Studies on the structure and biochemical characteristics of the N-terminal and Globin domains of human Androglobin

Jeffrey Oluwasemilore Asaju

A thesis submitted for the degree of PhD Biochemistry

School of Life Sciences
University of Essex

Date of Submission: July 2025

Acknowledgements

First, I am deeply grateful to God for the successful completion of my PhD. I also extend my heartfelt thanks to my family for their unwavering support and understanding throughout this journey. Their support over the years, especially during the intense period of thesis writing, allowed me to focus and persevere.

I would like to thank everyone at the University of Essex who supported me throughout my studies. A special shoutout goes to the past and present members of the 6.19 Lab. Thank you to Dr Michelle Simons, who took me under her wing when I first started and helped me settle in with warmth and guidance. To Dr Elizabeth Welbourn, the best lab technician anyone could ask for, thank you for always listening to my many complaints and offering help however you could. Dr Marcus Edwards, thank you for your invaluable assistance with SAXS, you're truly the best. Dr Sinan Battah, I will miss our spontaneous singing sessions. Hana and Mert, thank you for making my lab experience so memorable. A special thank you to Professor Michael Wilson for all the support with kinetics, your encouragement, and the impromptu grilling sessions that always kept me on my toes. I'll miss those moments.

I am also grateful to Dr Dima Svistunenko for his help with EPR experiments, Dr Sylwia Jencz for guidance on Western blot protocols, Dr Greg Brooke and Dr Gareth Wright for your support with cell work, and Dr Penny Beckerson for her assistance with my thesis. I genuinely could not have generated this volume of data without your collective expertise, suggestions, and willingness to help.

To all my friends in Essex, even though I can't mention everyone by name, each of you holds a truly special place in my heart. I loved our spontaneous lunch breaks, late night dinners, car karaoke sessions, and the unforgettable birthday celebrations. Especially my first one away

from family, spent by the lake, and the later celebration at the restaurant. You truly made those moments unforgettable. These are memories I will always cherish.

And now, to the coolest supervisor ever—thank you, Dr Brandon Reeder. You have no idea how much your belief in me and your support has meant. From our very first meeting during my undergraduate days, you have consistently encouraged me. Thank you for offering me the opportunity to pursue a PhD, for always being compassionate and understanding, for responding to my endless emails even when you weren't at your best, and for making time for me no matter how busy you were. Most of all, thank you for simply being kind. I honestly don't think I could have completed this PhD while also pursuing medical school without your unwavering support. You deserve all the flowers and more. I wish you nothing but the best in life.

Abstract

Androglobin (Adgb) is a multidomain, circularly permuted protein discovered in 2012. To date, there have been little significant progress in its experimental characterisation due to the lack of expression of stable forms of the protein and its domains. Recently, an examination of the protein sequence alignments have enabled the successful expression of the globin domain (Adgb-GD) of androglobin. This thesis examined methods to optimize protein expression for producing a functional holo Adgb-GD protein suitable for biochemical analysis and characterisation.

Biochemical characterisation revealed that Adgb is expressed predominantly as a hexacoordinate globin, both in the ferrous and ferric oxidation states, exhibiting optical characteristics similar to those of neuroglobin (Ngb) and cytoglobin (Cygb). However, the ferric state of Adgb exhibits pentacoordinate-like features, resembling the behaviour observed in Cygb, suggesting a potentially more complex coordination dynamics that is later explored in this study.

The role of the disulfide bond in Adgb-GD ligand binding was also investigated. The findings demonstrate that the disulfide bond not only contributes to the stabilisation of the heme pocket by forming an 'artificial' CD loop typically absent due to the circular permutation of the Adgb globin domain but also plays a critical role in modulating ligand-binding behaviour, including potential interactions with other residues involved in protein-ligand interactions.

Structural insights obtained from SEC-SAXS and SEC-UHPLC experiments, along with studies on the kinetics of ligand-binding, indicate that Adgb-GD exists as a heterogeneous population with multiple conformers. Moreover, Adgb-GD exhibited robust *in vitro* nitrite reductase activity, suggesting a potential redox regulation physiological role for the protein *in vivo*.

To our knowledge, this is the first study to experimentally investigate the N-terminal, calpain-like region of Adgb. The findings support previous studies that suggest this region possesses autoproteolytic activity. Additionally, a novel nucleophilic cysteine site is proposed at Cys206, based on its high sequence conservation and the observation that mutation at this position enabled, for the first time, the successful expression of the full-length N-terminal region of the protein.

Table of contents

List of figures

List of tables

List of Equations

Abbreviations

Adgb – Androglobin

Adgb-GD – Androglobin Globin Domain

ALA – 5-aminolaevulinic acid

C132G – Mutated Androglobin N-terminal domain functional cysteine 132 for glycine

C181G – Mutated Androglobin N-terminal domain functional cysteine 181 for glycine

C206G – Mutated Androglobin N-terminal domain functional cysteine 206 for glycine

C406G – Mutated Androglobin N-terminal domain functional cysteine 406 for glycine

CD – Circular dichroism

CO – Carbon monoxide

CO_2 – Carbon dioxide

Cygb – Cytoglobin

DNA –Deoxyribose nucleic acid

EPR – Electron paramagnetic resonance

GFP – Green fluorescent protein

H824C – Mutated Androglobin Globin domain proximal histidine for cysteine

H824Y – Mutated Androglobin Globin domain proximal histidine for tyrosine

Hb – Haemoglobin

HPLC – High performance liquid chromatography

HS – High-spin

IPTG – Isopropyl β-D-1-thiogalactopyranoside

kDa – Kilodalton

LB – Luria Bertani

LDH – Lactate dehydrogenase

LS – Low-spin

Mb – Myoglobin

MW – Molecular weight

MWCO – Molecular weight cut-off

Ngb – Neuroglobin

NO – Nitric oxide

PAGE –Polyacrylamide gel electrophoresis

PBS – Phosphate buffered saline

PCR – Polymerase chain reaction

Q792G – Mutated Androglobin Globin domain distal glutamine for glycine

Q792H – Mutated Androglobin Globin domain distal glutamine for histidine

ROS – Reactive oxygen species

SAXS – Small Angle X-ray Scattering

SASBDB - Small Angle Scattering Biological Data Bank

SDS-PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis

TCEP – Tris(2-carboxyethyl)phosphine

TCEP-reduced – Disulfide reduced

TFA – Trifluoroacetic acid

WT – Wildtype

CHAPTER I: General

Introduction

1.1 Background on Globins

Globins are among the most extensively studied proteins in biological and medical sciences. Nonetheless, research interest in them continues to grow. Over the past decade, more than 40,000 globin-related publications have been indexed on PubMed. They can be found in the three big domains of life namely: the Archaea, Bacteria and the Eukarya (Vinogradov *et al*., 2006). The widespread occurrence of globins in the three big domains of life demonstrate that globins diverge from a common ancestor that was present in the last common ancestor of all living things 1.5 billion years ago (Burmester *et al*., 2014).

Globins are typically composed of 120 to 200 amino acids and adopt a conserved tertiary structure consisting of 6 to 8 α-helical segments, conventionally labelled A to H (Bolognesi *et al*., 1997). A defining feature of globins is the presence of a heme prosthetic group, which includes a centrally coordinated iron atom bound within a protoporphyrin IX ring. This heme group is embedded within a characteristic 3-over-3 α-helix sandwich fold, providing the structural framework essential for ligand binding and the protein's functional versatility (Bolognesi *et al*., 1997).

All globins can be generally classified into two groups: single chain globins and globin domains within chimeric proteins (Vinogradov *et al*., 2006). Single domain globins comprises of: vertebrate globins (e.g. α- and β-globins in haemoglobin (Hb), myoglobin (Mb), neuroglobin (Ngb), cytoglobin (Cygb)), metazoan Hbs (both intra and extracellular), plants globins (e.g. symbiotic Hb (SHbs), and nonsymbiotic Hbs (NsHbs) or phytoglobins (Pgbs)) and finally the globins from algae and bacteria (e.g. the structurally different truncated Hb (truHb) with a 2/2 alpha helical fold) (Pesce *et al*., 2007). Chimeric globin domains contrastingly are part of larger proteins and usually appear at the N-terminal end. The two major types are: Flavohemoglobins (FHbs) (~400 amino acids long), with an NAD and FAD binding

C-terminal domain (coenzymes involved in redox reactions) (Kuriyan *et al*., 1988), and the Globin-coupled sensors (GCSs) with C-terminal regions that vary greatly in length and function (from ~50 to over 700 amino acids) (Thijs *et al*., 2007). Androglobin (Adgb) is the latest example of chimeric globins.

Overall, all globins fall into one of three evolutionary lineages: 3/3 FHbs and sensor globins (SDgbs) found in both eukaryotes and bacteria, the 3/3 GCSs and Pgbs, and finally the 2/2 Hbs that are structurally distinct, and the only lineage found in all three domains of life; bacteria, archaea, and eukaryotes (Vinogradov *et al*., 2006) (see Figure 1.1.1A). The 3/3 is the 3-over-3 α-helix sandwich fold that is made of 8 alpha-helices, labelled A through H described above (e.g. Figure 1.1.1B). These helices are folded into two layers of three helices each, stacked like a sandwich with the remaining helices stabilizing the structure (Lecomte *et al*., 2005). Whereas the 2/2 is an alternative globin fold, found mostly in bacteria and plants, and are structurally distinct from the 3/3 globins (Wajcman *et al*., 2009). They have two layers of two helices and are smaller than the structures from 3/3 fold. However, these proteins can still bind heme and can perform similar functions (like gas transport or sensing) (Wajcman *et al*., 2009).

Flavohemoglobin has been shown to be increasingly tied to virulence whereas the N-terminal globin domain of the globin coupled sensors can act as a sensor to stimulate a signal-controlled pathway e.g. FixL, a putative oxygen sensor that modulates the activity of FixJ, a transcriptional activator of nitrogen fixation FixK gene (Wajcman *et al*., 2009). Here, oxygen binding switches off the activity of FixL thus preventing the phosphorylation of FixJ and inhibiting this transcriptional activator (Wajcman *et al*., 2009). Both flavohemoglobin and GCS do not occur in vertebrates.

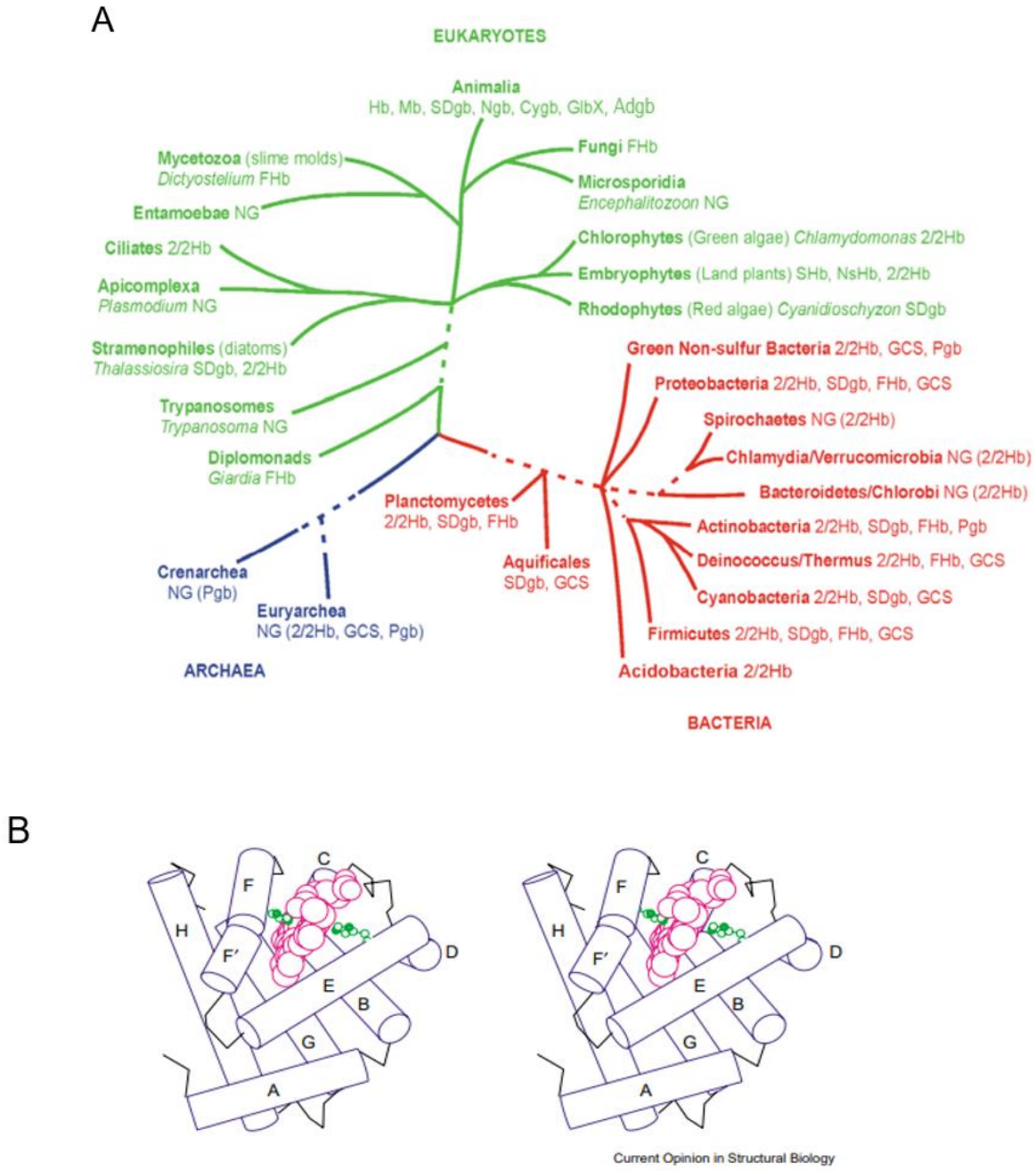


Figure 1.1.1 Adapted images representing globin proteins.

(A) Phylogenomic profile of globins that show the distribution of globins across the three kingdoms of life based on Vinogradov *et al*. (2006) and Baldauf et. al. (2004). (B) The structure (stereoview) of sperm whale myoglobin; PDB code: 1MBD (Lecomte *et al*., 2005). This image shows all 8 helices labelled A-H. The heme groups are shown as large spheres. Proximal and distal histidine are shown in green. The heme consists of an organic component (protoporphyrin) and a central iron atom. Including the 4 coordinate bonds with protoporphyrin, the central iron can form 2 more bonds. These binding sites are called the fifth and sixth coordination sites (Bracke *et al*., 2018).

1.2 The Discovery of Globins: Focus on Myoglobin, Neuroglobin, and Androglobin

The advances in bioinformatics and various high-throughput sequencing over the past five decades have driven the discovery of new globins within the globin superfamily, expanding beyond Hb and Mb discovered over a century ago. The newly discovered globins in vertebrates includes Adgb (Hoogewijs *et al*., 2012), Cygb (Kawada *et al*. 2001, Trent *et al*. 2002), globin E (GbE) (Kugelstadt *et al*. 2004), globin X (GbX) (Fuchs *et al*., 2006), globin Y (GbY) (Roesner *et al*., 2005) and Ngb (Burmester *et al*. 2000).

Mb (from sperm whale) was the first protein structure ever determined by X-ray analysis, then at 6 Å resolution (Kendrew *et al*., 1958) but quickly improved to 2 Å (Kendrew *et al*., 1960). Mb is a typical globular protein found in many animal cells, capable of reversibly binding to oxygen in a manner similar to Hb (Kendrew *et al*., 1958). It's about ~152 aa long with molecular weight (Mw) of 17 kDa (Kendrew *et al*., 1958). The 3D structure of Mb consists of a monomeric subunit containing a single heme group embedded in a hydrophobic pocket between helices E and F. It adopts the classic 3-over-3 α-helical sandwich characteristic of the globin fold, as illustrated in Figure 1.1.1B (Lecomte *et al*., 2005). In humans, Mb gene is located on chromosome 22, specifically at the 22q12.3 band (Jeffreys *et al*., 1984). Key residues include the proximal histidine (HisF8), which is directly coordinated to the iron atom of the heme, and the distal histidine (HisE7), which stabilises ligand binding (Kerr *et al*., 1985).

Ngbs are predominantly expressed in the brain particularly at the frontal lobe, the subthalamic nucleus and the thalamus (Burmester *et al*., 2000). They are monomeric proteins around ~151 aa long with Mw of ~17 kDa (Burmester *et al*., 2000). They also comprise eight α-helices with an embedded heme group, arranged in the classical 3-over-3 α-helical sandwich typical of globin folds (Pesce *et al*., 2004). In humans, Ngb gene is located on the long arm of

chromosome 14 at position q24.3 (Ascenzi *et al*., 2016). It comprises three conserved introns (named B12.2, E11.0, and G7.0) and four exons: exon 1 spans nucleotides 2271 – 2734, exon 2 spans 4200 – 4311, exon 3 spans 4941 – 5060, and exon 4 spans 6856 – 8041 (Ascenzi *et al*., 2016). In WT Ngb, the heme iron is hexacoordinate, bound to the four pyrrole nitrogen atoms of the heme plane, the distal histidine (His E7, position 64), and the proximal histidine (His F8, position 96) (Ascenzi *et al*., 2016). This contrasts with the sperm whale Mb (Figure 1.1.1B), in which the heme iron typically exhibits a pentacoordinate geometry, lacking the distal histidine coordination under resting conditions (ligand free) (Ascenzi *et al*., 2016).

Adgb, a multidomain and circularly permuted globin unique to metazoans, was discovered in 2012 and is predominantly expressed in the testis tissue of both humans and mice (Hoogewijs *et al*., 2012). Its delayed discovery was primarily due to its circular permutation, which caused it to be overlooked in conventional globin motif searches within genetic databases (Hoogewijs *et al*., 2012). However, advances in bioinformatics tools and the rise of high-throughput sequencing technologies in recent years have significantly facilitated the identification of novel proteins, including Adgb. Recent discoveries have revealed that it is also significantly expressed in the female reproductive tract, lung epithelial cells and in the brain ependymocytes and α-tanycytes, specifically associated with cell types forming the motile cilia (Koay *et al*., 2021). The Adgb gene was first found in the genome of sea urchins and lancelets (a subphylum (Cephalochordata) of small translucent marine primitive chordate animals) and it was shown to be highly conserved throughout the metazoan tree of life (Hoogewijs *et al*., 2012). The complete human Adgb gene spans 216,460 bp on chromosome 6q24.3 of which 5,004 bp are coding and contain 36 exons (Hoogewijs *et al*., 2012). There is also a candidate CpG island spanning the first exon and its 5’ region (Hoogewijs *et al*., 2012). This region is rich in cytosine and guanine residues and may play a role in the regulation of the Adgb gene. Four of the thirty-five introns present in Adgb gene are located in globin domain section of the

gene while eight of the introns are located in the calpain domain of the N-terminal region (Hoogewijs *et al*., 2012).

1.3 The Domains of Androglobin

Adgb is distinctive in several aspects to a typical globin (Hoogewijs *et al*., 2012). It's a chimeric protein around ~1667 aa long (human) with a centrally positioned globin domain (Reeder *et al*., 2024). The protein was described as containing four structural domains: a N-terminal region that contain a calpain-like cysteine protease domain, ~ 300 residues without known motifs/ domains, a globin domain that is circularly permuted with a calmodulin binding domain situated between the H and A helices, and finally the C-terminal region is reported to contain sequences for a coiled-coil region, a nuclear localization signal (NLS) and an ER membrane endoplasmic reticulum retention signal (Hoogewijs *et al*., 2012) (Figure 1.3.1). The globin domain is predicted to consist of an eight alpha-helical structure (termed A to H) with a 3-on-3 alpha-helical fold typical of Mb- and Hb-like heme proteins that encloses the heme moiety (Reeder *et al*., 2024).

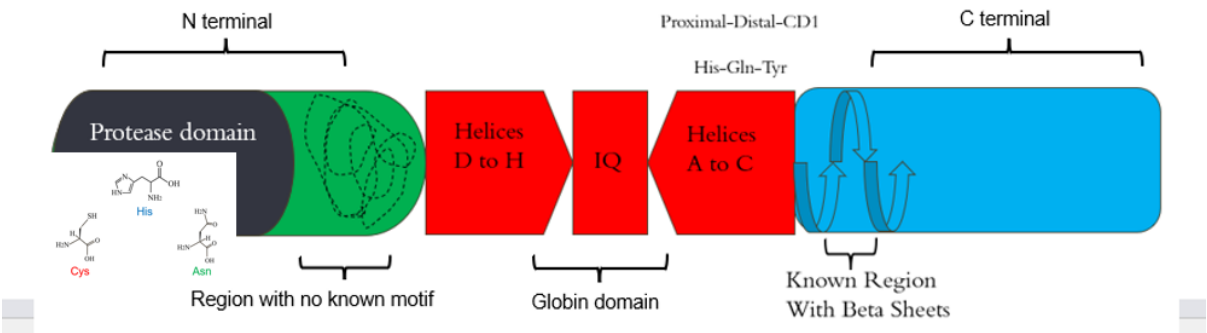


Figure 1.3.1 The overall domain architecture of Androglobin.

1.3.1 The N-terminal domain

Phylogenetic analysis of the calpain-like domains of Adgb alongside 14 human calpains revealed that the N-terminal region of Adgb contains a calpain-7-like domain, showing close similarity to human calpain-7 (Hoogewijs *et al*., 2012). Calpains are family of calcium dependent cysteine proteases that are involved in the intracellular processes of proteins. Their activation is often triggered by calcium influx and oxidative stress (Osako *et al*., 2010). Generally, calpains are made up of two subunits; a large catalytic subunit and a regulatory subunit with both subunits having a penta-EF hand domain that contributes to its activation via Ca^{2+} binding (Hanna *et al*., 2008). Calpain 7 is atypical in the calpain family because it lacks the penta-EF hand domain (Osako *et al*., 2010).

Human calpain 7 comprises of an N-terminal regulatory domain containing two microtubule-interacting and transport (MIT) domains (MITa and MITb), a calpain-like cysteine protease domain in the middle, and a C2-like domain at its C-terminus often involved in calcium binding (Osako *et al*., 2010) (see Figure 1.3.1.1A). The MIT domain has been shown to play a critical role in the activation of calpain 7 (Osako *et al*., 2010). *In vitro* studies suggest that upon binding to MIT interacting motifs (MIMs) from ESCRT-III related protein (i.e. endosomal sorting complex required for transport protein) e.g. increased sodium tolerance 1 (IST1), the MIT domain either dissociates from the protease domain to expose the catalytic core or induces a conformational change that renders the catalytic site active (Osako *et al*., 2010). This suggests that the N-terminal region of calpain 7 might have evolved to acquire specialised mechanisms for interacting with the ESCRT system, with its tandem MIT domains facilitating unique interactions that distinguish its physiological functions from those of calpains in lower eukaryotes (Osako *et al*., 2010). Therefore calpain 7 appears to participate in the ESCRT system, which is essential for key cellular processes such as multivesicular body formation, cytokinetic abscission, and viral budding (Osako *et al*., 2010). Moreover, the

catalytic core of calpain 7 contain a catalytic triad (Arthur *et al*., 1995): Cys290, His458 and Asn478 as indicated by the closed triangles in Figure 1.3.1.1A.

The N-terminal calpain-like protease domain of Adgb shares some homology with the catalytic domain of human calpain 7. However, it lacks the C2-like and MIT domains present in calpain 7 (Bracke *et al*., 2018). The previous study by Hoogewijs (2012) reported that the calpain domain of Adgb retains only one of the three canonical active site residues typical of calpains, a cysteine (Cys132), while several histidine and asparagine residues were present at non-conserved positions. In the catalytic triad of cysteine proteases, cysteine typically serves as the nucleophile, initiating catalysis by attacking the carbonyl carbon of the substrate's peptide bond cleavage (Moldoveanu *et al*., 2002). Histidine functions as the general base, activating the thiol group of cysteine by proton abstraction cleavage (Moldoveanu *et al*., 2002). Asparagine, acting as a stabilizer, helps to orient and stabilize the imidazole ring of histidine through hydrogen bonding and also contributes to stabilizing the oxyanion intermediate formed during the transition state of peptide bond cleavage (Moldoveanu *et al*., 2002).

In calpain 7, although the catalytic triad (Cys-His-Asn) is conserved, the active site remains inaccessible in the resting state, requiring regulatory activation via MIT–MIM interactions with ESCRT-III-related proteins, rather than traditional calcium binding (Osako *et al*., 2010). Adgb lacks the C2-like and MIT domains found in calpain 7, potentially leaving its catalytic core more structurally exposed. This difference may allow Adgb to adopt a distinct activation mechanism and catalytic behaviour, diverging from calpain 7's tightly regulated, conformation-dependent activation pathway. The activation pathway of the Adgb calpain domain, especially mechanisms involving the cysteine nucleophile site, is explored in greater detail in Chapter 3.

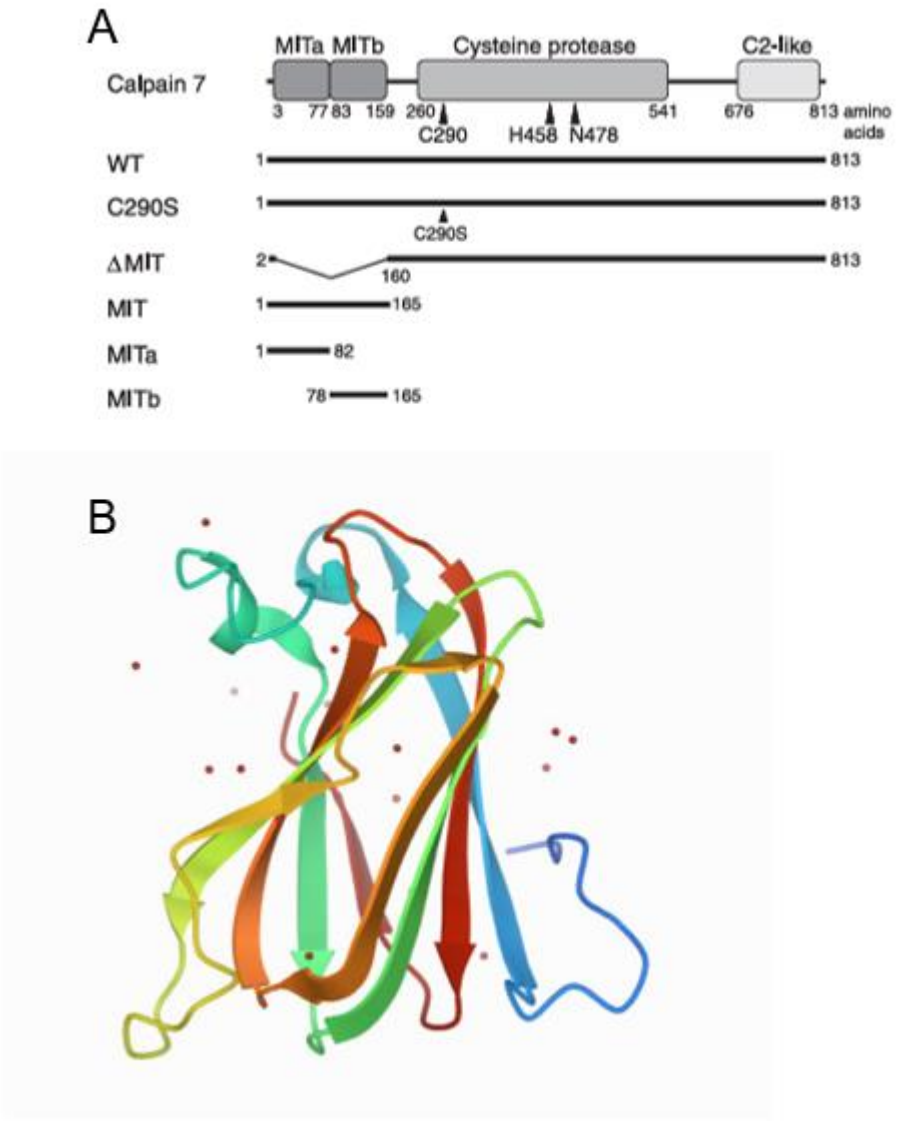


Figure 1.3.1.1 Schematic representation of calpain 7 from Osaka *et al*. (2010).

(A) Structure of calpain 7: N-terminal region containing two MIT domains (MITa and MITb), a calpain-like cysteine protease domain (Cys290, a putative catalytic Cys) and a C2-like domain at its C-terminus. Catalytic triad residues are indicated by closed triangles. (B) The Structure of the Distal C2-like structure of Human calpain 7 from the PDB database 2QFE. Human calpain 7 functions as a calcium-regulated, non-lysosomal thiol protease involved in biological processes such as promoting epithelial cell migration and undergoing self-proteolysis (Campbell and Davies, 2012).

1.3.2 The Globin domain

The Adgb globin domain contain the 8 helices of the typical globin fold, labelled A - H but is circularly permutated. Circular permutation in protein structure is because of rearrangement of amino acids such that the amino and carboxy-terminal regions are interchanged (Lo *et al*., 2008). The globin domain is split into two parts by the calmodulin binding IQ motif (Hoogewijs *et al*., 2012). It's been proposed that this IQ motif might have been acquired by exon-shuffling events (Hoogewijs *et al*., 2012). This region may mediate binding to calmodulin that reacts to calcium ion levels.

There are several amino acids that are highly conserved in all globins. These amino acids can be considered essential as they remain largely unchanged throughout evolution and therefore are fundamental to the function of globins (Vuletich *et al*., 2006). Often there are examples of some variation to these amino acids, but they retain activity and binding. Among these residues includes the three amino acids generally conserved at the heme binding site of all globins; histidine (His) at the proximal F8 site, phenylalanine (Phe) at the CD1 interhelical region and another histidine (or glutamine) molecule at the distal E7 site (Vuletich *et al*., 2006). Other highly conserved residues include leucine (Leu) at the B10 site, a leucine at the F4 site, and aproline (Pro) at the C2 site and phenylalanine at the CD4 site among others (Ota *et al*. 1997).

Some of these amino acids e.g. Leu (B10), Phe (CD1), Leu (F4), Pro (C2), and Phe (CD4), are conserved despite not interacting directly with the heme group (Ota *et al*., 1997). This conservation is attributed to their roles in maintaining structural stability: Leu (B10), Phe (CD1), and Leu (F4) contribute through hydrophobic side chain packing, while Pro (C2) and Phe (CD4) are stabilised by hydrogen bonding (Ota *et al*., 1997). Specifically, Pro (C2) is supported by the N terminal edge of an α helix, and Phe (CD4) is stabilized through backbone

hydrogen bonding with Phe (CD1) (Ota *et al*. 1997). Leu (B10) also plays a role in regulating ligand binding to the heme iron (Wajcma *et al*., 2009). His (E7) and His (F8) are at the distal and proximal heme binding sites respectively and are conserved primarily due to their essential roles in heme-related function (Ota *et al*., 1997).

As stated above, a His residue is typically conserved at both the proximal F8 and distal E7 positions in most pentacoordinate and hexacoordinate globins. While Adgb retains the conserved proximal His, it features a Gln residue at the distal E7 position instead of His (Hoogewijs *et al*., 2012). Notably, Gln is a recognized variant at the E7 position in both vertebrate and invertebrate globins, indicating that this substitution in Adgb may still support functional versatility at the distal site. For example, Ascaris Hb (AscHb), an invertebrate globin found in the peri-enteric fluid of parasitic nematode Ascaris suum (Das *et al*., 2000) and the vertebrate Mb of Asian elephants (Romera-Herrera *et al*., 1981; Bartnicki *et al*., 1983) both have Gln replacing the role of His in the function of their globin. In AscHb, the distal E7 Gln is paired up with Tyr at the B10 position to form an interlaced hydrogen binding network that stabilises heme-bound oxygen. This causes AscHb to have a high oxygen affinity (Das *et al*., 2000). Whereas in Asian elephant Mb, Gln interacts strongly with polar ligands to reduce the rates of heme autoxidation and CO dissociation without impairing Mb and oxygen reaction (Romera-Herrera *et al*. 1981; Bartnicki *et al*. 1983).

One of the most conserved residues in globins is PheCD1. This residue plays a vital role in anchorage and binding of the heme within the heme pocket (Hargrove *et al*., 1994; Roesner *et al*., 2004). Hb variants with mutations in this position, such as Hb Hammersmith and Hb Little Venice, invariably leads to instability of the protein (Akiyama *et al*., 2005). Recent work from our lab identified tyrosine as the residue at the CD1 site in Adgb, replacing the typically conserved Phe at this position following sequence reassessment (Reeder *et al*., 2024). The alignment of Adgb with related proteins, specifically focusing on the Adgb globin

domain (Adgb-GD), was re-evaluated using a novel helix-based alignment method effective within the twilight zone of low sequence similarity (Reeder *et al*., 2024). This approach redefined the sequence spanning the C helix (Figure 1.3.2.1C), therefore positioning a tyrosine residue at the CD1 site of the human Adgb-GD (Reeder *et al*., 2024). It also generated an alternate helical alignment around the C and D helical regions i.e. previously it was A-B helices on the C-terminal side and the C-H helices in the N-terminal side (Figure 1.3.2.1A) (Hoogewijs *et al*., 2012). However, in the recent study, the C helix is shifted to the C terminal side rather than the N-terminal side i.e. A-C helices are on the C-terminal side while the rest of the helices, D-H are on the N-terminal side as shown in the Figure 1.3.2.1B (Reeder *et al*., 2024). The structural models for the heme pocket helix sequence orders are shown in Figure 1.3.2.1D and Figure 1.3.2.1E, for the original and revised sequences respectively. The revised sequence displays a heme pocket architecture more comparable to that of heme-containing proteins such as Ngb than the original sequence (Figure 1.3.2.1E vs Figure 1.3.2.1F).

Additionally, the revised alignment revealed the potential presence of an intramolecular disulfide bond linking the N- and C-terminal regions of the heme-binding domain (Figure 1.3.2.2), thereby stabilizing the CD loop and surrounding heme pocket. This was confirmed through a dithiodipyridine titration to determine the surface exposed cysteines (Reeder *et al*., 2024). This disulfide bond was absent in the original sequence alignment (Hoogewijs *et al*., 2012). Only two cysteine residues, Cys787 and Cys978, are surface-exposed and positioned closely enough to form an intramolecular disulfide bond within the CD loop region of the protein (Figure 1.3.2.2C) (Reeder *et al*., 2024).

Although tyrosine and phenylalanine share structural similarities, both being hydrophobic and aromatic, the presence of tyrosine at this position is however unique in eukaryotic protein (Reeder *et al*., 2024). However, it is found in nonvertebrate species, including *Mycobacterium tuberculosis* (Ouellet *et al*., 2007) and *Caenorhabditis elegans*

(Hoogewijs *et al*., 2007). Tyr in the CD1 position of prokaryotic Hbs does not significantly affect heme binding, but can affect oxygen binding affinity (Reeder *et al*., 2024).

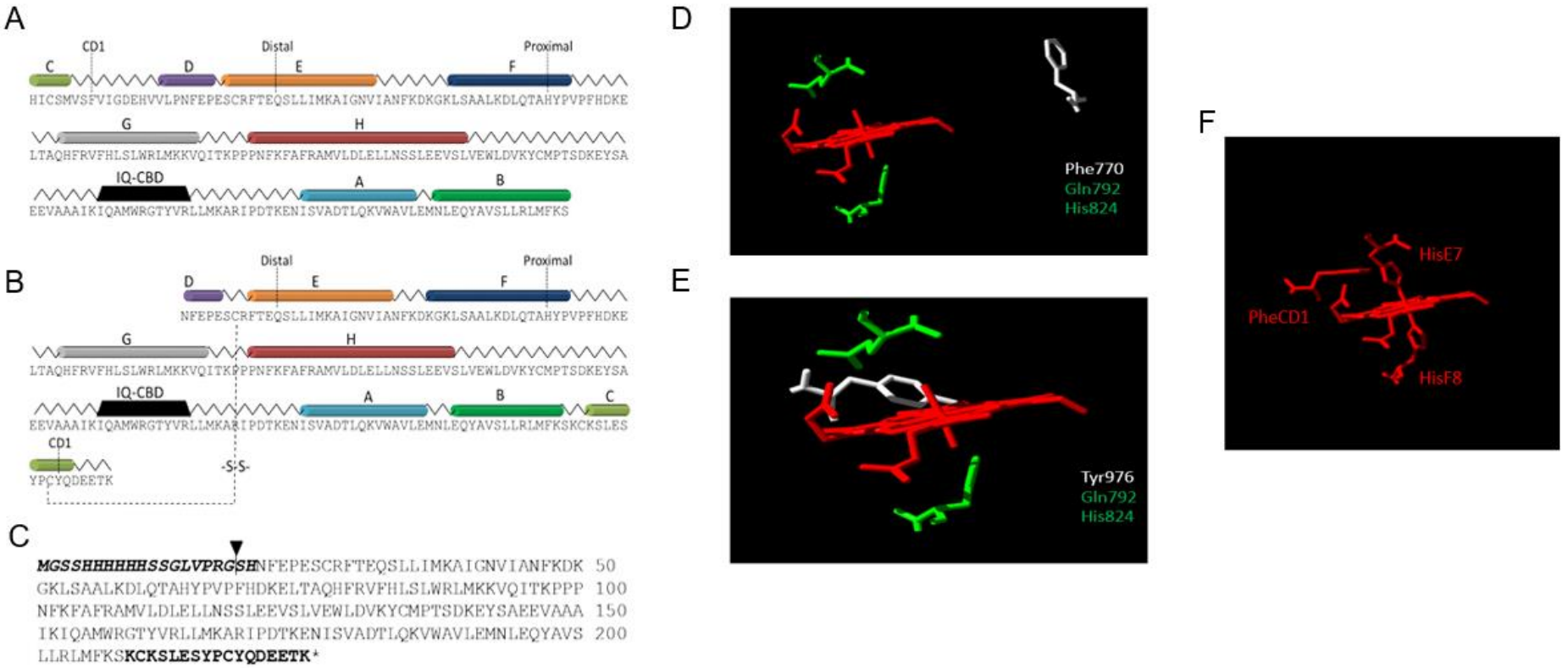


Figure 1.3.2.1 Androglobin globin domain structure alignment (structural models) adapted from Reeder *et al*. (2024).

(A) Original alignment reported by Hoogewijs *et al*. (2012). A-B helices are in the C-terminal direction, and the C-H helices are in the N-terminal direction. (B) The revised alignment reported by Reeder *et al*. (2024). C helix is shifted to the C-terminal direction i.e. A-C helices are in the C-terminal direction and D-H helices are in the N-terminal direction. (C) Protein sequence as expressed including N-terminal His-tag sequence and linker highlighted in bold italics, thrombin cleavage site indicated by arrow and redefined C helix region at the C terminus in bold. (D) Model heme pocket structure displaying the two heme iron coordination and the CD1 amino acid (white) used for stabilising the heme binding from the original Adgb sequence in A, (E) the revised sequence in B, and (F) human Ngb (PDB ID: 1OJ6). These structures (D – F) are predicted models generated using the I-TASSER server and visualised using Swiss-PDB Viewer.

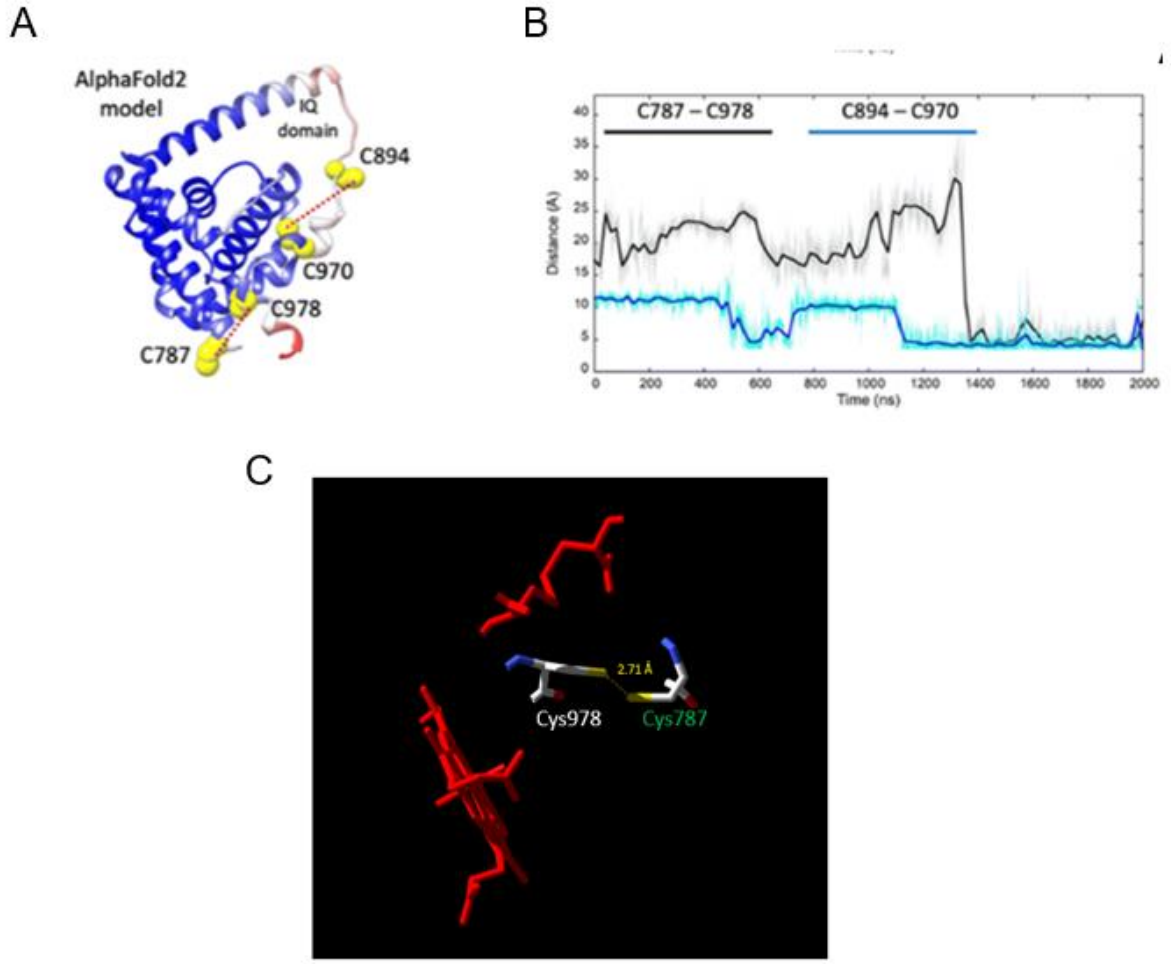


Figure 1.3.2.2 Cysteine residues involved in disulfide bridge formation in androglobin. Image adapted from Reeder *et al*., (2024).

(A) The four cysteine residues of the domain and their distance are shown on the AlphaFold 2 model. (B) The distance between pairs of cysteine residues during the MD simulation of the AlphaFold 2 model (C787 – C978 in black and C894 – C970 in blue). Both C787 – C978 and C894 – C970 moved close enough to form an intramolecular disulfide bond in the MD simulation of the AlphaFold 2 structure after 1.4 μs. (C) The structural model illustrates the distance between Cys787 and Cys978 following adjustment of the axial positions of the cysteine sulphide groups. The distance between cysteines is shown to be below the PDB cut-off value for disulfide bonds formation (< 3.0 Å) (Sun *et al*., 2017). This is compared to the two Cys residues of Ngb 4MPM (red).

1.3.3 The C-terminal domain

The Adgb C-terminal region contains a significant predicted disordered section of ~500 amino acids (Figure 1.3.3.1). It also has at least three ordered sections with potential regulatory roles; a nuclear localization signal (NLS), an endoplasmic reticulum (ER) membrane retention signal and a coiled coil region (Hoogewijs *et al*., 2012). Coiled coils can function as molecular spacers and can be used as vesicle tethering mediators thus suggesting that this region can be a scaffolding region. The NLS can mediate the import of proteins into the nucleus by acting as a molecular tag recognized by the nuclear import machinery (Lange *et al*., 2007). The ER membrane retention signal contributes to transmembrane localization by facilitating the retrieval of ER membrane proteins from the Golgi apparatus back to the ER (Nilsson *et al*., 1989). These features support the hypothesis that this region may also participate in protein–protein interactions or dimerization, as previously suggested (Hoogewijs *et al*., 2012).

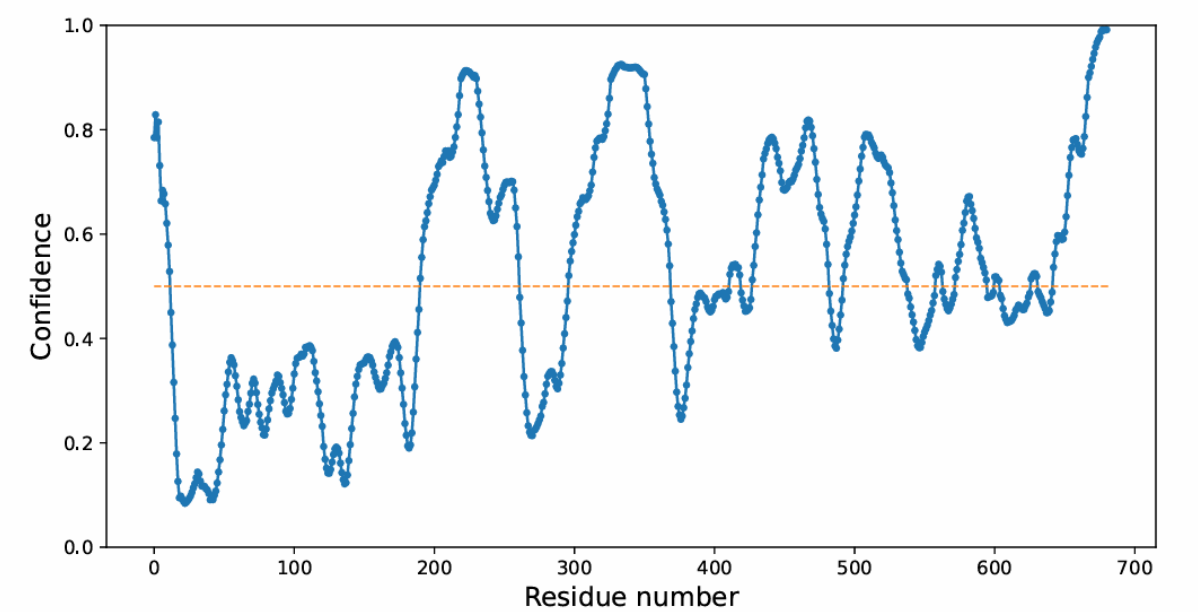


Figure 1.3.3.1 Showing the disorder prediction of the C-terminal region following analysis with the protein disorder prediction system (PrDOS). Residues falling below the threshold line are predicted to be ordered, while those above the threshold line are predicted to be disordered.

1.4 Globins and their ligands

The heme group in globins comprises an organic component (protoporphyrin IX), and a central iron atom (Fischer and Zeile, 1929). The protoporphyrin structure consists of four pyrrole rings connected by methine bridges, forming a planar tetrapyrrole ring system. The iron atom at the centre forms six coordinate bonds: four in the plane of the heme with the nitrogen atoms of the pyrrole rings, and two axial positions perpendicular to the plane (Schulz and Schirmer, 1984). These axial positions are referred to as the fifth (proximal) and sixth (distal) coordination sites. In globins, the proximal site is typically occupied by a histidine residue (His F8), while the distal site is available for ligand binding, such as oxygen, carbon monoxide, or nitric oxide (Kendrew *et al*., 1958; Perutz *et al*., 1960).

The heme iron in globins can cycle between reduced (ferrous) and oxidized (ferric and ferryl) states, with the oxidation state capable of binding different ligands. O_2, CO and NO are example of physiological ligands of the reduced form of the iron, although NO can also bind to the ferric form of the iron (Freitas *et al*., 2008). The heme iron can also form ferryl as part of the redox activity involving peroxidases (George and Irvine, 1956). Hb, Mb and GbE are characterized by the pentacoordinate bonds of the iron atom of the heme in the absence of oxygen. The fifth coordination site is occupied by the imidazole ring of the proximal histidine on the protein (Storz *et al*., 2011). The proximal histidine is highly conserved in globins and can be found on the 8th amino acid of helix F (F8). In their deoxy state, the sixth coordination site of the ferrous iron atom is accessible to oxygen binding or other gaseous ligands bindings like CO (Storz *et al*., 2011). Contrastingly, in the absence of exogenous ligands, the iron atom in the heme of Ngb, Cygb and Adgb are hexacoordinated in both the ferric and ferrous states of the iron atom (Pesce *et al*. 2002, Reeder *et al*. 2024). Both of the proximal and distal binding sites are occupied by the N-atom of the imidazole ring of a histidine or glutamine (Pesce *et al*., 2002). Oxygen binding and other exogenous ligand binding on hexacoordinate globins requires

the displacement of the distal histidine (Storz *et al*., 2011). Currently, no data exist for the coordination status of globin Y (Burmester *et al*., 2014).

The widespread presence of globins across various lineages in all three kingdoms of life has prompted questions about their functional roles. In the animal kingdom, the principal role of 3-over-3 Hbs is oxygen transport while the 2-over-2 Hbs are usually involved in NO detoxification (Wajcman *et al*., 2009). Some species however develop special evolution pattern for Hbs thus allowing their carriers to occupy new ecological niches (Wajcman *et al*., 2009).

Both Hbs and Mbs have the ability to function as redox enzymes reacting with NO, sulphides and peroxidases. Hbs reacts with NO in one of the two possible reactions: the binding of NO to deoxyferrous Hb or the conversion of NO to nitrate by oxyferrous Hb. Under physiological conditions, vertebrates Hb can influence the scavenging of NO by acting as a reductant or by influencing NO transport through the formation of S-nitroso-Hb (SNO-Hb) that help deliver NO to areas of hypoxia or restricted blood flow causing changes in vasoactivity (see Figure 1.4.1) (Jensen, 2009).

Mbs can act as an oxygen storage protein in some animals e.g. diving animals that enhances oxygen diffusion in high oxygen consuming environments. Mb main function is to facilitate oxygen diffusion across a gradient, specifically from the blood to the cytochrome c oxidase within mitochondria. They can also function as an intracellular scavenger of NO under aerobic condition through the reaction of oxyferrous globin reacting with NO to form ferric Mb and nitrate. This prevents the production of potentially damaging reactive oxygen species (ROS). Under hypoxic condition, Mbs can also generate NO via nitrite reactions, thus leading to the vasodilation (Reeder, 2010).

As described earlier, mammalian Ngb are expressed in the central nervous system, peripheral nervous system and endocrine tissues such as adenohypophysis, adrenal gland,

testes, the pancreatic islets of Langerhans etc, with the highest expression observed in the retina (Hankeln *et al*., 2008). Ngb can bind reversibly to oxygen at the ferrous heme iron with affinity (p50 ≈ 2 torr) which is comparable to that of Mb [p50 ≈ 1 torr (1 torr = 133 Pa)]. This supports the suggestion that Ngb functions in the brain through mimicking Mb in myocytes (storage, transport and intracellular diffusion of oxygen). However, this hypothesis is unlikely due to Ngb high autooxidation rate (within minutes) and low average concentration in the brain (Brunori *et al*., 2005). It is also suggested that Ngb is involved in NO scavenging in the neurons even though the binding affinity of NO to ferrous Ngb is very low compared to pentacoordinate Hbs and Mbs due to the internal His (E7) protecting the ferrous iron (Fe^{2+}) (Hankeln *et al*., 2008).Evidence suggests that the most critical function of Ngb involves the Ngb protein acting as a stress sensor and stress-inducible molecule. It appears to function as a compensatory protein in response to hypoxic, ischemic, and oxidative injuries by activating survival and anti-apoptotic pathways (Ascenzi *et al*., 2016).

