and chimeric proteins, which contain an N-terminal globin domain and a C-terminal FAD-binding domain. The second lineage consists of GCSs (globin coupled sensors), which act as chimeric gene regulators generally found in Bacteria and Archaea (Fig. 1.1.1). These two lineages encompass the canonical 3-over-3  $\alpha$ -helical fold globins. Whereas, the other lineage consists of single-domain globins, which holds a 'truncated', 2-over-2  $\alpha$ -helical fold (Vinogradov et al., 2007). These globins do not contain an N-terminal  $\alpha$ -helix, to which it is substituted with a proximal haem binding F helix, which has an extended non-helical loop (pre-F) (Hade et al., 2017). The Venn diagram (Fig. 1.1.1) shows that only one out of three globin lineages contain the 2-over-2  $\alpha$ -helical fold (Vinogradov et al., 2006).

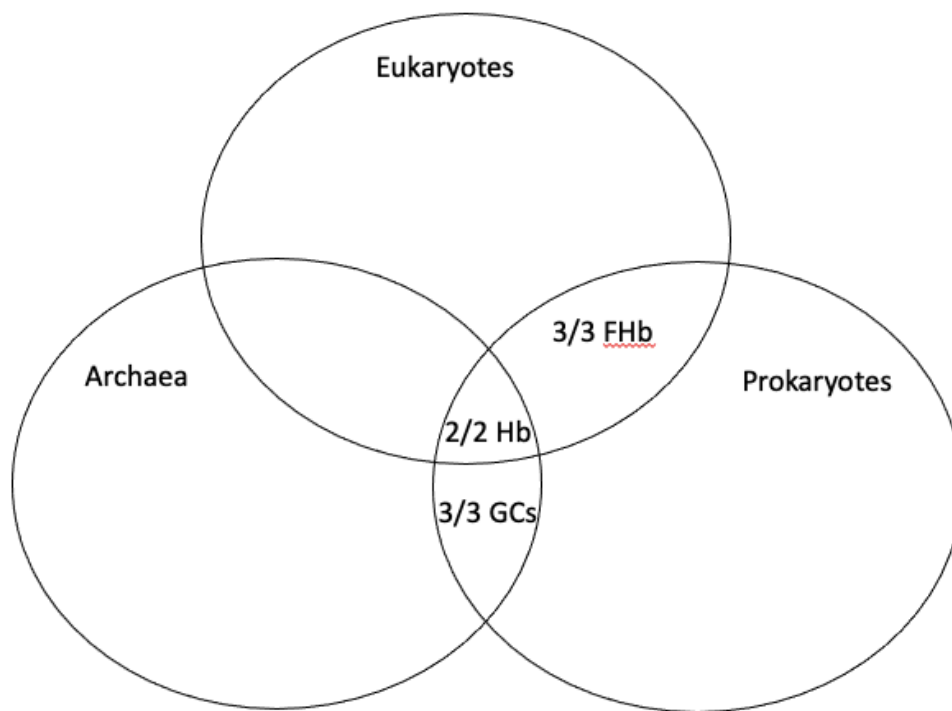


**Figure 1.1.1. A Venn diagram portraying the three globin lineages, flavohaemoglobins (FHbs), Globin coupled sensors (GCs) and truncated haemoglobins (2/2 Hbs) in the three kingdoms of life.** The types of Hbs that are selective to each kingdom of life. The ratio indicates the type of  $\alpha$ -helical fold the globin lineages contain, such as 2/2 represents a 2-over-2  $\alpha$ -helical fold and 3/3 represents 3-over-3  $\alpha$ -helical fold (Vinogradov et al., 2006).

Globins are considered to be small metalloproteins that are approximately comprised of 150 AAs (amino acids) and composed of eight  $\alpha$ -helical segments (Burmester and Hankeln, 2014). Interestingly, throughout the last century or so, haemoglobin (Hb) and myoglobin (Mb) were the only globins to be extensively studied, which facilitated the understanding between function and structure (Storz et al., 2011). However since the 1970s, further additions have been made to the globin family and the ensuing years have brought to light new globin members (Vinogradov et al., 2006). The main animal globins have been identified as, myoglobin, haemoglobin, neuroglobin (Ngb), cytoglobin (Cygb), globin X (GbX), globin Y (GbY), eye-globin or globin E (GbE) and most recently androglobin (Adgb) in 2012 (Burmester and Hankeln, 2014). The vertebrate-specific globins are predominately expressed in tissues and are involved in cellular processes associated with transport, storage, detoxification, sensing and signalling of oxygen as well as other reactive species. Therefore, they are recognised as multifunctional proteins that are vital for cellular metabolism, with further functions that may be unknown (Götting and Nikinmaa, 2015).

### 1.1.1 *The globin fold*

The globin superfamily has a classical 3-over-3  $\alpha$ -helical sandwich structure which is referred to as the globin fold (Burmester and Hankeln, 2014). This was first acknowledged and described by John Kendrew in 1958 (Kendrew et al., 1958). Bioinformatics has shown globins to display remarkable sequence diversity between family members. Despite their evolutionary distance, they contain a few well-conserved residues that maintain the globin fold and stability (Isogai et al., 1999). In general, globins customarily consist of eight  $\alpha$ -helices (labelled A-H) and a haem prosthetic group with a central ferrous iron ( $\text{Fe}^{2+}$ ) and a protoporphyrin IX configuration (Burmester and Hankeln, 2014). This haem pocket commonly binds to ligands such as oxygen in a reversible manner (Kendrew et al., 1958). There are a few well-conserved amino acid residues in globins that help maintain and stabilise the globin fold (Isogai et al., 1999). One residue conserved across all globins is the proximal histidine (F8), located adjacent to the  $\text{Fe}^{2+}$  (ferrous iron) in the haem pocket. Most globins contain a distal histidine residue (E7) however, there is some variation between species, for example, elephant Mb contains a distal glutamine residue at the E7 position instead of a histidine (Dikshit et al., 1998). The most highly conserved residue is phenylalanine in the

inter-helical CD1 region which stabilises the haem pocket (Burmester and Hankeln, 2014), the only exception is Tyr found in prokaryotic globins (Howes et al., 2011).

### 1.1.2 The haem Pocket and co-ordination of globins

The *alpha*-helices of globins surround a cavity that contains the haem group. The schematic structure of the haem pocket is illustrated in Figure 1.1.2.1. The haem pocket itself is four pyrrole rings linked by methylene bridges to form a protoporphyrin IX configuration, complexed with a coordinated central iron ion (Fig 1.1.2.1A). The central iron atom sits out of plane and is coordinated to four of the pyrrole nitrogen atoms from the protoporphyrin ring. The fifth coordination site is held by the proximal histidine (Fig. 1.1.2.1B) and the sixth site takes on an exogenous ligand/ the distal histidine (Aronson et al., 1994). In addition, the iron atom can exist in different oxidation states such as ferrous ( $\text{Fe}^{2+}$ ), ferric ( $\text{Fe}^{3+}$ ) and ferryl ( $\text{Fe}^{4+}$ ). The ferryl state is unstable and rarely exists. The protoporphyrin IX group carries two negative charges, whereas the ferrous iron has two positive charges (hence  $\text{Fe}^{2+}$ ). The opposite charges provide the haem group with an overall neutral charge that binds reversibly to neutrally charged ligands such as oxygen, carbon monoxide (CO) and, nitric oxide (NO). When the iron is in the ferric state ( $\text{Fe}^{3+}$ ), the overall charge of the haem group becomes slightly positive (He et al., 2000). Therefore, negatively charged ligands such as cyanide and azide bind better to the ferric iron state (Aronson et al., 1994).

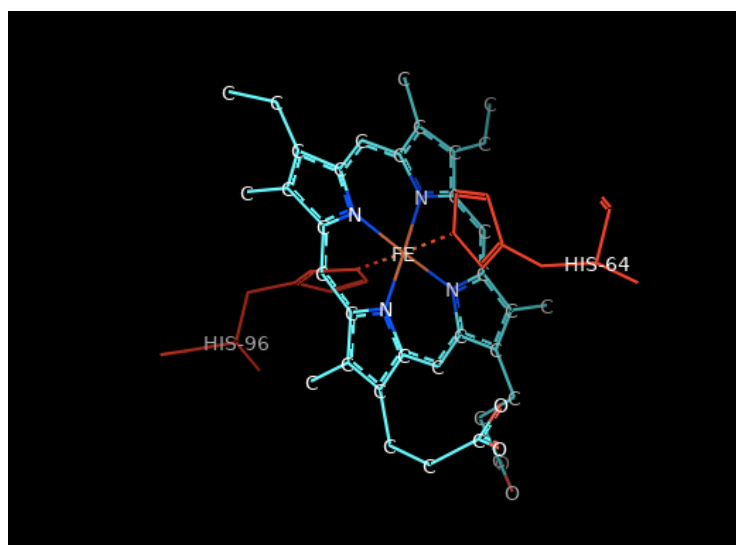
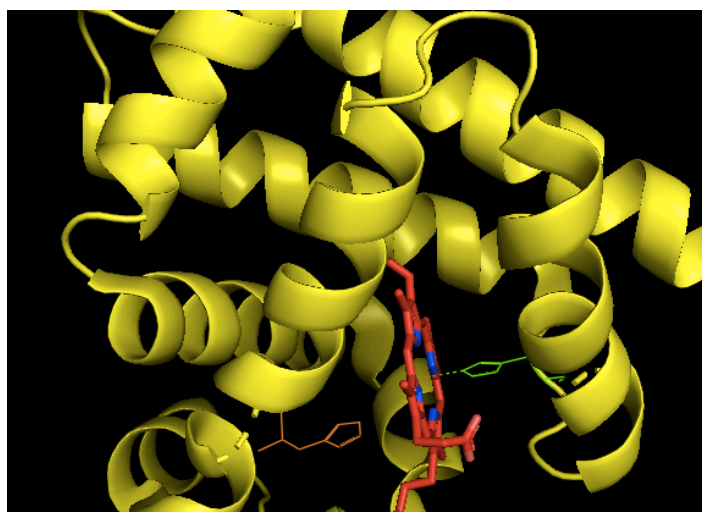
**A****B**

**Figure 1.1.2.1. The structure haem group and haem coordination.** (A) shows a skeletal structure of the haem group (protoporphyrin IX). The diagram shows that the iron ion is usually coordinated to four pyrrole nitrogens and the proximal histidine. The sixth coordination site is open in pentacoordinate globins (ferric and ferrous states) but is occupied in the ferric high-spin hexacoordinate globins and low-spin hexacoordinate globins. The haem group was derived from a monomer of Ngb (PDB code: 4MPM) (Woody and Hsu, 1971). (B) The ferrous iron coordination state of Cygb (in pink, left) representing hexacoordination and the ferrous iron coordination state of Mb (in yellow, right) representing pentacoordination. The distal His (orange) is not coordinated to the 6<sup>th</sup> haem position in Mb as it is in Cygb. In both cases, the proximal His residue (green) is bound to the iron (Cygb PDB code: 1URY and Mb PDB code: 3RGK). The models were constructed on PyMOL molecular graphic system v.2.3.5.

Furthermore, globins differ in haem coordination chemistry designating them as either pentacoordinate or hexacoordinate globins (Uppal et al., 2016). Therefore, when the iron is in the ferrous state (deoxygenated state), Hb, Mb and, GbE are pentacoordinated. The sixth binding site of  $\text{Fe}^{2+}$  is unoccupied and available for reversible binding to a potential ligand. Meanwhile, Ngb, GbX, Adgb and, Cygb are hexa-coordinated which means the sixth coordination site of  $\text{Fe}^{2+}$  is bound to the distal histidine (or distal glutamine for Adgb) in the protein chain. Pentacoordinate species usually have a broad visible absorption peak around ~ 500-635 nm indicating the haem iron is in the high-spin electronic configuration during the ferric and ferrous oxidation states. The His coordination to the sixth axial position converts the haem iron to a low spin electronic configuration in both oxidation states, resulting in split visible peaks near 560 and 530 nm for the ferrous state and stronger absorbance observed in the ferric state (Kakar et al., 2010). The distal histidine is located on the conserved E7 position for all globins (Burmester and Hankeln, 2014). The displacement of the distal histidine by a ligand, triggers a conformational change to occur. Ligands are considered to be competitive with the distal histidine, where binding is dictated by affinity (De Sanctis et al., 2004). Ligand kinetic and affinity rates are limited due to the rate of dissociation of the distal histidine (Ceccarelli and Scorciapino, 2012). Haem coordination chemistry for previous and newly established globins can be determined by absorbance spectroscopy. The haem coordination chemistry is determined by the splitting of  $\alpha/\beta$  bands, which helps to assign coordination of the central iron atom, usually in the ferrous state (deoxygenated state). This method can also be used to distinguish between hexacoordinate and pentacoordinate haem species in the absence of exogenous ligands and is monitored by UV-vis, fluorescence, CD or NMR spectroscopy. Lastly, pentacoordinate globins are mostly responsible for conveying and depositing oxygen (Song et al., 2019). Whereas the function of hexacoordinated globins has not yet been clarified, although some of the physiological functions may incorporate the protection of cells under hypoxic/ischemic conditions, being a scavenger of reactive oxygen species (ROS) and/or averting programmed cell death (Kim et al., 2019).

## 1.2 Cellular role and reactions of globins

### *1.2.1 Storage and Transport.*

Early into vertebrate evolution, adequate oxygen supplies in support of aerobic metabolism were formed via the development of circulatory systems. This system allowed sufficient delivery of oxygen to respiring cells, in cooperation with specialised oxygen transport and storage proteins. The development of this occurred due to an increase in body size and internal complexity, which made passive diffusion insufficient to meet the aerobic demands of animal life (Hoffmann et al., 2011). Globins are expressed abundantly in species and vastly in a tissue-specific manner. They are involved in cellular processes associated with oxygen transport and storage, detoxification of ROS and, nitrogen species, sensing and cellular signalling (Burmester and Hankeln, 2014). Primarily, globins are recognised for their respiratory functions. For instance, Hb is the most widely distributed globin and is the only globin to be expressed in vertebrate erythrocytes (Götting and Nikinmaa, 2015). The absorbance properties of the haem porphyrin moiety in Hb is accountable for the putative red pigment observed in vertebrate blood (Dwivedi et al., 2019). Vertebrate Hb transports oxygen from respiratory surfaces such as lungs, gills and, skin to the internal organs by the circulatory system. Typical Hbs of Gnathostomata (jawed vertebrates) are tetramers, composed of four subunits. The study of Hb has shed light on the relationship between ligand binding and affinity of the protein. The interaction between the subunits of Hb essentially reduces the oxygen-binding affinity of Hb, which is lower compared to globin monomers (e.g. Mb). The purpose of this is to allow cooperative binding that is linked to the evolution and survival in extreme environments. For example, some vertebrates can survive in high altitude habitats, which are more physiologically challenging (Burmester and Hankeln, 2014).

## 1.3 Androglobin

### *1.3.1 The Structure and Evolution of Androglobin*

A novel discovery of chimeric proteins with a putative N-terminal calpain-like domain, a globin domain and, an IQ calmodulin-binding motif was determined in 2012 by Hoogewijs *et al* (Burmester and Hankeln, 2014). It is the fifth mammalian globin to be identified by extensive *in silico* searches of unannotated globins in deuterostome sequence data (Hoogewijs et al., 2012). This was further verified by the identification of orthologues in the genomes of metazoan lineages and choanoflagellates (Burmester and Hankeln, 2014). Due to its highly abundant expression in mammalian testis, the name 'Androglobin', was proposed for this novel protein. Androglobin is also expressed in the heart, brain and, lungs (Huang et al., 2014). The complete *Adgb* genomic sequence was obtained from blast searches of sea urchins and amphioxus genomes, displaying a gene sequence of 216460 bp and is located at the 6q24.3 locus (Hoogewijs et al., 2012). In humans, the protein is 1667 AAs (Reeder, 2017). Phylogenetic analysis of Ngb, GbX and Adgb show major clades that are dissimilar to other globin family members, providing implications of their early evolutionary origin. Interestingly, the identification of Ngb, GbX, and Adgb in both vertebrates and invertebrates, provides evidence these globins may have existed before the separation of Protostomia and Deuterostomia. In addition, all three globins are hexa-coordinated, indicating that this was the original coordination state in the primary deoxy globin. Therefore, the typical penta-coordination of most globins may be a derivative that has evolved (Burmester and Hankeln, 2014). Adgb is subject to different animal lineages, as its ancestral lineages trace back to choanoflagellates and metazoan taxa. Furthermore, the modular structure comprises of four segments. The N-terminal calpain-like domain is ~350 residues, then a region of unknown motifs/ domains ~300 residues in length proceeds, followed by a circularly permuted globin domain that is ~150 residues and, lastly an uncharacterised C-terminal domain which is ~750 residues (Hoogewijs et al., 2012).

### 1.3.2 The haem domain of Androglobin

A schematic representation of the Adgb multi-domain is shown in Figure 1.3.2.1. The Adgb globin domain satisfies the criteria of a typical globin fold, observed in its other family members (Burmester and Hankeln, 2014). It is predicted that Adgb contains the 3-over-3  $\alpha$ -helical architecture (Ascenzi et al., 2016). The globin domain associated with Adgb consists of 8 alpha helices (A-H). The initial study displayed that Adgb possesses an unusual arrangement, by which it is circularly permuted and split into two sections. Helices A-B is located on the C-terminal side and is distinctively separated from helices C-H by an IQ calmodulin-binding motif, as shown in Figure 1.3.2.1 and Appendix 10 as the Zurich model. The shifted A-B helical segment was confirmed by the presence of a conserved B12.2 intron position, which is seen in other globins (Hoogewijs et al., 2012). The Zurich model sequence is the only published sequence analysis for Adgb haem domain thus far. The recombinant human Adgb haem domain (C-H helices) has an absorbance spectrum that is characteristic of hexa-coordination (Burmester and Hankeln, 2014). Despite its permutation, alignments of the split Adgb globin domain compared to Mb, Cygb and, Ngb mammalian sequences suggest that Adgb corresponds to a typical 'globin fold' tertiary structure. Molecular modelling of the isolated C-H region displayed the ability to produce a globin fold alone. Furthermore, comparisons of 3D structures of the shortened Adgb globin domain with the human Cygb globin domain were completed. The alignments revealed that the Adgb haem domain model only consisted of 6 alpha helices (A and D helices were absent), the results presented a longer H-helix which indicates a short globin domain and stable globin fold in Adgb (Hoogewijs et al., 2012).

The globin domain of Adgb contains functionally conserved residues for anchorage and stability such as, the proximal His (histidine) F8, which forms a coordination bond with the central iron atom and a distal Gln (glutamine) residue instead of the customary distal His E7 seen in other globins. Therefore, Adgb has a hexa-coordination scheme of [Gln]-Fe-[His] which is proposed from the in the sequence alignments in Figure 1.3.2.1 with experimental data in Figure 4.1.1 on the purified protein confirming this (Hoogewijs et al., 2012). This hypothesis was also tested by 3D modelling (Chapter 3) and UV/Vis spectroscopy (Chapter 4). The initial study by Hoogewijs *et al* indicated a conserved Phe (residue 770) in the inter-

helical CD region. However, it has been speculated that a Tyr (tyrosine) is a candidate residue that may lie in the CD1 position instead to provide anchorage for enhanced function (personal communication with Dr Brandon Reeder). It is widely known that bacterial globins, for example, globins present in *Mycobacterium leprae*, *Corynebacterium diphtheriae* and, *Streptomyces coelicolor* contain a Tyr residue in the CD1 position. These species have selectively adapted to host a Tyr instead of Phe (phenylalanine) residue to support its unique functional roles in CO, oxygen and, NO binding through its stronger hydrogen bonding network with the haem group (Mukai et al., 2002). The initial study in 2012 by Hoogewijs *et al* proposed a structure for androglobin, created using 34 Adgb sequences, since then a considerable amount more Adgb sequences have been made available from a range of different species, which helped to provide a new consensus sequence by Dr Reeder and Professor Reynolds (Appendix 11, the Essex model sequence), whereby Adgb was aligned to  $\alpha$ -Hb,  $\beta$ -Hb, Cygb, Mb and Ngb and profiles for helices A-B and E-H were generated for each protein. The profile alignment between Adgb and each of  $\alpha$ -Hb and  $\beta$ -Hb, cytoglobin, myoglobin and neuroglobin was carried out by doing all possible pairwise alignments and counting the votes for the top alignment, to which consensus alignment was determined by multiplying the votes together. In this consensus sequence helices A-C (at the C-terminus instead of N-terminus) are segregated from helices D-H (at the N-terminus) by the IQ-calmodulin binding motif therefore, the D helix becomes the first helix at the N-terminus from the haem domain. The shift in the new consensus sequence provides a new CD1 candidate, Tyr975 (Appendix 11) (personal communication with Dr Reeder and Professor Reynolds, unpublished work).

Genes that were separate and form the chimeric protein. androglobin and calpain structures are compared here.

IQ motif mediates binding of calmodulin

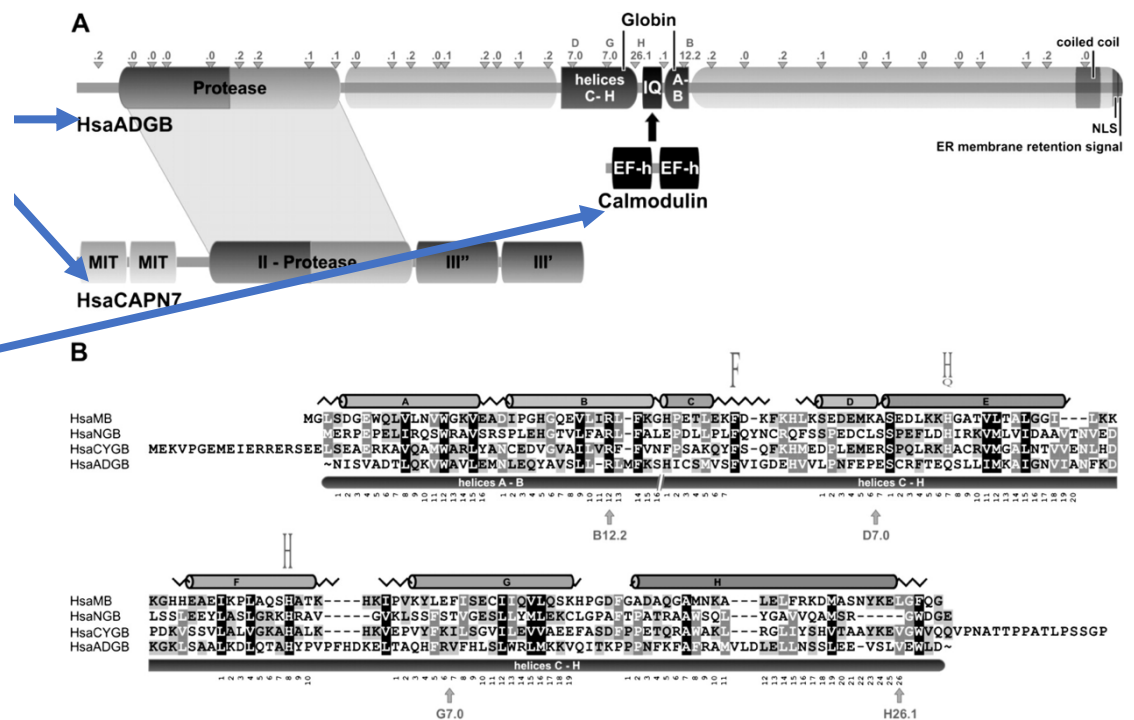


**Figure 1.3.2.1. A schematic representation of the domain structure of Adgb generated by Hoogewijs et al. (A)** Adgb contains a circularly permuted globin domain that is split by the IQ motif. On the C-terminal side, it appears to have a coiled-coil region, NLS and, an ER membrane retention signal. The small triangles, located above the domain indicate the position of introns. Below the domain, there is another schematic diagram of the human calpain structure. This demonstrates a clear similarity of the calpain protease region and the protease region of Adgb. **(B)** shows the sequence alignments of the Adgb globin domain with Ngb, Cygb and, Mb. The key conserved residues are indicated on the  $\alpha$ -helical structure above the sequences. The grey arrows underneath represent further intron positions in the globin domain of Adgb. Diagram taken and modified from the original study by Hoogewijs et al (Hoogewijs et al., 2012).

### 1.3.3 The calpain domain

The domain in Figure 1.3.2.1, shows that Adgb owns a distinct calpain-like domain (Burmester and Hankeln, 2014). Calpains are considered to be cysteine proteases regulated by calcium ions, which modulate the function/localisation of intracellular proteins by protein cleavage (Briz and Baudry, 2017). They also influence pathological signalling that is associated with diabetes, stroke, AD, muscular dystrophy, and traumatic brain injuries (Huang and Wang, 2001). Calpains exist as heterodimers and are made up of four domains. The catalytic domain II contains 3 active site residues Cys (cysteine), His and Asn (Asparagine). The Cys is located in the subdomain IIa and His and Asn are in the subdomain IIb (Strobl et al., 2000). Eugene searches have shown that the N-terminal of Adgb is homologous to the catalytic domain II of human calpains and is closely related to human calpain-7. Further evidence via sequence analysis presented a 47% and 43% sequence identity to the subdomains IIa and IIb in calpain-7. The calpain region of Adgb contains a conserved Cys residue in the same position, identical to the subdomain IIa of calpain. The other active site residues (His and Asn) are located in non-standard positions. This insinuates Cys-protease activity may be associated with Adgb's function (Hoogewijs et al., 2012).

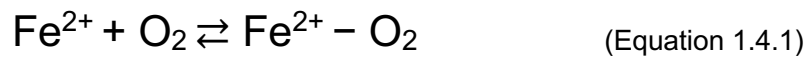
### 1.3.4 Other domains of Adgb

In addition, Adgb contains a calmodulin-binding IQ motif that interrupts the globin helices (Hoogewijs et al., 2012). Calmodulin is well known for mediating the effects of  $\text{Ca}^{2+}$ . It is a critical cellular second messenger that promotes highly localised calcium signalling for regulation of molecular motor activity, cytoskeletal organisation and mitosis (Bähler, and Rhoads, 2002). Figure 1.3.2.1 shows the Ef-hand motifs, demonstrating calmodulin's ability to bind to the IQ-calmodulin binding motif. In this instance, it is assumed that functions of Adgb may be regulated by  $\text{Ca}^{2+}$  (Calcium) and by the IQ calmodulin-binding motif (Hoogewijs et al., 2012). The binding affinity of calmodulin to the IQ-calmodulin binding motif is thought to vary with  $\text{Ca}^{2+}$  concentration indicating the regulatory role of calmodulin associated with Adgb. The IQ region of Adgb is a sequence where the first two residues correspond to conserved isoleucine and

glutamine residues. This sequence acts as a binding site for the EF-hands of calmodulin (Bähler, and Rhoads, 2002). Supplementary to this, Figure 1.3.2.1 shows that Adgb contains an overlapping ER (endoplasmic reticulum) membrane retention signal at the C-terminus and a coiled-coil motif which could be involved in protein dimerization, a suitable candidate for NLS (nuclear localisation signal) (Hoogewijs et al., 2012).

#### 1.4 Allosteric regulation of globins.

When the haem pocket interacts with an external ligand like oxygen, there is an effect on both structure and affinity in Hb. Equation 1.4.1 demonstrates the reversible binding of oxygen to the ferrous haem iron (Fago, 2017).



Reversible oxygen-binding causes an allosteric transition between two states (R and T state). This leads to a shift in the quaternary structure of Hb (Perutz, 1970). One state of Hb is called the tense state (T state) which is stabilised by the formation of salt bridges between the carboxyl-terminal residues of the subunits (Rodwell and Kennelly, 2003). When the oligomeric protein is in this state, it obtains a relatively low oxygen affinity. The relaxed state (R-state) is stabilised by oxygen binding, causing the cavity of the tetramer to become smaller. This state has a higher affinity for oxygen binding (Pennelly et al., 1975). The mechanism was first described in 1965 as the MWC model (Monod-Wyman Changeux), also known as the concerted model. The model proposed that the structural conformational changes in Hb drive ligand binding in a cooperative manner (Monod et al., 1965). The organs that require higher oxygen tensions ( $p\text{O}_2$ ) usually favour the R state (oxygen uptake). This is as opposed to organs favouring the T-state which release oxygen and have a lower  $p\text{O}_2$  (Fago, 2017). The transition between the states produces a quaternary R-T equilibrium which can be represented by an oxygen saturation curve (Pennelly et al., 1975). This saturation curve is demonstrated in blue in Figure 1.4.1 below showing the cooperative binding of Hb, where Hb stores oxygen in the lungs and releases oxygen at peripheral tissues, which is permitted by the transition between states. Ultimately, the MWC model provides an understanding of the effects of ligand binding for oligomeric proteins such as Hb. However, for monomeric proteins oxygen-binding is different. Mb is more suitable for oxygen storage rather than transport. The oxygen binding curve for Mb is hyperbolic (Fig 1.4.1) where Mb can readily offload oxygen whilst at rest. The  $p\text{O}_2$  decreases in the muscle tissue during strenuous exercise, as oxygen is being diffused into mitochondria to drive ATP synthesis. Figure 1.4.1 also shows that the oxygen-binding of Hb (tetramer) results in greater oxygen delivery compared to monomeric Mb (Rodwell and Kennelly, 2003).

Adgb has a similar globin fold to Hb and Mb however, it is hexa-coordinated unlike Hb and Mb (Burmester and Hankeln, 2014). Additionally, hexacoordinated proteins (such as Adgb) undertake rapid auto-oxidation of the ferrous state to the ferric iron state, relative to Hb and Mb (Dewilde et al., 2001). Whilst in the ferric state oxygen is unable to reversibly bind to Adgb therefore, this indicates that the prime function of Adgb is not associated with oxygen, despite containing a globin fold similar to Mb and Hb (Das et al., 2000) In addition, Adgb is predominately expressed in the testes (Hoogewijs et al., 2012), where oxygen levels are known to be lower (~10-15 mm Hg), this further suggests that the physiological function of Adgb is not oxygen transport (Lysiak et al., 2000).

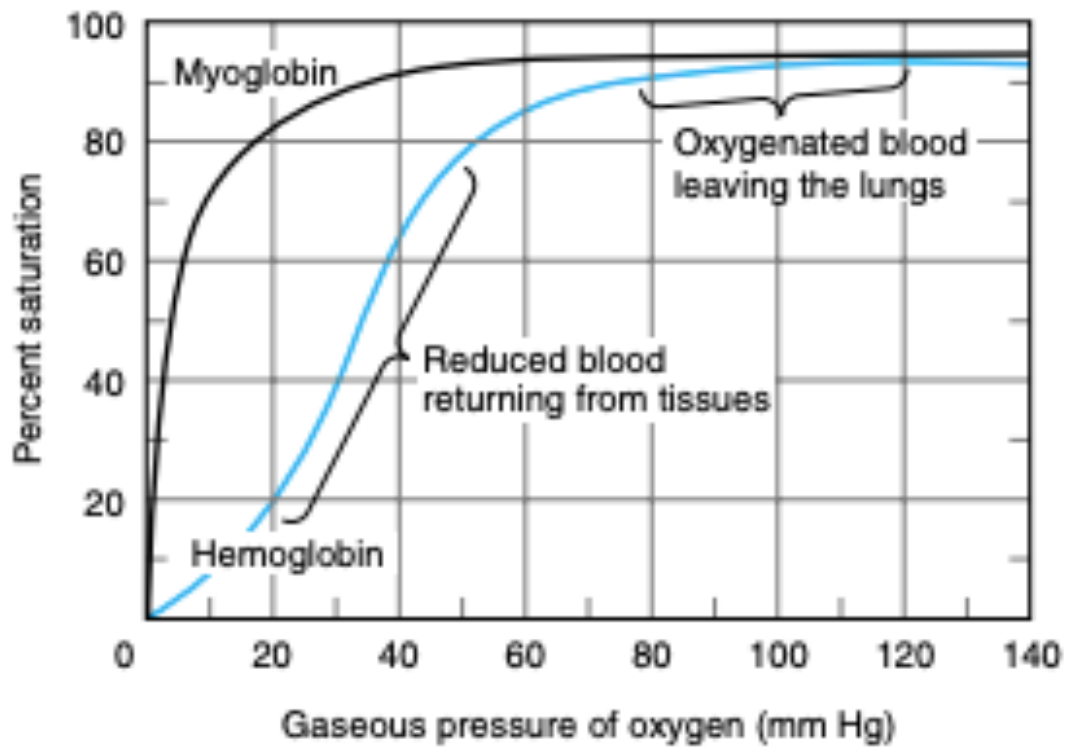


**Figure 1.4.1. Oxygen saturation curves of haemoglobin and myoglobin.** The blue curve represents oxygen binding of Hb which is sigmoidal (cooperative ligand binding), whereas the black curve is hyperbolic representing the oxygen binding of Mb (Rodwell and Kennelly, 2003).

## 1.5 The role of globins in Nitric oxide metabolism

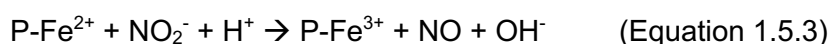
Nitric oxide is considered a critical signalling molecule and is strictly mediated in eukaryotic systems by globins (Tejero and Gladwin, 2014). NO plays an important role in vascular physiology as an autocrine and paracrine vasodilator molecule (Minnecci, 2008). Besides well-defined oxygen transport functions, Hb functions in nitric oxide (NO) metabolism as well. Oxy-Hb can scavenge excess toxic NO and convert it to nitrate via a dioxygenation reaction as shown in Equation 1.5.1. (Cosby et al., 2003).



The rate of NO deoxygenation *in vivo* is effectively reduced by compartmentalisation of Hb within erythrocytes (Azarov et al., 2005). This restriction by Hb in erythrocytes reduces the rate of NO inactivation by ~1000 fold, allowing NO to regulate vasodilation (Minnecci, 2008). Deoxy-Hb can also reduce nitrite back to NO (nitrite reductase activity) in the blood during physiological hypoxia to allow vasodilation (Huang et al., 2005). In addition, Mb is instrumental in reducing the toxicity of NO. Oxy-Mb is able to detoxify NO in cells with high metabolic activities by its dioxygenase enzyme activity (Burmester and Hankeln, 2014). On the other hand, deoxy-Mb obtains nitrite reductase activity, whereby it can generate bioactive NO by reducing nitrite in a similar way to Hb during cellular hypoxia (Hendgen-Cotta et al., 2010). The ability of Mb to mediate hypoxic vasodilation in the vascular system has biomedical implications for microvascular dysfunction (Burmester and Hankeln, 2014). NO regulation by Mb and Hb are highly suited for this role, as they are abundant in cells that help compensate for slow nitrite reduction rates and have a reduction system allowing them to regenerate the reduced form (Tejero and Gladwin, 2014).

The NO scavenging functions are conceivable for hexacoordinate globins such as, Ngb and Cygb which function as NO dioxygenases (Reeder and Ukeri, 2018). Ngb scavenges NO in neuronal cells and provides neuronal protection from harmful NO molecules (Burmester and Hankeln, 2004). Ngb plays a key role in brain homeostasis, whereby Ngb serves as a sensor to hypoxic stress. In addition, the NiR

activity of Ngb can inhibit mitochondrial respiration. In the presence of nitrite, oxygen consumption is suppressed and a direct link between mitochondrial integrity and Ngb *in vivo* has been established. For example, Ngb is associated with the reduced oxidative damage induced by reactive nitrite species (Baez et al., 2016). Cygb can be reduced by ascorbate and cytochrome b<sub>5</sub>, which can allow for NO dioxygenation and the production of NO in hypoxic conditions (Tejero and Gladwin, 2014). Under normal physiological conditions, Ngb and Cygb scavenge NO (Equation 1.5.2) however, in hypoxia both globins are able to switch function to a nitrite reductase and produce NO (Equation 1.5.3) (Reeder and Ukeri, 2018).



In comparison, Ngb and Cgb are slower nitrite reductases than Mb yet hexacoordinated plant Hbs are faster nitrite reductases than Mb (Tejero et al., 2015). A similar protection property of Ngb has been detected in FHbs, to protect from macrophages. *E.coli* expresses FHb in response to increasing NO concentrations. The mechanism involves scavenging NO and reducing it to N<sub>2</sub>O (Nitrous oxide) during hypoxia. The NO dioxygenase activity of FHbs is found to be more than 20-fold higher than that of Mb or Hb (Bonamore and Boffi, 2008). The NiR activity of haem containing globins represents conserved biochemical roles from an early evolutionary era (pre-aerobic periods) which could suggest that the primary form of NO synthases is preserved in modern-day globins during normoxic and hypoxic conditions (Tiso et al., 2012). The prime function of Adgb is likely associated with NO as it is hexa-coordinated globin, which is usually associated with a high autoxidation potential. Adgb is also expressed highly in the testes with implications to NO homeostasis in the testes and spermatogenesis (Hoogewijs et al., 2012).

## 1.6 Thiol groups in globins

Vertebrates surviving in different environments possess differently adapted Hbs. For example, freshwater turtle Hbs have a high level of Cys residues (reactive thiol groups) compared to other existing vertebrates. The presence of these residues is responsible for protecting against ROS in the mitochondria. This process occurs during the recommencement of respiration after hypoxic periods. Therefore, the abundance of thiols in *Trachemys* and *Chrysemys* Hbs are significant for biochemical adaptation to hypoxia, as such vertebrates experience fluctuations between hypoxic and normoxic conditions. This is considered to be an adaptive strategy for oxidative defence. Furthermore, thiol groups partake in the oxygenation-linked S-nitrosation (addition of NO to a thiol group) in Hb. This contributes to hypoxic vasodilation which causes the release of S-NO. This process is observed in the Mb of salmon fish too (Fago, 2017). Cys residues have essential functions associated with protein structure by conferring stability through the formation of disulfide bonds (Meitzler et al., 2013). Reactive Cys residues have been detected in members of the globin family, for example, irreversible oxidation of  $\beta$ Cys93 in Hb can lead to the collapse of the Hb structure and can cause the loss of the haem group. Interestingly, foetal Hb (HbF) contains the same reactive Cys in the  $\gamma$ Cys93 position. Site-directed mutagenesis studies have demonstrated the effectiveness of the  $\gamma$ Cys93 residue in HbF. For instance,  $\gamma$ Cys93 in HbF was mutated to an Ala residue (removal of  $\gamma$ Cys93) and oxidative instability was characterised by accelerated oxidation rates. The addition of a surface-exposed Cys at the  $\alpha$ 19 position, caused an alteration in regular electron transfer, providing another oxidative and disulfide bonding site between Hb subunits, ultimately enhancing the stability of HbF (Kettisen et al., 2018).

Ngb and Cygb possess surface-exposed cysteine residues capable of forming a disulfide bond. Reduction of the disulfide bond in Ngb increases the affinity of the distal histidine by a factor of 10, prompting the release of oxygen (Hamdane et al., 2005). The surface-exposed cysteines in human Ngb and Cygb are fundamental for ligand binding and the enzymatic properties of the proteins (Reeder and Ukeri, 2018). The cysteine residues are found in pairs; for Cygb the disulfide bond is distributed across two helices Cys38 on helix B and Cys83 on helix E which are highly conserved throughout mammals. For human Ngb

the surface-exposed cysteines (Cys46 and Cys55) are located on the C and D helix (CD loop) which forms a disulfide bond upon oxidation. These residues are highly conserved across Ngb in other vertebrates (Reeder, 2017). The 3D structure of human Ngb is shown in Figure 1.6.1, displaying the cysteines in the CD loop region, that form a disulfide bond in Ngb (Hamdane et al., 2005). The oxidation of cysteines in the CD loop of Ngb (Figure 1.6.1) affects the NiR activity of Ngb. A two-fold decrease in NiR activity has been detected in Ngb when the disulfide bond is reduced (Reeder and Ukeri, 2018). The redox state of surface-exposed Cys in Ngb modulates the NiR, which enhances the formation of NO. Disulfide bond formation is accompanied by a decrease in the distal His binding affinity to haem iron. Therefore, the redox state of Cys can influence the distal His and nitrite affinity to the haem iron. The oxidised disulfide-bridged Ngb also exhibits a higher affinity for nitrite than the thiol-reduced form, allowing Ngb to inhibit mitochondrial respiration in the presence of nitrite. Therefore, the oxidation of surface-exposed Cys in the CD loop of Ngb helps to regulate the hypoxic inhibition of cellular respiration by increasing Ngb's affinity to nitrite (Tiso et al., 2011).

The globin domain of Adgb is similar to that of Ngb and shows affinity to the Ngb lineage. With Adgb's implications to NO homeostasis and relation to Ngb, this indicates that Adgb globin domain could have a redox-regulated function as postulated for Ngb (Hoogewijs et al., 2012). In the Essex model sequence, shown in Appendix 11, it is thought that potential Cys residues in the CD region could form a disulfide bond between Cys786 (D helix) and Cys977 (C helix) similar to the surface-exposed cysteines detected in Ngb (Tiso et al., 2011). The presence of an intramolecular exposed disulfide bond in Adgb could provide stability to the haem pocket as well as regulate conceivable redox functions of Adgb.

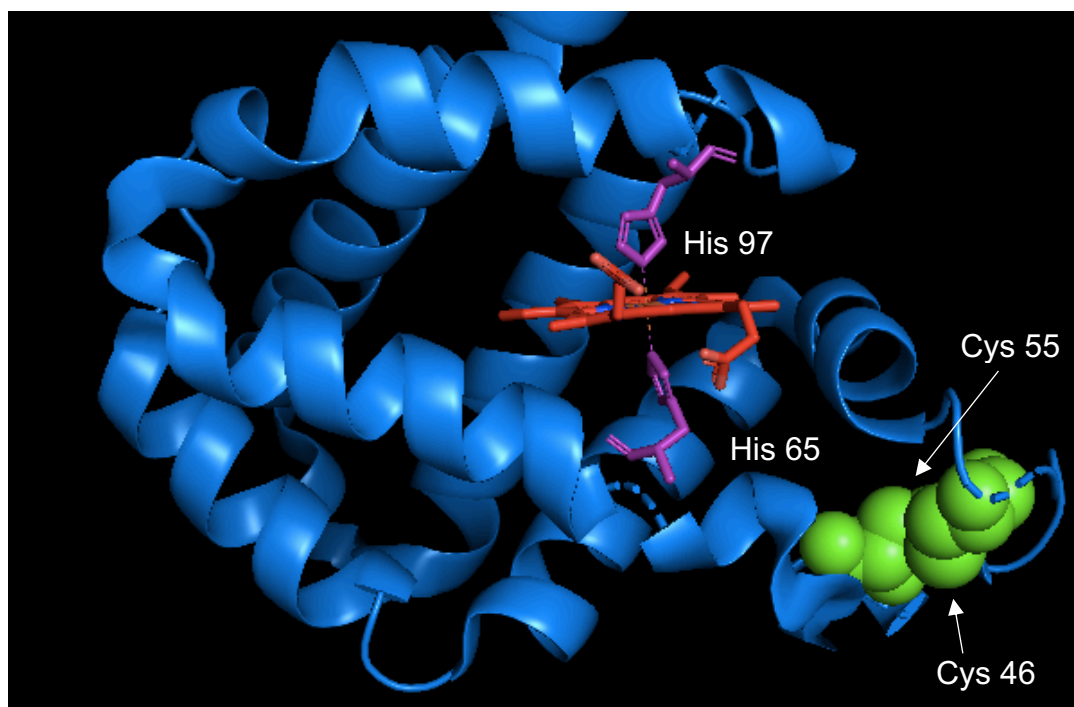


**Figure 1.6.1. A 3D diagram of the structure for human Ngb.** The haem pocket of the protein is shown in blue. The haem group is shown in red as well as the distal His (E7) and proximal His (F8), shown in purple. The proximal and distal histidines help to stabilise the protein. The surface-exposed cysteines in Ngb are located in the CD loop (shown in green). The CD7 and D5 Cys sites indicate where the disulfide bond is formed upon oxidation. The model were constructed on PyMOL molecular graphic system v.2.3.5 (Hamdane et al., 2005).