Cygb shares some features with Ngb such as high thermal stability. Cygb is specifically expressed in the cytoplasm of cells that are actively engaged in the production of extracellular matrix components in visceral organs (Hankeln *et al*., 2008). They are also expressed in fibroblasts and related cell types in a broad variety of splanchnic organs like the liver, heart, muscle, gut, kidney, lung and the pancreas (Hankeln *et al*., 2008). Not much is known about the functional roles of Cygb, however recent studies have shown that Cygb contributes to NO availability in the vasculature, which is central to normal blood vessel function through regulation of smooth muscle cell tone and physiological response (Lilly *et al*., 2018). Further experiments have also suggested that Cygb may play a protective role under oxidative stress conditions (Beckerson *et al*., 2014) e.g. controlling NO homeostasis in the cell under oxidative and ischemic conditions (Reeder and Ukeri, 2018).

It has been suggested that the globin domain of Adgb perform roles other than O_2 delivery (Hoogewijs *et al*., 2012). Recent study by Reeder *et al*. (2024) indicated that Adgb could potentially participate in NO sensing or regulation and that O_2 delivery was unlikely, also due a high autooxidation rate. It was also reported to have an important role in the late phases of spermatogenesis (Bracke *et al*., 2018). Adgb function may be linked to the chromatoid body of male germ cells, potentially through interactions with TDRD6 and PIWIL1 (Bracke *et al*., 2018). Alternatively, Adgb may exert crucial redox-regulatory function via its globin domain during sperm capacitation. It remains unclear whether Adgb acts earlier in spermatogenesis, or Adgb is cleaved, and only the globin domain persists in mature spermatozoa following proteolytic cleavage (Bracke *et al*., 2018).

Redox regulation plays a vital role in spermatogenesis, particularly during the process of sperm capacitation. During spermatogenesis, the body naturally produces reactive oxygen species (ROS), which, at controlled levels, are essential for initiating capacitation, a sequence of morphological and metabolic transformations required for spermatozoa to acquire fertilization capability (Aitken *et al*., 2012). Over the course of capacitation, there is a progressive increase in various ROS, including superoxide anion, hydrogen peroxide, nitric oxide, and peroxynitrite (Bracke *et al*., 2018). While moderate ROS levels support physiological signalling, excessive accumulation leads to oxidative stress, which can cause sperm abnormality ultimately impairing fertility (Aitken *et al*., 2015). Oxidative stress arises when ROS generation exceeds and overwhelms the body's natural antioxidant defence mechanisms (Bracke *et al*., 2018). In sperm cells, ROS are primarily produced via two systems: the NADPH oxidase system located at the plasma membrane and the NADH-dependent oxidoreductase system within the mitochondria (Bracke *et al*., 2018).

The identity of the sperm oxidase responsible for the production of superoxide anion or nitric oxide during capacitation in eukaryotes, remains elusive (Bracke *et al*., 2018). The globin

domain of Adgb has been proposed as a potential candidate for this function (Bracke *et al*., 2018). This is because there are parallels between Adgb-GD and globin 12 (GLB 12), a hexacoordinated globin in *Caenorhabditis elegans* that exhibits redox activity and plays a crucial role in reproductive processes (De Henau *et al*., 2015). GLB 12 generates superoxide, which is regulated by both intracellular and extracellular superoxide dismutase (SOD) (Sasha De Henau *et al*., 2015). This enzymatic modulation creates a transmembrane hydrogen peroxide gradient that functions as a redox signal to influence reproductive processes (De Henau *et al*., 2015). Similarly, Adgb has been shown to interact with SPATA20, a testis-enriched protein containing a thioredoxin-like domain (Bracke *et al*., 2018). Thioredoxins are proteins that are involved in redox signalling and facilitates the reversible oxidation of cysteine thiol groups to form disulfide bonds (Bracke *et al*., 2018). This reaction involves the transfer of two electrons and two protons, playing a key role in maintaining cellular redox homeostasis (Shi *et al*., 2004).

Other roles of Adgb includes having cellular roles in cilia biogenesis and function (Koay *et al*. 2021). It has been identified as a direct transcriptional target of FOXJ1, a key regulator of motile ciliogenesis, within a complex regulatory network (Koay *et al*., 2021). Moreover, Adgb appears to participate in redox and calcium signalling (Nie *et al*., 2025). Adgb nitrite reductase (NIR) activity is significantly enhanced at least ~2.2-fold higher upon binding with calmodulin (CaM), suggesting a calcium-dependent regulatory mechanism (Nie *et al*., 2025). Furthermore, Adgb has demonstrated potential oncogenic behaviour in glioma, a type of brain cancer derived from glial cells. Silencing Adgb expression in glioma cell lines led to reduced cell proliferation and increased apoptosis, indicating its involvement in cell cycle regulation and tumour suppression (Huang *et al*., 2014). Finally, Adgb is subject to transcriptional regulation by epigenetic mechanisms in non-reproductive tissues (Chen *et al*., 2020). For example, in pancreatic cancer, the long non-coding RNA STXBP5-AS1 has been

shown to suppress Adgb expression by recruiting the histone methyltransferase EZH2 to its promoter, thereby promoting DNA methylation and transcriptional silencing neighbour Adgb gene (Chen *et al*., 2020). Epigenetic analyses revealed hypermethylation of the Adgb promoter region, suggesting that Adgb expression may be silenced through DNA methylation. This opens up potential therapeutic avenues, including the use of DNA demethylating agents (Chen *et al*., 2020).

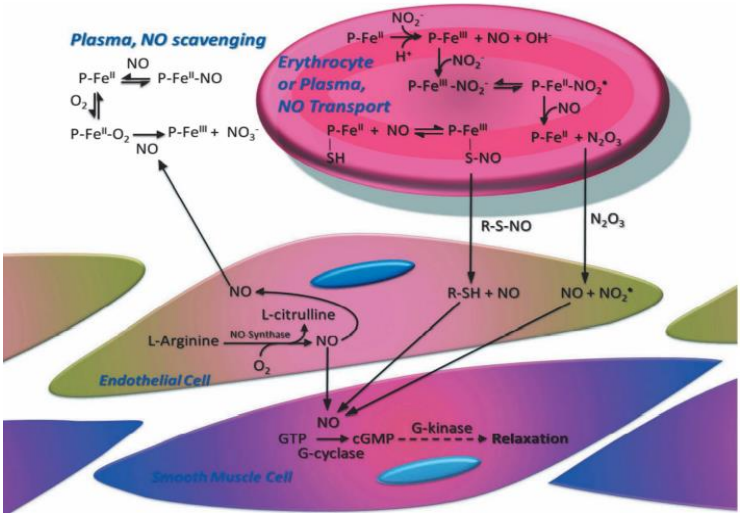


Figure. 1.4.1. Summary of haemoglobin reaction with nitric oxide taken from Reeder, 2010.

The reaction of NO with Ferryl [Fe(IV)] Hb is shown above. This is a scavenging reaction of Hb with NO yielding a peroxynitrite intermediate before finally producing ferric protein and nitrite. This reaction is very rapid and does not require hypoxic conditions. The physiological relevance of the scavenging reaction of Hb is most relevant outside of the erythrocyte (red blood cell), such as flowing cell lysis or through addition of cell-free haemoglobin-based oxygen carriers. The physiological relevance of NO-delivery system is still under debate because numerous studies using sensitive methods to detect NO liberation from SNO-Hb have reported extremely low or undetectable concentrations of SNO-Hb *in vivo*, which is too low for any significant effect on vasoregulation. Scavenging or supply to endothelial cells affects the NO signalling for smooth muscle contraction (Reeder, 2010).

1.5 Summary: Thesis objectives and study outline

Adgb remains a relatively novel protein, with essentially no structural and very little functional data. Until recently, generating a stable form of the protein either full length or part of the protein, e.g. globin domain, has proved challenging. Advances in understanding its helical alignment have now made recombinant production feasible (Reeder *et al*., 2024). However, the protein's large size and autoproteolytic properties still precludes efficient expression in *Escherichia coli (E.coli)* and other organisms e.g. Sf9 insect cells (Bracke *et al*., 2018) . Consequently, this study adopts a domain-based approach, segmenting Adgb into its individual structural regions. The focus of this PhD research is specifically on the N-terminal and globin domains.

The overall aim of studying the N-terminal domain of Adgb is to characterize its calpain-like region and assess whether this domain is responsible for the proteolytic activity that may hinder the successful expression of the full-length protein. Additionally, the study aims to elucidate the potential structural relationship between the N-terminal domain and the globin domain. This will be investigated by mutating the nucleophilic cysteine residue within the catalytic triad, which is essential for initiating proteolysis in calpain proteins. While previous studies have suggested that the N-terminal domain may possess autoproteolytic activity (Hoogewijs, 2012; Bracke *et al*., 2018), no experimental validation has been reported to support this hypothesis. Moreover, the failure to successfully produce the globin domain even after removal of the N-terminal region left previous results inconclusive and open to interpretation. This project represents the first in-depth investigation of the N-terminal domain of Adgb.

The overall aim of working with the globin domain is to gain insight into its behaviour *in vitro*, particularly its interaction with physiological ligands such as NO, as a means to infer

its potential biological function *in vivo*. Biochemical characterization will further explore structural aspects of the globin domain, including the presence of conformational variants such as those involving intramolecular disulfide bonds and whether these conformers exhibit distinct functional behaviours. These properties will be probed using ligand binding reactions with CO and CN. Although initial studies have described some characteristics of the Adgb globin domain (Reeder *et al*., 2024; Nie *et al*., 2025), no comprehensive or detailed biochemical analysis has been conducted to date. The ultimate aim is to integrate findings from both domains to construct a foundational understanding of the full-length protein, paving the way for future structural resolution of Adgb.

The objectives of the work presented in this thesis are as follows:

- ***Objective 1*:** To identify the optimal method for recombinant expression of heme-containing Adgb-GD in *E. coli*, and to characterize the coordination states of Adgb both *in vivo* and *in vitro*
- ***Objective 2*:** To investigate the interactions between different conformers of Adgb-GD and its non-physiological ligands (CO and CN) to provide structural information
- ***Objective 3*:** To investigate the redox property of Adgb-GD and its physiological ligands (NO and lipids) to provide information about potential function
- ***Objective 4*:** To understand how proximal and distal mutants regulate ligand binding
- ***Objective 5*:** To investigate the proteolytic properties of the N-terminal domain by attempting its recombinant expression
- ***Objective 6*:** To understand how cysteine nucleophiles regulate the autoproteolytic nature of the N-terminal domain and locating the key active site amino acid
- ***Objective 7*:** To investigate the structure of the N-terminal and globin domain through SAXS

E.coli is a perfect host for large scale recombinant expression for small globular proteins such as globins because they usually don't have any specific requirement in terms of folding or post-translational processing (Bracke *et al*., 2018). *E. coli* is the first host used for recombinant protein expression and it's the leading host organism for most scientific application of protein expression. The B strains of *E. coli* (in particular BL21 DE3) are very popular hosts for protein expression. This is because it is easy to manipulate and grow quickly in inexpensive media due to its short doubling time. Recombinant proteins are often directed to the cytoplasm in *E. coli* as that's the location that produce the most yield. They can also be directed to the periplasm and to lesser extent secreted (Brondyk, 2009). The most popular expression system used in *E. coli* is the T7 expression system. In this system, the gene of interest is cloned behind a promoter recognised by the phage T7 RNA polymerase (T7 RNAP). These T7 expression hosts e.g. pET28a carry a chromosomal copy of phage T7 RNA polymerase gene and a T7 lac operon. Therefore, the system can be induced by isopropyl β-D-1 thiogalactopyranoside (IPTG) (Rosano and Ceccarelli, 2014). After induction, T7 RNA polymerase is expressed, and transcription commences. pET28a expression vector have a strong inducible T7 promotor thus it was used for the recombinant expression of all the proteins in this research thesis.

Mammalian expression systems are the preferred system for the expression of mammalian proteins. However, expressing recombinant proteins in mammalian cells is considered the least efficient way of expressing recombinant protein at scale. Mammalian expression systems can be used to produce proteins transiently or through stable cell lines, where the expression construct is integrated into the host genome. Recent technological advances have significantly improved the expression levels of proteins in both the transient and stable cell lines. Their expression level meets the criteria set for the recombinant expression of

protein for research purposes in terms of cost-effectiveness, simplicity, and speed of the process in conjunction with adequate yields of the product (Geisse and Fux, 2009).

Mammalian expression systems provide the physiological relevant environment that promote post-translational processing and activity hereby making it the ideal environment for the expression of mammalian proteins. Transient expression involves rapid expression of protein in the host cell followed by cell lysis. The protein yield in this expression is very variable based on transfection efficacy but typically low as it is in a mixture of lysed cellular protein, nucleic acids, and viral particles. Contrastingly stable expression is used for the expression of large quantities of proteins. Stable expression results from integration into the host-cell genome of the gene for the expression of the heterologous protein. This involves the transfer of the foreign DNA, along with the relevant DNA-based signals for transcription by RNA polymerase II, into the chromosomal DNA of the host cell (e.g. Chinese hamster ovary (CHO)). The integrated gene is transcribed efficiently, and the protein is expressed persistently over many generations by the host cell (Gray, 1997). The disadvantage of using this method is the time required to obtain stable, productive cell banks which can take months. In contrast transient transfection method is relatively easy but technically challenging to scale up. However, the evaluation when using transient expression system can be made in less than 2 weeks (Brondyk, 2009). Therefore, this method is the most suitable for *in vivo* protein expression in this thesis as we only require proteins in small quantity such as western blot analysis. Furthermore, previous studies have shown that Adgb undergoes slower autoproteolysis in insect cell lines compared to *E. coli* (Bracke *et al*., 2018). This slower degradation during mammalian cell expression may be advantageous for investigating the potential physiological roles of both the full-length protein and its proteolytic fragments in Eukaryotic cells.

CHAPTER 2: Expressing Androglobin domains in *Escherichia coli* and Human Embryonic Kidney cells (HEK293T) mammalian cells

2.1 Introduction to prokaryotic and eukaryotic protein expression.

Amino acids are building block of proteins and are present in every cell. Regulation of protein synthesis is fundamental to the functioning of cells. Proteins are synthesised by a complex but well understood multistep system which can be summarised into three principal steps: transcription, translation and protein folding.

Protein expression starts with transcription which is the transfer of genetic information in the DNA into single stranded messenger RNA (mRNA). The mRNA contains the same genetic information as the DNA. The message coded by the mRNA is then translated into proteins. A subpopulation of proteins requires post translational modification (PTM) in order to function correctly. They mainly occur in the endoplasmic reticulum or Golgi bodies of cells. However, in prokaryotes these PTM are less common and generally limited. In contrast, eukaryotes exhibit a higher frequency of PTMs, including methylation, phosphorylation, and the addition of complex molecules such as glycosylation (Figure 2.1.1) (Wang *et al*., 2013). The most predominant PTM in Adgb may be proteolytic cleavage, however, the function of which is unknown.

Protein folding involves folding protein to its correct shape. The order by which an amino acid is linked determines the final structure of the protein. The stages of protein folding include secondary, tertiary and sometimes quaternary folding. A change in the amino acid sequences causes a mutation which changes the structure of the protein. This can lead to a loss of function for the protein as each protein in the body differs in its amino acids thus their shape.

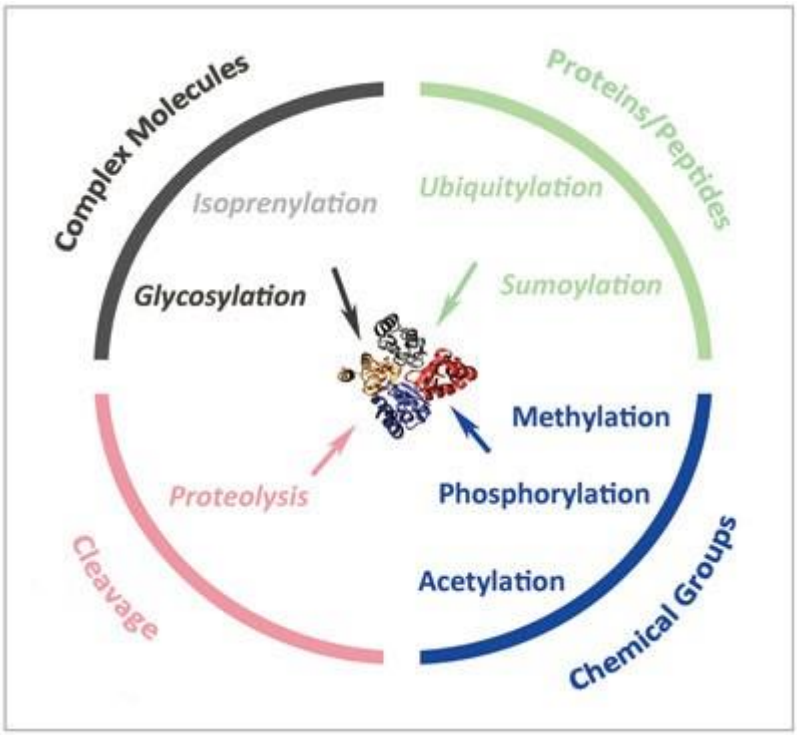


Figure 2.1.1. Overview of various post-translational modifications that can occur following protein translation.

In eukaryotic cells, post-translational modifications primarily occur in the endoplasmic reticulum, with some continuing in the Golgi apparatus. These modifications dynamically influence the structure, stability, and function of proteins within the cell. In the case of androglobin, proteolytic cleavage appears to be the predominant form of post-translational modification. Image from Wang, Peterson, and Loring, 2013.

Previous expressions of Adgb in eukaryotes have been in yeasts (*Pichia pastoris*) and baculovirus infected insect cells (Bracke *et al*., 2018). The study was unable to detect the expression of Adgb globin domain recombinant protein in either the secreted or intracellular fractions when expressed in *P. pastoris*. In contrast, the insect cell system yielded expression of the full-length protein, but with very low levels of the globin domain. However, the full-length protein also showed complexity in its expression as a time-dependent truncation for the full-length protein was observed (Bracke *et al*., 2018). This is likely a result of proteolysis by intracellular proteases or by auto-cleavage by the N-terminal calpain-like domain as reported (Bracke *et al*., 2018). In addition, the full length of the protein is large for recombinant expression systems so there have been problem with expressing the full size of the protein in *E. coli*. When expressed in multiple expression systems, it was still unsuccessful possibly due to the complexity of the globin domain structure. For example, the circular permutation of the globin adds some extra complexity in the expression.

As a result, in addition to classic recombinant *E. coli* expression systems, eukaryotic mammalian host cells systems (HEK293T cells) were used for protein expression via transient transfection. These mammalian expression systems support more complex post-translational modifications and possess superior protein folding machinery for mammalian proteins. By utilising this system, the aim was to express the complex, full-length Adgb and the Adgb globin domain. This approach enabled the investigation of protein interactions both *in vitro* and *in vivo*, while also providing insights into the potential effects of post-translational modifications of Adgb.

2.2 Materials and Methods

2.2.1 Materials

All primers were synthesised by Eurofins Genomics Europe. The plasmid for the Globin domain of androglobin was synthesised by Epoch Life Science, Texas, USA. Proximal and distal heme ligand mutants were produced by site directed mutagenesis using Agilent Quikchange II method. Human cytoglobin in pET28a was synthesised by Epoch Life Science (USA) with codon optimisation for expression in *E. coli*. The neuroglobin (Ngb) gene in pET28a was a kind donation of Prof Worrall, University of Essex. Protein ladder used is ThermoFisher PageRuler™ Prestained Protein Ladder, 10 to 180 kDa protein ladder. Kanamycin sulphate and Luria-Bertani (LB) media used were purchased from ThermoFisher Scientific.

2.2.2 Expression of the Globin domain of Androglobin, Cytoglobin and Neuroglobin in *Escherichia coli*

All three globins were expressed in *E. coli* using essentially identical protocols. All three genes which included sequences of the Adgb globin domain, Ngb and Cygb, each contained in the pET28a vector, were transformed into *E. coli* BL21 (DE3) competent cells (0.1 mL) by heat shock method which involves incubating on ice for 30 min then heat shock at 42 °C, 90 s. The cells were returned to ice for 2 minutes, followed by the addition of 0.4 mL of LB media. They were then incubated at 37 °C with shaking at 180 rpm for 50 minutes.

The cells were plated on an agar plate containing 50 µg/ml kanamycin sulphate and incubated overnight (16 - 18 h) at 37 °C. A colony from each plate was grown overnight (16 - 18 h) in a shaking incubator (180 rpm) at 37 °C in LB media (100 mL containing 50 µg/mL

kanamycin sulphate). Subsequently, 10 mL of the cell suspension was split and transferred into 4 x 1.4L LB flasks containing 50 µg/ml kanamycin sulphate and grown in a shaking incubator (180 rpm) at 37 °C. Induction of protein expression was initiated after achieving an optical density of A600nm: ~0.8 – 1.0. IPTG (500 µM) was used to induce the cells containing Adgb globin domain, Ngb and Cygb genes. Ferric citrate (100 µM) and 5-aminolaevulinic acid (ALA) (250 µM) were used to augments heme synthesis. Carbon monoxide (CO) was bubbled through the media for 1 minute to stabilise the heme binding to the globin. The CO also prevents redox chemistry of the heme with oxygen, thus preventing autoxidation and radical damage that occur following the autoxidation. The flasks were then sealed with a rubber bung and incubated in a shaking incubator (180 rpm) at 37°C overnight.

The cells were then harvested by centrifugation (4000 g, 20 min, 4 °C). The pellets were resuspended in media and transferred to a 50 mL falcon tube where they are re-centrifuged (4000 g, 20 min, 4 °C) and frozen at -80 °C for storage. The concentration of Ngb, Cygb and ferrous Mb were determined by measuring the absorbance of their heme Soret bands. Extinction coefficients used were $\varepsilon_{414\ nm} = 129\ mM^{-1}cm^{-1}$ (Lardinois *et al*., 2008) for Ngb, $\varepsilon_{428\ nm} = 165\ mM^{-1}cm^{-1}$ (Beckerson *et al*., 2014) for Cygb, and $\varepsilon_{435\ nm} = 121\ mM^{-1}cm^{-1}$ for Mb (Antonini and Brunori, 1971 & Gonzalez *et al*., 2011), respectively.

2.2.3 Purification of Globin domain of Androglobin, Cytoglobin and Neuroglobin expressed in *Escherichia coli*

Frozen cells were thawed and resuspended in 20 mM NaPi 500 NaCl pH 7.4 buffer before being subjected to lysis using an Avestin Emulsiflex C3 homogenizer at ~15000 psi, 3 passes. The lysates were clarified by centrifugation (40000 g, 30 min, 4°C, (or 40000 g, 30

min, 22°C for Adgb globin domain)). All proteins were His-tagged, enabling the purification of supernatants through nickel affinity column chromatography.

Ngb and Cygb were purified using a GE Healthcare immobilised metal (nickel) affinity column via the AKTA pure chromatography system at 4 °C. During the purification, the 5 mL nickel column was pre-calibrated with buffer (20 mM NaPi with 500 mM NaCl, pH 7.4). Following loading of the cell lysate, the proteins were washed with 20 mM NaPi 500 mM NaCl 20 mM imidazole, pH 7.4 buffer before eluting with elution buffer (20 mM NaPi 500 mM NaCl 500 mM imidazole, pH 7.4 buffer) in a linear gradient over 10 column volumes (50 mls). Imidazole was removed by dialysis (8000 Da MWCO dialysis tubing, buffer: 1 mM sodium tetraborate, pH 10, three changes; a change approximately every 1.5 h and finally a change overnight in a 4 °C cold room).

Adgb GD was also purified using a GE Healthcare immobilised metal (nickel) affinity column; however, this procedure was more complex due to the native instability of the protein domain. This added complexity arose from the necessity of incorporating SDS at a final concentration of 1 % into the purification process to stabilise the protein during purification, while ensuring that neither the column with the pre-calibrated buffer nor the supernatant after centrifugation precipitated due to the presence of SDS. Four different methods for purifying Adgb-GD were evaluated and compared based on the quality and yield of the protein they produced. Additionally, their practicality and ease of use were assessed, considering the mentioned limitations.

i. The first method involved purification using an AKTA system. Initially, Adgb is loaded onto the nickel column via peristaltic pump with a loading buffer containing 20 mM NaPi, 500 mM NaCl, and 0 mM imidazole at pH 7.4. The column is then transferred to the AKTA system, where it is washed with 20 mM NaPi, 500 mM NaCl, and 20 mM imidazole. Finally, the protein is eluted using a linear gradient of 20 mM NaPi, 500 mM

NaCl, and 500 mM imidazole over 10 column volumes (50 ml) and then dialysed under similar condition as Ngb and Cygb (8000 Da MWCO dialysis tubing, buffer: 1 mM sodium tetraborate, pH 10, three changes; a change every 1.5 h and finally a change overnight in a cold room). SDS precipitates at cold temperatures, but was not included in the first purification method, eliminating restrictions on its ease of use. Therefore, purification with AKTA system under cold condition is possible. Purification with an AKTA system ensures quality control and enhances protein purity (via linear gradient elution) compared to other methods.

ii. A second method utilised only a peristaltic pump connected to the column for protein purification. SDS was added to the lysed cells to a final concentration of 1% before centrifuging at 40000 g, 30 min, 22 °C. Following this, purification was performed in water baths maintained at 30 °C to prevent SDS precipitation. However, it introduced logistical challenges due to the additional setup required. This process includes loading the protein onto the nickel column with loading buffer (1 % SDS 20 mM NaPi with 500 mM NaCl and 0 mM imidazole, pH 7.4) via the peristaltic pump. The column is then washed with the loading buffer followed by the washing buffer (1 % SDS 20 mM NaPi with 500 mM NaCl and 20 mM imidazole, pH 7.4 (in excess of 10 column volumes (100 % gradient)) and subsequently eluted in a single step using the elution buffer (1 % SDS 20 mM NaPi with 500 mM NaCl and 500 mM imidazole, pH 7.4 (15 ml)) also via the peristaltic pump. The eluted sample is then dialysed under similar condition as (i) to remove imidazole and SDS. The dialysed protein was centrifuged to remove precipitated SDS and protein. The soluble protein was then concentrated using Merck Amicon™ Ultra-15 Centrifugal Filter Units.

iii. A third method also used a peristaltic pump to purify the protein like in method (ii) for similar reasons but utilises the same buffer to load the protein onto the nickel column

and for washing which was 1 % SDS 20 mM NaPi with 500 mM NaCl and 20 mM imidazole, pH 7.4. The rationale was to use low concentration of imidazole in the lysis process which displaces the weak interaction of endogenous proteins and thus minimizing unwanted protein binding to the column. Additionally, due the lysate had been preconditioned in low imidazole, it would theoretically enhance selectivity and ensure less contaminants stick to the column in the first place and thus the wash and elution steps become more effective. Elution is also in a single step using the elution buffer (1 % SDS 20 mM NaPi with 500 mM NaCl and 500 mM imidazole, pH 7.4) over 10 ml via a peristaltic pump. The eluted sample is then dialysed and concentrated under similar condition as method (ii). The key difference between Method ii and Method iii lies in the timing of introducing the imidazole-SDS mixture. In Method ii, imidazole is introduced to SDS during the washing phase, whereas in Method iii, it is added immediately after lysis. Since this mixture increases the risk of precipitation, it was best to delay its addition for as long as possible. Precipitation can complicate the purification process, reducing ease of use and introducing logistical challenges, such as clogging or damaging the column.

iv. The final method utilised urea to purify the protein using similar processes mentioned above under nondenaturing conditions with an AKTA system where the resuspension and loading buffer is 8 M urea 20 mM NaPi with 500 mM NaCl and 0 mM imidazole pH 7.4. The washing buffer is 8 M urea 20 mM NaPi with 500 mM NaCl and 20 mM imidazole pH 7.4 and finally the elution buffer is 8 M urea 20 mM NaPi with 500 mM NaCl and 500 mM imidazole pH 7.4. The eluted sample is then dialysed under similar condition as (i) to remove imidazole and urea. Since SDS was not used, it also eliminates restrictions on its ease of use making purification possible using the AKTA system.

2.2.4 Calculating the theoretical molecular weight of Androglobin

The purified proteins retained their histidine tags, which were not removed post purification process. Consequently, the molecular weights were calculated with the histidine tags included. Their theoretical molecular weights were calculated by uploading the protein and His-tag sequence into the Expasy server (https://web.expasy.org/compute_pi/) and choosing average M_W calculation.

2.2.5 Site directed mutagenesis of Androglobin Globin Domain

Single point mutations to the Androglobin globin domain were created by site-directed mutagenesis. These included mutations to the heme iron distal ligand Q792G and Q792H in addition to H824C and H824Y for the heme iron proximal histidine ligand. These mutants were expressed in BL21 (DE3) using the methods mentioned above (section 2.2.2) and are purified using the method of Adgb purification that included SDS (method #ii chapter 2.2.3).

2.2.6 Heme reconstitution of the globin domain

Following cell lysis and purification, the heme moiety is lost and the protein purified is that of the apo-protein. Therefore, heme needed to be reconstituted. The concentration of the apo form of the globin was measured at 280 nm using an extinction coefficient of 32.68 $mM^{-1}cm^{-1}$. The extinction coefficient was calculated by inputting the sequence of the protein (including His-tag) into Expasy server (Expasy - ProtParam) (Gasteiger *et al*., 2003). Based on this concentration, the amount of heme needed to reconstitute into the protein was calculated. The optical properties for a concentration of the reconstituted ferric Adgb-GD were obtained

using the Agilent Cary 50 UV-Vis spectrophotometer set to scan between 350 – 700 nm. The apo form of the WT globin (5 μM) was diluted in 0.1 M NaPi buffer at pH 7.4 and was titrated with heme. The heme solution was prepared by dissolving bovine hemin (Sigma-Aldrich, UK) in 5 mM NaOH. Undissolved hemin was removed via centrifugation. The concentration of hemin in this stock solution was determined using the extinction coefficient ϵ_{403nm}=170 mM^{-1} cm^{-1} in 80 % DMSO (Miller *et al*., 1999).

2.2.7 Determining the extinction coefficient of holo Adgb GD

The concentration of holo Adgb GD was calculated using Agilent 1290 ultra high-pressure liquid chromatography (UHPLC) fitted with an Agilent 1260 diode array spectrophotometer. The column used was a reverse-phase was Agilent Zorbax 300 Å StableBond C3, 4.6 mm x 250 mm, with 4.6mm x 12.5mm guard column. The solvents used were (A) 0.1 % trifluoroacetic acid (TFA) and (B) acetonitrile containing 0.1 % TFA. The gradient used was adapted from Reeder *et al*. 2001 where the gradient was initially set at 35 % solvent B (for 10 minutes) and then increased to 37 % (5 minutes), 40 % (1 minute) and finally 43 % (10 minutes). The flow rate was 1 mL min^{-1} at 25 °C (Reeder *et al*., 2001). The heme concentration was calculated from the integrated area under the peak and compared to the integral of heme peak of known concentration of Mb. This heme concentration was then used to calculate the extinction coefficients from the optical spectra of Adgb WT and mutants from the same sample that was analysed by UHPLC. This method is similar in principle to the pyridine hemochromagen assay (Barr *et al*., 2015). Mb used in this experiment was from horse (equine) Mb.

2.2.8 Size exclusion chromatography

Size exclusion chromatography (SEC) was carried out with Agilent 1290 infinity UHPLC with a 1260 diode array detector. Proteins with known molecular weight were used as a control at 2 mg/ml and they include Bovine Serum Albumin (BSA), Mb, Cytochrome c and ferric cyanide. This is to calibrate the size exclusion column and also provide a platform to calculate the molecular weight of Adgb. Isocratic mobile phase buffer was 0.1 M NaPi at pH 7 (and high salt buffer 150 mM NaCl). The column used was Agilent Advance BioSec 130 Å, 2.7 µm, 7.6 mm x 300 mm.

2.2.9 SDS-PAGE analysis

The supernatant and pellet from each protein sample were analysed by SDS-PAGE to check for expression of the corresponding recombinant proteins. The samples were prepared under reducing conditions (2x loading dye (0.5 M Tris pH 6.8, 10 % SDS, containing bromophenol blue (0.2 %), 20 % glycerol, 10 % v/v β-mercapto-ethanol) in a 1:2 (sample: reducing solution) ratio and incubated at 99 °C for 2 min. Denatured proteins were then separated by electrophoresis (150 V, 60 minutes) using a 15 % acrylamide resolving gel with 5 % stacking gel. Following electrophoresis, the gel was washed three times with water for five minutes on a rocker (20 rpm) before staining with Biosafe Coomassie stain for 1 h (gentle rocking, 20 rpm). Following staining, the gel was washed with water for 30 minutes to remove excess stain before storage in water. Visualisation was performed by imaging via Fusion Fx Vilber Lourmat imaging system or an iPhone camera system.

2.2.10 Western blot

Throughout this study, Western blotting was used to detect His-tagged fragments using a 6X His-Tag monoclonal antibody (His.H8) conjugated to HRP (ThermoFisher), unless stated otherwise. This process involves rinsing the gel briefly in water (3 times, 5 minutes each) and then equilibrating the gel in chilled transfer buffer (25 mM Tris, 192 mM glycine, 0.1 % SDS and 20 % methanol) for 15 minutes. Additionally, PVDF membranes (0.45 μm) were pre-wetted in 100 % methanol for 1 min to activate the membrane and then rinsed in water prior to its equilibration. PVDF membranes were equilibrated by soaking in chilled transfer buffer (25 mM Tris, 192 mM glycine, 0.1 % SDS and 20 % methanol). Proteins were transferred to the PVDF membrane using a BioRad Mini Trans-Blot cell according to manufacturer's instructions (wet transfer) with electrophoresis at constant 100 V for 60 minutes. Following transfer, the membrane was blocked overnight with blocking buffer (10 % w/v skimmed powdered milk and 0.1 % w/v tween-20 in phosphate-buffered saline (PBS)). The membrane blot was washed twice in PBS for 15 minutes to remove all traces of blocking buffer. 6X His-Tag monoclonal antibody (His.H8) conjugated with HRP from Thermofisher, UK was diluted 1:5000 in phosphate buffered saline supplemented with 0.1 % tween-20 (PBS-T) and incubated with the membrane for one hour. The membrane was washed five times with phosphate buffered saline supplemented with 0.1 % (w/v) Triton X-100 for at least 5 min per wash. To generate a chemiluminescent signal, SuperSignal West Pico PLUS chemiluminescent substrate was used according to manufacturer's instructions. Visualisation was performed by imaging via Fusion Fx Vilber Lourmat imaging system.

2.2.11 Mammalian cell protein expression

Human Embryonic Kidney cells (HEK293T, ATCC) were cultured and maintained in culture medium DMEM supplemented with 10 % foetal bovine serum (FBS) and 50 µg/mL gentamicin. They were incubated at 37 °C in a humidified atmosphere with 5 % CO_2. Prior to transfection, cells were seeded onto a 6 well plate at density 1 x 10^6 per cm^2 of the dish surface area. The cells were transfected using a calcium phosphate precipitation method of DNA delivery (Koay et., 2021) where overnight medium is replaced with cell medium containing 25 µM chloroquine. Transfection mixture consisting of 50 % (v/v) 1.5 µg plasmid DNA in 250 mM $CaCl_2$ and 50 % (v/v) HBS buffer (50 mM HEPES, 280 mM NaCl, 10 mM KCl, 1.5 mM Na_2HPO_4, 12 mM glucose), was mixed and incubated for five minutes before adding to the cell media to a final volume of 10 % (v/v). Plasmid DNA vector (pcDNA 3.1+) contained the GFP-tagged gene for Adgb-GD, Ngb, or Mb was transfected into HEK293T cells. An additional control plasmid containing only the gene for GFP-protein and an empty vector (pcDNA 3.1 plasmid, Thermo-Fisher, UK) was also used. Transfected cells were incubated at 37 °C, 5 % CO_2 for 24 h before the chloroquine containing medium was replaced with fresh DMEM without chloroquine. Cells were continued to be incubated for a further 24 - 48 h. Imaging was done using EVOS M5000 instrument (ThermoFisher Scientific) using the brightfield and GFP (482 nm excitation / 524 nm emission) fluorescence filter.

2.2.12 Spectroscopic analysis of cell protein expression *in situ*

In situ spectroscopy of HEK293T cells with overexpressed globins were measured by an OLIS RSM 1000 CLARITY spectrophotometer with light-integrating sphere. Prior to this, the optical spectrum for known concentration of Mb was taken with an Agilent Cary 50 UV-Vis-NIR spectrophotometer. The optical spectrum of the same protein sample was then

measured using the OLIS RSM 1000 CLARITY spectrophotometer. The light integrating sphere has a variable pathlength, so the Mb was used to calibrate the Olis spectrophotometer according to the Javofi method (Javofi *et al*., 2005).

After cell samples expressing proteins were prepared using the methodology above (section 2.2.11), the DMEM media was removed, and the cells were washed with PBS. Cells were harvested 48 hours post-transfection by gently pipetting up and down in 1x PBS, allowing them to be resuspended in the well without the use of trypsin. The optical spectra of the cell suspensions were measured on the OLIS RSM 1000. The scans are repeated 4 times for all samples at 413 nm. The photomultiplier tube (PMT) high volts (HV) value was adjusted to value above 7 to maximise transmission at wavelength 350 – 700 nm for 350 points. All samples were analysed in PBS. Measurements were taken for the protein present in its native oxidation state in the cells suspension, followed by the deoxyferrous state of the protein subsequent to the addition of dithionite, and the CO-bound state after gently bubbling ~1 ml CO gas through each sample. Quantitative spectrophotometry analysis using integrating cavities (Javofi *et al*., 2005) was used to analyse the spectrum produced by the OLIS RSM 1000. The plotted data was fitted to the inverted expression $A = \frac{1}{a1}[\left(\frac{I0}{I}\right)^{\Upsilon} - 1]$. Where A = absorbance, a1= offset constant, ϒ= constant. Also $\frac{I0}{I} = e^{A}$ and A is the absorbance of the plots (Javofi *et al*., 2005).

2. 3 Results

2.3.1 Recombinant expression of Androglobin protein domains and other globins in *Escherichia coli.*

The *E. coli* expression system is often unsuitable for producing large and complex proteins, such as the 190 kDa full-length Adgb protein, due to the absence of organelles like the endoplasmic reticulum and Golgi apparatus, which are essential for proper protein folding (Dang, 2024). Consequently, individual domains of Adgb (globin domain) were isolated and expressed separately using the *E. coli* system. However, despite this approach, prior attempts to express the Adgb globin domain in *E. coli* were unsuccessful (Bracke *et al*., 2018). It was not possible to purify the globin domain as a soluble protein, even under modified conditions, such as the use of denaturing agents like 8 M urea (Bracke *et al*., 2018). In this experimental chapter, globin domain of Adgb was expressed using a reassessed alignment sequence Adgb-GD gene (Reeder *et al*., 2024), and its behaviour during expression in the *E. coli* expression system was evaluated.

Following the transformation of pET28a vectors containing Ngb and Cygb gene in *E. coli* BL21 (DE3) competent cells, followed by incubation and protein induction as detailed in method section 2.2.2, the globin proteins were successfully expressed and purified. Cygb is known to be expressed in monomeric and disulfide-linked dimeric forms (Beckerson *et al*., 2014). Some preparations exhibit primarily dimeric forms (Lechauve *et al*., 2010), whereas our expression method yields primarily monomeric protein. Monomeric Cygb can be separated from the dimeric Cygb by running the purified protein on a G75 column, however this was not executed in this study. Soluble fractions from the purification were screened using SDS-PAGE and their gel are shown in Figure 2.3.1.1 Analysis showed typical bands that can be observed for soluble protein i.e. Ngb and Cygb. Following analysis of the gel using gel analyser software,

the protein was confirmed to be ~19 and 26 kDa for Ngb and Cygb respectively, which is close to the theoretical molecular weight (MW) 16.9 and 21.4 respectively kDa + 2.0 kDa Histag (Expasy).

The expression of Adgb GD in *E. coli* proved more challenging to express and purify. This is partly due to the circular permutation of the globin domain which decreases the stability of the protein, plus this is a single domain of the complete protein being expressed. Other laboratories have attempted to express this protein but have been unsuccessful (Bracke *et al*., 2018). Following induction of the pET28a vector containing the Adgb-GD gene with the realigned sequence identifying an alternative sequence corresponding to the C helix sequence (Reeder *et al*., 2024), Adgb-GD protein expression was not observed in lane A of Figure 2.3.1.2A showing that protein was either not expressed or insoluble. This lane corresponded to the results following the purification method (i) in section 2.2.3 where SDS was not used. It can be observed that the cell pellet showed a band corresponding to the Adgb-GD predicted molecular weight, but not the cell lysate. This confirmed that in the absence of SDS the protein is insoluble. Soluble Adgb-GD protein was detected following the use of method (ii) and (iii) using SDS as described in section 2.2.3 (lanes B and C respectively in Figure 2.3.1.2A) with lane B producing a purer protein compared to lane C. After concentrating the protein following purification using method #2, darker band and a dimer was detected at lane normal GD. There's also the presence of 2 band in the empty pET vector.

The calculated MW of Adgb-GD including the his-tagged was 25.9 kDa. Following analysis of the gel, the protein migrated to a MW of ~23 kDa, close to the predicted MW. The Western Blot also shows a single band at the approximate position 23 kDa (Figure 2.3.1.2B). This result is within an acceptable range of error. Factors such as shifts based on intrinsic net charge or aggregation can account for slight variations between observed and calculated protein MW in gel electrophoresis and western blot. Additionally excess SDS may also still be bounded

to the protein, potentially causing it to migrate faster. Denaturing concentration of urea was also used to purify the protein, and this is shown in Figure 2.3.1.3 (method iv, section 2.2.3). The purified sample purified with 8 M (Fraction 1) urea shows fewer bands with one main band at 23 kDa. This suggests that purification using urea is equally effective as purification using SDS, as demonstrated in the same figure for comparison.

Thus, it was established that method (ii) was an efficient method for generating high purity yield for the globin domain. This method was used for the expression of the mutants Adgb-GD Q792G, Q792H, H824C and H824Y. The four mutants showed the same migration and high purity compared to the WT protein, with the same MW as shown in Figure 2.3.1.1.

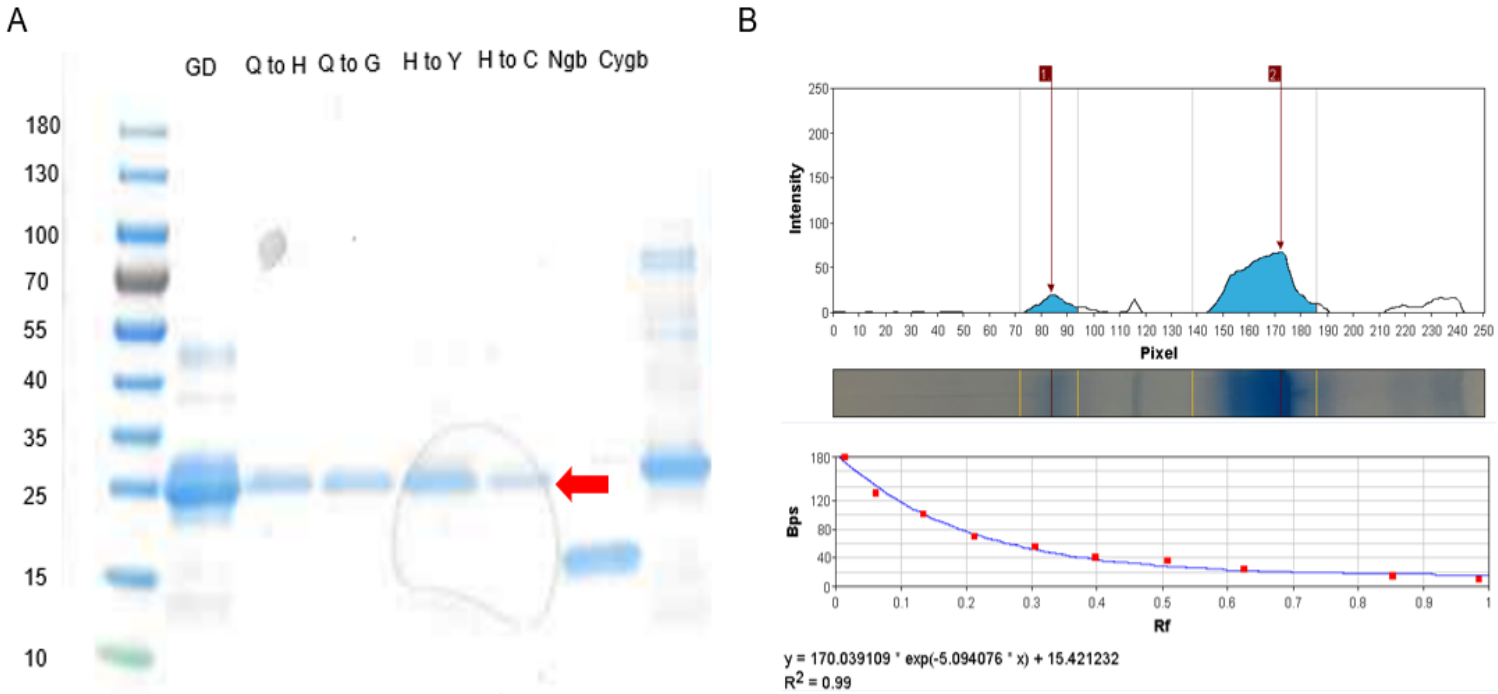


Figure 2.3.1.1. SDS-PAGE of recombinantly expressed globins. Wild type and mutants of the Adgb globin domain, Ngb and Cygb are shown.

The glutamine (Q) and histidine (H) mutants are the distal and proximal mutants respectively. The gel revealed mutants and Adgb-GD WT all expressed a band at ~23 kDa (red arrow) following analysis with gel analyzer (*GelAnalyzer v23.1.1*) (B). The image above also shows the expression of Ngb and Cygb, which were purified and used as positive controls. The Mw of the protein was confirmed to be ~19 and 26 kDa for Ngb and Cygb respectively. Theoretical Mw (including the his-tagged) for Ngb and Cygb is 18.0 kDa and 21.4 kDa respectively (Expasy).

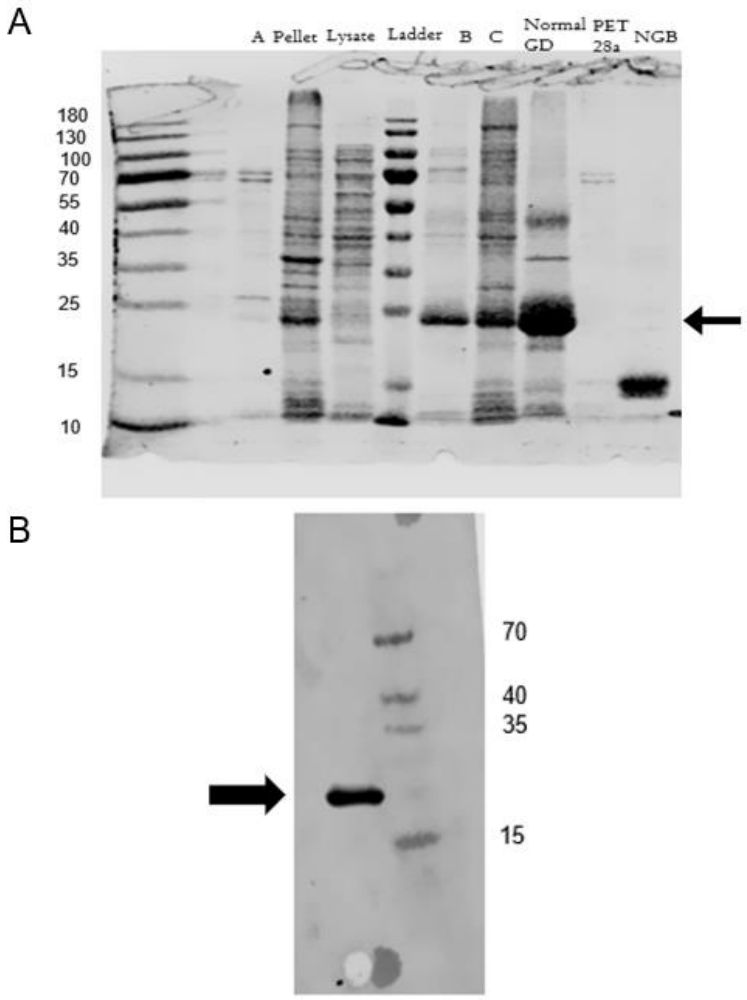


Figure 2.3.1.2. Recombinant expression of Androglobin globin domain in *E. coli* using the three purification methods described in section 2.2.3.

(A) SDS-PAGE of expression of wild type Adgb globin domain. lane A = Adgb globin domain purified with the first method (section 2.2.3), lane B = Adgb globin domain purified with the second method (section 2.2.3) and lane C = Adgb globin domain purified with the third method (section 2.2.3). Normal GD is a descriptive used specifically here to describe the concentrated protein from method (ii). The arrow shows Adgb globin domain expressed and purified at 23 kDa his-tagged included. (B) Western blot of the GD (purified from method #2) against the ThermoFisher BenchMark™ His-tagged protein standard. A single band was observed at approximately the Mw of the protein (23 kDa) as previously mentioned.

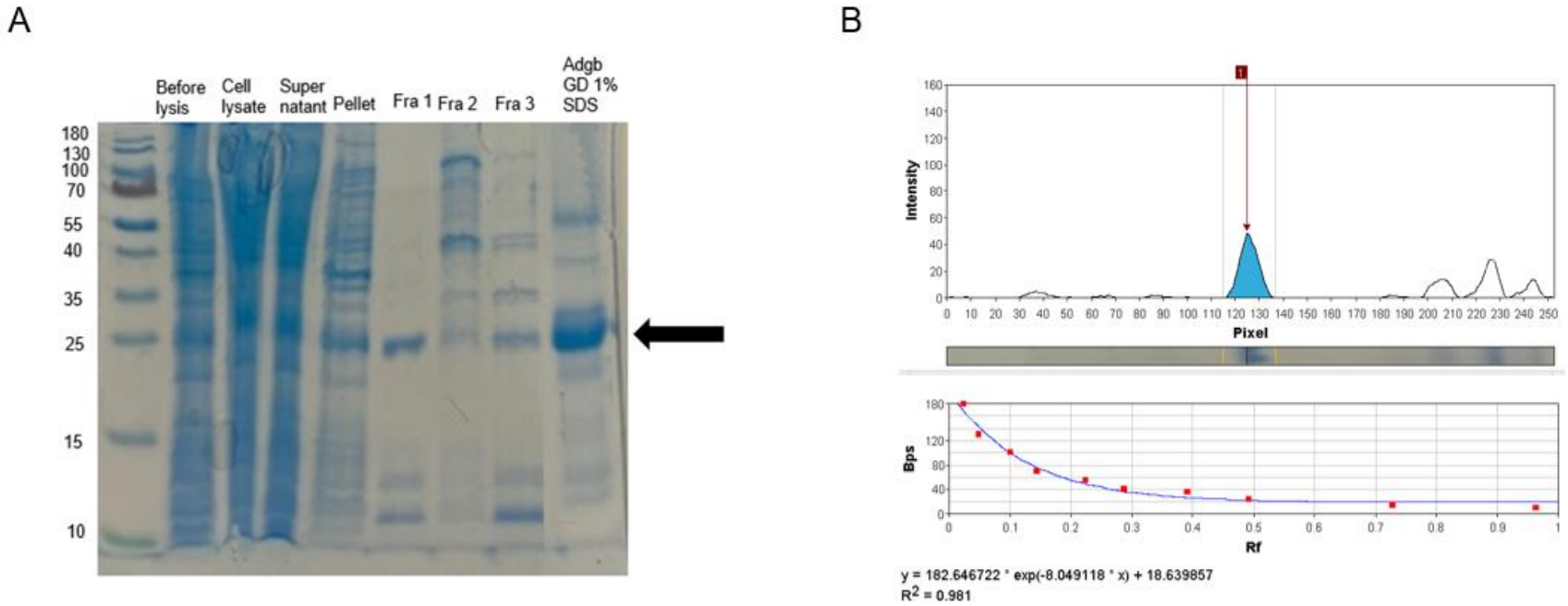


Figure 2.3.1.3. SDS PAGE analysis of Adgb globin domain purified with 8 M urea.

(A) The purity and expression level of the protein purified with urea is compared with the protein purified in 1 % SDS for the same amount of protein. The analysis shows higher purity (fraction 1 or fra 1) among the three fractions collected when the protein is purified with urea however with lower yield. (B) The molecular weight of the protein purified in urea (fraction 1) is 23 kDa following analysis with gel analyzer (*GelAnalyzer v23.1.1*). Fractions 1 – 3 represent distinct fractions collected at various peaks during purification using a linear gradient on the AKTA pure chromatography system.

2.3.2 Heme reconstitution following expression of the globin domain

The addition of 1 % SDS detergent (or 8M Urea) during the purification made the protein stable in solution, consequently making it possible to purify the soluble form of the protein. However, a consequence of this was that the heme was lost during purification. Therefore, after the surfactant was removed by dialysis, the apo form of the protein needed to be reconstituted with hemin to produce the holo protein.

The UV-vis optical spectrum of the protein was taken following expression and purification of Adgb-GD protein. The spectra demonstrated that the protein is primarily in the apo form, as indicated by the lack of heme spectra in the visible part of the spectrum (Figure 2.3.2.1). The reason is currently unknown however all current evidence points to heme loss during the purification stage of the protein, potentially resulting from the unusual circular permutation of the globin domain. There is a peak at 280 nm which indicates the presence of Adgb GD protein aromatic amino acids. Using the primary sequence of the protein, the approximate theoretical extinction coefficient of this apo-protein at 280 nm is 32680 M^{-1} cm^{-1}. This extinction coefficient was used to estimate the concentration of the apo-protein and hence the theoretical amount of heme required to saturate the protein. Equation 2.3.2.1 was used for this calculation.

$$Volume\ of\ heme\ to\ be\ added\ for\ reconstitution = \frac{total\ volume\ of\ apoprotein}{dilution\ factor\ (df)}\ where\ \ df\ is \frac{stock\ concentration\ of\ heme}{concentration\ protein\ in\ the\ cuvette}$$

Equation 2.3.2.1

Figures 2.3.2.2A and 2.3.2.3A display the optical changes observed during the titration to reconstitute the ferric and ferrous forms of heme into the apo-protein. The spectra showed a gradual shift at the 413 nm and 425 nm wavelength for the insertion of ferric and ferrous hemin

respectively to the apo-protein. There was a linear increase in the peak intensity at these wavelengths. This indicates that the rise in absorbance is an accumulation of product in the bound form i.e. heme bounded to Adgb-GD (holo-protein), correlating with the increasing amount of heme. A single shift in wavelengths was observed as shown in figure 2.3.2.2A and 2.3.2.3A when the protein becomes saturated with heme.

The plot of the difference between the two wavelengths (413 – 438 nm and 421 - 650 nm in the ferric and ferrous protein respectively) versus the heme bound to the protein in figure 2.3.2.2B and 2.3.2.3B offers clearer insight into heme binding, as it reveals the saturation concentration of heme in relation to the protein. The solid lines and dash lines represent heme insertion into globin (low heme concentration) and excess heme (high heme concentration) respectively. The heme should bind to the apoprotein in a 1:1 stoichiometry and this ratio was verified in the experiments. From this, it may be concluded that the protein was correctly folded following purification. It was confirmed from the optical spectra of the purified protein with 1:1 heme:protein reconstitution (*vide infra*). The concentration of the protein can be recalculated from the amount of heme that binds to the protein as shown in both figures (Figure 2.3.2.2B and 2.3.2.3B). Therefore, this result was used to calculate the heme concentration required to reconstitute the holo-protein in subsequent preparations of Adgb-GD. The findings of this study also indicate the excess heme is not in solution as exhibited by the emergence of a heme spectrum that does not conform to the free heme in solution. It is likely that there is a secondary binding site, likely characterised by low affinity binding to the surface of the protein. After reconstitution in a 1:1 heme to protein ratio, the protein gives a bright red colour as shown in Figure 2.3.2.4, similar to other hexacoordinate globins. Pentacoordinate ferric protein are brown.

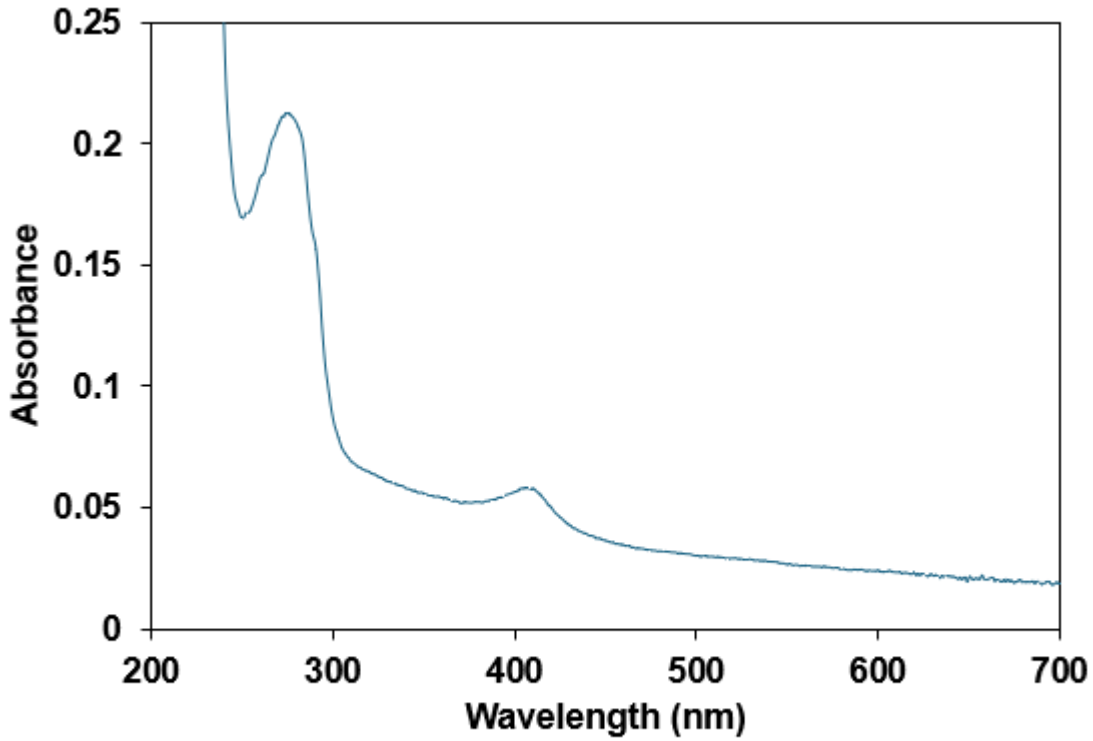


Figure 2.3.2.1 The Apo protein form of the globin domain of androglobin.

Extinction coefficient at 280 nm was calculated to be 32680 M-1 cm-1 using the primary sequence of the globin domain (Expasy server). The heme content is very low (< 5 %) confirming that the protein is primarily in the apo form following purification with SDS.

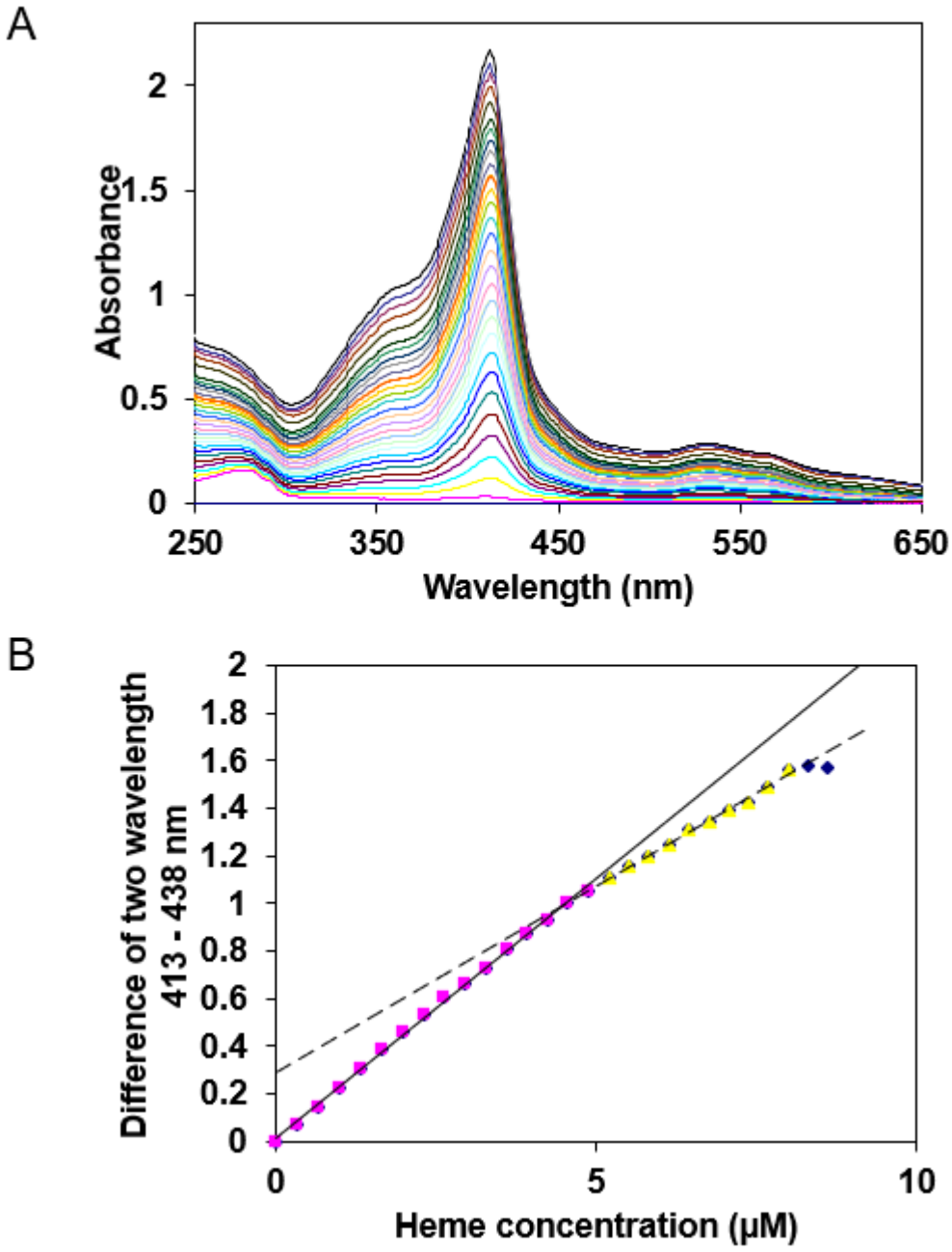


Figure 2.3.2.2. Ferric heme reconstitution of the apo WT Adgb GD 4.8 µM.

(A) The optical absorption spectra. Soret peak is at 413 nm in the spectra. (B) Plot of the difference of two wavelength 413 nm and 438 nm. Straight line = low heme concentration. dash line = high heme concentration.

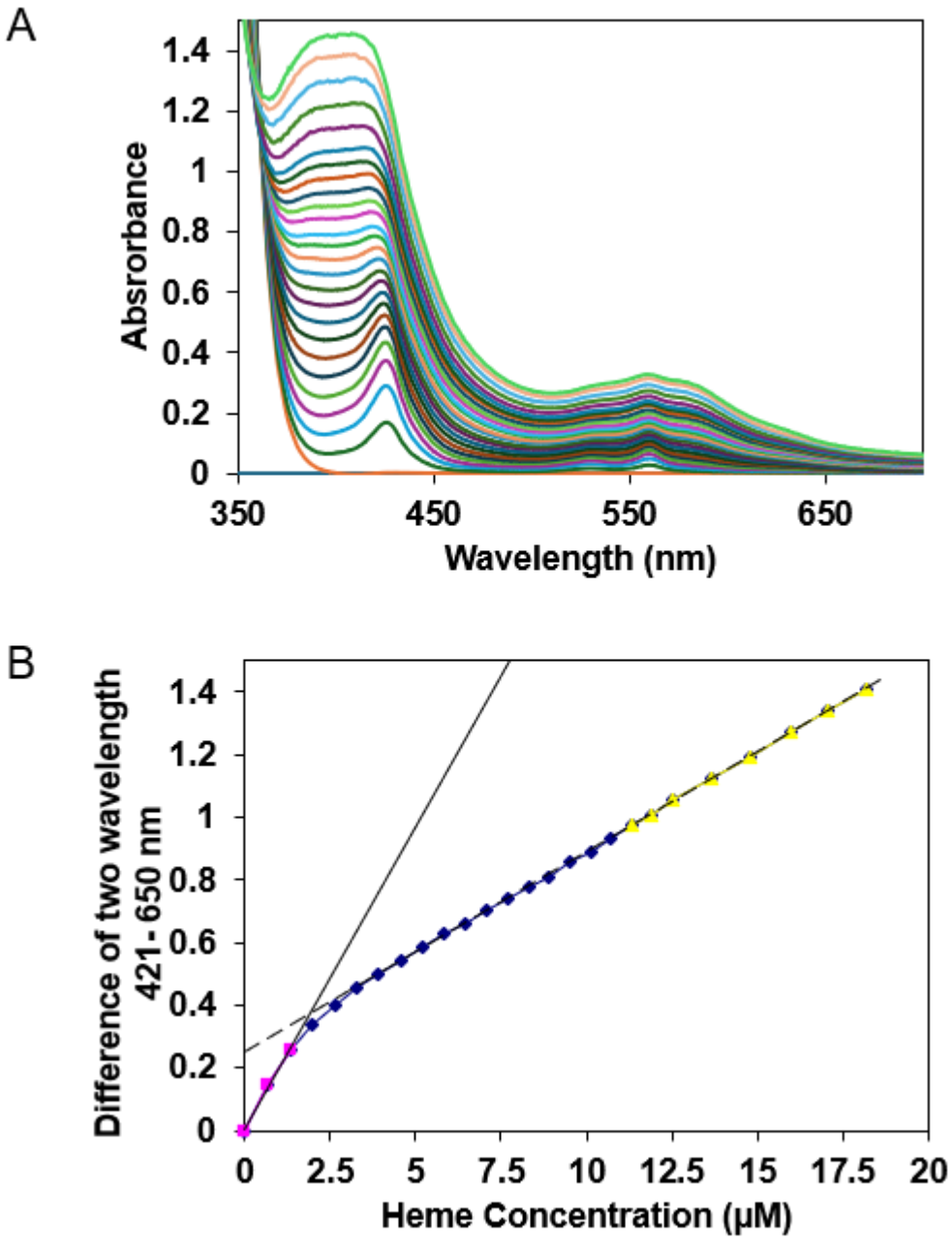


Figure 2.3.2.3. Ferrous heme reconstitution of apo WT Adgb GD (2.0 µM).

(A) Optical absorption spectra in the presence of dithionite. Soret peak is at 425 nm in the spectra. (B) Plot of the difference between two wavelength 421 nm and 650 nm. Straight line = low heme concentration. dash line = high heme concentration.

Figure 2.3.2.4 Photograph of ferric WT Adgb globin domain following heme reconstitution.

2.3.3 Optical characteristics and extinction coefficients of Adgb-Globin Domain.

After reconstitution of the apo-Adgb with heme, the optical spectra of the holo form of the protein (WT and mutants) and other heme proteins (Mb and Ngb) were recorded in the ferric (Fe^{3+}), deoxyferrous (Fe^{2+}) and deoxyferrous-CO (Fe^{2+}-CO) forms. Ferric and deoxyferrous Mb exhibited the typical pentacoordinate optical spectra, as shown in Figure 2.3.3.1A. The ferric Mb plot showed a Soret band at 409 nm and weaker Q bands at 502 nm and 630 nm. The ferrous Mb showed Soret band at 434 nm and a distinct Q band peak at 560 nm. Finally, the deoxyferrousCO-bound Mb has a Soret peak at 423 nm with Q bands at 543 nm and 580 nm which is indicative of a six coordinate species (Vickery *et al*., 1976). Ferric and deoxyferrous Ngb, in contrast, is a hexacoordinate protein, with its optical spectra presented in Figure 2.3.3.1B. Ferric Ngb showed a Soret band at 410 nm with weak Q bands at 530 nm and 560 nm which is indicative of a six-coordinate species as shown in the Figure 2.3.3.1B (Uno *et al*., 2004). The ferrous Ngb has Soret band at 425 nm with Q band at 531 nm and 560 nm which is a characteristic of hexa-coordination (Uno *et al*., 2004). The deoxyferrousCO-bound Ngb has Soret band at 417 nm and Q bands at 540 nm and 567 nm which indicate a hexacoordinate low spin state (Uno *et al*., 2004) (Table 2.3.3.1).

The WT ferric Adgb (Figure 2.3.3.1C) show a Soret band peak at 413 nm with weak Q bands in the visible region at 531 nm and 559 nm. This is typical of a hexacoordinate heme protein like Ngb shown in Figure 2.3.3.1B, as predicted previously (Hoogewijs *et al*., 2012). However, a small 630 nm component suggests that the protein has some pentacoordinate-like features. Overall, ferric Adgb show optical characteristics that is similar to those observed in ferric Ngb, conforming a primarily hexacoordinate heme iron geometry. In the reduced state, WT ferrous Adgb shows a right-shift in the spectra with Soret band at 425 nm. The Q bands appear more clearly in the visible region, with sharp peaks at 530 nm and 560 nm. These two

peaks are indicative of a hexacoordinate heme iron, almost identical to the optical features observed in Ngb. The WT deoxyferrous CO-bound Adgb show a Soret peak left-shifted of the deoxyferrous at 421 nm. There is also a right-shift in the Q bands at 539 nm and 570 nm, closely resembling the spectral features observed in both Mb and Ngb, typical of hexacoordinate CO bound spectra. This is also evidence that CO binds to Adgb similar to other globin's including Mb and Ngb. Overall, the spectral changes observed in Adgb is typical of hexacoordinate proteins like Ngb but with some minor pentacoordinate features for the ferric oxidation state, similar to that noted for cytoglobin (Cygb) (Sanctis *et al*., 2004).

Figures 2.3.3.2 show the optical spectra of Adgb-GD with proximal and distal mutations. Overall, the spectra of the ferric, ferrous, and deoxyferrous-CO states of the mutants are largely consistent with those of the WT protein, exhibiting similar characteristics with the only exception being the disappearance of the peak at 630 nm. Specifically, their Soret bands were observed at 413 nm, 425 nm, and 421 nm for their ferric, ferrous, and deoxyferrous-CO states, respectively, along with their corresponding Q bands. Their extinction coefficients are also described below (Table 2.3.3.1). The Q792G mutant (Figure 2.3.3.2A) demonstrates a marginally less pronounced Q band near the ~560 nm region in the ferrous form.

The optical spectra of the disulfide reduced Adgb-GD (TCEP-reduced) ferric, ferrous and deoxyferrous-CO states are shown in Figure 2.3.3.3. The Soret peak for the ferric, ferrous and deoxyferrous-CO was identical to the WT. However, the ferric spectrum differed significantly from the typical WT Adgb-GD spectrum: It exhibited an additional broad feature between 365 – 375 nm and changes to the Q-bands showing peaks at 630 nm and approximately ~490 nm which are close to the Q-bands of pentacoordinate heme iron e.g. Mb (Figure 2.3.3.1A). The ferrous and deoxyferrous-CO spectra however closely resembled those of the WT Adgb-GD. Therefore, while the ferrous spectrum remains hexacoordinate, the loss of the disulfide bond results to a significant shift to the pentacoordinate form in the ferric state.

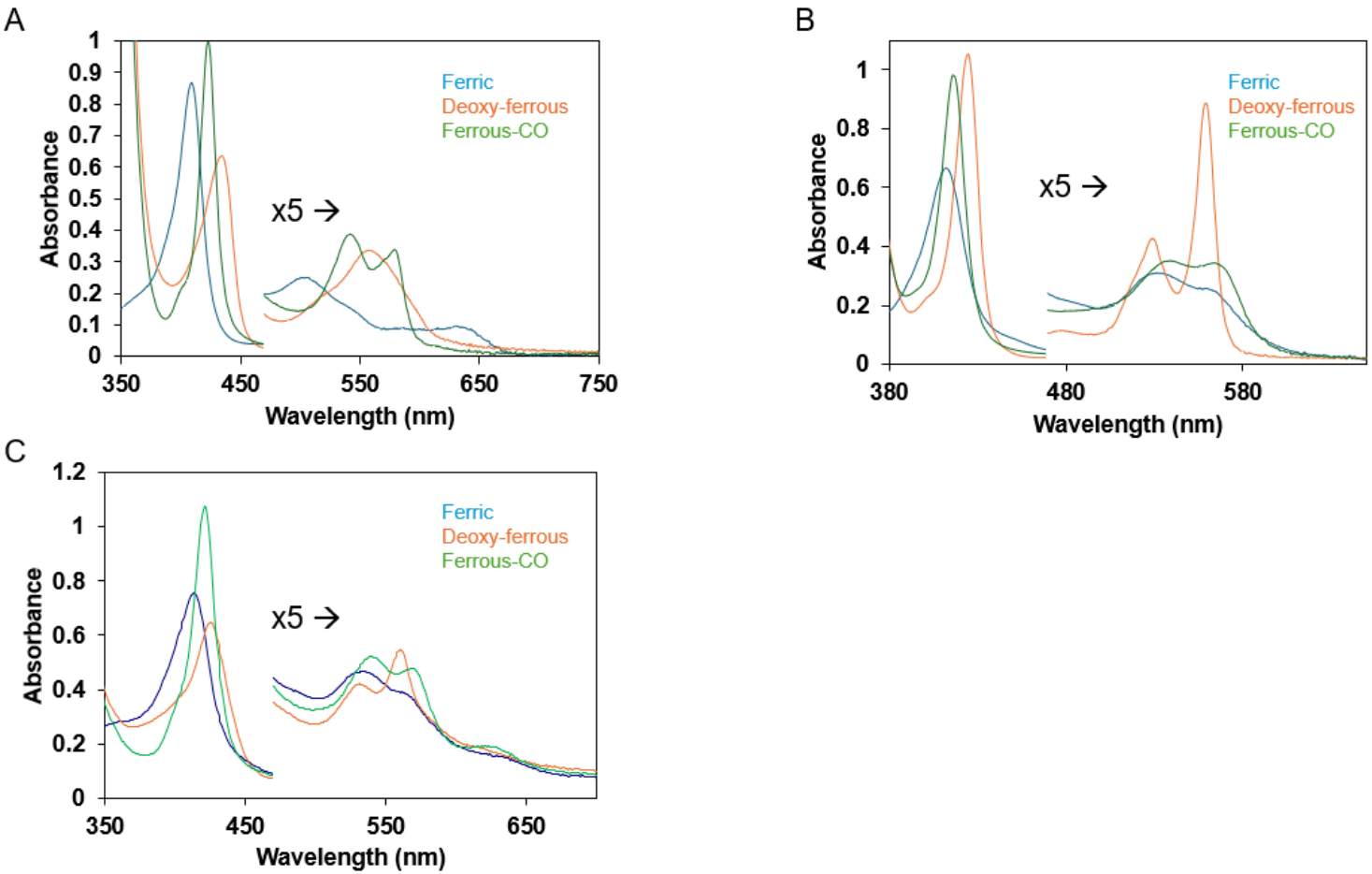


Figure 2.3.3.1 The optical absorption spectra of globins bound to heme (0.1 M NaPi buffer at pH 7.4), measured on Agilent Cary 50 UV-Vis Spectrophotometer.

(A) Equine Mb (5.3 µM). (B) neuroglobin (5 µM) (C) WT Adgb globin domain. The blue is the ferric, orange is the deoxy-ferrous and green is the ferrous-CO. Visible region has been enlarged by a factor of 5 for clarity. Changes in the Soret and Q bands between the ferric, ferrous, and CO-bound states reflect the different coordination environments and spin state of the heme iron during its transition among these states.

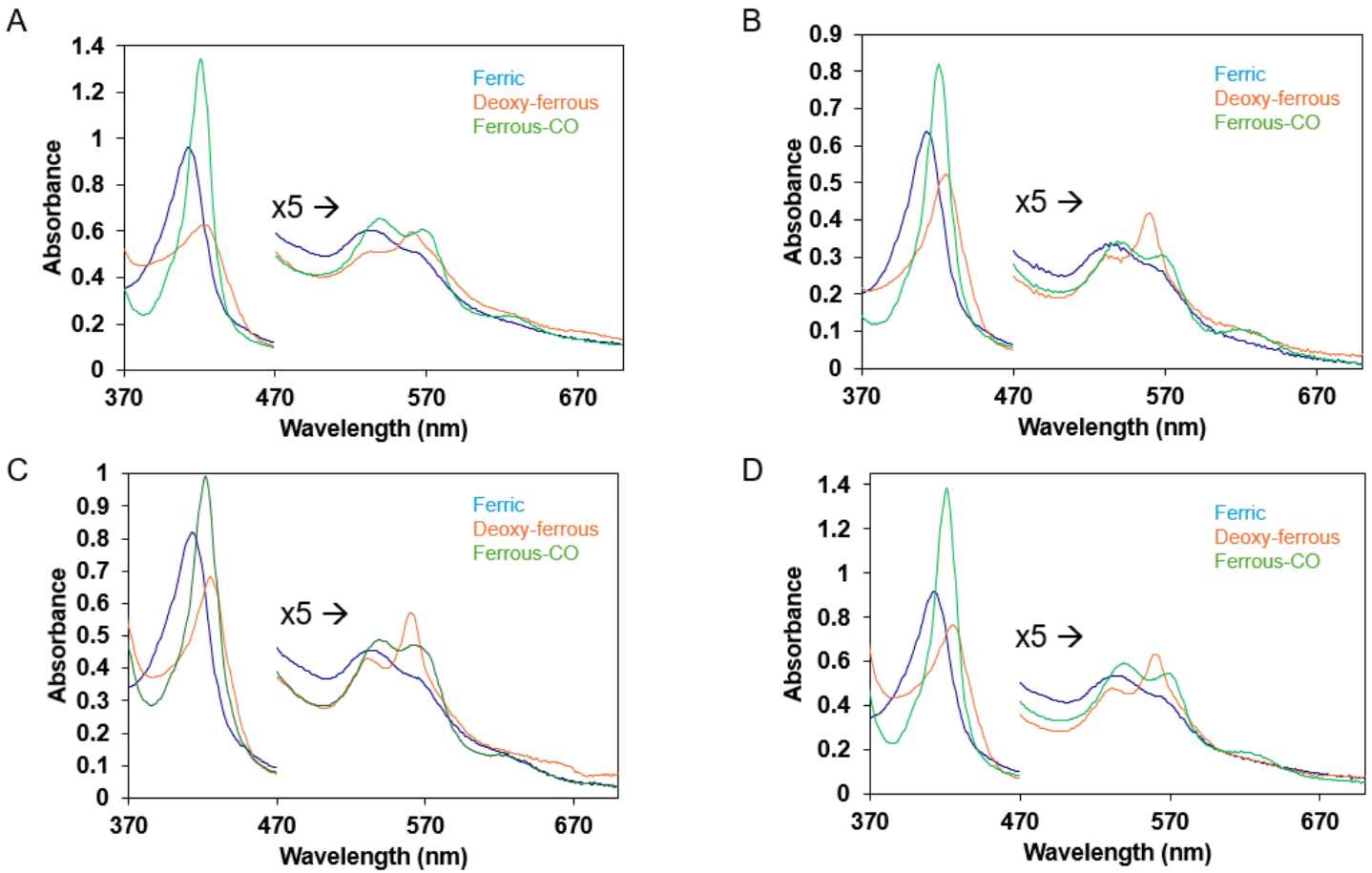


Figure 2.3.3.2 The optical absorption spectra of mutant Adgb-GD bound to heme (0.1 M NaPi buffer at pH 7.4), measured on Agilent Cary 50 UV-Vis Spectrophotometer.

(A) Mutant Q792G Adgb-GD protein. (B) Mutant Q792H Adgb-GD protein. (C) Mutant H824C Adgb-GD protein. (D) Mutant Q824Y Adgb-GD protein. The blue is the ferric, orange is the deoxy-ferrous and green is the ferrous-CO. Visible region has been enlarged by a factor of 5 for clarity.

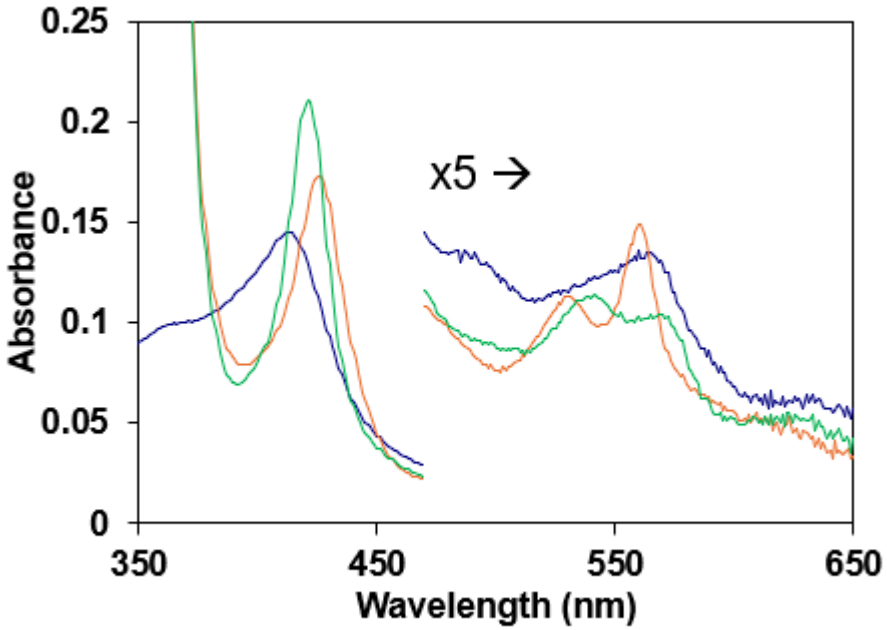


Figure 2.3.3.3 The optical absorption spectra of disulfide reduced (TCEP) Adgb-GD bound to heme (0.1 M NaPi buffer at pH 7.4), measured on Agilent Cary 50 UV-Vis Spectrophotometer.

The blue is the ferric, orange is the deoxy-ferrous and green is the ferrous-CO. Visible region has been enlarged by a factor of 5 for clarity.

Figure 2.3.3.4 show the chromatograms of WT Adgb GD compared to Mb at 280 nm and 400 nm following analysis with reverse phase HPLC as described in the methods (section 2.2.7). Two separate peaks were observed at 280 nm for Adgb-GD (Figure 2.3.3.4A), one corresponding to the heme (elution time = 14.82 mins) and the other to the apo form of the protein (elution time = 29.26 mins). This pattern is similar to that observed for Mb also shown in the same Figure 2.3.3.4A with the elution time for the heme essentially identical to that of the Adgb, but with an apoprotein peak at 22.64 mins, reflecting the different amino acid compositions of the two proteins. Therefore, this shows that both samples contain a heme and protein component. Only one peak showing the presence of heme was observed in the spectra at 400 nm for Adgb GD (elution time = 14.82 mins), which is essentially identical to that observed in Mb (elution time = 14.54 mins) as shown in the Figure 2.3.3.4B. This is expected as the heme for Mb and Adgb are both iron-protoporphyrin IX. The area under this chromatographic peak (400 nm) corresponds to the relative heme concentration. Conventionally, heme quantification involves acidification and extraction of the heme component into pyridine, followed by optical measurement of the pyridine fraction. This is the pyridine hemochromogen assay (Barr *et al*., 2015). In this experiment, however, the HPLC method employed acetonitrile/water/trifluoroacetic acid (TFA) for heme extraction from the protein, replacing the conventional pyridine/hydrochloric acid (HCl) approach, thereby enabling heme concentration measurement directly in the acetonitrile/water/TFA solvent. The exact heme concentration of Adgb GD was determined by:

$$[heme]\ in\ sample: \frac{heme\ peak\ area\ of\ Adgb\ sample}{heme\ peak\ area\ of\ a\ known\ [Mb]} * [Mb]$$

Equation 2.3.3.1

Where the heme peak area is calculated by the integration of the heme peak. The calculated concentration of the Adgb sample was then used to calculate the concentration of the Adgb protein (WT and mutants).

An optical spectrum of the same sample used for the HPLC analysis, ferric Adgb-GD, was taken with a UV-vis Spectrophotometer (Agilent Cary 50) and the dilution of the protein (heme) in the cuvette was calculated (*y*) i.e. how many times was it diluted. This information (*y*) was used to calculate the dilution factor (*k*, in μM), which was subsequently utilised to determine the extinction coefficient derived from the optical spectra of the corresponding sample:

$$k = \frac{concentration\ of\ heme}{y}$$

Equation 2.3.3.2

$$extinction\ coefficient = \frac{absorbance}{k}\ x\ 10^6$$

Equation 2.3.3.3

To summarise, the area underneath the heme peak was measured for Adgb-GD and compared to the Mb sample to determine the concentration of Adgb. This data was subsequently used to calculate the extinction coefficient from the optical spectrum obtained from the same sample for all transition states of WT Adgb; where the ferric Soret peak (413 nm) is 127.6 $mM^{-1}\ cm^{-1}$, ferrous (425 nm) is 109.2 $mM^{-1}\ cm^{-1}$ and the CO-bound Adgb (421 nm) is 181.5 $mM^{-1}\ cm^{-1}$ (Table 2.3.3.1).

Figures 2.3.3.5 and 2.3.3.6 show the chromatograms of Adgb GD mutants compared to Mb and Adgb WT at 280 nm and 400 nm respectively following analysis with reverse phase HPLC. There are no significant differences between the chromatogram of the WT and the mutants for the heme peak at 400 nm. The proximal chromatograms for H824Y and H824C, as

well as the distal Q792C, exhibited partial splitting of the heme peak due to a slight void to the guard column prior to its replacement. Additionally, due to an elongated method from 30 - 40 minutes the apoproteins are broader that observed in Figure 2.3.3.4. The extinction coefficients for the mutants are also listed in Table 2.3.3.1. Interestingly, the disulfide reduced version (using TCEP) of ferrous Adgb WT has similar extinction coefficient with the WT protein (109.2 mM^{-1} cm^{-1}). The chromatogram for the disulfide reduced version of Adgb WT compared to Mb at 280 nm and 400 nm are given in Figure 2.3.3.7. In summary, the heme peaks all showed the spectra for isolated heme and was used for protein concentration determination assuming that the heme bound to the protein in a 1:1 stoichiometry.

Table 2.3.3.1. Shows the calculated extinction coefficients and wavelength maxima of Adgb WT, mutants and other globins. They are calculated by averaging the extinction coefficients calculated after every independent cycle over their respective number of experiments to the nearest hundred (n=3).

Protein	Oxidation and ligand state	Soret peak, nm (mM^{-1} cm^{-1})	Q bands: Beta peak, nm	Q bands: Alpha peak, nm	Ref
Human WT Adgb-GD	Ferric	413 (127.6)	531	559	This study
	Ferrous	425 (109.2)	530	560	
	Ferrous-CO	421 (181.5)	539	570	
Distal Q792G	Ferric	413 (131)	531	559	This study
	Ferrous	425 (85.1)	530	560	
	Ferrous-CO	421 (183.6)	539	570	
Distal Q792H	Ferric	413 (135.5)	531	559	This study
	Ferrous	425 (111.5)	530	560	
	Ferrous-CO	421 (174.3)	539	570	
Proximal H824C	Ferric	413 (120.7)	531	559	This study
	Ferrous	425 (97.9)	530	560	
	Ferrous-CO	421 (142.6)	539	570	
Proximal H824Y	Ferric	413 (129.7)	531	559	This study
	Ferrous	425 (102.8)	530	560	
	Ferrous-CO	421 (183)	539	570	
Disulfide-reduced Adgb with TCEP	Ferrous	425 (109.2)	530	560	This study
Equine Mb	Ferric	408 (188)	502	630	E. Antonini and M. Brunori, 1971. Reeder *et al*., 2024.
	Ferrous	435 (121)		560	
	Ferrous-CO	424 (207)	540	579	
Human Ngb	Ferric	414 (129)	530	560	Nicolis *et al*., 2007.
	Ferrous	425	528	558	

	Ferrous-CO	420 (119)	538	561	Fago *et al*., 2004. Hamdane *et al*., 2003. Reeder *et al*., 2024 This study

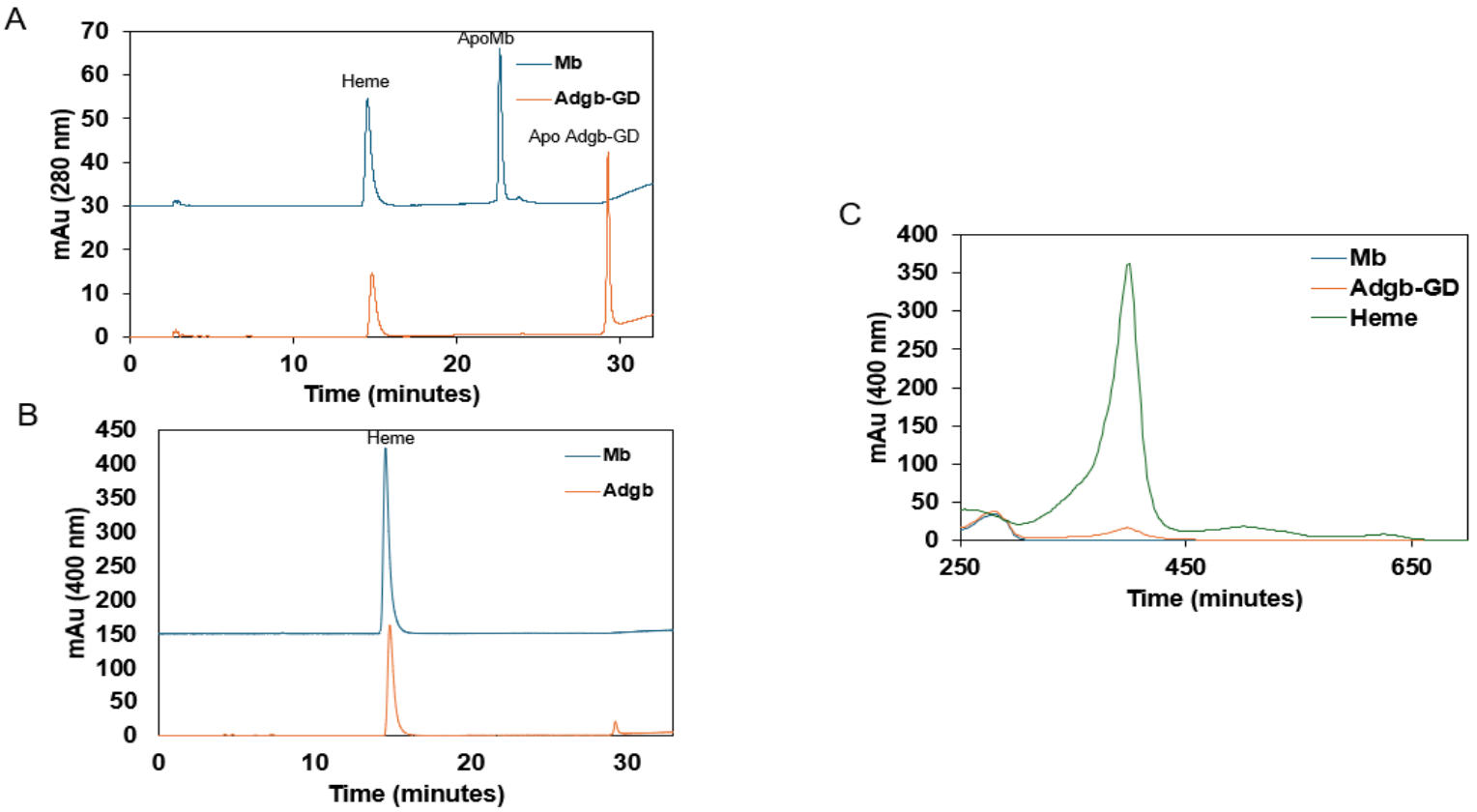


Figure 2.3.3.4 HPLC analysis of the heme content within the globin domain of WT androglobin following reconstitution, with Horse Mb (50 µM) used as a control for comparison against unknown concentration of Adgb.

The analysis was repeated with four independent samples. The area underneath the heme peak at (A) 280 nm (B) 400 nm of both samples Mb and Adgb GD was determined using Agilent Chemstation software and compared. The area under the heme peak at 400 nm was used to determine the extinction coefficient from the optical spectra obtained from the same samples using a UV-Vis spectrophotometer (Agilent Cary 50). (C) HPLC chromatogram spectra of the apoprotein forms of Mb and Adgb-GD. The traces were offset for clarity.

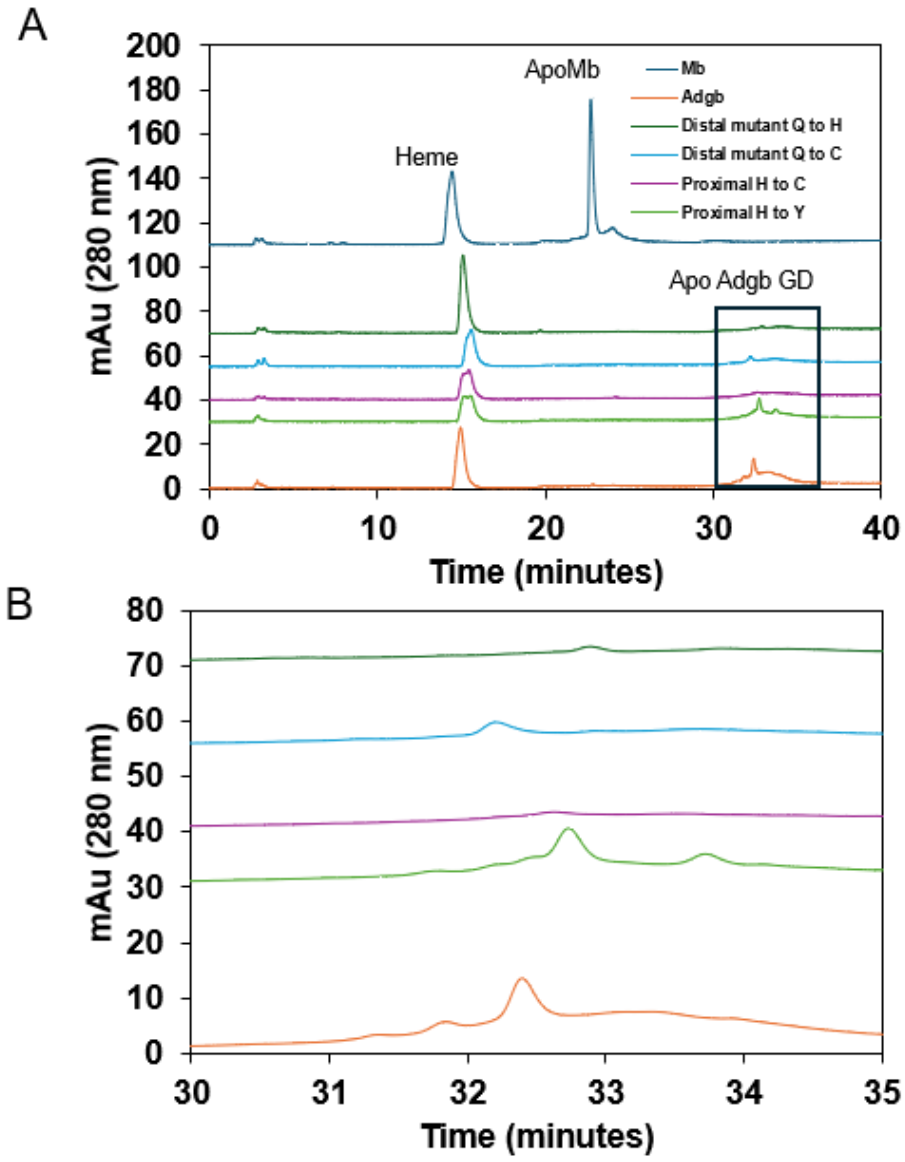


Figure 2.3.3.5 HPLC analysis (280 nm) of the heme content.

(A) The WT globin domain of androglobin and its mutants following reconstitution, with equine Mb used as a control for comparison. Equine Mb (50 μM) and unknown concentration of Adgb WT and mutants was analysed with reversed phase HPLC. The analysis was repeated with two independent samples. The separation of the Apo proteins at 280 nm in the two samples was consistently observed. (B) Expanded view of the box showing the chromatogram of the apo Adgb-GD protein. The traces were offset for clarity.

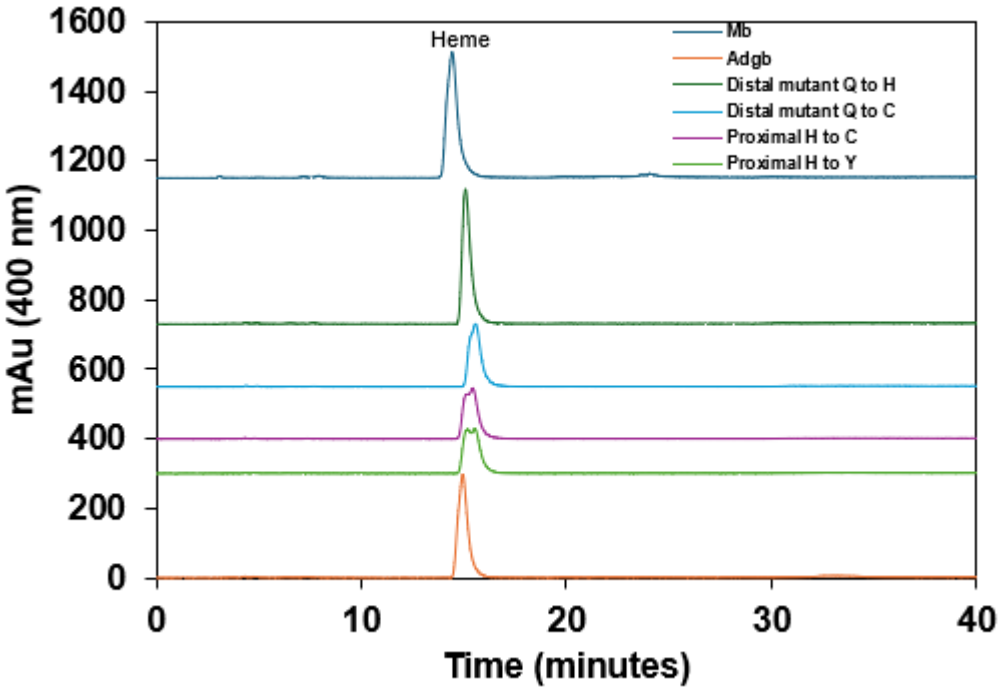


Figure 2.3.3.6. HPLC analysis (400 nm) of the heme content of the WT globin domain of androglobin and its mutants following reconstitution, with equine Mb used as a control for comparison.

Equine Mb (50 μM) and unknown concentration of Adgb WT and mutants was analysed with reversed phase HPLC. The analysis was repeated with two independent samples. The area under the heme peak at 400 nm for the Adgb GD mutants was determined and it was used to determine the extinction coefficients of the optical spectra. The traces above were offset for clarity.

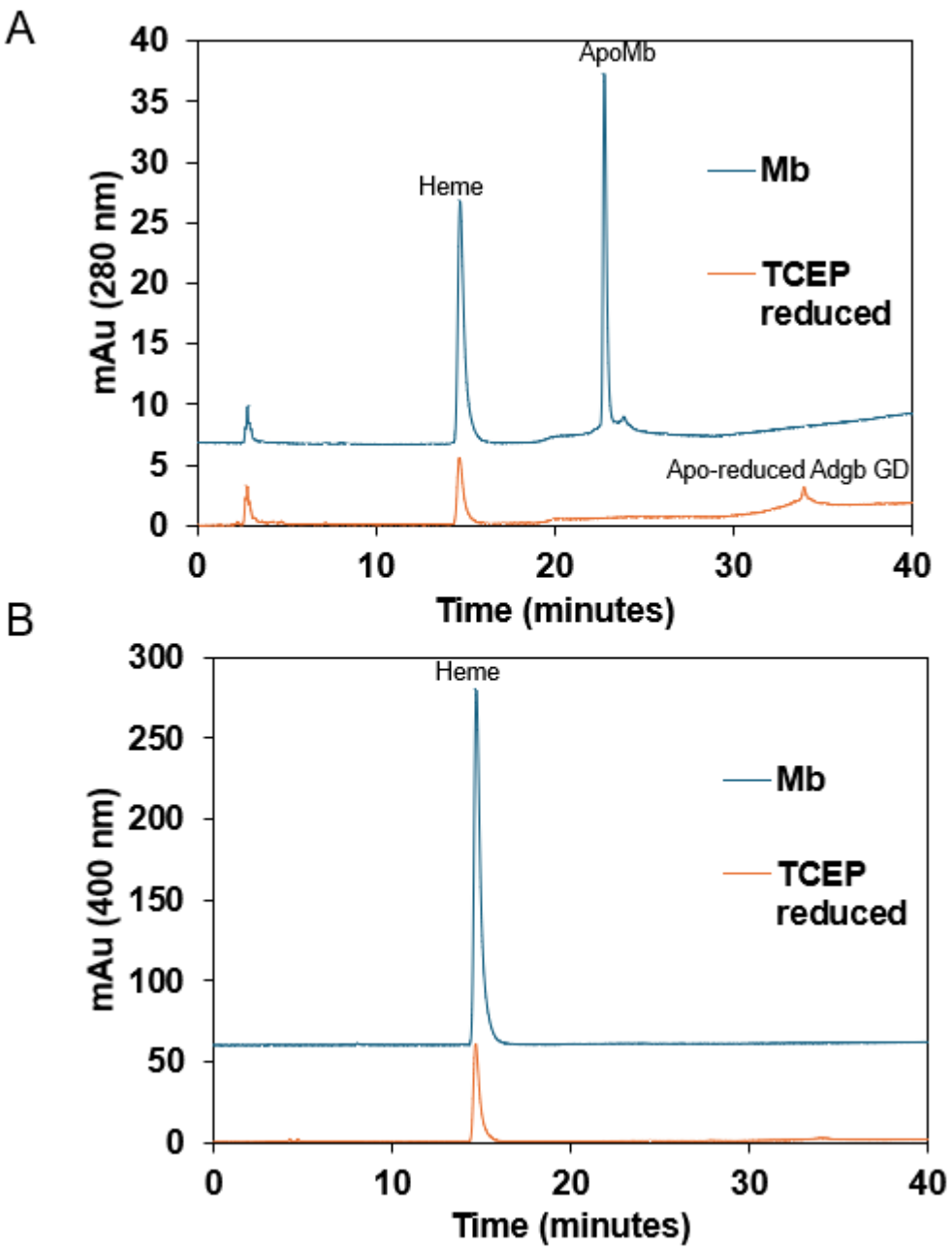


Figure 2.3.3.7. HPLC analysis of the heme content of the disulfide reduced Adgb (with TCEP|) following reconstitution, with equine Mb used as a control for comparison. Equine Mb (50 µM) and unknown concentration of the disulfide reduced Adgb was analysed with reversed phase HPLC.

(A) 280 nm. (B) 400 nm. The extinction coefficient of the disulfide reduced Adgb was calculated using the area underneath the peak at 400 nm and compared to the optical spectra of the holo protein. The traces above were offset for clarity.

2.3.4 Determining the oligomeric state of Adgb-GD

As mentioned earlier, no laboratory has successfully expressed a stable form of Adgb-GD protein prior to these studies, leaving its biophysical properties largely unknown until now. The optical properties described in Section 2.3.3 suggest that this protein is hexacoordinate globin with a heme geometry consisting of a proximal and distal amino acid ligand. Hexacoordinate proteins like Cygb and Ngb are predominantly monomeric under physiological conditions but can form dimers or higher-order oligomers under specific circumstances e.g. under non-reducing conditions, putative Cygb dimers were detected, likely resulting from intramolecular disulfide bond formation (Lechauve *et al*., 2010). Adgb-GD is a hexacoordinate protein with an intramolecular disulfide bond, which may enhance its propensity to form dimers and higher-order oligomeric complexes.