### 1.7 Biomedical implications of globins.

Defects present in globins can lead to pathological disorders and have negative effects on disease states. The majority of globin-related diseases are inherited yet some may sporadic or determined by environmental stimuli. One common blood disorder caused by the abnormal structural integrity of Hb is Thalassemia. Thalassemia is the result of the absence of one or more  $\alpha/\beta$  subunit chains in Hb (known as  $\alpha/\beta$  Thalassemia). This can occur from 750 different mutations (Vinogradov, and Moens, 2008). Individuals affected suffer from severe anaemia and hepatosplenomegaly (swelling of the liver and spleen), subsequent to this life expectancy is shortened (Wang et al., 2012). Bone marrow transplantations throughout an individual's life are the only current treatment option available (Vinogradov, and Moens, 2008). However, the recent generation of genetically corrected iPSCs (induced pluripotent stem cells) derived hematopoietic stem cells, has potentially provided a solution for curing this blood disease (Wang et al., 2012). Complications associated with Mb is usually associated with severe injury to the skeletal muscle which causes rhabdomyolysis. Rhabdomyolysis is the result of rapid skeletal muscle breakdown that causes the leakage of toxic cellular contents, entering the circulatory system. In muscles, Mb is bound to plasma globulins and during rhabdomyolysis Mb levels elevate. Mb then precipitates in the glomerular filtrate of the kidneys, which is highly toxic and causes acute renal failure. When treated early, rhabdomyolysis has a good prognosis and a low mortality rate (8-10%) (Zutt et al., 2014). Moreover, it has been reported that when levels of Ngb have been significantly reduced the risk of developing Alzheimer's disease (AD) is increased. Alzheimer's disease is the most common neurodegenerative disease associated with elderly people and is caused by the plaque deposits of  $\beta$ -amyloid and neurofibrillary tangles. Therefore, the age-related decline of Ngb is correlated with the development of AD (Chen et al., 2012). Likewise, it has been found that the overexpression of Ngb can counteract this disease and protect against AD. Therefore, Ngb could be a potential therapeutic target for preventing the accumulation of  $\beta$ -amyloid accumulation in AD (Brittain et al., 2010). Furthermore, Cygb has an interesting ability to suppress tumour growth however, the changes in expression can encourage carcinogenesis. Therefore, tumour growth is sensitive to Cygb concentration which is evident by reduced expression of Cygb in cancer cells. When *Cygb* was overexpressed a reduction in tumour growth is also experienced,

demonstrating Cygb's ability to act as a tumour suppressant and an oncogene. The loss of Cygb expression in cancer cells can be attributed to the loss of heterozygosity (LOH) and hypermethylation of CpG islands in the promoter region of Cygb gene. Therefore, quantitative research of hypermethylation levels at the Cygb promoter region in different cancers would make Cygb a new biomarker for cancer research (Bholah et al., 2015).

Like Cygb, Adgb has been reported to play a role as an oncogene, encouraging cell proliferation in cancer cells. (Reeder, 2017). This was observed in a recent study whereby Adgb RNA was silenced in glioma cell lines. The Adgb concentrations and colony growth were monitored in cancer cell lines, U87, U25 and, U373. The results are shown in Figure 1.7.1 where there appears to be a difference in the growth of glioma cell lines (shAdgb) and control cells. Cell lines with Adgb knockdown have a slower proliferation rate. This study also revealed that as Adgb levels were decreased, cell apoptosis was induced causing a reduction in colony formation of glioma cell lines. The results further presented an altered phosphorylation activity of key cell signalling proteins associated with the mTOR and MAPK pathways. This study implies that Adgb has wider functions that are associated with the progression of cancer (Huang et al., 2014). Therefore, the results provide further incentive to analyse other globins in tumour biogenesis research. For example, Cygb which already plays a putative role in cancer development could provide an insight into Adgb's role in tumour proliferation (Altinoz et al., 2016).

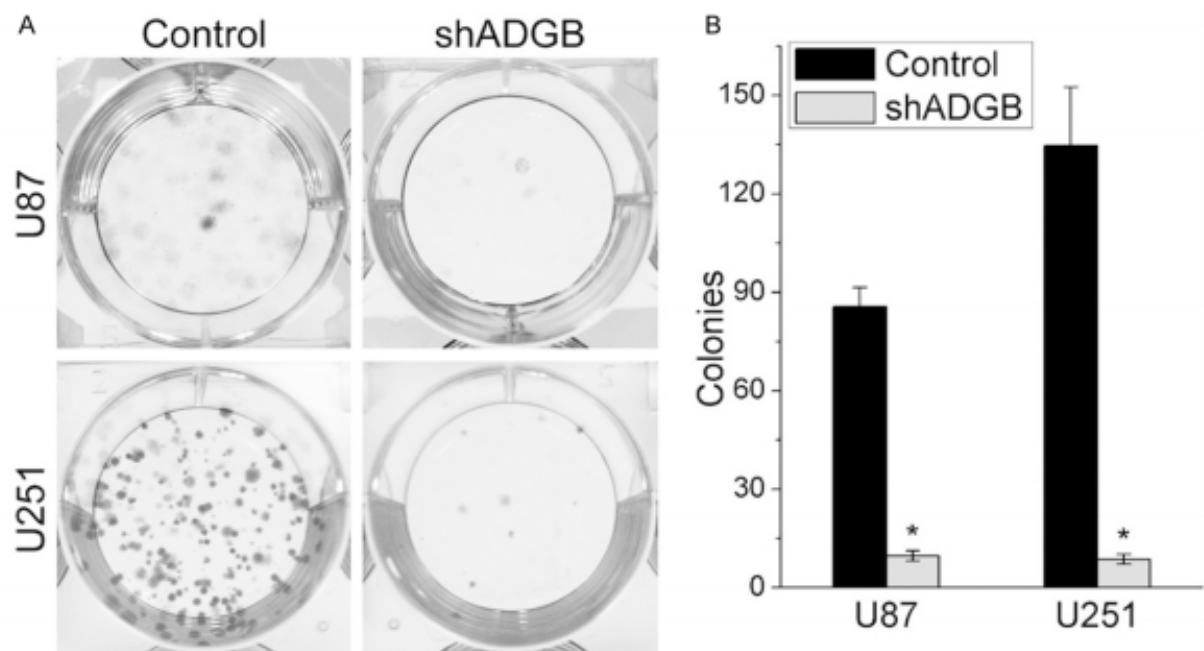


**Figure 1.7.1. Adgb knockdown inhibits cell proliferation of glioma cell lines U251 and U87. (A)** Stained images of glioma cells in a 6-well plate have been transfected with control or shAdgb lentiviruses. **(B)** The quantification of colony number. The chart represents the colony counts with error bars from the results displayed in **(A)** (Huang et al., 2014).

## 1.8 Functions of Androglobin

The key function of Adgb is yet to be determined however, some aspects of its structure give an indication of its function. The haem coordination of Adgb suggests that it may play a role in regulating redox signalling as well as oxygen sensing functions. This has been seen in Ngb to which Adgb is related to (Hoogewijs et al., 2012). It is also well known that Adgb is expressed abundantly in the testes therefore, it is assumed that spermatogenesis may involve Adgb (Gleiner et al., 2012).

### *1.8.1 Androglobin's role in Spermatogenesis.*

Upon its discovery in 2012, Adgb mRNA expression was measured in mouse tissue using RT-PCR. Expression was found to be highly abundant in the testes, with a 10-fold lower expression in the lungs and tissues. Analysis of testes development revealed an elevation of Adgb concentration during the production of postmeiotic spermatids, which continued into adulthood of the mice. Therefore, expression of Adgb is highly linked to the late stages of spermatogenesis and germ cell development (Hoogewijs et al., 2012). The key stages of spermatogenesis are summarised in Figure 1.8.1.1. Spermatogenesis is a process that involves the differentiation of spermatogonial stem cells into spermatocytes via mitosis. This produces haploid spermatids that endure in meiosis to form spermatozoa (sperm cells) and occurs in the seminiferous tubules of the testes (Nishimura and L'Hernault, 2017). The Adgb expression levels were also found to be greater in fertile mice compared to infertile mice, indicating Adgb's importance during spermatogenesis, as well as stimulating new biomedical implications to male infertility. Moreover, spermatogenesis requires oxidative energy metabolism that incorporates vast oxygen consumption for sperm formation and the generation of free radicals. This is toxic to both sperm and hormones associated with the testes. The seminiferous tubules are avascular therefore, oxygen is supplied via diffusion which can allow the organ to experience hypoxia easily. In this environment, globins usually sense oxygen or detoxify the ROS, by overexpression in these circumstances (Hoogewijs et al., 2012). NO also regulates the motility of sperm. For instance, low and extremely high NO concentrations decrease sperm motility, which contributes to male infertility. Hence, NO concentration determines sperm motility and viability. NO

can directly influence the motility of sperm by inhibiting mitochondrial respiration, which causes the deterrence of electron transfer during respiration. The concentration of NO is also dependent on the amount of NO that can be produced (usually by Nitric oxide synthases). With an abundant concentration of Adgb in the testis and possible redox-regulated functions, similar to Ngb, Adgb likely plays a role in spermatogenesis (Rosselli et al., 1998)

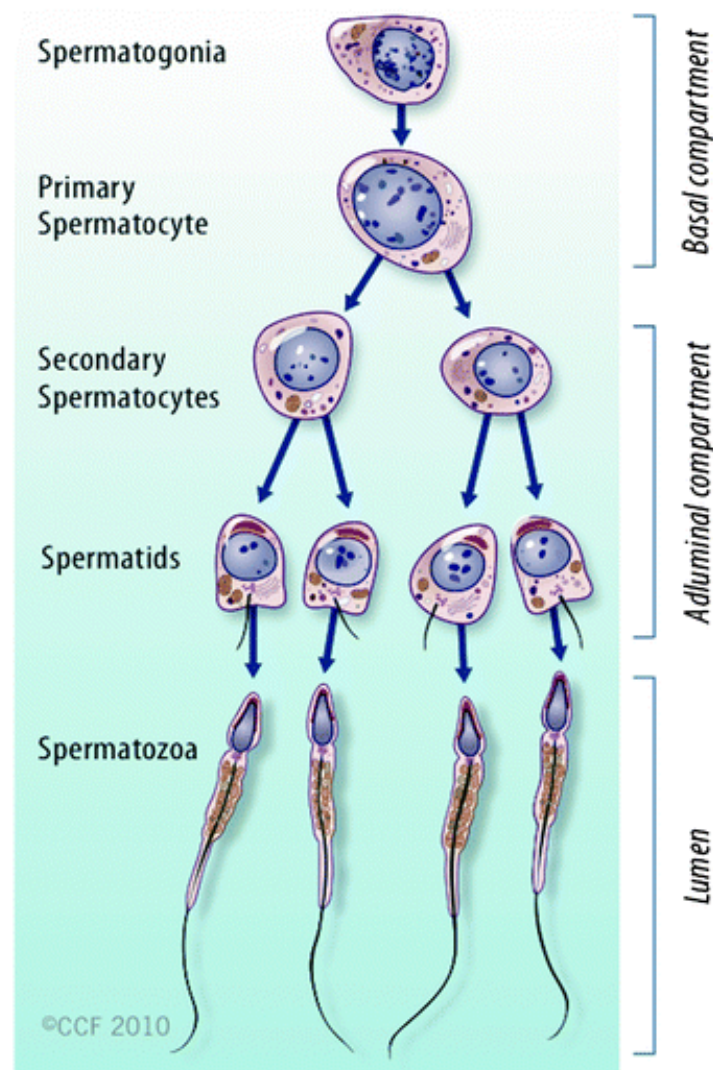


**Figure 1.8.1.1. A schematic overview of major events in spermatogenesis.** The initial stage of spermatogenesis is the division of diploid spermatogonia into primary spermatocytes via mitosis. The spermatocytes located in the basal compartment in Sertoli cells undergo meiosis to produce haploid spermatocytes in the adluminal compartment of Sertoli cells. This sub-process is called spermatocytogenesis. Subsequently, spermatids elongate to form spermatozoa by a morphological development phase called spermiogenesis. Nuclear transformations that remodel and compact chromatin occur during spermiogenesis. Spermatozoa leave the Sertoli cells through the lumen of the seminiferous tubules to the testis for final maturation. At this juncture, spermatozoa acquire motility and acrosomal function (Sharma and Agarwal, 2011).

## 1.9 Aims and Objectives

Since its discovery in 2012 by *Hoogewijs et al*, there is little research surrounding Adgb pertaining to its structural and functional properties, mainly due to caveats associated with generating the full-length protein by recombinant techniques and instabilities caused by the circularly permuted haem domain (Bracke et al., 2018). Generating a stable purified form of the protein and haem domain by recombinant techniques would permit further biochemical characterisation such as confirming the coordination of the haem domain which until now has only been done on the cell lysate Adgb (Hoogewijs et al., 2012). Additionally, the purified haem domain will be used to determine whether the protein can bind to NO comparably to other globin members. What also remains absent from the current literature is a 3D model structure of the haem domain, which could elucidate as to whether the protein shares similar features to other globin members such as containing surface-exposed Cys residues like observed in Ngb and Cygb something, that could be experimentally tested on the purified haem domain.

Furthermore, the initial study by Hoogewijs *et al* proposed a structure for androglobin (the Zurich model, Appendix 10), created using 34 Adgb sequences since then a considerable amount more Adgb sequences have been made available from a range of different species, which helped to provide a new consensus sequence, formulated by Dr Reeder and Professor Reynolds (Appendix 11, the Essex model sequence) Adgb (Hoogewijs et al., 2012). In this consensus sequence helices, A-C (at the C-terminus) are segregated from helices D-H (at the N-terminus) by the IQ-calmodulin binding motif whereby the sequence shift provides a new CD1 candidate, Tyr975 (personal communication with Dr Reeder and Professor Reynolds, unpublished work). Therefore, using the Essex model it can be hypothesised that the Tyr 975 residue could be the CD1 residue as opposed to Phe 770 in the Zurich model. Re-examining the sequence homology of key amino acids using all Adgb sequences available will determine the critical amino acids of the haem domain; the proximal and distal haem residue, and the CD1 residue.

Therefore, the main objectives of the project are as following:

1. Re-examine the critical amino acids of Adgb-HD to confirm the proximal and distal haem residue, and CD1 residue using multiple sequence alignment and all available Adgb sequences (Chapter 3).
2. Construct a 3D model structure for the Essex model sequence to visualise its characteristic globin fold despite being circularly permuted and provide insight into the haem pocket and potential surface-exposed Cys residues (Chapter 3).
3. Produce a stable purified form of the haem domain to confirm androglobin's haem coordination, measure the effect of distal mutations and Adgb's nitrite reductase activity and NO binding by spectroscopy, and experimentally identify the surface exposed Cys residues hypothesised by 3D modelling (Chapter 4).
4. Generate the full-length androglobin protein by recombinant techniques and attempt to express the full-length protein for further structural and functional characterisation (Chapter 4).

## Chapter 2. Materials and Methods

### 2.1 Materials

Kanamycin sulfate, sodium nitrite, Imidazole, sodium tetraborate, TCEP (Tris(2- carboxyethyl)phosphine), 4,4'-dithiodipyridine, TEMED (N,N,N',N'-Tetramethyl ethylenediamine), bromophenol blue sodium salt, dimethyl sulfoxide (DMSO) and, 30 % Acrylamide were purchased from Sigma-Aldrich, UK. Isopropyl beta-D-1 thiogalactopyranoside (IPTG), 5-aminolaevulinic acid (ALA) and, Ampicillin were from Molekula, UK. Ferric citrate, agarose, sodium dodecyl sulfate (SDS), 10% Ammonium persulfate (APS), ethanol, hydrochloric acid (HCl), Tris-base, glacial acetic acid, sodium dithionite, glycine, glycerol and, sodium hydroxide (NaOH) was purchased from Fisher, UK. Luria-Bertani (LB) media was from Melford Laboratories Ltd, UK. Carbon monoxide was purchased from BOC, UK. Dithiothreitol-Clelands Reagent (DTT) was from Fluka. The PROLI NONOate was from Cayman chemicals and the ethidium bromide was from Severn Biotech, UK.

### 2.2 Molecular cloning

#### *2.2.1 Primer Design*

Primers were designed to truncate the full androglobin gene sequence for effective expression in competent *E.coli* cells. Primers were designed using the PrimerX website (<https://www.bioinformatics.org/primerx/>) for DNA amplification by PCR (polymerase chain reaction). The DNA sequence for the full androglobin protein was obtained from the Uniprot database (<https://www.uniprot.org/>) and was used as a DNA template in PCR. The primer sequences were synthesised by Eurofins Genomics. The primer's physical properties were generated by the synthesis report for the oligonucleotides (Table 2.2.1.1). Primers were designed to truncate the full-length protein after the haem domain (Adgb-HD-F/R primers) as well as, 200 aa past the haem domain (Adgb-HD200-F/R primer) (Table 2.2.1.1) by the addition of a TAA stop codon. The application of mutations to the Adgb-HD sequence are shown in Appendices 8 and 9.

**Table 2.2.1.1. Properties of forward (F) and reverse (R) primers for the truncation of the full androglobin gene sequence by PCR.** An adenine (A) or Thymine (T) nucleotide mutation (bold and underlined) was substituted/inserted into the DNA sequence of Adgb for specific truncation after the HD (Adgb-HD-F/R primers) and 200aa after the HD (Adgb-HD200-F/R primer).

| Oligo Name   | DNA Sequence (5'→3')                           | Length | Tm (°C) | GC (%) |
|--------------|------------------------------------------------|--------|---------|--------|
| Adgb HD F    | GCTTTCGCGGACTA <u><b>A</b></u> ACAGTCACGTACC   | 28     | 68      | 53.6   |
| Adgb HD R    | GGTACGTGACTGT <u><b>T</b></u> TAGTCCGCGAAAGC   | 28     | 68      | 53.6   |
| Adgb HD200 F | CCATTCTGCTTCCT <u><b>T</b></u> AAAAAAGAGCAGGAG | 28     | 65.1    | 46.4   |
| Adgb HD200 R | CTCCTGCTCTTTTT <u><b>A</b></u> AGGAAGCAGAATGG  | 28     | 65.1    | 46.4   |

### *2.2.2 Polymerase chain reaction*

PCR was used to truncate the full Adgb gene in the pET28a vector using the primers designed in Table 2.2.1.1 and Pfu Turbo DNA polymerase (Agilent). The primers were used to apply a stop codon (TAA) after the HD and 200 aa after HD by PCR. The pET28a-full Adgb plasmid was used as a template in the PCR. In each reaction, 50 ng of DNA template and 125 ng of forward and reverse primers were used. The final volume was made to 50  $\mu$ l for each reaction. PCR reactions for each forward and reverse primer set were made in duplicate, with/without 10% DMSO to prevent secondary structures of the DNA template from forming and thus to improve PCR yield. The PCR reaction conditions are shown in Table 2.2.2.1 and PCR was completed in a Thermocycler (Barloworld Scientific).

After the PCR, the reaction mixtures were treated with DpnI restriction enzyme (Thermo Scientific) for cleavage of methylated recognition sites in the DNA template sequence, so that gene transcription is not repressed. Thus producing an unmethylated PCR product containing the desired mutation (Table 2.2.2.2).

**Table 2.2.2.1. The PCR conditions for each stage of the PCR reaction 1-4.** The steps include initial denaturing, denaturing, annealing, elongation and, final elongation.

| Stage                | Temperature (°C) | Time   | Number of cycles |
|----------------------|------------------|--------|------------------|
| Initial denaturation | 95               | 2 min  | 1                |
| Denaturing           | 95               | 50 s   | 20               |
| Annealing            | 55               | 1 min  |                  |
| Elongation           | 68               | 13 min |                  |
| Final Elongation     | 68               | 13 min | 1                |

**Table 2.2.2.2. The volumes, concentrations and, reagents for the restriction enzyme digest of the 50 µl PCR reaction mixtures using DpnI.**

| Reaction components | PCR reaction mixture (µl) |
|---------------------|---------------------------|
| DpnI RE             | 1                         |
| DpnI RE buffer      | 5                         |
| Total               | 56                        |

Samples were incubated at 37 °C for 2 h however, after the initial 30 min a 5 µl aliquot was taken from each PCR mixture and stored at -20 °C.

### 2.2.3 Transformation of plasmid DNA into XL1B or BL21 DE3 *E. coli* cells.

For each transformation, 2  $\mu\text{l}$  of plasmid DNA ( $\sim 30\text{-}150\text{ ng }\mu\text{l}^{-1}$ ) or ligation mixture was incorporated into 100  $\mu\text{l}$  of thawed *E. coli* competent cells (BL21 DE3 expression cells from Sigma-Aldrich or XL1B cloning cells from Agilent Technologies). The cells were incubated on ice for approximately 30 min and then immediately heat-shocked at 42 °C for 45 s in a water bath. They were immediately placed on ice for a further 2 min and 900  $\mu\text{l}$  of sterile LB media was added to each transformation mixture. This was incubated at 37 °C, shaking at 160 rpm for 50 min. The transformation mixture was plated onto LB agar plates containing a suitable antibiotic for colony growth and selection. All transformation mixtures were spread on to plates in a class II MSC (Microbiology Safety Cabinet) to prevent contamination. The plate was incubated O/N (overnight) at 37 °C.

For Kanamycin Sulfate antibiotic, 50 mg  $\text{ml}^{-1}$  of was added to LB media and Agar plates to a final concentration of 50  $\mu\text{g ml}^{-1}$  for pET28a vector, pET28a-HD (Haem Domain), pET28a-full length Adgb and, pET28a-truncated Adgb DNA.

For Ampicillin antibiotic, 100 mg  $\text{ml}^{-1}$  of was added to LB media and Agar plates to a final concentration of 100  $\mu\text{g ml}^{-1}$  for pcDNA 3.1 His-tagged full Adgb DNA.

### 2.2.4 Inoculation of liquid culture

Isolated colonies were selected using sterile pipette tips and placed in 10 mL LB universal bottles or 100 mL LB conical flasks containing 100 or 50  $\mu\text{g ml}^{-1}$  of ampicillin or kanamycin sulfate, respectively. Cultures were incubated O/N at 37 °C, shaking at 100 rpm.

### *2.2.5 DNA miniprep*

DNA was extracted from liquid cell cultures using the Thermo Scientific GeneJET Plasmid Miniprep Kit. The cultures were centrifuged for 10 min at 4000 rpm. The protocol was followed as per the manufacturer's instructions ([Plasmid Miniprep protocol](#)).

### *2.2.6 Determination of DNA concentration*

The concentration of the eluted DNA was measured on the Thermo Scientific NanoDrop 2000c UV-Vis Spectrophotometer. Prior to measuring, 2 µl of the appropriate buffer was used as a blanking solution to calibrate the machine. DNA samples (~2 µl) were measured at 260 nm 3 times to obtain an average DNA concentration. As for cell cultures, 1 ml of cell culture in cuvettes were used to obtain the optical density at 600 nm.

### *2.2.7 Restriction enzyme digests of DNA*

Several restriction enzyme digestions of DNA were performed throughout this project. Initially, a double and triple digest was completed on the pET28a vector and the pcDNA 3.1 (+) His-tagged full Adgb plasmids, respectively to prepare plasmids for ligation. The linearized pET28a vector was dephosphorylated by an alkaline phosphatase digestion reaction. After the ligation of the vector and insert (section: 2.2.8), a single digest was completed on the pET28a full Adgb recombinant plasmid for verification (All restriction enzyme digestion reaction tables are located in Appendix 3).

### *2.2.8 DNA ligations.*

The digested His-tagged full Adgb from the pcDNA 3.1 (+) vector was inserted into the dephosphorylated pET28a vector at a 2:1 insert to vector molar ratio. The ligation was completed using T4 DNA ligase (Thermo Scientific). Two reactions were set up for the ligation so that one could be incubated at 22 °C for

1 h and the other was incubated at 16 °C O/N. The ligation reaction mixtures are shown in Table 2.2.8.1. The ligated plasmids were transformed into XL1B *E.coli* cells (section: 2.2.3) and spread onto kanamycin sulfate (50 µg ml<sup>-1</sup>) plates. The plates were incubated O/N at 37 °C. A single colony was selected and subjected to liquid inoculation (section: 2.2.4). The ligation product was purified by DNA Miniprep (section: 2.2.5).

**Table 2.2.8.1. Volume and Concentrations of the reagents included in the ligation reactions.**

| Reaction Components           | 2:1 reaction (μl) |
|-------------------------------|-------------------|
| Linear Vector DNA (100 ng)    | 2                 |
| Insert (200 ng)               | 2.3               |
| 10X T4 DNA ligase buffer      | 2                 |
| T4 DNA ligase (1000 units/ml) | 1                 |
| H <sub>2</sub> O              | 12.7              |
| Total                         | 20                |

### *2.2.9 Agarose gel electrophoresis*

DNA fragments were purified on a 1% agarose gel. Agarose gels were prepared by dissolving 0.3 g of 3% Agarose in 30 ml of 1x TAE buffer. DNA samples were stained in 6X TriTrack DNA Loading Dye (contains tracking dyes, Xylene Cyanol FF, Bromophenol Blue, and Orange G). The stained samples were loaded adjacent to 1 Kb Generuler (ladder and loading dye from ThermoFisher Scientific) to measure the size of DNA bands produced (Appendix 1). Electrophoresis was performed at 80 V (~50 mA) for 45 min. Gels were immediately stained for an hour in  $10 \text{ mg m}^{-1}$  Ethidium Bromide for optimal fluorescence of DNA. DNA band sizes were visualised under UV lighting and gel images were captured using a mobile camera phone.

### *2.2.10 Gel extraction*

DNA was extracted from agarose gels using the QIAquick Gel Extraction Kit from ThermoFisher. DNA bands visualised under UV light were exercised using a sharp scalpel and DNA extraction was followed as per the manufacturer's instructions ([Gel extraction protocol](#)).

### *2.2.11 Sanger sequencing*

DNA samples ( $30 \text{ }\mu\text{l}$  of  $30\text{-}100 \text{ ng }\mu\text{l}^{-1}$ ) were sequenced by Eurofins for confirmation of the full Adgb gene present in the pET28a vector and to check that all mutations to truncate Adgb-HD had been successful. A sequencing primer (TGGGCCGTTCTGGAGATGAACTTG) was synthesised by Eurofins. Sequencing was also used to confirm the successful ligation of the full Adgb gene in the pET28a vector with no contamination or undesired mutations (using a standard primer pBR3- TCCCCATCGGTGATGTC). The DNA sequence results were viewed using SnapGene Viewer software.

## 2.3 Protein synthesis

### *2.3.1 Androglobin haem domain protein expression and purification*

The pET28a vector containing the haem domain, full/truncated Adgb, haem domain with distal mutations were expressed in BL21 DE3 cells by transformation using the heat shock method (section: 2.2.3). Cell cultures were prepared as in section: 2.2.4 and were incubated O/N at 100 rpm at 37 °C with kanamycin sulfate. Protein synthesis was induced when the optical density of cells was between 0.8-1 at 600 nm, by adding 500  $\mu$ M of 200 mM IPTG, 250  $\mu$ M of 100 mM ALA and 50  $\mu$ M of 100 mM ferric citrate. Ferric citrate and ALA were used to facilitate haem production. After, CO was bubbled through each flask for 30 s and sealed. CO was utilised for the stabilisation of synthesised protein as most globins can autoxidise to form ferric (met) haem iron plus superoxide. CO ligated ferrous holoprotein is stabilised during expression and does not autoxidise (Beckerson et al., 2015). The flasks were incubated O/N (approx. 18 h) at 100 rpm at 37 °C. Cells were harvested by centrifuging at 4000 g at 4 °C for 25 min. The supernatant was discarded and cells were resuspended in water and frozen at -20 °C. The wet pellet was thawed and lysed using the Avestin Emulsiflex C3 homogenizer at 15000-20000 psi. The pellet was subjected to 3 passes and centrifuged at 20,000 g at 4 °C for 25 min. The supernatant was applied to an immobilised Nickel ( $\text{Ni}^{2+}$ ) affinity column (5 ml IMAC, GE Healthcare). The column was treated with 20 mM sodium phosphate buffer containing 500 mM NaCl and 20 mM imidazole pH 7.4. At the same time, the pellet was treated with 0.5 % SDS and 2 M imidazole and centrifuged (20,000 g, 20 min at 4 °C). The bound protein was eluted using 20 mM sodium phosphate buffer containing 500 mM NaCl and 500 mM imidazole. The imidazole was removed via dialysis using 1 mM borax buffer pH 10, where the buffer solution was changed three times over 24 h.

Due to haem loss during the purification process, the elution was treated with 5 mM Hemin dissolved in 10 mM NaOH and 1.1x excess was added based on the extinction coefficient of the protein at 280 nm ( $\epsilon_{280\text{nm}}=32.5 \text{ mM}^{-1} \text{ cm}^{-1}$ ). The extinction coefficient was calculated using the ProtParam tool on the Expasy website for determination of physical and chemical parameters (<https://web.expasy.org/protparam/>). This was incubated O/N at 4 °C to allow the binding of the haem group. The solution was then applied to the

Sephadex G25 5x1 cm column (gel filtration) using 100 mM sodium phosphate buffer pH 7.4. to remove excess haem. The protein was then concentrated using an Amicon Ultra-14 Centrifugal Filter Column with a 3 kDa molecular weight cut-off and re-applied to the Sephadex G25 column. The final protein concentration of the purified deoxyferrous protein was determined using the Agilent Cary 50 UV-Visible spectrophotometer at 426 nm ( $\epsilon_{426\text{nm}}=108 \text{ mM}^{-1} \text{ cm}^{-1}$ ). The extinction coefficient was previously determined by comparing optical spectra with measurement of the haem content by HPLC (high-performance liquid chromatography) and hemochromogen assay (personal communication with Dr Brandon Reeder to whom determined the extinction coefficient of Adgb).

### 2.3.2 SDS PAGE analysis of Androglobin protein

Proteins were subjected to SDS-PAGE gel electrophoresis where 8 % and 12 % acrylamide gels were utilised. Cracking buffer was used as a loading buffer for samples and the components of the resolving and stacking gel are shown in Table 2.3.2.1. Protein was dissolved in cracking buffer and then boiled at 95 °C for 10 min prior to loading. Samples were loaded adjacent to 5  $\mu\text{l}$  of Thermofisher Pageruler Prestained protein ladder (10-180 kDa, see Appendix 2). Gels were placed in a tank containing running buffer at 140 mV until stained bands had reached the end of the gel. The gel was then stained in Coomassie blue stain for an hour and de-stained in buffer O/N. All components used in the gel and buffers are shown in Table 2.3.2.1 and gel images were captured using a mobile phone camera.

**Table 2.3.2.1. Reagents used in the preparation of gel and buffers for SDS-PAGE gel electrophoresis.**

| Reagent         | Components                                                                                                                                                           |
|-----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Resolving gel   | 1.5 M Tris/HCl at pH 8.8, 10 % SDS, 30 % Acrylamide/Bis-acrylamide, 10 % APS, TEMED and, dH <sub>2</sub> O                                                           |
| Stacking gel    | 0.5 M Tris/HCl at pH 6.8, 10 % SDS, 30 % Acrylamide/Bis-acrylamide, 10 % APS, TEMED and, dH <sub>2</sub> O                                                           |
| Cracking buffer | 1 mL 1 mM Tris/HCl at pH 6.8, 2 mL 10 % SDS, 5 mL Glycerol, 10 mg 0.05% Bromophenol Blue, 11.8 mL dH <sub>2</sub> O and a few DTT crystals before adding to samples. |
| Running buffer  | 3 g Tris base, 14.45 g Glycine, 0.5 g of SDS dissolved in 1 L of dH <sub>2</sub> O                                                                                   |
| De-stain buffer | 7.5 mL Glacial Acetic acid, 100 mL Ethanol added to 1000 mL dH <sub>2</sub> O.                                                                                       |

When preparing gels, TEMED and APS were added last to induce polymerisation of the gel.

## 2.4 Determination of Androglobin haem domain disulfide bonds

Free cysteine sulfhydryls were determined using a dithiodipyridine assay. To reduce any disulfide bonds, 5 mM TCEP was added to 5 mM Adgb and mutants. Samples were incubated for 1 h at room temperature. TCEP was removed using a 5x1 cm Sephadex G25 column and the protein was eluted from the column using 100 mM sodium phosphate buffer pH 7.4. The protein was titrated with ~0.5  $\mu$ M 4,4'-dithiodipyridine in 5  $\mu$ l increments and the spectrum was measured on the Agilent Cary 50 UV-Visible spectrophotometer. A titration curve was produced using the spectra at 324 nm and fitted to a fractional saturation curve (as described in Equation 2.4.1) using the least-squares and solver method in Microsoft Excel. Where  $[P_T]$  is the total protein concentration,  $[S]$  is the total ligand concentration, and  $K$  is the dissociation constant (Beckerson et al., 2015).

$$y = \frac{([P_T] + [S] + K) - \sqrt{([P_T] + [S] + K)^2 - 4([P_T][S])}}{2[P_T]} \quad (\text{Equation 2.4.1})$$

## 2.5 Nitrite reductase and Nitric oxide binding analysis

The distal mutants (Q791H/G) and Adgb-HD (5  $\mu$ M) were reacted with sodium nitrite at varying concentrations (100  $\mu$ M - 1 mM) in 100 mM sodium phosphate buffer pH 7.4. A stock solution of 1 M sodium nitrite was diluted with 100 mM sodium phosphate buffer pH 7.4 to produce the aforementioned final concentrations. The sodium nitrite working stocks were then added to the protein in aliquots of 1.4 ml in independent experiments. Sodium dithionite (~5 mM) was added separately to the nitrite working stock solution and protein before analysis, which were mixed together in the cuvette and the optical spectra was recorded immediately following this. The optical spectra were recorded using an Agilent 8453 diode array spectrophotometer fitted with a multi-cell carriage and temperature was maintained using a water bath at 25 °C. This provided time courses, from which rate constants were determined at (390 nm - 428 nm) by using the least-squares and solver method using Microsoft Excel.

Optical changes of NO binding were also analysed on an Agilent Cary 50 UV-Visible spectrophotometer for the distal mutants and Adgb-HD (3 ml aliquots of 5  $\mu$ M protein in 100 mM sodium phosphate buffer pH 7.4) at different oxidation states. Samples were pre-treated with ~5 mM sodium dithionite for ~1 min for ferrous state analysis and absorbance measurements were taken, 40  $\mu$ M of PROLI NONOate was prepared in 100 mM sodium phosphate buffer pH 7.4 to a concentration of 80  $\mu$ M NO, then was added to the ferrous protein solution and NO binding absorbance was monitored.

## 2.6 Sequence Homology Analysis

All mammalian Adgb amino acid sequences were obtained (in the FASTA format) from the NCBI website (<https://www.ncbi.nlm.nih.gov/>) and were aligned using ClustalWS multiple sequence alignment on the Jalview software (V.2.11.1.0). As well as this, four key amino acids were analysed using pairwise sequence analysis. The human Adgb proximal histidine (His824), the distal glutamine (Gln792), the CD1 phenylalanine (Phe770) and, the predicted CD tyrosine residues were compared to 362 mammalian Adgb sequences and a single data series was created using Microsoft Excel to present links of common residue patterns to a specific clade/species. To further represent this, amino acid charts were created using the WebLogo-Berkeley (V.2.8.2) website (<https://weblogo.berkeley.edu>). A phylogenetic tree was also created to compare the evolutionary relationship between Adgb with other mammalian globin members using the MEGA software (V.10.1.8).

## 2.7 Molecular modelling

The 3D tertiary structure of the globin domain for Adgb was predicted by submitting sequences (Table 2.7.1) to the [I-TASSER](#) and [PHYRE2](#) protein fold recognition server to attain templates with the highest homology to the sequences in Table 2.7.1. The templates were then assembled in PyMOL (V.2.3.5) as well as structurally compared to human Ngb in DeepView (V.1.6). The 3D structure of Ngb was extracted from PDB (protein data bank) using its PDB code, 4MPM. Two models were produced and derived from the Essex and Zurich model sequences (Appendices 12 and 13). Ngb was also used to extract and apply a haem group to both structures.

**Table 2.7.1. Amino acid sequences for the globin domain of Androglobin for 3D modelling.** The amino acid positions of the sequences are shown in bold.

| Sequence Name                                  | Amino Acid Sequence                                                                                                                                                                                                                                      |
|------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| C/D to H helix                                 | <b>781'</b><br>NFEPESCRFTEQSLLIMKAIGNVIANFKDKGKLSAALKDLQTAHYPVPFH<br>DKELTAQHFRVFHLSLWRLMKKVQITKPPPNFKFAFRAMVLDLELLNSS<br>LEEVS LV <b>-886'</b>                                                                                                          |
| A to B helix                                   | <b>936'</b><br>NISVADTLQKVWAVLEMNLEQYAVSLLRLMFKSKCKSLESYPCYQDEE<br>TK <b>-985</b>                                                                                                                                                                        |
| A to C helix                                   | <b>936'</b><br>NISVADTLQKVWAVLEMNLEQYAVSLLRLMFKSHICSMVSFVIGDEHV<br>LP <b>-780'</b>                                                                                                                                                                       |
| A to B helix (including the IQ-binding motif). | <b>781'</b><br>NFEPESCRFTEQSLLIMKAIGNVIANFKDKGKLSAALKDLQTAHYPVPFH<br>DKELTAQHFRVFHLSLWRLMKKVQITKPPPNFKFAFRAMVLDLELLNSS<br>LEEVS LVEWLDVKYCMPTSDKEYSAEEVAAAIIQAMWRGTYVRLLMK<br>ARIPDTKENISVADTLQKVWAVLEMNLEQYAVSLLRLMFKSKCKSLESY<br>PCYQDEETK <b>-985</b> |
| Calmodulin IQ-binding motif                    | <b>891'</b> VKYCMPTSDKEYSAEEVAAAIIQAMWRGTYVRLLMKARIPDTKE<br><b>-935</b>                                                                                                                                                                                  |

## Chapter 3. Bioinformatic analysis and 3D modelling of Androglobin haem domain

The sequence homology of key amino acids of the haem domain in other species will be examined to determine the critical amino acids located in the haem domain of Adgb. Phylogenetic analysis will also be conducted to demonstrate the phylogenetic relationship between Adgb and other human globin members. In addition, a 3D prediction model of the haem domain of Adgb derived from the Essex model sequence (Appendix 11) will be completed. A 3D model of the haem domain will demonstrate the characteristic globin fold; despite being circularly permuted.

### 3.1 Sequence homology analysis of key residues in the haem domain of Androglobin

The globin domain for most globins contains eight  $\alpha$ -helical segments (named A-H), which has a characteristic 3-over-3  $\alpha$ -helical sandwich structure referred to as the globin fold. The globin domain of Adgb satisfies the criteria of the globin-fold; however, it is circularly permuted and is split into two parts by an IQ-binding motif (Burmester and Hankeln, 2014). In the first and only structural study of Adgb, it was found that the C-H helices are located on the N-terminal side, and helices A-B has been shifted to the C-terminal side, as shown in Appendix 10 (Hoogewijs et al., 2012).

The original globin sequence alignment was conducted using 34 Adgb sequences from species identified as having the Adgb gene and was found to contain several highly conserved and functionally important residues; Phe770 in the CD helix region and the proximal His824 (F8), which coordinates the haem iron. Adgb also has a distal haem position that coordinates the haem iron to form a hexacoordinated haem arrangement. The distal position is a Gln792 (E7) instead of the more common distal His. These key residues are seen in most Adgb's (Appendix 7) (Hoogewijs et al., 2012). Since the original alignment by Hoogewijs et al, there are many more sequences available for the haem domain alignment of Adgb to be re-evaluated. A re-evaluation of the Adgb HD was completed and the results of this study are shown in Figure 3.1.1. The results show a more in-depth sequence analysis of the key amino acids that are highly conserved throughout the globin superfamily.

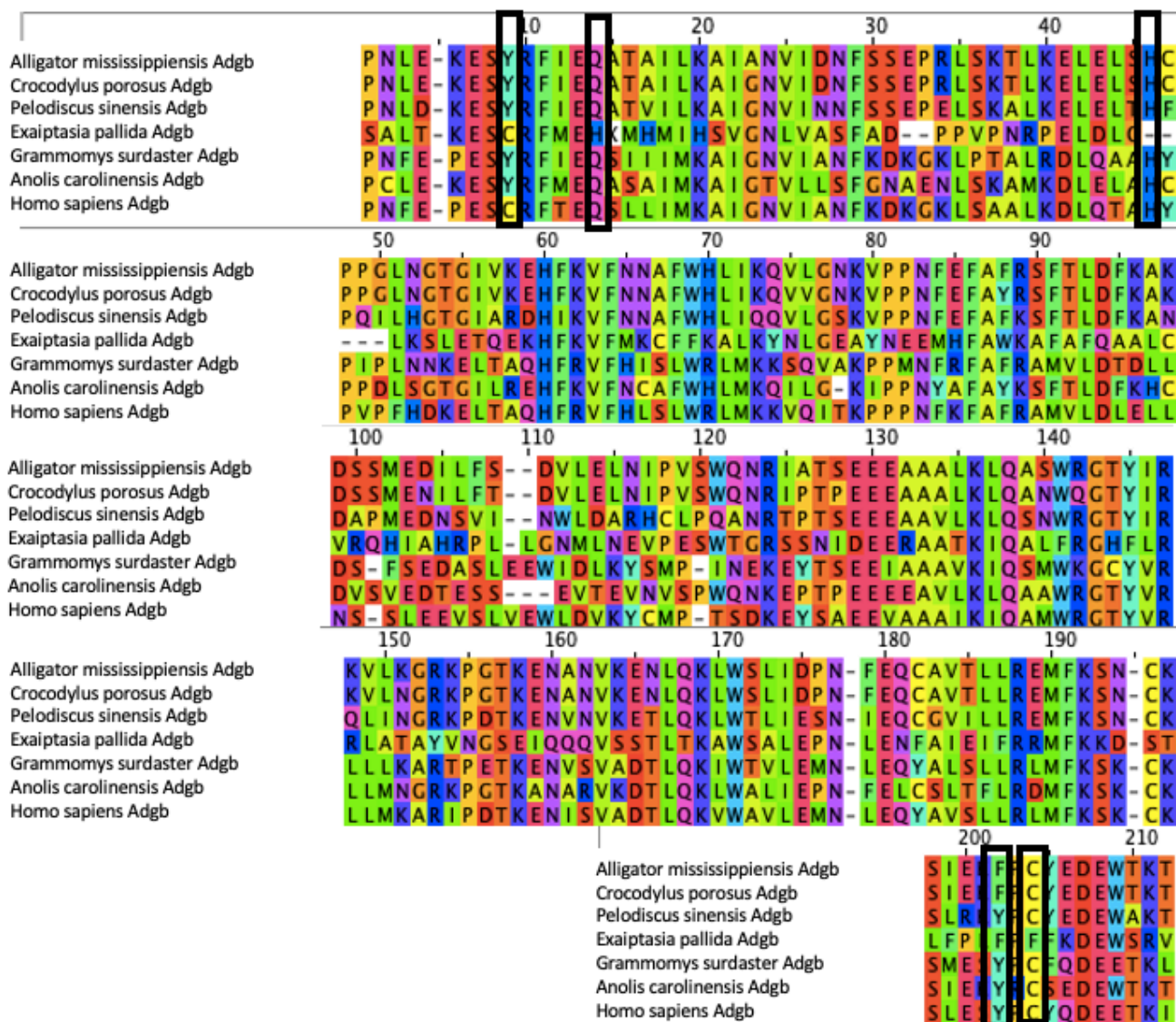


**Figure 3.1.1. Multiple sequence alignment of the Androglobin haem domain.** Adgb sequences from 362 different species were aligned using ClustalWS on Jalview software. Sequences with 5-10% consensus causing large gaps were removed. The alignment shows a selected set of specie from the complete 362 sequences. The selected sequences were used to represent all classifications animal classes that were analysed. The key amino acids that have been analysed throughout this study are indicated in black boxes, as well as the position of surface-exposed cysteines that form a disulfide bond (Chapter 4).