Figure 2.3.4.1 shows the size exclusion chromatogram (SEC) of the control samples at different wavelengths following analysis by UHPLC (Agilent technology 1290 infinity) in 0.1 M phosphate buffer. The chromatogram (220 nm) showed that bovine serum albumin dimer (BSA-D) was first to be eluted at 5.89 mins, followed by bovine serum albumin monomer (BSA-M) (6.59 min), myoglobin dimer (Mb-D) (7.33 min), myoglobin monomer (Mb-M) (8.27 min), cyt c (8.59 min) and finally ferric cyanide (12.81 min). The retention times, along with the MW and log MW of the protein control, are provided in Table 2.3.4.1. The plot of absorbance (220 nm) as a function of elution time in Figure 2.3.4.2 shows the chromatogram analysis of Adgb-GD. The fastest elution was at 5.86 min which was also the major peak on the chromatogram. The peak at retention time (5.86 min) when compared to the minor shoulder peaks (highlighted within the box, Figure 2.3.4.2B) at retention times 6.49, 6.95 and 7.42 min, was found to be significantly higher, indicating that it is the predominant component of the sample.

The log Mw of control samples vs retention time graph was plotted in Figure 2.3.4.3. This resulted in a straight line with the equation : y = -0.38x + 7.38. This was used to calculate the log molecular weight and oligomeric state for each of the retention time of Adgb-GD using the following formulas:

$$x = Elution\ time\ (Adgb - GD) * gradient$$

Equation 2.3.4.1.

Where x = control. Therefore y = mx + c *using the equation of the straight line (*y = *-0.38x + 7.38) gives the* y *value of log*M_W.

$$M_W = 10\ \hat{}\ y$$

Equation 2.3.4.2.

Where MW = molecular weight at retention time, y = y value or log M_W.

$$\mathbf{J} = \frac{M_W^{rt}}{M_W^M}$$

Equation 2.3.4.3.

Where J = oligomeric state, M_W^{rt} = molecular weight at retention time, M_W^M = calculated molecular weight of monomeric protein i.e. Adgb-GD (23 kDa, from section 2.3.1).

The oligomeric state for each of the retention time of Adgb-GD is provided in Table 2.3.4.2. The y value for the major component of Adgb-GD which is the log MW derived from ***Equation 2.3.4.2*** fits the trend line as shown in Figure 2.3.4.3 enabling calculation of the approximate MW of Adgb-GD in the UHPLC-SEC sample. This gave a MW of 135.023 kDa

and thus enabled the use of ***Equation 2.3.4.3*** to calculate the oligomeric state of the major component which was a hexamer.

As SDS PAGE and Western blot rule out covalent oligomerisation, the nature of the oligomerisation was further explored. To investigate the role of ionic interactions in Adgb-GD oligomerisation, the protein was re-analysed under high-salt buffer conditions, which are known to disrupt salt bridges between protein subunits. This approach was intended to help determine whether the formation of the hexamer was primarily driven by hydrophobic interactions or ionic (salt bridge) interactions. High salt concentrations would be expected to interfere with ionic bonds but would likely have minimal impact on hydrophobic patches located on the protein surface. Figure 2.3.4.4 presents the results of this analysis. The primary retention time of Adgb-GD in high-salt buffer (Figure 2.3.4.4A) was observed at 5.86 minutes, matching the retention time recorded under native conditions. This peak also represents the most abundant species in the sample, consistent with a hexameric form. In addition to the primary peak, two smaller peaks were detected at 7.43 minutes and 9.94 minutes (Figure 2.3.4.4B), corresponding to dimeric and monomeric forms of the protein, respectively. To estimate the molecular weights of these species and confirm their oligomeric states, log molecular weight (Mw) of control proteins was plotted against retention time (Figure 2.3.4.4C). The retention times of the three peaks observed in the high-salt buffer analysis were mapped onto it trendline which enabled accurate molecular weight estimations which is provided in Table 2.3.4.1.

It can then be surmised from these results that the isolated globin domain of the protein predominantly exists as a hexamer, primarily through hydrophobic interactions. There are other smaller peaks corresponding to other oligomeric states of the protein as shown in Table 2.3.4.2. The result indicates that the monomeric form of the protein exists in a very small amount.

Overall, the current data following the size exclusion analysis showed that in 0.1 M phosphate buffer under low and high-salt buffer condition, the protein is primarily a multimer.

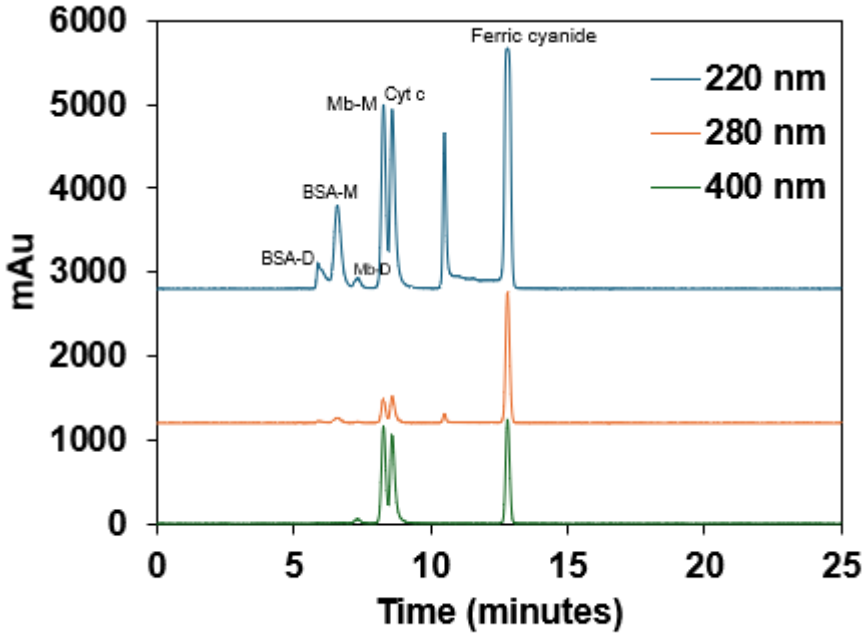


Figure 2.3.4.1. The chromatogram of the protein mixture used for the control in the UHPLC size exclusion chromatography.

The size exclusion was done in non-denaturing (native) conditions using 0.1 M NaPi at pH 7. The concentration of the samples used was 2 mg/ml for bovine serum albumin (BSA), myoglobin (Mb), cytochrome c (Cyt c) and ferric cyanide. Traces were offset for clarity.

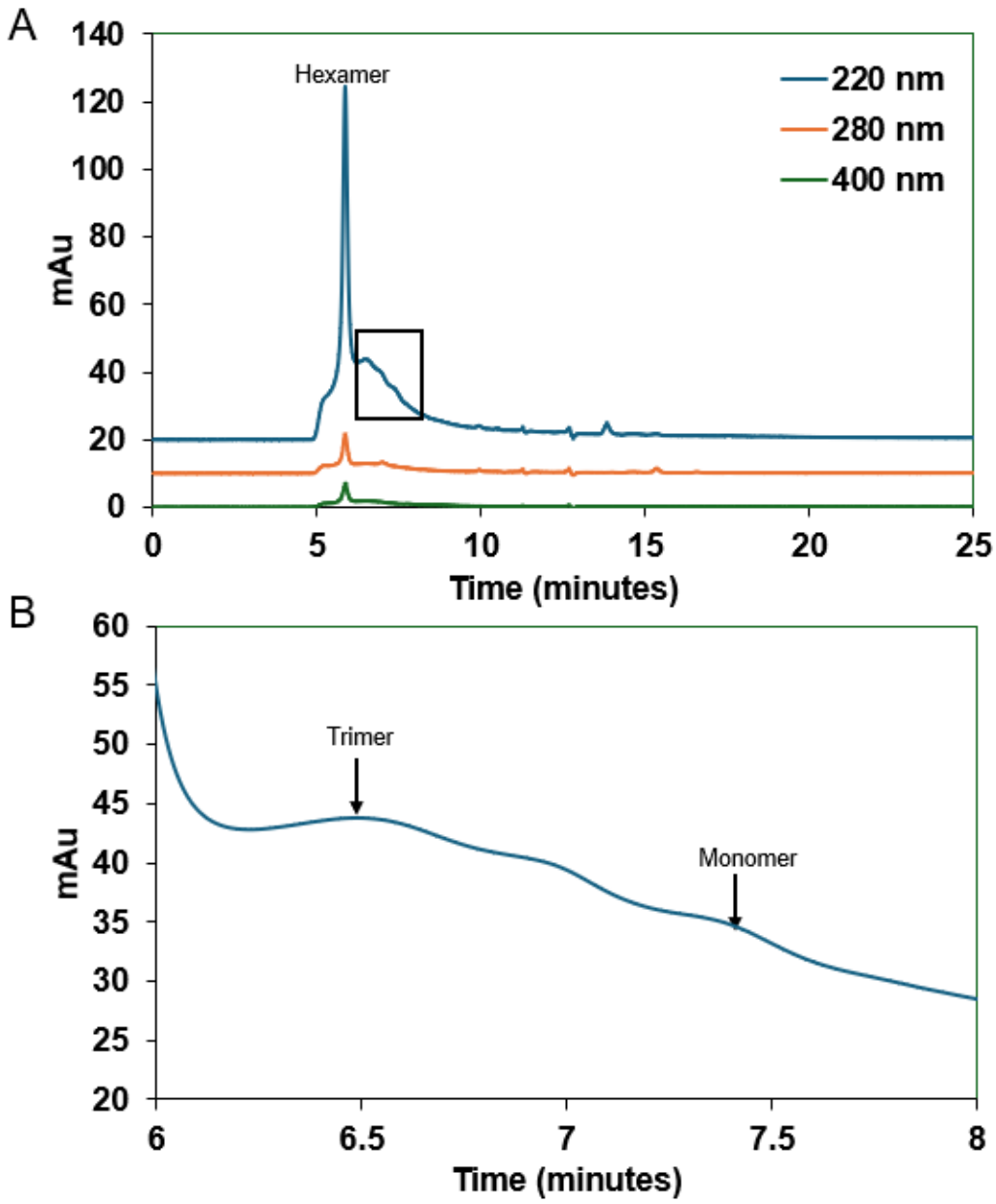


Figure 2.3.4.2. The chromatogram of Adgb at 220 nm, 280 nm and 400 nm taken after UHPLC chromatography.

(A) The main retention time was shown at 5.856 mins which corresponds to a hexamer being eluted. Peak at 5.856 mins is the highest among all the peaks therefore it is the most abundant and the major component in the sample. (B) The subsequent minor peaks appear as a shoulder of the main peak (box), having longer retention times and thus are smaller oligomers. The traces were offset for clarity.

Table 2.3.4.1. The control proteins and compound analysed using HPLC on a chromatography column, with their molecular weights, the logarithm of those weights, and their corresponding peaks recorded.

Proteins	Molecular weight (Da)	Log Mw	Elution time (mins)
BSA Dimer (BSA-D)	132860.6	5.123	5.885
BSA Monomer (BSA-M)	66430.3	4.822	6.592
Mb Dimer (Mb-D)	35134	4.546	7.325
Mb Monomer (Mb-M)	17567	4.245	8.272
Cytochrome c	12384	4.093	8.592
Ferricyanide	281.24	2.449	12.805

Table 2.3.4.2. Shows the time of Adgb-GD elution where each time point corresponds to the peak shown in figure 2.3.4.2 and 2.3.4.4. The data provide information about the oligomeric state of the protein calculated from the formular in ***Equation 2.3.4.3***. The stable state of the protein is a hexamer.

Adgb elution time (mins) (low or/and high salt buffer conditions)	Calculated average oligomeric state
5.856 (low and high salt)	5.87
6.488 (low salt)	3.36
6.949 (low salt)	2.23
7.416 (low salt)	1.48
7.431 (high salt)	1.46
9.938 (high salt)	0.16

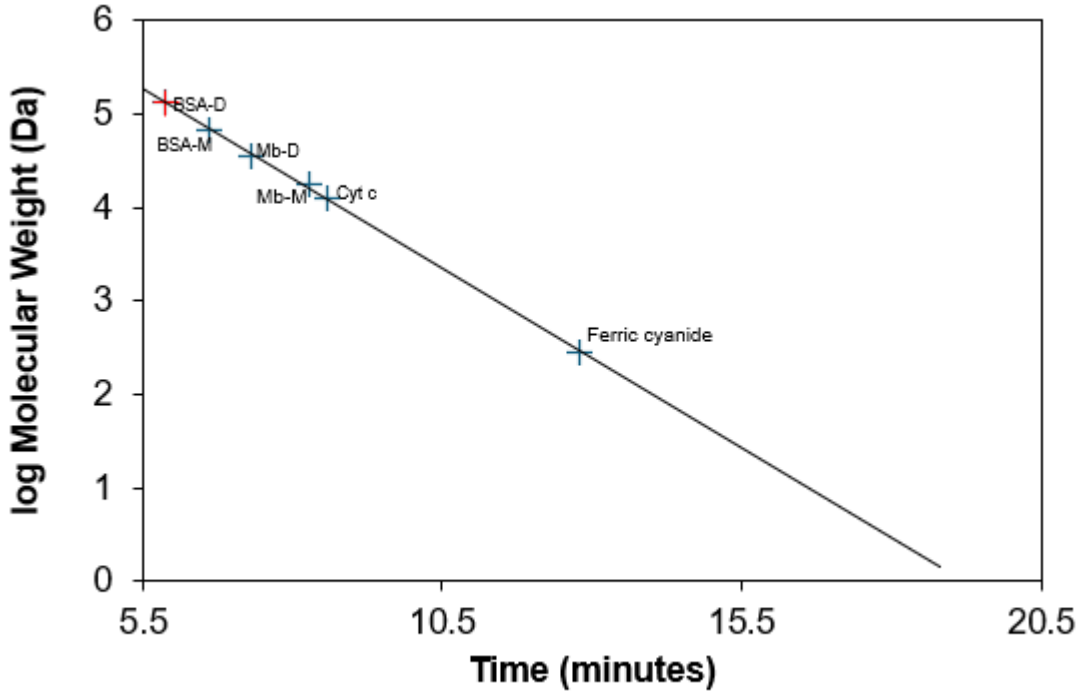


Figure 2.3.4.3. The log Mw of control samples vs retention time graph.

The equation of the straight line: y = -0.38x + 7.38 for the control samples was utilised to calculate the molecular weight and oligomeric state of the most stable state (hexamer) of androglobin-GD using the ***Equations 2.3.4.1 to 2.3.4.3.*** The y value which is the log molecular weight derived from ***Equation 2.3.4.2*** was plotted onto the graph vs its retention time 5.856. The hexamer state log MW of Adgb-GD (red cross) is expected to align with the trendline of log MWs for ferric cyanide, cyt c, Mb-monomer (Mb-M), Mb-dimer (Mb-D), BSA-monomer (BSA-M), and BSA-dimer (BSA-D), and it does. The result indicate that the stable form of Adgb-GD is a hexamer.

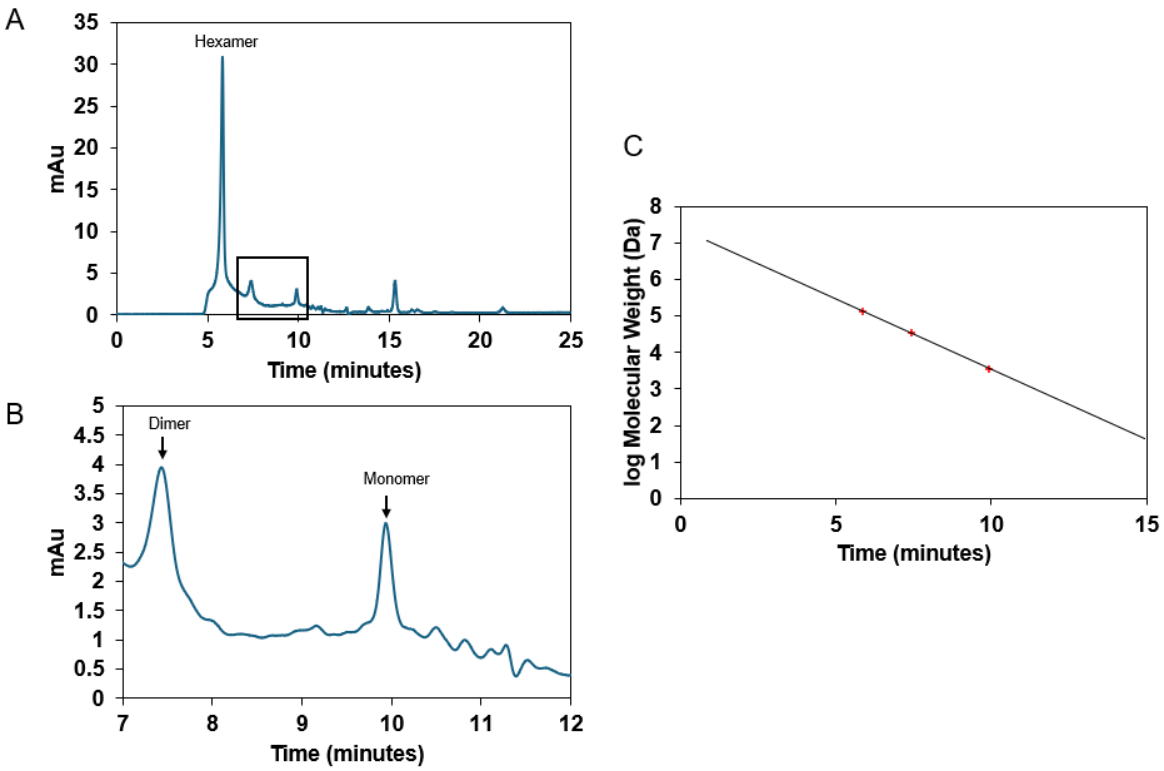


Figure 2.3.4.4. Analysis of Adgb-GD in high salt buffer.

(A) The chromatogram of Adgb at 220 nm taken after UHPLC chromatography. The main retention time was shown at 5.858 mins which corresponds to a hexamer being eluted. (B) The subsequent minor peaks (box), having longer retention times and thus are smaller oligomers. (C) The log Mw of control samples vs retention time graph in high salt buffer. The equation of the straight line was also: y = -0.38x + 7.38 for the control samples and it was utilised to calculate the molecular weight and oligomeric state of the of the three peaks using the ***Equations 2.3.4.1 to 2.3.4.3.*** The three y value which is the log molecular weight derived from ***Equation 2.3.4.2*** was plotted onto the graph vs their retention time. They align with the trendline of log MWs for ferric cyanide, cyt c, Mb-M, Mb-D, BSA-M and BSA-dimer BSA-D. The result indicate that the stable form of Adgb-GD is a hexamer.

2.3.5 Characterising the expression of Adgb-GD in mammalian cells

The expression of Adgb-GD gene in eukaryotic cells has encountered similar challenges to those observed in *E. coli*. To date, efforts to express the globin domain have been largely unsuccessful, either resulting in expression failure, as observed in *P. pastoris*, or yielding very low expression levels, as reported in baculovirus-infected *Sf9* insect cells (Bracke *et al*., 2018). In this experimental study, Adgb-GD was expressed using the higher eukaryotic expression system HEK293T, which may represent a viable alternative for globin domain expression.

The expression of the protein was observed a total of 48 h (24 h with chloroquine and 24 h without) after the transfection of Adgb-GD GFP-tagged plasmid DNA. Due to the C-terminal GFP tag on this protein, the protein produced a green fluorescence in the cells which typically indicate successful transfection and expression for the GFP tagged protein, in this case Adgb-GD domain. The images of these green fluorescent cells are presented in Figure 2.3.5.1.

The image in Figure 2.3.5.1 is typical to that seen in other heme proteins that are GFP tagged as shown in Figures 2.3.5.2 and 2.3.5.3 for Ngb and Mb respectively. Figure 2.3.5.2 and Figure 2.3.5.3 depicts captured image of Ngb and Mb expression in HEK293T cells after 48 h using the transfection method given by Koay et., 2021. The captured images showed similar expression levels for all three cells with high expression in some cells and low or no expression in others depending on the transfection efficiency. This heterogeneity is typical of transiently transfected cells. Normal GFP was also expressed using this method as a control for the experiment. The fluorescent level in the control was the same as that observed in the heme protein as displayed in Figure 2.3.5.4. This indicates that the expressed protein using Adgb-GD GFP-tagged, Mb GFP-tagged and Ngb GFP-tagged plasmid DNA are GFP tagged protein. This

is also supported by the absence of GFP signal during the expression of the empty PCDNA 3.1 plasmid. A slightly reduced expression level was observed for Mb-GFP, likely attributable to decreased transfection efficiency of the plasmid.

Unfortunately, the lack of His-tag on these proteins made it difficult to perform Western blot analysis with the available 6X His-Tag monoclonal antibody (His.H8) and HRP (ThermoFisher) in the laboratory. An alternative approach using a GFP antibody was attempted but did not work immediately, requiring further exploration time to optimise transfection. Nonetheless, expression of Adgb-GD appears feasible though this method.

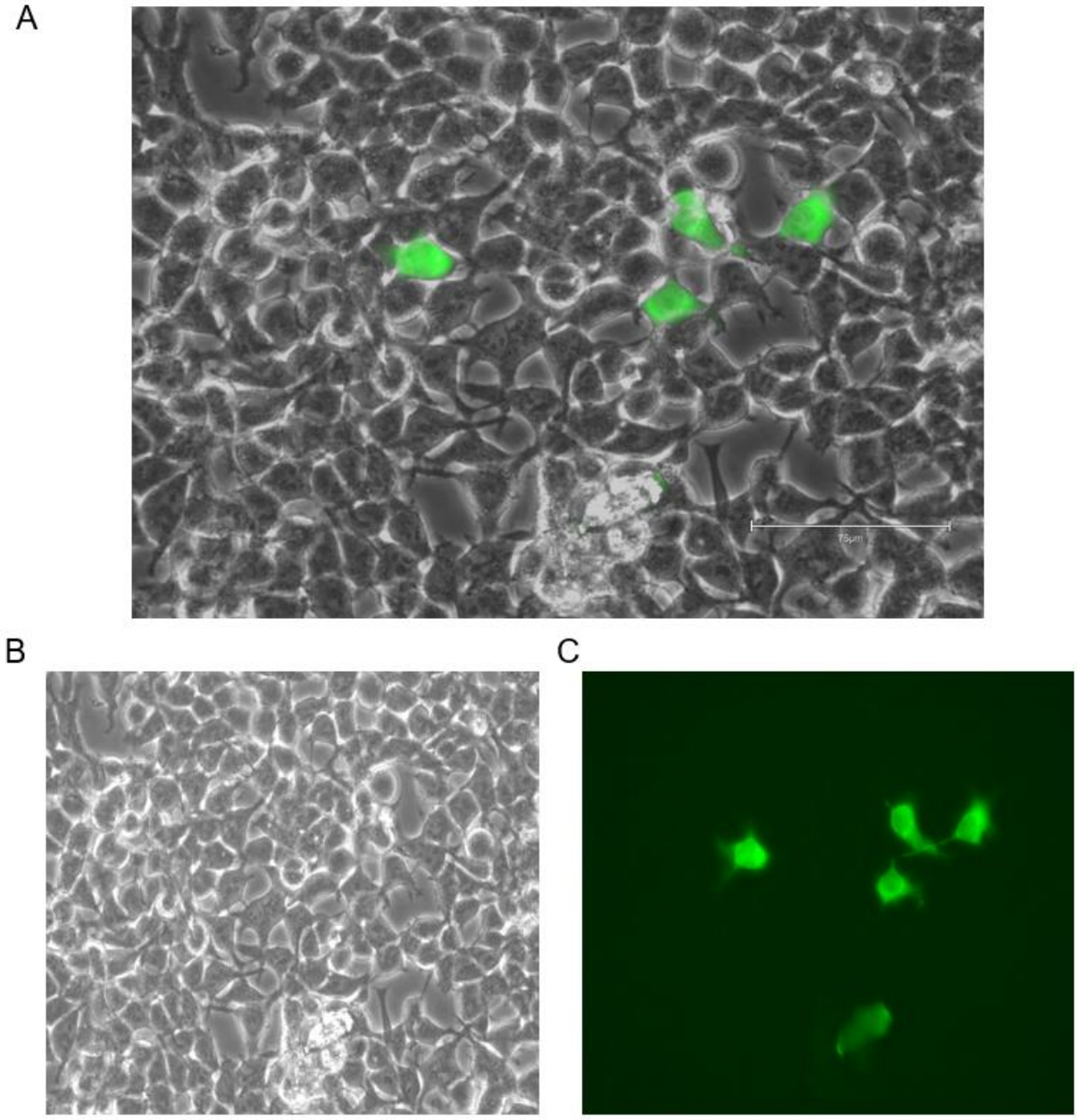


Figure 2.3.5.1. The expression of GFP-tagged Adgb GD in HEK293T cells captured on the EVOS M5000 invitrogen using the brightfield and GFP filters system 48 h after transfection using the method described in section 2.2.11.

(A) Composite of brightfield and green fluorescence (ex 475 nm and emission 524 nm). The protein will produce a green fluorescence in the cells as shown in the image due to the GFP tag. (B) Brightfield view. (C) green fluorescence view (ex 475 nm and emission 524 nm).

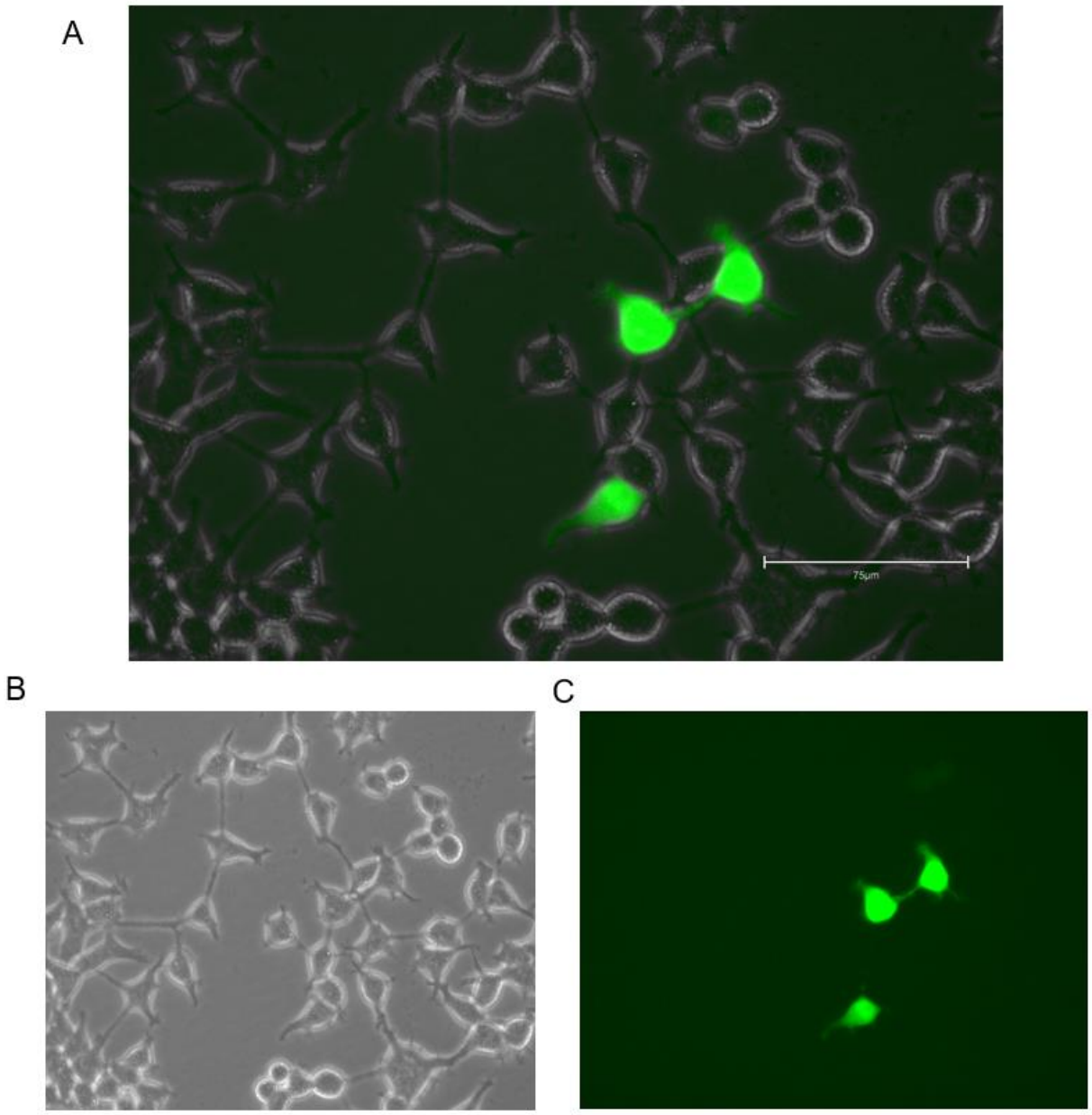


Figure 2.3.5.2. The expression of Ngb GFP-tagged in HEK293T cells captured on the EVOS M5000 invitrogen by ThermoFisher Scientific system using the brightfield and GFP filters system.

The conditions are identical to those described in Figure 2.3.5.1. (A) Composite of brightfield and green fluorescence (ex 475 nm and emission 524 nm). (B) Brightfield view. (C) green fluorescence view (ex 475 nm and emission 524 nm).

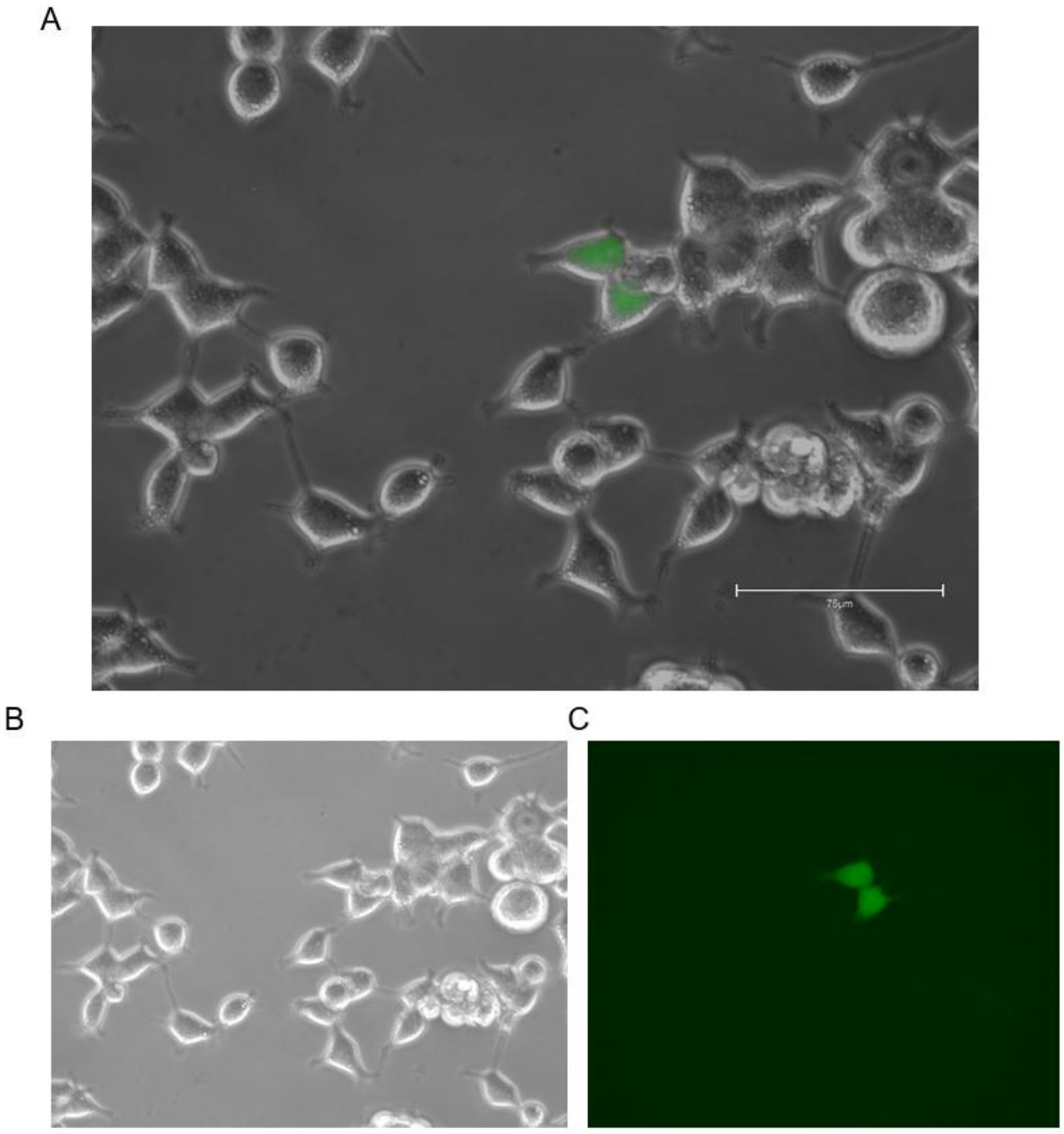


Figure 2.3.5.3. The expression of Mb GFP-tagged in HEK293T cells captured on the EVOS M5000 invitrogen by thermoFisher Scientific system using the brightfield and GFP filters system.

The conditions are identical to those described in Figure 2.3.5.1. (A) Composite of brightfield and green fluorescence (ex 475 nm and emission 524 nm). (B) Brightfield view. (C) green fluorescence view (ex 475 nm and emission 524 nm).

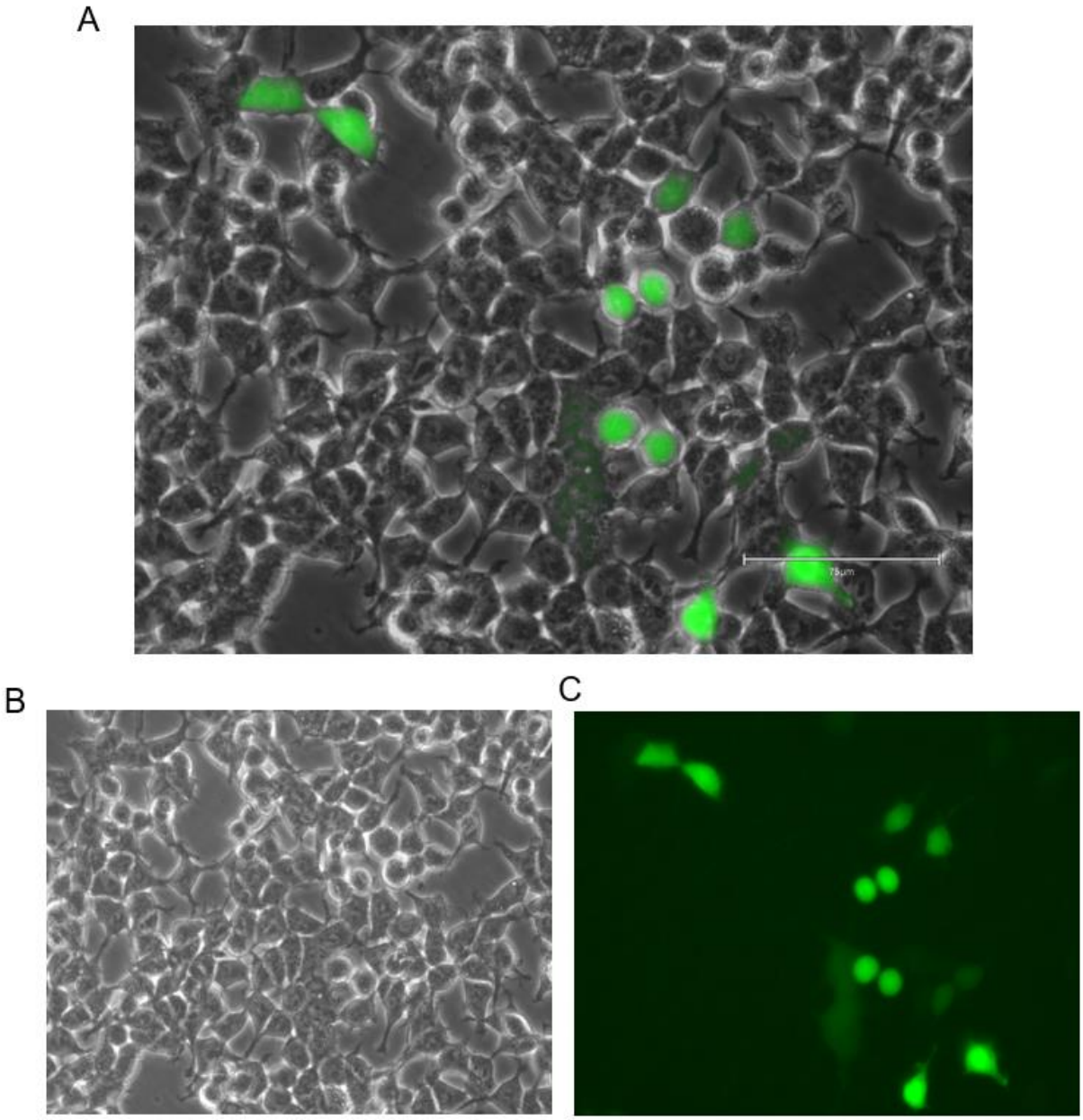


Figure 2.3.5.4. The expression of GFP in HEK293T cells captured on the EVOS M5000 invitrogen by thermoFisher Scientific system using the brightfield and GFP filters system.

The conditions are identical to those described in Figure 2.3.5.1. (A) Composite of brightfield and green fluorescence (ex 475 nm and emission 524 nm). (B) Brightfield view. (C) green fluorescence view (ex 475 nm and emission 524 nm).

2.3.6 Investigating the expression level of the heme proteins using the OLIS

Following the successful expression of the GFP-tagged protein, the next step was to assess whether the oxidation and ligation states could be determined *in situ* using the Olis light integrating spectrophotometer, potentially informing on likely oxidation and ligation states *in vivo*. The spectra obtained represent measurements of the whole cell. Known concentrations of ferric Mb, initially measured with Agilent Cary 50 UV-Vis Spectrophotometer (Figure 2.3.6.1A), were examined on the OLIS RSM 1000. This was to calibrate the Olis spectrophotometry due to the variable pathlength of the Olis instrument light integrating sphere. An optical spectrum consistent with the characteristic Mb spectra obtained using the Agilent Cary 50 UV-Vis Spectrophotometer was observed (Figure 2.3.6.1B). The ferrous spectrum of the same sample was also acquired using the Agilent Cary 50 UV-Vis Spectrophotometer (Figure 2.3.3.1A) and subsequently utilised to calculate the concentration of Mb retrospectively, following the methodology described by Javorfi *et al*., 2005 quantitative analysis method (Figure 2.3.6.1C).

The absorbance vs wavelength graph of the Adgb-GD protein expressed in cells, as demonstrated by the OLIS RSM 1000 spectra in Figure 2.3.6.2, exhibits similarities to the WT, as depicted in these same figures. Baseline shifts made analysis more difficult, particularly in the Q-band area in the visible region at 500 – 600 nm. However, it is clear that the spectra from the harvested Adgb-GFP cells, closely resembled the optical spectrum of the ferric Adgb-GD, with a Soret peak at 413 nm and two peaks in the Q band region. Upon reduction of the protein in the harvested cells with sodium dithionite, the iron appeared to be destabilised, as shown in Figure 2.3.6.2B. Although there was a red shift to the right, it differed from the ferrous WT (425 nm) (Figure 2.3.3.1C), with the optical shift occurring at a lower absorbance and at 418 nm. No changes in the Q-band at the visible region was observed. The CO seemed to stabilise

the iron in the protein, and this is observed by the spectral shift at 421 nm (Figure 2.3.6.2C). This shift at 421 nm is typical of that observed in Adgb WT CO-bound spectrum. The corrected absorbance data from the OLIS RSM 1000 CLARiTY, plotted against wavelength for ferric Adgb-GD, indicated a concentration of 0.354 μM in approximately 300 μL of the sample. Overall, Adgb-GD expressed *in situ* within cells appears primarily in the ferric hexacoordinate state.

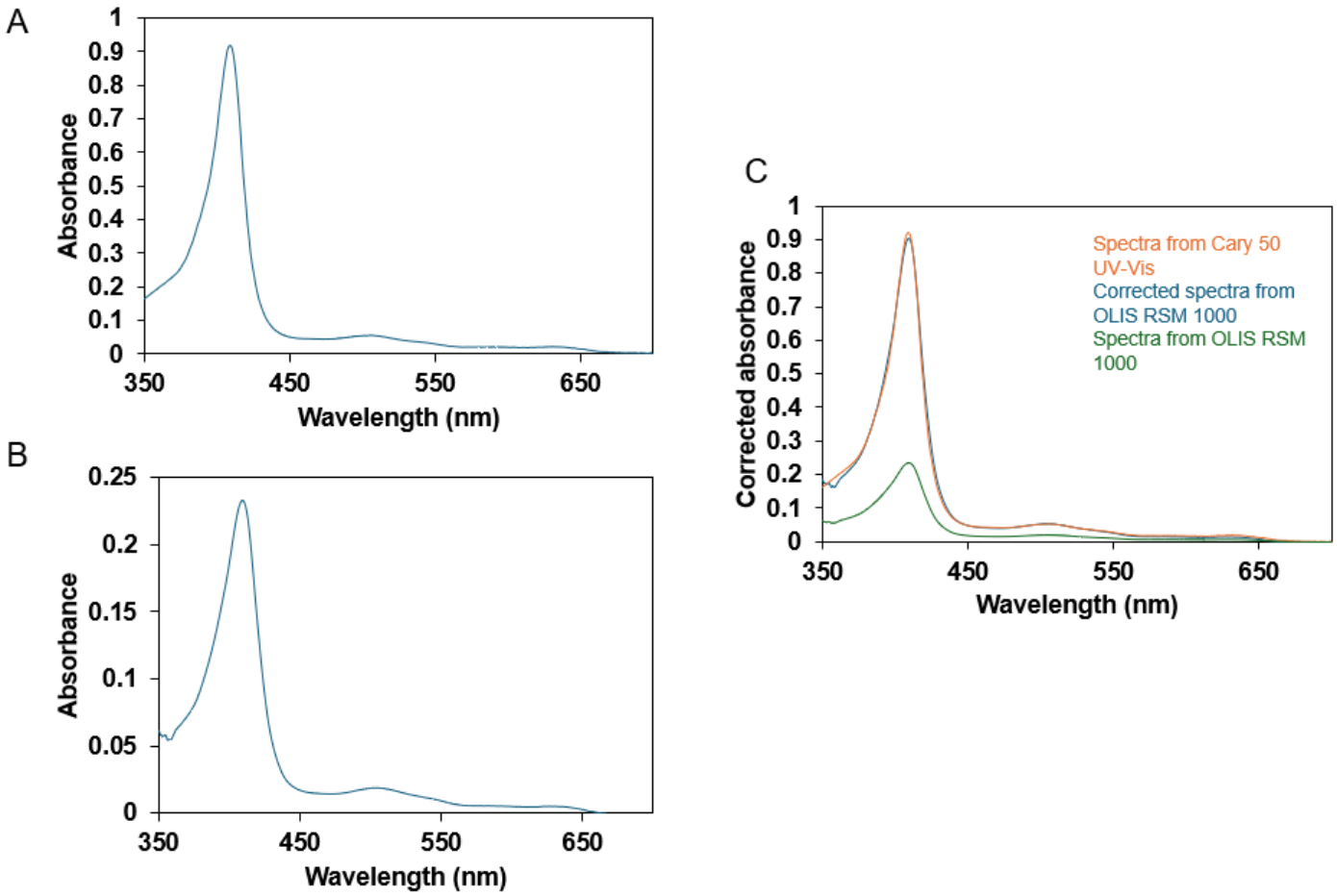


Figure 2.3.6.1. Optical spectrum of ferric Equine Mb (5.3 µM).

From (A) Agilent Cary 50 UV-Vis Spectrophotometer and (B) OLIS RSM 1000 Spectrophotometer. (C) Mb (5.3 µM) Cary 50 spectrum of ferric Mb (orange) and from Olis (green) with corrected Olis spectrum using the Javorfi method, described earlier in section 2.2.12, shown in blue.

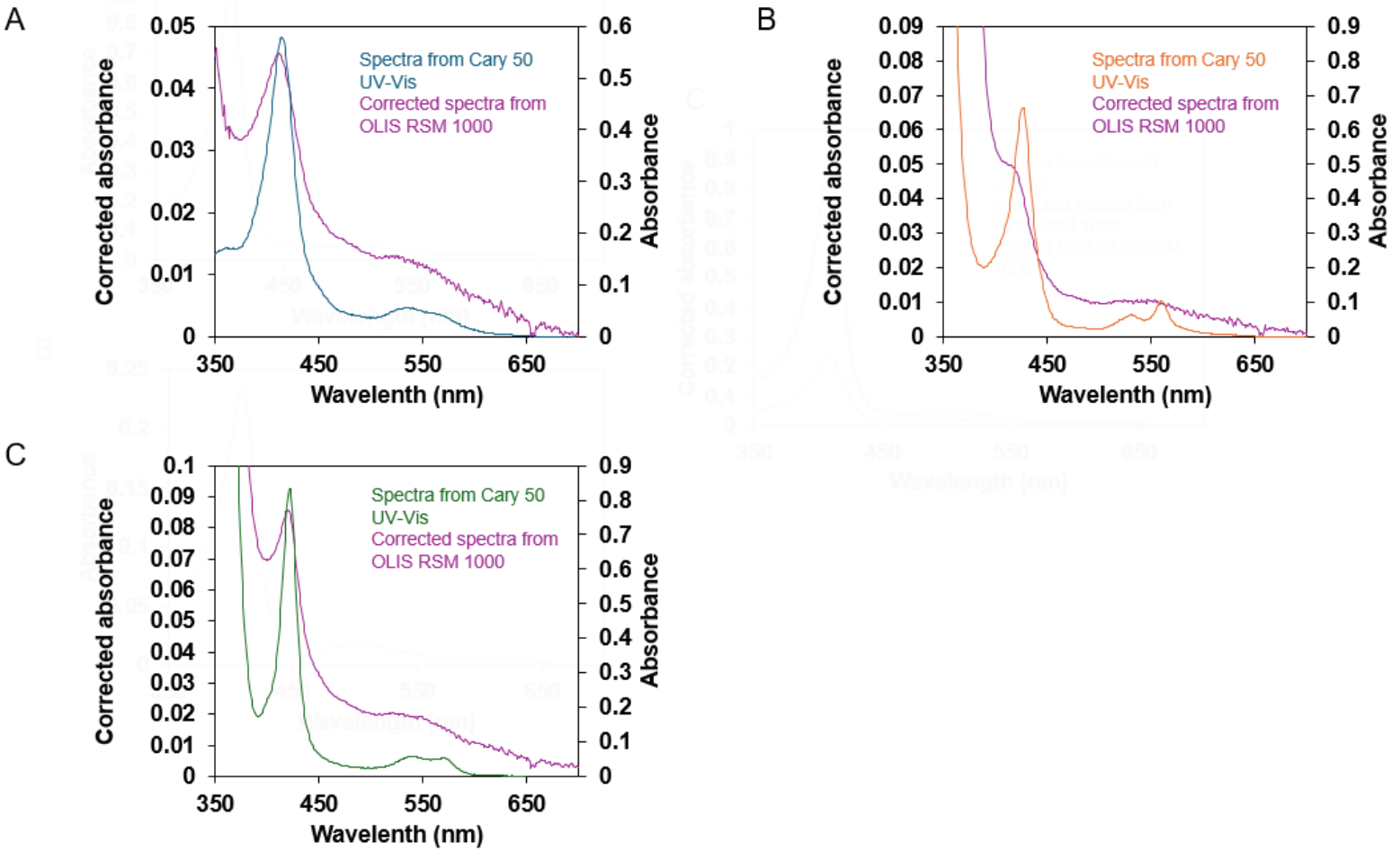


Figure 2.3.6.2. Optical spectrum of WT Adgb-GD (5 µM).

Showing corrected plot of absorbance vs wavelength of (A) ferric Adgb, (B) ferrous WT Adgb-GD and (C) CO-bound Adgb, following analysis using the same condition as figure 2.3.6.1.

2.4 Discussion

2.4.1 Expressing Androglobin Globin Domain in *Escherichia coli* and Human Embryonic Kidney cells

E .coli is the most popular host for recombinant protein expression due to propensity of being easy, quick and cheap. *E. coli* also have a fast growth rate with a generation time spanning 20-40 minutes under optimised condition and is in suspension, unlike many eukaryotic expression systems, making them much higher density overall. *E. coli* can be readily manipulated (e.g. transfection) due to the comprehensive understanding of their biology and the availability of numerous well-developed tools for molecular manipulation (Bhatwa *et al*., 2021).

E. coli is the perfect host for large scale recombinant expression for small globular proteins such as globins because *E. coli* usually doesn't have any specific requirement in terms of folding or post-translational processing (Bracke *et al*., 2018). However, protein expression in *E. coli* cells is frequently limited by the aggregation of proteins into inclusion bodies, particularly if the protein being expressed is not extremely stable. Inclusion bodies are nuclear or cytoplasmic aggregates primarily composed of proteins, often forming as by-products of cellular metabolism (Bracke *et al*., 2018). Protein expression in *E. coli* is also often constrained by the size of the DNA plasmid. Consequently, Adgb was subdivided into its predicted individual domains prior to expression to improve the likelihood of successful protein production. This strategy was adopted due to the inability to achieve successful expression of the full length Adgb protein (~190 kDa) in *E. coli* (Bracke *et al*., 2018). The full length Adgb protein, however, was successfully expressed in the eukaryotic *Sf9* cell system; but it was reported to undergo truncation over time. This truncation was attributed to proteolysis by intracellular proteases or autocleavage by the N-terminal calpain-like domain (Bracke *et al*.,

2018). The autoproteolysis in the N-terminal domain is examined in detail in Chapter 3 of this thesis.

In previous studies, the globin domain was expressed as an insoluble aggregate (inclusion bodies) in *E. coli*, making purification extremely difficult even under modified conditions (Bracke *et al*., 2018). Strategies to support purification of unstable proteins such as the addition of reducing agents, detergents, or adjustments to the pH or salt concentrations of the resuspension buffer did not improve the protein's stability. Mutagenesis was also performed to replace the cysteine residues in the globin domain, which could potentially cause intermolecular disulfide bridge formation and subsequent precipitation, with serine. However, despite these modifications, the globin domain remained unstable following purification (Bracke *et al*., 2018).

Despite the realigned sequence of the globin domain, identifying and alternative sequence for the C helix for the protein, the purification of the globin domain from *E.coli* initially remained problematic (Reeder *et al*., 2024). Four strategies were tested for the purification of Adgb-GD in this study (section 2.2.3). The results from the first method were the most comparable to those of previous studies, as it resulted in insoluble aggregates following purification. This is evident from the absence of any band in the SDS-PAGE gel analysis (Figure 2.3.1.2A). However, the second, third, and fourth methods resulted in the expression of the Adgb-GD protein as a distinct band at ~23 kDa was observed (Figure 2.3.1.2A and Figure 2.3.1.3). The main difference between the first method and the others was the inclusion of surfactants and denaturing agents, specifically SDS in the second and third methods, and urea in the fourth method. SDS is an effective anionic surfactant that prevents or reverses aggregation in the protein of interest i.e. Adgb-GD. The use of SDS in this purification process however introduced significant logistical challenges, as it precipitates in the cell lysate at temperatures below ~30 °C, likely from interactions with endogenous *E.coli* proteins. This

is why the purification process (method 2 and 3) were performed in water baths maintained at 30 °C to prevent SDS precipitation. The greater purity being attributed to the additional washing step included in the methodology (section 2.2.3 and Figure 2.3.1.2). However, the purest sample was obtained when the protein was purified under denaturing conditions using 8 M urea. This may be partially due to the ease of introducing a linear gradient rather than a step purification using the AKTA pure chromatography system as SDS precipitation tended to block the ATKA system. However, it was more likely that the SDS in combination with the higher binding and elution temperatures, or the use or denaturing urea, resulted in the prevention of Adgb-GD binding to other endogenous *E.coli* proteins. This suggests that the Adgb-GD tends to bind other proteins. Whether this is a property that leads to a binding to a partner protein *in vivo* or more likely binding to another part of the Adgb N or C terminal domains, remains unknown.

The urea-based purification process yielded lower protein quantities during expression compared to the SDS based method for similar number of flasks as depicted in Figure 2.3.1.3. Consequently, the active purification step involving SDS was adopted as the primary method for expressing the Adgb-GD protein in *E. coli* throughout the study. Overall, the successful expression and purification of Adgb-GD in *E. coli* in this study was in part due to reassessment of the globin domain sequence alignment (Reeder *et al*., 2024) but also using SDS to stabilise the protein. Other common surfactant used for stabilising membrane proteins have been previously tested like triton x-100 and tween-20, however, SDS provided the best results (Reeder *et al*., 2024).

Yeast systems e.g. *Saccharomyces cerevisiae (S. cerevisiae)* and *Pichia pastoris* system (*P. pastoris)* are widely used model organism for protein expression in eukaryotic system due to its ease of cultivation and capability for post-translational modifications. *Crocodylus siamensis* Hb, a globin with strong antibacterial and anti-inflammation properties, is an

example of a globin that has been successfully expressed in the eukaryotic system *P. pastoris* (Anwised *et al*., 2016). Similarly to Adgb-GD expression in *E. coli*, the expression of Adgb-GD protein in eukaryotic systems had also proven to be unsuccessful. The expression either failed, as observed in *P. pastoris*, or it yielded very low expression levels, as reported in baculovirus-infected *Sf9* insect cells (Bracke *et al*., 2018). The protein was not detected in either the secreted fraction or the intracellular fraction following expression in *P. pastoris.* It was suggested that the protein was degraded due to improper protein folding (Bracke *et al*., 2018). A similar explanation was provided for the low-level expression observed in baculovirus-infected *Sf9* insect cells. Degradation of proteins due to improper folding appears to limit recombinant protein production in eukaryotic systems, as protein synthesis and secretion are inherently error-prone processes. Misfolded proteins are typically targeted for rapid degradation via the endoplasmic reticulum (ER)-associated protein degradation (ERAD) pathway; however, some misfolded proteins may be retained and accumulate in the ER, potentially inducing cellular stress (Li *et al*., 2022). It has been shown that recombinant protein expression in eukaryotic systems, such as yeast, can be enhanced. For instance, the production of human Hb was enhanced in the yeast system *S. cerevisiae* (Liu *et al*., 2014) by employing a novel approach. This strategy involved balancing the expression of alpha- and beta-globin chains and optimizing the heme biosynthetic pathway within *S. cerevisiae*. These modifications significantly improved the production of complex proteins like human Hb.

The expression of the Adgb-GD protein in HEK293T cells, a more advanced eukaryotic system compared to yeasts and baculovirus-infected *Sf9* insect cells, resulted in good yields of protein expression compared to other globins, as exhibited by Figures 2.3.5.1 vs 2.3.5.2 and 2.3.5.3. This expression, visualised in shown in Figure 2.3.5.1, highlights bright green fluorescence in surrounding cells, indicating successful uptake and expression of the Adgb-GD GFP-tagged gene. The GFP tag was engineered to be positioned at the C terminal end of the

globin sequence, meaning that the protein would not produce GFP fluorescence unless the entire sequence was successfully expressed by the system. Future work will involve western blot analysis to confirm the presence of the protein and provide additional missing information e.g. the Mw of the protein, thereby giving us a complete understanding of Adgb-GD expression in eukaryotic cells. Overall, the successful expression and isolation of protein in *E.coli*, together with the successful expression in human cell lines enabled the characterisation of the globin domain.

2.4.2 Biophysical characterisation of Androglobin Globin domain

Purification of Adgb-GD protein under SDS or urea conditions appeared to result in the loss of heme, although the exact cause of this phenomenon remains unknown. A similar observation was reported in a previous Adgb-GD expression study by Bracke *et al*. (2018). Due to the inability to express and purify a stable form of the protein, inclusion bodies were solubilised in 8 M urea, leading to the reported loss of heme during purification. While we successfully purified the stable protein, the majority of heme was lost during the process. To address this, heme was incorporated into the protein. Similar method of incorporating heme into globin proteins has been previously employed for early expression trials of Cygb (de Sanctis *et al*., 2003). In that case, overexpression of Cygb resulted in inclusion bodies, and excess hemin was added to the protein solubilised in 6 M guanidinium chloride, 50 mM Tris-HCl (pH 7.5), and 1 % mercaptoethanol to restore functionality (de Sanctis *et al*., 2003).

The optical spectra of reconstituted ferric, ferrous, and deoxyferrous CO-bound Adgb-GD showed typical features of a hexacoordinated globin. When compared to Ngb (Figure 2.3.3.1B), the Soret and Q bands for each oxidation state for Adgb -GD were very similar (see Section 2.3.3). Furthermore, the ferric extinction coefficient of both Adgb-GD (127.6 mM^{-1}

cm^{-1}) and Ngb (129 mM^{-1} cm^{-1}; Lardinois *et al*., 2008) are similar further supporting Adgb-GD hexacoordinate nature. However, the presence of the 630 nm component in the ferric state indicates the protein may develop some penta-like characteristics. This component seems to disappear in the mutants. In conclusion, the predominantly hexacoordinate nature of Adgb-GD aligns with previous predictions (Hoogewijs, 2012). The findings from this study also support the classification of Adgb as a hexacoordinate protein with pentacoordinate-like characteristics, similar to that observed for Cygb (Reeder *et al*., 2011) and some phytoglobins (Villar *et al*., 2021).

The OLIS RSM 1000 CLARiTY spectrophotometer enables rapid, non-derivatizing analysis of globin production and oxidation states in turbid whole-cell suspensions (Marcero *et al*., 2016). This part of the experiment aimed to investigate the native oxidation and ligation states of the overexpressed Adgb-GD *in situ,* providing insight into its likely states *in vivo*. This is the first study of its kind. Results indicate that Adgb-GD is primarily expressed as a hexacoordinate ferric form, suggesting this may be the main form of the protein *in vivo* although one must be cautious when comparing native levels of protein expression with that of highly overexpressed protein. Although the ferric Q-band in the visible region resembled that of ferric Ngb, high signal noise and a sloping baseline somewhat limits definitive conclusions. The ferrous spectra revealed heme destabilization following sodium dithionite treatment, likely due to partial reduction during expression and over-reduction by excess dithionite. CO addition appeared to stabilise the heme (Figure 2.3.6.2B vs 2.3.6.2C). The low stability and high autoxidation rate of the ferrous form further support the ferric state as the predominant form *in vivo* (Reeder *et al*., 2024).

2.4.3 Oligomeric state of the Adgb-GD

Most proteins fulfil their physiological roles either through interactions with other molecules or as stable molecular complexes formed with other proteins, which may exist as homo- or hetero-oligomers (Sakashita *et al*., 2018). Oligomeric proteins typically exist in at least two distinct states: monomeric and oligomeric forms (Sakashita *et al*., 2018). Consequently, a suitable technique is necessary to separate and quantify these states within a protein sample to determine the relative abundance of each.

Size-exclusion chromatography (SEC) is a highly effective technique for protein analysis, enabling accurate size profiling of protein samples due to its mild separation conditions, which allow for high resolution separations (Challener, 2014). SEC analysis of Adgb-GD under native and high-salt buffer conditions revealed Adgb-GD primarily forms hexameric multimers, with only trace amounts of monomer detected after refolding. Interestingly, no higher-order multimers were detected in the SDS-PAGE, except for dimers that occasionally appeared in low amounts. This is likely due to the SDS-PAGE being conducted under denaturing and reducing conditions, which are similar to those the protein experienced during purification.

Adgb-GD can combine not only with itself, but also with other proteins, carrying them through the purification steps until denaturation before electrophoresis. This suggests that the removal of SDS promotes the formation of large multimeric complexes due to areas of the protein's surface amino acids being hydrophobic or/and charged. This is confirmed by the electrostatic potential mapping (Figure 2.4.3.1) and hydrophobic patch analysis (Figure 2.4.3.2) of the Adgb-GD surface. Results from this analysis revealed that Adgb exhibits a significant surface electrostatic potential and a higher localised density of hydrophobic regions compared to typical monomeric globins, such as Mb, with large localised positively charged

and negatively charged regions of the protein. The data from the high salt SEC confirms that such salt bridges are not sufficient alone to explain the aggregation, suggesting that hydrophilic interactions also play a significant role. These features suggest a greater propensity for aggregation, likely through charge-based or hydrophobic interactions with itself or other proteins, implying that the native form of Adgb-GD would exist bound to other proteins *in vivo*. In its functional state as part of a larger Adgb protein, it is likely that Adgb-GD has interactions with these other Adgb domains. However, as there is yet any reliable structural information concerning these Adgb domains, including using AlphaFold 2 and 3 models, it is not possible to predict binding of the other Adgb domains to Adgb-GD. Nonetheless other partner protein cannot be discounted.

Further questions remain as to the role of calmodulin binding sequence in the structure and function of Adgb. Calmodulin typically functions in a Ca^{2+}-dependent manner but can also regulate the activity of various proteins in its Ca^{2+}-free form (Villalobo *et al*., 2019). Its structural flexibility allows calmodulin to adopt multiple conformations, enabling interactions with a wide range of target proteins. As a result, calmodulin can act as a linker between different proteins or between segments of the same polypeptide chain, such as within the globin domain of Adgb (Villalobo *et al*., 2019). This interaction contributes to the structural conformation of its target proteins and modulates their functionality. A recent study investigated calmodulin's role in Adgb highlighted its regulatory influence (Nie *et al*., 2024). The research demonstrated that the globin domain of Adgb interacts with calmodulin through an IQ motif, which enhances Adgb's nitrite reductase activity (Nie *et al*., 2024). Consequently, the absence of this IQ motif could lead to significant structural consequences.

The presence of disulfide bonds in Adgb-GD may also lead to protein dimerization, as observed in the hexacoordinate protein Cygb. Cygb contains two exposed cysteine residues (CysB2 and CysE9), which facilitate the formation of inter- or intramolecular disulfide bridges

(Lechauve *et al*., 2010). Mass spectrometry analysis in the study by Lechauve *et al*. (2010) confirmed dimer formation in Cygb under non-reducing conditions, driven by intramolecular disulfide bridges. The disulfide bond in Cygb impacts the positioning of the E-helix, thereby altering the affinity of the amino acid at the distal site, in their case distal histidine (Lechauve *et al*., 2010). A comparable mechanism may be at play in Adgb-GD, where the disulfide bond could influence the affinity of glutamine at the distal site, thereby contributing to the penta-like optical spectrum trend observed across the ferric, ferrous, and deoxyferrous CO oxidation states of Adgb-GD. This is explored later in chapter 4. However, the position of the disulfide bond is in an almost identical position to Ngb on the CD loop, an important component of the stabilisation of the heme pocket (Lei *et al*., 2019). Following the reassessment of the globin C helix sequence, the prediction is that Adgb lacks a native CD loop due to the circular permutation, making the disulfide additionally important for the stability of heme binding (Reeder *et al*., 2024) and explaining the heme loss during purification and the significantly altered optical properties when the disulfide is reduced (Figure 2.3.3.3)

The similarities between these hexacoordinate proteins allow predictions about the kinetics of Adgb-GD. The presence of disulfide bridges would likely result in heterogeneous kinetics, consistent with observations in other hexacoordinate proteins (Lechauve *et al*., 2010). This suggests that the disulfide bond would not only play a structural role but would also influences the functional and kinetic properties of Adgb-GD. Further structural insights and clarification of the oligomeric state of Adgb-GD were explored through BioSEC analysis in Chapter 5.

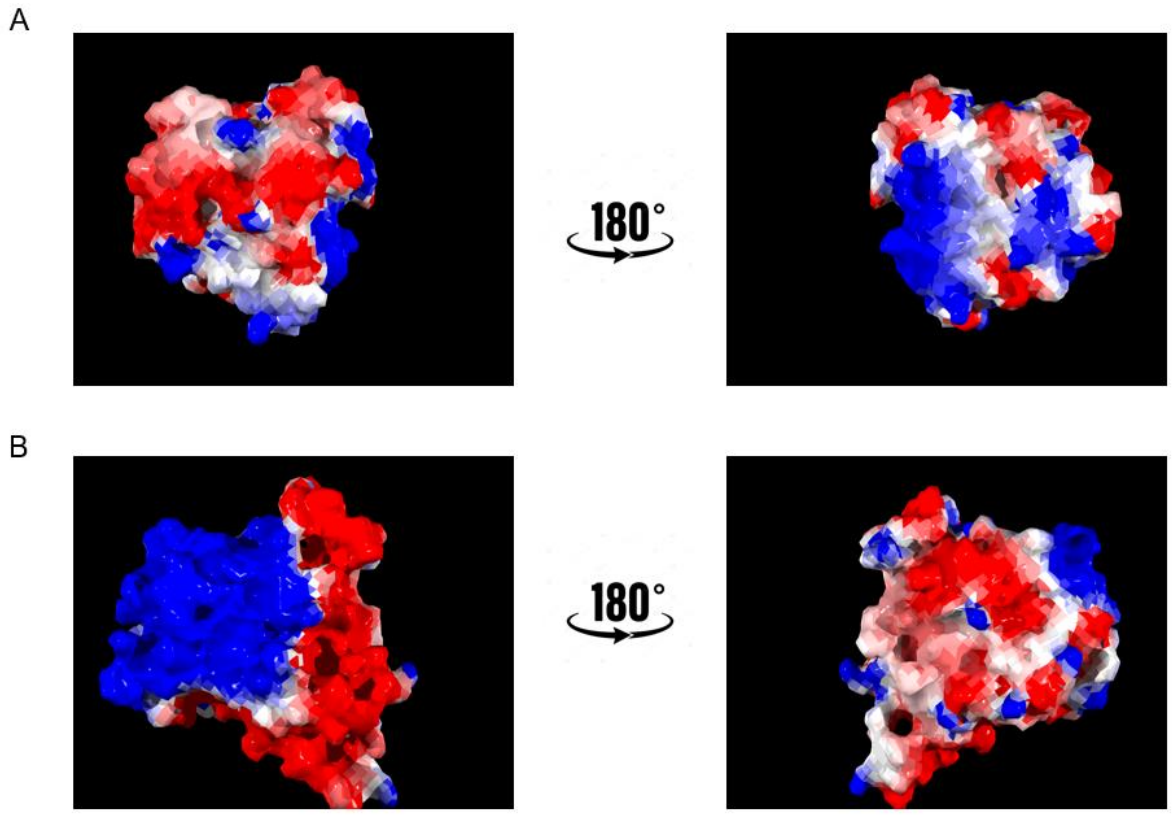


Figure 2.4.3.1. The distribution of electric charges on globular protein surfaces.

(A) Equine Mb 0° and 180° and (B) Human Adgb-GD 0° and 180°. Positively charged regions are denoted in blue, negatively charged regions are denoted in red and neutral in white. Adgb displays significant localised surface electrostatic potential compared to the more scattered potentials of Mb. The visualisation of structure files was captured using Swiss-PDB viewer. The 3D models of *Homo sapiens* Adgb-GD were aligned to the 3D structure of Equine Mb 1WLA via magic fit analysed by "compute electrostatic potential" using Swiss PDB viewer. The computational modelling employed the Poisson-Boltzmann method, which accounts for ionic shielding. The final image was prepared using Pov-Ray.

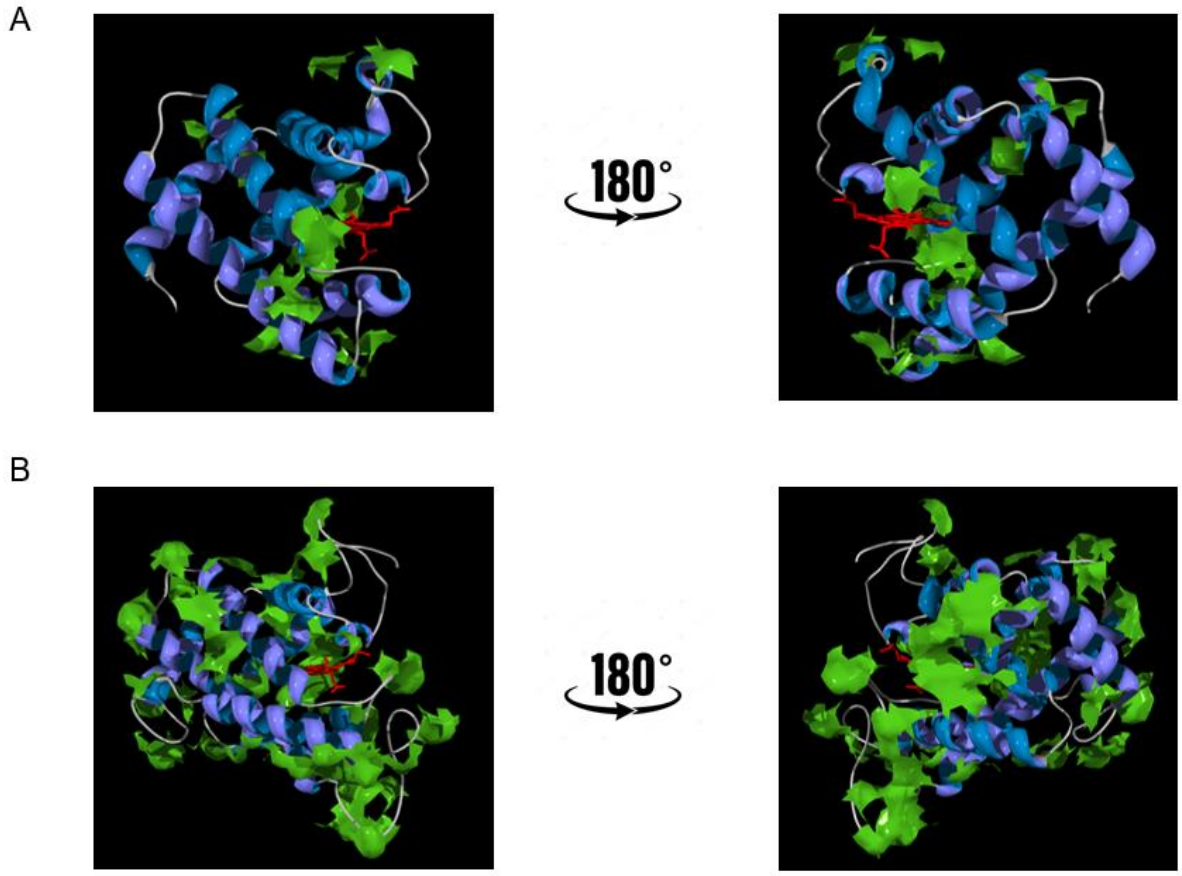


Figure 2.4.3.2. The distribution of hydrophobic patches on globular protein surfaces.

(A) Equine Mb 0° and 180° and (B) Human Adgb-GD 0° and 180°. The hydrophobic patches are denoted in green. Adgb shows a high density of hydrophobic regions on its surface compared to Mb. Visualisation was done using the similar methodology in 2.4.3.1. However, analysis was done using "compute detect hydrophobic patches" using Swiss PDB viewer.

CHAPTER 3: Investigation into the proteolytic activity in the N-terminal domain of Androglobin

3.1. Introduction

Protein function is dependent on the structure of the protein. Protein structure is established as α-helices and β-sheets fold into a stable 3D conformation. This folding is stabilised by interactions such as hydrogen bonds, ionic bonds, and disulfide bridges between secondary structural elements, ultimately determining the protein's function. Some proteins consist of multiple polypeptide chains thus forming a quaternary structure protein. An example of this protein is Hb, a heterotetramer protein that transport oxygen in the blood from the lungs to the body tissues and organs (Perutz *et al*., 1960). Adgb calpain-like N-terminal domain does not have a known 3D structure, and available secondary structure information is currently limited. Figure 3.1.1 shows the AlphaFold model of Adgb. While some regions are of high confidence, many part of the model are still predicted with low confidence. As mentioned earlier in chapter 1.3.1, the calpain-like domain of Adgb may have distinct mechanisms that regulates its interaction with the other domains of the protein, likely due to its structural arrangement. Therefore, the aim of this chapter is to improve the understanding of the underlying relationship between protein sequence, structure and function of the N-terminal region of Adgb.

This chapter examines the relationship between mutations and their roles in altering the functional properties of the N-terminal region including the way they affect Adgb functional properties. Most of the mutations in this experimental chapter will be site directed mutants, where missense changes in the base sequence of a DNA molecules results in the substitution of one amino acid for another in the gene. Another mutation used in this experimental chapter is the introduction of stop codons through nonsense mutations to generate truncated forms of the protein. Site directed mutagenesis has previously been used to study the catalytic activity of calpain domains including the study of the proteolytic mechanism of calpain 1 (Fernández-Montalván *et al*., 2004) by replacing cysteine at position 115 with alanine (C115A).

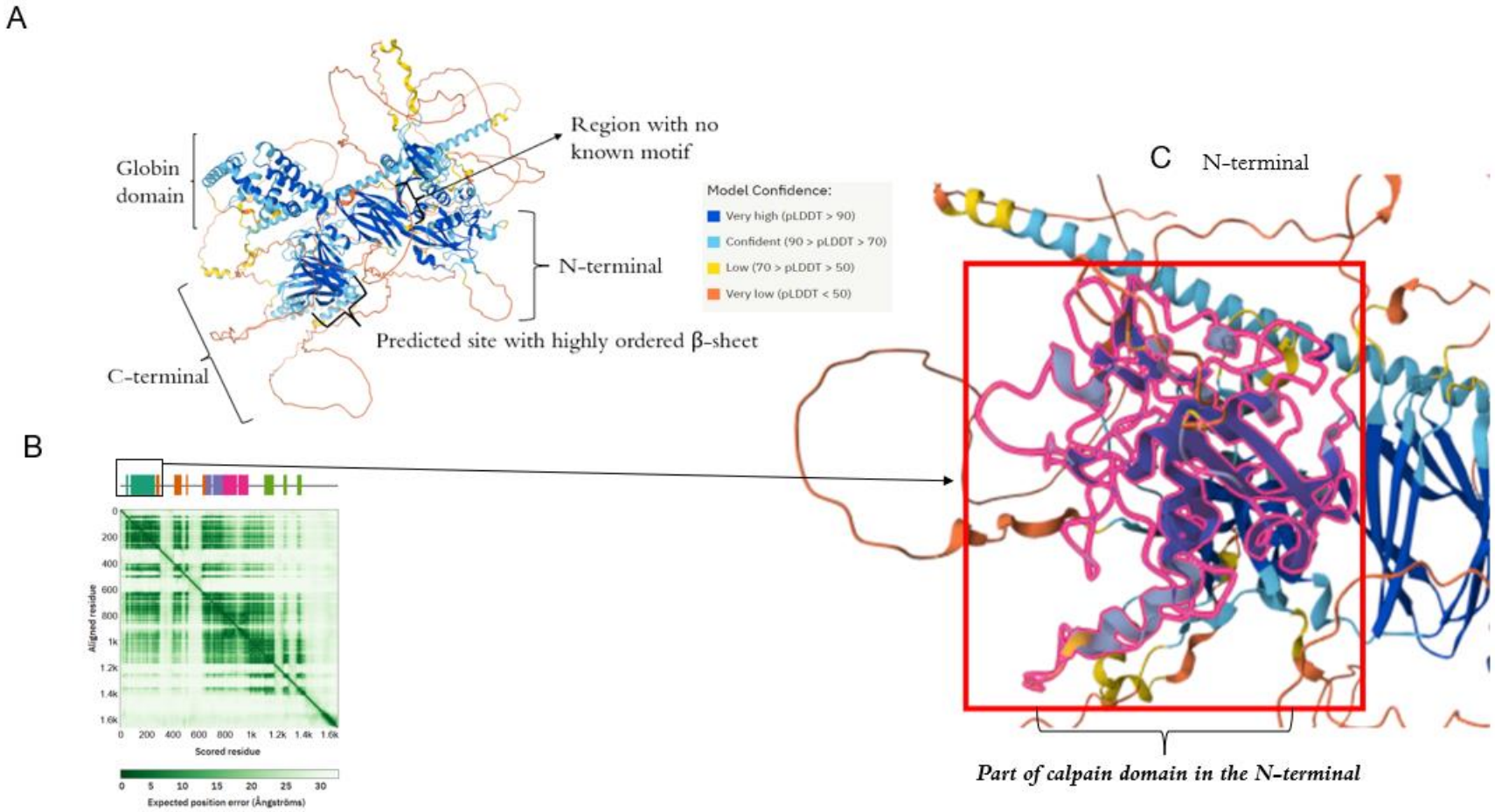


Figure 3.1.1. AlphaFold structure prediction of androglobin (Adgb) (Cheng, J *et al*., 2023, AlphaFold code AF-Q8N7X0-F1-v4).

(A) AlphaFold model: the colour scheme highlights regions of both high and low confidence in Adgb structural prediction. (B) TED Domains and Predicted Aligned Error (PAE) from AlphaFold. It shows that some of the predicted structures have low positional error (dark green), but other areas are light green, meaning that it has little confidence in how these structures may fit together into an overall structure. (C) Highlights part of the calpain domain.

The N-terminal region of Adgb was reported to be homologous to the calpain-like protease domain (CysPc) of human calpains (which includes a two-core domain, PC1 and PC2) of the catalytic domain (Hoogewijs., 2012). Specifically, the calpain domain in the N-terminal is closely related to human calpain 7 (Hoogewijs., 2012). As mentioned in section 1.3.1, the crystal structure of the catalytic domain of human calpain 7 revealed that the CysPc protease domain shares three active site residues across the two subdomains, with Cys in PC1 and His and Asn in PC2 as shown in Figure 3.1.2 (Strobl *et al*. 2000). Through sequence alignment, the calpain domain of Adgb appears to have only one of three calpain-typical active site residues (the Cys in PC1), but it contains several His and Asn residues at non-standard positions, i.e. their position in Adgb does not correspond to positions in previous sequence alignment studies (Hoogewijs *et al*., 2012).

In addition to the Cys in Adgb (Cys 132) at the position that has been proposed to align with CAPN7, there are 10 other Cys in the N-terminal region of Adgb with potential to be alternative active site nucleophiles. The proposed and alternative potential functional sites were examined in sequence alignments for conservation between animal species. Of the 11 Cys residues in the N terminal region of human Adgb, only four were highly conserved (see section 3.3 of this chapter). Mutagenesis was used to investigate the relationship between the structural changes in the N-terminal region and this region proposed function in regard to auto-proteolytic activity. Point mutations were introduced to replace the nucleophilic Cys with the aim to decrease auto-proteolytic activity. In addition, in-frame deletions, removing sections of the N-terminal domain without causing frameshifts, were also introduced to study their effect on the autoproteolytic activity of the N-terminal region. Studying proteins through in-frame deletion not only provide functional information but also serves as a valuable tool for investigating disease-related mechanisms. For example, in cystic fibrosis transmembrane conductance regulator (CFTR) protein, in-frame deletion of three nucleotides CTT results to the loss of

phenylalanine at position 508 and thus a silent codon change occurred for isoleucine-507 (I507-ATC→ATT), which affected the folding and function of the protein (Lazrak *et al*., 2013). This in turn led to defective chloride transport in epithelial cells causing cystic fibrosis. Therefore, identifying and studying in-frame deletions is a useful tool for understanding the pathophysiology of proteins and can provide insights into the functions of protein.

The first 1200 DNA bases in the N-terminal was characterised as the most likely part of the proteolytic catalytic region, corresponding to the first 400 amino acids. Therefore, deleting a section of nucleotides in this region including the main nucleophilic amino acid (cysteine or possible a serine) of the active site theoretically should inhibit autoproteolytic activity. Consequently, the mutant protein would be more likely to express as the full-length protein as it can no longer proteolyzed itself, thereby narrowing down the identity of key active site amino acids and potentially leading to insights into the specific roles of the full and fragmented protein.

The final type of mutation used to study the autoproteolytic activity of the N-terminal region was nonsense mutation. The mutation codes for a stop codon in a defined position in the sequence (in this case the middle). Like the other mutations, this mutation supports the characterisation of region on the N-terminal proteolytic activity. The functional changes of the mutant proteins were compared with the wild type (WT) protein and other globins.

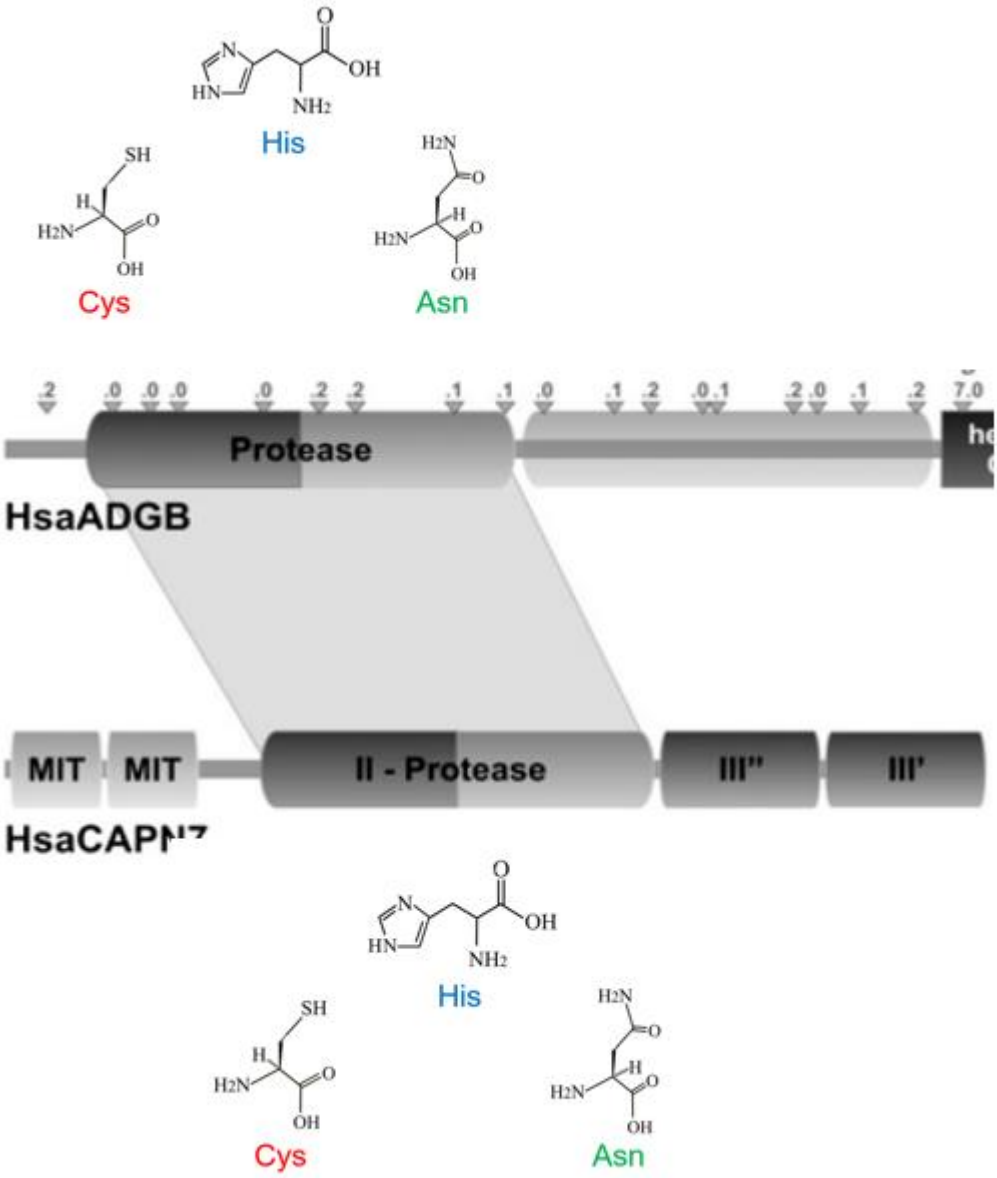


Figure 3.1.2 Diagram adapted from Hoogewijs *et al*. 2012 showing the N-terminal region of human Adgb (HsaADGB) and human Calpain 7 (HsaCAPN7).

The diagram includes the catalytic triad (Cys-His-Asn) in the protease domain of both Adgb and HsaCAPN7. This is because the calpain domain in Adgb N-terminal was proposed to be partially structurally comparable to human calpain 7 (Hoogewijs., 2012).

3.2. Materials and methods

3.2.1 Materials

All buffer materials were purchased from Sigma-Aldrich (Merck), Poole, UK. The restriction digest materials which include restriction enzymes, buffers and DNA ladder GeneRuler plus were purchased at ThermoFisher Scient UK. The PCR and ligation materials which consists of Phusion^{TM} High-Fidelity DNA polymerase, 5X Phusion^{TM} HF buffer, 10mM dNTPs, DMSO, 10x T4 DNA Ligase buffer and T4 DNA Ligase were purchased at ThermoFisher Scientific. Material for the site-directed mutagenesis including the PfuUltra High-Fidelity DNA polymerase, 10X reaction buffer, DpnI restriction enzyme and tango buffer were purchased from Agilent Technologies UK. The Adgb gene was synthesised by Epoch Life Sciences Inc, Missouri City, USA and incorporated into a mammalian expression vector pcDNA3.1+. Adgb variations, including the N-terminal and N terminal + globin domain, were produced by site directed mutagenesis. Cytoglobin in pET28a was synthesised by Epoch Biosciences (USA). The neuroglobin gene in pET28a was a kind donation of Prof Worrall (University of Essex). All primers were purchased at Eurofins Genomics Europe and designed using primerX (www.bioinformatics.org/primerx/). Protein ladder used is ThermoFisher PageRuler™ Prestained Protein Ladder, 10 to 180 kDa protein ladder. Western blot used the ThermoFisher BenchMark™ His-tagged protein standard as ladder.

3.2.2 Identification of conserved Cysteines in the N-terminal region of Androglobin

The 1667 amino acid human Adgb gene was aligned using multiple sequence alignment via ClustalO on Jalview (version 2.11.4.1), with manual adjustments made within Jalview. The N-terminal region sequences were then manually checked using Jalview for poorly aligned,

misidentified or incomplete sequences at any of the N-terminal regions, which were subsequently removed.

3.2.3 Engineering mutant Androglobin N-terminal recombinant expression vectors

3.2.3.1 Removal of extraneous bases

Previous cloning of the cDNA sequences of the N-terminal and N-terminal with globin domain (N-terminal-GD) into *E.coli* recombinant expression vector pET28a showed that extra bases, termed here as “extraneous DNA” had been inserted to the pET28a vector. This extraneous DNA was from the previous excision of the full protein and N-terminal region from pcDNA3.1+ into pET28a vector via restriction digest and ligation (NotI and BamHI). These extra bases are located between the thrombin cleavage site and the beginning of the protein sequence (see results section Figure 3.3.2.1). Therefore, the removal of this extraneous DNA, without causing a frameshift, was desirable. Both constructs ("N-terminal" and "N-terminal-GD") used the pET28a vector with a 6x His-tag and T7 promoter for expression in *Escherichia coli* BL21 (DE3) pLysS cells.

Extraneous DNA was removed via standard PCR using 5’ phosphorylated [PHO] primers as shown in the table below (Table 3.2.3.1). It was this corrected WT expression vector that was used for all subsequent site-directed mutagenesis and in-frame deletions. The primers used to remove the extraneous DNA and in-frame deletion constructs were designed using benchling (Cloud-based platform for biotech R&D | Benchling). The primers melting temperature (T_M) was calculated using the T_M calculator (Tm Calculator | Thermo Fisher Scientific - UK). The amplification of the fragments from extraneous DNA primers coding the N-terminal was carried out using the conditions in detailed in Table 3.2.3.2; a 3-step PCR

protocol via Phusion High-Fidelity DNA Polymerase (Thermofisher, UK). After PCR, the DNA was cleaned up using the DNA cleanup kit (Monarch® PCR & DNA Cleanup Kit) before ligation to remove contaminants from the amplified DNA. The ligation of the self-circularisation of linear DNA; N-terminal was done using T4 DNA ligase and T4 DNA ligase buffer according to manufacturer instruction. The mixtures were incubated for 1 hour at 22 °C and further incubated in the fridge at 7 °C overnight.

The ligated N-terminal domain DNA were firstly transformed into *E.coli* XL1B cells by heat shock method (42 °C, 90 s), plated and incubated overnight at 37 °C. Three colonies from each plates were grown overnight in a shaking incubator (180 RPM) at 37 °C in LB media (10 mL LB, 50 µg/mL kanamycin sulphate). The cells were isolated via centrifugation (4000 RPM, 23 °C, 10 min). The DNA were then prepared using a GeneJET Plasmid miniprep kit. Total DNA volume of 30 mL for each colony was sent for Sanger sequencing to confirm the absence of the extraneous DNA.

3.2.3.2 Introducing mutations into the N-terminal domain

After confirming the absence of the extraneous DNA, similar process was repeated to make the different in-frame deletion mutants on the N-terminal region. In total, nine in-frame deleted mutations were made using the primers shown in table (Table 3.2.3.1). Up to 30 µL of DNA for each frameshift were sent for analysis via Sanger sequencing to confirm the frameshift mutation. Samples from the nine mutants of the in-frame deletion mutation were analysed using agarose gel to verify the absence of the deleted sections. The gel was prepared by dissolving agarose in TAE buffer to make 1% agarose gel. The samples were then mixed with loading dye (GelRed) at 1:1 before loading on the gel.

Primers were used to enable the single base-pair site-directed mutagenesis at the four cysteine sites: Cys132, Cys181, Cys206 and Cys406 as well as the base-pair substitution for addition of a premature stop codon at position E395 (Table 3.2.3.1). Primers were also designed using benchling and synthesised by Eurofins Genomics. The amplification of the fragments from these primers were carried out using the conditions in QuikChange II site-directed mutagenesis handbook as shown in (Table 3.2.3.3). After the PCR reaction, the template DNA were digested by DpnI following incubation at 37 °C for 1 hour. 2 µL of the DpnI-treated DNA from each of the mutagenesis amplification reaction were transformed in XL1B competent cells by heat shock method (42 °C, 90 s), plated onto LB media agar plate including antibiotic (50 µg/mL kanamycin sulphate) and incubated overnight at 37 °C. A colony from each plates were grown overnight in a shaking incubator (180 RPM) at 37 °C in LB media (10 mL LB, 50 µg/mL kanamycin sulphate). The cells were isolated via centrifugation (4000 RPM, 10 min). The DNA were then purified using the miniprep kit (Qiagen GeneJet) and sent to Eurofins Europe for Sanger sequencing. Some samples containing Cys→Gly mutations towards the centre of the protein sequence (i.e. C406G) and stop mutation E395 were not covered by the standard sequencing primers by Eurofins. Therefore, a sequencing primer was designed using benchling (Adgb Seq primer mixer, Table 3.2.3.1) and premixed with the DNA for sequencing in a 1:1 ratio before sending for Sanger sequencing.

Table 3.2.3.1: Forward (F) and reverse (R) primers used in the mutagenesis study. [PHO] represents 5' phosphorylated primers and trunc represents the F and R primers used in their truncation.

Primers	Sequences
N-terminal Adgb extraneous-F	[PHO] GGATCCATGGCATCCAAGCAGACTAAAAAGAAAGAG
N-terminal Adgb extraneous-R	[PHO]GCGCGGCACCAGGCCGCTGC
Adgb-C132G-R	CATCAATTCGCTGCCAAGGAGATGCTCG
Adgb-C132G-F	CGAGCATCTCCTTGGCAGCGAATTGATG
Adgb-C181G-R	CTTTAACAGCCTTGCCCAGACTGTAAATGTG
Adgb-C181G-F	CACATTTACAGTCTGGGCAAGGCTGTTAAAG
Adgb-C206G-R	GATCTTCCGCCAACCTCCCATCCAGTAG
Adgb-C206G-F	CTACTGGATGGGAGGTTGGCGGAAGATC
Adgb-C406G-R	GGAAGTCTGTATCGCTGAGCTGCAATCAGACAGAGACTGCAC
Adgb-C406G-F	GTGCAGTCTCTGTCTGATTGCAGCTCAGCGATACAGACTTCC
Adgb-Nterm stop-R	GGAATACTGTACTTAAGAGCTCGGGCG
Adgb-Nterm stop-F	CGCCCGAGCTCTTAAGTACAGTATTCC
Adgb Seq primer mixer	GCAAACATTGACATTCATGTG
Adgb_Nterm_trunc_R1	[PHO]CTTGGATGCCATGGATCCGCG
Adgb_Nterm_trunc_R2	[PHO]GCCTGTTTTGTCCTTTTCCTTCGC
Adgb_Nterm_trunc_R3	[PHO]GGACAGTATTCCCCCGTTGAATATC
Adgb_Nterm_trunc_F1	[PHO]CCATGGGAACACATTTACAGTCTG
Adgb_Nterm_trunc_F2	[PHO]GAATTCGAGCTTTGGCCCATGCTG
Adgb_Nterm_trunc_F3	[PHO]GAGTCAAGTCTGACTACCTTGAAAGCC

Table 3.2.3.2 Overview of PCR three step protocol program for amplifying N-terminal and its various domains

Cycle step	**3-step protocol**		**Cycles**
	Temperature	Time (secs)	
Initial denaturation	98°C	30 s	1
Denaturation	98°C	8 s	30
Annealing	64°C	20 s	
Extension	72°C	240 s	
Final extension	72°C	480 s	1
Hold	4°C	Hold	Hold

Table 3.2.3.3 Cycling parameters for Quikchange II site-directed mutagenesis method from the Instruction Manual (Agilent).

Cycle step	**3-step protocol**		**Cycles**
	Temperature	Time (secs)	
Initial denaturation	98°C	30 s	1
Denaturation	95°C	30 s	18
Annealing	55°C	60 s	
Extension	68°C	480 s	
Final extension	68°C	480 s	1
Hold	4°C	Hold	Hold

3.2.4 Expression of N-terminal region of Androglobin

The protocol used in the expression of the N-terminal region of Adgb was essentially identical to the expression protocol for the globular protein, as described in the previous chapter (section 2.2.2). However, only IPTG was used to induce protein expression, as ALA and iron supplementation was not required for heme synthesis, or CO for heme stabilisation. The cells were harvested by centrifugation (8000 g, 20 min, 4 °C). The pellets were resuspended in media and transferred to a 50 mL falcon tube where they are re-harvested by centrifugation (8000 g, 20 min, 4 °C) and frozen at -80 °C for storage. The cell was resuspended in 20 mM NaPi 500 mM NaCl pH 7.4 buffer and then lysed using an Avestin Emulsiflex C3 homogenizer at ~15000 psi, 3 passes.

Purification was done via GE Healthcare nickel affinity column using similar condition for the purification of Ngb and Cygb (see section 2.2.2). Briefly, column was pre-calibrated with buffer (20 mM NaPi with 500 mM NaCl and 20 mM imidazole, pH 7.4). After loading the protein was washed with 20 mM NaPi 500 NaCl 20 mM imidazole pH 7.4 buffer (5+ column volumes) before eluting with 20 mM NaPi 500 mM NaCl 500 mM imidazole pH 7.4 buffer in a linear gradient over 50 mL. Imidazole was removed by dialysis (8000 MWCO 1 mM sodium tetraborate, pH 10, 3 changes; a change every ~1.5 h, with the final change before overnight in a cold room). Similar process was done for the expression of the N terminal + globin domain, however because this protein contains heme, ferric citrate (100 μM) and 5-aminolaevulinic acid (ALA) (250 μM), followed by bubbling with CO, was also used to induce the cells at the point of protein expression induction.

3.2.5 Expression of mutants Androglobin N-terminals in *Escherichia Coli*

After confirming the presence of the different mutations and frameshifts in the DNA via Sanger sequencing, the mutants were then transformed into *E.coli* BL21 (DE3) competent cells by heat shock method (90 s, 42 °C) and then same expression protocol was applied for the N-terminal protein, as described above (section 3.2.4). An empty pET28a vector and Ngb were used in parallel as negative and positive controls respectively. Following protein expression, the expressed proteins were purified using GE Healthcare immobilised metal (nickel) affinity column via the AKTA pure chromatography system under denaturation conditions following the method as described in section 2.2.3.

3.2.6 Auto-proteolytic analysis of the calpain-like N terminal region by Western blot

The analysis used to study the auto-proteolytic activity in the N-terminal region was to examine the peptide fragmentation and full protein expression by sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and western blot. The SDS-PAGE analysis was carried out following the protocol described in section 2.2.9, and the gel was subsequently stained with Biosafe Coomassie stain for 1 h (gentle rocking, 20 rpm). In parallel, a second SDS-PAGE gel was prepared and analysed using western blotting to further validate protein expression. The technique used here has been described in section 2.2.10. The antibody used throughout the study was also 6X His-Tag monoclonal antibody (His.H8) with conjugated HRP (ThermoFisher), unless stated otherwise. The specificity of the antibody-antigen interaction enables target protein to be identified in a complex protein mixture, in this case the fragments as a result of autoproteolytic activity, although only those with the His-tagged N-terminal

region. The fragment sizes were obtained using GelAnalyzer v23.1.1 (available at www.gelanalyzer.com) by Istvan Lazar Jr., PhD and Istvan Lazar Sr., PhD, CScas.

3.3. Results

3.3.1 Identifying which Cysteine is the Nucleophile responsible for the autoproteolytic activity in the N-terminal region via Sequence Alignment

The percentage homology of various N-terminal Adgb cysteines were analysed using multiple sequence alignment. An example of the alignment is shown in Figure in 3.3.1.1. Sequence alignments were used to calculate the percentage of sequences containing each specific cysteine residue. Figure 3.3.1.2 shows the percentage conservation of the specific cysteines from the multiple sequence analysis. Previous studies suggested Cys132 as the candidate cysteine for the active site on the catalytic triad (Hoogewijs *et al*., 2012). However, the availability of 243 sequences for alignment, which compared to the previous study where only 33 sequences were available (Hoogewijs *et al*., 2012), enabled a more accurate and comprehensive sequence alignment. Therefore, this site was an area of interest to re-examine and test by a mutagenesis study to examine its effect on the autoproteolysis of Adgb. The pie chart in Figure 3.3.1.2 exhibits sequence conservation of several cysteines from the sequence data of Figure 3.3.1.1. Overall, the results from the pie charts showed that Cys132 is not the most conserved cysteine residue among the four highly reserved cysteine sites in the N-terminal region, despite being the proposed cysteine in the catalytic triad in the 1st literature (Hoogewijs., 2012). It is conserved in 90.17% of the 243 species analysed, ranking third in conservation. In addition, the 3 of the 33 sequence alignments shown previously to not show a cysteine equivalent C132 of human Adgb, were dismissed as sequencing/alignment errors (Hoogewijs *et al*., 2012). Here, 24 sequences did not align with a cysteine at position 132, giving more confidence that conservation of a cysteine at this position is not universal.

In contrast, Cys181 and Cys206 cysteine sites are more highly conserved compared to Cys132. They are the most conserved across all the species reaching 99% and 97% respectively

as shown in Figure 3.3.1.2B and 3.3.1.2C. Therefore, suggesting that the cysteine residue at the catalytic triad could potentially be at these or other alternative positions. The alignment of these sites showed that alternative amino acids at these positions are also largely potential nucleophiles as shown in Figure 3.3.1.2B and 3.3.1.2C (Thr in position 181 and Ser in 206). Although the amino acids at these sites are infrequent (with Thr and Ser showing conservation rates of 0.85% and 2.49%, respectively), the observed conservation pattern remains significant. Nucleophilic amino acids like serine play crucial roles in various proteases, including chymotrypsin, trypsin, and elastase. This suggests that alternative amino acids at these positions in the Adgb of some species may also function as nucleophiles.

In contrast, when analysing the conservation patterns at Cys132 and Cys406, multiple alternative amino acids were observed. Many of these alternatives are not potential nucleophiles as shown in the sequence alignment (Fig 3.3.1.2A and 3.3.1.2D). Cys132 share its site with four amino acids including one nucleophile i.e. serine and it's conserved across 0.43% of the species. Whereas the site at Cys406 has nine other amino acids other than Cys (74.37%), including three nucleophiles: Lys (8.4%), Ser (2.94%) and Thr (0.84%). The conservation patterns described above can be better understood when represented in a pie chart, as illustrated in Figure 3.3.1.2B and 3.3.1.2C vs 3.3.1.2A and 3.3.1.2D. In conclusion, the charts clearly indicate that Cys181 and Cys206 are more likely to serve as alternative cysteine sites. Cys406 is less likely to be part of the catalytic triad due to its very low level of conservation compared to the other three especially Cys181 and Cys206. However, it is also possible that some species lack autoproteolytic activity.

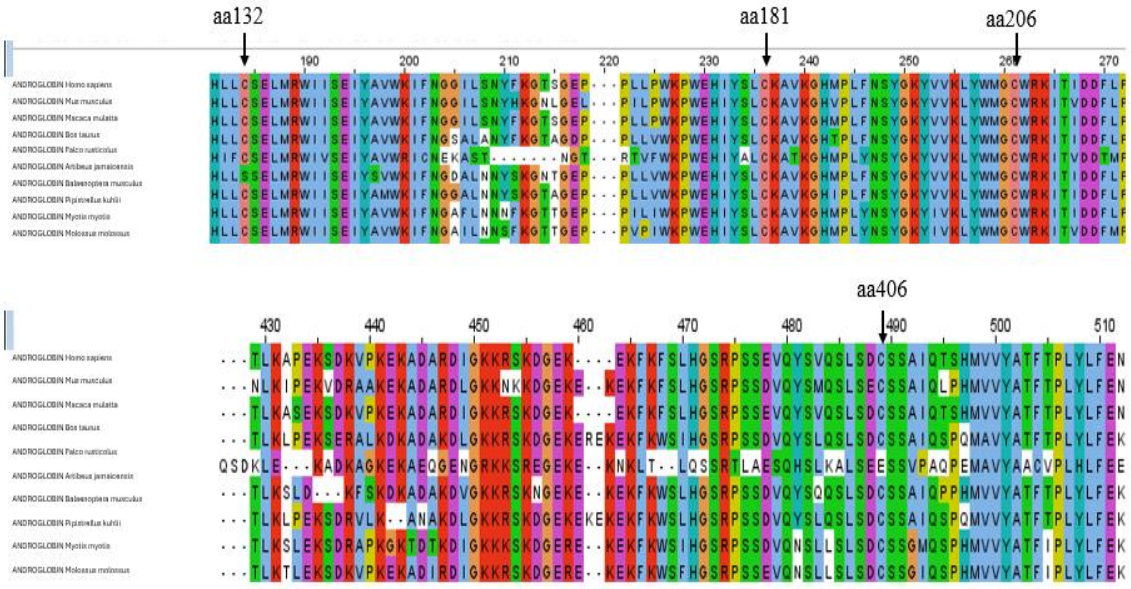


Figure 3.3.1.1. Multiple sequence alignment of *Homo sapiens* Adgb N-terminal domain and the Adgb globin domains of various animals.

This figure shows a subsection around ~200 aa section of the full-length sequence alignment. Alignment (from 243 sequences) was performed using ClustalO and adjusted manually with Jalview. Arrows point to the cysteines found at the N-terminal sites are alternative cysteine for the catalytic site of the N-terminal domain. All sequences were obtained from Uniprot.