Figure 3.1.1 shows a multiple sequence alignment (using ClustalWS) of a selection from 362 Adgb sequences from a range of species and also displays the alignment for the HD for different species as well as the *Homo sapiens* Adgb-HD sequence. The strictly conserved proximal histidine (highlighted in Fig 3.1.1) is known to link the haem iron to the polypeptide chain in almost all globins whereas the distal haem residue in hexa-coordinated globins (His) commonly plays a major role in determining the affinity and selectivity for ligands at the 6<sup>th</sup> iron-coordination position (Avella et al., 2014). In penta-coordinated globins such as Hb and Mb, the 6<sup>th</sup> coordination site is unoccupied by the distal His to allow quick accessible binding of a ligand (Burmester and Hankeln, 2014). Other than the significant distal and proximal haem residues, the CD1 Phe in globins provides stability for the haem group and pocket (Hargrove et al., 1994). It has previously been found that mutations of the CD1 Phe can often lead to haemoglobinopathies with severe effects. For example, Hb Louisville is also known as Hb Buccuresti, is a haemoglobinopathy where a point mutation [ $\beta$ 42(CD1)Phe→Leu] alters the haem contact at F42 of  $\alpha$ 1 $\beta$ 2 dimer, causing a lower oxygen binding affinity that can result in anaemia (Kamseng et al., 2017). Another Hb variant called Hb Hammersmith [ $\beta$ 42(CD1)Phe→Ser] increases haem loss from affected subunits which cause Heinz body formation (precipitated Hb) (Akiyama et al., 2006). This demonstrates the importance of the CD1 Phe residue for haem stability in globins.

All the Adgb-HD sequences were subjected to a pairwise sequence analysis with the *Homo sapien* Adgb sequence, in an effort to show which of the key amino acids in Adgb are highly conserved throughout all species. This was aimed at establishing key amino acids in the HD of Adgb to re-evaluate and verify the findings of the original alignment. The pairwise sequence analysis was completed on Jalview where the *Homo sapien* sequence was aligned with each of the 362 sequences. From this, a single data set was formed based on the number of species obtaining identical amino acids to the human sequence. The conservation of these amino acids was displayed as an amino acid chart to represent further sequence homology between Adgb sequences. The key amino acids analysed were: proximal His, distal Gln, and the CD Phe (Figure 3.1.2).

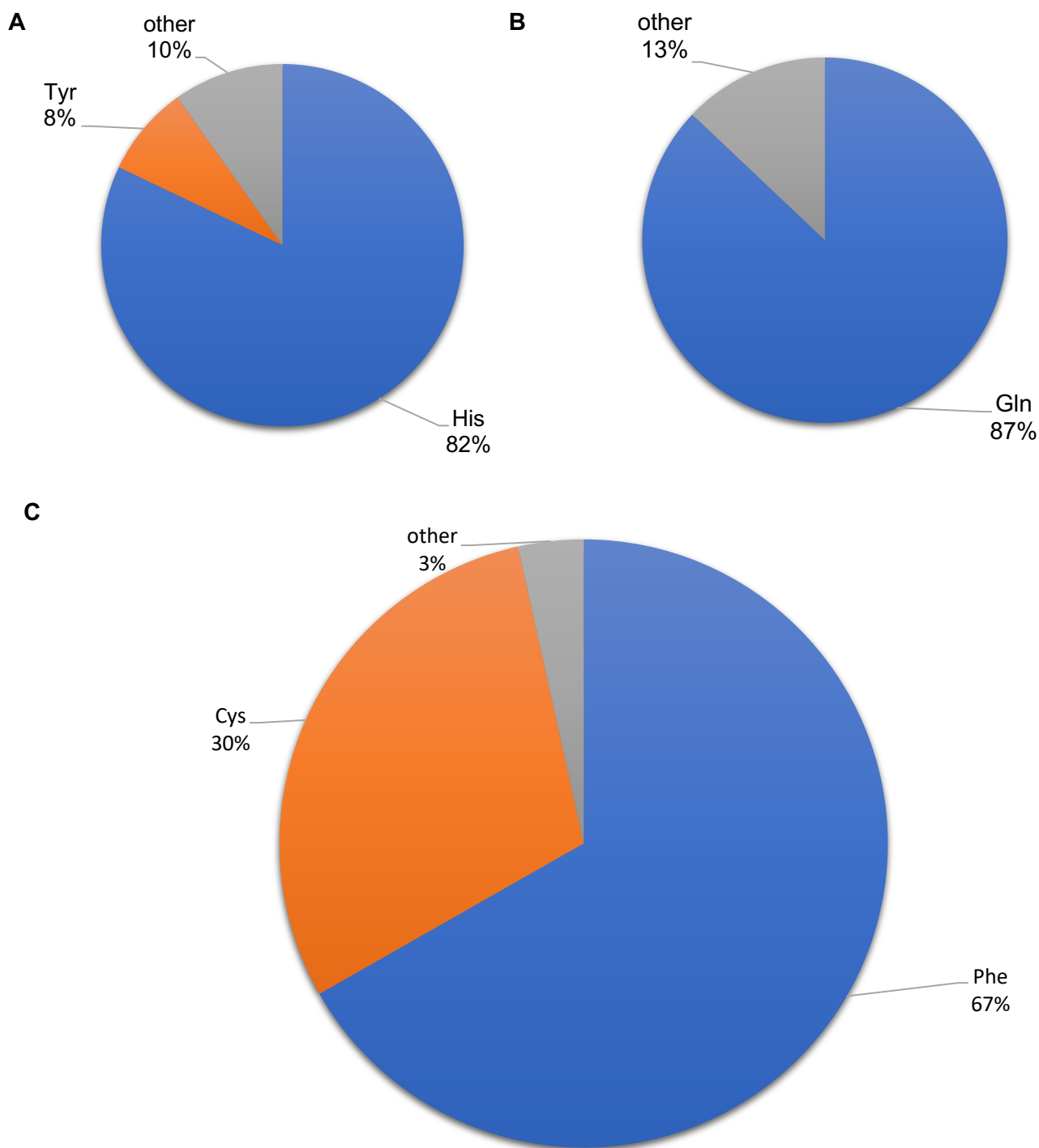


**Figure 3.1.2. Single data series representing the sequence homology of the three key amino acids in *Adgb*.** The three key amino acids that were analysed were the proximal His (**A**), the distal Gln (**B**) and, the CD Phe (**C**). The frequency for each of these key amino acids was recorded based on pairwise sequence alignments of the human *Adgb* sequence with each *Adgb* sequence from different species. The frequency was used to form a pie chart to demonstrate the conservation of residues throughout the species. The pie chart was formed using Microsoft Excel.

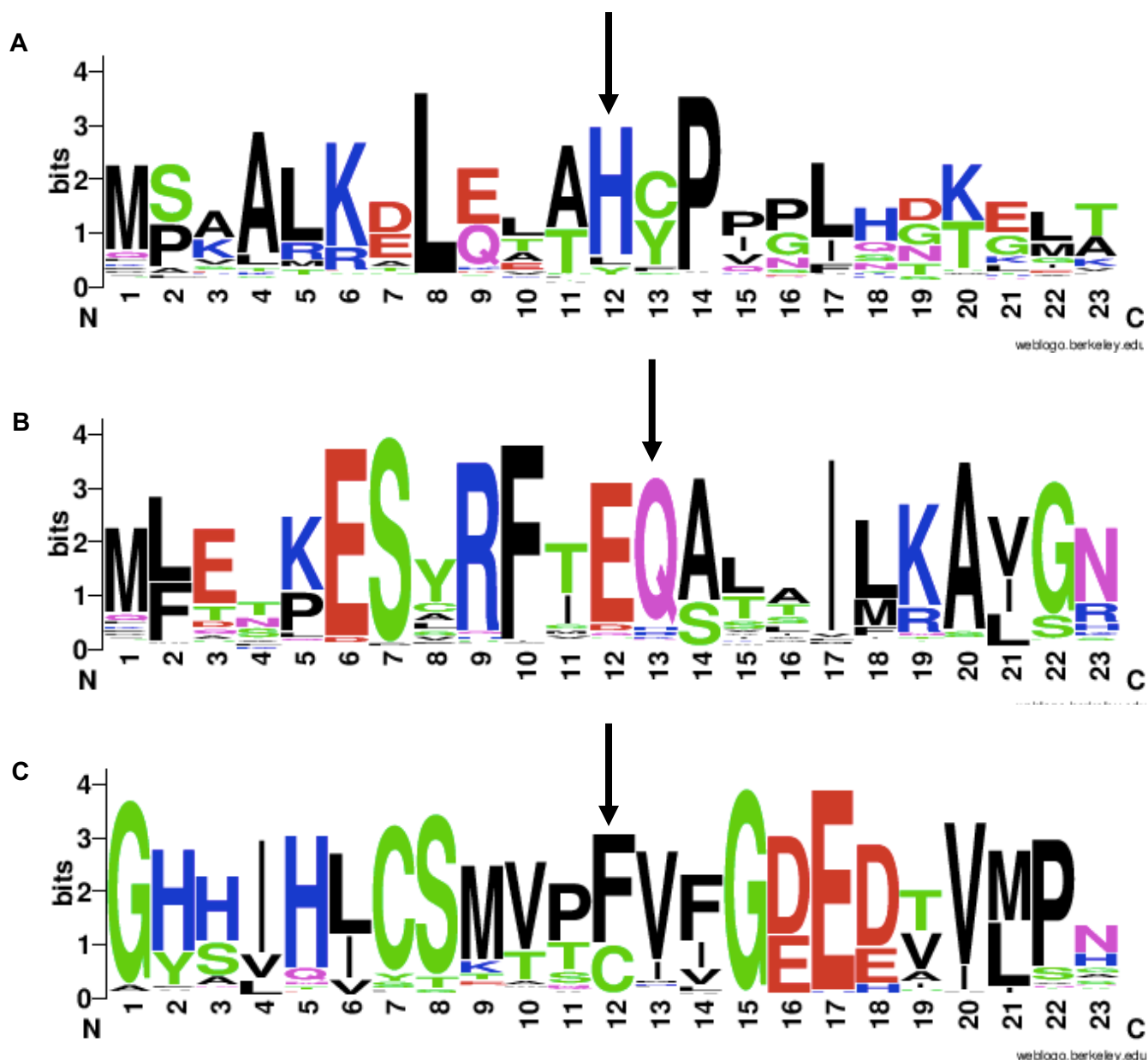


**Figure 3.1.3. Amino acid charts for the key amino acids in the haem domain of Androglobin.** A graphical representation of the key amino acid sequences from the pairwise sequence alignments. The charts illustrate the conservation for each amino acid across 362 different species for the Adgb sequence. The arrows indicate the positions for the proximal His (sequence range 780-802, F8 region) **A**, the distal Gln (sequence range 812-834, E7 region) **B**, and the CD1 Phe according to the Zurich model (sequence range 758-779, CD region) **C**, in the HD. Amino acids are coloured according to their chemical properties: polar amino acids (G, S, T, Y, C, Q, N) are green, basic (K, R, H) blue, acidic (D, E) red and hydrophobic (A, V, L, I, P, W, F, M) amino acids are black (Schneider and Stephens, 1990).

Figure 3.1.2 shows a single data series of the pairwise sequence alignment with the 362 Adgb sequences and the human Adgb sequence. From the alignments, the key amino acids: the proximal His, distal Gln and, CD Phe was analysed. The pie charts in Figure 3.1.2 show the percentage homology of each key amino acid across 362 different species. Figure 3.1.2A indicates that 82 % of the species analysed, contained a proximal His residue identical to the one observed in the human Adgb sequence. A very small amount obtaining a Tyr or Leu residue in the proximal position. Some of these may be a result of misalignment however, tyrosines are observed in the proximal position of some bacterial catalases (Murshudov et al., 1996). There was also a small percentage of other amino acids e.g. Arg and His (other amino acids make up 10 %) detected for the analysis of the distal Gln (Fig 3.1.2B), again due to the small percentages, it is likely that some of these may have resulted from sequence misalignment. However, Arg and His are both nitrogenous in comparison and are both viable distal amino acids. The majority of species (87 %) obtain a distal Gln residue, identical to the distal Gln residue in the human Adgb sequence. Figure 5.1.2C shows that only 67 % of the species contained a Phe in the nominated CD region. Interestingly, there was a significant percentage of species that contain a Cys residue (30 %) instead of a Phe. This cannot be explained through misalignment and would suggest that the CD residue may be a Cys instead of a Phe in some species, based on the assignment of this amino acid as the CD1 Phe.

Figure 3.1.3 shows an alternative graphical representation of the key amino acid sequences from the pairwise sequence alignments conducted. The overall height of the stack of amino acids indicates the sequence conservation at a certain position. Whereas the actual height of an amino acid specifies the relative frequency at that position in the sequence. Therefore, the amino acid charts provide a richer representation for the conservation of amino acids across the range of species. Figure 3.1.3A and B correspond to the results seen in Figure 3.1.2A and B. Figure 3.1.3C shows a visual depiction of the result in Figure 3.1.2C for the CD Phe. Figure 3.1.3C emphasises how Cys is commonly found in the CD position in an abundance of species (30 %) than a Phe.

The CD1 Phe is a critical amino acid in the haem pocket that maintains the stability of haem binding. The substitution of the CD Phe to a Cys has been previously reported for a natural Hb variant. Hb little Venice

[ $\beta$ 42(CD1)Phe→Cys] is a mutation that causes haemolytic anaemia and the formation of Heinz bodies (Henderson et al., 2016). Heinz body is the aggregation of denatured and precipitated Hb in erythrocytes (Tunsaringkarn et al., 2012). The substitution of the Phe to Cys removes contact with the haem leaving a gap at the surface of the pocket and results in instabilities of the haem group by changing the solubility within the haem pocket and capacity to bind to a ligand (Bartram et al., 2012). Due to the lack of research, it is not clear why having a Cys residue in the CD1 region is advantageous or why a significant amount of species contain this residue, especially if the Cys residue is associated with unstable haem as observed in Hb little Venice.

An alternative CD sequence is that of the Essex model, that through sequence alignment, the CD1 was proposed to be Tyr975 (Appendix 11). The assumption that Tyr975 is in the CD1 position in the haem domain was analysed by assessing its sequence homology and forming a single data series similar to Figure 3.1.2. In addition, an amino acid chart (like in Figure 3.1.3) was produced for consistency and further insight. This was done to see the conservation of Tyr975 across 362 species containing Adgb (Figure 3.1.4).

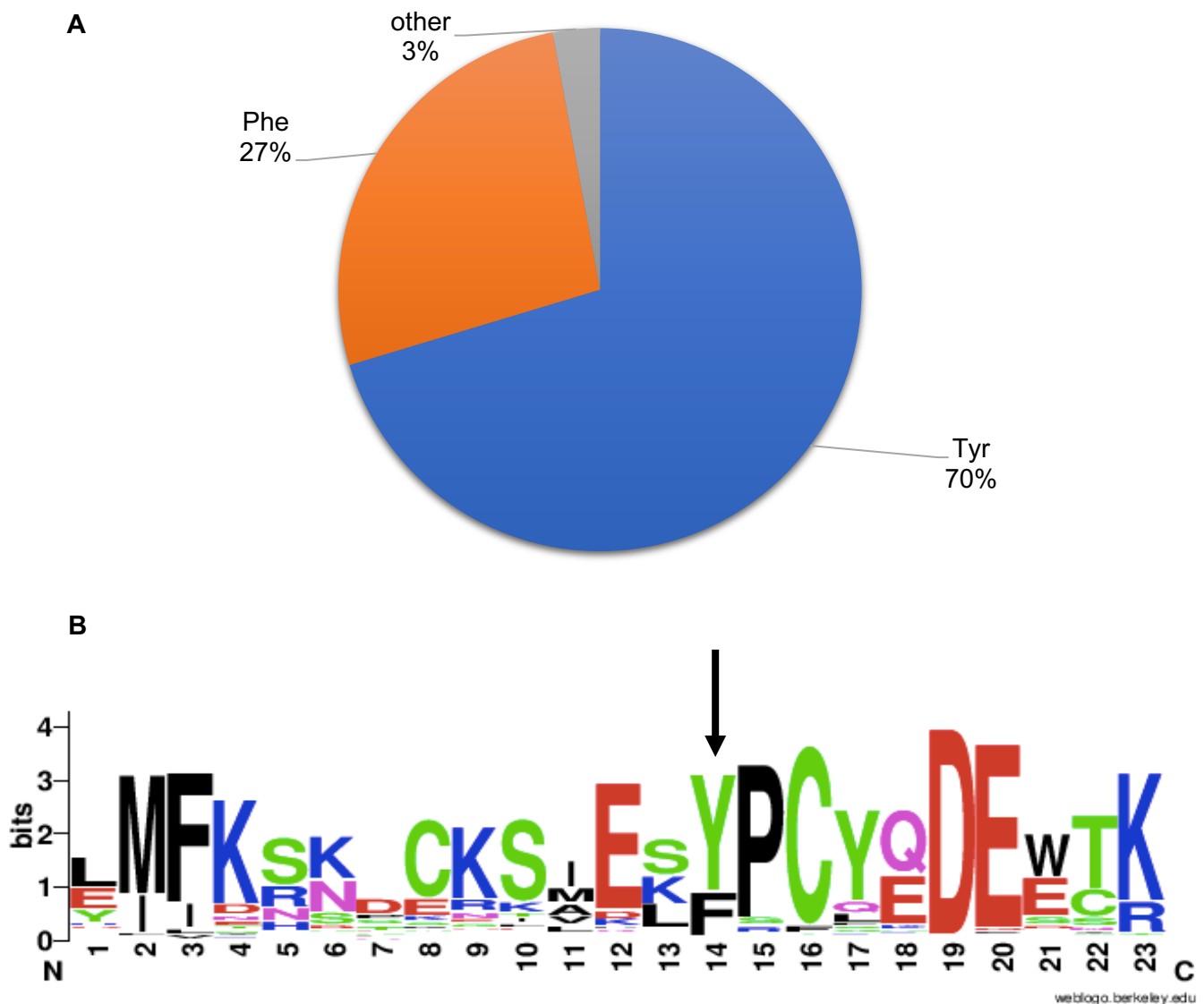


**Figure 3.1.4. Sequence homology analysis of Tyr975 in Androglobin across 362 species.** The frequency of Tyr975 was recorded based on pairwise sequence alignments of the human Adgb sequence with each Adgb sequence from different species. The frequency was used to form a pie chart (**A**) to demonstrate the conservation of Tyr across 362 different species. The pie chart was formed using Microsoft Excel. An amino acid chart (**B**) was generated to illustrate the conservation of Tyr in the CD1 position according to the Essex model (sequence range 964- 986). The black arrow indicates the position of Tyr (Y) in the sequence. The chart was developed using the Web Logo-Berkeley website. Amino acids are coloured according to their chemical properties: polar amino acids (G, S, T, Y, C, Q, N) are green, basic (K, R, H) blue, acidic (D, E) red and hydrophobic (A, V, L, I, P, W, F, M) amino acids are black (Schneider and Stephens, 1990).

Figure 3.1.4A shows a single data series of the pairwise sequence alignments for the 362 Adgb sequences with the human Adgb sequence for CD Tyr975 as well as, the homology of Tyr in the proposed CD1 position across different species. Here, 70 % of species contain a Tyr residue in the proposed CD1 position and 27 % a Phe residue, with the other 3 % as probable misalignments. Figure 3.1.4B illustrates that based on the proportion of species analysed, a Tyr (indicated by a black arrow in Fig 3.1.4B) is found with much higher frequency in sequences than Phe. Thus, using this sequence as the potential assignment of the CD region of the Adgb-HD, it is evident that most species would contain a Tyr in the CD1 position than Phe, which is more commonly found in globins in the CD1 position.

The frequency of CD1 Phe and Cys occurring instead of the CD1 Tyr975 (Essex model) and Phe770 (Zurich model) was determined to identify which CD1 residues were selective to which particular animal group. Pairwise sequence analysis was done for CD1 Tyr975 and all the matches that came up with a Phe instead of a Tyr were recorded (27 %, Fig 3.1.4A). From the 27% that was identified as having a Phe, the species were analysed. This method was also completed for CD1 Phe from the Zurich model to see if the reoccurring Cys (30 %, Fig 3.1.2C) was selective to a particular animal group. The number of Phe and Cys residues present was calculated from pairwise sequence alignments for CD1 Tyr975 and Phe770 and are shown as a bar chart in Figure 3.1.5.

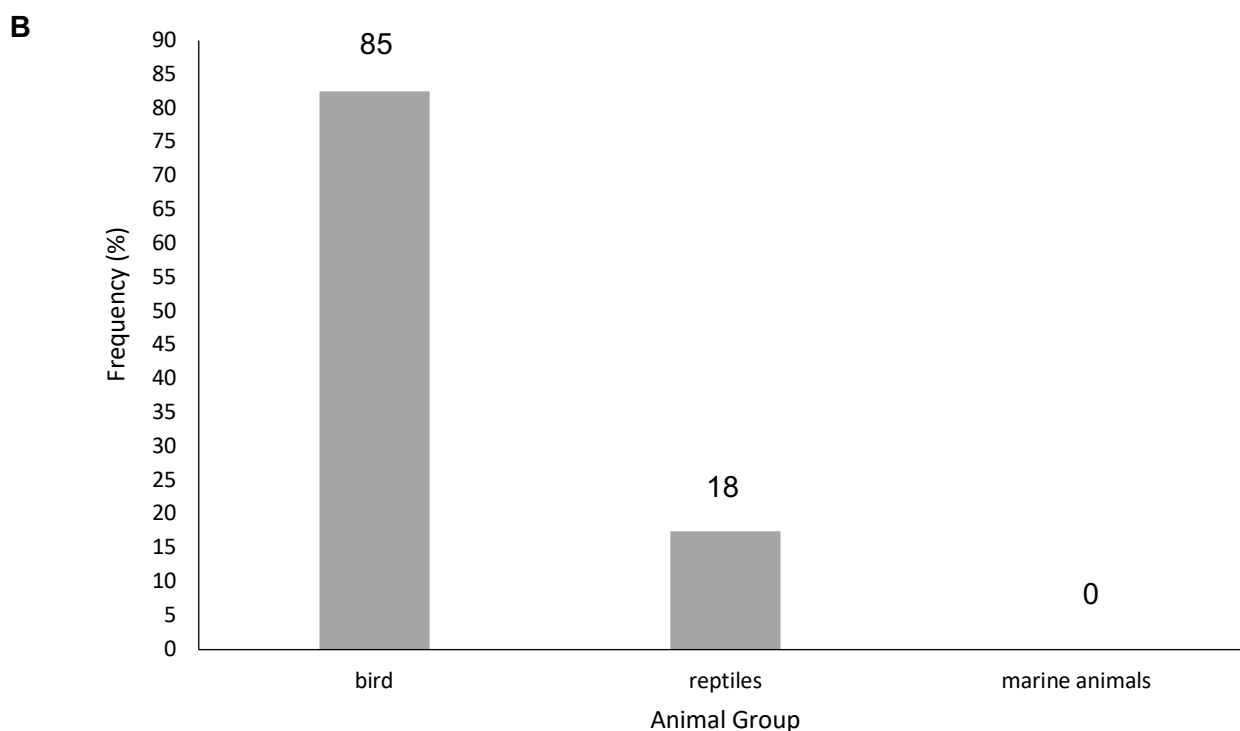
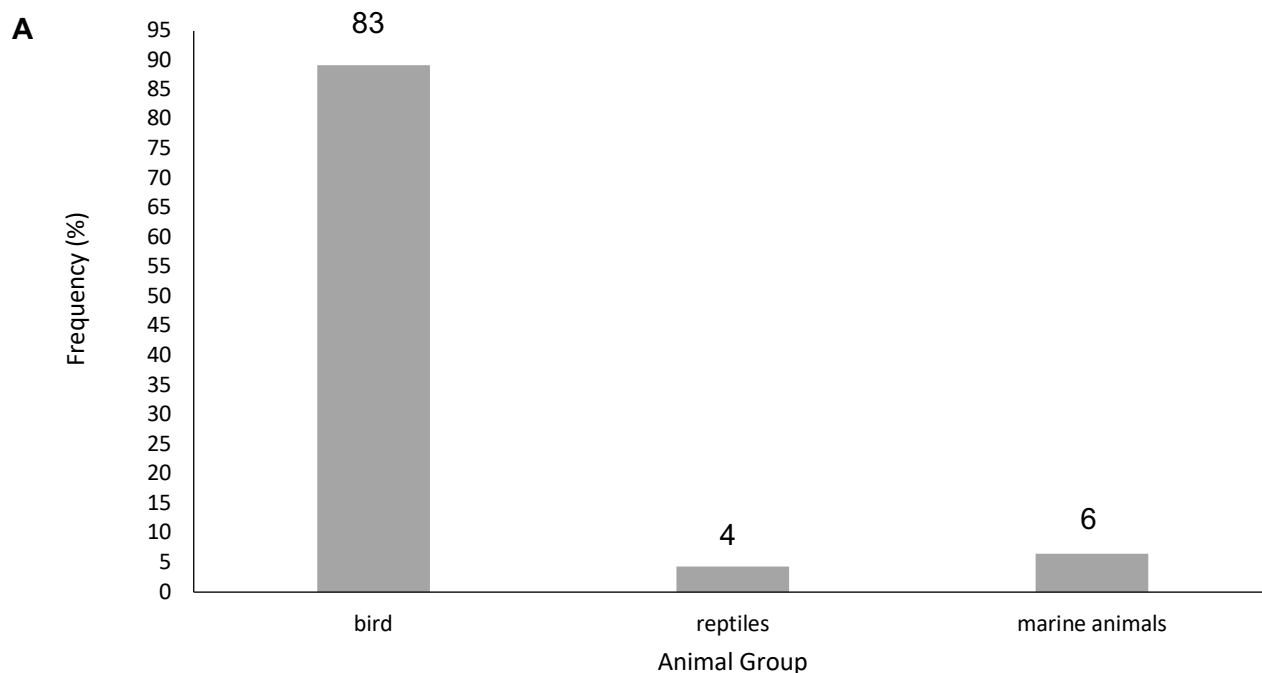


**Figure 3.1.5. Percentage frequency of animals containing a Phe instead of Tyr at position 975 in the Essex model and Cys instead of Phe at position 770 in the Zurich model.** The frequency was expressed as a percentage for the number of Phe residues found instead of a Tyr in the CD1 position (Essex model) with Adgb sequences from 362 species (**A**). The percentage was recorded for the number of Cys residues found instead of a Phe in the CD1 position (Zurich model) with Adgb sequences from 362 species (**B**). The numbers on top of the bar charts represent the absolute numbers of species recorded.

For the analysis of CD1 in the proposed 975 position (Essex model), 27 % (Fig 3.1.4A) of species was found to contain a Phe instead of a Tyr residue and Figure 3.1.4A shows that ~89 % is selective to birds. A few amount of reptiles (~4 %) and marine animals (~6 %) e.g. sea anemone, cucumber, and urchins also contained a Phe instead of the more common Tyr in Adgb-HD. The bird and reptiles belong to the Sauropsida clade. From the original paper by *Hoogewijs et al* from which the Zurich model sequence was constructed, it was also reported that in the Sauropsida clade animals are missing the Phe in the CD1 position. This was also seen for the Hemichordata clade (marine animals) for the Zurich model (Hoogewijs et al., 2012). However, in Figure 3.1.5A (marine animals) for Cnidaria (sea anemones) and Echinodermata (sea cucumbers and urchins), a Phe is found in these clades in the proposed CD1 position (amino acid 975, Essex model) (Hoogewijs et al., 2012). The results suggest that the birds and reptiles containing a Phe may be an exception and perhaps contain a variant resulting from the split from the common ancestor with mammals, as they are not commonly found to contain the CD1 Phe as predicted by *Hoogewijs et al*. Although, the original analysis was conducted on 34 different species and in this study, the analysis was conducted on 362 species. Therefore, the occurrence of Phe in the Sauropsida clade may be more legitimate since a larger analysis was conducted. The other animal groups not mentioned were not accounted for as they contained a Tyr residue and zero variation.

For the analysis of CD1 in the proposed 770 position (Zurich model), 30 % (Fig 3.1.2C) of species were found to contain a Cys residue instead of a Phe in the CD1 position. Figure 5.1.5B shows that this Cys (~82 %) is selective to birds. Reptiles (~17 %) also contain this Cys residue in the CD1 position. From the original paper by *Hoogewijs et al*, the Sauropsida clade (includes birds and reptiles) is shown to contain a Cys but in the calpain domain of Adgb, not the haem domain. Due to a high frequency of Cys in reptiles and birds, the results suggest that this is common in birds and some reptile species. The results seem likely as this analysis was conducted using 362 species as opposed to 34 species that were used in the original study by *Hoogewijs et al* (Hoogewijs et al., 2012). However, why many species of birds and reptiles contain the Cys variant remains unclear, especially the benefits of it or whether this is just a mutation. The other animal groups not mentioned were not accounted for as they contained a Phe residue and zero variation.

### 3.2 Phylogenetic analysis of Androglobin with other globin members.

Vertebrate globins share a common ancestor that was present 800 million years ago (Ascenzi et al., 2016). Hb, Mb, Ngb, Cygb, and Adgb are present in most gnathostome (jawed) vertebrates. However, Adgb is very distinct from the other globin members for example, even though it contains the conserved globin fold it is circularly permuted and has additional protein domains similar to flavoglobins (Burmester and Hankeln, 2014). Therefore, a cladogram (phylogenetic tree) was formed to present Adgb's phylogenetic relationship with the other globins members (Figure 3.2.1).

The human DNA for globin members (Hb $\alpha$ , Hb $\beta$ , Mb, Ngb, Cygb, and Adgb) was obtained from the Uniprot database. The sequences were then aligned using ClustalWS on Jalview. The aligned sequences were then imported onto the MEGAX software to form a cladogram to describe the phylogenetic relationship amongst human globins. Figure 3.2.1 presents the globins in their specific clades to demonstrate their evolution origin. This shows that Cygb and Mb form a common clade within Vertebrata suggesting that they are closely related and perhaps have a common ancestor. Hb $\alpha$  and Hb $\beta$  also share a common clade; Gnathostoma and have evolved in vertebrates (Pesce et al., 2002). Hb and Mb are penta-coordinated with their sixth binding site being unoccupied (Burmester and Hankeln, 2014). Ngb and Adgb are hexa-coordinated and emerged before the split between Deuterostomia and Protostomia. The remaining members are products of vertebrate-specific duplication events (Storz et al., 2013). This corroborates with the current understanding that Adgb has its own clade and does not have a close phylogenetic relationship with the other globin members.

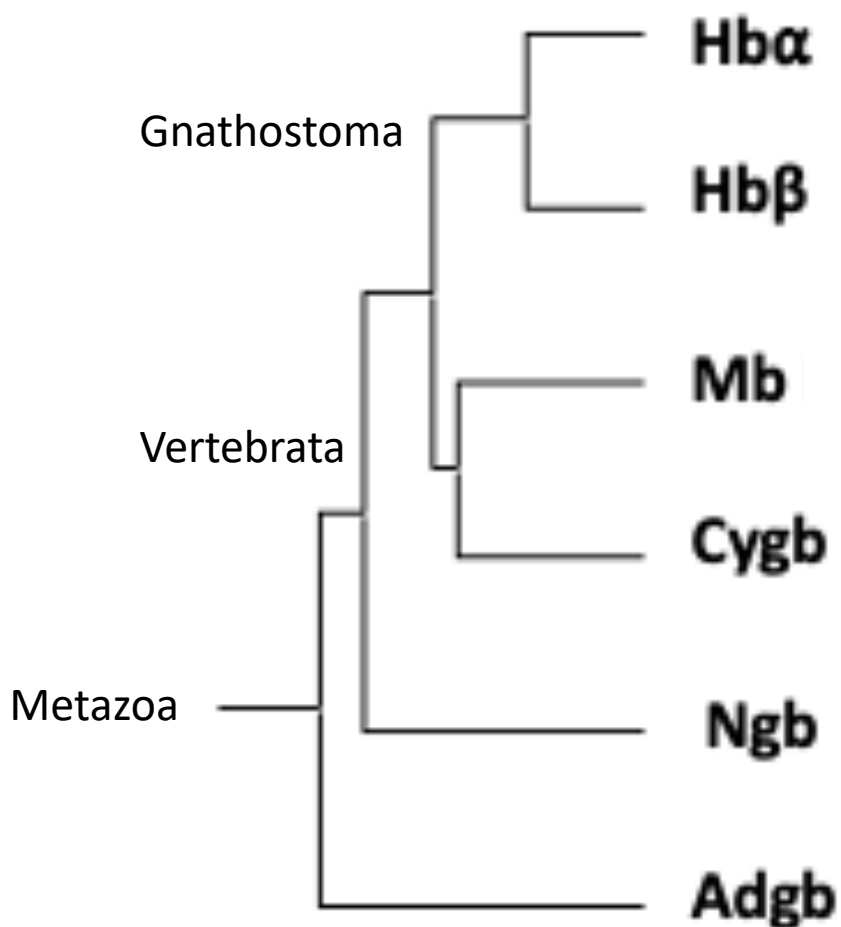


**Figure 3.2.1. Cladogram for the evolution of human globins.** The cladogram shows the ancestral precursor of haemoglobin alpha subunit (Hb $\alpha$ ), haemoglobin beta subunit (Hb $\beta$ ), myoglobin (Mb), neuroglobin (Ngb), cytoglobin (Cygb), and androglobin (Adgb) (Ascenzi et al., 2016). The cladogram was generated using the MEGAX software.

### 3.3 A 3D model prediction of Androglobin haem domain

There is currently no 3D model or X-ray crystal data for the haem domain of androglobin or any of the other domains of the protein (Hoogewijs et al., 2012). The construction of the two 3D model structures was based on the sequences in Table 2.7.1 (section 2.7 Molecular modelling), for the Essex and Zurich model sequences in Appendices 10 and 11. Templates of the structures were formed using the I-Tasser and Phyre2 servers, which are protein folding recognition servers to help generate helices and the IQ-calmodulin binding motif in the HD. I-TASSER (Iterative Threading ASSEmbly Refinement) is a server that provides protein structure prediction and structure-based function annotation. This server initially identifies structural templates from PDB by a multiple threading approach, with full-length atomic models constructed by iterative template-based fragment assembly simulations (Yang and Zhang, 2015). Phyre2 (Protein Homology/AnalogY Recognition Engine V2.0) is a tool online to predict and analyse protein structure, function and mutations to help build 3D models (Kelley et al., 2015).

Templates with the highest sequence identity to the model sequences were selected by I-TASSER and Phyre 2 providing the top 10 predicted protein structures. The models were constructed from two polypeptide fragments using Ngb as a template. Due to the haem domain being circularly permuted the sequences were put together in the order D-H (N-terminus) and A-C (C-terminus) in PyMOL to present the Essex model of Adgb-HD (Figure 3.3.1A). The same was repeated for the Zurich model but sequences was assembled according to Appendix 10 (Figure 3.3.1B). The threading templates used in the algorithms do not include haem groups. The haem group was extracted from the Ngb protein structure (PDB code: 4MPM).

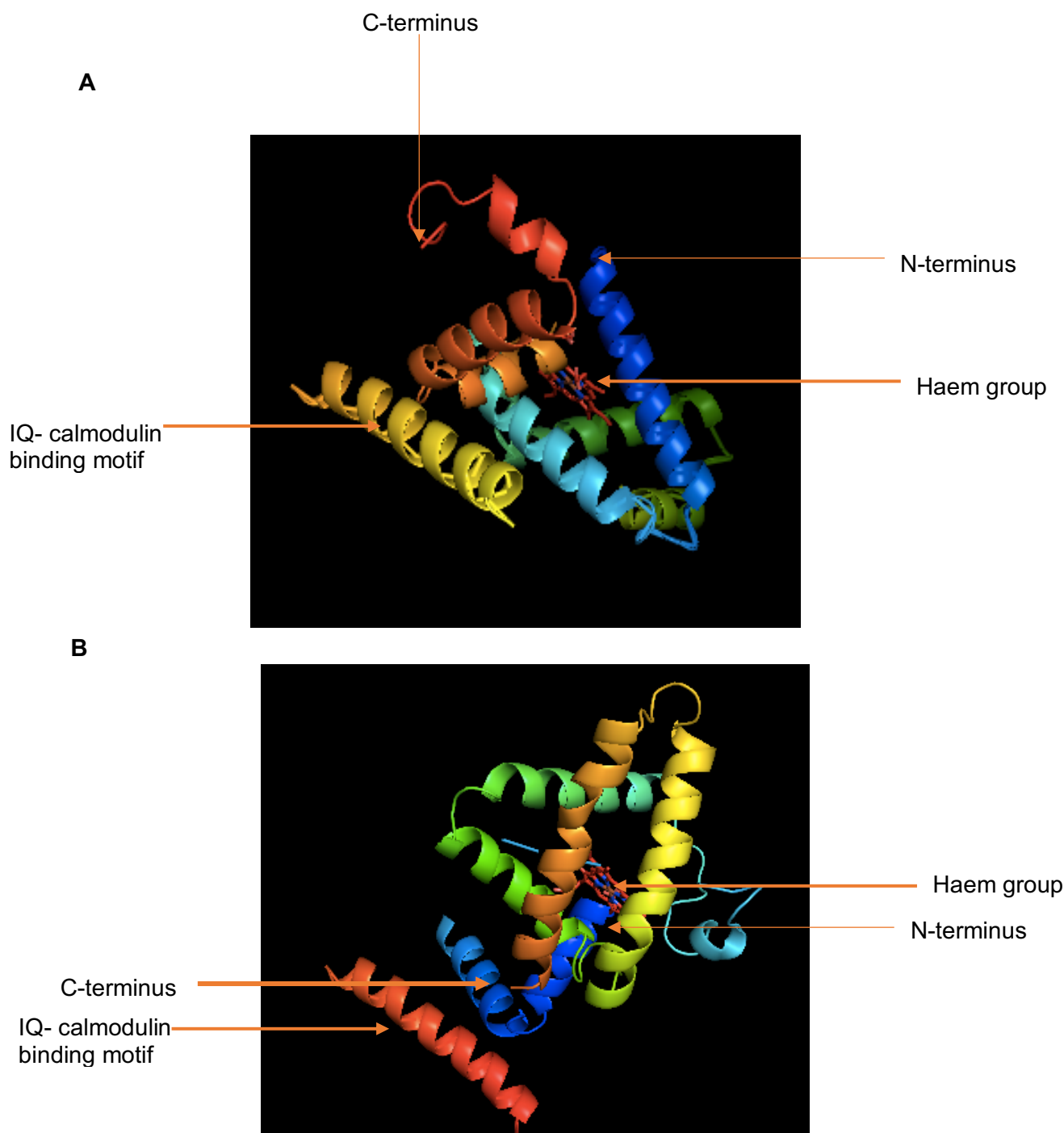


**Figure 3.3.1. 3D model of the Androglobin-haem domain, including the IQ-calmodulin binding motif.** The androglobin-haem domain is shown as a ribbon structure generated by PyMOL. **A**, A 3D model of Adgb-HD based on the Essex model sequence. The blue and green ribbon structures represent helices D-H whereas, the red and orange helices represent helices A-C. The yellow helix is the IQ-calmodulin binding motif (as indicated by the orange arrow). **B**, A 3D model of Adgb-HD based on the Zurich model sequence. The blue ribbon structures represent helices A-B whereas, the green and orange helices represent helices C-H. The red helix is the IQ-calmodulin binding motif (as indicated by the orange arrow)

and the porphyrin ring is shown in red in both images as a stick structure to represent the haem group. The N and C terminals are indicated by the orange arrows.

Figure 3.3.1 shows the typical 8  $\alpha$ -helices of the haem domain, with the addition of the alpha helix of the IQ-calmodulin binding motif (in Fig 3.3.1A&B). The IQ-calmodulin binding motif is separate; The I-TASSER and Phyre2 softwares cannot construct a model with the IQ sequence, so this was done separately and was manually placed into a position consistent with where the peptide should be. The helices were assembled separately (due to the circular permutation) in the order A-C at the C-terminus and D-H at the N-terminus to represent the Essex model, as depicted in the alignment diagram in Appendix 11. The helices for the Zurich model were assembled in the same way but different order for instance, A-B located at the C-terminus and C-H at the N-terminus (Appendix 10). The haem group is coordinated by the F8 proximal His and the distal Gln E7. Figure 3.3.1 shows that the ribbon view for both models has a classical 3-over-3  $\alpha$ -helical sandwich structure however, the models do not contain the common CD loop as seen in other globin members. The absence of the CD loop is due to the permutation so that the N and C terminal regions of the haem domain is in the place of the CD loop. The position of the IQ-calmodulin binding motif in Fig 3.3.1 is still uncertain, as its exact location in relation to the circularly permuted HD of Adgb is still unknown.

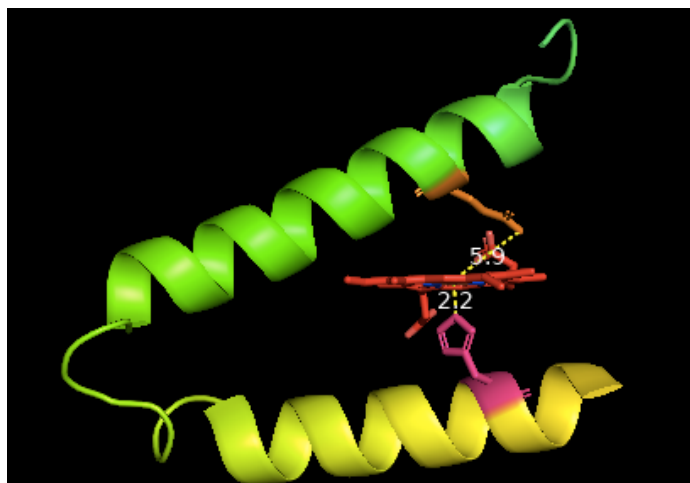
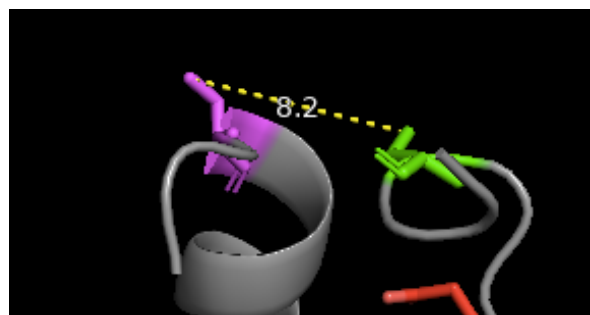
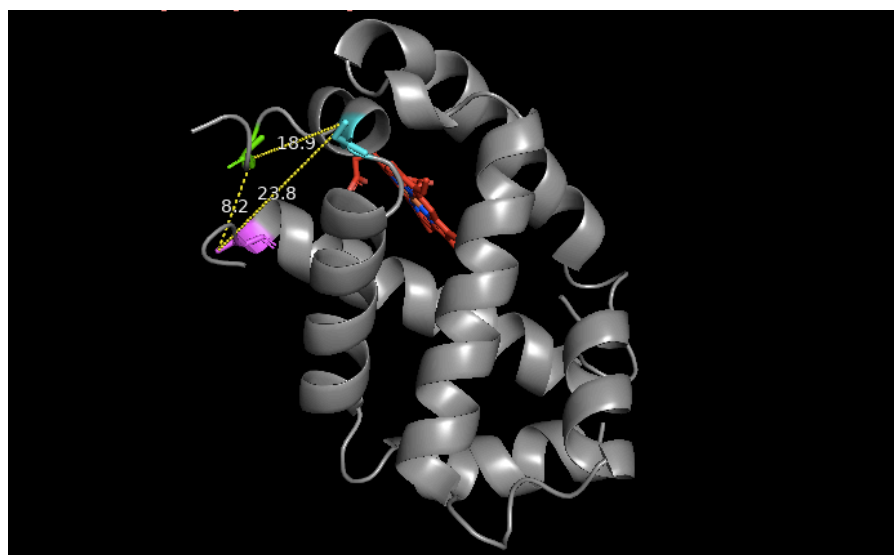
**A****B**

**Figure 3.3.2. A 3D model of Androglobin haem pocket and the potential disulfide bond formed in the CD region of the Essex model.** The proximal His824 on helix F is indicated in pink and the distal Gln 792 on Helix E is indicated in orange as well as relative interatomic distance to the central haem iron for each residue, (**A**). The predicted surface exposed cysteines in the Essex model that may form disulfide bonds are indicated in magenta (Cys786 on D helix) and green (Cys977 on C helix), as well as the relative interatomic distance (40.5 Å) between the two residues, (**B**). The third Cys residue, Cys970 in the haem domain is displayed in light blue, incapable of forming a disulfide bond on with the Cys residue on the D helix as it is not surface exposed. Bond length was measured in Å.

Figure 3.3.2A provides an insight into the haem pocket of Adgb-HD as well as the bond lengths for the proximal His F8 (2.2 Å) and distal Gln E7 (5.9 Å) coordinated to the central haem iron, suggesting that the proximal His residue is positioned closer and contains a smaller bond length than that of the distal Gln. In Ngb, the bond length between the distal His64 and the haem iron is ~ 2.7 Å thus, the distal haem position in Adgb-HD is greater than what is observed in hexacoordinated globins (Vallone et al., 2004). The lack of proximity between the haem iron and distal Gln residue could indicate a more pentacoordinate geometry for Adgb-HD. A bond length of 8.2 Å between two surface-exposed Cys (Cys 786 and Cys 977 shown in magenta and green respectively, Fig 3.3.2B) is indicative of a potential surface-exposed disulfide bond present on Adgb-HD for the Essex model, similar to a surface disulfide bond observed in Ngb. This suggests that a disulfide bond may form on the surface of Adgb-HD for the purpose of regulating ligand-binding properties like in Ngb and Cygb (Reeder and Ukeri, 2018). Adgb-HD contains a third Cys (position 970) as shown in light blue (Fig 3.3.2B); although it is not in close proximity with the other Cys residues (bond length of 18.9 and 23.8 Å to Cys 786 and Cys 977, respectively) Cys 970 was not considered as a possible residue for disulfide bond formation as it was not located on the surface of the HD and is too far from the other Cys residues to form a disulfide bond.

The globin fold model was superimposed onto a monomer of Ngb (PDB code: 4MPM). The IQ-calmodulin binding motif was not included in the alignment as its position was unknown. Ngb was selected for alignment to Adgb-HD, as Ngb is more closely related to Adgb than the other globin members, although the sequence homology of Ngb is low compared to Adgb-HD, as Adgb is very distantly related to the other globin members (Fig 3.2.1). The superimposition was completed on DeepView v.1.6 for both the Zurich and Essex 3D models (Figure 3.3.3).

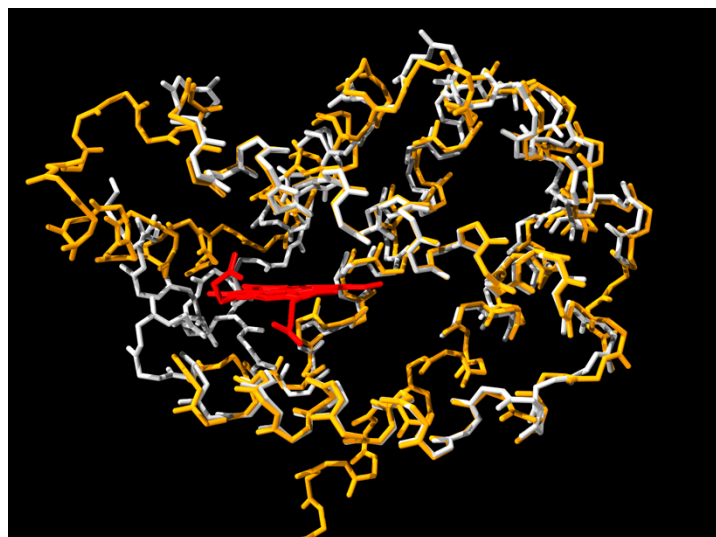
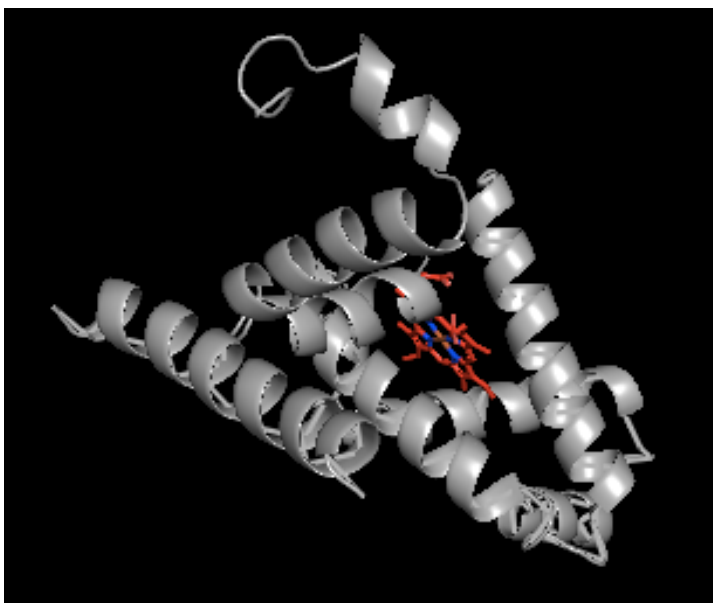
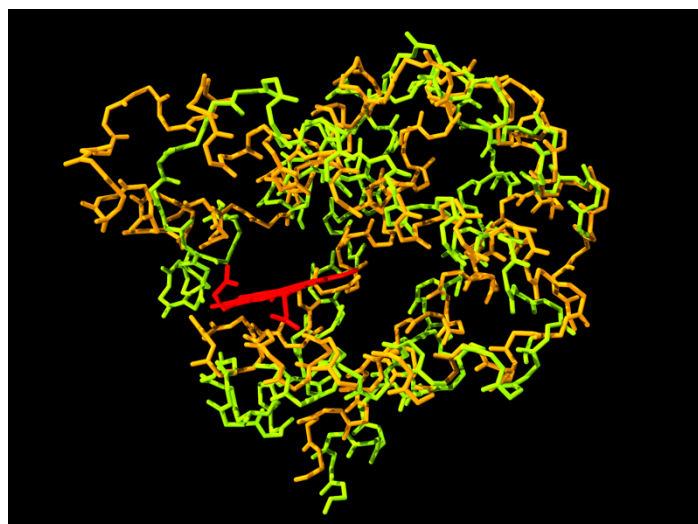
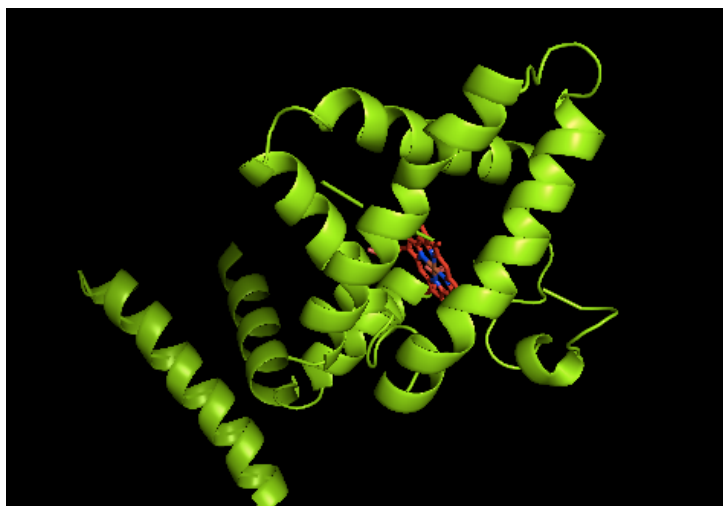
**A****B**

**Figure 3.3.3. 3D model of the Essex and Zurich models compared to Neuroglobin.** The Essex model of androglobin-haem domain (grey) is shown as a ribbon structure generated by PyMOL, as well as the alignment with Ngb (orange) (**A**). The Zurich model of androglobin-haem domain (green) is shown as a ribbon structure generated by PyMOL, as well as aligned with Ngb (orange) (**B**). Both models were superimposed onto Ngb for structural comparison. The porphyrin ring is shown in red as a stick structure to represent the haem group, for all 3D models.

Figure 3.3.3A shows the Essex model in grey and Ngb in orange. Both ribbon structures contain 8  $\alpha$ -helices representing the globin-fold. Pairwise sequence alignment of Adgb-HD and Ngb showed to have an 11.7 % sequence identity and sequence similarity of 15.6 %. The sequence identity and similarity scores suggest that Ngb and Adgb-HD are poorly homologous, which are below the 20-35 % considered to be in the 'twilight zone' protein sequence homology (Agüero-Chapin et al., 2020). Again, this low sequence homology could be due to Ngb and Adgb not being phylogenetically close (Figure 3.2.1). However, both the haem domain of Adgb and Ngb contain the same conserved globin-fold. Figure 3.3.3A shows a superimposition of the Essex model and Ngb which displayed that helices F-H in Adgb is superimposed on to Ngb. Therefore, helices F-H in the Essex model is homologous to Ngb helices. Adgb-HD does superimpose with the Ngb structure well but not completely, this could be because of the permutation in Adgb-HD and the poor sequence homology. Figure 3.3.3B shows the superimposition of the Zurich model and Ngb which displayed that helices E-H are superimposed on to Ngb but to a lesser extent than the Essex model, suggesting homology to some degree between the Zurich model and Ngb structure.

## Chapter 4. Structural and biochemical characterisation of Androglobin haem domain

In order to characterise the haem domain of androglobin, the human haem domain was recombinantly expressed with a C-terminal His tag and purified from *E.coli* as described in section 2.3.1, and the 280nm to Soret peak ratio suggested a reasonable degree of purity. Once purified the protein concentration was calculated using UV/Vis spectroscopy and the beer-lambert equation. The spectra for Adgb-HD and mutants confirmed that there was protein present in the solution. The DNA of mutations were sequenced by sanger sequencing prior to protein expression to confirm the distal haem position had been correctly mutated. In addition, the oligomeric state of the purified protein was assumed to be a monomer (personal communication with Dr Reeder, who had previously purified and observed Adgb-HD as a monomer via size exclusion chromatography). The recombinantly expressed haem domain and distal mutants were then optically characterised, as well as the NO binding and nitrite reductase activity was explored. Further study into whether androglobin contains any surface-exposed cysteines on the haem domain was completed for the recombinantly expressed haem domain.

### 3.1 Optical characteristics of Androglobin haem domain and distal mutants.

UV-Vis spectra were utilised to determine the optical characteristics of the expressed proteins; Adgb-HD (Wild type), Q791H, and Q791G (distal mutants). For Q791H Adgb, the distal Gln residue in androglobin was substituted to a histidine residue to compare and represent the more usual conformation of the distal haem ligand and for Q791G Adgb the distal Gln was substituted to glycine to try to create a pentacoordinate haem iron. Mutants were created by site-directed mutagenesis by Dr Brandon Reeder. The optical characterisation was completed before biophysical characterisation (Figures 4.1.1-4.1.3). Three different oxidation states were tested for Adgb-HD and mutants; the ferric state (as isolated), ferrous state (by dithionite addition), and the ferrous-CO state.

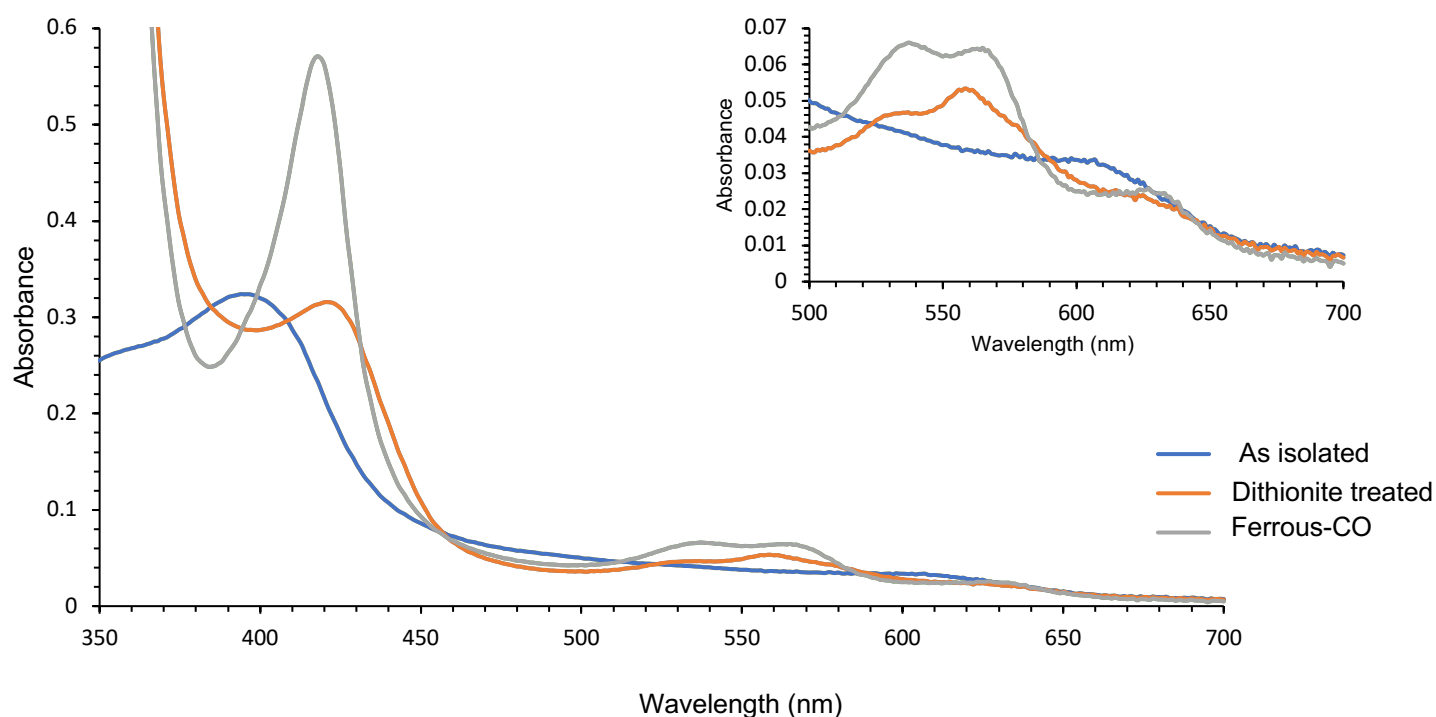


**Figure 4.1.1. UV-Vis spectra of Androglobin-haem domain (wildtype).** The optical characteristics for the recombinantly expressed Adgb-HD ( $\epsilon_{426\text{nm}}=108 \text{ mM}^{-1} \text{ cm}^{-1}$ ), 5  $\mu\text{M}$  of protein in 100 mM sodium phosphate buffer, pH 7.4. The spectra are shown for Adgb-HD in the as isolated (blue), dithionite treated (orange), and reduced ferrous-CO (grey) states. The ferric form was generated by the addition of CO during the induction of protein synthesis (section 2.3.1) so that the isolated protein would exist in the ferric state. After measurement of the protein as isolated the protein was reduced with dithionite and then CO was added to the dithionite treated protein.

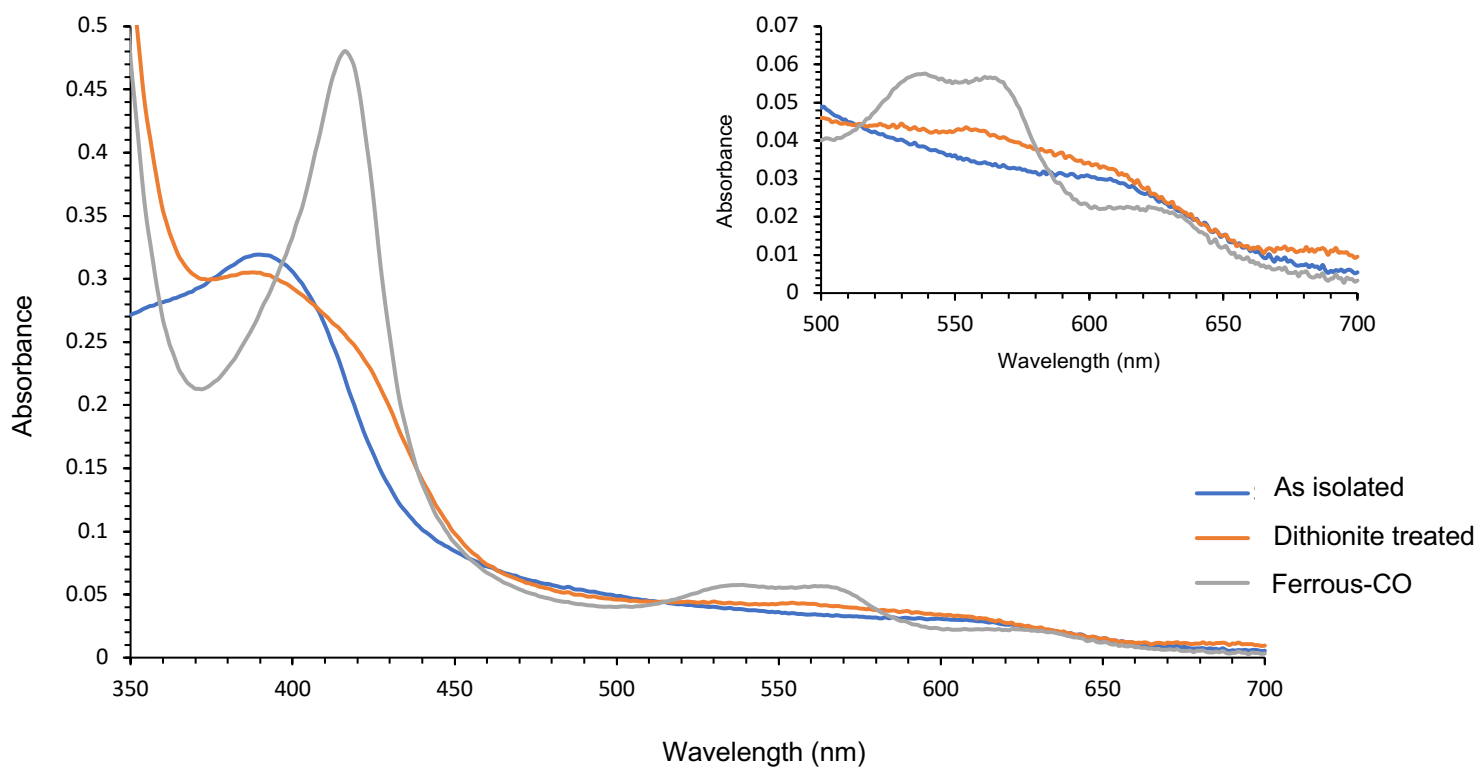


**Figure 4.1.2. UV-Vis spectra of Q791H Adgb distal mutant.** The optical characteristics for the recombinantly expressed Q791H Adgb distal mutant ( $\epsilon_{426\text{nm}}=108 \text{ mM}^{-1} \text{ cm}^{-1}$ ), 5  $\mu\text{M}$  of protein in 100 mM sodium phosphate buffer, pH 7.4. The spectra are shown for Q791H Adgb distal mutant in the as isolated (blue), dithionite treated (orange), and reduced ferrous-CO (grey) states. The ferric form was generated by the addition of CO during the induction of protein synthesis (section 2.3.1) so that the isolated protein would exist in the ferric state. After measurement of the protein as isolated the protein was reduced with dithionite and then CO was added to the dithionite treated protein.

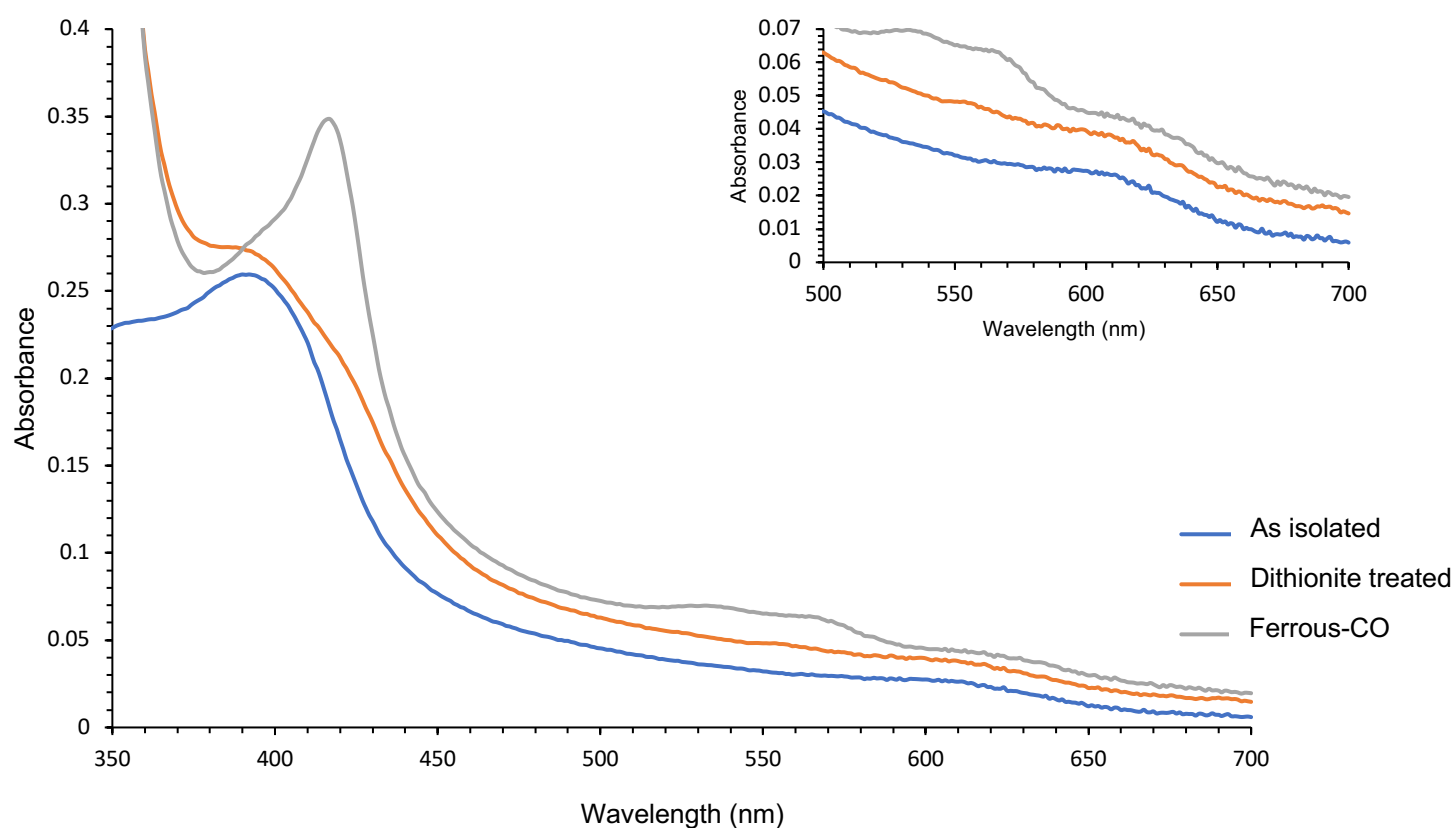


**Figure 4.1.3. UV-Vis spectra of Q791G Adgb distal mutant.** The optical characteristics for the recombinantly expressed Q791G Adgb distal mutant ( $\epsilon_{426\text{nm}}=108 \text{ mM}^{-1} \text{ cm}^{-1}$ ), 5  $\mu\text{M}$  of protein in 100 mM sodium phosphate buffer, pH 7.4. The spectra are shown for Q791G Adgb distal mutant in the as isolated (blue), dithionite treated (orange), and reduced ferrous-CO (grey) states. The ferric form was generated by the addition of CO during the induction of protein synthesis (section 2.3.1) so that the isolated protein would exist in the ferric state. After measurement of the protein as isolated the protein was reduced with dithionite and then CO was added to the dithionite treated protein.

The optical spectrum for Adgb-HD (Figure 4.1.1) showed a Soret peak at 396 nm for the ferric state (blue spectrum of protein as isolated from *E. coli*) and a broad visible peak at 608 nm, this peak is typical of a five-coordinate, high spin haem iron. Whilst other ferric globins, bands at around 540 and 580 nm are usually seen low spin haem iron. Such bands here are indistinct indicating a very little low spin haem iron (Sugita et al., 1971). The reduction of the protein by dithionite (orange spectra) led to a bathochromic shift, with a Soret peak of 421 nm. The visible peaks for the ferrous spectrum are between 535 and 560 nm, indicating at least a partial hexa-coordinated geometry which suggests that the protein may have only been partially reduced by dithionite. The binding of CO (grey spectra) resulted in a Soret peak of 417 nm, this was accompanied by visible peaks between 538 and 565 nm. When compared to other globins, the optical characteristics of Adgb-HD are somewhat similar to the optical properties of Ngb. For instance, the ferrous Ngb has a visible peak around 560 nm, similar to the visible peaks of Adgb-HD in the ferrous state. The spectra resemble pentacoordinate globin in the ferric state and a mixture of hexa- and penta-coordinated globin for Adgb-HD in the ferrous state as there are multiple peaks in the visible region for Adgb-HD (Dewilde et al., 2001). Therefore, the partial mix of states observed for the dithionite treated spectrum could suggest that the protein was not fully reduced by dithionite and was only partially reduced.

The optical spectra Q791H Adgb (Fig 4.1.2) showed a Soret peak for the ferric state (blue spectrum of protein as isolated from *E.coli*) of 390 nm and a low-intensity broad peak at 610 nm. The reduction of the protein by dithionite (orange spectra) led to a partial shift to form a shoulder at ~420 nm, consistent with a partial reduction of the haem. The broad visible peak for the ferrous spectrum was between 550 and 600 nm, the intensity of which was much lower compared to the Adgb-HD protein. However, the binding of CO (grey spectra) resulted in a large hyperchromic shift for the Soret peak of 417 nm and a bathochromic shift. This was accompanied by visible peaks of 540 and 570 nm.

For the Q791G Adgb distal mutant (Fig 4.1.3), similar results to the Q791H Adgb mutant were observed. Q791G Adgb showed a Soret peak for the ferric state (blue spectrum of protein as isolated from *E.coli*) of 391 nm and a broad peak around 610 nm. The protein treated by dithionite (orange spectra) led to a partial shoulder to form at ~420 nm which as discussed above this could be due to incomplete reduction

of the haem. The broad visible region for the ferrous spectrum showed peaks between 550 and 600 nm but the intensity of these were low compared to the Adgb protein. The binding of CO (grey spectra) once again resulted in a Soret peak of 416 nm and a bathochromic and large hyperchromic shift. This was accompanied by slightly broad visible peaks between 530 and 560 nm.

## 4.2 Determination of the cysteine state in the haem domain of Androglobin

A titration was completed to measure how many surface-exposed cysteines are present on Adgb-HD protein and their oxidation state as isolated. The ability of surface-exposed cysteines to form intramolecular bonds is an important feature for the ligand binding properties in Ngb and Cygb. For Cygb, the NiR activity increases 50-fold when the intramolecular disulfide bridge is formed between helix B and E (Reeder and Ukeri, 2018). In Ngb especially, the formation of a disulfide bridge between Cys46 (CD7) and Cys55 (D5) on the CD and D helix causes a decrease in the distal His64 (E7) binding affinity, sequentially increasing the protein's nitrite reductase activity (Guimarães et al., 2014). Therefore, the ligand binding properties of the closely related Adgb would be controlled by the coordination state of the Fe<sup>2+</sup> iron atom. For Adgb, the CD region is located on the N and C terminal segments of the haem domain, the presence of a disulfide bond could help stabilise haem binding to the haem pocket. The reactivity of the haem iron would be affected by the redox state of the cysteines as previously observed in Ngb and Cygb (Bocahut et al., 2009). The optical changes of the Adgb-HD were examined at 324 nm, as when dithiodipyridine reacts with free sulfhydryls, as suggested at 324 nm (Winther and Thorpe, 2014).

Figure 4.2.1 shows the spectral changes and the fractional saturation of dithiodipyridine binding to free sulfhydryl, produced when TCEP is applied for reduction of cysteines that may be in a disulfide bond. This was also completed without the TCEP reduction (Fig 4.2.2). TCEP was used to reduce any disulfide bonds formed between cysteine residues to become free sulfhydryls. The purpose of this experiment is to bind any surface-exposed cysteines to dithiodipyridine (which was recorded at 324 nm), enabling the number of surface-exposed cysteines on the HD of Adgb to be measured (Burns et al., 1991).

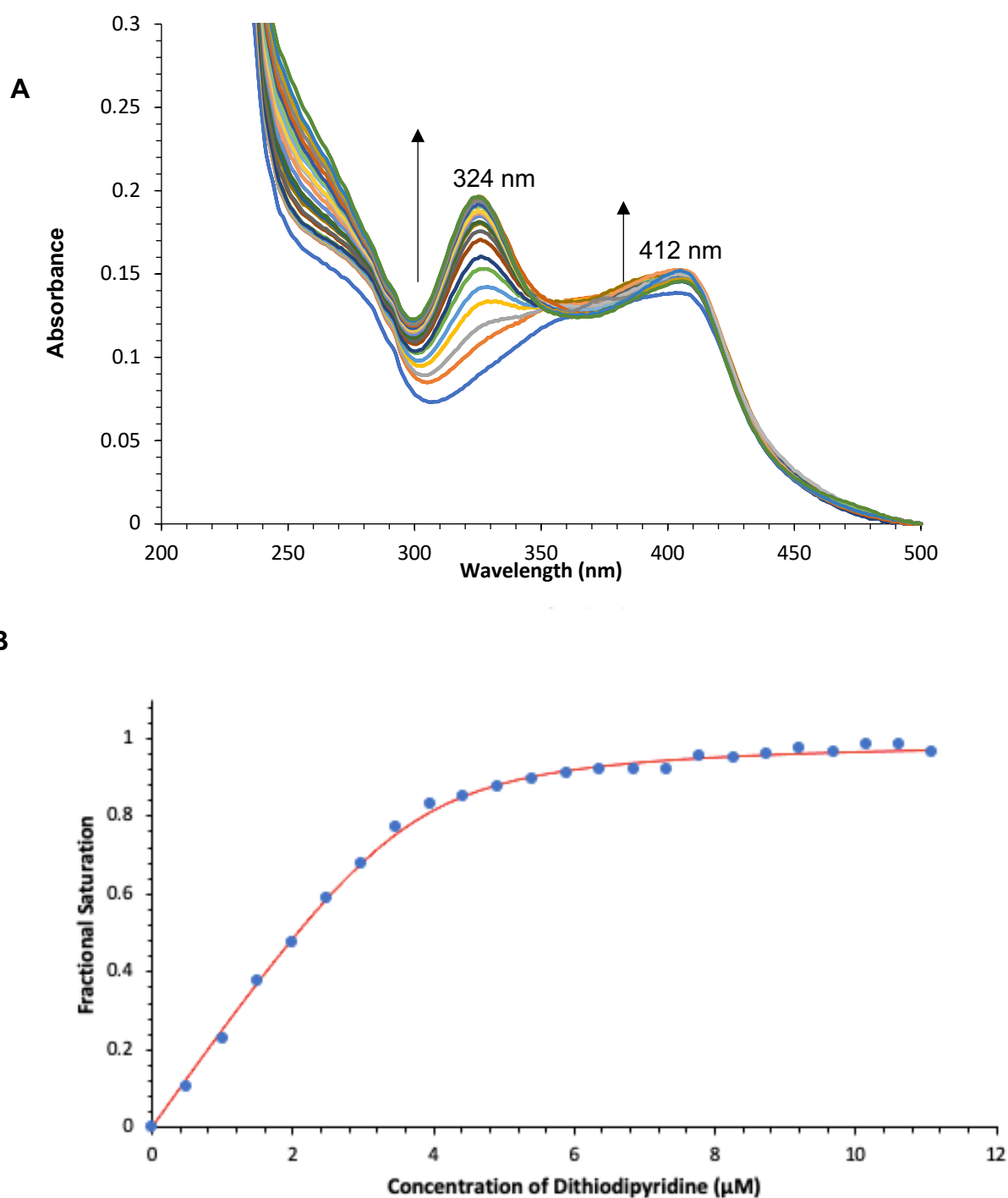


**Figure 4.2.1. The determination of the number of cysteines exposed on the surface of the haem domain in androglobin. (A),** the optical changes of 3.6  $\mu\text{M}$  Adgb-HD titrated with dithiodipyridine, showing binding to TCEP-reduced surface-exposed thiols **(B)**, the fractional saturation of the optical changes resulting from dithiodipyridine reacting with free thiols (blue dots), this was fitted to equation 2.4.1 (section 2.4) represented by the red line that used the  $K_D$  and total protein concentration as fitting

parameters. Fractional saturation was determined from the maximum absorbance change (at 324 nm) for the TCEP-treated protein.

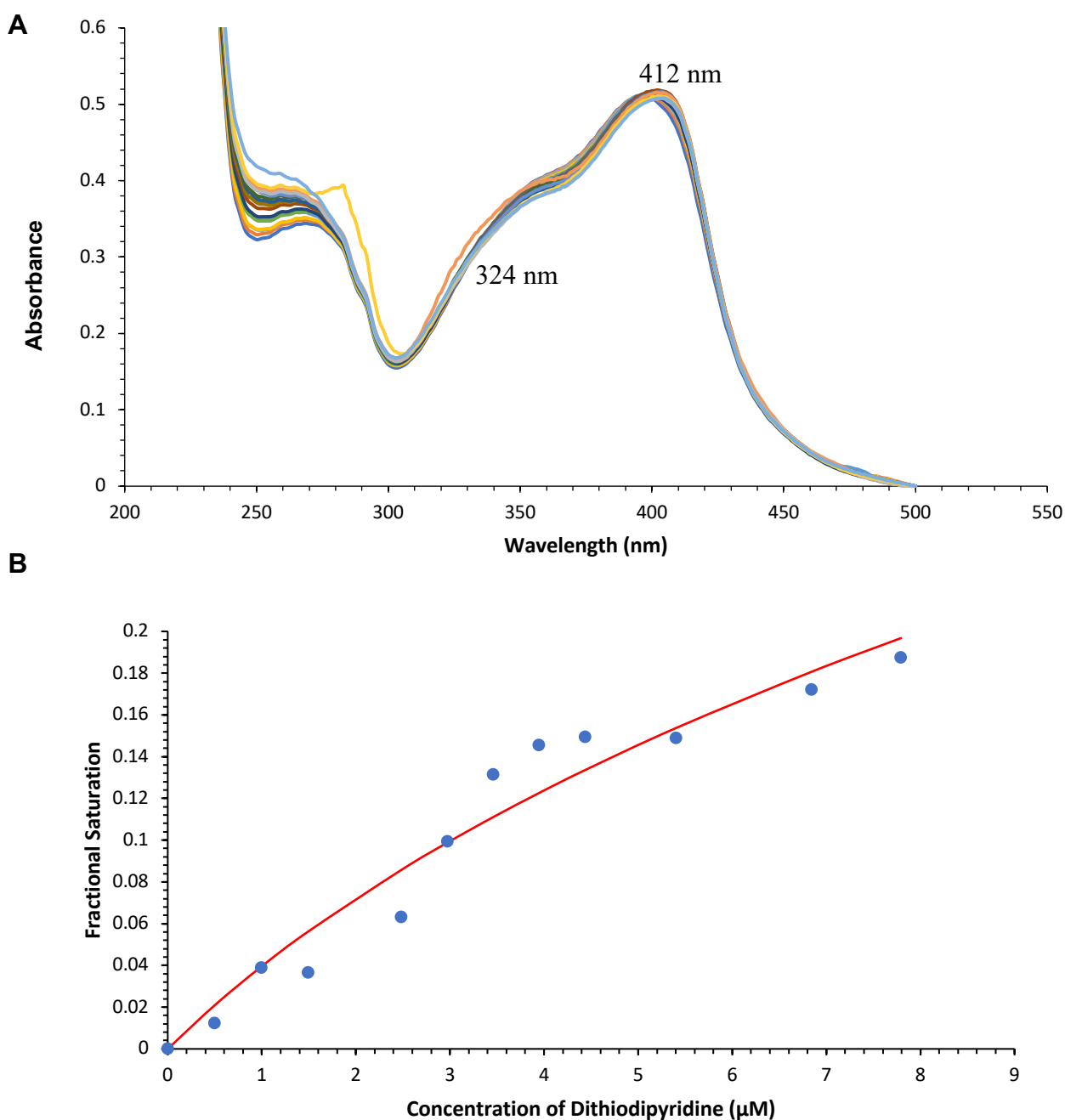


**Figure 4.2.2. The determination of the number of cysteines exposed on the surface of the Haem domain in androglobin.** **A**, The optical changes of 5.4  $\mu\text{M}$  of Adgb-HD titrated with dithiodipyridine, showing binding to surface-exposed thiols (without TCEP reduction). **B**, the fractional saturation of the optical changes resulting from dithiodipyridine reacting with free thiols (blue dots), this was fitted to equation 2.4.1 (section 2.4) represented by the red line that used the  $K_D$  and total protein concentration

as fitting parameters. Fractional saturation was determined from the maximum absorbance change (at 324 nm) for the non-TCEP-treated protein.

Figure 4.2.1A shows the optical changes at 324 nm, which shows that the addition of dithiodipyridine increases the absorbance at 324 nm as indicated by the black arrows in Fig 4.2.1A. Figure 4.2.1A, also shows a slight shift in the peak and absorbance increase at 412 nm, which is likely to be due to minor disturbances to the haem pocket structure resulting from the binding of dithiodipyridine to cysteines close to the haem pocket. The  $[P_T]$  (total protein concentration) was made to 3.6  $\mu\text{M}$  and the  $K$  value (dissociation constant) was 0.224  $\mu\text{M}$ . However, the Adgb Soret peak at  $\sim 1.6$ , gives an actual protein concentration of 1.5  $\mu\text{M}$  that reacts with dithiodipyridine ( $\epsilon_{426\text{nm}}=108 \text{ mM}^{-1} \text{ cm}^{-1}$ ). Therefore, Figure 4.2.1B shows that there is a 2:1 ratio of dithiodipyridine to Adgb protein (two thiols per haem) detected with the TCEP reduced protein suggesting that two reduced cysteines were accessible to the dithiodipyridine. From the Essex model sequence (Appendix 11) it is speculated that the disulfide bond detected is between Cys786 (D helix) and Cys977 (C helix) in the haem domain of Adgb, as predicted by 3D modelling (Chapter 3).

Figure 4.2.2 shows the determination of surface-exposed non-reduced cysteines, as isolated from *E.coli*. Figure 4.2.2A shows little optical change which indicates that the addition of dithiodipyridine did not have a significant effect on the absorbance at 324 nm without prior thiol reduction using TCEP. There was also a slight peak shift observed again at 412 nm. Figure 4.2.2B shows that there were no thiol groups detected, which means that, as isolated, all surface-exposed cysteines are oxidised and likely in disulfide bonds.

Therefore, the results suggest that the HD of Adgb contains a single disulfide bond, formed between N-terminal Cys786 (D helix) and C-terminal Cys977 (C helix) in the predicted CD region, assuming the protein exists as a monomer. The disulfide bond formed between the two Cys residues in the CD region are similar to those found in Ngb (in the CD loop) thus, it could be assumed that the presence of the surface-exposed Cys in Adgb-HD could provide stability for the haem pocket. The CD Cys residues were derived from the Essex model re-alignment sequence shown in Appendix 11. There are also three cysteines in the HD sequence and the last cysteine (Cys 970) is likely not surface exposed as discussed in Chapter 3, whereby Cys 970 is too far to form a disulfide bond (Fig 3.3.2).

### 4.3 Nitric oxide binding analysis of Androglobin haem domain and distal mutants

UV-Vis spectra were used to follow the optical changes of NO binding to Adgb-HD (WT, wild type), Q791H Adgb and, Q791G Adgb to provide further insight into the haem coordinate bonds involved in this. Figure 4.3.1, shows the absorbance spectra of NO binding to the ferrous-NO oxidation state for all three proteins. Ferrous state protein was achieved prior to NO addition by adding sodium dithionite to the protein mixture and then Proli NONOate to add NO to proteins.