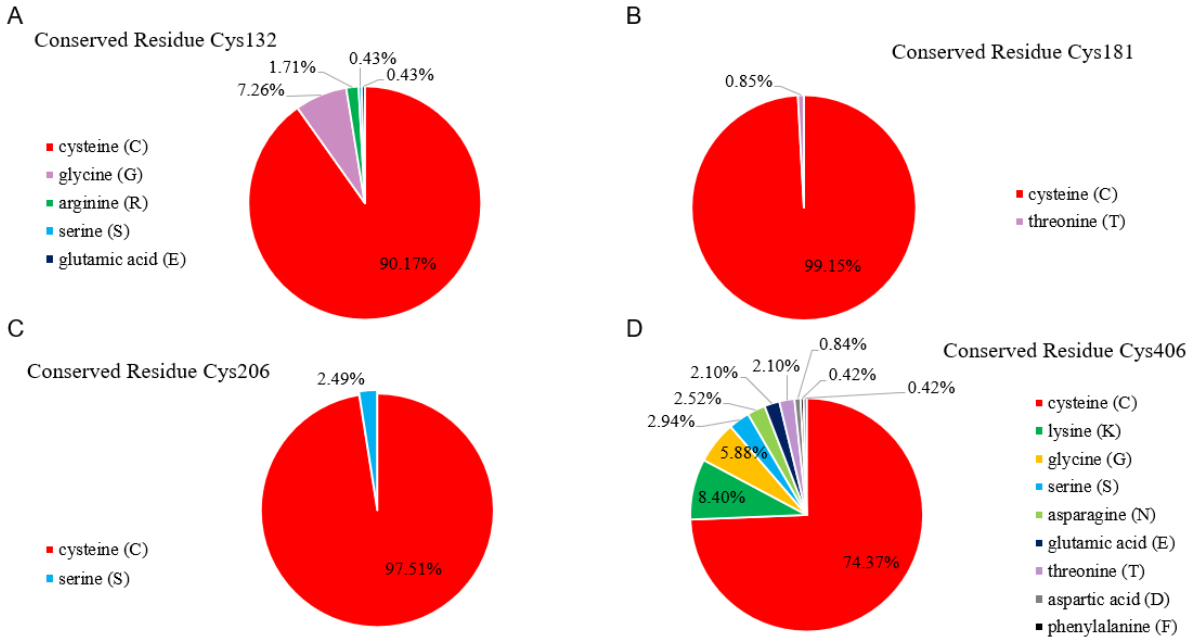


Figure 3.3.1.2. Pie charts showing the percentage of conserved cysteine residues found at the calpain-like protease domain of the N-terminal region of Adgb.

Where (A) Cys132, (B) Cys181, (C) Cys206, and (D) Cys406. Multiple sequence alignment of the full-length Homo sapiens Adgb, including the calpain-like protease domain were aligned with 242 other Adgb sequences. Key: the listed degree of amino acids conservation across the species corresponds to the percentages in descending order.

3.3.2 Design of N-terminal Androglobin Mutants for Protein Expression

Prior to the experiments presented in this study, the N-terminal region was excised from a pcDNA3.1+ vector into the pET28a vector, which resulted in the addition of extraneous DNA not part of the Adgb gene. Therefore, PCR using phosphorylated primers was used to remove this extraneous DNA present after the thrombin site at the N-terminal region (Figure 3.3.2.1). The removal of the extraneous DNA via PCR was successful as shown in the same Figure 3.3.2.1. Subsequent expression and characterisation studies were only carried out after the removal of the extraneous DNA.

Phosphorylated primers were also used to delete the specific DNA sections targeted for in-frame deletion mutations. There were nine mutations in total made from the six primers (three forward and three reverse) listed in Table 3.2.3.1. These nine mutations were made from combining the forward and reverse primers in various combinations as shown in Figure 3.3.2.2. The agarose gel showed successful PCR deletion of the specific DNA sections targeted for in-frame deletion mutations. Clear, distinct and bright bands were observed in all the lanes indicating that the amplification was successful with high concentration of DNA. Bands at approximately the predicted location corresponding to the expected base pair sizes of target fragments were observed. The positive control lane which contained the wild type N-terminal DNA (bp of 7642) used in this PCR showed the comparative location of the N-terminal fragments. The deleted sections were then confirmed via Sanger sequencing and the results for each of the nine mutations are shown from Figure 3.3.2.3. These figures display the full sequence of each mutant protein according to the combination mentioned in Figure 3.3.2.2 (appendix A2.1). It also showed that all the deletion are in-frame with the original gene and no frameshift occurred during the process of deleting the sections (appendix A2.1).

Single base mutations were used to change the cysteines at the four positions reported in the method section to glycine (C132G, C181G, C206G and C406G), as well as the mutation of E395 to a stop codon. This was designed to decrease the autoproteolytic activity through removing the nucleophilic active site amino acids. Sanger sequencing showed successful substitutions of the four-cysteine position to glycine as well as the E395 to a stop codon (Figure 3.3.2.3, 3.3.2.4 and appendix A2.1.3B).

See appendix (Figure A2) for a more comprehensive detailed figure containing annotations of the N-terminal region sequences and the different primers combinations and their respective cysteine sites.

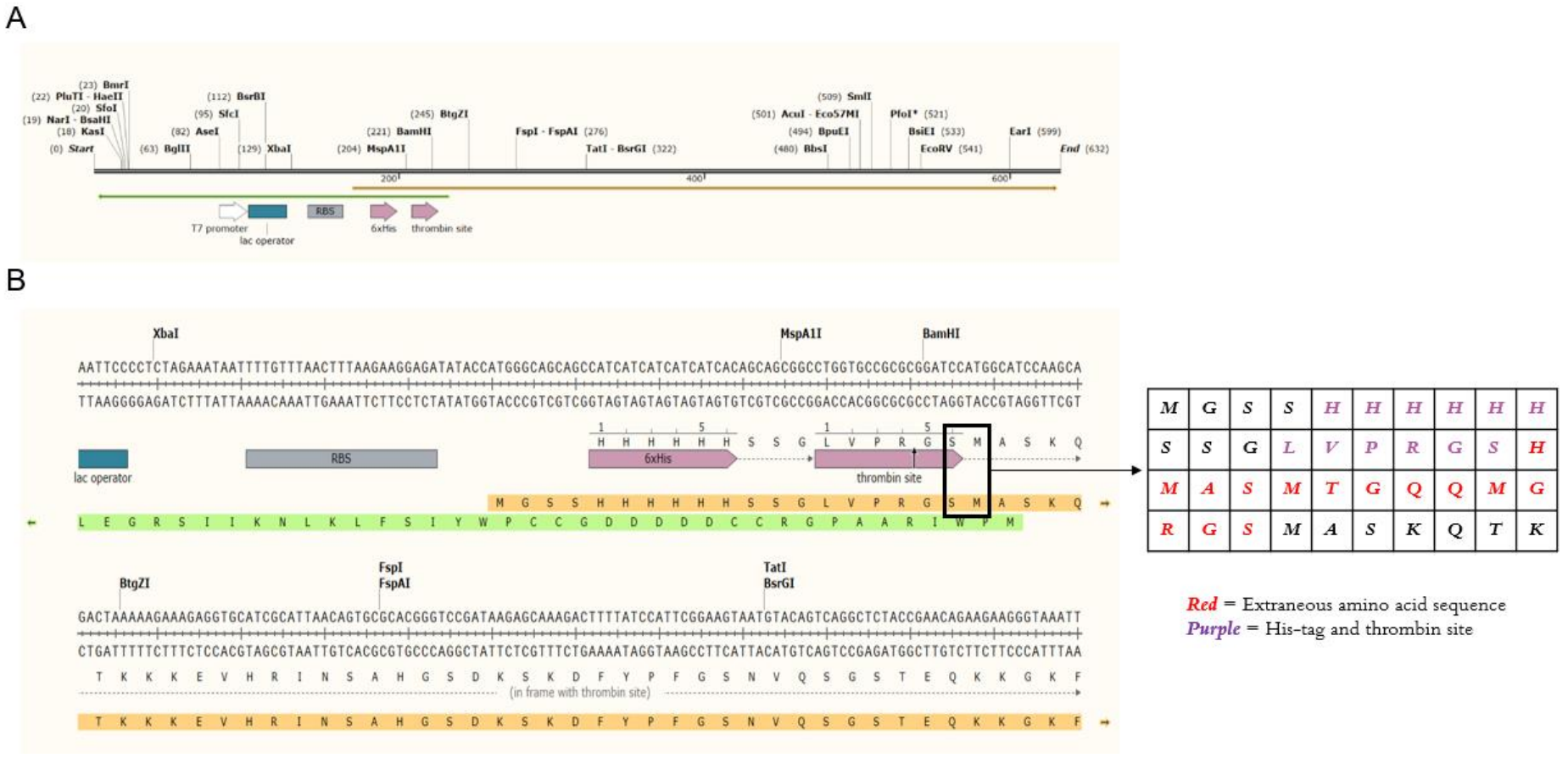
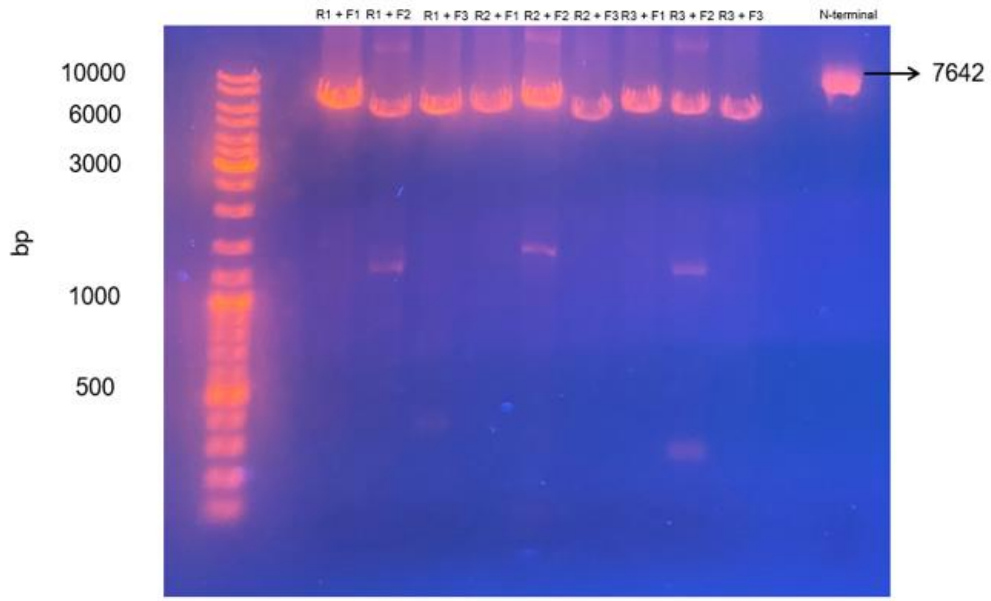


Figure 3.3.2.1. The amino acid sequences of the first ~120 bp of the N-terminal region.

The sequence within the ~120 bp of the N-terminal region contain the extraneous amino acids. It also contains the His-tag and thrombin amino acid sequence (purple font, inset table). The extraneous sequence, located between the thrombin cleavage site and the start of the Adgb sequence, is highlighted in red in the inset table. The DNA sequences used for protein expression were those in which the extraneous DNA between the thrombin site and the N-terminal region had been deleted.

A



B

Primers combination	*Number of key cysteines site cut out in protease region*	Predicted size (bp)
Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F1	1	7138
Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F2	3	6961
Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F3	3	6622
Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F1	1	7348
Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F2	3	7171
Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F3	3	6832
Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F1	0	7594
Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F2	2	7417
Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F3	2	7078

Figure 3.3.2.2. Amplified PCR products following deletion mutagenesis.

(A) PCR products were obtained from combining 6 primers in different combination as shown in the table (B). 1 % Agarose gel was used. N-terminal lane was used for the control (wild type).

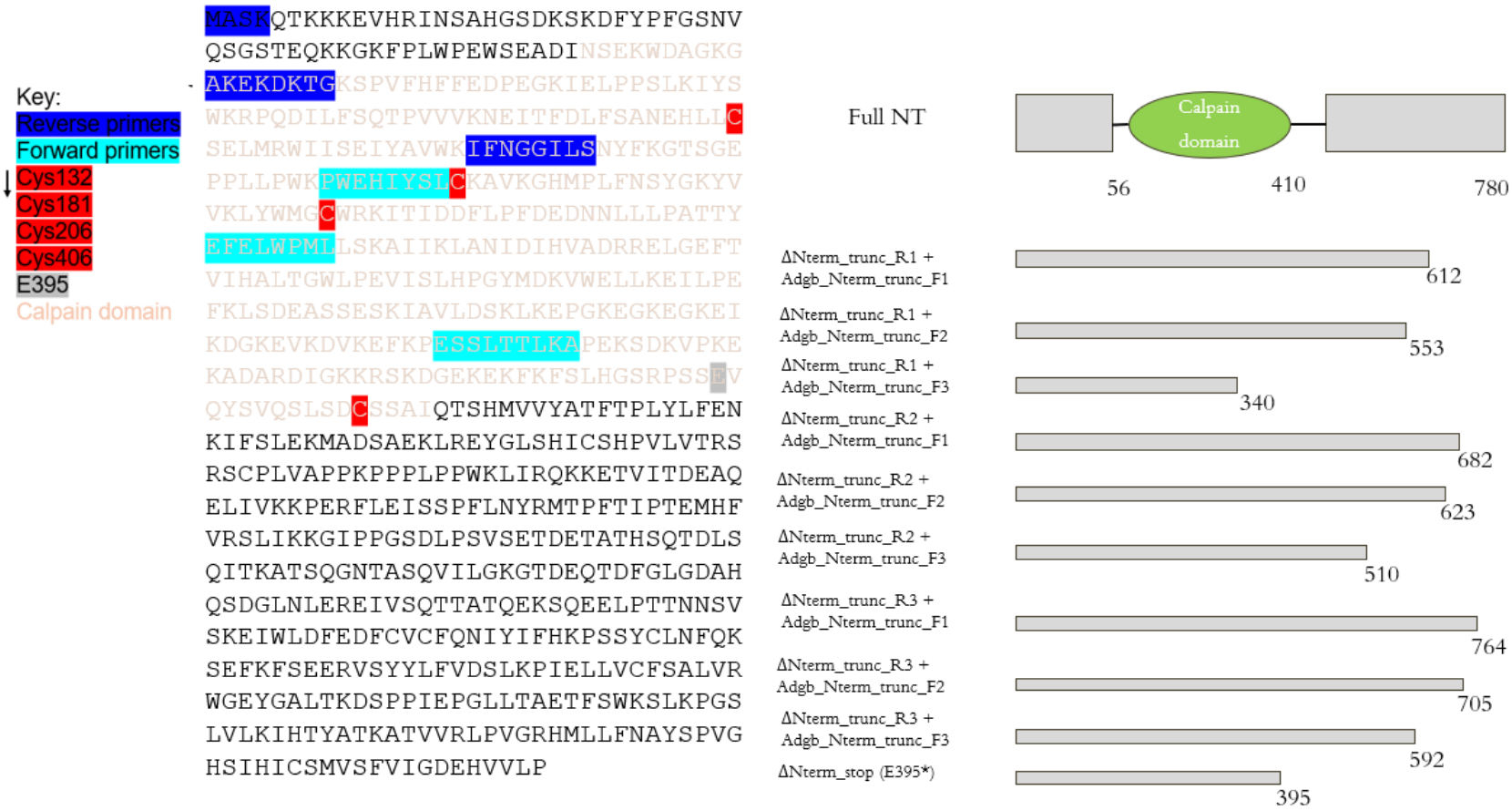


Figure 3.3.2.3. Show the truncated mutants derived from Figure 3.3.2.2 after Sanger sequencing.

The sequence shows all the key residues in the N-terminal domain including Cys132, Cys181, Cys206 and Cys 406. The truncated mutant derivatives after using the primers in table 3.2.3.1 is also depicted.

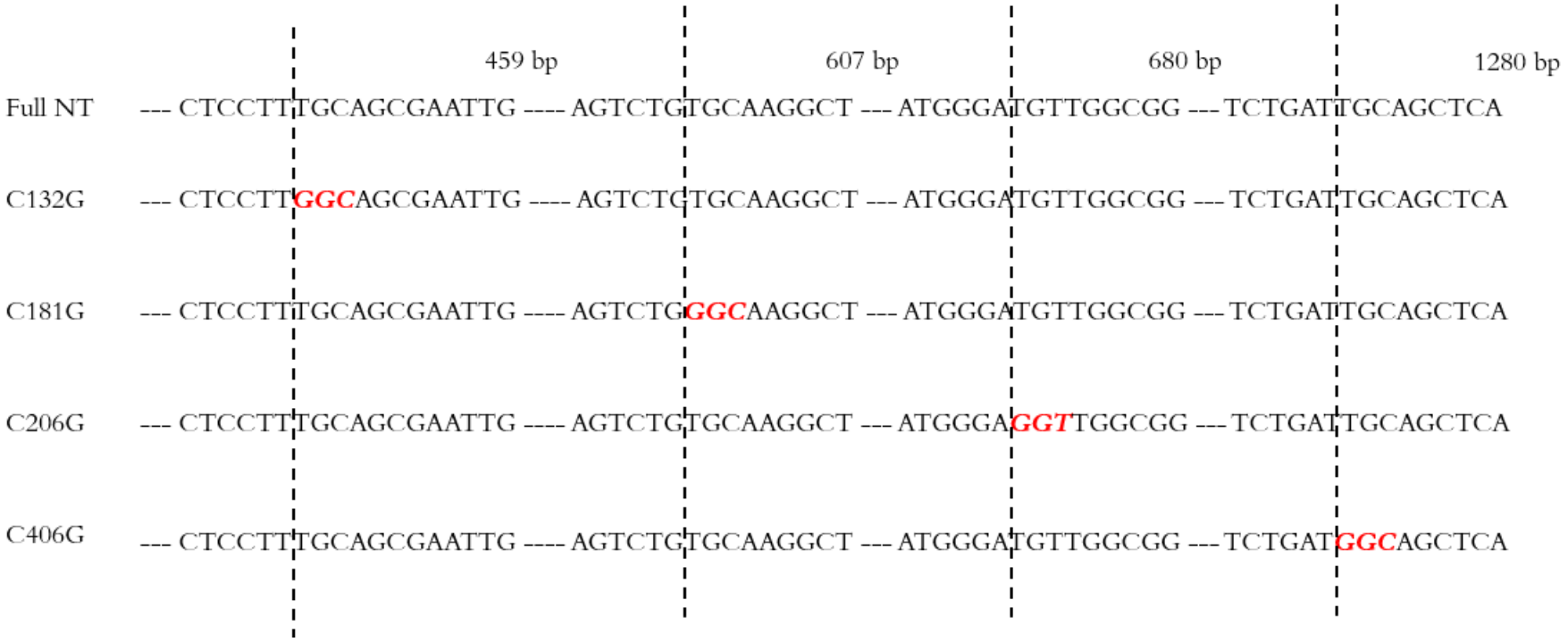


Figure 3.3.2.4._DNA sequences showing the successful mutation following site directed mutagenesis and Sanger sequencing.

The point cysteine mutation to glycine for Cys132, Cys181, Cys206 and Cys406 are highlighted in red.

3.3.3 Recombinant Expression of Androglobin N-terminal domains

The N-terminal domain appears to be proteolytic following *E.coli* expression as shown in Figure 3.3.3.1. The theoretical Mw of the N-terminal region was 88.4 kDa + 2 kDa His-tag according to inputting the sequence into the Expasy server. There are multiple bands observed after SDS–PAGE analysis, as shown in Figure 3.3.3.1A, all of which were significantly below the theoretical Mw for the peptide expressed at ~50, ~25 and ~15 kDa. This may be due to *E. coli* halting expression beyond a certain protein length, resulting in the production of a truncated form in the recombinant expression system. However, it is more likely that the calpain active site influences the N-terminus, leading to self-cleavage of the protein (autoproteolysis), which in turn resulted in the appearance of multiple low Mw bands during the analysis. Alternatively, this could also be due to the recombinant protein co-eluting with endogenous *E. coli* proteins, which are then separated from the recombinant protein during the gel electrophoresis. After western blot analysis, where the antibody binds to the N-terminal His-tag, the blot only produced a single band at ~50 kDa. This is shown in Figure 3.3.3.1B, demonstrating that this is likely a proteolyzed Adgb fragment that contains the N-terminal His-tag. Note that the N-terminal fragments without the His-tag would not be visible on the western blot gel. The results indicates that the additional bands observed in the SDS PAGE were either co-eluted endogenous proteins or other fragments of the N-terminal domain lacking the His-tag.

An alternative possibility to auto-proteolysis is that *E.coli* serine proteases may result in the proteolytic attack of the recombinant protein following lysis, explaining the SDS-PAGE and Western blot data. To decrease the effects of proteolytic activity from *E.coli*, the serine protease inhibitor phenylmethylsulfonyl fluoride (PMSF) was added immediately to the cell lysate during cell following lysis to inhibit exogenous proteolytic activities and to see if the inhibitor may impact on the potential auto-proteolytic activity of the N-terminal region of

Adgb. Similar gel electrophoresis results were obtained as shown in the Figure 3.3.3.2A, showing that this was likely not the case that serine-based proteolytic enzymes were a cause of the N-terminal Adgb fragmentation.

To examine if the extra bands on the SDS PAGE resulted from aggregation with co elution endogenous *E. coli* proteins through purification, the N-terminal Adgb was also purified under denaturing conditions using 8 M urea, as shown in Figure 3.3.3.2B. The lane shows clearer bands overall, with a prominent single band at approximately 27 kDa, different to the 50 kDa observed for the protein purified under non-denaturing conditions. Faint bands were observed in the western blot. As a preliminary study aimed at finding the correct method for purifying the proteolytic fragment, this procedure requires further optimisation. Similar results to the WT N-terminal domain were observed for the N-terminal-GD construct during SDS-PAGE analysis, as shown in Figure 3.3.3.2C. Specifically, fragments were observed at ~53, ~22 and ~19 kDa with the theoretical M_W of 112.1 kDa + 2 kDa His-tag, again highlighting the susceptibility of Adgb to fragment following purification.

These data indicate that the Adgb N-terminal fragment co-elutes with endogenous *E. coli* proteins and potentially other Adgb fragments. Alternatively, the appearance of smaller fragment under urea-treated condition may imply an additional or secondary proteolytic cleavage; however, this is unlikely since any proteolytic enzymes should also be denatured under such conditions.

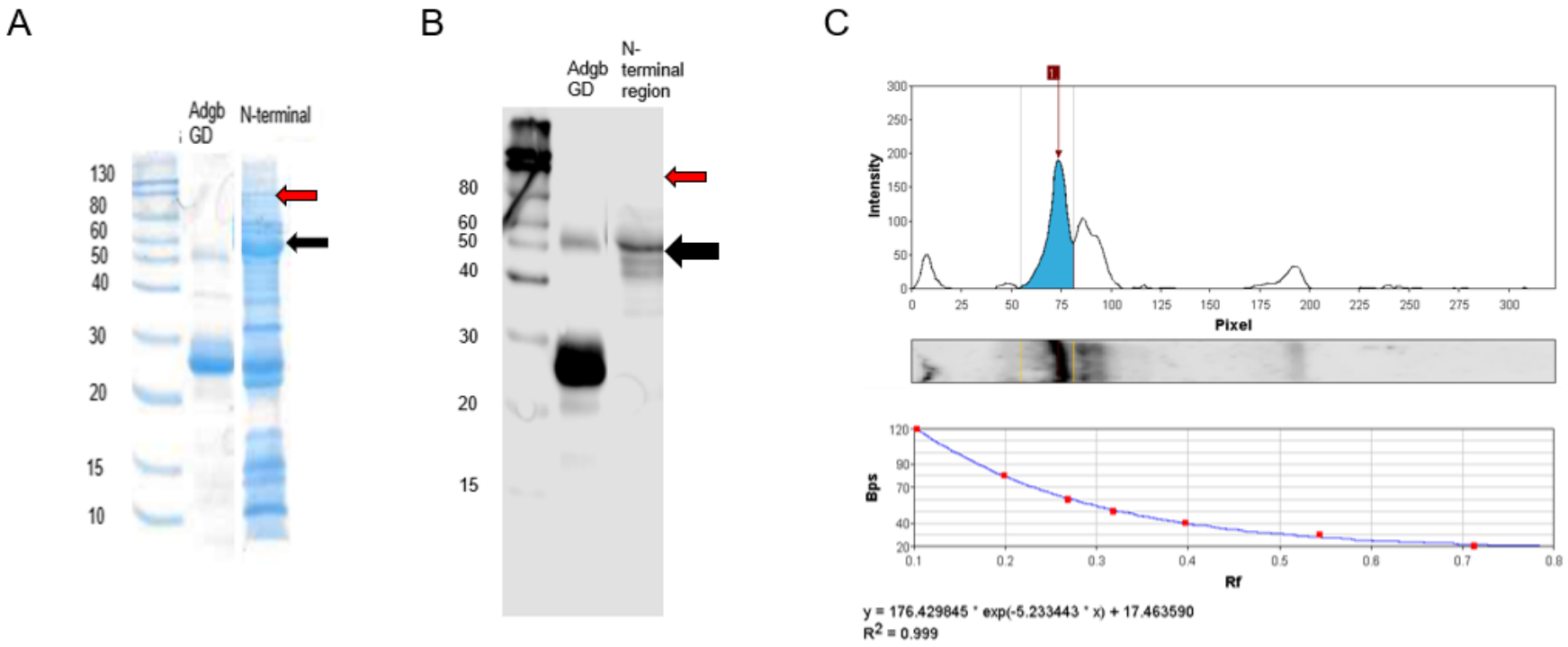


Figure 3.3.3.1 Analysis of auto-proteolytic activity of the N terminal, calpain-line domain of Adgb.

(A) SDS-PAGE and (B) Western blot analysis of Adgb N-terminal region. The marker lane showed bands at known molecular weights, allowing estimation of the sizes of the protein bands in the sample lanes. The purified samples showed multiple bands in the SDS-PAGE analysis with the main band at ~50 kDa (black arrow) according to gel analysis via *GelAnalyzer v23.1.1* as shown in (C). The theoretical M_w of the N-terminal region (+ His-tag) is 90.4 kDa (red arrow).

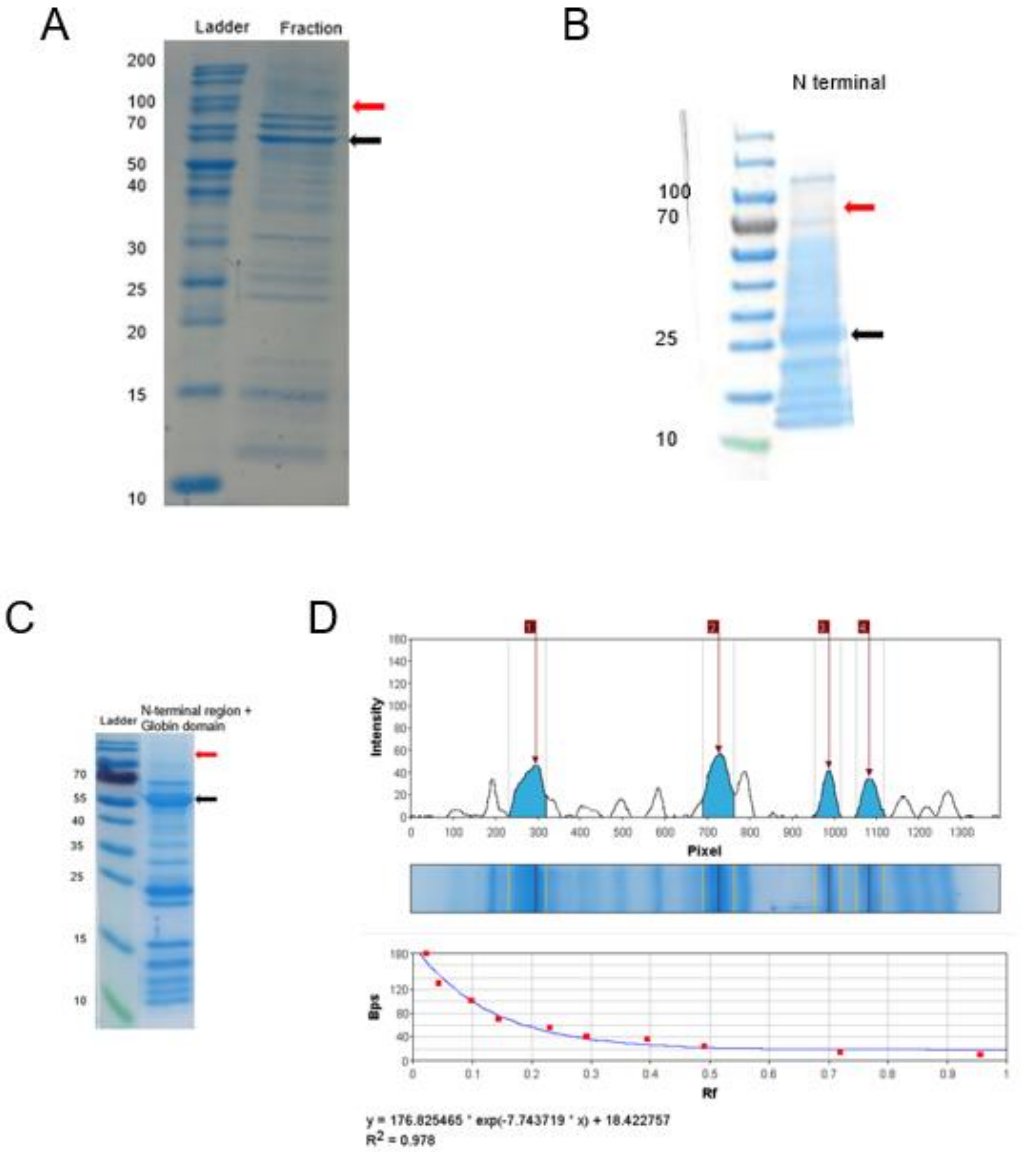


Figure 3.3.3.2. SDS-PAGE analysis of Adgb N-terminal region and N-terminal + globin domain under different conditions.

Resolving and stacking gels contained 15 % and 5 % acrylamide/bis-acrylamide, respectively. (A) SDS-PAGE of the purified N-terminal region with PMSF, a serine protease inhibitor. Multiple bands were observed, similar to those seen without PMSF (see Figure 3.3.3.1). (B) SDS-PAGE of the N-terminal region purified with urea, showing multiple bands with a main band at approximately 27 kDa (black arrow). The theoretical molecular weight (Mw) of the N-terminal region is 90.4 kDa (red arrow). (C) SDS-PAGE of the N-terminal + globin domain, displaying a similar migration pattern with multiple bands and a dominant band around 53 kDa (black arrow). (D) Densitometric analysis of (C) using *GelAnalyzer v23.1.1*, indicating the main band at ~53 kDa. The theoretical Mw of the His-tagged N-terminal + globin protein is 114.1 kDa (red arrow).

3.3.4 Mutagenesis Studies to identify the Nucleophilic Cysteine responsible for N-terminal autoproteolytic activity

The impact of these four cysteine sites (Cys132, Cys181, Cys206 and Cys406) on the autoproteolytic activity of the N-terminal domain were also assessed via site-directed mutagenesis. The purified C132G mutant protein was analysed using a 15% acrylamide gel (according to protocol as described in previous chapter section 2.2.9). Multiple bands, similar to the bands detected during the expression of the N-terminal region, were observed (Figure 3.3.4.1A). However, western blot shows fragments at ~47 kDa. This is essentially identical to the proteolytic fragment expressed by the WT N-terminal region protein above (Figure 3.3.3.1B). This is then shown in Figure 3.3.4.1B. This demonstrated that the mutation of C132G had no effect on the proteolytic activity of the N-terminal region. The results from the C132G mutagenesis experiment, as well as Cys181 and Cys206 cysteine sites being considerably more highly conserved, suggests that the cysteine residue at the catalytic triad could potentially be at these or other alternative positions.

The SDS-PAGE analysis of the C406G revealed similar results to the C132G mutant, displaying multiple bands similar to those observed during the expression of the N-terminal region and C132G mutant (Figure 3.3.4.1). It also produced similar proteolytic fragment at ~47 kDa in their western blot analysis (Figure 3.3.4.1B). This result aligns with the earlier conclusion in section 3.3.1, suggesting that this site is less likely to be part of the catalytic triad due to its low level of conservation.

Multiple bands were also observed in the SDS-PAGE analysis of C181G and C206G (Figure 3.3.4.1C) with marginal higher purity. Initially no distinct band was observed following western blot as depicted in Figure 3.3.4.1B. However, after increasing the antibody concentration from 1:5000 to 1:2000 dilution, multiple distinct fragments were observed in

both lanes of the blot (3.3.4.1B vs 3.3.4.1C). C181G appears to follow familiar pattern as observed in C132G and C406G with fragments observed at Mw (kDa) 43, 30 and 14. There was no evidence that this mutation expressed the full-length protein based on these results. Contrastingly, C206G mutation seemed to partially express the full-length protein as there was a band at ~87 kDa. The theoretical Mw of the N-terminal region according to Expasy server is 90.4 kDa. However, there were 3 distinct additional bands on this blot for C206G with each presumably corresponding to 3 fragments. There are further additional faint bands at 48 and 44 kDa, which corresponds to where the proteolytic fragments are normally seen in the WT N-terminal protein and the other mutants. The appearance of multiple bands in C206G indicate that this mutation does impact, but does not fully inhibit, the autoproteolytic activity of the protein as the other bands corresponds to other his-tagged fragments. Overall, these results are consistent with the earlier predictions based on the analysis of the cysteine conservation pattern which is Cys181 and Cys206 are more likely to serve as the alternative cysteine sites.

Figure 3.3.4.2 shows the gel analysis of the expression of various fragments following in-frame deletions (as shown in Figures 3.3.2.3) and purification. Unfortunately, no information about the proteolytic nature of the protein including the effects of removing more than one cysteine sites and the effect of removing a section of the proteolytic domain was obtained as depicted by the presence of multiple bands throughout the samples in the figure. However, supporting information about the nature of the protein was observed. The appearance of multiple bands indicate that the N-terminal and its variants may have high affinity of binding to endogenous *E.coli* proteins. Therefore, the denaturation agent urea was used to purify the proteins. Figure 3.3.4.3 shows the SDS-PAGE gel analysis of the N-terminal and its variants following purification with urea. A clear difference was observed when comparing the samples purified with urea to those purified without urea. A good example is the lane with fragment 6

(fig. 3.3.4.2 (#6) vs fig. 3.3.4.3 (urea fragment 6)) where the sample became significantly more pure following purification in the presence of urea.

Fragment 6 was chosen because it lacks three out of the four cysteines sites including Cys206. Therefore, on the premise that the expressed fragment would be observable in the absence of all predicted proteolytic cysteine nucleophiles, it was added as part of the protein to be purified with urea. No definitive conclusions could be drawn from the result, aside from the observation that purification in the presence of urea yielded a significantly purer sample. This process still requires further optimisation, as time limited the experiment development of this process. However, from the current data this process shows potential that denaturant agent like urea can decrease the affinity of N-terminal and its variant binding to *E.coli* protein. Additionally, a band around 42 kDa, corresponding to the theoretical weight of the full-length protein fragment (44.7 kDa) following the E395 mutation to a stop codon (E395U), was observed in the 'urea stop' lane in Figure 3.3.4.3A. However, the band appeared with low intensity, as illustrated in Figure 3.3.4.3B. A band at higher intensity was also observed at ~26 kDa on the same lane in Figure 3.3.4.3B. The appearance of this band at ~26 kDa therefore supports that autoproteolytic activity occur within the first ~400 amino acid which could be the proteolytic region.

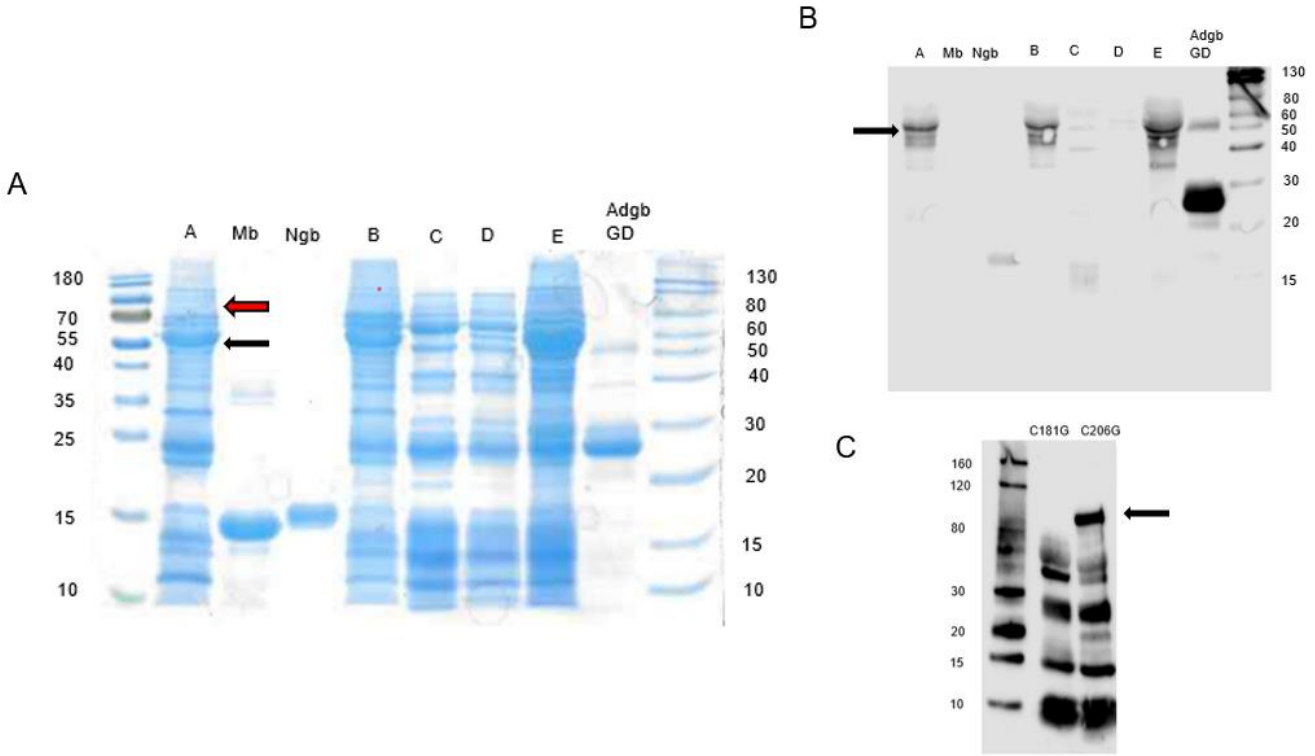
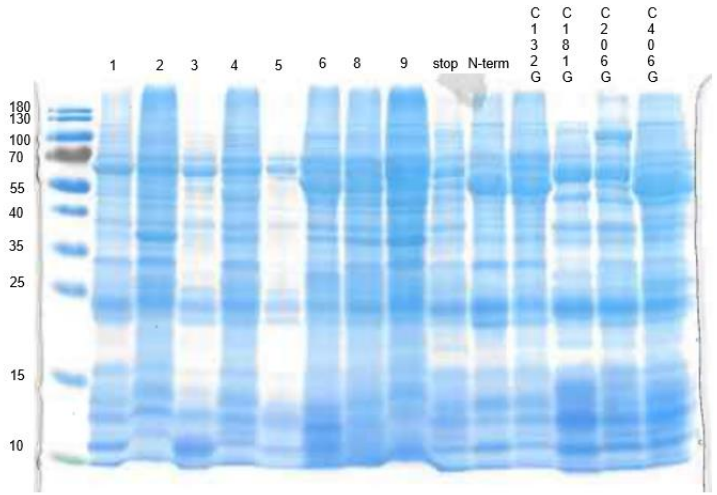


Figure 3.3.4.1. SDS-PAGE analysis of the WT Adgb N-terminal domain and mutants C132G, C181G, C206G, and C406G.

(A) Lanes: A = N-terminal, B = C132G, C = C181G, D = C206G, E = C406G. Control proteins Mb (~16.9 kDa) and Ngb (16 kDa) showed bands at ~14 kDa and ~15.3 respectively. Similar-sized bands were seen for N-terminal Adgb, C132G, and C406G. (B) Western blot detected a ~47 kDa fragment in C132G and C406G, matching the proteolytic fragment seen in the WT (Figure 3.3.3.1). Faint bands appeared for Ngb, C181G, and C206G, while Mb (no histag, negative control) showed no signal, ruling out non-specific binding. (C) C181G and C206G mutants was analysed again after increasing the antibody concentration to 1:2000. Three distinct bands were observed in C181G according to gel analyzer (appendix A3). They have Mw (kDa) of 43, 30 and 14. The fourth band may be a smear. In contrast, there are 5 bands (including 4 distinct bands) in C206G with Mw (kDa) of 87, 48, 42, 28 and 14. From this result it can be concluded that C206G mutation partially expressed the full-length N-terminal protein. They were all run alongside ThermoFisher BenchMark™ His-tag ladder for comparisons.



Fragment	Primers combination	Theoretical molecular weight (kDa)
1	Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F1	71..3
2	Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F2	64.2
3	Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F3	51.4
4	Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F1	79.1
5	Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F2	72.1
6	Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F3	59.2
7	Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F1	88.7
8	Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F2	81.7
9	Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F3	68.8
Stop E395U		44.7
N-terminal region WT		88.4
N-terminal region WT + GD		112.1

Figure 3.3.4.2. SDS-PAGE analysis of the different in-frame deletion from fragments 1 – 9 including the mutation with the stop codon as described in section 3.2.3.2 and shown in Figure 3.3.2.3 – 3.3.2.4.

Distinct bands corresponding to various molecular weights were compared to known molecular weights i.e. the ladder. Multiple bands characterised by low purity were observed in all the lanes which is different from the expected results. Bands at Mw (kDa) 1 = 71.3, 2 = 64.2, 3 = 51.4, 4 = 79.1, 5 = 72.1, 6 = 59.2, 7 = 88.7 (not shown), 8 = 81.7, 9 = 68.8 and stop E395U = 44.7 were expected. However, these results support the hypothesis that N-terminal binds to endogenous *E.coli* protein.

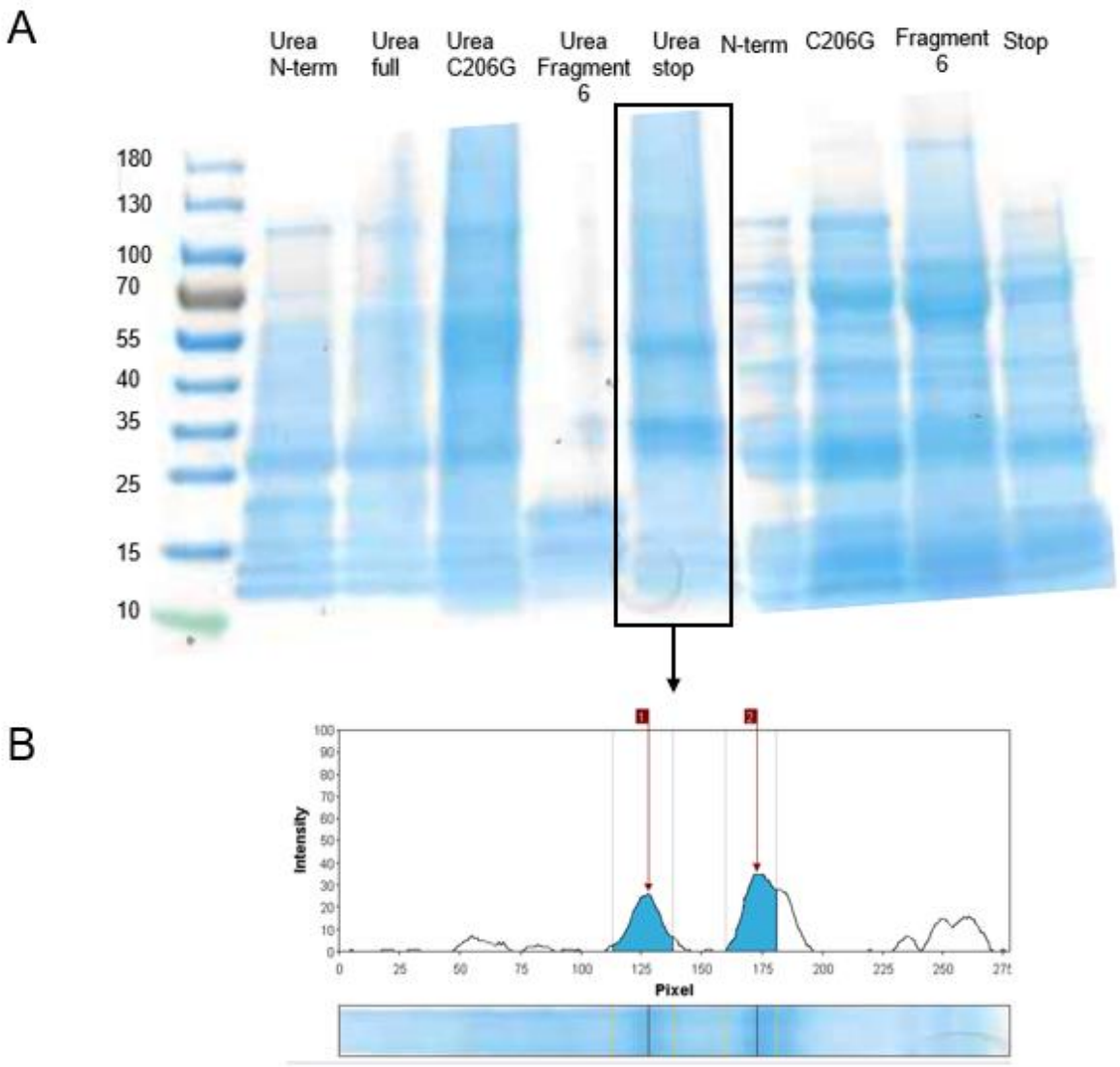


Figure 3.3.4.3. SDS-PAGE analysis of different N-terminal variants following purification with urea.

(A) The urea components are compared to the proteins purified without the denaturing agent (on the right). Clearer bands were observed in the fractions purified with urea compared to the fraction purified without urea. No further conclusions can be drawn at this stage, as these data are preliminary. The process still requires optimisation; however, these results suggest an alternative way of purifying the N-terminal region. (B) The gel analysis reveals two bands at 42 and 26 kDa in the urea-purified sample of stop codon E395U. The appearance of a band at ~42 indicates the presence of the full-length of the expressed protein with the stop codon E395U however the presence of another band at ~26 suggests the presence of a lysed fragment.

3.4. Discussion

In the absence of crystal structures of the protein or high-quality structural models, including current AlphaFold models, site directed mutagenesis would be the method of choice to target and confirm active site residues, specifically the cysteine nucleophile. This method has previously been used to identify the nucleophile of the proteolytic active site in human calpain 1 by substituting cysteine 115 with alanine (C115A) (Fernández-Montalván *et al*., 2004). This study aimed to determine the position of the active site nucleophile cysteine through inactivating the protease function of the calpain domain. The experiment in this chapter attempted to emulate the study by Fernández-Montalván *et al*. 2004, to investigate the significance of the cysteines in the proteolytic region of Adgb N-terminal domain by substituting the cysteines with glycine. Glycines, like alanines, are not nucleophiles because they lack side -OH or -SH chains with reactive functional group capable of donating an electron pair to an electrophile (Otto and Schirmeister, 1997). The C115A mutation in human calpain-1 inactivated the proteolytic activity without significantly altering the overall structure of the enzyme (Fernández-Montalván *et al*., 2004). The mutation disrupted electrostatic interactions (e.g., salt bridges) at the domain IIb–III interface, demonstrating that cysteine 115 is essential for the catalytic mechanism and highlighting its critical role within the catalytic triad.

The results presented in this chapter suggests that the recombinant N-terminal Adgb protein is proteolytically digested, but not by other serine-based proteolytic enzymes from *E.coli* as the serine protease inhibitor, PMSF, did not inhibit proteolysis (section 3.3.3). This supports the previous study suggesting that the protein is autoproteolytic (Bracke *et al*., 2018). Furthermore, the purification under denaturing conditions suggests that other peptides, be it from the recombinant Adgb or from endogenous *E.coli* proteins, co-elute with the fragment (Figure 3.3.3.2B). Further studies should incorporate surfactants like SDS and denaturing agents such as urea to enhance the purification and analysis of Adgb N-terminal proteins.

The proposed calpain domain of Adgb N-terminal was proposed to be closely related to human calpain 7 (Hoogewijs., 2012). Similar to Adgb, the function of mammalian calpain 7 is still largely unknown (Maemoto *et al*., 2013). The active site of calpain-7, like other members of the calpain family, is highly dependent on its tertiary structure for both activity and substrate recognition (Osako *et al*., 2010). In human calpain-7, for example, the active site comprises a catalytic triad consisting of cysteine, histidine, and asparagine (Bracke *et al*., 2018), which must be precisely positioned within the 3D structure to enable catalysis. This precise alignment can only occur when the protein folds into its correct tertiary conformation, making it challenging to predict cleavage sites based solely on sequence data. Given the limited tertiary structure information available for Adgb (e.g., AlphaFold predictions Figure 3.1.1), past and current identification of its active site has relied primarily on sequence alignments.

Unlike many proteases, calpains lack a well-defined consensus cleavage motif and their substrate recognition mechanisms remain poorly understood (duVerle *et al*., 2019; Gal *et al*., 2025). While PEST motifs (regions enriched in proline, glutamic acid, serine, and threonine) contribute to recognition for some substrates (Wang *et al*., 2003), numerous studies show they are not required in many cases (Molinari *et al*., 1995; Carillo *et al*., 1996). Instead, recognition relies on weak interactions between an atom in the peptide bonds of a substrate and an atom of calpains' subsite residues (Shinkai-Ouchi *et al*., 2016).

This weak specificity may stem from calpain biological function: they cleave diverse substrates across many cellular processes, suggesting their binding sites evolved to recognize broad peptide sequences rather than specific residues like in trypsin or caspases (Craik *et al*., 1985; Demon *et al*., 2009; duVerle *et al*., 2019). Consequently, predicting cleavage sites in calpains requires complex combinatorial analysis of sequences, as experimental data show cleavage across varied sequences without strong position-specific residue preferences (duVerle *et al*., 2019). CalCleaveMKL, a machine learning-based prediction tool was developed to

address this challenge but it predictive accuracy remains limited and does not reach complete reliability (Sorimachi *et al*., 2012).

Previous site directed mutagenesis studies have indicated that Cys290 of calpain 7 plays a critical role in calpain 7 auto-proteolytic activity (Osako *et al*., 2010). The presence of this Cys resulted to ~45 kDa fragments being observed during western blot analysis due to the autoproteolytic nature of the protein (Figure 3.4.1). The mutation of Cys290S altered the Mw of this fragment to ~97 kDa (Figure 3.4.1). This result suggested that some calpain 7 proteases contain more than one active site and one of them lies at the N-terminal boundary of the protease domain which is responsible for the ~45 kDa fragment (Osako *et al*., 2010). The result from the Cys290S mutagenesis experiment in calpain 7 offers a plausible explanation for the western blot findings of Adgb WT (Figure 3.3.3.1B) and its variant constructs (C132G, C181G, C206G and C406G) in Figure 3.3.4.1. The presence of multiple bands may be due to multiple active sites within the N-terminal domain, assuming it has been confirmed that the bands are not the result of co-elution with endogenous *E. coli* proteins.

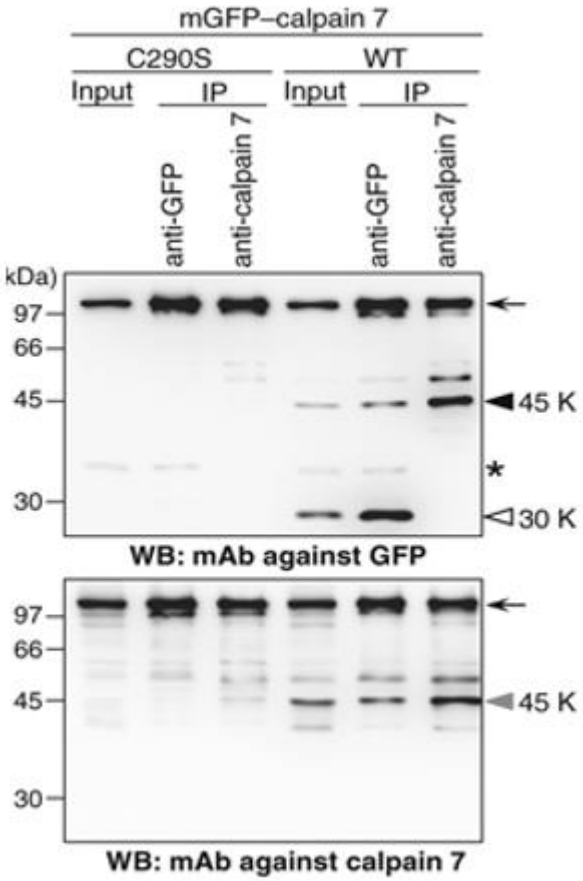


Figure 3.4.1. Image from Osako *et al*. (2010). This figure demonstrates that the Cys290S mutation effectively halts the auto-proteolytic activity of mGFP-calpain 7 fusion protein.