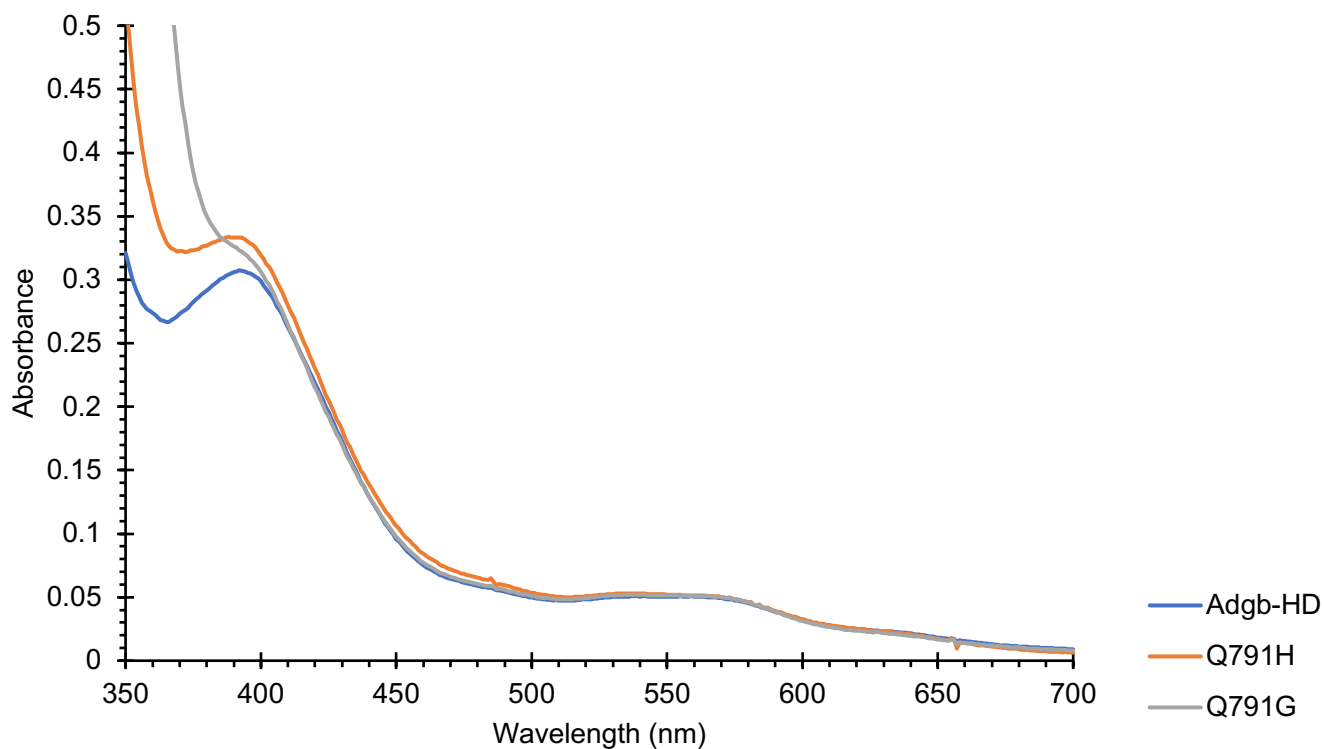
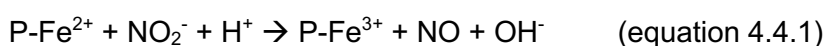


**Figure 4.3.1. The absorbance spectra for ferrous state of Adgb-HD (WT), Q791H Adgb, and Q791G Adgb when NO is bound.** For each protein, 5  $\mu\text{M}$  of protein in 100 mM sodium phosphate buffer, pH 7.4 was used. The blue spectrum represents Adgb-HD protein, the orange spectrum represents Q791H Adgb and the grey spectrum is Q791G Adgb with the addition of 40  $\mu\text{M}$  Proli NONOate to give a final concentration of 40  $\mu\text{M}$  NO.

Figure 4.3.1 shows that when Proli NONOate added NO did bind to each protein. For Adgb-HD (blue spectra) a peak at 394 nm formed. For Q791H Adgb (orange spectra), NO binds at 394 nm and 390 nm for Q791G Adgb (grey spectra) which is very similar to NO binding for Adgb-HD spectra. The visible peaks for all three proteins are also similar and show a broad peak between 540 and 480 nm. All three spectra for the three proteins are similar and the results suggest that NO binding occurs in a penta-coordinated state of the globin.

#### 4.4 Nitrite Reductase activity of the haem domain of Androglobin and distal mutants.

The nitrite reductase (NiR) activity was determined for the Adgb-HD to provide further insight into the protein's potential cellular function as a reductase. Many globins are believed to be involved in NO homeostasis through their NiR activity, acting as NO reservoirs by reducing nitrite ( $\text{NO}_2^-$ ) (Corti et al., 2016). The NiR role of globins has not yet been fully determined. NO plays an important role in vascular physiology as an autocrine and paracrine vasodilator molecule (Minnecci., 2008). Penta-coordinate globins such as Mb have a NiR role whereby they produce bioactive NO by reducing nitrite during cellular hypoxia (Hendgen-Cotta et al., 2010). Ngb, a six-coordinate globin can also reduce nitrite to NO (Reeder and Ukeri, 2018). Ngb plays a key role in brain homeostasis, whereby Ngb serves as a sensor during hypoxic stress (Baez et al., 2016). Under normal physiological conditions, Ngb can scavenge NO however, in hypoxia it can switch function to a nitrite reductase and produce NO (Reeder and Ukeri, 2018). Factors such as haem accessibility and redox potential have been related to NiR rates and studies of Ngb have shown that the ability of the distal haem residue can modulate NiR rates. Globins can catalyse the reduction of nitrite to produce NO according to equations 4.4.1 and 4.4.2 below. The reaction of nitrite with a globin (P) oxidises the haem iron to its ferric form and reduces nitrite to produce NO (equation 4.4.1). This reaction is highly pH-dependent as indicated by  $\text{H}^+$  in equation 4.4.1. In the presence of excess sodium dithionite, the ferric protein is reduced back to its ferrous protein form and allow NO to bind (equation 4.4.2) (Tejero et al., 2015). The second reaction *in vivo* normally takes place in hypoxic conditions, which is why dithionite was applied (to remove oxygen) and to ensure that the end product was not a mixture of ferrous-NO and ferric protein (Tiso et al., 2011).



The distal mutants (Q791H/G Adgb) and Adgb-HD (5  $\mu$ M) were reacted with sodium nitrite at varying concentrations (100  $\mu$ M - 1 mM) in 100 mM sodium phosphate buffer pH 7.4. The 1 M stock solution of sodium nitrite was diluted appropriately in buffer to produce the final concentrations mentioned. The sodium nitrite working stocks were then added to the protein in aliquots of 1.4 ml in independent experiments. Sodium dithionite was added separately to the nitrite working stock solution and protein before analysis, which was mixed together in the cuvette and the optical spectra was recorded immediately following this. Furthermore, the NiR activity was determined for distal mutants such as Q791H Adgb and Q791G Adgb, to compare the role and importance of the distal position in the haem group of androglobin on its NiR activity. Despite the optical spectra of the mutants presenting partial haem reduction/mix of penta- and hexacoordinated states, NiR activity analysis was completed on the mutants on the basis that Adgb could also be hexacoordinated. Therefore, for the Q791H Adgb mutant, the distal Gln in androglobin was mutated to histidine to compare it to other globins with a bis-His configuration, as this residue is common in most globin members. For Q791G Adgb the distal Gln was mutated to glycine to examine the effect of pentacoordination of the protein on the NiR activity and NO binding.

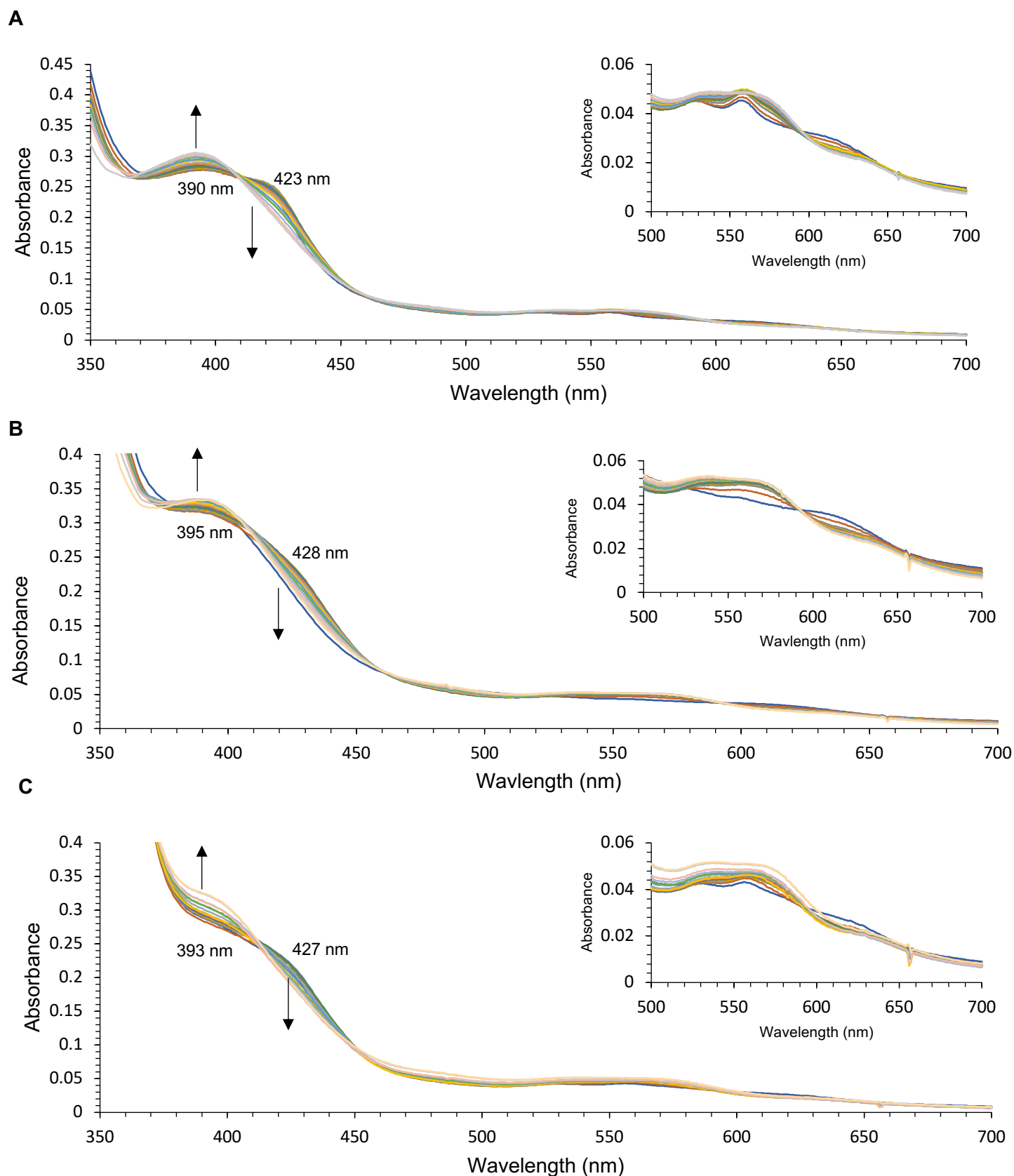


**Figure 4.4.1. Optical changes resulting from NiR activity in the deoxyferrous-haem domain of Adgb-HD and distal mutants. A,** Nitrite reductase activity of Adgb-HD. **B,** Nitrite reductase activity of distal mutant, Q791H Adgb . **C,** Nitrite reductase activity of the distal mutant, Q791G Adgb.

For Adgb-HD peaks at 390 nm for the ferric state and 423 nm for the ferrous-NO state which formed when nitrite and dithionite was added, respectively (Fig 4.4.1A). A decrease in absorbance was observed as a function of time and the absorbance becomes broad in the visible region ~540 – 600 nm. A similar pattern emerges for both distal mutants; Q791H Adgb mutant for the ferric state and ferrous-NO state absorbance was observed at 395 nm and 428 nm, respectively. For Q791G Adgb mutant for the ferric state and ferrous-NO state absorbance was observed at 393 nm and 427 nm, respectively (Fig 4.4.1B&C).

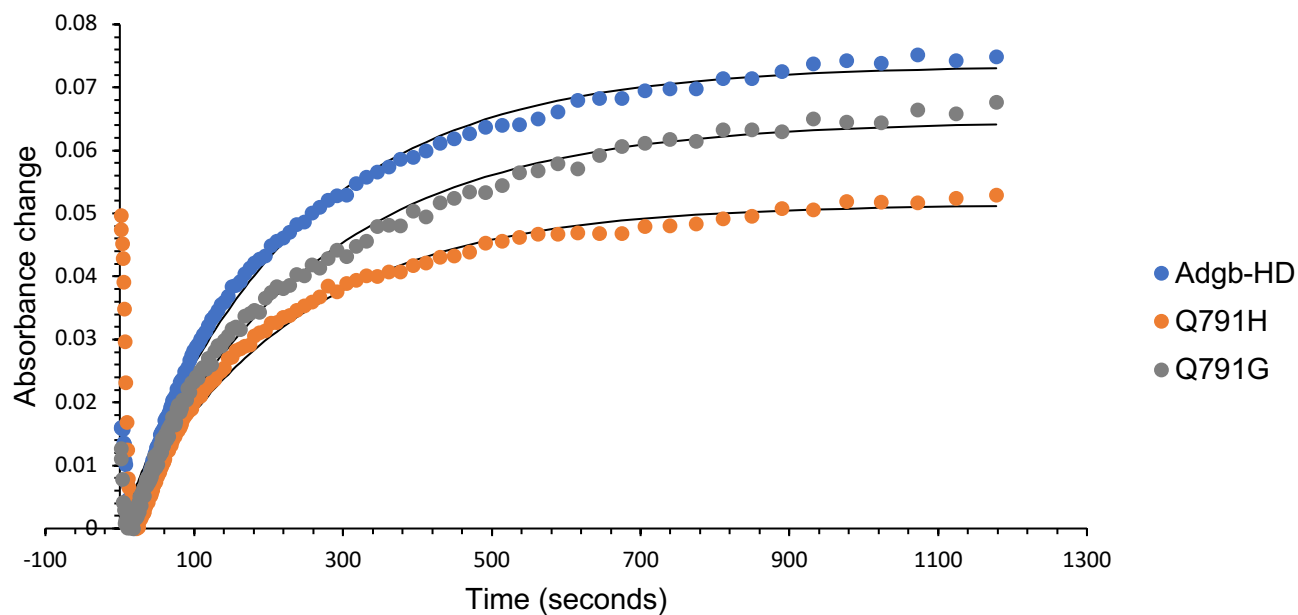


**Figure 4.4.2. The time course for nitrite reductase activity for Adgb-HD (blue), Q791H Adgb (orange) and, Q791G Adgb (grey) distal mutants.** The time courses for Adgb ( $0.0044 \text{ s}^{-1}$ ), Q791H Adgb ( $0.0044 \text{ s}^{-1}$ ) and Q791G Adgb ( $0.0040 \text{ s}^{-1}$ ) mutants, where the absorbance at 390 nm was measured over 20 min, generated by adding 500  $\mu\text{M}$  of sodium nitrite to 5  $\mu\text{M}$  of each protein in 100 mM sodium phosphate buffer, pH 7.4.

Figure 4.4.2A shows that the Adgb-HD (blue) and Q791G Adgb (grey) have similar changes absorbance in amplitude compared to the G791H Adgb (orange) distal mutant, which has a greater absorbance change within the initial part of NO binding (during the first 25 s). An absorbance decrease of 0.05 was observed in the first 25 s for Q791H Adgb. Whereas, for Adgb-HD and Q791G Adgb a smaller absorbance decrease of  $\sim 0.01$  was observed in the first 25 s. The time course was measured at 390 nm to analyse nitrite bound ferrous species of Adgb, which is usually very short lived in other globins and the only reference to a stable form is in horse heart Mb at 100 K. Therefore, the initial change due to nitrite bound to the ferrous species of Adgb is unlikely, as NO binding was measured at room temperature and not 100 K as previously done for horse heart Mb (Yi et al., 2010).

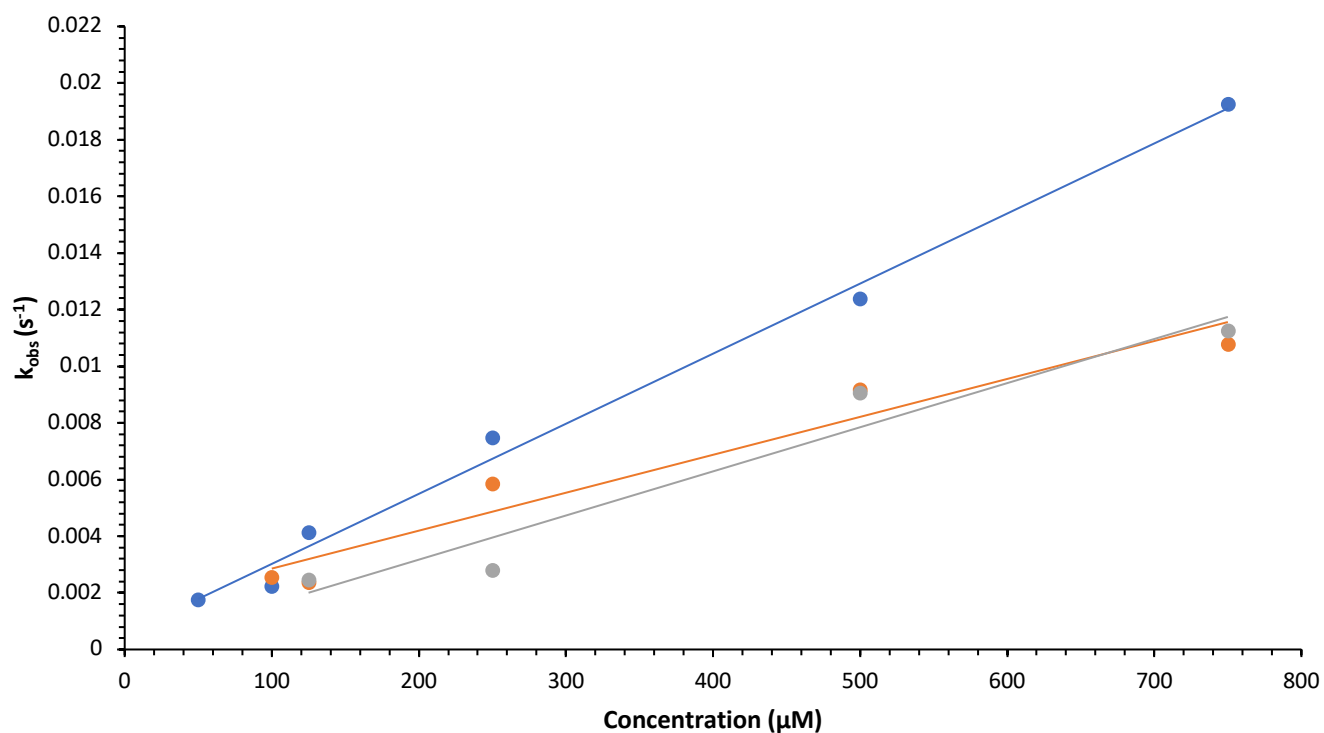


**Figure 4.4.3. The pseudo-first-order rate for Adgb-HD (blue), Q791H Adgb (orange) and, Q791G Adgb (grey) NiR activity and for the formation of ferrous-NO.** For Adgb-HD and distal mutants, 5 μM of protein in 100 mM sodium phosphate buffer at pH 7.4 was used. The rate constants were measured at varied nitrite concentrations (125 μM – 750 μM). The second-order rate constants were determined from this graph and calculated on Microsoft Excel.

Figure 4.4.3 shows that as the concentration of nitrite increases as does the rate for the WT and distal mutants. Therefore, the second-order rate constants were determined as a function of nitrite concentration. The second-order rate constant for the Adgb-HD was  $24.7 \text{ M}^{-1}\text{s}^{-1} \pm 1.07$ , for Q791H Adgb the rate constant was  $13.4 \text{ M}^{-1}\text{s}^{-1} \pm 1.87$ , and for Q791G Adgb the rate constant was  $15.6 \text{ M}^{-1}\text{s}^{-1} \pm 2.66$ . All proteins were fitted to a linear relationship consistent with previous studies whereby this has typically been a linear relationship between rate and nitrite concentration. The rate constant of Adgb-HD is greater than other pentacoordinate globins for example, equine Mb with a reaction rate of  $2.9 \text{ M}^{-1}\text{s}^{-1}$  which is  $\sim 8$ -fold lower than Adgb-HD rate constant. This indicates Adgb-HD has a high NiR activity compared to other pentacoordinate globins (Reeder and Ukeri, 2018). The Adgb-HD has the greatest NiR activity than the distal mutants since Adgb-HD has a higher second-order rate constant. For the distal mutants, they appear to have similar rate constants with Q791G Adgb only being slightly higher however, the difference is not significant. Therefore, this suggests that the mutation of the distal haem position to a histidine or glycine residue has little effect on NiR activity. Overall, Figure 4.4.1 shows that Adgb-HD and distal mutants all show Soret peaks at  $\sim 390$ - $395 \text{ nm}$  following the reaction with nitrite, which supports a pentacoordinate NO-bound arrangement in all three proteins (deoxyferrous-NO is formed) (Sun et al., 2016). The results (Fig 4.4.3) suggest that Adgb-HD has the greatest NiR activity compared to the distal mutants and is capable of reducing nitrite slightly quicker than Q791H Adgb and Q791G Adgb.

For the spectra at 390 and 428 nm (Fig 4.4.1), for all three proteins, an isosbestic point is observed, which indicates that a transition between ferrous and ferrous-NO, with no intermediate forms. For globins such as Ngb and Cygb, NO binds to hexacoordinated species but in comparison, Adgb and the mutants a penta-coordinate species was detected at 390 nm indicating that there is no endogenous ligand bound from the protein. The results suggest that the NiR activity could be similar to that of cytochrome *c'*, where NO binds to the proximal side. This reaction is biphasic, where the NO-hexa-coordinated haem is converted to a penta-coordinate haem, and NO competes with the proximal His residue for binding to the proximal haem position resulting in an absorption maxima of 395 nm, similar to Adgb-HD (Hough and Andrew, 2015). However, since no significant change is seen in the NiR activity for the mutants compared

to Adgb-HD therefore, it shows that the distal amino acid does not affect NO binding and this data, while not proving the hypothesis that NO binds to the proximal side, does not discount it either.

#### 4.5 Cloning the full Androglobin gene from a eukaryotic expression system vector (pcDNA 3.1 +) into an *E.coli* expression vector (pET28a).

There is currently little known about the full androglobin protein as it is relatively challenging to generate the full-length protein by recombinant biology techniques. The unstable protein expression is thought to be because of the permutation in the globin domain of Adgb. This instability has been observed in other engineered circular permuted globins such as permuted Mb (Ribeiro and Ramos, 2005). androglobin is considered to be a very large protein for recombinant expression by *E.coli* (1667 aa and 189.7 kDa) hence, the difficulties in isolating the full-length protein and the circularly permuted haem domain have been experienced.

The full-length Adgb gene was present in a pcDNA 3.1 (+) vector, used for eukaryotic expression. The pcDNA vector typically expresses proteins in lower yields due to difficulty in culturing cells in large numbers. To express the full Adgb protein by molecular cloning, the pET28a vector (Appendix 4) was double digested with BamHI and NotI restriction enzymes, to linearize the plasmid and to create sticky ends for ligation (Figure 4.5.2). The pcDNA 3.1(+)His-tagged full Adgb plasmid (Appendix 5) was digested separately with BamHI, AseI, and NotI restriction enzymes to effectively cut the Adgb gene out of this plasmid (Fig 4.5.3). Both digestions were verified by Agarose gel electrophoresis and DNA was isolated via gel extraction methods. The digested full-length Adgb gene insert was ligated into the pET28a vector by ligation as shown schematically in Fig 4.5.1. Figure 4.5.1 shows all restriction sites, for instance the pcDNA 3.1 (+) His-tagged full Adgb plasmid contains AseI sites as well as BamHI and NotI sites. AseI restriction enzyme was used to cut the plasmid at multiple AseI sites which would create smaller bands and help distinguish the band containing the full-length Adgb gene. Otherwise, the pcDNA plasmid would just be cut in half if only BamHI and NotI restriction enzymes were used, making it difficult to extract the DNA containing the full-length Adgb gene, due to the similar lengths of the two fragments.

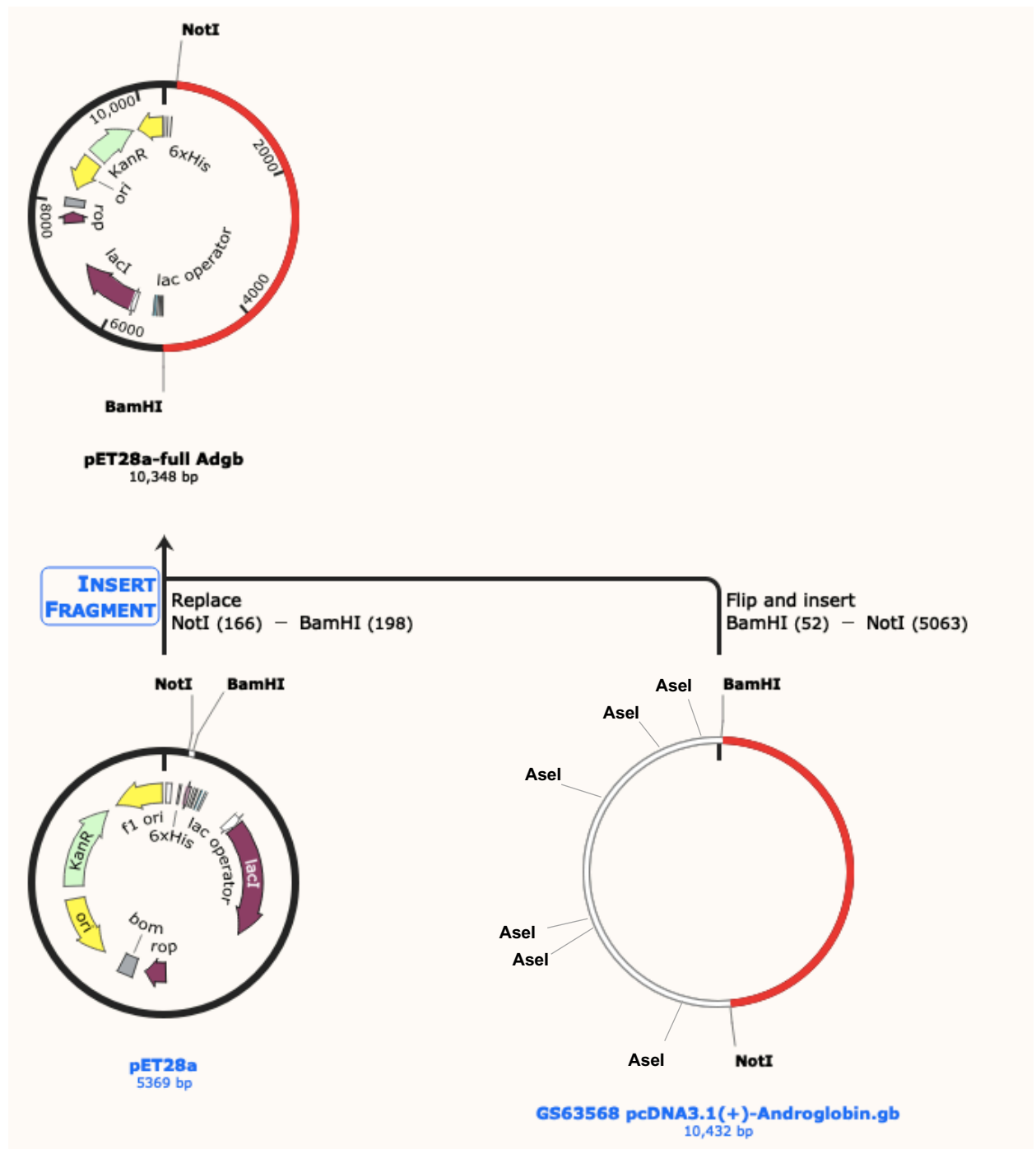


**Figure 4.5.1. Schematic representation of the ligation of the pET28a vector and the full Adgb gene insert.** The full Adgb gene (in red) insert from pcDNA 3.1 (+) full androglobin plasmid (right) was ligated into the pET28a vector (left) at a 2:1 insert-to-vector molar ratio, to produce the recombinant plasmid pET28a-full Adgb (top left). Plasmid maps were designed using SnapGene View software.

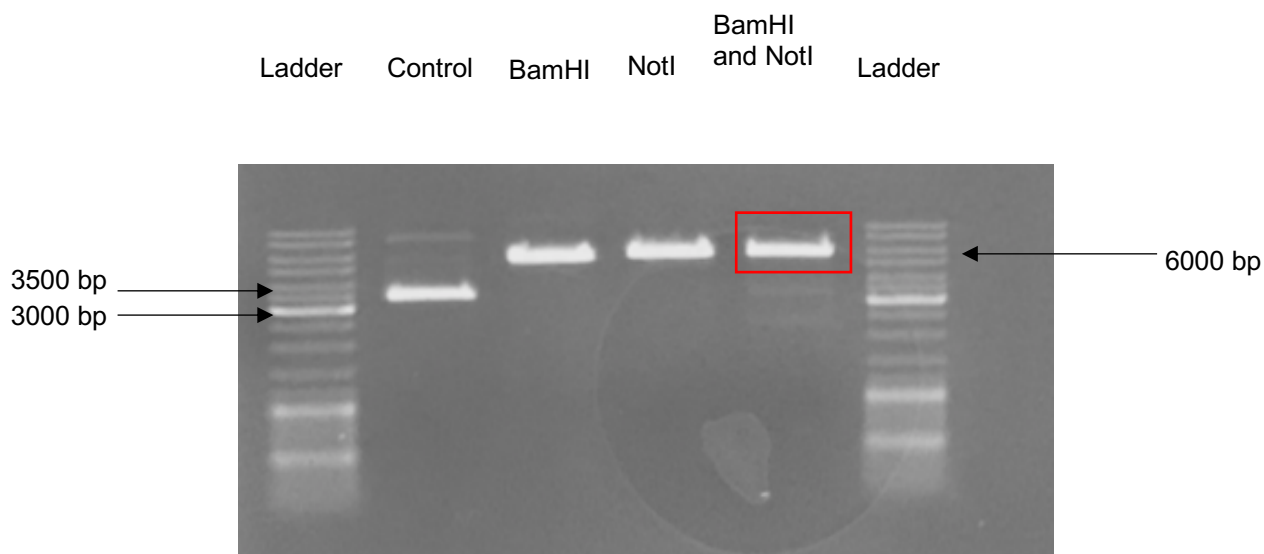


**Figure 4.5.2. Agarose gel electrophoresis for the pET28a vector digested with BamHI and NotI restriction enzymes.** Samples were loaded on a 1% Agarose gel against a 10 kb generuler (ladder). Lane 2 represents a negative control (no restriction enzyme added) whereas, lane 3/4 are positive controls containing one of each restriction enzyme. After electrophoresis, gels were stained with ethidium Bromide and the gel image was captured. The section highlighted in red was cut out of the gel for purification and Adgb gene insert by ligation.

A negative control containing no restriction enzymes was used therefore, the thick band in this lane represents the circular DNA, whereas the separate BamHI and NotI lanes contained one of each restriction enzyme for positive controls showing thick linear DNA bands for each. The pET28a plasmid is 5369 bp therefore, the band at ~6000 bp confirms both restriction enzymes are effectively competent at digesting the plasmid and the (Fig 4.5.2). The pET28a vector is theoretically 5369 bp therefore, the band at ~6000 bp for the double digest with BamHI and NotI indicates that the pET28a vector was successfully digested by both enzymes. BamHI and NotI restriction sites are 32 bp apart (Appendix 4) which implies that the double digestion will not have a large effect on the plasmid size and will remain ~6000 bp. The band highlighted in red from Figure 4.5.2 was excised from the gel and the DNA was extracted from the gel.

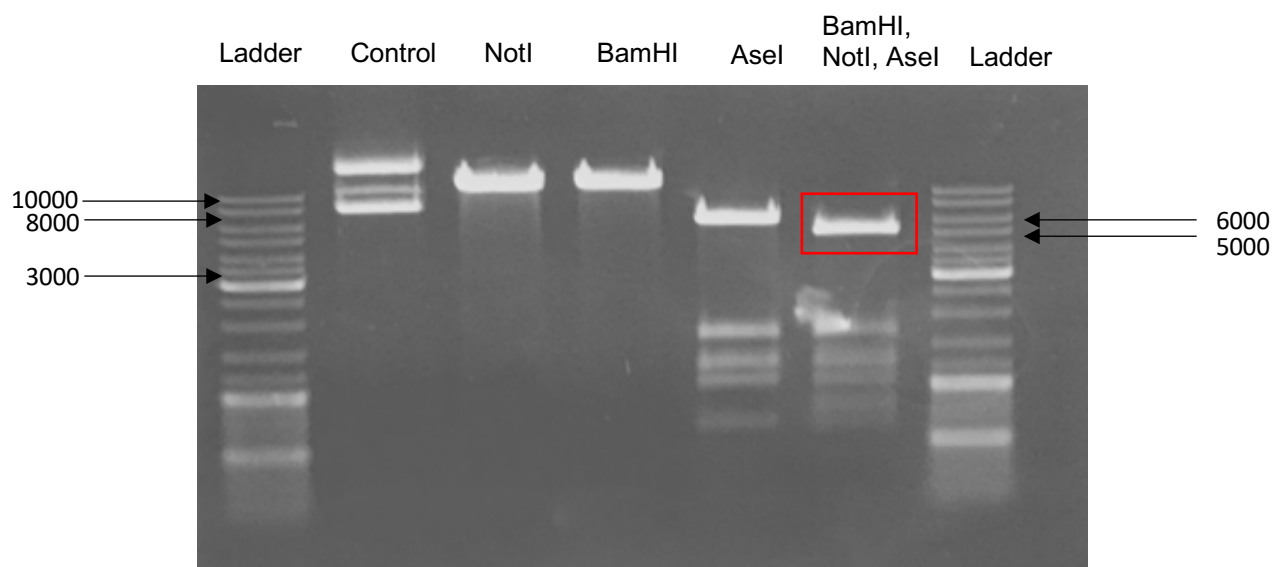


**Figure 4.5.3. Agarose gel electrophoresis for the pcDNA 3.1 His-tagged full Adgb plasmid digested with BamHI, NotI, and AseI restriction enzymes.** Samples were loaded on a 1% Agarose gel against a 10 kb generuler (ladder). Lane 2 represents a negative control reaction (no restriction enzyme added) whereas, lane 3-4 are positive controls containing one of each restriction enzyme. After electrophoresis, gels were stained with ethidium Bromide and the gel image was captured. The section highlighted in red was cut out of the gel for purification and ligation into the pET28a vector.

Figure 4.5.3 shows that for each positive control the respective restriction enzymes successfully linearized the plasmid, verifying that the enzyme cleaved the DNA. For the negative control and the positive control reactions (that contained only NotI and BamHI) a distinctive band at 10 kb was present, which corresponds to the correct (theoretical) size of the pcDNA 3.1 (+) His-tagged full Adgb plasmid (~10,4316 bp) as shown in Appendix 5. As for the sample containing only AseI enzyme, the plasmid has six AseI restriction sites (Appendix 5). Therefore, a smaller band at ~7 kb had formed as well as, multiple smaller DNA bands. The multiple DNA band fragments represent where the plasmid was cut multiple times with AseI restriction enzyme. Lane 6 in Figure 4.5.3 represents the triple digest. A thick band formed at 5 kb which corresponds to the size of the Adgb gene (5063 bp, Fig 4.5.1). The smaller DNA bands in lane 6 represent DNA fragments that were cut at the multiple AseI sites (Fig 4.5.1). The DNA band highlighted in red in Figure 3.5.3 was exercised from the gel and the DNA was extracted.

The full Adgb gene insert was then ligated into the digested and dephosphorylated pET28a vector at 16°C O/N at a 2:1 insert-to-vector molar ratio. The ligation DNA was then transformed into XL1B competent *E.coli* cells and DNA from one colony was cultured. The isolated DNA concentration was 92.5 ng  $\mu\text{l}^{-1}$ . The ligation reaction was verified by Agarose gel electrophoresis (Fig 4.5.4) to check that the correct size (10,348 bp, Fig 4.5.1) of the recombinant DNA was produced. The recombinant plasmid (pET28a-full Adgb, is shown schematically in Appendix 6) was cut with BamHI restriction enzyme to linearize it.

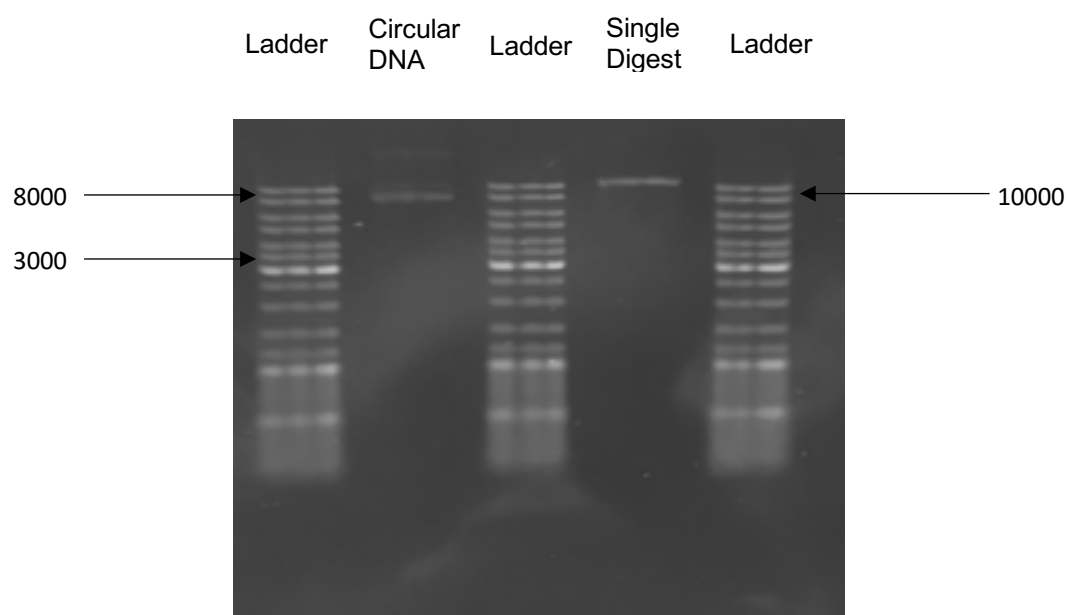


**Figure 4.5.4. Agarose gel electrophoresis for the pET28a-full Adgb recombinant plasmid digested with BamHI.** Samples were loaded on a 1% Agarose gel against a 10 kb generuler (DNA ladder). Lane 2 represents a negative control reaction (no restriction enzyme added). Lane 4 represents a single restriction site digest for the recombinant plasmid using BamHI restriction enzyme. After electrophoresis, gels were stained with ethidium Bromide and gel image was captured.

Figure 4.5.4 shows that a band at ~ 8 kb was produced for the control reaction. This contained no restriction enzyme. A single distinctive band was produced at 10 kb indicating that the full Adgb gene had been successfully incorporated into the pET28a vector. The digested pET28a vector and Adgb gene insert are both ~5 kb (Fig 4.5.1) therefore, the incorporation of both by ligation forms a ~10 kb recombinant plasmid which corresponds to the theoretical size (10,348 bp shown in Appendix 6).

After this, the recombinant DNA (pET28a-full Adgb) was transformed into competent XL1B *E.coli* cells for DNA propagation. The transformation mixture was then spread on to kanamycin sulfate plates and incubated O/N at 37°C. Two white colonies had formed and the DNA was extracted from a single colony. The DNA was measured to be 160.4 ng  $\mu\text{l}^{-1}$  by spectroscopy and 30  $\mu\text{l}$  of this was sent for forward and reverse DNA sequencing. Sequencing confirmed the presence of the full-length Adgb gene situated in the pET28a vector. A pBR3 primer was used instead of a T7 primer since the protomer region for this was too close to the open reading frame (ORF). This allowed for the full-length Adgb gene to be sequenced as well as avoiding obscure reads. After, the pET28a-full Adgb protein was expressed in BL21 DE3 cells to check (by SDS-PAGE gel electrophoresis) that the protein would express (Figure 4.5.5). The gel did not show sufficient levels of protein expression over that of the native proteins (full length protein with His-tag is >190 kDa), so expression of the full length protein cannot be confirmed. However, this was not evidence against the production of the full-length Adgb protein as the Adgb-HD protein (control) showed expression to which we know expresses.

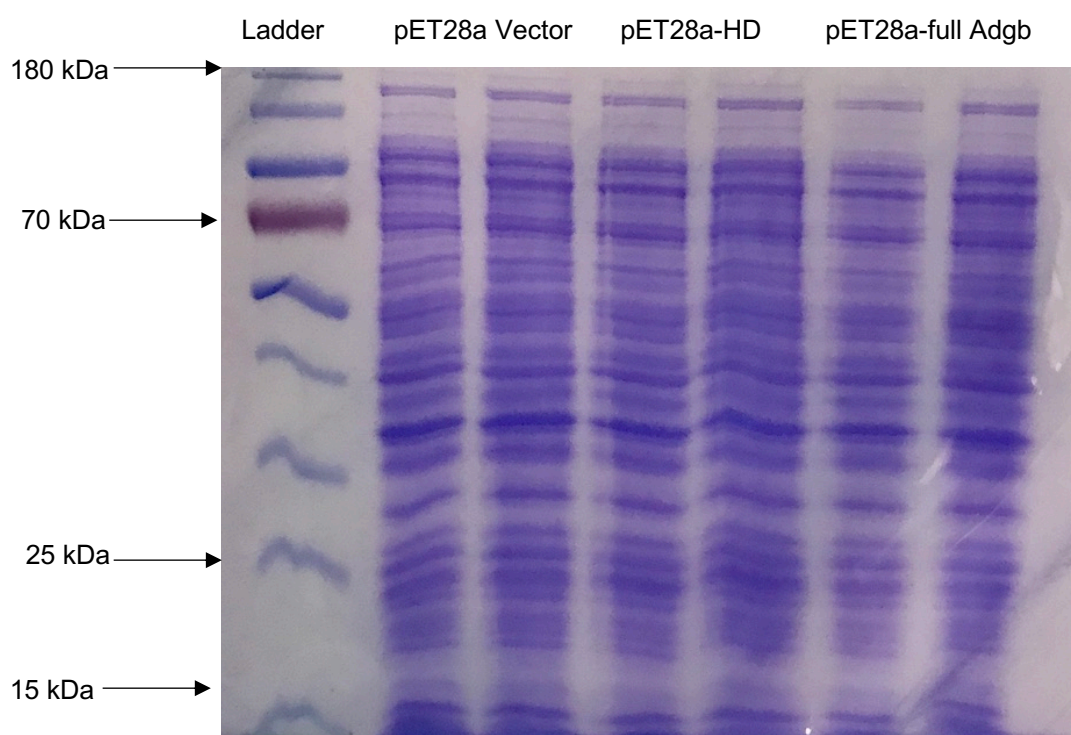


**Figure 4.5.5. SDS-PAGE gel electrophoresis for the expression analysis of pET28a-full Androglobin protein.** Samples were diluted in cracking buffer and loaded (in duplicate) on a 12% Acrylamide gel against a 180 kDa pre-stained protein ladder. Three samples were run in duplicate, pET28a vector (control), pET28a-HD (control) and, pET28a-full Adgb. After electrophoresis, gels were stained in Coomassie Blue stain and de-stain buffer. The presence of multiple thick bands in each lane, indicating that protein expression for the pET28a full Adgb is possible. The pET28a and pET28a-HD were used at positive controls and for comparison.

#### 4.6 Truncation of the pET28a-full Adgb Plasmid.

To increase the chance of expressing the protein under recombinant conditions, the pET28a-full Adgb DNA was truncated in one of two ways, both of which removes some or all of the C-terminal domain. It has been found that 200 aa preceding the HD, there is ordered secondary structures however after this, disordered secondary structures (section 4.3) were encountered which could be the cause for unstable protein folding and purification making biochemical characterisation of the full Adgb protein highly unlikely. Therefore, Adgb was truncated 200 aa after the HD as well as, exactly after the HD for comparison of the effect of disorder on protein expression and folding (schematic diagrams in Appendices 8&9).

Primers were designed to truncate the full Adgb by applying a TAA stop codon 200 aa after the HD (HD200 F/R primers) and precisely after the HD (HD F/R primers) (for primer sequences see Table 2.2.1.1 in section 2.2). The aim was to produce two truncated versions of the Adgb proteins; pET28a-truncated Adgb where the protein is truncated immediately following the haem domain and pET28a-truncated Adgb(200), where the protein is truncated 200 aa after the haem domain. The pET28a-full Adgb recombinant plasmid was used as a DNA template in PCR and Pfu DNA polymerase was used for DNA replication (section 2.2.2). Four PCR reactions were made to contain different primer pairs and 0 or 5 % DMSO combinations. PCR reaction 1 (Adgb HD F/R primers without DMSO), PCR reaction 2 (Adgb HD F/R primers with DMSO), PCR reaction 3 (Adgb HD200 F/R primers without DMSO) and, PCR reaction 4 (Adgb HD200 F/R primers with DMSO). DMSO was added to prevent secondary structures of the DNA template from forming by altering the  $T_M$  value of the primer to change the optimum annealing temperature (Chester and Marshak, 1993). After the PCR, one of each reaction was treated with DpnI restriction enzyme to remove the parent template (Fig 4.6.1).

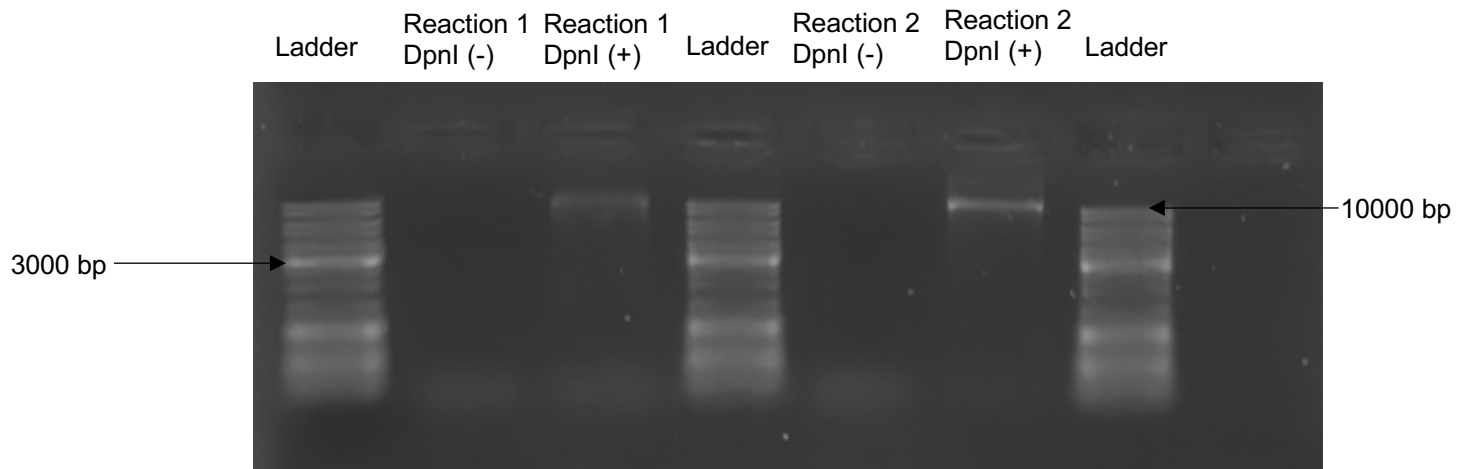
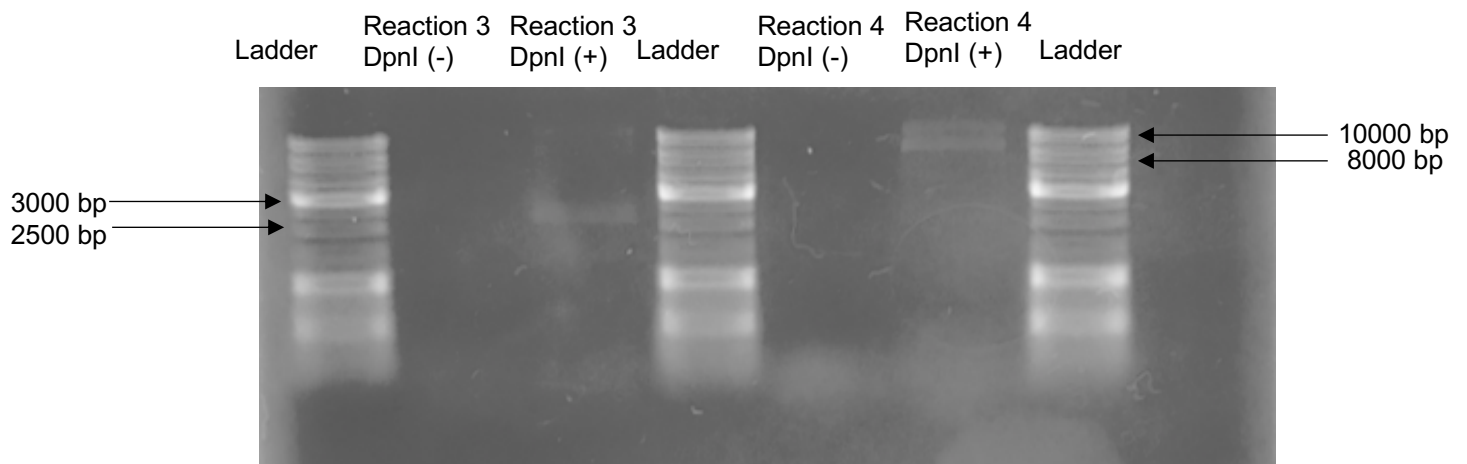
**A****B**

**Figure 4.6.1. Agarose gel electrophoresis for the truncation of the pET28a-full Adgb by PCR.**

Samples were loaded on a 1% Agarose gel against a 10 kb generuler (DNA ladder). Reaction 1-4 were run in duplicate and one of each reaction contained (+) DpnI restriction enzyme (**A**, reactions 1 and 2. **B**, reactions 3 and 4). After electrophoresis, gels were stained with Ethidium Bromide and gel image was captured.

Figure 4.6.1 shows a distinct band at ~10 kb for reaction 1 and reaction 2 (which contained 5 % DMSO) treated with DpnI (Fig 4.6.1A) confirming the truncation exactly after the HD was successful. As for Figure 4.6.1B, reactions 3 and 4 ( which contained 5 % DMSO) had been treated with DpnI had successfully been mutated via PCR, as DNA bands were produced. Figure 4.6.1, shows that the PCR reactions for all that did not contain DMSO were unsuccessful, suggesting DMSO optimised DNA yield. The DNA from the bands treated with DpnI in Figure 4.6.1 was extracted and purified by gel extraction. The truncated DNA samples were sequenced to check that the TAA stop codon had been correctly positioned (sequencing primer TGGGCCGTTCTGGAGATGAACTTG). The results confirmed that both truncations; exactly after and 200 aa after the HD had occurred and that the TAA stop codon was correctly positioned for reactions 2 and 4 reactions. After, a very small amount of the pET28a-truncated Adgb (reaction 2) and pET28a-truncated Adgb(200) (reaction 4) protein was expressed in BL21 DE3 cells to check (by SDS-PAGE gel electrophoresis) that there were no challenges in protein expression (Figure 4.6.2). The gel appeared to be a negative result but was not evidence against the production of the truncated Adgb proteins as expression had occurred.

The molecular weight for both truncated proteins was determined by Expasy ProtParam bioinformatic resource tool. The theoretical molecular weight for pET28a-truncated Adgb is 112.1 kDa and for pET28a-truncated Adgb(200) it is 134.9 kDa to which the gels failed to show.

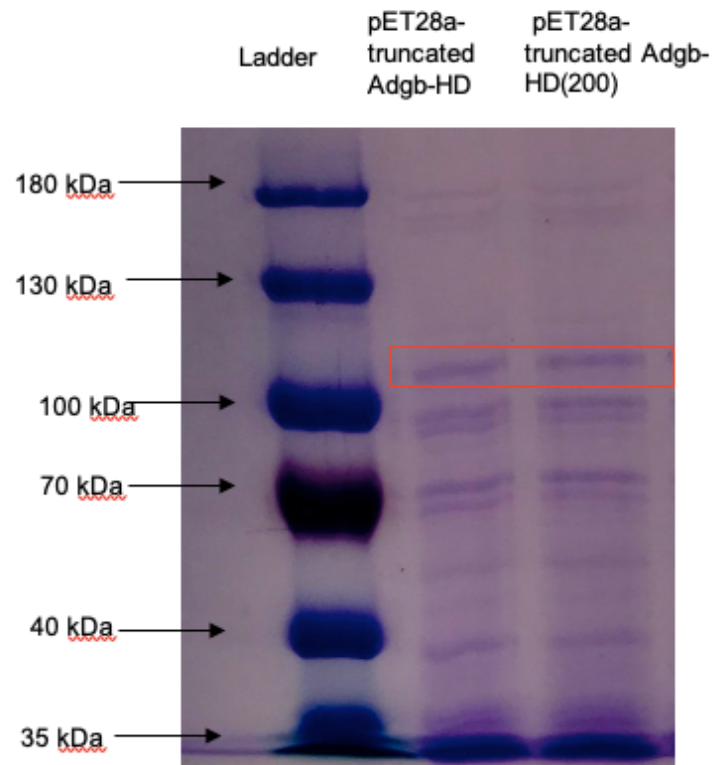


**Figure 4.6.2 SDS-PAGE gel electrophoresis for the expression analysis of truncated pET28 full Adgb protein.** Samples were diluted in cracking buffer and loaded on an 8% Acrylamide gel against a 180 kDa pre-stained protein ladder. Lane 2 is pET28a-truncated Adgb- HD (derived from PCR reaction 2), lane 3 is pET28a-truncated Adgb-HD(200) (derived from PCR reaction 4 respectively). After electrophoresis, gels were stained in Coomassie Blue stain and de-stain buffer. The bands corresponding to the molecular weight of truncated proteins is highlighted in red. The theoretical molecular weight for pET28a-truncated Adgb is 112.1 kDa and for pET28a-truncated Adgb(200) it is 134.9 kDa. Protein bands (highlighted in red) in each lane at just above ~100 kDa had been produced and correspond to the molecular weight of both truncated proteins.

## Chapter 5. Discussion

In this project, the purified Adgb-HD protein was structurally and biochemically characterised by establishing the key residues critical to the haem domain, producing a 3D model of the haem domain, experimentally testing the biochemical avenues of the haem domain such as NiR activity, NO binding, and surface-exposed cysteines, as well as clone the full Adgb gene insert into the pET28a vector. It was recognised that the haem domain of Adgb satisfies the criteria of the globin-fold however, it is circularly permuted and is split into two parts by an IQ-calmodulin-binding motif (Burmester and Hankeln, 2014). The original study of Adgb demonstrated that helices C-H is located on the N-terminal side and helices A-B is shifted to the C-terminal side (Appendix 10, the Zurich model sequence) which was constructed from NCBI conserved blast searches from 34 genomes across the metazoan taxa (Hoogewijs et al., 2012). Currently, there are more Adgb sequences available on the NCBI database, which has allowed for a re-alignment to be completed and more sequences to be compared. A reassessment of the CD helix sequence was completed by Dr Reeder and Professor Reynolds, prompting the formation of the Essex model sequence (shown in Appendix 11). The reassessment (Essex model sequence) was found to possess a different arrangement to the Zurich model sequence, whereby the Essex model sequence places the C helix on the C terminal side. Therefore, D-H helices are located on the N-terminal side and are split by the IQ-calmodulin-binding motif from helices A-C (personal communication with Dr Reeder and Professor Reynolds, unpublished work). The new consensus sequence was further analysed in this project to establish and identify the key amino acids which found Adgb to contain a proximal His, a distal Gln residue, and CD1 Tyr residue, as opposed to a CD1 Phe like previously reported (Hoogewijs et al., 2012). Phylogenetic analyses suggest that Ngb, GbX, and Adgb have an earlier evolutionary origin, consistent with previously reported literature. Ngb, GbX, and Adgb genes have been identified in both invertebrates and vertebrate animals implying that they emerged before the separation of Protostomia and Deuterostomia (Burmester. and Hankeln, 2014). Predicting a 3D model of the haem domain for both the Essex and Zurich model sequence (Appendices 10 and 11) was done so by molecular modelling. The models formed were aligned to Ngb, (Adgb is more closely related to Ngb than other globin members) for homologous structural comparison. Molecular modelling of the Essex model also provided insight into the

distal and proximal haem positions insinuating potential pentacoordinate haem geometry for Adgb and the possible presence of surface-exposed Cys in close proximity to be able to form a disulfide bond, similar to that of Ngb and Cygb (Reeder and Ukeri, 2018). The project then proceeds to apply experimental approaches to Adgb for further characterisation, whereby a stable form of the haem domain of androglobin was expressed and purified as well as the recombinant form of the full-length protein and partially truncated the protein was expressed. The expression of the full-length protein has not yet been achieved by recombinant methods, largely due to the large size of the full-length Adgb protein (1667 aa, 189.7 kDa) and it's circularly permuted haem domain causing the protein to be rather unstable (Burmester and Hankeln, 2014). Although, the haem domain of Adgb has been previously expressed, but this only went as far as being observed in the cell lysate due to instability. The lysate absorption spectrum was characteristic of a six-coordinate haem, presumably with the iron coordinated to the proximal His and distal Gln (Hoogewijs et al., 2012). However, in this study Adgb-HD coordination examined on the purified haem domain protein via optical spectroscopy to reveal a mix of penta-/hexa-coordinated species. Furthermore, Adgb is predominantly expressed in mammalian testes, and reported to be associated with stages of spermatogenesis (Hoogewijs et al., 2012). Therefore, NO which plays a key function is sperm motility could be linked to the NiR activity of Adgb in the testes, potentially acting as a reservoir for NO (Lee and Cheng, 2004). This evidently appears to be likely, as Adgb-HD was measured to have a high NiR activity compared to other globin members (Corti et al., 2016). Lastly, the measurement of surface-exposed cysteines on the haem domain and as hypothesised by molecular modelling was detected experimentally on the haem domain of Adgb.

## 5.1 Sequence homology of amino acids in Androglobin haem domain

In the Zurich model sequence (Appendix 10) the key residues were found to be the distal Gln792 (E7), the proximal His824 (F8), and the Phe770 (CD1). The key residues in the original study were established by extensive in silico analyses and further BLAST searches in genomes of over 30 Adgb sequences (Hoogewijs et al., 2012). Since then more sequences have been made available for the haem domain to be re-aligned and the key residues to be verified. In this study, 362 sequences were analysed as opposed to 34 sequences using multiple sequence alignment (Fig 3.1.1), pairwise sequence alignments (Fig 3.1.2), and amino acid charts to represent conservation of the key amino acids (Fig 3.1.3). The results in Figure 3.1.2&3.1.3 for the proximal His and distal Gln showed they were highly homologous across all species analysed. However, 30 % of the species analysed contained a Cys residue in the CD position (Fig 3.1.2C&3.1.3C) instead of a Phe residue (as indicated by the Zurich model sequence). The benefits of this variant as well as why it may occur in a high frequency of species remains unclear. A Hb variant called Hb little Venice [ $\beta$ 42 Phe  $\rightarrow$  Cys] found in a Caucasian male infant who suffers from chronic haemolytic anaemia and Heinz body formation, contains a Cys residue instead of a Phe. Although, the patient's oxygen dissociation curve showed no alteration in oxygen affinity, chronic symptoms were still experienced. The Hb little Venice showed the Phe residue to be a crucial amino acid in the haem pocket, maintaining solubility and stability. The substitution of the Phe by the Cys removes contact with haem and leaves a gap at the surface of the haem pocket, causing instability. Therefore, a single substitution of Phe by Cys is detrimental to the structure of the haem group (Bartram et al., 2012). It is not clear why a CD1 Cys residue is advantageous, especially when a CD1 Cys residue is associated with an unstable haem group.

Most Adgbs contain a distal Gln (E7) residue instead of the more common His residue which is observed frequently in other globins hence, a hexa-coordination scheme of [Gln]-Fe-[His] (Hoogewijs et al., 2012). The distal His residue is common in vertebrate globins and stabilises ligands at the E7 position via a hydrogen-bonding network. On the other hand, the E7 position in most invertebrate globins is occupied by a Gln residue. For example, ligand stabilisation by the distal Gln has been detected in bacterial globins found in *Lucina pectinata* and *Ascaris suum*. Interestingly, the distal Gln does not provide significant

stability to ligands due to conformational restraints imposed by the polypeptide and in some bacterial globins, this region (E7-10) of the polypeptide is disordered, resulting in the distal Gln being rotated out of the haem pocket (Yeh et al., 2000). Occasionally a distal Gln residue has been detected in a few vertebrates. The African elephant (*Loxodonta africana*) is a paradigm of a vertebrate, whereby its Mb contains a distal Gln (Teixeira et al., 2011). The invariant residue could represent an adaptation to hypoxia, a common feature in many invertebrate globins (Projecto-Garcia et al., 2017). The proximal His residue also plays a key role in mediating the binding of co-factors to the protein and modulating binding properties of the haem group (Moore and Fasan, 2019). Mutations of the proximal His to an Ala residue showed a ~30 fold lower catalytic activity in the ferric state. This observation suggests that the proximal His residue is important for the stabilisation and the influence of oxygen binding. Therefore, ligand interaction is significantly influenced by the proximal His residue (Roy et al., 2014).

The Phe in the inter-helical CD1 region is crucial for providing stability to the haem group (Hargrove et al., 1994). Mutations of the CD1 Phe has been reported to lead to pathologies with detrimental effects. The Hb Louisville mutation [ $\beta$ 42(CD1)Phe→Leu] alters the haem's contact at F42 of the  $\alpha$ 1 $\beta$ 2 dimer causing a reduced oxygen affinity but normal Bohr effect. Sufferers of this unfortunate mutation experience haemolytic crisis and anaemia. Other Hb variants of the CD1 position have been previously reported and are similar to Hb Louisville, for example, Hb Hammersmith [ $\beta$ 42 (CD1)Phe→Ser] and Hb Sendagi [ $\beta$ 42(CD1)Phe→Val]. Therefore, the CD1 Phe residue in the haem pocket is extremely crucial for the stability of globins (Kamseng et al., 2017). Overall the results confirmed that the key amino acids of the haem pocket are certainly the proximal His and distal Gln due to their high frequency with a substantial amount of species containing a Cys in the CD position.

The Essex model sequence proposes that Tyr975 is a potential candidate for the CD1 position. Due to the shift in C and D helices from the reassessment of the Zurich model sequence, the Essex model sequence indicates that a Tyr residue is more likely than a Phe in the CD1 position. Pairwise sequence comparisons using the 362 Adgb sequences showed a high percentage frequency of Tyr residues and approximately one-third of the species containing Phe residues (Fig 3.1.4). The CD1 Tyr residue has been

detected in prokaryotes and was first observed in *Mycobacterium tuberculosis*. The CD1 Tyr residue has also been found in other species such as *Mycobacterium leprae*, *Mycobacterium avium*, *Corynebacterium diphtheriae*, and *Streptomyces coelicolor* (Mukai et al., 2002). The CD1 Tyr in prokaryotes produces a strong hydrogen-bonding network with the B10 Tyr to stabilise the binding of O<sub>2</sub> and CO in the distal haem position (Milani et al., 2003). Furthermore, the presence of a distal Gln in Hb of a parasitic nematode obtains higher oxygen affinity, which is due to the stabilisation of the oxygen molecule by the distal Gln and B10 Tyr which donate two hydrogens (De Baere et al., 1994). However, Adgb does not contain a B10 Tyr and therefore, this interaction with either the CD1 Tyr or distal Gln should be ruled out (Hoogewijs et al., 2012). The findings suggest that the role of the CD1 Tyr may not affect ligand binding due to the absence of the B10 Tyr that contributes to the hydrogen bonding network. The Adgb-HD was used experimentally (in Chapter 4) and was stably expressed using the Essex model sequence (Appendix 11), which contains a CD1 Tyr instead of the Phe. However, attempts to express the HD using the Zurich model sequence (Appendix 10) has previously resulted in unstable expression (Hoogewijs et al., 2012). The results imply that CD1 Tyr may have an important role in the stable expression and perhaps the stability of the overall haem group. Overall, the key residues identified in the Adgb sequence by multiple sequence alignments are the proximal His (F8), the distal Gln (E7), and the CD1 Tyr.

To compare the occurrence of the CD1 Phe and Tyr in Adgb sequences, it was discovered that ~30 % of species contained a Cys residue instead of a CD1 Phe (Fig 3.1.2C) and a Phe instead of the CD1 Tyr (Fig 3.1.4A). Therefore, the frequency of Cys and Phe was examined to determine if they were selective to a particular animal group. For the species that contained a Phe instead of a Tyr residue for the Essex model sequence, the majority of them were selective to birds, a few reptiles and marine animals (Fig 3.1.5A). Birds and reptiles belong to the Sauropsida clade (Agius et al., 2019). In comparison to the original paper by Hoogewijs et al, the Sauropsida clade is missing a Phe in the CD1 position (Hoogewijs et al., 2012). However, the original sequence comparison was conducted using Adgb sequences from fewer sequences than the 362 sequences used in this study which could justify why the CD1 Phe reoccurs more. Therefore, the presence of a Phe residue in the position of the CD1 Tyr is mostly found in birds and reptiles (Sauropsida clade). For the species that contained a Cys residue in place of the Phe in the CD1

position for the Zurich model sequence, a large majority was selective to birds as well as a few reptiles (Fig 3.1.5B), therefore this is common in the Sauropsida clade. In the original study, the frequency of Cys residue was detected in the Sauropsida clade but located in the calpain domain, not the haem domain (Hoogewijs et al., 2012). Therefore, the Cys residue found in the CD1 position in this study belongs in the Sauropsida clade and is an authentic observation as a large number of comparison sequence alignments were completed. The restriction to animal groups has been observed in other globins as well. For example, phyletic distributions of GbY has been found selective to *Xenopus* (Amphibians) and reptiles such as lizard and platypuses, while GbE is only expressed in birds (Dröge and Makołowski, 2011). GbX is selective to marine fish and amphibians but is absent in higher vertebrates (Burmester and Hankeln, 2004). Data compiled from databases of genome and transcriptome indicates that Adgb is generally selective to marine fishes (e.g. Lampreys, Cartilaginous, Ray-finned fishes, lungfishes, and Coelacanth), amphibians, reptiles, birds, and mammals (e.g. Monotremes, Marsupials, and Placentals). Therefore, Adgb has a widespread distribution across all animal groups. Ngb and Cygb (other hexa-coordinated globins) are also commonly present in similar animal groups to Adgb which could suggest that they have phylogenetic associations (Burmester. and Hankeln, 2014).

## 5.2 Phylogenetic distribution of mammalian globins

Hb, Mb, Ngb, Cygb, and Adgb are present in most gnathostome (jawed) vertebrates (Burmester. and Hankeln, 2014). The cladogram in Figure 3.2.1 demonstrates the phylogenetic relationship of Adgb with the other mammalian globin members, demonstrating that Cygb and Mb form a common clade, as do Hb $\alpha$  and Hb $\beta$  which are both penta-coordinated proteins (Fig 3.1.2). This is consistent with previous analyses (Burmester. and Hankeln, 2014). Phylogenetic analysis suggests that the genes of Cygb and Mb would have separated 450 million years ago which was supported by the chromosomal regions for *Cygb* and *Mb* genes containing paralogous stretches of genomic DNA, thought to have originated by an ancient duplication event (Pesce et al., 2002). Ngb and Adgb diverged before the split between Deuterostomia and Protostomia. The remaining members are products of vertebrate-specific duplication events (Storz et al., 2013). Figure 3.1.2 indicated that Ngb and Adgb have distinct clades suggesting early evolutionary

origin and may have evolved from a common ancestor by independent duplication events which is consistent with *Hoogewijs et al* (Hoogewijs et al., 2012). Ngb and Adgb are reported to be hexa-coordinated and during early evolution when atmospheric oxygen levels increased, complex life forms evolved and rendered oxygen supply which triggered the evolution of penta-coordinate globins (Ascenzi et al., 2016). In addition, the last common globin ancestor may have had respiratory roles involving oxygen, which may have been retained in Ngb and Adgb. These globins are expressed at lower levels than Mb and Hb and are restricted to certain cell types/cellular compartments (Burmester and Hankeln, 2004). Therefore, early globins (e.g. Adgb and Ngb) are essential components of the gene repertoire of globin lineages and their loss has been substituted for another well-adapted globin to fulfil essential functions *in vivo* (Burmester. and Hankeln, 2014). From the cladogram (Fig 3.2.1) it can be concluded that Adgb is a part of a monophyletic group and has an early evolutionary origin compared to the other globin members. Adgb does not have a close phylogenetic relationship with the other globin members but does show an affinity to the Ngb lineage as supported by *Hoogewijs et al* (Hoogewijs et al., 2012).

### 5.3 Molecular modelling of Androglobin haem domain

The 3D model of the Essex and Zurich Adgb model sequence was successfully generated (Fig 3.3.1). The models demonstrated an absence of the CD loop but did possess a globin fold like seen in other globin members. The IQ-calmodulin binding motif was initially incorporated however, the position of the motif relative to the haem domain was undetermined. Future studies could try and fit the IQ-calmodulin binding motif to a calmodulin that contains an IQ-sequence for example, calmodulin found in *Homo sapiens* (PDB code: 2F3Y) which could provide more information regarding the 3D structure (Fallon et al., 2005) and how the calmodulin-bound IQ peptide interacts with the rest of the haem domain. The 3D model of the haem domain of Adgb (both Essex and Zurich models) were compared to a monomer of Ngb (Fig 3.3.3). The superimposition in Figure 3.3.3A&B confirms that Adgb has a globin fold similar to Ngb. The similarity and sequence identity of Ngb and Adgb demonstrated that structural homology was less than 20 %. The N-terminus helices of the Essex and Zurich model with Ngb were comparable, suggesting that

F to H helices in the Essex model can produce a globin fold better than that of the Zurich model. This implies that despite being split and circularly permuted Adgb is still capable of forming a globin fold. The effects of permutation have been investigated previously by *Ribero et al* to analyse whether the order of secondary structure segments is critical for protein folding, as well as stability (Ribeiro and Ramos, 2005). A circular permuted Mb was engineered where A-B helices were shifted to the C-terminus, similar to the haem domain of Adgb (Bracke et al., 2018). This study found that permutations when compared to the native Mb protein did not cause large structural changes. In the permuted Mb, longer-range interactions were detected in residues between G and H helices, as well as G and A helices than in the wild type protein. The interactions were necessary for haem stability therefore, the study proposed that permutation puts a strain on these interactive sites causing instability (Ribeiro and Ramos, 2005). Overall, molecular modelling demonstrated the folding of the circularly permuted Adgb-HD based on the Essex model sequence. For further analysis of Adgb, the production of mutant where the IQ-calmodulin binding motif is shifted to the N/C terminus and comparing it to the native state of Adgb may provide more detail about protein folding and stability.

The haem pocket of the Essex model (Fig 3.3.2A) showed that the distal Gln had a bond length of 5.9 Å in which the distal haem residue is comparatively more distant than other distal haem residues in hexacoordinated globins such as the distal His in Ngb (~2.7 Å) and is closer to resembling a pentacoordinate globin (e.g. the bond length for distal Gln in Adgb is similar to the distal His in human Mb which is ~4.5 Å) (Vallone et al. 2004; Hubbard et al., 1990). The 3D haem domain model for the Essex model demonstrated Cys 786 and Cys 977 to be within close proximity to form a surface-exposed disulfide bond in the CD region, similarly to Ngb and Cygb (Fig 3.3.2B). The bond length between the two aforementioned Cys residues was 8.2 Å whereby both Cys residues were comparatively closer to one another than the third Cys residue in the haem domain, Cys 970. Therefore, molecular modelling of the Essex model haem domain predicted a pentacoordinate-like haem geometry and a possible surface disulfide bond formed between Cys 786 and Cys 977.

A 3D model of the haem domain for Adgb has not yet been published, nor has a crystal structure been produced. Therefore, a 3D structure of the haem domain in Adgb was generated using protein templates constructed on I-Tasser and Phyre2 servers that can drive initial template structures closer to their native states (Zhang, 2009). The helices were assembled separately due to the permutation. The helices were assembled on PyMOL according to the Essex model sequence (Appendix 11). PyMOL was selected as the appropriate software to build and modify models, as it has the best facilities for generating protein structures and provides a full python interface for programmers (Lai et al., 2004). Another suitable modelling software that was used was DeepView, which has high-resolution visualisation of complex molecular structures such as Adgb haem domain (Klosowski et al., 2002).

#### 5.4 Absorbance properties of Androglobin and distal mutants

The absorbance spectra for Adgb-HD and distal mutants were characterised on the purified protein (section 4.1). The optical characteristics of three oxidation states (ferric, ferrous, and ferrous-CO) for all proteins were analysed (Fig 4.1.1-4.1.3). For the Adgb-HD ferric state, this showed a spectrum (Fig 4.1.1) typical of a five coordinate, high spin haem iron. Whilst other ferric globins, bands at around 540 and 580 nm are usually seen, low spin haem iron. Such bands here are indistinct indicating a very little low spin haem iron (Sugita et al., 1971). The spectrum for the reduction of the protein indicated partial hexacoordinated geometry, suggesting partial reduction by sodium dithionite. Going forward, the manner of adding dithionite to ensure complete protein reduction should be optimised for example, incubating the protein in sodium dithionite instead of measuring the protein's absorbance immediately after addition of dithionite. For the Q791H Adgb distal mutant, the ferric spectrum was similar to that of Adgb-HD in the ferric state (Fig 4.1.2). The reduction of ferric Q791H Adgb led to a partial shift to form a shoulder at ~420 nm, with lower intensity than that of Adgb-HD protein. For the Q791G Adgb distal mutant (Fig 4.1.3), similar results to the Q791H Adgb mutant were observed. The slightly broad visible region indicates a mix of pentacoordinate and hexacoordinate states for Adgb-HD, Q791H Adgb, and Q791G Adgb in the ferrous state, likely resultant of incomplete reduction by sodium dithionite. The visible region of the ferrous-CO

spectrum for Adgb-HD and distal mutants were typical of most globins (Cuypers et al., 2017). The spectral properties of the ferrous-CO spectrum for Adgb-HD and distal mutants are also similar to the ferrous-CO state observed in Ngb and distal mutation of Ngb (distal His mutated to a Leu) (Dewilde et al., 2001). Overall, no significant change was seen in the distal mutants in comparison to the Adgb-HD CO-bound protein.

The absorption spectra of penta-coordinated ferric haem proteins have been reported to display broad high-spin marker bands at either 600–605 or 630 nm together and the Soret peaks at ~395–409 nm (Boffi et al., 1999). The distal histidine (E7) in penta-coordinate globins is not in direct contact with the iron atom but is still important for ligand stabilisation (Tejero et al., 2015). Histidine coordination to the sixth distal position converts the haem iron to a low spin configuration in both ferric and ferrous oxidation states. The absorbance is stronger in the visible region in the ferric state, and the ferrous visible absorbance band forms two peaks at ~560 and 530 nm (Kakar et al., 2010). Adgb-HD in this case could be a mix between penta-coordinate and hexacoordinate. This observation could be due to the purified protein being unstable and partial reduction by sodium dithionite. An alternative approach would be to use the cell lysate of the expressed protein like used in the original study of Adgb by *Hoogewijs et al* (Hoogewijs et al., 2012).

### 5.5 Measurement of surface-exposed Cysteines in Androglobin haem domain

In the Essex model sequence (Appendix 11) the potential Cys residues in the proposed CD region may form a disulfide bond between Cys786 (D helix) and Cys977 (C helix), similar to the surface-exposed cysteines in the haem domain of Ngb and as speculated by the molecular visualisation of the Essex model sequence (Chapter 3) (Reeder and Ukeri, 2018). The surface-exposed Cys residues were measured by the addition of 4,4'-dithiodipyridine which bound to any free sulfhydryls on the surface of the haem domain (Fig 4.2.1) (Winther and Thorpe, 2014). The experiment was also repeated when cysteines were oxidised (bound in a disulfide bond) for comparison. Figure 4.2.1B showed that there was a 2:1 ratio of thiols per haem. Whereas, the titration without the reduction by TCEP (Fig 4.2.2) showed, no thiols were present. The conclusion for this experiment is that Adgb-HD contains a single disulfide bond.

To ensure the Cys residues were free to bind to dithiodipyridine, Adgb-HD was incubated with TCEP to reduce any disulfide bonds (Burns et al., 1991). Previously DTT has been commonly used as a sulfhydryl reductant, but is affected by the presence of metal chelates such as haem porphyrins and is less stable than TCEP therefore, TCEP was used as a reducing agent in this study (Getz et al., 1999). The results confirmed that for Adgb-HD there are two sulfhydryl's per haem (2:1 ratio) and no free Cys was observed in the titration without TCEP reduction (Fig 4.2.1&4.2.2). The results strongly suggest that Adgb-HD contains a single disulfide bond between Cys786 and Cys977, which are in close proximity to form an intramolecular disulfide bond on the surface of the haem domain.

Cysteine residues present on the surface of the haem domain of globins are important for influencing the haem group's ligand-binding properties (Reeder and Ukeri, 2018). The Cys residues form intramolecular disulfide bonds that stabilise the haem group and regulate the activity of the protein (Le Quyen et al., 2016). Human Ngb contains two important surface-exposed cysteines (Cys-46 and Cys-55) located in the C and D helix (CD loop region) which forms a disulfide bond upon oxidation. Disulfide bond formation in Ngb is accompanied by a decrease in distal histidine binding affinity to the haem iron from ~3000 to 280 ( $k_{on}/k_{off}$  values). The decrease in binding affinity increases the penta-coordinate transition as well as it's affinity to other ligands such as oxygen and NO (Tiso et al., 2011). The cysteine oxidation in Ngb exhibits a higher NiR activity by increasing the protein's affinity to nitrite and when reduced exerts a twofold decrease in NiR activity compared to the reduces cysteines (Reeder and Ukeri, 2018). Providing there is a disulfide bond in Adgb and that there are known influences on ligand binding there is scope to study the effect of Cys on the NiR activity in the future.

## 5.6 NO binding and nitrite reductase activity of Androglobin haem domain and distal mutants.

In this study, NO binding of Adgb-HD and distal mutants; Q791H Adgb and Q791G Adgb were analysed (Fig 4.3.1). Distal mutants were used to compare how the distal position affects NO binding. For Adgb-

HD in the ferrous-NO oxidation state, a Soret peak at 394 nm was observed (Fig 4.3.1). The results confirm that Adgb-HD binds to NO in hypoxic conditions as a pentacoordinate species. This observation is dissimilar to other globins such as Ngb, which are typically in a hexacoordinated form when bound to NO (Soret band at 417 nm and visible bands between 540-570 nm) (Van Doorslaer et al., 2003). As for the distal mutants (Fig 4.3.1), there was no substantial change observed in spectra when NO was bound. It may be hypothesised that NO binds to the proximal side of the haem iron, like that of cytochrome *c'* and does not bind to the distal position, except maybe as an unobserved kinetic intermediate. It would be expected that NO binding to the distal side of the protein would be influenced by the distal amino acid. As NO binding in Adgb is independent of the distal haem amino acid composition. The results support the binding of NO to the proximal side, but this is not definitive. Previous studies of NO binding to the proximal haem position has been reported in cytochrome *c'* (found in *Shewanella frigidmarina* and *Alcaligenes xylosoxidans*) (Hough and Andrew, 2015). The proposed binding of NO in this way is shown schematically in Figure 5.6.1.

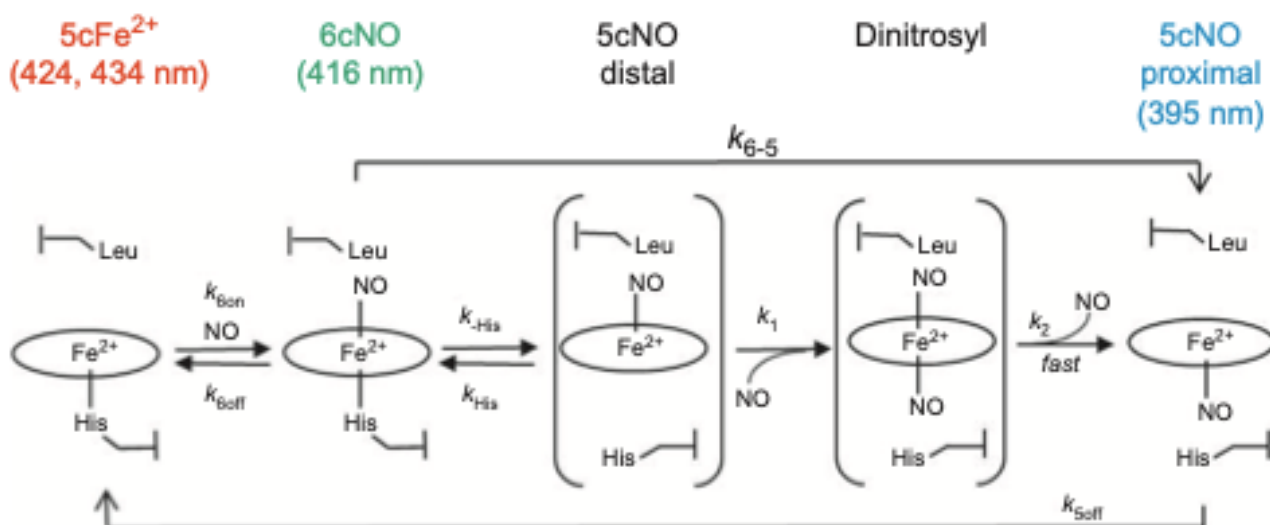


**Figure 5.6.1. Schematic diagram of the mechanism for the formation of the proximal penta-coordinate-NO haem in cytochrome c' of *Alcaligenes xylosoxidans* (contains a distal Leu residue instead of the Gln residue of Adgb-HD).** Absorption maxima of the haem intermediate species are shown above the scheme. The mechanism may be similar to that observed for Adgb-HD therefore, as the spectra and wavelength maxima are similar (Hough and Andrew, 2015).

Figure 5.6.1 shows the proximal five-coordinate-NO formation of cytochrome *c'* in *Alcaligenes xylosoxidans* (contains a distal Leu residue). The Figure shows a biphasic reaction where the six-coordinate-NO is converted to a five-coordinate haem via an extra step associated with NO-Fe<sup>2+</sup>-His. The schematic diagram of the mechanism demonstrates that the proximal five-coordinate-NO species requires two NO molecules and a series of distal haem-NO intermediates. NO binds to the distal haem site which results in dissociation of the proximal His and formation of a dinitrosyl intermediate (five-coordinate intermediate species formed). A second NO molecule competes with the proximal His residue and then binds to the proximal haem position. The distal NO molecule dissociates, resulting in a proximal five-coordinate-NO species with a Soret peak of 395 nm (Hough and Andrew, 2015). This absorption maxima is essentially identical to that observed in this study for the ferrous-NO spectrum. The transition from distal to proximal five-coordinate NO arrangement is believed to be used for selective NO responses (Herzik et al., 2014). Therefore, it can be proposed that NO binding is not regulated by the distal haem position and the haem intermediates may not have been observed by the absorbance spectra, suggesting NO association and dissociation from the distal position could be rapid in Adgb and hence unobserved. An alternative approach would be to use a stopped-flow spectrophotometer, which has been used previously to measure the NiR activity of Cygb and Ngb. This technique could potentially allow for the observation of transient intermediates like the distal-bound NO species (Reeder and Ukeri, 2018). The NiR activity for Adgb-HD and distal mutants; Q791H Adgb and Q791G Adgb were analysed (section 4.4). The optical changes for the NiR activity for Adgb-HD and distal mutants displayed similar peaks (absorbance maximum) at ~390 nm and ~426 nm (Fig 4.4.1). The optical changes were consistent with the NO binding spectra (Fig 4.3.1). The results are characteristic of a penta-coordinate species indicating ferrous-NO is formed (Kim-Shapiro and Gladwin, 2014). Time courses (Fig 4.4.2A) for the NiR activity revealed similar changes in absorbance in amplitude across all three proteins and variable absorbance changes within the first 25 s. The time course was measured at 390 nm to analyse nitrite bound ferrous species of Adgb, which is known to be very short lived in other globins, with the only recorded stable existence found in horse heart Mb at 100 K. Therefore, the initial absorbance change at 390 nm is very unlikely nitrite bound ferrous species.

The NiR rate constants for Adgb-HD and distal mutants were determined and are linear as a function of nitrite concentration (Fig 4.4.3). The second-order rate constant for Adgb-HD was  $24.745 \text{ M}^{-1}\text{s}^{-1}$ , similar to the rate constant for NiR activity reported for Globin X in Zebrafish erythrocytes ( $26.7 \text{ M}^{-1}\text{s}^{-1}$ ), which was recorded at  $37^\circ\text{C}$  as oppose to  $25^\circ\text{C}$  for Adgb-HD. This second-order rate constant is 5-10 fold faster than Mb and T-state Hb rate constant for NiR activity and 25-to-50-fold faster than Ngb and Cygb dimer or monomer with free sulfhydryls. The NiR rate constant of Adgb-HD implies that Adgb has a specialised function as a nitrite reductase during hypoxia. Therefore, Adgb is able to generate NO from nitrite with stronger effects than other globin members but similar to Globin X (Corti et al., 2016) and Cygb with an intramolecular disulfide bond (Reeder and Ukeri, 2018). The distal mutants contained a lower rate constant than Adgb-HD ( $13.392 \text{ M}^{-1}\text{s}^{-1}$  for Q791H Adgb and  $15.583 \text{ M}^{-1}\text{s}^{-1}$  for Q791G Adgb), suggesting a slightly slower NiR activity than that of Adgb-HD however, this was not substantially different. A similar result has been observed for *Arabidopsis* Hbs where the distal haem position was mutated to a Leu; H69L mutant (rate constant of  $10.6 \text{ M}^{-1}\text{s}^{-1}$ ). This study suggested that this slightly lower rate could be from a polarity difference in the distal cavity (Kumar et al., 2016). It has been proposed that the polarity of the haem pocket is important for regulating efficient NO detoxification, as seen in Hb in *M. tuberculosis*. The haem pocket may be facilitated by the polar environment of the oxy-ferric iron, whereby the cavity may provide easy movement access for molecules to the distal haem position (Milani et al., 2004). Therefore, decreased polarity may reduce the movement of nitrous acids to the distal haem position ultimately, decreasing the rate of nitrite reduction (Kumar et al., 2016). The nitrite reductase activity for Adgb distal mutants is greater than that of pentacoordinate Mb mutants H64L and H64V ( $1.1 \times 10^4$  and  $0.35 \text{ M}^{-1}\text{s}^{-1}$ , respectively) (Matsui et al., 1997; Yi et al., 2009). Overall the mutation of the distal haem position to a His/Gly did not have a large impact on the NiR activity in this case. Although, no substantial change was seen in the NiR activity for the distal mutants compared to Adgb-HD. Overall, the results indicate that the distal ligand does not affect NO binding.