The calpain is a fusion of mGFP-Calpain and antibody. Cleared lysates from cells expressing mGFP–calpain 7 or mGFP–calpain7 C290S were subjected to immunoprecipitation (IP) with anti-GFP serum or polyclonal antibody against calpain7, followed by western blot analysis with monoclonal antibodies against GFP and against calpain7, respectively.

Mutating the cysteine residues Cys132 and Cys406 did not prevent the auto-proteolytic activity of the N-terminal domain of Adgb protein as observed in Figure 3.3.4.1, unlike the results described in the study by Osako *et al*., 2010 (Figure 3.4.1). The results from these mutagenesis studies therefore suggest that Cys mutation at position Cys132 and Cys406 have no effect on the proteolytic activity of the N-terminal region. This is because successful targeting of the cysteine active site nucleophile should have prevented the autocatalytic activity of Adgb N-terminal region and therefore the expected band on the SDS-PAGE gel should be the expected Mw of the full-size N-terminal region, which is ~90.4 kDa. This data casts doubt on the previous proposal that the active site cysteine was C132 (Hoogewijs *et al*., 2012). Additionally, the SDS-PAGE gel analysis continues to display multiple bands.

The revised sequence conservation analysis, based on a larger dataset, strongly suggested that Cys181 was the most likely candidate for the cysteine in the catalytic triad, as it is highly conserved at >99%, compared to Cys132, which is only 90% conserved across species (Figure 3.3.1.2). However, the mutagenesis results did not support this hypothesis. Only the mutation at C206G seemed to affect the autolytic activity of the protein (Figure 3.3.4.1C), potentially supporting a role in the catalytic activity. However, this mutation may have led to an altered the protein structure, resulting in a decreased proteolytic activity. This could be related to the previously proposed multiple active site hypothesis, as Cys181 and Cys206 are highly conserved and could both potentially function as active sites. Furthermore, depending on their proximity within the active site, one cysteine might compensate for the other in the event of a mutation. This suggests that removing one cysteine may not completely inhibit the autoproteolytic activity. The appearance of multiple bands on their blot (Figure 3.3.4.1) instead of a single band supports this theory. Therefore, the experiments with the protein expression following the deletions and further truncation of the Adgb through introduction of a stop codon were used to investigate this possibility further. This experiment was meant to help characterize

the proteolytically active region within the N-terminal domain and provide clearer insight into the location of the active site(s), if not identify them directly.

It is worth noting that although antibody sensitivity may explain the lack of observable bands for C181 and C206 in the initial western blot analysis, this is unlikely to be the case for the WT N-terminal samples. At a higher antibody concentration, no full-length WT protein was detected; instead, it produced a cluttered blot with patterns similar to those in Figure 3.3.4.1. This suggests that the overarching issue across the experiment is not antibody sensitivity, but the absence of a clean experimental readout of proteolysis.

Deletions have previously been used to study the intermolecular proteolytic activity of calpain 7 (Maemoto *et al*., 2013). In this approach, a C-terminally truncated form of ALG-2-interacting protein X (ALIX), containing the Bro1 domain and part of the V domain, was used as a substrate for calpain 7. This also involved using three deletion mutants from the full-length ALIX protein: ΔPRR (154 amino acids deleted), ΔC (209 aa), and NT (445 aa), as shown in Figure 3.4.2. These mutants and the full length were then fused with bovine α_{s1}-casein or human calpastatin domain 1 (castD1) to create non-physiological substrates. Among the constructs tested, cleavage was only observed in the castD1-ΔC fusion, suggesting that calpain-7 has limited proteolytic activity and specific substrate preferences. In the Adgb study, similar in-frame deletions were successfully introduced into the N-terminal region. Unfortunately, due to time constraints, the intended objectives of this part of the study could not be fully achieved. The SDS-PAGE gel analysis of the expressed protein variants (fragments 1 – 9 and E395U) indicated the need for a more effective purification strategy as it appeared that endogenous *E.coli* proteins were co-eluting with the Adgb fragments.

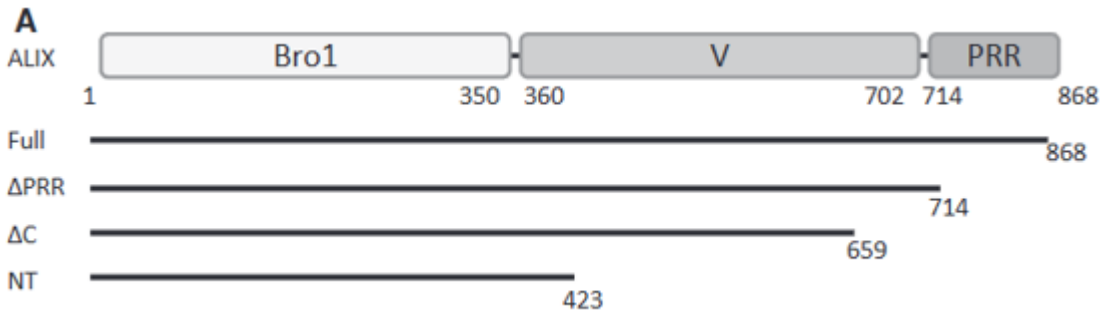


Figure 3.4.2, Image from Maemoto *et al*. (2013), illustrates the three deletion mutants derived from ALG-2-interacting protein X (ALIX).

ALIX comprises three key domains involved in membrane trafficking: the Bro1 domain, which interacts with CHMP4 proteins; the V domain, which binds viral late domains and syntenin; and the proline-rich region (PRR), which associates with proteins such as TSG101, ALG-2, endophilins, and CIN85 (Maemoto *et al*. 2013).

Although the results presented in this chapter do not allow for a definitive conclusion regarding the precise location of the proteolytic active site(s), the data challenges the previous proposed Cys132 site as the primary active nucleophile. Furthermore, the data from the purification of the E395U (stop codon) construct using urea as a denaturant (Figure 3.3.3.7) suggest that the active region lies within the first 400 amino acid residues of Adgb. This region includes Cys206, the candidate cysteine identified in this study as potentially part of the catalytic triad responsible for autolysis in the N-terminal domain. This process still requires further optimization before we can gain in-depth understanding of the characteristic of this region. Additionally, we propose using SDS or guanidinium chloride for purification because it's cheaper, readily available and most importantly, this study showed that it works very well with Adgb. SDS can be a denaturant at high concentrations. However, the concentration of SDS needed to unfold the protein still needs further investigation. This can be done by circular dichroism to monitor the secondary structure or by PAGE to monitor efficiency of the purification.

To date, only one study (Hoogewijs *et al*., 2012) has proposed that Adgb may function as a calpain-like cysteine protease. However, the sequence alignment between Adgb and known calpains is limited and does not align well with the histidine and asparagine residues typically found in the catalytic triad. While it remains possible that Adgb is a serine rather than a cysteine protease, the lack of detailed structural information on the N-terminal domain makes it difficult to confidently identify potential active sites (with dozens of potential active sites). Although the use of PMSF during purification helped rule out the contribution of exogenous serine proteases to the fragmentation observed, it does not definitively exclude Adgb itself from being a serine protease, especially since PMSF was not present during expression in *E. coli*. This uncertainty further supported the rationale for employing deletion constructs to broadly explore possible active regions.

Future experiments could include designing and expressing the Adgb N-terminal domain containing multiple tags. For example, a His-tag at the beginning of the N-terminal region and a Strep-tag at the end of the N-terminal region. This multi-tag approach would allow for more complete detection and analysis of fragments that are typically lost during purification when using the His-tag alone. Preliminary design work on constructing this plasmid has been carried out using Golden gate PCR, in which the His-tag Adgb gene from *E.coli* was transferred into *Shewanella oneidensis* MR-1 (a kind donation from Dr Marcus Edwards, University of Essex), containing a strep-tag. Alternatively, immunoprecipitation techniques using polyclonal antibodies could be used similar to those in the calpain 7 study by Osako *et al*. (2010) instead of His or Strep tags. Additionally, CalCleaveMKL could have served as a useful tool for predicting cleavage sites in the N-terminal region, although this approach is not entirely reliable and its accuracy is not 100 %.

CHAPTER 4: Kinetic studies on Androglobin Globin Domain

4.1. Introduction

The globin domain of androglobin (Adgb-GD) is reported to have a hexacoordinated heme geometry (Hoogewijs *et al*., 2012). The expression of Adgb-GD in isolation shows that a disulfide bond takes the place of the CD loop, supporting the heme pocket structure and stabilising the globular domain (Reeder *et al*., 2024). The recent stable expression of the globin domain has allowed the characterisation of the protein in regard to ligand binding and other kinetics studies. Spectroscopic studies have been used to characterise the ligand binding kinetics of heme proteins and demonstrate their effect on the coordination state of haem proteins. The research into the redox activities of globular proteins via spectroscopy have presented insights into the diversity and complexity of globin-associated redox chemistry and hence potential physiological roles (Reeder, 2017). For example, it has been previously demonstrated that lipid binding to Cygb induce a change in the heme coordination of the protein thus transforming the heme iron from a hexacoordinate to pentacoordinate state (Reeder *et al*., 2011). Furthermore, it was observed that this transformation occurs only under the ferric oxidation state and may be associated with redox-linked cell signalling mechanisms, potentially via the high lipid peroxidase activity that Cygb exhibits (Reeder *et al*., 2011). This lipid-induced transformation of Cygb iron coordination is a distinctive feature of this globular protein and may be related to its physiological function as similar conditions do not induce the same transformation in Ngb, another hexacoordinate globin (Reeder *et al*., 2011). This behaviour offered insights into the redox chemistry of Cygb and Ngb, which served as comparative models for investigating lipid peroxidase via model cell oxidation (liposomes) in Adgb-GD in this chapter.

In addition to examining lipid peroxidase activity, this chapter will be exploring the ligand binding kinetics of Adgb-GD with CO, CN, and NO, as well as investigating the nitrite reductase activity of Adgb-GD. A series of rationally designed mutants will also be utilised to

explore Adgb-GD redox activity and molecular function. Currently, the suggested physiological function of Adgb includes ciliogenesis and spermatogenesis-associated role in mammals. However, the true molecular action of Adgb remains unknown. Investigating the redox chemistry of Adgb-GD *in vitro* with various ligands will provide insights into its dynamics and reactivity, explaining some of its potential molecular roles and the mechanisms regulating its redox function.

4.2. Methodology

4.2.1. Recombinant expression of globin proteins

The materials used to produce Adgb-GD and other globin proteins in this study were previously outlined in chapter 2, section 2.2.1. The procedures for protein expression and purification were also previously described in chapter sections 2.2.2 and 2.2.3.

4.2.2 Electron paramagnetic resonance spectroscopy.

Ferric monomeric WT Adgb-GD, the disulfide reduced Adgb-GD WT protein (using TCEP) and the four Adgb-GD mutant protein were diluted in 0.1 M NaPi buffer at pH 6, pH 7 and pH 8, all to a final concentration of 80 μM, before being transferred to Electron Paramagnetic Resonance (EPR) tubes (Wilmad SQ Glass). Tubes were flash-frozen in methanol cooled on dry-ice before being transferred to liquid nitrogen (77 K) for storage. The TCEP reduced Adgb-GD WT protein was prepared by incubating the protein in 1 mM TCEP for 1 h then the protein was dialysed (8 kDa MWCO dialysis tubing) overnight in 1 mM Borax to remove the TCEP. The EPR spectra were acquired using a Bruker EMX X-band spectrometer, equipped with a high-quality spherical ER 4122 SP9703 resonator and an Oxford Instruments liquid helium cooling system. Measurements were conducted with a modulation frequency of 100 kHz. Precise g-factor values were determined utilising the instrument's built-in microwave frequency counter and a 2,2-diphenyl-1-picrylhydrazyl (DPPH) powder standard, which has a known g value of 2.0027 ± 0.0002 (Nugent, 1994). The quantitation of the heme forms of Adgb-GD i.e. high-signal (HS) and low-signal (LS) were studied using EPR spectra (Svistunenko *et al*., 2000).

4.2.3 Measurement of carbon monoxide binding kinetics to heme iron.

The rapid binding kinetics of CO to Mb, Ngb and Adgb-GD WT proteins were studied by monitoring the optical spectra (between 350 - 700 nm) generated using an Applied Photophysics SX20 stopped-flow UV-visible spectrophotometer. The protein was diluted to a final concentration of 5 μM in 0.1 M NaPi pH 7.4 and mixed with few grains of sodium dithionite. CO-saturated stock solution was prepared by degassing 0.1M sodium phosphate pH 7.4 buffer using a glass tonometer connected to a supply of argon gas. The buffer was cycled between vacuum and argon several times to remove oxygen. The deoxygenated buffer was then saturated with or without CO and transferred anaerobically to a 10 mL glass syringe before addition of a few grains of sodium dithionite at 20°C. The stopped-flow apparatus included a rapid mixing system to combine the ferrous protein with the deoxygenated CO solution. The reaction was initiated by rapidly mixing the two solutions in the stopped-flow chamber in a 1:1 volume ratio. Therefore, the final concentration of the protein was 2.5 μM and the CO was 25 μM – 200 μM after mixing. The data were fitted to a double or quadruple exponential equation using Pro-Data and ProKineticist IV software by Applied Photophysics. The fits were performed after reducing the data by singular value decomposition to remove high frequency noise. The processed data were then globally fitted to their mechanistic models to determine rate constants.

4.2.4 Measurement of cyanide binding kinetics to heme iron.

The optical spectra of cyanide (CN) binding kinetics to Mb, Ngb and Adgb-GD WT was monitored using a UV-Vis detector set to the absorbance 350 - 700 nm using an Applied Photophysics SX20 stopped-flow UV-visible spectrophotometer (20 °C). The ferrous proteins were prepared similarly to the CO binding experiment (section 4.2.3). 1 M sodium CN stock

was prepared in 10 mM sodium hydroxide solution and diluted in buffer 0.1 M NaPi pH 7.4. The stopped-flow rapidly combined the ferrous protein with the various concentrations of CN to reach final concentration of 5 μM for the protein and 1 – 20 mM in 0.1 M NaPi pH 7.4 buffer after mixing. These data were fitted to a triple exponential equation using Pro-Data and ProKineticist IV software by Applied Photophysics. This fit was achieved by reducing the data by singular value decomposition and globally fitted to determine rate constants.

4.2.5 Nitrite reductase activity measurement

An Agilent 8453 diode array spectrophotometer was used to measure the nitrite reductase (NiR) activity of Mb, Ngb and Adgb-GD WT. The final concentration of the ferrous protein used was 5 μM. The ferrous protein was thoroughly mixed with $NaNO_2$ by pipetting in a 1 ml quartz cuvette. The NiR activity of Mb, Ngb and Adgb-GD spectra were monitored (380 - 750 nm) and two wavelengths, 419 nm and 438 nm, were used for the kinetic analysis for Mb, whereas 413 nm and 426 nm were used for Ngb, and 412 nm and 438 nm for Adgb-GD respectively. NiR activity of the different globins were measured over a range of sodium nitrite concentrations (0 mM – 2 mM).

4.2.6 NO binding kinetics to heme iron

NO solution was prepared by diluting proli-NONOate, an NO donor, in a degassed buffer solution (NaPi 0.1 M pH 7.4). The proli-NONOate was previously dissolved stored under nitrogen or argon at -70 °C as a 40 mM stock solution in 10 mM NaOH. Before use, the concentration of Proli-NONOate (dissolved in NaOH) was determined using an extinction coefficient of 8500 M^{-1} cm^{-1} at 250 nm. When diluted in the pH 7.4 buffer, the NO rapidly dissociates from the proli-NONOate, with a theoretical release of two NO molecules per proli-

NONOate NO donor. However, experimentally this was previously determined to be closer to 1.8:1 (Simons *et al*., 2018). The final concentration of the ferrous Adgb-GD and Mb were 10 μM, 5 μM after mixing in the stopped flow apparatus (20 °C). The published second-order rate constant for NO binding to Mb, 1.7 x 10^7 M^{-1} s^{-1} (Moore and Gibson, 1976) was used to correct NO concentrations, resulting from NO depletion due to residual oxygen remaining in the system after degassing, typically <10 μM.

4.2.7 Liposome oxidation

An Agilent diode 8453 array spectrophotometer was used to measure liposome oxidation by various globins. Ferric Mb, Cygb (homodimer), Ngb, Adgb-GD WT and their derivatives were used at a final concentration of 2.5 μM in 0.1 M sodium phosphate buffer (pH 7.4). The heme proteins were incubated with multilamellar liposomes (100 μg/μl). The oxidation of the liposomes was monitored optically by the formation of lipid-based conjugated dienes at 234 nm. The liposome was prepared by dissolving lecithin (5 mg/ml) in 0.1 M sodium phosphate buffer pH 7.4 and then sonicated in a sonicating water bath for up to 10 minutes until no particulates can be seen and the solution appears pearlescent. Following addition with hemoproteins, changes in conjugated diene directly relating to the oxidation of the liposomes (Reeder and Wilson, 1998), were monitored at 234 nm (ε_{234nm}=25000 M^{-1} cm^{-1}, 11000 M^{-1} cm^{-1} using the three-point baseline drop using 220, 234 and 250 nm).

4.2.8 Visualisation

Plots of the full optical transitions following ligand binding were visualised using RStudio ggplot2 (version 2022.12.0-353. exe; RStudio Team, 2020).

4.3 Results

4.3.1 EPR

The X-band EPR spectra of ferric WT Adgb-GD, Adgb-GD proximal and distal mutants, and the disulfide reduced Adgb-GD protein (with TCEP) at pH 6 - 8 values are shown in Figure 4.3.1.1 - 3. Overall, the paramagnetic heme iron EPR spectra exhibit a pattern characteristic of mixture of high-spin (S = 5/2, HS) and low-spin (S=1/2, LS) signals, typical of that of many globins (Svistunenko, 2021).

The HS components, typically associated with pentacoordinate heme iron configuration, showed signals at g ≈ 5.95 (two perpendicular components) and g ≈ 2 (parallel) in all Adgb-GD protein forms across different pH values (Figures 4.3.1.1 - 3). Also, low-spin signals associated with hexacoordinate heme iron were observed at g = 2.95 and g = 2.26, corresponding to the g_x and g_y components in all the Adgb-GD protein forms across different pH values (Figures 4.3.1.1 - 3). The third g_z component expected at $g \leq 2$ was however not observed in the WT, proximal, and distal mutants as it was too broad, but this is not untypical (Svistunenko *et al*., 2007). Although in the disulfide reduced protein (with TCEP), there is some evidence for a minor second LS signal at g = 2.43 as shown in Figure 4.3.1.1 - 3.

The WT HS and LS spectra exhibit essentially the same shape at different pH values as demonstrated in figure 4.3.1.4. This shows that pH affects how strong the signal is, but not the overall shape of the spectra. Since the EPR line shape reflects the local environment of the heme, the conservation of line shape suggests that the overall heme structure remains essentially the same within the pH range measured. Therefore, the pH-dependent intensity changes seen for the WT in Figures 4.3.1.1 - 3 are more likely due to minor changes within the heme pocket, such as protonation of nearby amino aid side chains, which affects the heme-protein coordination.

For the WT, a strong HS signal at g = 5.95 was observed at acidic pH (pH 6), which decreased by over 70% at pH 7, followed by a slight increase at pH 8. In contrast, the LS signal at g = 2.26 showed a slight increase from pH 6 to pH 7, before decreasing again at pH 8. These trends are more clearly illustrated in the bar chart showing changes in relative EPR signal intensities across the pH range (Figure 4.3.1.5), demonstrating that pH influences the relative populations of the HS and LS species while preserving the overall heme environment. These results indicate that while the HS signal is consistently present across all pH conditions, its signal intensity at g = 5.95 is pH-dependent, reaching its lowest intensity at neutral pH, which is close to the normal physiological cytosolic (pH ~7.4). The majority of biochemical experiments in this study were conducted at pH 7.4.

To probe this pH-dependent behaviour, mutations in residues surrounding the heme pocket were examined. Two mutants of the heme iron distal amino acid were chosen, mutating the WT distal glutamine to a glycine (Q792G) and histidine (Q792H) variants. The rationale for these mutations is that the Q792G should be more HS compared to WT due to the lack of side chain to coordinate to the heme iron, and the Q792H perhaps more LS due to the more typical distal ligand for globins. Additionally, two proximal histidine mutants were also chosen, based on alternative heme-protein ligation in other heme proteins such as cysteine (H824C) or tyrosine (H824Y) often observed in cytochromes (Manole *et al*., 2015).

Overall, the EPR spectrum of the mutants follows a similar pattern to that of the WT protein, as shown in Figure 4.3.1.1 - 3. A similar pattern was expected for the distal mutant Q792H, since glutamine and histidine are similar in size and both have neutral side chains under near neutral to alkaline pH conditions (histidine side chain pKa ≈ 6.04). However, it was surprising that the Q792G mutant showed a pattern comparable to the WT, despite lacking a side chain. In summary, the distal mutants showed a stronger HS signal at g = 5.95 under acidic pH conditions compared to the proximal mutants, whereas the proximal mutants exhibited a

stronger HS signal at g = 5.95 under alkaline pH. Overall, the HS signal is consistently present across all pH conditions for both proximal and distal mutants.

LS signals are also consistently observed across all pH conditions for both proximal and distal mutants and are more prominent than those of the WT, as illustrated in Figure 4.3.1.5. As in the WT, the signal at g = 2.26 is strongest at pH 7 for all mutants relative to acidic and alkaline conditions.

The EPR spectrum of the disulfide reduced protein (with TCEP) shows some differences from the WT and mutants, but overall, closely resembles them, highlighting the importance of the C894/C970 cysteine pair. The reduction in the disulfide of the cysteine pair C894/C970 still resulted in a HS signal at g = 5.95 and g = 2, indicative of pentacoordinate heme proteins (Figure 4.3.1.1 - 3). However, the intensity of the high-spin signal at g = 5.95 is significantly lower across all pH conditions compared to the WT and mutant proteins. This does not mean that the HS species is lower overall, as the signal is broader. This suggests that the HS heme component is still relatively high, but that the conformation of the heme binding pocket compared to the unreduced protein is significantly different. Furthermore, the g = 5.95 HS signals appear to have equal intensity at all pH values observed, showing that the protein now lacks the pH effects seen in the WT and mutant proteins with oxidised cysteines (4.3.1.5).

The relative intensity of the LS signals also appears similar or slightly higher when compared to the Cys oxidised and mutant proteins, further indicating that the ratio of the LS to HS signal is relatively unchanged (Figure 4.3.1.5). Again, the LS signals do not appear to be significantly affected by pH.

In summary, the mutations have minimal impact on the EPR spectra, while the cysteine oxidation state has a more pronounced effect. Furthermore, although the overall spectra of WT

and mutants are similar, their relative intensities differ significantly, with distal mutants showing the most pronounced changes.

Additionally, the optical spectrum of ferric WT Adgb-GD and the mutants at room temperature (see Figure 2.3.3.1 and 2.3.3.2) aligns more closely with that of a hexacoordinate LS protein, with only some minor features suggesting a small subpopulation of pentacoordinate HS (in the WT). This discrepancy between optical and EPR spectra was also observed for ferric Cygb but was deemed a freezing artifact (Beckerson *et al*., 2014). Moreover, there is an additional very small signal at g = 4.3, corresponding to the presence of non-heme ferric ions in a rhombic coordination (Svistunenko *et al*., 2007). This is a typical small component in native and recombinant proteins (Svistunenko *et al*., 2007).

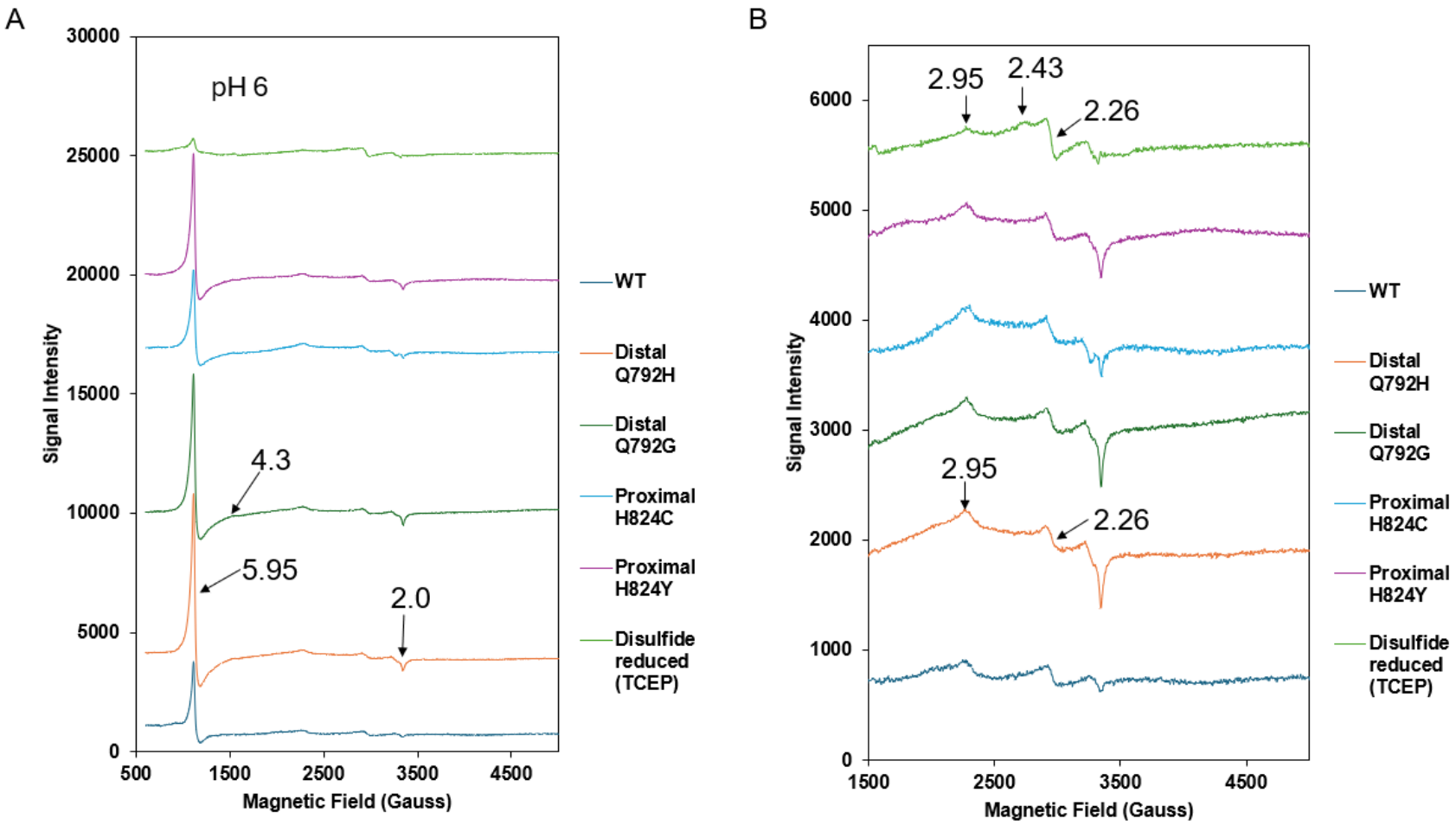


Figure 4.3.1.1. The EPR spectra of 80 µM ferric Adgb-GD forms and 50 µM TCEP-reduced Adgb-GD at pH 6.

The spectra were recorded at 10 K, 3.16 mW microwave power and 5 G modulation amplitude. (A) Shows both high spin (g=5.95, 2.0) and low spin (g=2.95, 2.26) Fe^{3+} EPR signals at pH 6. (B) Expanded view of the low spin signals of (A). All spectra are offset for clarity.

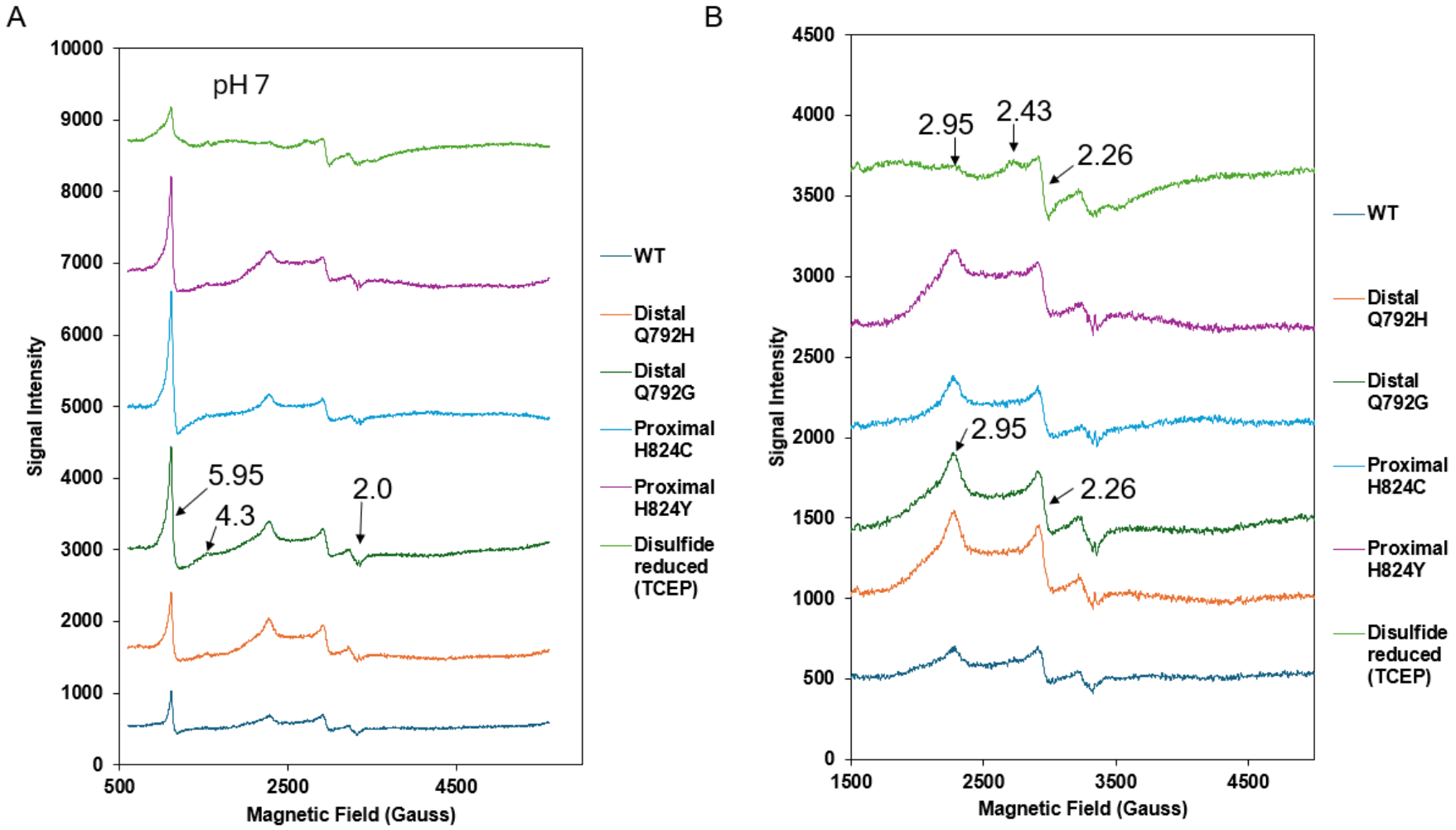


Figure 4.3.1.2. The EPR spectra of 80 µM ferric Adgb-GD forms and 50 µM TCEP-reduced Adgb-GD at pH 7.

The spectra were recorded at 10 K, 3.16 mW microwave power and 5 G modulation amplitude. (A) Shows both high spin (g=5.95, 2.0) and low spin (g=2.95, 2.26) Fe^{3+} EPR signals at pH 7. (B) Expanded view of the low spin signals of (A). All spectra are offset for clarity.

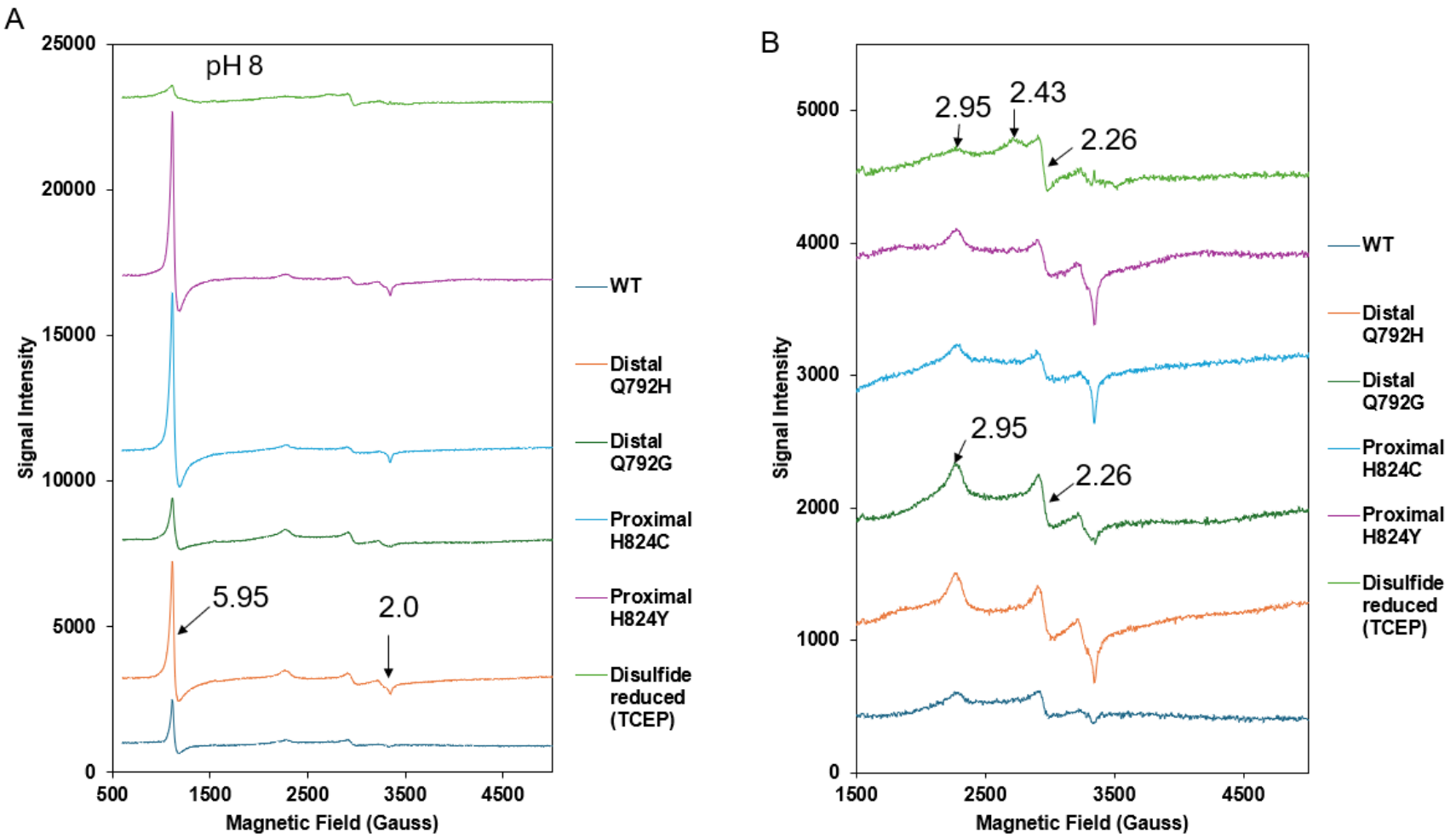


Figure 4.3.1.3. The EPR spectra of 80 µM ferric Adgb-GD forms and 50 µM TCEP-reduced Adgb-GD at pH 8.

The spectra were recorded at 10 K, 3.16 mW microwave power and 5 G modulation amplitude. (A) Shows both high spin (g=5.95, 2.0) and low spin (g=2.95, 2.26) Fe^{3+} EPR signals at pH 8. (B) Expanded view of the low spin signals of (A). All spectra are offset for clarity.

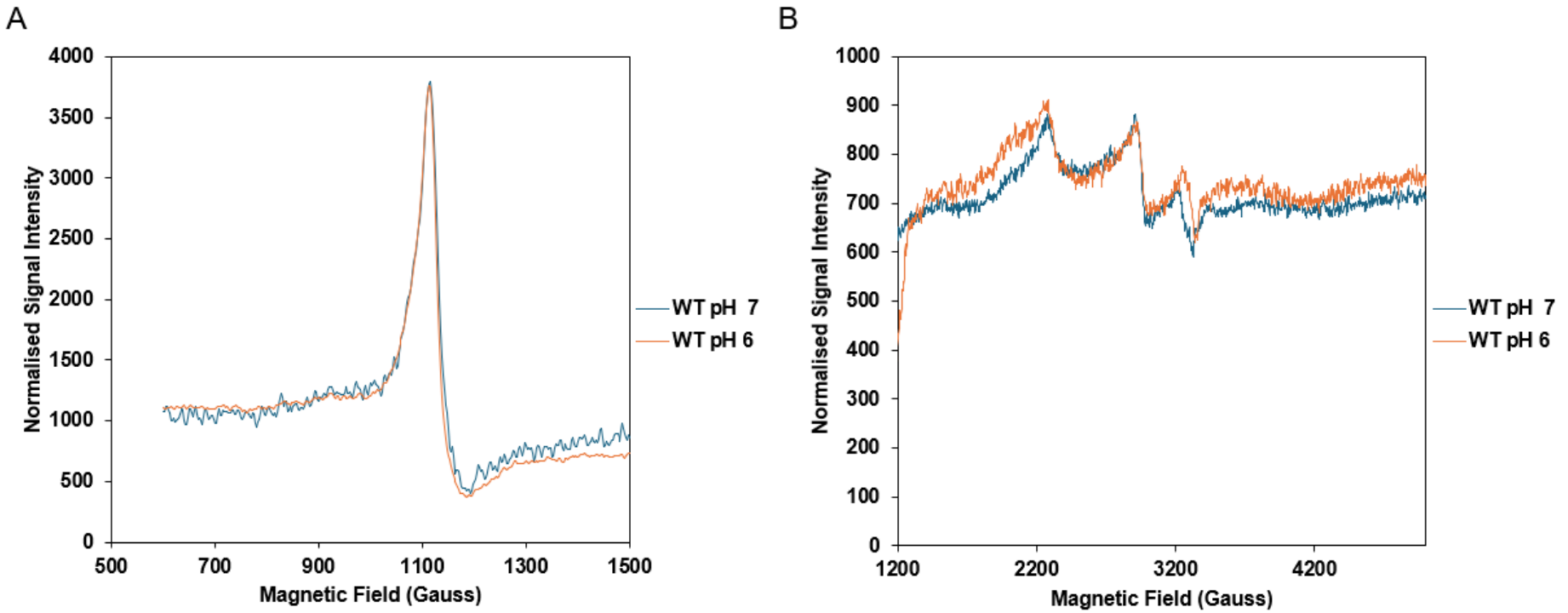


Figure 4.3.1.4. Overlay comparison of high-spin (HS) and low-spin (LS) signal profiles of 80 µM ferric Adgb-GD WT under pH 6 and pH 7 conditions.

Where (A) HS WT comparison and (B) LS WT comparison. The WT HS and LS spectra have essentially the same shape, which means that the changes observed in their biochemistry cannot be attributed to a shape change but rather to differences in peak–trough intensity. The signal intensities have been normalised.

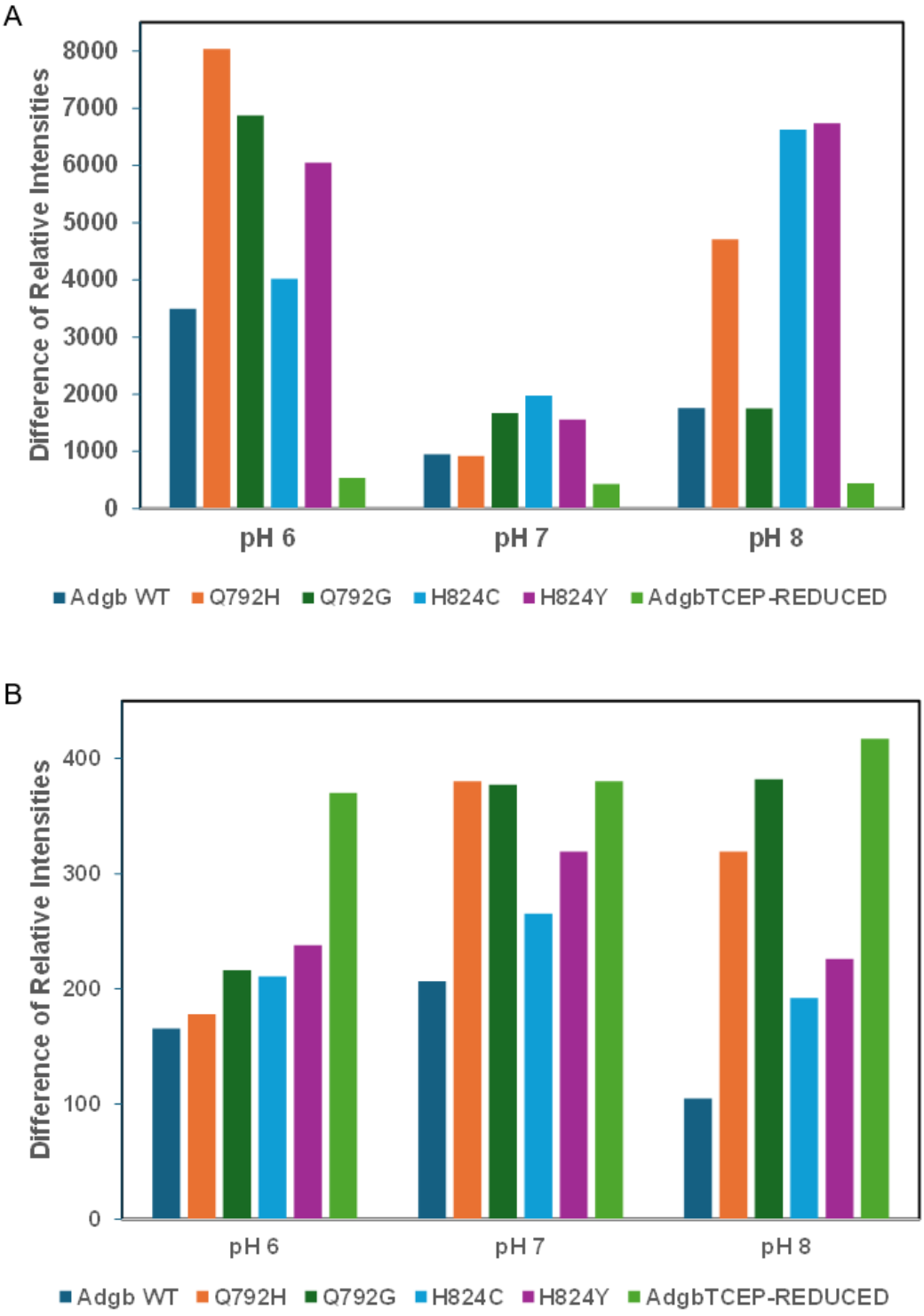


Figure 4.3.1.5. Bar chart showing the difference of the relative intensities (peak to trough) of the EPR signals for the HS and LS signal at different pH.

(A) High-spin signal (5.95) of the Adgb-GD WT, mutants and TCEP-reduced proteins at pH 6, pH 7 and pH 8. (B) Low-Spin signal (2.26), also in similar condition.

4.3.2 Kinetics of CO binding to androglobin WT vs TCEP reduced Androglobin globin domain

To probe the effect of the oxidation state of Adgb-GD cysteines on potential protein structure and function, the effects of the oxidation state of the C894/C970 disulfide on ligand binding was examined by stopped flow spectroscopy. On rapidly mixing ferrous Adgb-GD with solution containing known concentrations of CO in NaPi buffer pH 7.4, optical transitions were observed as shown in Figure 4.3.2.1A - B. The data shown in this figure is consistent with the previous CO spectra observed in Figure 2.3.3.1C. The kinetic traces fit to a double exponential at 434 nm and is exhibited in Figure 4.3.2.1C. The time course show that CO binding to Adgb-GD comprises at least two phases. However, as shown by comparison with Figure 4.3.2.1A, the final spectrum remained the same, otherwise the isosbestic point would not be so evident at 404 nm, 429 nm, 453 nm, 554 nm and 569 nm. On changing the CO concentration, as expected the kinetics changed, but the optical spectra transition remained the same, i.e. a ferrous to CO complex formed with no evidence of a distinct intermediate species.

Figure 4.3.2.2 presents the plot of the rate constant versus CO concentration, revealing a linear relationship between the rate constant for CO binding and CO concentration in both phases. The second-order rate constant for fast phase and slow phase are 5.99 x 10^5 M^{-1} s^{-1} and 6.75 x 10^4 M^{-1} s^{-1} respectively. It showed similar biphasic trend as Ngb with both a fast and slow phase (Hamdane *et al*., 2003; Uzan *et al*., 2004). The second-order rate constant for the fast phase was reported to be 5 × 10^7 M^{-1} s^{-1} (Uzan *et al*., 2004). Comparable values were observed for the fast phase for Mb, Ngb and WT Adgb-GD (Table 4.3.2.1).

The kinetics of CO binding to the disulfide reduced form of Adgb-GD (with TCEP) shows similar linear dependency as a function of CO concentration. There were also similar optical changes to that of the WT Adgb-GD spectra observed on rapidly mixing the disulfide-

reduced ferrous protein with CO. This is illustrated in Figure 4.3.2.3. These figures showed that CO binds to the TCEP reduced form of the protein forming two optical species, namely ferrous and the ferrous-CO complex formed after binding to CO. An example time course at 434 nm is shown in the same figure (Figure 4.3.2.3C).

These time courses showed that the kinetic comprises of 4 phases, all with the same spectrum as it can be supported by the appearance of isosbestic points on the plot absorbance vs wavelength shown in Figure 4.3.2.3A and 4.3.2.3B.

Additionally, no optical differences were observed during the reaction at either high or low concentrations CO concentration. All four phases showed a linear dependency as shown in Figure 4.3.2.4 with second-order rate constant of 1.53 x 10^5 M^{-1} s^{-1}, 2.39 x 10^4 M^{-1} s^{-1}, 6.3 x 10^3 M^{-1} s^{-1} and 8 x 10^2 M^{-1} s^{-1} for their k_1, k_2, k_3 and k_4 respectively. The second-order rate constant for the fast phase of the disulfide reduced protein is also comparable to the WT and other globins. The association rate for CO binding to the disulfide reduced protein is ~3.9 magnitudes slower when compared to the WT. This effect is likely attributable to the reduction of the disulfide bond.

Reduction of the disulfide bond in Ngb and Cygb generally leads to faster ligand binding and increased affinity, regardless of the ligand type, for example, oxygen (Hamdane *et al*., 2003). This redox-dependent enhancement in ligand reactivity is considered a hallmark feature of Ngb (Bocahut *et al*., 2012). In contrast, the absence of the internal disulfide bridge contributed to slower CO reactivity in Adgb. This may be due to the presence of a more stable distal glutamine iron coordination in Adgb, in the absence of an internal disulfide bridge. Similar effect was reported in Ngb binding to CN after the disulfide bond has been reduced (Bocahut *et al*., 2012).

To probe the structure and heterogeneity of the ferrous protein, the data was also analysed using global fitting. The global fitting for the disulfide reduced protein reaction with CO is shown in Figure 4.3.2.5. The global fitting analyses revealed that the reduced ferrous species (TCEP reduced) does not exist as a single conformation, but rather as a family of closely related forms. These conformers exhibit highly similar spectra, yet each display slight spectral variations and distinct second-order binding constants. A comparable global fit pattern was also observed for the WT protein (Appendix A4). The variability in second-order rate constants for the slow phases of the TCEP-reduced form further supports this hypothesis. Additionally, the absence of discernible intermediates in the spectral data reinforces the presence of multiple pre-existing conformations rather than sequential intermediates. In conclusion, the CO probe revealed the presence of multiple conformers, with their binding rates differing depending on the presence or absence of the disulfide bond.

Table 4.3.2.1 The table showing the ligand binding activity of various members of the androglobin family. All are human globins unless otherwise stated.

Ligand	Protein	k (M^{-1} s^{-1})				Temp	Ref
		k_1	k_2	k_3	k_4		
CO	Adgb-GD WT (434 nm)	5.99 x 10^5	6.75 x 10^4			20	This study
	Adgb-GD disulfide reduced (with TCEP) (434 nm)	1.53 x 10^5	2.39 x 10^4	6.3 x 10^3	8 x 10^2	20	This study
	Sperm whale Mb	5 x 10^5					Olson and Phillips. (1996)
	hNgb	5 x 10^7					Uzan *et al*. (2004)
CN	Adgb-GD WT	1.62 x 10^2	8.5			20	This study
	Mutant Adgb Q792G	1.95 x 10^2	4.2			20	This study
	Mutant Adgb H824C	2.2 x 10^2	14.1			20	This study
	Equine Mb	9.97 x 10^2				20	This study
	hNgb	3.8	0.681			20	This study
NiR activity	Adgb-GD WT	72.7				25	This study
	Mutant Adgb Q792G	60.2				25	This study
	Mutant Adgb Q792H	58.0				25	This study
	Mutant Adgb H824C	44.5				25	This study
	Mutant Adgb H824Y	81.2				25	This study
	Equine Mb	9.2				25	This study

	hNgb	2.27	0.32			25	This study
NO	Adgb-GD WT	1.39×10^6	1.32×10^4			20	This study
	Equine Mb	1.14×10^6	6.45×10^4			20	This study
	hNgb	21 ± 1	2.9 ± 0.4				Herold *et al*. (2004)

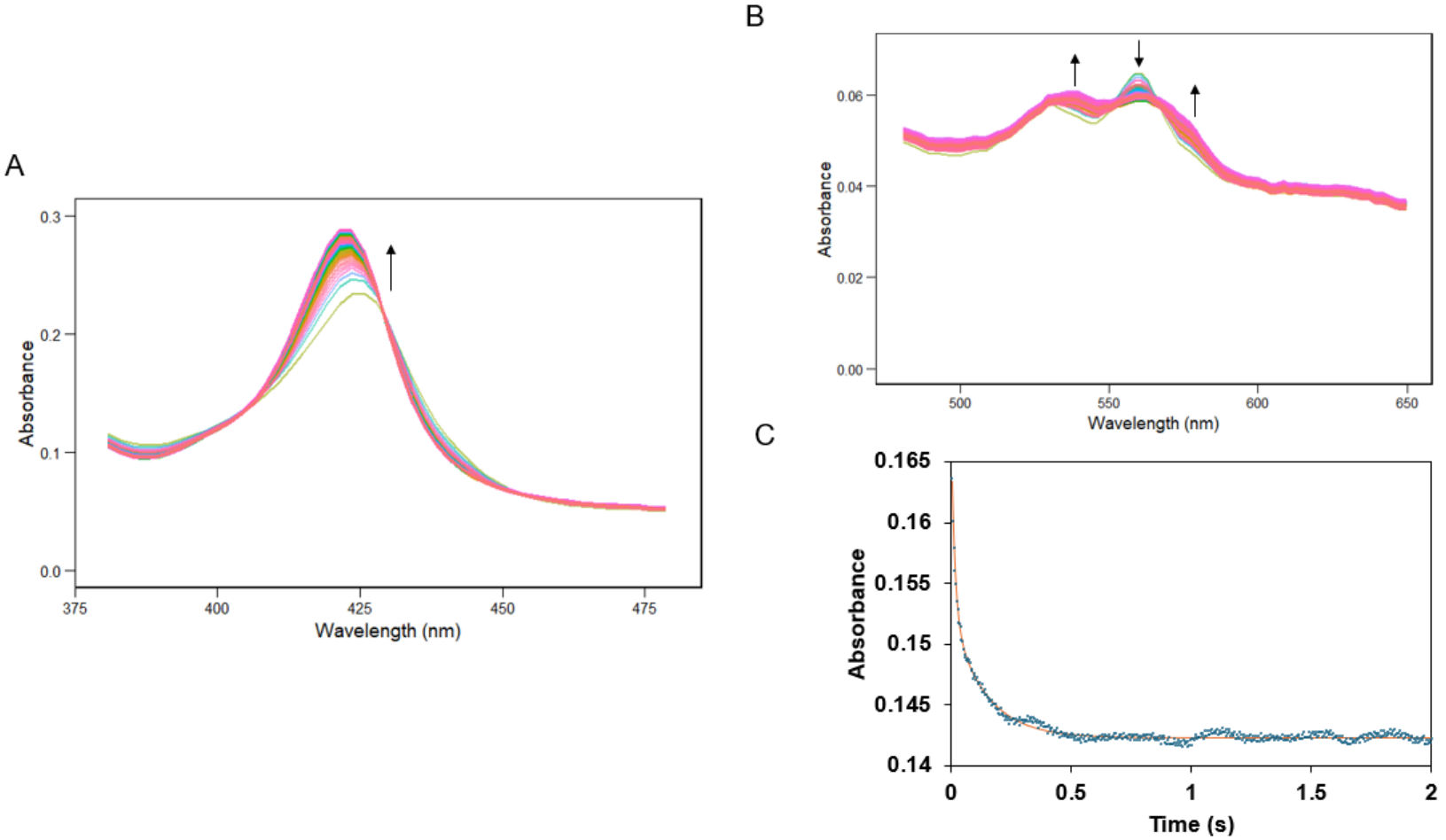


Figure 4.3.2.1 The time dependent spectra of CO bound to Adgb-GD (2.5 µM) at 100 µM CO concentration.