The reduction of nitrite to NO by most globins is considered an important physiological reaction. For example, Hb NiR activity has been proposed to regulate blood pressure, hypoxic vasodilation, platelet activation, and cellular resilience to hypoxia (Tejero et al., 2015). Once NO is produced it diffuses into the

plasma and smooth muscle, acting as a signalling molecule through the activation of soluble guanylate cyclase (sGC). The activation of this triggers the production of a second messenger; cGMP, and subsequent inhibition of calcium release from the cell. This results in decreased platelet aggregation. Therefore, in low (hypoxic) oxygen conditions, Hb's role as a nitrite reductase to generate NO is triggered and prevents platelet aggregation (Corti et al., 2016). Globins catalyse the reduction of nitrite to NO according to Equation 3.4.1 and 3.4.2. In the presence of nitrite, the haem iron is oxidised to the ferric oxidation state and produces NO (Equation 4.4.1). Sodium dithionite was used to provide hypoxic conditions like those met by globins (Tiso et al., 2011). The sodium dithionite was added to the protein and nitrite solution separately which were mixed together in the cuvette and absorbance was recorded immediately. The sodium dithionite reduces ferric protein back to the ferrous oxidation state so that there is not a mix of ferrous-NO and ferric protein (Tiso et al., 2011) and to allow NO to bind to the haem (Equation 4.4.2), but does not alter the kinetics of NO formation (Tejero et al., 2015).

NO has been recognised to have severe effects on sperm function including the facilitation of sperm motility. High NO concentrations have been found to correlate with decreased sperm motility resulting in infertility issues in men and compromised spermatogenesis (Yeoman et al., 1998). As established Adgb has high NiR activity and could therefore be linked to NO homeostasis in the testes. Adgb contains a calpain-like domain which is a calcium-dependent cysteine protease whereby it is activated by intracellular  $\text{Ca}^{2+}$  (Chua et al., 2000). In addition, the presence of the IQ-calmodulin binding motif is a  $\text{Ca}^{2+}$  sensing protein that transduces intracellular  $\text{Ca}^{2+}$  signals into biological responses. It is well known that calcium levels are associated with endothelial NO concentrations (Porter et al., 2020). For example, in the testes, NO is produced from L-arginine by oxidation. The reaction is catalysed by NO synthase (NOS), which requires calmodulin to enhance  $\text{Ca}^{2+}$  signalling in Sertoli, Leydig, and germ cells, including spermatozoa (Auharek et al., 2012). It has been found that anti-hypertensive medications such as calcium ion-channel blockers effectively increase motile sperm count, improve spermatogenesis, and reverse male infertility (Shiraishi and Matsuyama, 2018). The original study of Adgb, discovered Adgb expression levels in spermatozoa to have functional importance in spermatogenesis (Hoogewijs et al., 2012). Therefore, the high NiR activity of Adgb implies that Adgb has a role in NO production and may be linked to

spermatogenesis, as well as association with the IQ-calmodulin-binding motif in the haem domain that regulates intracellular  $\text{Ca}^{2+}$ . The NO dioxygenase activity of most globins, consuming NO to form nitrate, are faster than their respective NiR activities (Reeder and Ukeri, 2018). However, NOD activity requires aerobic conditions and a stable oxyferrous protein. Like Ngb, the stability of the oxyferrous Adgb HD is very low (personal communication), suggesting that Adgb does not function as an NOD enzyme. The result of the extremely low oxyferrous stability could potentially promote Adgb as an oxygen sensor, however, the mechanism by which this would operate is unknown.

### 5.7 Expression and truncation of the full Androglobin protein in the pET28a vector.

In this study, the expression of the full-length protein was attempted based on the Essex model re-alignment sequence as the expression of Adgb-HD had been successful using this re-alignment (Appendix 11). The pET28a vector was successfully prepared by double restriction enzyme digestion (Fig 4.5.2) and the gene insert encoding the His-tagged full Adgb protein was successfully exercised from the pcDNA 3.1 (+) vector by triple restriction enzyme digest (Fig 4.5.3). The full Adgb gene was successfully ligated into the pET28a vector (Fig 4.5.4). Expression of the pET28a full Adgb protein was attempted by SDS-PAGE gel electrophoresis (Fig 4.5.5), whereby samples were diluted in cracking buffer and loaded (in duplicate) on to a 12% Acrylamide gel and stained in Coomassie Blue stain. Three samples were run in duplicate, pET28a vector (control), pET28a-HD (control) and, pET28a-full Adgb. The presence of multiple thick bands in each lane, indicating that protein expression for the pET28a full Adgb is possible. The pET28a and pET28a-HD were used at positive controls and for comparison. The results did not show sufficient levels of protein expression over that of the native proteins therefore, expression of the full length protein was not confirmed. An alternative approach would be to purify the protein so that a single band would form corresponding to the molecular weight of the full-length protein and compare that to the cell lysate as well as the positive control (Adgb-HD, purified protein). The recombinant expression of the full-length protein of Adgb has not yet been achieved and the challenges faced in expression by recombinant biology techniques are thought to be because of the permutation in the haem domain of Adgb, as well as the size of the large protein (189.7 kDa). The expression of circularly permuted Mb has been previously

explored by *Ribeiro et al* (Ribeiro and Ramos, 2005). An engineered circular permuted Mb was produced and contained A-B helices that had been shifted to the C-terminus similar to the HD of Adgb. An attempt to express the permuted Mb in *E. coli* resulted in the formation of inclusion bodies and required *in vitro* refolding. Despite containing normal absorbance spectra to Mb WT, the permuted Mb was found to be unstable and formed insoluble aggregates (Bracke et al., 2018). Precipitation (insoluble aggregates) was encountered for expression of Adgb-HD similar to the described study therefore, a low percentage of SDS (strong ionic detergent) was used for protein renaturing (He and Ohnishi, 2017).

A pET28a cloning vector was used in this study and has been widely used for the expression of recombinant protein in *E. coli*. Target genes are easily cloned in to pET plasmids under the control of a strong inducible bacteriophage T7 promoter. Transcription of target genes is completed by T7 RNA polymerase under the *lac* operon control (Mierendorf et al., 2000). The pET28a vector (Appendix 4) can also express a protein with a C-terminal His-tag, which was useful for binding to immobilised metal ions during purification by chromatography, something that was utilised in this project (Bracke et al., 2018). The BL21 DE3 strain of *E. coli* is used commonly in conjunction with pET vectors as it's able to produce T7 polymerase from  $\lambda$ -lysogen DE3 in the host and its expression can be controlled by the IPTG-inducible *lac* UV5 promoter, therefore it was utilised to express Adgb-HD, the full-length protein, and truncated proteins (Miroux and Walker, 1996). Protein expression was induced by IPTG; an analogue of lactose. When IPTG is supplied to the host it induces gene expression (Young and Deiters, 2007). ALA is a universal precursor involved in haem synthesis and is converted to a protoporphyrin IX (haem) therefore, ALA is important for facilitating haem production (Severance and Hamza, 2009). Ferric citrate was also added for cellular iron supplementation during haem synthesis (Ryu et al., 2018).

For further structural analysis the full protein was truncated twice; exactly after the HD and 200 aa after the HD (schematic diagrams in Appendix 8 and 9). The presence of disordered secondary structures in the full protein is proposed to cause improper protein folding during purification. PCR reactions that contained 5 % DMSO were effective and both truncations were successful (Fig 4.6.1). Protein expression was attempted to see if both truncated proteins stably expressed and visualised by SDS-PAGE gel

electrophoresis (Fig 4.6.2) whereby samples were diluted in cracking buffer and loaded (in duplicate) on to a 12% Acrylamide gel and stained in Coomassie Blue stain. Two bands formed above 100 kDa, corresponding to the molecular weights of pET28a-truncated Adgb and pET28a-truncated Adgb(200). The presence of multiple thick bands in each lane, suggestive of protein expression for the truncated proteins is possible however, purifying the protein to reveal a single band would confirm it has expressed stably. To further verify stable protein expression, expression studies would be required as well as protein purification to see if the formation of inclusion bodies would be encountered and if the protein could be purified as a stable protein. The Adgb-HD could be used as a positive control as this has already been successfully expressed and purified. Other routes can be considered for stable expression such as shifting inclusion bodies to a soluble protein state. This has previously been achieved by changing expression temperature, media, *E. coli* strains, and using molecular chaperones to prevent aggregation. Furthermore, another expression system could be considered, one that resembles closer to a mammalian cell such as a baculovirus-insect cell system. Insects as host represent a higher eukaryotic system, capable of complex post-translational modifications to allow better protein folding. As well as, accommodating larger DNA inserts for the production of multi-subunit protein complexes. The baculovirus-insect cell expression system has been widely used before for recombinant expression of high molecular weight proteins (Bracke et al., 2018). As Adgb is considered a high molecular weight protein (189.7 kDa) this could be a relevant option.

## Chapter 6. Conclusion

To conclude, this project identified the proximal His, distal Gln, and CD1 Tyr (Tyr 975) residues as critical amino acids of the haem domain by analysing homology of the residues in 362 Adgb sequences via sequence alignments and measuring the frequency of residues, ultimately highlighting the importance of the CD1 residues and how variations can lead to detrimental effects, indicating that the CD1 residue is vital for maintaining haem stability. The frequency of Cys/Phe770 and Phe/Tyr975 was examined to determine if they were selective to a particular animal group. For the species that contained either a Phe instead of a Tyr for the Essex sequence or a Cys instead of a Phe for the Zurich sequence, a clear majority of the species belong to the Sauropsida clade (observed predominantly in birds and reptiles). The phylogenetic relationship of Adgb with the other mammalian globin members was determined as not having a close phylogenetic relationship with the other globin members but Adgb does show an affinity to the Ngb lineage (as previously reported). Molecular 3D modelling of the haem domain illustrated Adgb to possess a characteristic globin fold, but the 3D haem domain derived from the Essex model appeared to be a slightly better fit than the Zurich model. Additionally, 3D modelling of the haem domain demonstrated a lack of proximity between the haem iron and distal Gln indicating a pentacoordinate haem geometry which was experimentally tested by optical spectra on the purified haem domain protein to reveal a mix of penta-/hexa-coordinated ferrous species. The presence of two surface-exposed Cys close enough to form a disulfide bond (similar to Ngb) was visualised in the 3D model as well, which was experimentally detected in the Adgb-HD protein. The NO binding spectrum suggested that the distal haem position could be independent of NO binding and perhaps the proximal haem position is related to NO binding and dissociation. Overall, the results from the distal mutants showed that the distal ligand does not largely affect NO binding. The rate constants for the NiR activity of Adgb-HD showed Adgb to possess a higher NiR activity than other mammalian globin members and is similar to that of Globin X, indicating a potential specialised function in nitrite reduction during spermatogenesis. The ligation of the full Adgb gene into the pET28a vector and truncation of the full gene was accomplished by molecular cloning techniques however, challenges were still encountered during protein expression. The expression and purification of the full-length protein could allow for further analysis such as expression studies and biochemical

characterisation to be pursued. Overall this study provides a deeper insight into the structure and the biochemical properties of haem domain of Adgb, an unusual globin.

## References

- Agius, J.E., Phalen, D.N., Rose, K. and Eden, J.S., 2019. New Insights into Sauropsid *Papillomaviridae* Evolution and Epizootiology: Discovery of Two Novel Papillomaviruses in Native and Invasive Island Geckos. *Virus Evolution*, **5**(2), pp.1-11.
- Agüero-Chapin, G., Galpert, D., Molina-Ruiz, R., Ancede-Gallardo, E., Pérez-Machado, G., De la Riva, G.A. and Antunes, A., 2020. Graph Theory-Based Sequence Descriptors as Remote Homology Predictors. *Biomolecules*, **10**(1), pp.1-32.
- Akiyama, M., Murayama, S., Yokoi, K., Yanagisawa, T., Hattori, Y., Yamashiro, Y., Eto, Y. and Fujisawa, K., 2006. Hemoglobin Hammersmith [ $\beta$  42 (CD1) Phe→ Ser] Causing Severe Hemolytic Anemia in a Japanese girl. *Pediatric Blood & Cancer*, **47**(6), pp.839-841.
- Altinoz, M.A., Elmaci, I., Ince, B., Ozpinar, A. and Sav, A.M., 2016. Hemoglobins, Hemorphins, and 11p15.5 Chromosomal Region in Cancer Biology and Immunity with Special Emphasis for Brain Tumors. *Journal of Neurological Surgery Part A: Central European Neurosurgery*, **77**(3), pp.247-257.
- Aronson, H.E.G., Hendrickson, W.A. and Royer Jr, W.E., 1994. Quantification of Tertiary Structural Conservation Despite Primary Sequence Drift in The Globin Fold. *Protein Science*, **3**(10), pp.1706-1711.
- Auharek, S.A., Lara, N.L.M., Avelar, G.F., Sharpe, R.M. and Franca, L.R., 2012. Effects of Inducible Nitric Oxide Synthase (iNOS) Deficiency in mice on Sertoli Cell Proliferation And Perinatal Testis Development. *International Journal of Andrology*, **35**(5), pp.741-751.
- Avella, G., Ardiccioni, C., Scaglione, A., Moschetti, T., Rondinelli, C., Montemiglio, L.C., Savino, C., Giuffrè, A., Brunori, M. and Vallone, B., 2014. Engineering the Internal Cavity of Neuroglobin

Demonstrates the Role of the Haem-sliding Mechanism. *Acta Crystallographica Section D: Biological Crystallography*, **70**(6), pp.1640-1648.

Azarov, I., Huang, K.T., Basu, S., Gladwin, M.T., Hogg, N. and Kim-Shapiro, D.B., 2005. Nitric Oxide Scavenging by Red Blood Cells as a function of Hematocrit and Oxygenation. *Journal of Biological Chemistry*, **280**(47), pp.39024-39032.

Baez, E., Echeverria, V., Cabezas, R., Ávila-Rodriguez, M., Garcia-Segura, L.M. and Barreto, G.E., 2016. Protection by Neuroglobin Expression in Brain Pathologies. *Frontiers in Neurology*, **7**, pp.146-156.

Bähler, M. and Rhoads, A., 2002. Calmodulin Signalling via the IQ Motif. *FEBS Letters*, **513**(1), pp.107-113.

Bartram, J., Old, J., Phelan, L., Layton, M. and de la Fuente, J., 2012. Novel Haemoglobin Variant [ $\beta$ 42 Phe $\rightarrow$  Cys]: Hb Little Venice. *Blood*, **120**(21), pp.1013.

Beckerson, P., Reeder, B.J. and Wilson, M.T., 2015. Coupling of Disulfide Bond and Distal Histidine dissociation in Human Ferrous Cytochrome Regulates Ligand Binding. *FEBS Letters*, **589**(4), pp.507-512.

Beckerson, P., Wilson, M.T., Svistunenko, D.A. and Reeder, B.J., 2015. Cytochrome Ligand Binding Regulated by Changing Haem-co-ordination in Response to Intramolecular Disulfide Bond Formation and Lipid Interaction. *Biochemical Journal*, **465**(1), pp.127-137.

Bholah, T.C., Neergheen-Bhujun, V.S., Hodges, N.J., Dyal, S.D. and Bahorun, T., 2015. Cytochrome as a Biomarker in Cancer: Potential Perspective for Diagnosis and Management. *BioMed Research International*, **2015**, pp.1-6.

Bocahut, A., Bernad, S., Sebban, P. and Sacquin-Mora, S., 2009. Relating the Diffusion of Small Ligands in Human Neuroglobin to its Structural and Mechanical Properties. *The Journal of Physical Chemistry B*, **113**(50), pp.16257-16267.

Boffi, A., Das, T.K., Della Longa, S., Spagnuolo, C. and Rousseau, D.L., 1999. Pentacoordinate Hemin Derivatives in Sodium Dodecyl Sulfate Micelles: Model Systems for the Assignment of the Fifth Ligand in Ferric Heme Proteins. *Biophysical Journal*, **77**(2), pp.1143-1149.

Bonamore, A. and Boffi, A., 2008. Flavohemoglobin: Structure and Reactivity. *IUBMB Life*, **60**(1), pp.19-28.

Bracke, A., Hoogewijs, D. and Dewilde, S., 2018. Exploring Three Different Expression Systems for Recombinant Expression of Globins: *Escherichia coli*, *Pichia pastoris* and *Spodoptera frugiperda*. *Analytical Biochemistry*, **543**, pp.62-70.

Bracke, A., Van Raemdonck, G., Peeters, K., Punjabi, U. and Dewilde, S., 2018. Expression Analysis of Androglobin in Human Spermatozoa and Testis Tissue. PhD Thesis, University Antwerp, Belgium.

Brittain, T., Skommer, J., Raychaudhuri, S. and Birch, N., 2010. An Antiapoptotic Neuroprotective Role for Neuroglobin. *International Journal of Molecular Sciences*, **11**(6), pp.2306-2321.

Briz, V. and Baudry, M., 2017. Calpains: Master Regulators of Synaptic Plasticity. *The Neuroscientist*, **23**(3), pp.221-231.

Burmester, T. and Hankeln, T., 2004. Neuroglobin: a Respiratory Protein of the Nervous System. *Physiology*, **19**(3), pp.110-113.

Burmester, T. and Hankeln, T., 2014. Function and Evolution of Vertebrate Globins. *Acta Physiologica*, **211**(3), pp.501-514.

Burns, J.A., Butler, J.C., Moran, J. and Whitesides, G.M., 1991. Selective Reduction of Disulfides by Tris (2-Carboxyethyl) Phosphine. *The Journal of Organic Chemistry*, **56**(8), pp.2648-2650.

Ceccarelli, M. and Scorciapino, M.A., 2012. MD Simulations of Plant Hemoglobins: the Hexa-to Penta-Coordinate Structural Transition. *Biophysical Journal*, **102**(3), pp.1105-1114.

Chen, L.M., Xiong, Y.S., Kong, F.L., Qu, M., Wang, Q., Chen, X.Q., Wang, J.Z. and Zhu, L.Q., 2012. Neuroglobin Attenuates Alzheimer-like Tau Hyperphosphorylation by Activating Akt Signalling. *Journal of Neurochemistry*, **120**(1), pp.157-164.

Chester, N. and Marshak, D.R., 1993. Dimethyl sulfoxide-mediated primer T<sub>m</sub> reduction: A Method for Analyzing the Role of Renaturation Temperature in the Polymerase Chain Reaction. *Analytical Biochemistry*, **209**(2), pp.284-290.

Chua, B.T., Guo, K. and Li, P., 2000. Direct Cleavage by the Calcium-activated Protease Calpain can lead to Inactivation of Caspases. *Journal of Biological Chemistry*, **275**(7), pp.5131-5135.

Corti, P., Xue, J., Tejero, J., Wajih, N., Sun, M., Stolz, D.B., Tsang, M., Kim-Shapiro, D.B. and Gladwin, M.T., 2016. Globin X is a Six-coordinate Globin that Reduces Nitrite to Nitric Oxide in Fish Red Blood Cells. *Proceedings of the National Academy of Sciences*, **113**(30), pp.8538-8543.

Cosby, K., Partovi, K.S., Crawford, J.H., Patel, R.P., Reiter, C.D., Martyr, S., Yang, B.K., Wacławski, M.A., Zalos, G., Xu, X. and Huang, K.T., 2003. Nitrite Reduction to Nitric Oxide by Deoxyhemoglobin Vasodilates the Human Circulation. *Nature Medicine*, **9**(12), pp.1498-1505.

Cuypers, B., Vermeylen, S., Hammerschmid, D., Trashin, S., Rahemi, V., Konijnenberg, A., De Schutter, A., Cheng, C.H.C., Giordano, D., Verde, C. and De Wael, K., 2017. Antarctic Fish Versus Human Cytoglobins—the Same but yet so Different. *Journal of Inorganic Biochemistry*, **173**, pp.66-78.

Das, T.K., Weber, R.E., Dewilde, S., Wittenberg, J.B., Wittenberg, B.A., Yamauchi, K., Van Hauwaert, M.L., Moens, L. and Rousseau, D.L., 2000. Ligand binding in the ferric and Ferrous States of *Paramecium* Hemoglobin. *Biochemistry*, **39**(46), pp.14330-14340.

De Baere, I., Perutz, M.F., Kiger, L., Marden, M.C. and Poyart, C., 1994. Formation of Two Hydrogen Bonds from the Globin to the Heme-linked Oxygen Molecule in *Ascaris* Hemoglobin. *Proceedings of the National Academy of Sciences*, **91**(4), pp.1594-1597.

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), pp.643-651.

Dewilde, S., Kiger, L., Burmester, T., Hankeln, T., Baudin-Creuza, V., Aerts, T., Marden, M.C., Caubergs, R. and Moens, L., 2001. Biochemical Characterization and Ligand Binding Properties of Neuroglobin, a Novel Member of the Globin Family. *Journal of Biological Chemistry*, **276**(42), pp.38949-38955.

Dikshit, K.L., Orii, Y., Navani, N., Patel, S., Huang, H.Y., Stark, B.C. and Webster, D.A., 1998. Site-directed Mutagenesis of Bacterial Hemoglobin: the Role of Glutamine (E7) in Oxygen-binding in the Distal Heme Pocket. *Archives of Biochemistry and Biophysics*, **349**(1), pp.161-166.

Dröge, J. and Makalowski, W., 2011. Phylogenetic Analysis Reveals Wide Distribution of Globin X. *Biology Direct*, **6**(1), pp.1-5.

Dwivedi, A.K., Awasthi, N., Yadav, R., Chaudhary, S., Shukla, A. and Kumar, N., 2019. Spectroscopic Analysis of Porphyrin (C<sub>20</sub>H<sub>12</sub>N<sub>4</sub>) Ring Studied by DFT Methodology. *Journal of Information and Computational Science*, **9**(8), pp. 89-99.

Fago, A., 2017. Functional Roles of Globin Proteins in Hypoxia-tolerant Ectothermic Vertebrates. *Journal of Applied Physiology*, **123**(4), pp.926-934.

Fallon, J.L., Halling, D.B., Hamilton, S.L. and Quiocho, F.A., 2005. Structure of Calmodulin Bound to the Hydrophobic IQ Domain of the Cardiac Cav1. 2 Calcium Channel. *Structure*, **13**(12), pp.1881-1886.

Getz, E.B., Xiao, M., Chakrabarty, T., Cooke, R. and Selvin, P.R., 1999. A Comparison Between the Sulfhydryl Reductants Tris (2-Carboxyethyl) Phosphine and Dithiothreitol for use in Protein Biochemistry. *Analytical Biochemistry*, **273**(1), pp.73-80.

Gleiner, E., Herlyn, H., Zimmerling, S., Burmester, T and Hankeln, T., 2012. Testes-specific Hemoglobins in Drosophila Evolved by a Combination of Sub-and Neofunctionalization after Gene Duplications. *BMC Evolutionary Biology*, **12**(1), pp. 34-46.

Götting, M. and Nikinmaa, M., 2015. More than Hemoglobin—the Unexpected Diversity of Globins In Vertebrate Red Blood Cells. *Physiological Reports*, **3**(2), pp.1-8.

Guimarães, B.G., Hamdane, D., Lechauve, C., Marden, M.C. and Golinelli-Pimpaneau, B., 2014. The Crystal Structure of Wild-type Human Brain Neuroglobin Reveals Flexibility of the Disulfide Bond that Regulates Oxygen Affinity. *Acta Crystallographica Section D: Biological Crystallography*, **70**(4), pp.1005-1014.

Hade, M.D., Kaur, J., Chakraborti, P.K. and Dikshit, K.L., 2017. Multidomain Truncated Hemoglobins: New Members of the Globin Family Exhibiting Tandem Repeats of Globin Units and Domain Fusion. *IUBMB Life*, **69**(7), pp.479-488.

Hamdane, D., Kiger, L., Dewilde, S., Uzan, J., Burmester, T., Hankeln, T., Moens, L. and Marden, M.C., 2005. Hyperthermal Stability of Neuroglobin and Cytoglobin. *The FEBS Journal*, **272**(8), pp.2076-2084.

Hargrove, M.S., Krzywda, S., Wilkinson, A.J., Dou, Y., Ikeda-Saito, M. and Olson, J.S., 1994. Stability of Myoglobin: a Model for the folding of Heme Proteins. *Biochemistry*, **33**(39), pp.11767-11775.

He, C. and Ohnishi, K., 2017. Efficient Renaturation of Inclusion Body Proteins Denatured by SDS. *Biochemical and Biophysical Research Communications*, **490**(4), pp.1250-1253.

He, F., Hendrickson, C.L. and Marshall, A.G., 2000. Unequivocal Determination of Metal Atom Oxidation State in Naked Heme Proteins: Fe (III) Myoglobin, Fe (III) Cytochrome c, Fe (III) Cytochrome b5, and Fe (III) Cytochrome b5 L47R. *Journal of the American Society for Mass Spectrometry*, **11**(2), pp.120-126.

Henderson, S.J., Timbs, A.T., McCarthy, J., Gallienne, A.E., Proven, M., Rugless, M.J., Lopez, H., Eglinton, J., Dziedzic, D., Beardsall, M. and Khalil, M.S., 2016. Ten years of Routine  $\alpha$ - and  $\beta$ -globin Gene Sequencing in UK Hemoglobinopathy Referrals Reveals 60 Novel Mutations. *Hemoglobin*, **40**(2), pp.75-84.

Hendgen-Cotta, U.B., Flögel, U., Kelm, M. and Rassaf, T., 2010. Unmasking the Janus Face of Myoglobin in Health and Disease. *Journal of Experimental Biology*, **213**(16), pp.2734-2740.

Herzik, M.A., Jonnalagadda, R., Kuriyan, J. and Marletta, M.A., 2014. Structural Insights into the Role of Iron-histidine Bond Cleavage in Nitric Oxide-induced Activation of H-NOX Gas Sensor Proteins. *Proceedings of the National Academy of Sciences*, **111**(40), pp.4156-4164.

Hoffmann, F.G., Opazo, J.C. and Storz, J.F., 2011. Whole-genome Duplications Spurred the Functional Diversification of the Globin Gene Superfamily in Vertebrates. *Molecular Biology and Evolution*, **29**(1), pp.303-312.

Hoogewijs, D., Ebner, B., Germani, F., Hoffmann, F.G., Fabrizius, A., Moens, L., Burmester, T., Dewilde, S., Storz, J.F., Vinogradov, S.N. and Hankeln, T., 2012. Androglobin: a Chimeric Globin in Metazoans that is Preferentially Expressed in Mammalian Testes. *Molecular Biology and Evolution*, **29**(4), pp.1105-1114.

Hough, M.A. and Andrew, C.R., 2015. Cytochromes c': Structure, Reactivity and relevance to Haem-based Gas Sensing. In *Advances in Microbial Physiology*, **67**, pp.1-84.

Howes, B.D., Giordano, D., Boechi, L., Russo, R., Mucciacciaro, S., Ciaccio, C., Sinibaldi, F., Fittipaldi, M., Martí, M.A., Estrin, D.A. and di Prisco, G., 2011. The Peculiar Heme Pocket of the 2/2 Hemoglobin of Cold-Adapted *Pseudoalteromonas haloplanktis* TAC125. *JBIC Journal of Biological Inorganic Chemistry*, **16**(2), pp.299-311.

Huang, B., Lu, Y.S., Li, X., Zhu, Z.C., Li, K., Liu, J.W., Zheng, J. and Hu, Z.L., 2014. Androglobin Knockdown Inhibits Growth of Glioma Cell Lines. *International Journal of Clinical and Experimental Pathology*, **7**(5), pp.2179-2184.

Huang, Y. and Wang, K.K., 2001. The Calpain Family and Human Disease. *Trends in Molecular Medicine*, **7**(8), pp.355-362.

Huang, Z., Shiva, S., Kim-Shapiro, D.B., Patel, R.P., Ringwood, L.A., Irby, C.E., Huang, K.T., Ho, C., Hogg, N., Schechter, A.N. and Gladwin, M.T., 2005. Enzymatic Function of Hemoglobin as a Nitrite Reductase that Produces NO under Allosteric Control. *The Journal of clinical investigation*, **115**(8), pp.2099-2107.

Hubbard, S.R., Hendrickson, W.A., Lambright, D.G. and Boxer, S.G., 1990. X-ray Crystal Structure of a Recombinant Human Myoglobin Mutant at 2· 8 Å Resolution. *Journal of Molecular Biology*, **213**(2), pp.215-218.

Isogai, Y., Ota, M., Fujisawa, T., Izuno, H., Mukai, M., Nakamura, H., Iizuka, T. and Nishikawa, K., 1999. Design and Synthesis of a Globin Fold. *Biochemistry*, **38**(23), pp.7431-7443.

Kakar, S., Hoffman, F.G., Storz, J.F., Fabian, M. and Hargrove, M.S., 2010. Structure and Reactivity of Hexacoordinate Hemoglobins. *Biophysical Chemistry*, **152**(1-3), pp.1-14.

Kamseng, P., Trakulsrichai, S., Trachoo, O., Yimniam, W., Panthan, B., Jittorntam, P., Niparuck, P., Sanguanwit, P., Wananukul, W. and Jindadamrongwech, S., 2017. Low Oxygen Saturation and Severe Anemia in Compound Heterozygous Hb Louisville [ $\beta$ 42 (CD1) Phe→ Leu] and Hb La Desirade [ $\beta$ 129 (H7) Ala→ Val]. *Hematology*, **22**(2), pp.114-118.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J., 2015. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nature Protocols*, **10**(6), pp.845-858.

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), pp.662-666.

Kettisen, K., Strader, M.B., Wood, F., Alayash, A.I. and Bülow, L., 2018. Site-directed Mutagenesis of Cysteine Residues Alters Oxidative Stability of Fetal Hemoglobin. *Redox Biology*, **19**, pp.218-225.

Kim-Shapiro, D.B. and Gladwin, M.T., 2014. Mechanisms of Nitrite Bioactivation. *Nitric Oxide*, **38**, pp.58-68.

Kim, J., Fukuda, Y. and Inoue, T., 2019. Crystal structure of Kumaglobin: a Hexacoordinated Heme Protein from an Anhydrobiotic Tardigrade, *Ramazzottius Varieornatus*. *The FEBS Journal*, **286**(7), pp.1287-1304.

Klosowski, J.T., Kirchner, P.D., Valuyeva, J., Abram, G., Morris, C.J., Wolfe, R.H. and Jackman, T., 2002. Deep view: High-resolution Reality. *IEEE Computer Graphics and Applications*, **22**(3), pp.12-15.

Kumar, N., Astegno, A., Chen, J., Giorgetti, A. and Dominici, P., 2016. Residues in the Distal Heme Pocket of Arabidopsis non-symbiotic hemoglobins: implication for nitrite reductase activity. *International journal of molecular sciences*, **17**(5), pp.1-16.

Lai, P., Kaplan, W., Church, W.B. and Wong, R.K., 2004. Informative 3D Visualization of Multiple Protein Structures. In *Proceedings of the Second Conference on Asia-Pacific Bioinformatics*, **29**, pp. 201-208.

Le Quyen, B.T., Hoai, N.T.L. and Van Thanh, N., 2016. Influence of Disulfide Bridge on the Structural Stability of Human Neuroglobin: A Molecular Dynamics Simulation Using Latest Data Entry. *Communications in Physics*, **26**(2), pp.151-157.

Lee, N.P. and Cheng, C.Y., 2004. Nitric Oxide/Nitric Oxide Synthase, Spermatogenesis, and Tight Junction Dynamics. *Biology of Reproduction*, **70**(2), pp.267-276.

Lysiak, J.J., Nguyen, Q.A.T. and Turner, T.T., 2000. Fluctuations in Rat Testicular Interstitial Oxygen Tensions are linked to Testicular Vasomotion: Persistence after repair of Torsion. *Biology of Reproduction*, **63**(5), pp.1383-1389.

Matsui, T., Ozaki, S.I. and Watanabe, Y., 1997. On the Formation and Reactivity of Compound I of the His-64 Myoglobin Mutants. *Journal of Biological Chemistry*, **272**(52), pp.32735-32738.

Meitzler, J.L., Hinde, S., Bánfi, B., Nauseef, W.M. and de Montellano, P.R.O., 2013. Conserved Cysteine Residues provide a Protein-protein Interaction Surface in Dual Oxidase (DUOX) Proteins. *Journal of Biological Chemistry*, **288**(10), pp.7147-7157.

Mierendorf, R.C., Morris, B.B., Hammer, B. and Novy, R.E., 2000. Expression and Purification of Recombinant Proteins using the pET System. *The Nucleic Acid Protocols Handbook*, pp. 947-977.

Milani, M., Pesce, A., Ouellet, Y., Dewilde, S., Friedman, J., Ascenzi, P., Guertin, M. and Bolognesi, M., 2004. Heme-ligand Tunneling in Group I Truncated Hemoglobins. *Journal of Biological Chemistry*, **279**(20), pp.21520-21525.

Milani, M., Savard, P.Y., Ouellet, H., Ascenzi, P., Guertin, M. and Bolognesi, M., 2003. A TyrCD1/TrpG8 Hydrogen Bond Network and a TyrB10—TyrCD1 Covalent Link Shape the Heme Distal Site of *Mycobacterium tuberculosis* Hemoglobin O. *Proceedings of the National Academy of Sciences*, **100**(10), pp.5766-5771.

Minneci, P.C., Deans, K.J., Shiva, S., Zhi, H., Banks, S.M., Kern, S., Natanson, C., Solomon, S.B. and Gladwin, M.T., 2008. Nitrite Reductase Activity of Hemoglobin as a Systemic Nitric Oxide Generator Mechanism to Detoxify Plasma Hemoglobin Produced During Hemolysis. *American Journal of Physiology-Heart and Circulatory Physiology*, **295**(2), pp.743-754.

Miroux, B. and Walker, J.E., 1996. Over-production of Proteins in *Escherichia coli*: Mutant Hosts that Allow Synthesis of Some Membrane Proteins and Globular Proteins at High Levels. *Journal of Molecular Biology*, **260**(3), pp.289-298.

Monod, J., Wyman, J. and Changeux, J.P., 1965. On the Nature of Allosteric Transitions: a Plausible Model. *J Mol Biol*, **12**(1), pp.88-118.

Moore, E.J. and Fasan, R., 2019. Effect of Proximal Ligand Substitutions on the Carbene and Nitrene Transferase Activity of Myoglobin. *Tetrahedron*, **75**(16), pp.2357-2363.

Mukai, M., Savard, P.Y., Ouellet, H., Guertin, M. and Yeh, S.R., 2002. Unique Ligand– Protein Interactions in a New Truncated Hemoglobin from *Mycobacterium tuberculosis*. *Biochemistry*, **41**(12), pp.3897-3905.

Murshudov, G.N., Grebenko, A.I., Barynin, V., Dauter, Z., Wilson, K.S., Vainshtein, B.K., Melik-Adamyanyan, W., Bravo, J., Ferrán, J.M., Ferrer, J.C. and Switala, J., 1996. Structure of the Heme d of *Penicillium vitale* and *Escherichia coli* Catalases. *Journal of Biological Chemistry*, **271**(15), pp.8863-8868.

Nishimura, H. and L'Hernault, S.W., 2017. Spermatogenesis. *Current Biology*, **27**(18), pp. 988-994.

Pennelly, R.R., Tan-Wilson, A.L. and Noble, R.W., 1975. Structural States and Transitions of Carp Hemoglobin. *Journal of Biological Chemistry*, **250**(18), pp.7239-7244.

Perutz, M.F., 1970. Stereochemistry of Cooperative Effects in Haemoglobin: Haem–haem Interaction and the Problem of Allostery. *Nature*, **228**(5273), pp.726-734.

Pesce, A., Bolognesi, M., Bocedi, A., Ascenzi, P., Dewilde, S., Moens, L., Hankeln, T. and Burmester, T., 2002. Neuroglobin and Cytochrome b5. *EMBO Reports*, **3**(12), pp.1146-1151.

Porter, J.J., Jang, H.S., Haque, M.M., Stuehr, D.J. and Mehl, R.A., 2020. Tyrosine Nitration on Calmodulin Enhances Calcium-Dependent Association and Activation of Nitric-Oxide Synthase. *Journal of Biological Chemistry*, **295**(8), pp.2203-2211.

Projecto-Garcia, J., Le Port, A.S., Govindji, T., Jollivet, D., Schaeffer, S.W. and Hourdez, S., 2017. Evolution of Single-domain Globins in Hydrothermal Vent Scale-worms. *Journal of Molecular Evolution*, **85**(5-6), pp.172-187.

Reeder, B.J. and Ukeri, J., 2018. Strong Modulation of Nitrite Reductase Activity of Cytochrome b5 by Disulfide Bond Oxidation: Implications for Nitric Oxide Homeostasis. *Nitric Oxide*, **72**, pp.16-23.

Reeder, B.J., 2017. Redox and Peroxidase Activities of the Hemoglobin Superfamily: Relevance to Health and Disease. *Antioxidants & Redox Signalling*, **26**(14), pp.763-776.

Ribeiro, E.A. and Ramos, C.H., 2005. Circular Permutation and Deletion Studies of Myoglobin Indicate that the Correct Position of its N-terminus is required for Native Stability and Solubility but not for Native-like Heme Binding and Folding. *Biochemistry*, **44**(12), pp.4699-4709.

Rodwell, V.W. and Kennelly, P.J. (2003) Proteins: Myoglobin & Hemoglobin. 26<sup>th</sup> Edition, *Murray, R.K.; Granner, D.K.; Mayes, P.A.*, pp.40-48, New York, Lange Medical Books.

Rosselli, M., Keller, R.J. and Dubey, R.K., 1998. Role of Nitric Oxide in the Biology, Physiology and Pathophysiology of Reproduction. *Human Reproduction Update*, **4**(1), pp.3-24.

Roy, J., Santara, S.S., Bose, M., Mukherjee, S., Saha, R. and Adak, S., 2014. The ferrous–dioxy Complex of Leishmania Major Globin Coupled Heme Containing Adenylate Cyclase: The Role of Proximal Histidine on its Stability. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, **1844**(3), pp.615-622.

Ryu, M.S., Duck, K.A. and Philpott, C.C., 2018. Ferritin Iron Regulators, PCBP1 and NCOA4, Respond to Cellular Iron Status in Developing Red Cells. *Blood Cells, Molecules, and Diseases*, **69**, pp.75-81.

Schneider, T.D., and Stephens, R.M. (1990). <https://weblogo.berkeley.edu/>. *Nucleic Acids Res.* Last accessed: 02.11.2020.

Severance, S. and Hamza, I., 2009. Trafficking of Heme and Porphyrins in Metazoa. *Chemical Reviews*, **109**(10), pp.4596-4616.

Sharma, R. and Agarwal, A., 2011. Spermatogenesis: An Overview. In *Sperm Chromatin* (pp. 19-44). Springer, New York, NY.

Shiraishi, K. and Matsuyama, H., 2018. Effects of Medical Comorbidity on Male Infertility and Comorbidity Treatment on Spermatogenesis. *Fertility and Sterility*, **110**(6), pp.1006-1011.

Song, S., Starunov, V., Bailly, X., Ruta, C., Kerner, P., Cornelissen, A. and Balavoine, G., 2019. Globins and Hemogenic Cells in the Marine Annelid *Platynereis Dumerilii* Shed New Light on Blood Evolution in Bilaterians. *BioRxiv*, **1**, pp.1-46.

Storz, J.F., Opazo, J.C. and Hoffmann, F.G., 2011. Phylogenetic Diversification of the Globin Gene Superfamily in Chordates. *IUBMB Life*, **63**(5), pp.313-322.

Storz, J.F., Opazo, J.C. and Hoffmann, F.G., 2013. Gene Duplication, Genome Duplication, and the Functional Diversification of Vertebrate Globins. *Molecular Phylogenetics and Evolution*, **66**(2), pp.469-478.

Strobl, S., Fernandez-Catalan, C., Braun, M., Huber, R., Masumoto, H., Nakagawa, K., Irie, A., Sorimachi, H., Bourenkow, G., Bartunik, H. and Suzuki, K., 2000. The Crystal Structure of Calcium-free Human m-calpain Suggests an Electrostatic Switch Mechanism for Activation by Calcium. *Proceedings of the National Academy of Sciences*, **97**(2), pp.588-592.

Sugita, Y., Nagai, M. and Yoneyama, Y., 1971. Circular Dichroism of Hemoglobin in Relation to the Structure Surrounding the Heme. *Journal of Biological Chemistry*, **246**(2), pp.383-388.

Sun, Y., Benabbas, A., Zeng, W., Muralidharan, S., Boon, E.M. and Champion, P.M., 2016. Kinetic Control of O<sub>2</sub> Reactivity in H-NOX Domains. *The Journal of Physical Chemistry B*, **120**(24), pp.5351-5358.

Teixeira, K.N., Oliveira, J.S., Souza, K.N., De Moura, J., Brito, C.A., Vidigal, T.H., Santos, A.M. and Santoro, M.M., 2011. Comparison of the biochemical and Molecular Properties of Myoglobins from Three Biomphalaria Species. *Biochemical Systematics and Ecology*, **39**(4-6), pp.581-586.

Tejero, J. and Gladwin, M.T., 2014. The Globin Superfamily: Functions in Nitric Oxide Formation and Decay. *Biological Chemistry*, **395**(6), pp.631-639.

Tejero, J., Sparacino-Watkins, C.E., Ragireddy, V., Frizzell, S. and Gladwin, M.T., 2015. Exploring the Mechanisms of the Reductase Activity of Neuroglobin by Site-directed Mutagenesis of the Heme Distal Pocket. *Biochemistry*, **54**(3), pp.722-733.

Tiso, M., Tejero, J., Basu, S., Azarov, I., Wang, X., Simplaceanu, V., Frizzell, S., Jayaraman, T., Geary, L., Shapiro, C. and Ho, C., 2011. Human Neuroglobin Functions as a Redox-regulated Nitrite Reductase. *Journal of Biological Chemistry*, **286**(20), pp.18277-18289.

Tiso, M., Tejero, J., Kenney, C., Frizzell, S. and Gladwin, M.T., 2012. Nitrite Reductase Activity of Nonsymbiotic Hemoglobins from *Arabidopsis thaliana*. *Biochemistry*, **51**(26), pp.5285-5292.

Tunsaringkarn, T., Soogarun, S., Rungsiyothin, A. and Palasuwan, A., 2012. Inhibitory Activity of Heinz Body Induction *in vitro* Antioxidant Model and Tannin Concentration Of Thai mimosaceous Plant Extracts. *Journal of Medicinal Plants Research*, **6**(24), pp.4096-4101.

Uppal, S., Kumar, A., Shandilya, M., Mukhi, N., Singh, A.K., Kateriya, S., Kaur, J. and Kundu, S., 2016. Penta-and Hexa-coordinate Ferric Hemoglobins Display Distinct pH Titration Profiles Measured by Soret Peak Shifts. *Analytical Biochemistry*, **510**, pp.120-128.

Vallone, B., Nienhaus, K., Matthes, A., Brunori, M. and Nienhaus, G.U., 2004. The Structure of Carbonmonoxy Neuroglobin Reveals a Heme-sliding Mechanism for Control of Ligand Affinity. *Proceedings of the National Academy of Sciences*, **101**(50), pp.17351-17356.

Van Doorslaer, S., Dewilde, S., Kiger, L., Nistor, S.V., Goovaerts, E., Marden, M.C. and Moens, L., 2003. Nitric Oxide Binding Properties of Neuroglobin a Characterization by EPR and Flash Photolysis. *Journal of Biological Chemistry*, **278**(7), pp.4919-4925.

Vinogradov, S.N. and Moens, L., 2008. Diversity of Globin Function: Enzymatic, Transport, Storage, and Sensing. *Journal of Biological Chemistry*, **283**(14), pp.8773-8777.

Vinogradov, S.N., Hoogewijs, D., Bailly, X., Arredondo-Peter, R., Gough, J., Dewilde, S., Moens, L. and Vanfleteren, J.R., 2006. A Phylogenomic Profile of Globins. *BMC Evolutionary Biology*, **6**(1), pp 1-17.

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), pp.132-142.

Wang, Y., Zheng, C.G., Jiang, Y., Zhang, J., Chen, J., Yao, C., Zhao, Q., Liu, S., Chen, K., Du, J. and Yang, Z., 2012. Genetic Correction of  $\beta$ -thalassemia Patient-specific iPS Cells and its use in Improving Hemoglobin Production in Irradiated SCID Mice. *Cell Research*, **22**(4), pp.637-648.

Winther, J.R. and Thorpe, C., 2014. Quantification of Thiols and Disulfides. *Biochimica et Biophysica Acta (BBA)-General Subjects*, **1840**(2), pp.838-846.

Woody, R.W. and Hsu, M.C., 1971. Origin of the Heme Cotton Effects in Myoglobin and Hemoglobin. *Journal of the American Chemical Society*, **93**(14), pp.3515-3525.

Yang, J and Zhang, Y. (2015). <https://zhanglab.ccmb.med.umich.edu/I-TASSER/> . *Nucleic Acids Research*. Last accessed: 02.11.2020.

Yeh, S.R., Couture, M., Ouellet, Y., Guertin, M. and Rousseau, D.L., 2000. A Cooperative Oxygen Binding Hemoglobin from *Mycobacterium tuberculosis* Stabilization of Heme Ligands by a Distal Tyrosine Residue. *Journal of Biological Chemistry*, **275**(3), pp.1679-1684.

Yeoman, R.R., Jones, W.D. and Rizk, B.M., 1998. Evidence for Nitric Oxide Regulation of Hamster Sperm Hyperactivation. *Journal of Andrology*, **19**(1), pp.58-64.

Yi, J., Heinecke, J., Tan, H., Ford, P.C. and Richter-Addo, G.B., 2009. The Distal Pocket Histidine Residue in Horse Heart Myoglobin Directs the O-binding Mode of Nitrite to the Heme Iron. *Journal of the American Chemical Society*, **131**(50), pp.18119-18128.

Yi, J., Orville, A.M., Skinner, J.M., Skinner, M.J. and Richter-Addo, G.B., 2010. Synchrotron X-ray-induced Photoreduction of Ferric Myoglobin Nitrite Crystals gives the Ferrous Derivative with Retention of the O-bonded Nitrite Ligand. *Biochemistry*, **49**(29), pp.5969-5971.

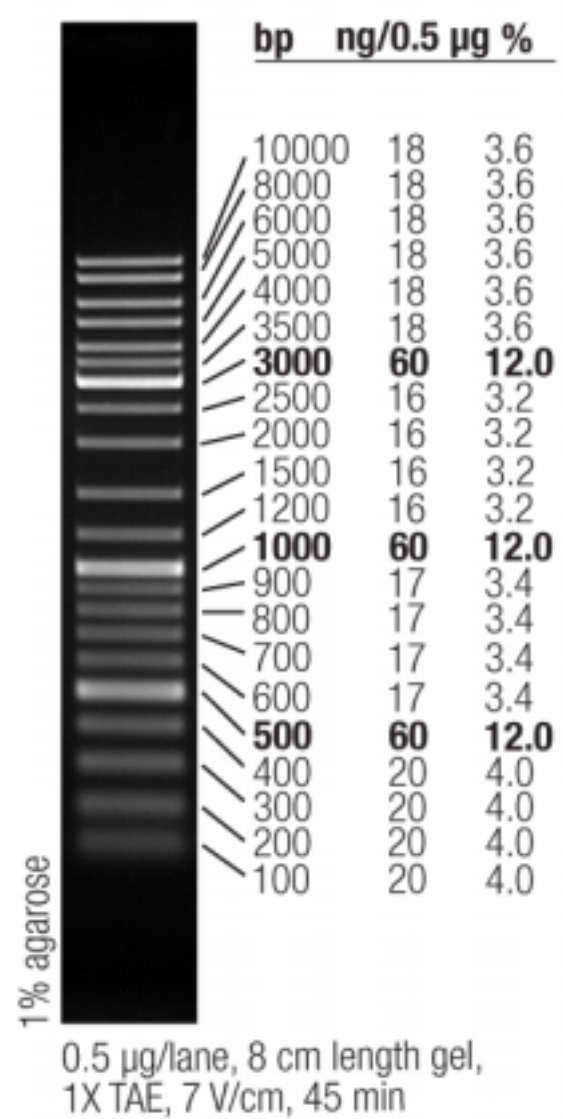
Young, D.D. and Deiters, A., 2007. Photochemical Activation of Protein Expression in Bacterial Cells. *Angewandte Chemie International Edition*, **46**(23), pp.4290-4292.

Zhang, Y., 2009. I-TASSER: Fully Automated Protein Structure Prediction in CASP8. *Proteins: Structure, Function, and Bioinformatics*, **77**(9), pp.100-113.

Zutt, R., Van Der Kooi, A.J., Linthorst, G.E., Wanders, R.J.A. and De Visser, M., 2014. Rhabdomyolysis: Review of the Literature. *Neuromuscular Disorders*, **24**(8), pp.651-659.

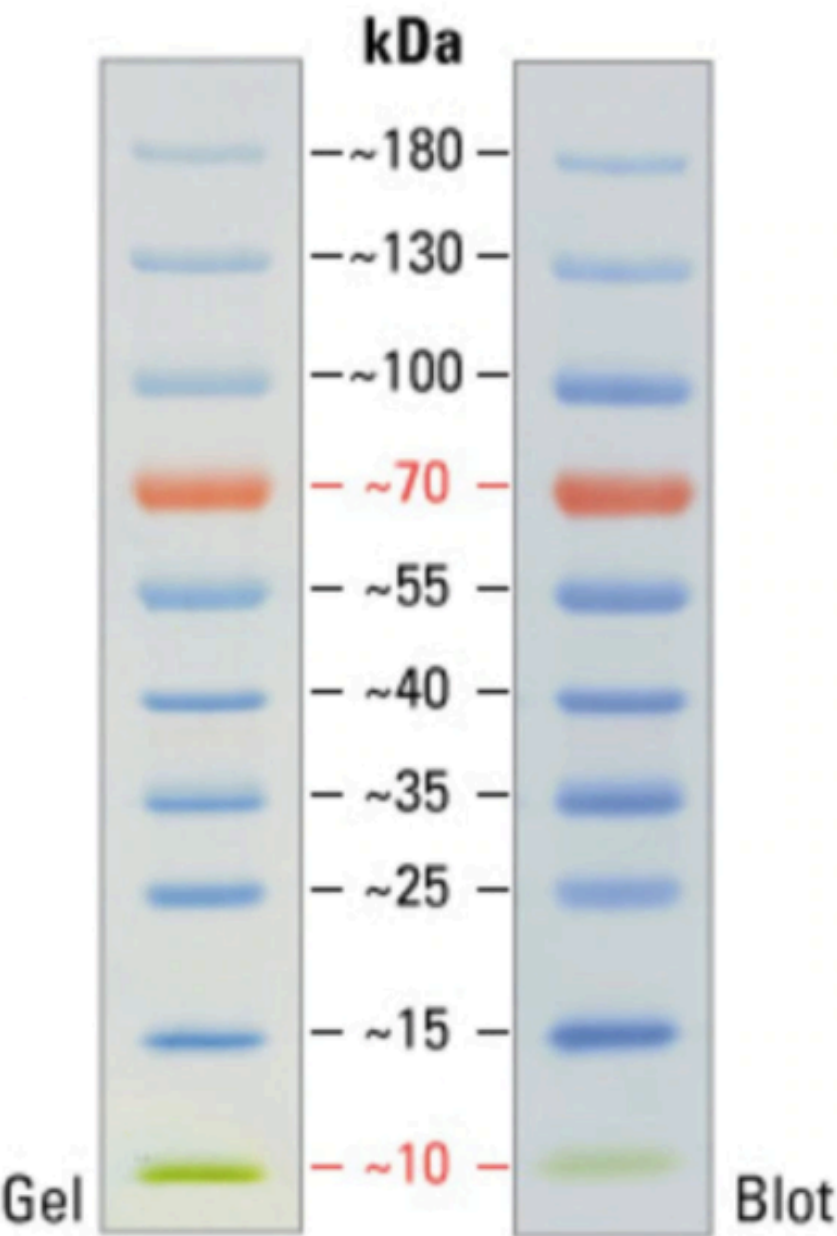
**Appendix 1. 10 Kb generuler DNA ladder Mix**

The GeneRuler was supplied by Thermo Fisher Scientific SM0333 for the approximate quantification of DNA band sizes on an agarose gel. It contains three distinctive reference bands at 3000, 1000 and 500 bp for easy orientation.



**Appendix 2. 180 kDa Thermo Fisher Pageruler prestained protein ladder**

The prestained protein ladder (10 to 180 kDa) was supplied by Thermo Fisher scientific 26616 for the approximate quantification of protein band sizes on an SDS-PAGE gel. The band profile contains two distinctive bands; ~70 kDa (red), ~10 kDa (green) and, the rest of the other bands are blue.



### Appendix 3. Restriction enzyme digestion reaction volume, concentration and reagents.

**Table I.** The volume, concentration and, reagents for the double restriction enzyme digest of PET28a vector using BamHI and NotI.

| Reaction components | Control (μl) | BamHI control (μl) | NotI control (μl) | BamHI and NotI (μl) |
|---------------------|--------------|--------------------|-------------------|---------------------|
| DNA (1μg/μl)        | 8.9          | 8.9                | 8.9               | 8.9                 |
| BamHI RE (10 U/μl)  | -            | 2                  | -                 | 2                   |
| NotI RE (10 U/μl)   | -            | -                  | 2                 | 2                   |
| EcoRI buffer        | 2            | 2                  | 2                 | 2                   |
| H <sub>2</sub> O    | 9.1          | 7.1                | 7.1               | 5.1                 |
| Total (μl)          | 20           | 20                 | 20                | 20                  |

Each reaction was centrifuged for 10 s and incubated 3-4 h at 37 °C in an Eppendorf Thermomixer block heater.

**Table II.** The volume, concentration and, reagents for the triple restriction enzyme digest of pcDNA 3.1 His-tagged full Adgb insert using AseI, BamHI and, NotI.

| Reaction components | Control (μl) | BamHI control (μl) | AseI control (μl) | NotI control (μl) | BamHI, AseI and NotI (μl) |
|---------------------|--------------|--------------------|-------------------|-------------------|---------------------------|
| DNA (1μg/μl)        | 5.8          | 5.8                | 5.8               | 5.8               | 5.8                       |
| BamHI RE (10 U/μl)  | -            | 2                  | -                 | -                 | 2                         |
| NotI RE (10 U/μl)   | -            | -                  | -                 | 2                 | 2                         |
| AseI RE (10 U/μl)   | -            | -                  | 2                 | -                 | 2                         |
| EcoRI buffer        | 2            | 2                  | 2                 | 2                 | 2                         |
| H <sub>2</sub> O    | 12.2         | 10.2               | 10.2              | 10.2              | 6.2                       |
| Total (μl)          | 20           | 20                 | 20                | 20                | 20                        |

Each reaction was centrifuged for 10 s and incubated 3-4 h at 37 °C in an Eppendorf Thermomixer block heater.

**Table III.** The volume, concentration and, reagents for nucleic acid dephosphorylation of the linearized PET28a Vector using alkaline phosphatase.

| Reaction components                                         | Volumes ( $\mu$ l) |
|-------------------------------------------------------------|--------------------|
| 1 $\mu$ g of Linear DNA                                     | 8.8                |
| 10X reaction buffer for AP                                  | 2                  |
| Fast AP thermosensitive alkaline phosphatase (1 U/ $\mu$ l) | 1                  |
| H <sub>2</sub> O                                            | 8.2                |
| Total ( $\mu$ l)                                            | 20                 |

The reaction was centrifuged and incubated at 37 °C for 10 min. The reaction was stopped by heating the reaction at 75 °C for 5 min.

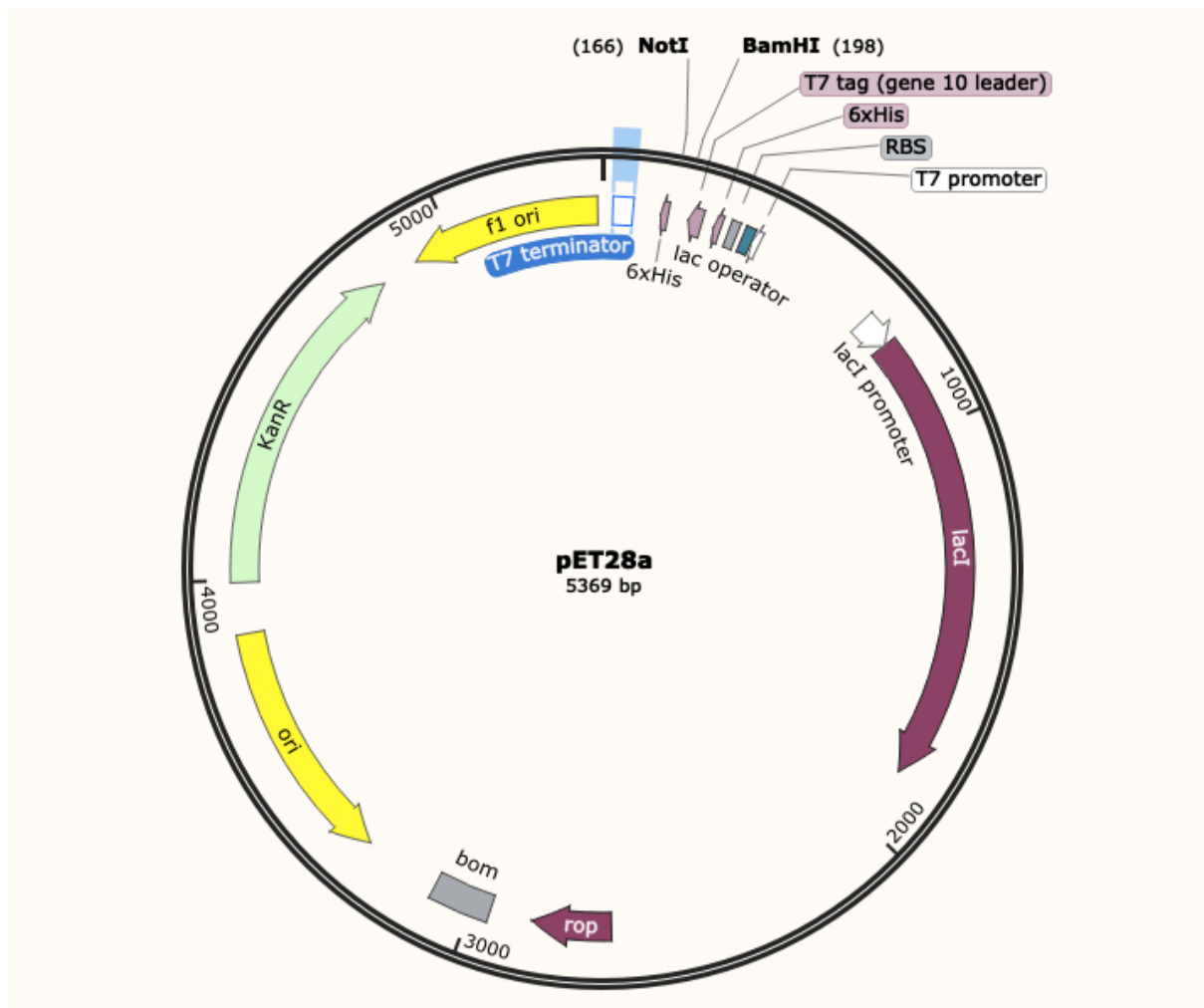
**Table IV.** The volume, concentration and, reagents for the single restriction enzyme digest of the full Adgb in the PET28a vector using BamHI.

| Reaction components | Control-circular DNA (μl) | Single digest with BamHI (μl) |
|---------------------|---------------------------|-------------------------------|
| DNA (1μg/μl)        | 10.8                      | 10.8                          |
| BamHI RE (10 U/μl)  | -                         | 1                             |
| 10X BamHI buffer    | 2                         | 2                             |
| H <sub>2</sub> O    | 7.2                       | 6.2                           |
| Total (μl)          | 20                        | 20                            |

Each reaction was centrifuged for 10 s and incubated 3-4 h at 37 °C in an Eppendorf Thermomixer block heater.

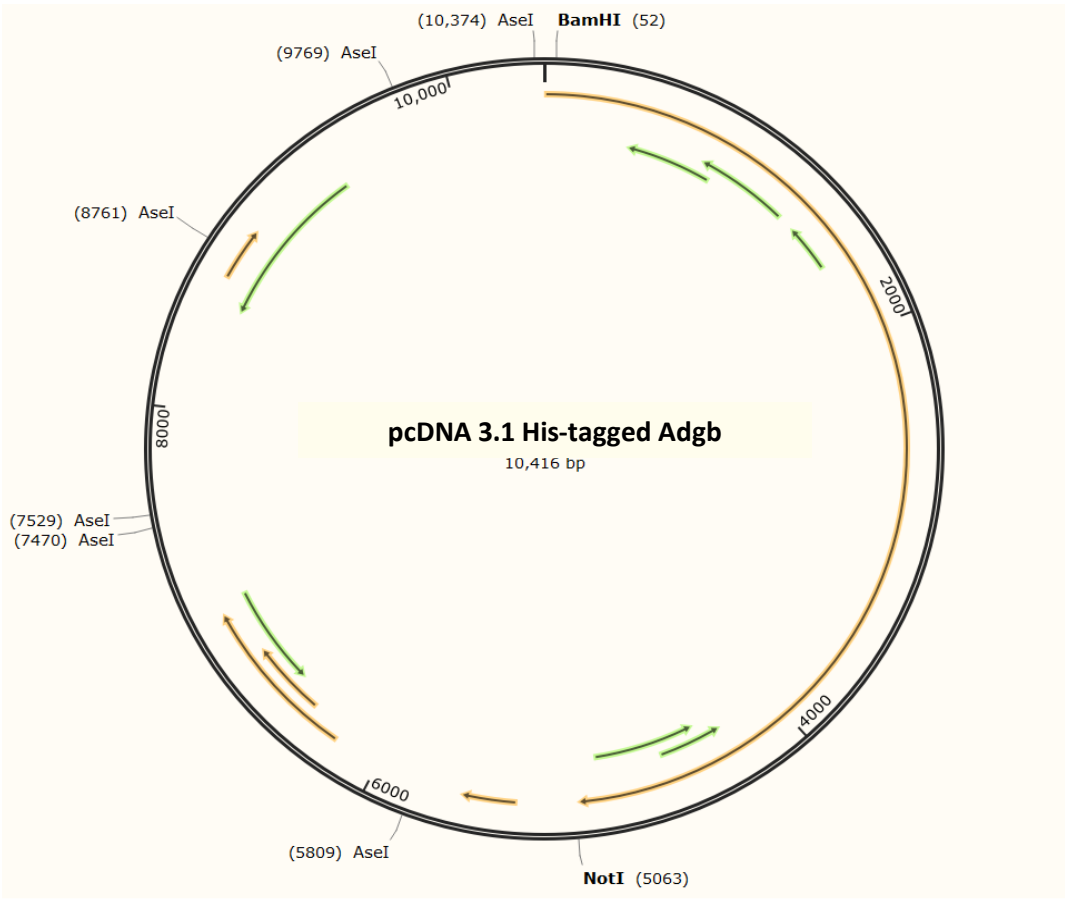
#### Appendix 4. pET28a vector plasmid map

The pET28a vector is shown below to be 5369 bp. The plasmid was developed using SnapGene viewer software. The plasmid shows the location of restriction enzyme sites; BamHI and NotI used to prepare the plasmid by double restriction enzyme digest for ligation with the full Adgb gene insert. The pET28a vector contains a kanamycin resistance gene (green), the origin of replication sites (yellow), bom site (the basis of mobility region) in grey, rop protein site for maintaining low copy plasmids (purple), LacI gene (purple), and LacI promoter (white) that encodes the Lac repressor protein used for protein induction, T7 promoter (white) for T7 RNA polymerase, RBS (ribosome binding site) in blue and, a 6xHis affinity tag (pink) used for specific purification.



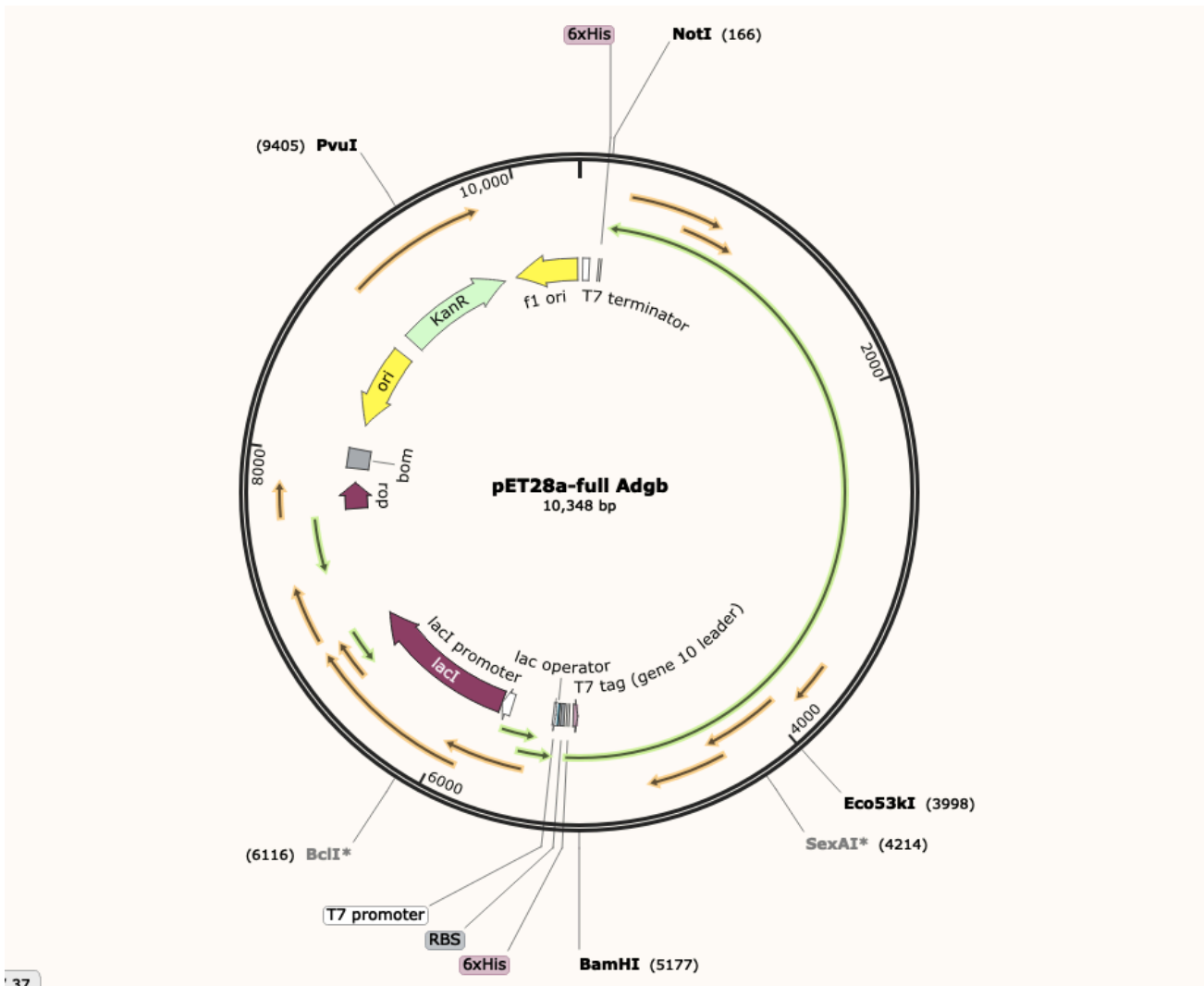
**Appendix 5. pcDNA 3.1 His-tagged full Androglobin plasmid map.**

The pcDNA 3.1 (+) His-tagged full Adgb plasmid is shown below to be ~10,416 bp. The plasmid was developed using SnapGene viewer software. The plasmid shows the location of restriction enzyme sites; BamHI, NotI and, AseI used to cut the Adgb gene out of the plasmid by triple restriction enzyme digest. The pcDNA 3.1 (+) His-tagged full Adgb plasmid contains an Ampicillin resistance gene (green) and the gene for Adgb (yellow).



**Appendix 6. pET28a-full Androglobin recombinant plasmid map**

The pET28a-full Adgb recombinant plasmid is shown below to be ~10,348 bp. The plasmid was developed using SnapGene viewer software. The plasmid shows the location of restriction enzyme sites; BamHI and NotI where the full Adgb gene insert was ligated into the vector. The recombinant plasmid contains a Kanamycin resistance gene (green) and the gene for the full-length Adgb (green) that has a 6xHis tag.



## Appendix 7. The full Androglobin amino acid sequence

The full Adgb amino acid sequence (>NP\_078970.3, *Homo sapiens*) downloaded from the NCBI website (<https://www.ncbi.nlm.nih.gov/>). The globin domain is highlighted in bold as well as the significant residues present in the globin domain. The sequence shows the distal Gln (blue) and the proximal histidine (pink). The full sequence is 1667 aa and the protein is 189.7 kDa, determined by ExPASy ProtParam bioinformatic resource tool (<https://web.expasy.org/cgi-bin/protparam/protparam>).

MASKQTKKKEVHRINSAHGSDDKSKDFYPFGSNVQSGSTEQKKGKFPLWPEWSEADINSEKWDAGKGAK  
DKTGKSPVFHFFEDPEGKIELPPSLKIYSWKRPQDILFSQTPVVVKNEITFDLFSANEHLLCSELMRWII  
SEIYAVWKIFNGGILSNYFKGTSGEPLLPWKPEHIYSLCKAVKGHMPLFNSYGYVVKLYWMGCWRKI  
TIDDFLPFDEDNNLLLPATTYEFELWPMLLSKAIKLANIDIHVADRRELGEFTVIHALTGWLPEVISLH  
PGYMDKVWELLKEILPEFKLSDEASSESKIAVLDSKLKEPGKEGKEGKEIKDGKEVKDVKEFKPESSLT  
LKAPEKSDKVPKEKADARDIGKKRSKDGEKEKFKFSLHGSRPSSEVQYSVQSLSDCSSAIQTSHMVVYAT  
FTPLYLFFENKIFSLEKMADSAEKLREYGLSHICSHPVLVTRSRSCPLVAPPKPPPLPPWKLIRQKKETVI  
TDEAQELIVKKPERFLEISSPFLNYRMTPFTIPTMHFVRS LIKKGI PP GSDLPSVSETDETATHSQTDL  
SQITKATSQGNTASQVILGKGTDEQTDGFLGDAHQSDGLNLEREIVSQTTATQEKSQEELPTTNNSVSKE  
IWLDFEDFCVCFQNIYIFHKPSSYCLNFQKSEFKFSEERVSYLLFVDSLKPIELLVCFSALVRWGEYGAL  
TKDSPPIEPGLLTAETFSWKS LKPGSLVLKIHTYATKATVVRLPVGRHMLLFNAYSPVGHSIHICSMVSF  
VIGDEHVLPN**FEPESCRFTEQ****SLLIMKAIGNVIANFKDKGKLSAALKDLQTA****YPVPFHDKELTAQHFR**  
**VFHLSLWRLMKKVQITKPPPNFKFAFRAMVLDLELLNSSL EEVSLVEWLDVKYCMPTSDKEYSAEEVAAA**  
**IKIQAMWRGTYVRLMKARIPDTKENISVADTLQKVWAVLEMNLEQYAVSLLRLMFKSKCKSLESYPCYQ**  
**DEETK**IAFADYTVTYQEQQPNSWFIVFRETFVLVHQMILVPKVYTTLPICILHIVNNDTMEQVPKVFQKV  
VPYLYTKNKKGYTFVAEAFTGDTYVAASRWKLRLIGSSAPLPCLSRDSPCNSFAIKEIRDYYIPNDKKIL  
FRYSVKVLTPQPATIQVRTSKPDAFIKLQVLENEETMVSSSTGKGQAIIPAFHFLKSEKGLSSQSSKHILS  
FHSASKKEQEVYVKKKAAQGIQKSPKGRAVSAIQDIGLPLVEEETTSTPTREDSSSTPLQNYKYIIQCSV  
LYNSWPLTESQLTFVQALKDLKKSNTKAYGERHEELINLGSPDSHTISEGQKSSVTSKTTRKGKEKSSEK  
EKTAKEKQAPRFEPQISTVHPQQEDPNKPYWILRLVTEHNESELFVKKDTERADEIRAMQAWETTEPG  
RAIKASQARLHYLSGFIKKTSDAESPPISESQTKPKEEVETAARGVKEPNSKNSAGSESKEMTQTGSGSA  
VWKKWQLTKGLRDVAKSTSSESGGVSSPGKEEREQSTRKENIQTGPRTRSPTILETSPRLIRKALEFMDL  
SQYVRKTDTDPLLQTDLELNQQQAMQAAEIIHQFRQHRTRVLSIRNIDQEERLKLKDEVLDMYKEMQDSDL  
EARQKIFDIREEYRNKLLEAEHLKLETLAAQEAAMKLETEKMTAPADTQKKKKGKKK

## Appendix 8. The truncated Androglobin haem domain amino acid sequence

The C-terminus and globin domain of Adgb amino acid sequence (>NP\_078970.3, *Homo sapiens*) downloaded from the NCBI website (<https://www.ncbi.nlm.nih.gov/>). The full Adgb sequence was truncated by designing primers to apply a TAA stop codon exactly after the HD (highlighted in yellow). The globin domain is highlighted in bold. The sequence is 985 aa and is 112.1 kDa, determined by Expasy ProtParam bioinformatic resource tool (<https://web.expasy.org/cgi-bin/protparam/protparam>).

```

MASKQTKKKEVHRINSAHGSDKSKDFYPFGSNVQSGSTEQKKGKFPLWPEWSEADINSEKWDAGKGAKEK
DKTGKSPVFHFFEDPEGKIELPPSLKIYSWKRPQDILFSQTPVVVKNEITFDLFSANEHLLCSELMRWII
SEIYAVWKIFNGGILSNYFKGTSGEPLLPWKWPWEHIYSLCKAVKGHMPLFNSYGKYVVKLYWMGCWRKI
TIDDFLPFDEDNLLLPATTYEFELWPMLLSKAI I KLANIDIHVADRRELGEFTVIHALTGWLPEVISLH
PGYMDKVWELLKEILPEFKLSDEASSES K IAVLDSKLKEPGKEGKEGKEIKDGKEVKDVKEFKPESSLTT
LKAPEKSDKVPKEKADARDIGKKRSKDGEKEKFKFSLHGSRPSSEVQYSVQSLSDCSSAIQTSHMVVYAT
FTPLYLFFENKIFSLEKMADSAEKLREYGLSHICSHPVLVTRSRSCPLVAPPKPPPLPPWKLI R QKKETVI
TDEAQELIVKKPERFLEISSPFLNYRMTPTFTIPTMHFVRSLIKKGIPPGSDLPSVSETDETATHSQTDL
SQITKATSQGNTASQVILGKGTDEQTD FGLGDAHQSDGLNLEREIVSQTTATQEKSQEELPTTNNSVSKE
IWLDFEDFCVCFQNIYIFHKPSSYCLNFQKSEFKFSEERVSYL FVDSLKPIELLVCF SALVRWGEYGAL
TKDSPPIEPGLLTAETFSWKS LKPGSLVLKIHTYATKATVVR L PVGRHMLLFNAYSPVGHSIHICSMVSF
VIGDEHVVL PNFEPESCRFTEQSLLIMKAIGNVIANFKDKGKLSAALKDLQTAHYVPVPFHDKELTAQHFR
VFHLSLWRLMKKVQITKPPPNFKFAFRAMVLDLELLNSSLEEVSLVEWLDVKYCMPTSDKEYSAEEVAAA
IKIQAMWRGTYVRLLMKARIPDTKENISVADTLQKVWAVLEMNLEQYAVSLLRLMFKSKCKSLESYPCYQ
DEETK

```

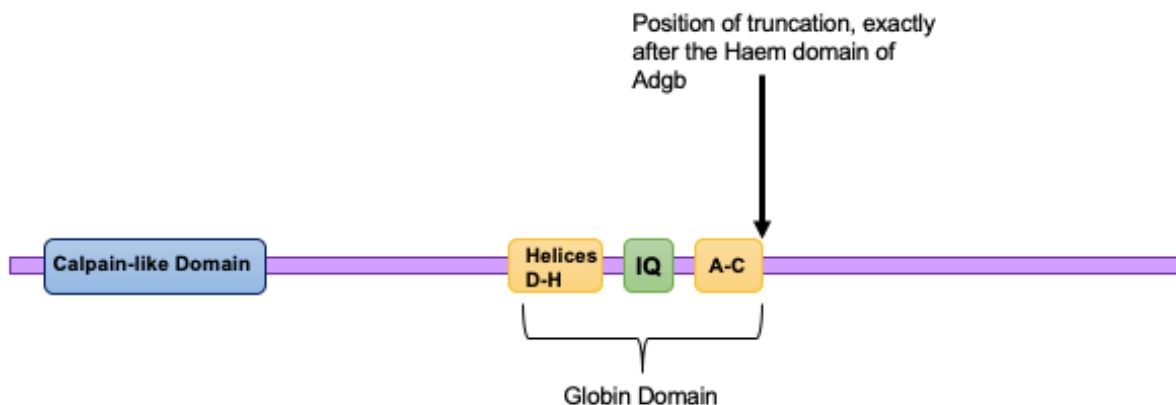


**Figure I. A schematic diagram of Adgb as well as, the position of truncation indicated by a black arrow.** The diagram shows the calpain-like domain (blue) and the 8  $\alpha$ -helices (orange) in the haem domain that is split by an IQ-calmodulin binding motif (green).

## Appendix 9. The truncated Androglobin haem domain amino acid sequence

The amino acid sequence of the C-terminus and 200 aa after the HD for Adgb (>NP\_078970.3, *Homo sapiens*) downloaded from the NCBI website (<https://www.ncbi.nlm.nih.gov/>). The full Adgb sequence was truncated by designing primers to apply a TAA stop codon 200 aa after the HD (highlighted in yellow). The globin domain is highlighted in bold. The sequence is 1185 aa and is 134.9 kDa, determined by Expsy ProtParam bioinformatic resource tool (<https://web.expasy.org/cgi-bin/protparam/protparam>).

```
MASKQTKKKEVHRINSAHGSDKSKDFYPFGSNVQSGSTEQKKGKFPLWPEWSEADINSEKWDAGKGAKEK
DKTGKSPVFHFFEDPEGKIELPPSLKIYSWKRPQDILFSQTPVVVNEITFDLFSANEHLLCSELMRWII
SEIYAVWKIFNGGILSNYFKGTSGEPLLPWKPWEHIYSLCKAVKGHMPLFNSYGKYVVKLYWMGCWRKI
TIDDFLPFDEDNNLLLPATTYEFELWPMLLSKAI IKLANIDIHVADRRELGEFTVIHALTGWLPEVISLH
PGYMDKVWELLKEILPEFKLSDEASSESIAVLDSKLKEPGKEGKEGKEIKDGKEVKDVKEFKPESSLTT
LKAPEKSDKVPKEKADARDIGKKRSKDGEKEKFKFSLHGSRPSSEVQYSVQSLSDCSSAIQTSHMVVYAT
FTPLYLFFENKIFSLEKMADSAEKLREYGLSHICSHPVLVTRSRSCPLVAPPKPPPLPPWKLIQKKETVI
TDEAQELIVKKPERFLEISSPFLNYRMTPTFTIPTEMHFVRS LIKKGIPPGSDLPSVSETDETATHSQTDL
SQITKATSQGNTASQVILGKGTDEQTDGFLGDAHQSDGLNLEREIVSQTTATQEKSQEELPTTNNVSKE
IWLDFEDFCVCFQNIYIFHKPSSYCLNFQKSEFKFSEERVSYLFFVDSLKPIELLVCF SALVRWGEYGAL
TKDSPPIEPGLLTAETFSWKS LKPGSLVLKIHTYATKATVVRLPVGRHMLLFNAYSPVGHSIHICSMVSF
VIGDEHVLPNFEPESCRFTEQSLLIMKAIGNVIANFKDKGKLSAALKDLQTAHYVPVFHDKELTAQHFR
VFHLSLWRLMKKVQITKPPPNFKFAFRAMVLDLELLNSSLEEVS LVEWLDVKYCMPTSDKEYSAEEVAAA
IKIQAMWRGTYVRLLMKARIPDTKENISVADTLQKVWAVLEMNLEQYAVSLLRLMFKSKCKSLESYPCYQ
DEETKIAFADYTVTYQEQQPNSWFI VRETFLVHQDMILVPKVYTTLPICILHIVNNDTMEQVPKV FQKV
VPYLYTKNKKGYTFVAEAFTGDTYVAASRWKLRLIGSSAPLPCLSRDSPCNSFAIKEIRDYYIPNDKKIL
FRYSVKVLTPQPATIQVRTSKPDAFIKLQVLENEETMVSS TGGKQAIIPAFHFLKSEKGLSSQS
```

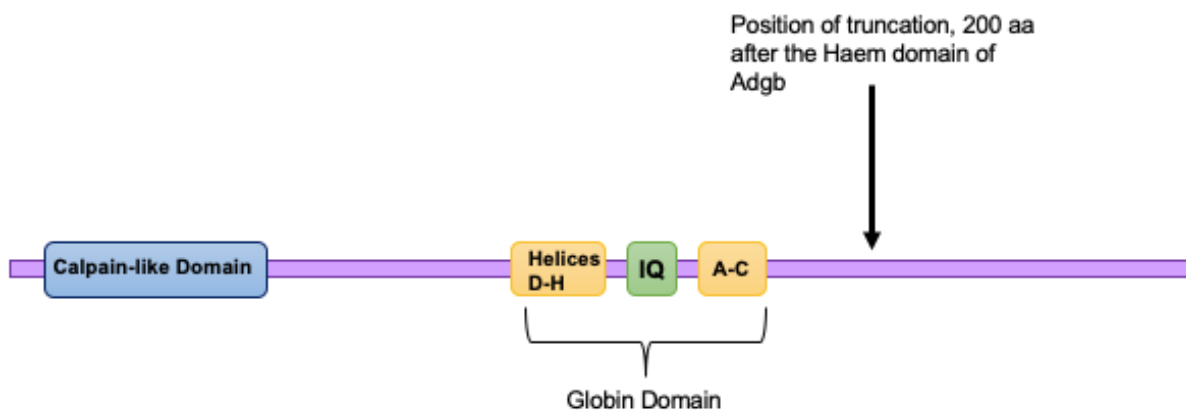


**Figure II. A schematic diagram of Adgb as well as, the position of truncation indicated by a black arrow.** The diagram shows the calpain-like domain (blue) and the 8  $\alpha$ -helices (orange) in the haem domain that is split by an IQ-calmodulin binding motif (green).

## Appendix 10. The original Zurich alignment of Androglobin haem domain structure

The original alignment of Adgb-HD sequence alignment, the figure has been modified from the *Hoogewijs et al* (2012) original paper. The globin  $\alpha$ -helical structure is illustrated on top of the sequence alignment. The original alignment shows 8  $\alpha$ -helices and the IQ- calmodulin binding motif that splits helices C-H (N-terminus) from A-B helices at the C terminus. The conserved residues CD Phe (F) is shown in green on the N-terminal side, the distal Gln (Q) E7 is shown in blue, and the proximal His (H) F8 is shown in pink. The original alignment that was first published by *Hoogewijs et al* is referred to as the Zurich model throughout this project.

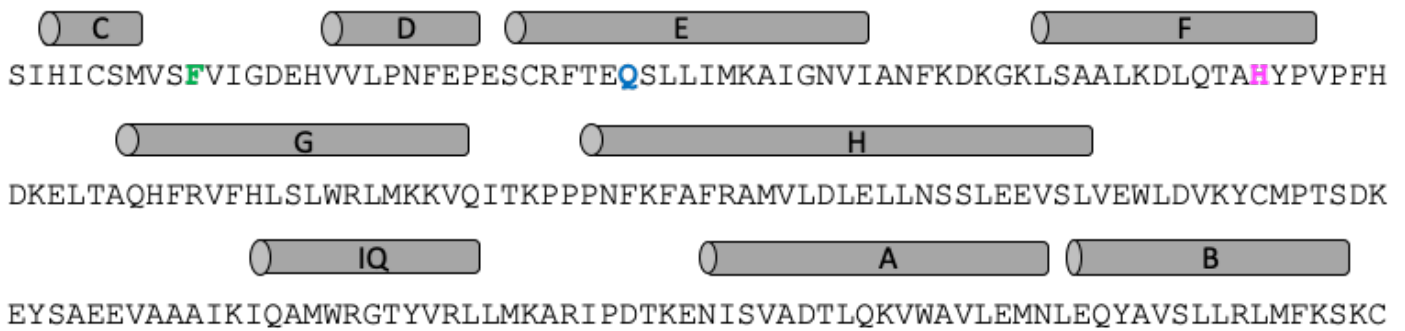


**Figure III. The Zurich model sequence.** The original alignment of Adgb-HD sequence alignment with the  $\alpha$ -helical structure illustrated on top of the sequence alignment. The conserved residues CD Phe (F) is shown in green, the distal Gln (Q) E7 is shown in blue, and the proximal His (H) F8 is shown in pink, as well as surface-exposed cysteines in purple.

## Appendix 11. The re-alignment of Androglobin haem domain structure

An alternative alignment of Adgb-HD is shown below, the revised alignment was completed by Dr Brandon Reeder. The schematic sequence alignment was formed to illustrate the difference between the Zurich model and the Essex model (revised alignment shown below). The globin  $\alpha$ -helical structure is illustrated on top of the sequence alignment. The alignment shows 8  $\alpha$ -helices and the IQ- calmodulin binding motif that splits helices D-H (N-terminus) from A-C helices at the C-terminus. The conserved distal Gln (Q) E7 is shown in blue and the proximal His (H) F8 is shown in pink. The revised alignment found the CD residue to be a Tyr (Y) highlighted in red instead of a Phe, which is located on the C-terminal side instead of the N-terminal side. Two Cys residues are highlighted in purple, as the alignment predicts a potential disulfide bond between regions C and D for haem stabilisation. This stabilisation is usually provided by the CD loop in other globin members (Hargrove et al., 1994). This alternative alignment is referred to as the Essex model throughout this project.



**Figure IV. The Essex model sequence.** The original alignment of Adgb-HD sequence alignment with the  $\alpha$ -helical structure illustrated on top of the sequence alignment. The conserved residues CD Tyr (Y) is shown in green, the distal Gln (Q) E7 is shown in blue, and the proximal His (H) F8 is shown in pink, as well as surface-exposed cysteines in purple