(A) Optical spectra of WT Adgb-GD visualised using RStudio. The time dependent spectrum exhibits a set of isosbestic points at: 404 nm, 429 nm, 453 nm therefore suggesting that only a single optical transition between two species is present, namely the ferrous species and the ferrous-CO bond complex formed. (B) Highlight the visible region and the last two isosbestic point at 554 nm and 569 nm. (C) Time course at 100 µM CO concentration and the fits of the data (double exponential).

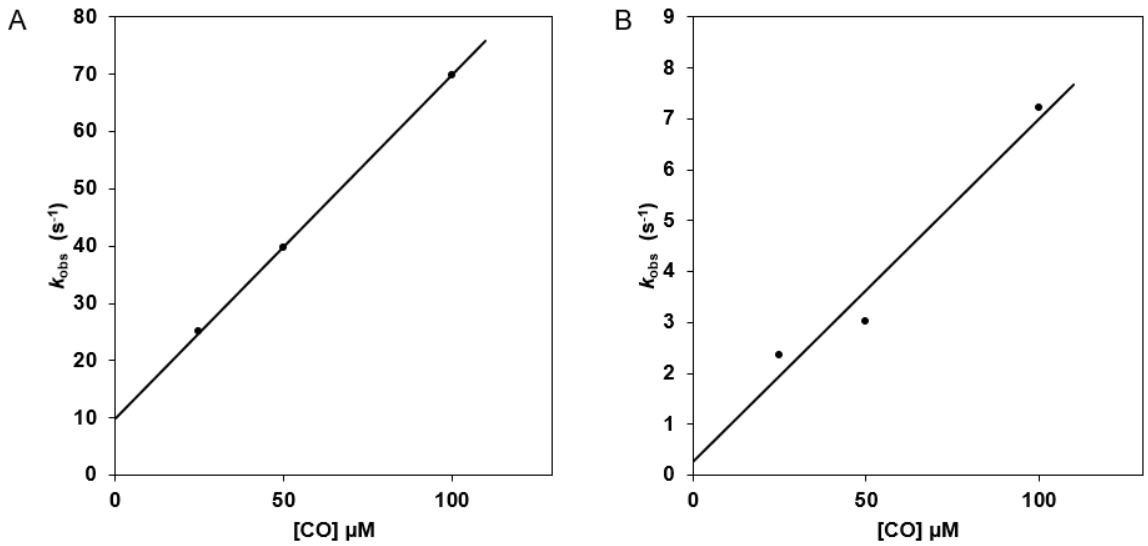


Figure 4.3.2.2. Adgb-GD WT binding kinetics with CO at 434 nm.

(A) Fast phase k_1; of ferrous WT with CO with final concentration of 2.5 µM after mixing in the stopped flow chamber reduced with excess dithionite in a degassed 0.1M sodium phosphate pH 7.4 buffer. $k_{on} = 5.99 \times 10^5$ M^{-1} s^{-1}, $k_{off} = 9.94$ s^{-1}. (B) Slow phase k_2; $k_{on} = 6.75 \times 10^4$ M^{-1} s^{-1}, $k_{off} = 0.26$ s^{-1}.

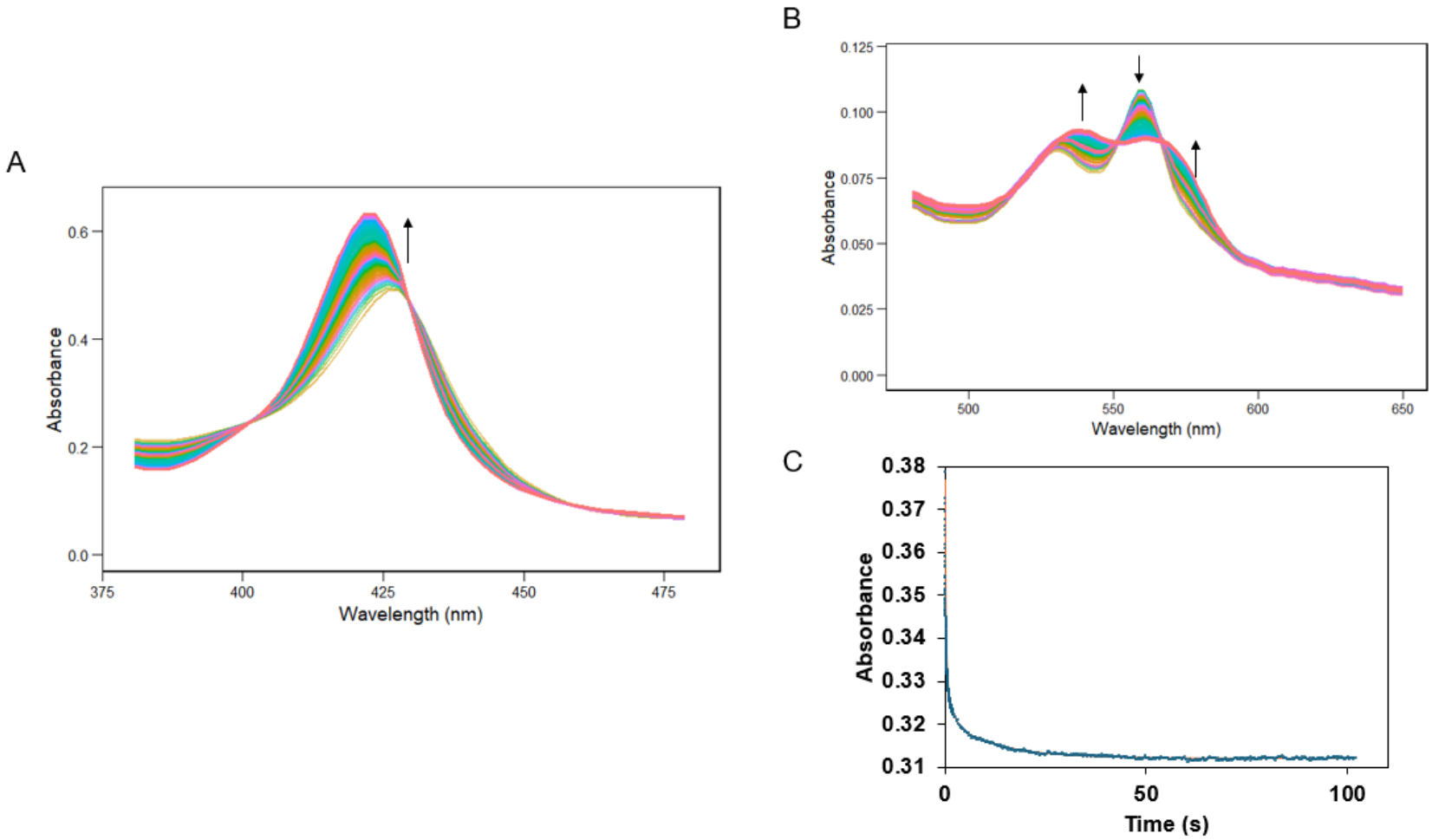


Figure 4.3.2.3. The time dependent spectra of CO binding to disulfide reduced Adgb-GD (with TCEP) (2.5 µM) at 100 µM CO concentration.

(A) Optical spectra of disulfide reduced Adgb-GD (with TCEP) visualised using RStudio. The spectrum also exhibits similar set of isosbestic points as the WT. (B) Highlight the visible region and the last two isosbestic point similar to the WT. (C) The time course for 100 µM CO concentration and the fits of the data for wavelength at 434 nm (4 exponentials).

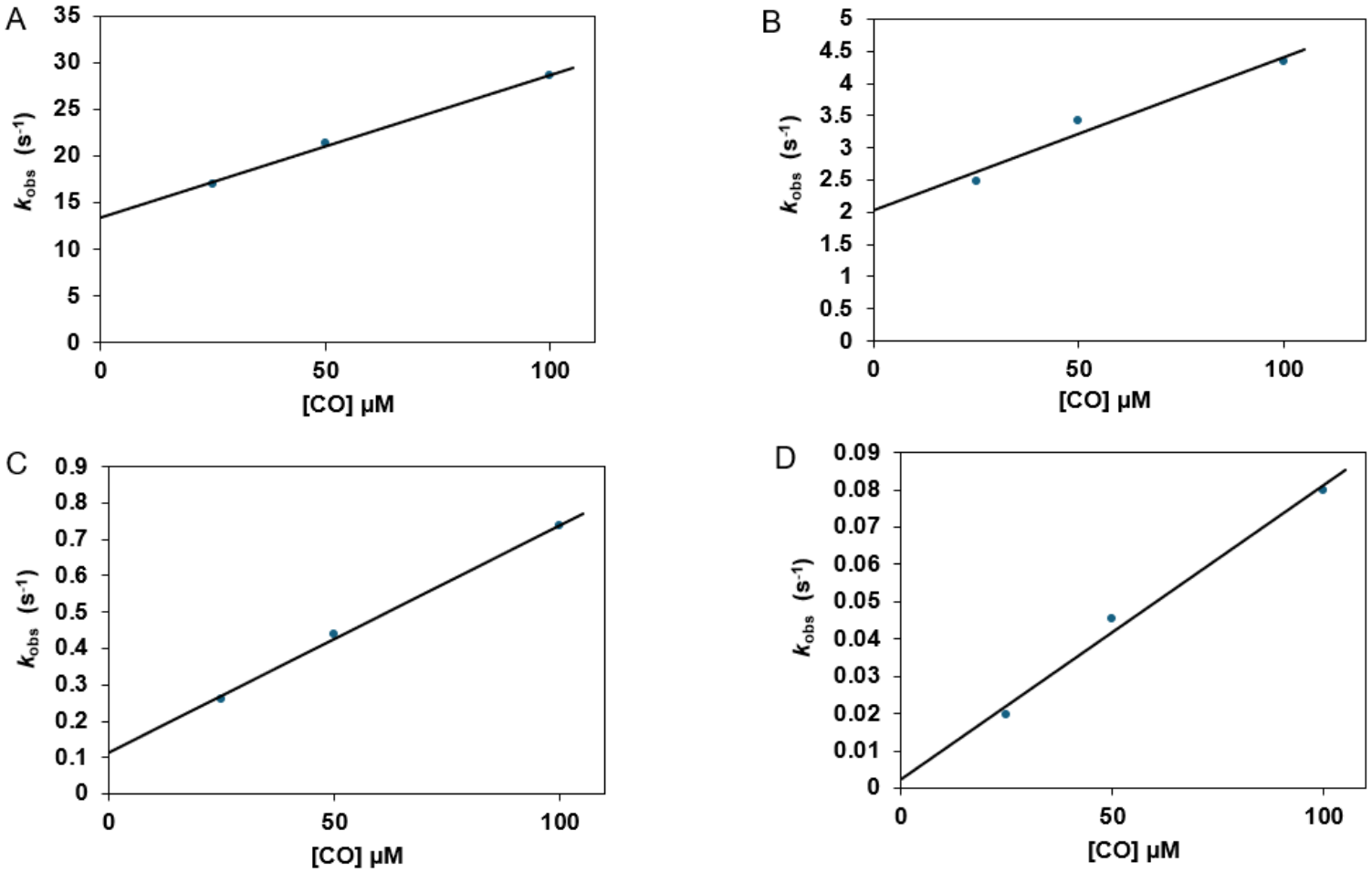


Figure 4.3.2.4. Disulfide reduced Adgb globin domain (with TCEP) binding kinetics with CO at 434 nm.

(A) Fast phase k_1; of ferrous WT with CO with final concentration of 2.5 µM after mixing in the stopped flow chamber reduced with excess dithionite in a degassed 0.1M sodium phosphate pH 7.4 buffer. k_{on} = 1.53 x 10^5 M^{-1} s^{-1}, k_{off} = 13.38 s^{-1}. (B) Slow phase k_2; k_{on} = 2.39 x 10^4 M^{-1} s^{-1}, k_{off} = 2.02 s^{-1}. (C) Slow phase k_3; k_{on} = 6.3 x 10^3 M^{-1} s^{-1}, k_{off} = 0.11 s^{-1}. (D) Slow phase k_4; k_{on} = 8 x 10^2 M^{-1} s^{-1}, k_{off} = 0.0025 s^{-1}.

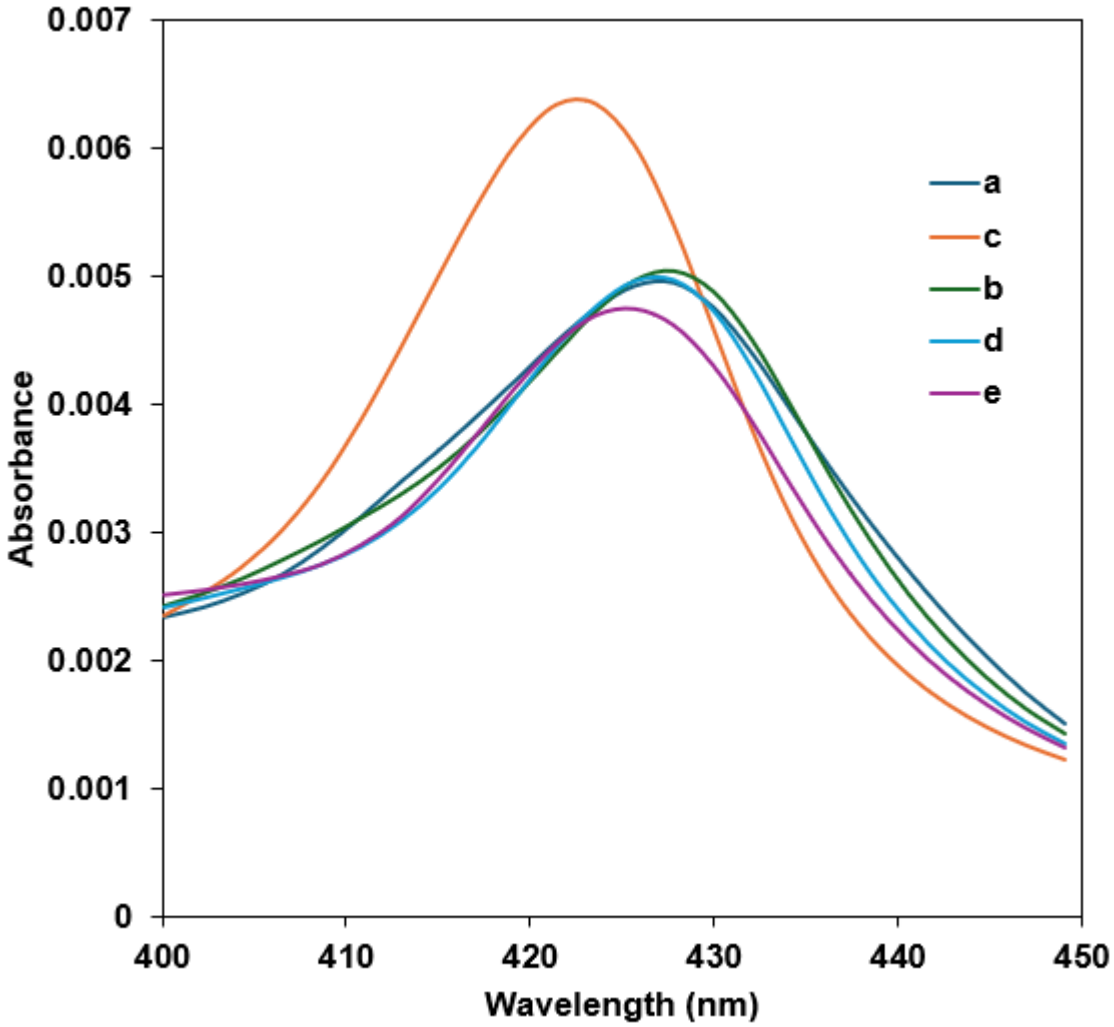
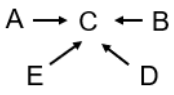


Figure 4.3.2.5. Global analysis fit of TCEP reduced Adgb-GD protein.



It was fitted to an sequential mechanism between 400 – 450 nm at CO concentration of 100 µM. The fit indicates the presence of more than one form of the protein in the reaction.

4.3.3 Kinetics of CN binding to WT androglobin globin domain, myoglobin and neuroglobin

To investigate the heme pocket environment of Adgb-GD and other globins (Mb and Ngb), and to gain deeper insight into the thermodynamics and kinetics of their interactions, the rate at which cyanide (CN) binds to these globins was examined. Following the rapid mixing of ferric Mb or ferric Ngb (5 μM) with solution containing cyanide (KCN) in NaPi buffer pH 7.4, optical transitions were observed showing that CN binds to these heme proteins as shown in the appendix, A5. Their time courses showed CN binding to Mb and Ngb comprises a single kinetic phase and a double exponential function respectively.

Table 4.3.2.1 provides a helpful summary of rate constants. Mb shows a linear dependence of the rate constants on CN concentration with a second-order rate constant of 9.97 x 10^2 M^{-1} s^{-1} (Figure 4.3.3.3A) which is slightly faster than the 4.0 x 10^2 M^{-1} s^{-1} reported previously (Antonini *et al*., 1971). Ngb exhibited a biphasic reaction, with both fast and slow phases of CN binding showing linear dependence, as previously reported by Herold *et al*. (2004). The second-order rate constants determined in this study are 3.8 M^{-1} s^{-1} and 0.681 M^{-1} s^{-1}, which is again slightly higher than the corresponding fast and slow phase rate constants 1.65 ± 0.06 M^{-1} s^{-1} and 0.37 ± 0.02 M^{-1} s^{-1} reported previously by Herold *et al*. (2004).

Figure 4.3.3.1 shows the optical transition and kinetic traces for CN binding to Adgb, monitored at 421 nm. The data were fitted using a triple exponential equation. Adgb displays ligand-binding behaviour comparable to that of Ngb, characterised by low affinity for CN and low binding across a broad concentration range. However, its binding appears slightly greater than that of Ngb, exhibited by larger optical changes. Additionally, as observed with Ngb, CN binding becomes less apparent at lower concentrations, with no detectable binding at 1 mM. No changes were observed in the visible region, and the spectrum closely resembles that of

Ngb, indicating similar behaviour and suggesting that Adgb is also hexacoordinate in its ferric form. However, Adgb appears to exhibit at least one isosbestic point in its optical spectra, resembling that of Mb (Figure 4.3.3.1A vs. appendix A5.1).

There are at least two distinct conformational states of the protein that interact with the ligand at different rates in the WT Adgb and they displayed a linear dependence on cyanide concentration (Figure 4.3.3.2). The rate constant for the third phase was too small to be reliably measured and was therefore considered negligible. The second-order rate constants derived from the linear fits of the fast and slow binding phases was 1.62×10^2 M^{-1} s^{-1} and 8.5 M^{-1} s^{-1} respectively. These results indicate that the faster binding phase of Adgb-GD to CN is comparable to that of Mb, albeit approximately seven-fold slower. In contrast, the slow binding phase resembles that of Ngb fast binding phase but occurs at a rate nearly two times faster. This dual behaviour suggests that Adgb-GD may exhibit characteristics of both pentacoordinate and hexacoordinate species. In summary, Adgb-GD demonstrates a rapid binding phase akin to Mb and a slower phase more similar to the fast phase observed in Ngb, pointing to a complex binding mechanism that reflects mixed coordination states.

To gain insights into the potential complexity of binding mechanisms occurring at the heme pocket, the kinetics of CN binding was investigated with both distal and proximal site mutants. Overall, the mutants exhibited optical spectra trends similar to those of the WT Adgb-GD, and their kinetic data were best fitted using a triple exponential model. Their optical transition and kinetic traces for 20 mM CN binding, monitored at 421 nm are shown in Figure 4.3.3.3 and 4.3.3.4. The main difference is in their binding affinities for CN. All variants showed no perceived optical changes at 2 mM CN or below, with the Q792G mutant showing no detectable binding even at 5 mM. The lack of binding observed at low ligand concentrations suggests that the mutant protein has a lower affinity for the ligand than the WT.

CN binding with Q792G and H824C mutants are the most similar to the WT protein. Their observed rate displayed a linear dependency on CN concentration with a fast and slow phase (Figure 4.3.3.5). Their third phase was also too small to be reliably measured and was therefore considered negligible. Their second-order rates were comparative with the WT: 1.95 x 10^2 M^{-1} s^{-1} and 4.2 M^{-1} s^{-1} for Q792G and 2.2 x 10^2 M^{-1} s^{-1} and 14.1 M^{-1} s^{-1} for H824C. In contrast, both mutants Q792H and H824Y displayed zero-order rates. Although their second-order rate constant appeared negative (i.e. -1.95 × 10^2 $M^{-1}{\cdot}s^{-1}$ and -22 $M^{-1}{\cdot}s^{-1}$ for Q792H mutant, and -46.8 $M^{-1}{\cdot}s^{-1}$ and -19.9 $M^{-1}{\cdot}s^{-1}$ for H824Y mutant), their data still fit well to a straight line with no dependence on CN concentration. Their third phases were also too small to be reliably measured and was therefore considered negligible. This suggests that their reaction phase follows zero-order kinetics, indicating that their rate is independent of ligand concentration at this stage.

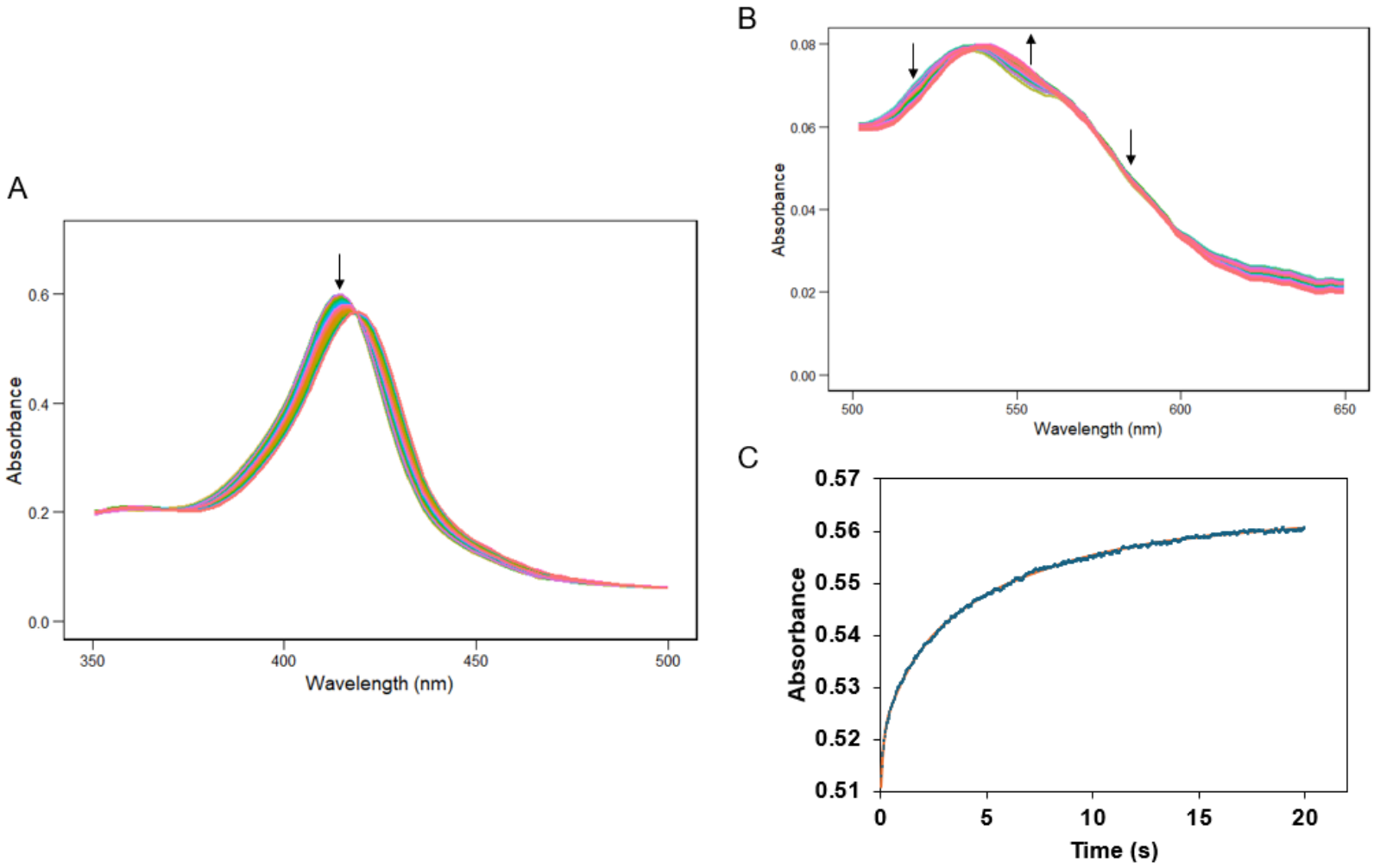


Figure 4.3.3.1 The optical spectra of CN binding to WT Adgb-GD (5 µM) at 20 mM CN concentration visualised using RStudio.

(A) Optical spectra of WT Adgb-GD visualised using RStudio. The ferric species and the ferric-CN bond complex formed. (B) Highlight the visible region. (C) The time course for CN binding to WT Adgb-GD at 421 nm. Colour code: fit = orange line, dataset = blue dots.

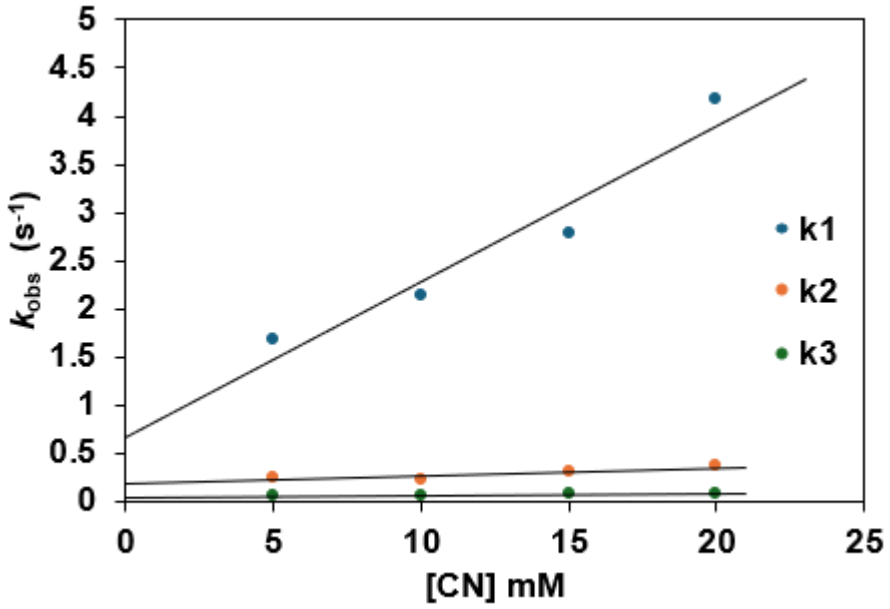


Figure 4.3.3.2. Global fitting of WT Adgb-GD binding kinetics with CN.

The plot revealed a linear dependence of rate constants on CN concentration. Rate constant, of Adgb-GD binding with CN in the stopped flow chamber in a 0.1M sodium phosphate pH 7.4 buffer. Fast phase k_1; k_{on} = 1.62 x 10^2 M^{-1} s^{-1}, k_{off} = 0.67 s^{-1}. Slow phase k_2; k_{on} = 8.5 M^{-1} s^{-1}, k_{off} = 0.18 s^{-1}.

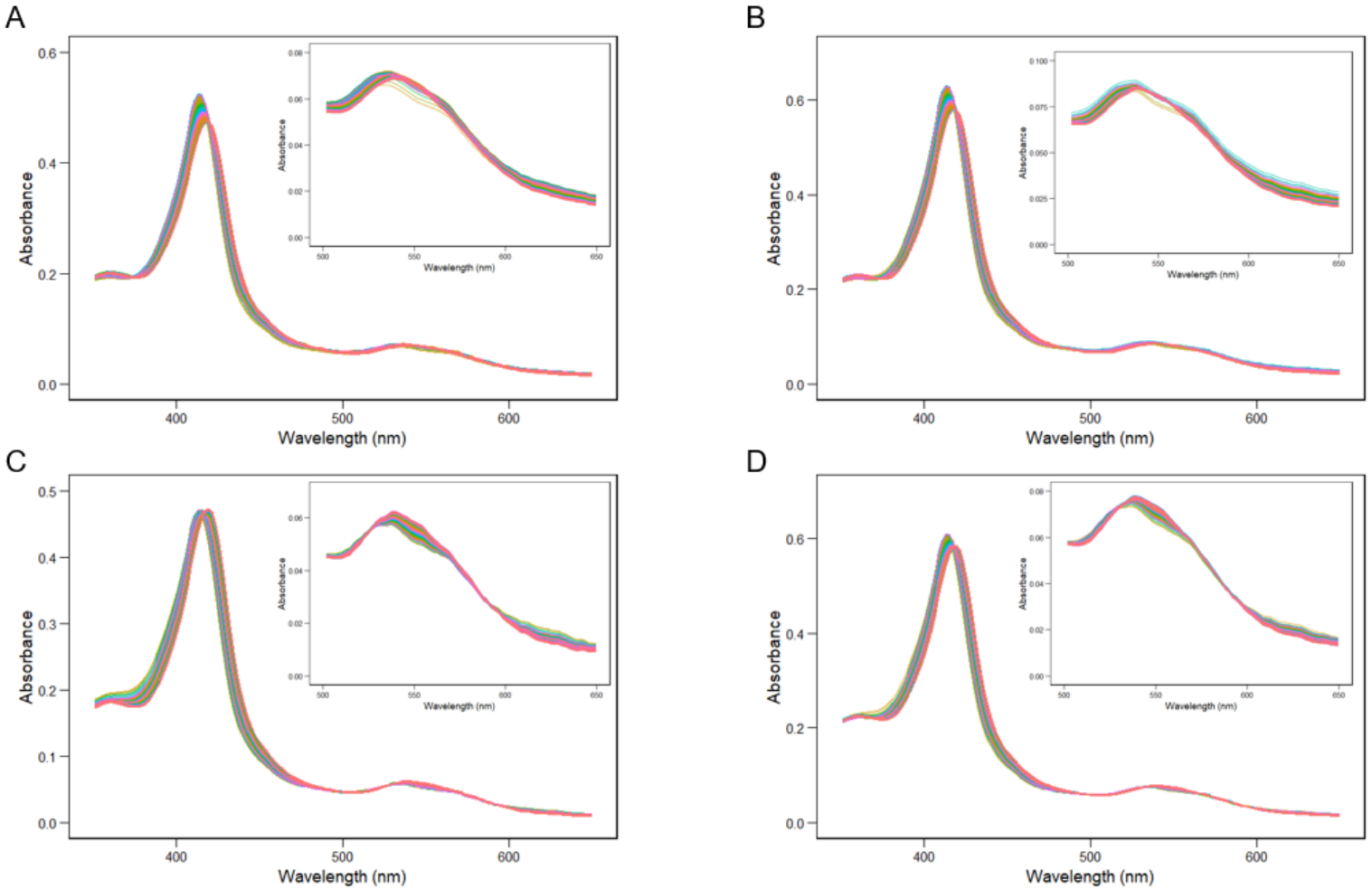


Figure 4.3.3.3 The optical spectra of CN binding to mutant Adgb-GD (5 µM) at 20 mM CN concentration visualised using RStudio. (A) Mutant Q792G Adgb-GD. (B) Mutant Q792H Adgb-GD. (C) Mutant H824C Adgb-GD. (D) Mutant H824Y Adgb-GD.

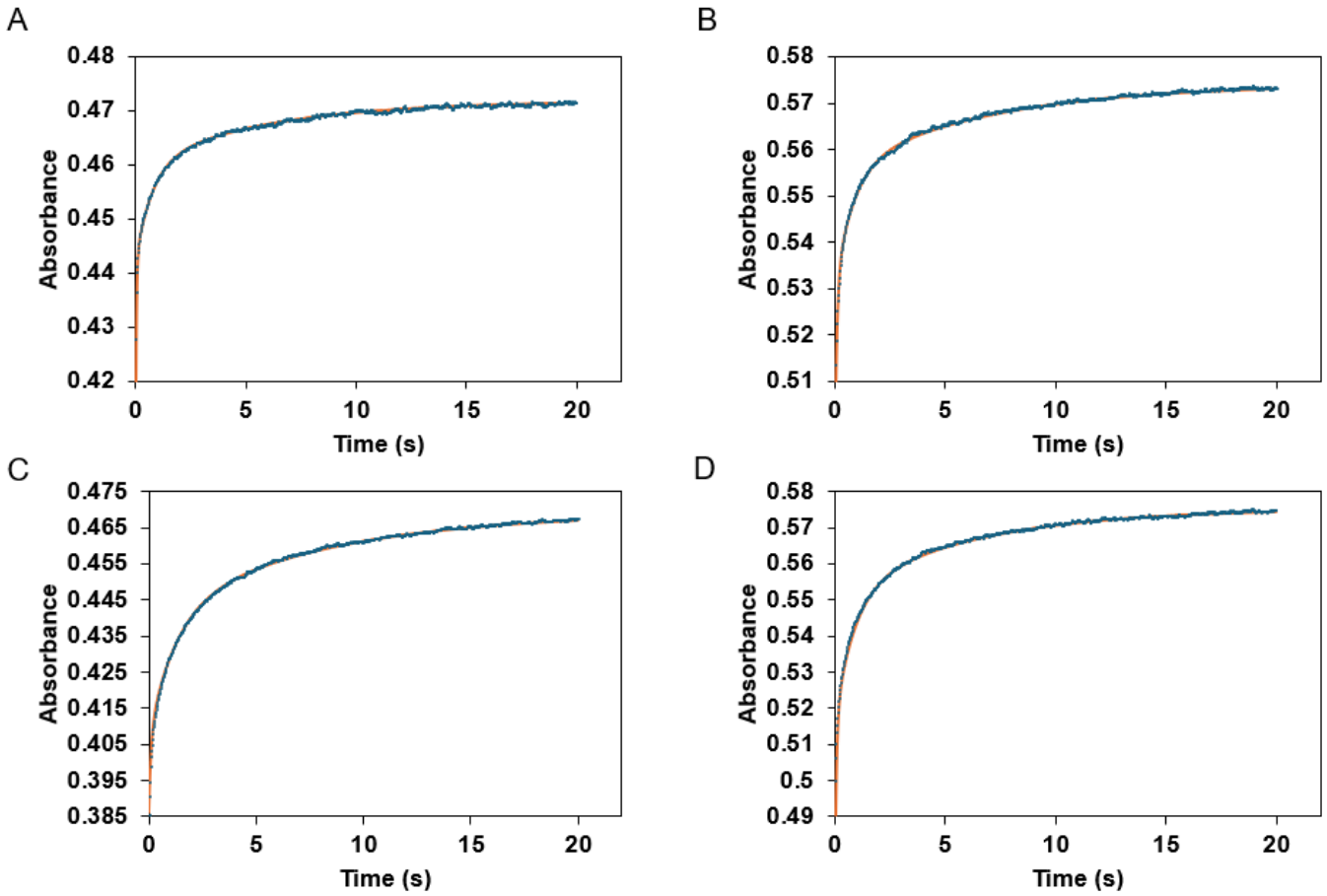


Figure 4.3.3.4. The time course for CN binding to mutant (5 µM) at 20 mM CN concentration.

(A) Mutant Q792G Adgb-GD. (B) Mutant Q792H Adgb-GD. (C) Mutant H824C Adgb-GD. (D) Mutant H824Y Adgb-GD. Colour code: fit = orange line, dataset = blue dots.

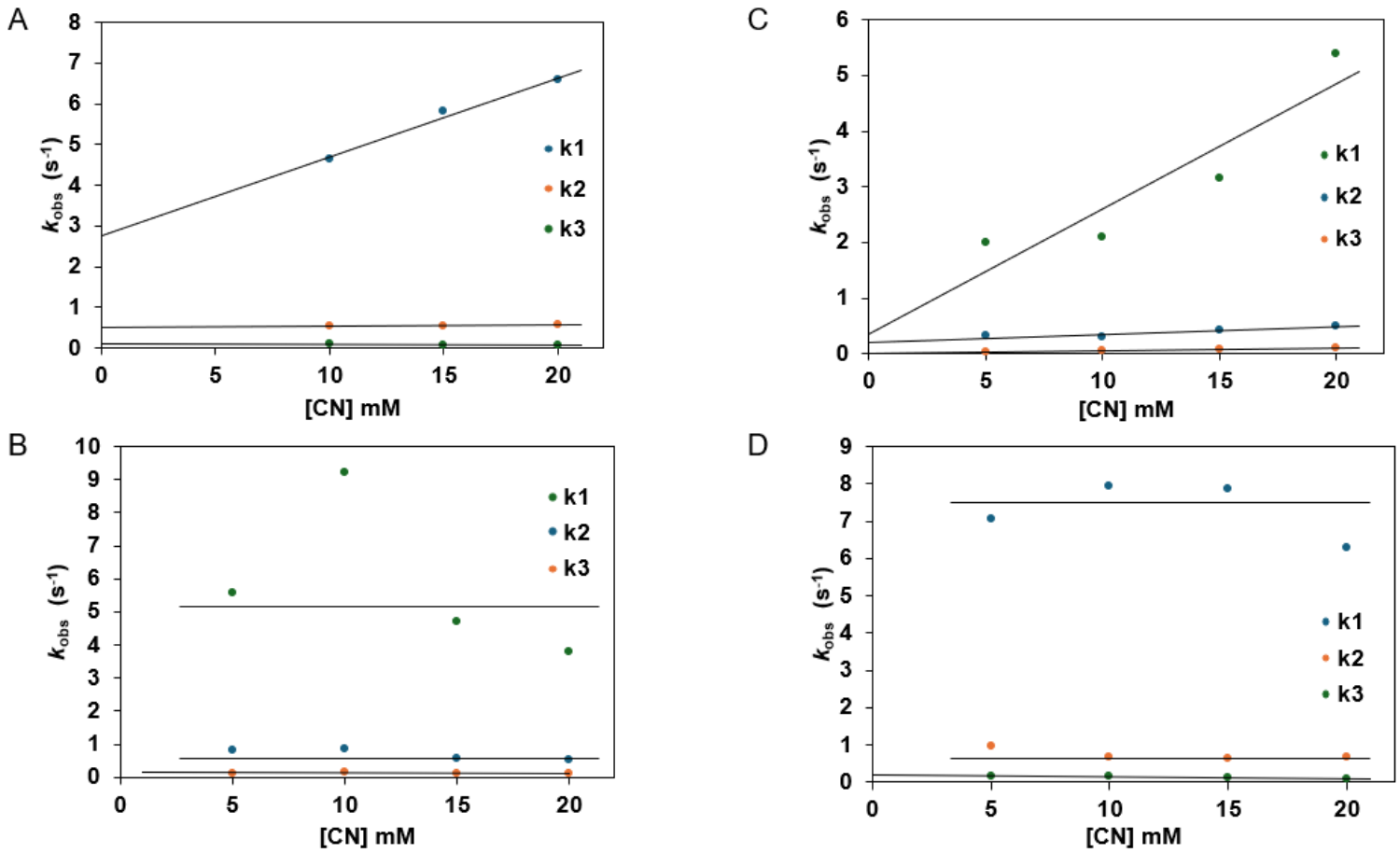


Figure 4.3.3.5. Global fitting of Mutant Adgb-GD binding kinetics with CN.

(A) Mutant Q792G Adgb-GD revealed a linear dependence of rate constants on CN concentration in similar condition as 4.3.3.2. Fast phase k_1; k_{on} = 1.95 x 10^2 M^{-1} s^{-1}, k_{off} = 2.76 s^{-1}. Slow phase k_2; k_{on} = 4.2 M^{-1} s^{-1}, k_{off} = 0.49 s^{-1}. (B) Mutant Q792H revealed a zero order reaction. (C) Mutant H824C Adgb-GD also revealed a linear dependence of rate constants on CN concentration in similar condition in (A). Fast phase k_1; k_{on} = 2.2 x 10^2 M^{-1} s^{-1}, k_{off} = 0.35 s^{-1}. Slow phase k_2; k_{on} = 14.1 M^{-1} s^{-1}, k_{off} = 0.21 s^{-1}. (D) Mutant H824Y revealed a zero order reaction.

4.3.4 Nitrite Reductase activity of myoglobin, neuroglobin and WT androglobin globin domain

To assess the potential role of Adgb as a nitrite reductase (NiR) enzyme involved in generating nitric oxide (NO) *in vivo*, the reaction between deoxyMb, deoxyNgb and deoxyAdgb-GD and nitrite was examined.

In the presence of sodium dithionite in sodium phosphate buffer (pH 7.4, 25 °C), the observed optical changes in the NiR reaction of deoxyMb indicate a transition from deoxyferrous Mb to NO-bound ferrous Mb (appendix A6). If the reaction was in the absence of dithionite, it will lead to a mixture of ferric Mb and NO-bound ferrous Mb (Shiva *et al*., 2007). However, the presence of dithionite reduces the ferric Mb generated from the reaction of nitrite and deoxyferrous Mb, leading to only NO-bound ferrous Mb being observed. Nonetheless in each system, the kinetics of the NiR activity are identical, meaning that the rate constants of the NiR activity of the globins are not affected by the presence or absence of dithionite (Shiva *et al*., 2007).

Table 4.3.2.1 provides a clear summary of the rate constants for the globin proteins. The corresponding time courses were fitted to either single- or double-exponential functions to determine these constants. The observed rate constants were then plotted against nitrite concentration, with the slope of the resulting line used to derive the NiR activity.

Mb revealed a linear trend with second-order rate constant of 9.2 $M^{-1}s^{-1}$, which is comparable to the previously reported value of 12.4 $M^{-1}s^{-1}$ (pH 7.4, 37 °C) and 6 $M^{-1}s^{-1}$ (pH 7.4, 25 °C) (Shiva *et al*., 2007). Whereas Ngb showed a biphasic dependence characterized by fast and slow kinetic phases. Both fast and slow phases followed pseudo-first-order time dependence on nitrite, though the overall nitrite reduction kinetics were slower compared to Mb (appendix A6).

The calculated second-order rate constants were 2.27 $M^{-1}s^{-1}$ for the fast phase and 0.32 $M^{-1}s^{-1}$ for the slow phase. The second-order rate constant for the fast phase is comparable to that reported in an earlier study for Ngb NiR activity ($5.1 \pm 0.4\ M^{-1}\ s^{-1}$; Petersen, Dewilde, and Fago, 2008). In contrast, the second-order rate constant for the slow phase observed in this study closely aligns with the fast-phase rate constant recently reported for Ngb under similar conditions ($0.12 \pm 0.02\ M^{-1}\ s^{-1}$ at pH 7.4, 25 °C or $0.26 \pm 0.02\ m^{-1}s^{-1}$ at 37 °C; Tiso *et al*., 2011). The slow phase in this study accounts for approximately 86% of the total reaction, whereas the fast phase contributes only about 14% making the slow phase the dominant part of the reaction.

The reaction of deoxyAdgb-GD with nitrite in the presence of sodium dithionite under identical conditions to those used for Mb NiR reaction showed optical transitions similar to the optical spectra from Mb and Ngb NiR reaction (Figure 4.3.4.1A). Therefore, indicating that deoxyAdgb-GD get reduced by nitrite to form a NO-bound ferrous protein. The formation of Adgb-GD-NO complex led to a shift in the Soret peak from 427 nm to 390 nm, along with the disappearance of both the beta and alpha bands at 530 nm and 560 nm respectively to yield a broad plateau between ~510 - 580 nm (Figure 4.3.4.1A). These changes generated isosbestic points at wavelength: 410 nm, 453 nm, 518 nm and 570 nm and are highlighted in the difference spectra in Figure 4.3.4.1B.

The time course at two wavelengths (412 nm and 438 nm) was fitted to a double exponential (Figure 4.3.4.1C) to calculate the rate constants. The observed rate constants were then plotted against nitrite concentration (Figure 4.3.4.1D). The nitrite reductase activity for the protein was obtained from the slope of this line. The second-order rate constant is 72.7 $M^{-1}s^{-1}$ for the fast phase and followed a pseudo-first-order time dependence on nitrite similar to Ngb. The second-order rate constant for the fast phase observed in this study is significantly higher than that of other globins and is approximately three times faster than the rate recently

reported for the Adgb NiR reaction (25 $M^{-1}s^{-1}$, Brandon *et al*., 2024). The fast phase accounts for approximately 83% of the total reaction, whereas the slow phase contributes only about 17% indicating that the fast phase represents the bulk of the conversion of deoxyferrous Adgb-GD to NO bound.

However, for the slow phase, the rate constant was essentially independent of nitrite concentration across the range of concentrations examined. This may reflect slow conformational changes within the protein or suggest that nitrite binding is not the rate-limiting step, but rather the dissociation of a distal amino acid ligand. An alternative explanation, given the findings in Chapters 2 and 5 indicating that Adgb-GD may be heterogeneous in nature, is that the slow phase represents binding at an alternative site. This site may involve a more tightly bound endogenous distal ligand.

To investigate this theory, the NiR activity of the mutant proteins was examined. Figures 4.3.4.2 and 4.3.4.3 present the time courses and corresponding data fits for the mutants i.e. Q792G, Q792H, H824C and H824Y, measured at the same wavelengths as the WT Adgb-GD (412 nm and 438 nm). The results revealed that all mutant proteins exhibited a biphasic kinetic pattern, with only the fast phases showing linear concentration dependency and pseudo-first-order time dependence on nitrite (Figure 4.3.4.2B and E, 4.3.4.3B and E vs Figure 4.3.4.2C and F, 4.3.4.3C and F). The slow phase exhibited nitrite concentration independence, mirroring the trend observed in the WT protein. Their second-order rate constant for the fast phase observed were: 60.2 $M^{-1}s^{-1}$ for Q792G mutant, 58 $M^{-1}s^{-1}$ for Q792H mutant, 44.5 $M^{-1}s^{-1}$ H824C and finally 81.2 $M^{-1}s^{-1}$ for H824Y mutant. These second-order rate constants are all comparable to those of the Adgb-GD WT protein and, like the WT, remain several orders of magnitude faster than those observed for Mb and Ngb.

Overall, the results from this study indicate that Adgb functions effectively as a nitrite reductase, with the introduced mutations having minimal impact on its enzymatic activity.

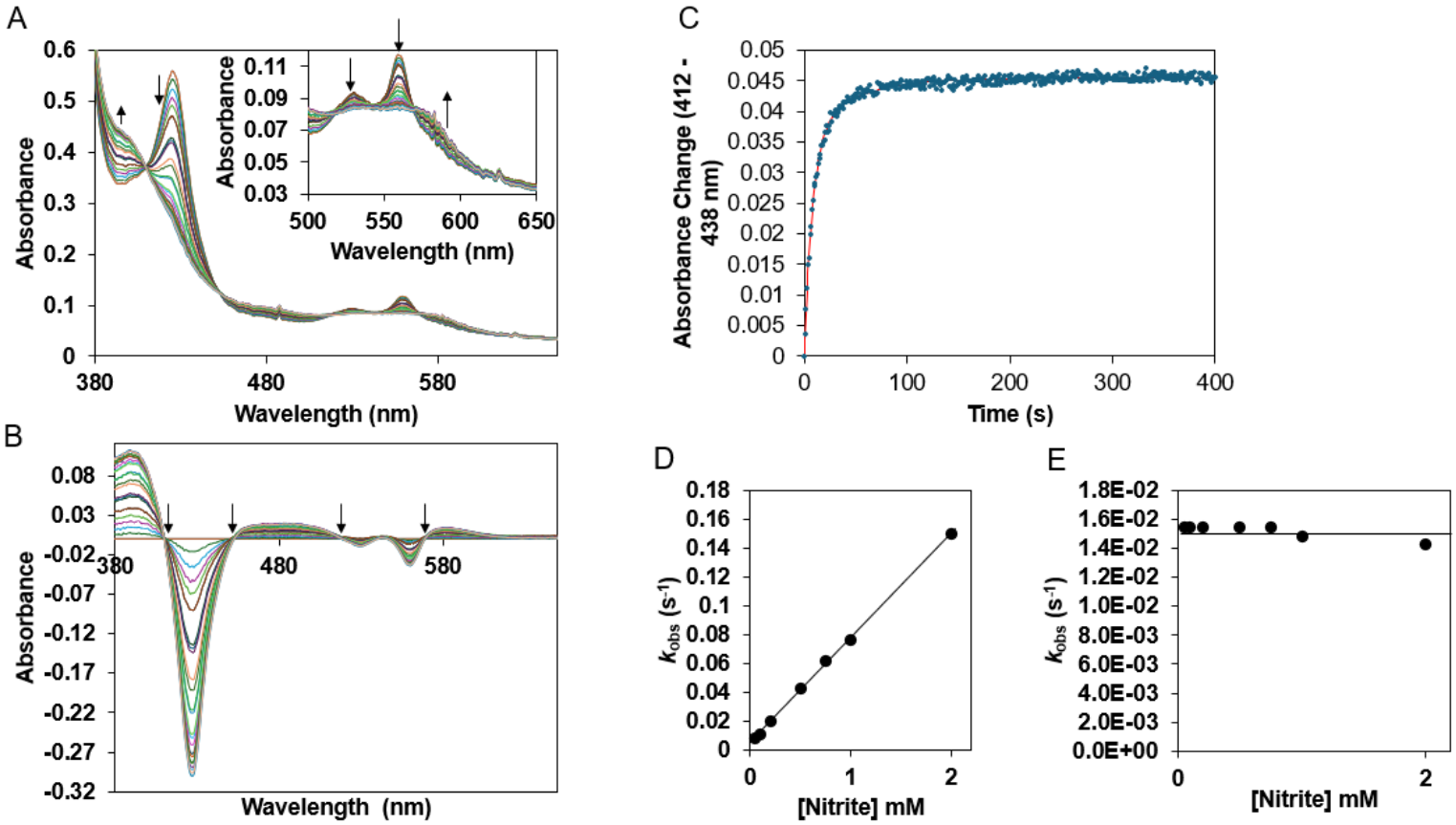


Figure 4.3.4.1. Optical changes and kinetics of nitrite reaction with Human deoxyAdgb-GD.

(A) UV-Vis spectrum of the anaerobic reaction of 5 µM deoxy-Adgb with 0.1 mM nitrite, pH 7.4, at 25 °C in the presence of dithionite. Inset, absorbance changes at the Q band (enlarged). (B) difference spectra of A. Arrow point to isosbestic points in A. (C) Rate constants were obtained by fitting the time course (412 – 438 nm) to a double exponential function (2 mM nitrite). Colour code: fit = red, dataset = blue dots. D, E: Plots of observed rate constant (k_{obs}) versus nitrite concentration; the second-order bimolecular rate constant obtained from the linear fit of both data where (D) Fast phase: 72.7 $M^{-1}s^{-1}$ and k_{off} = 0.0048 s^{-1} and (E) Slow phase: zero order reaction.

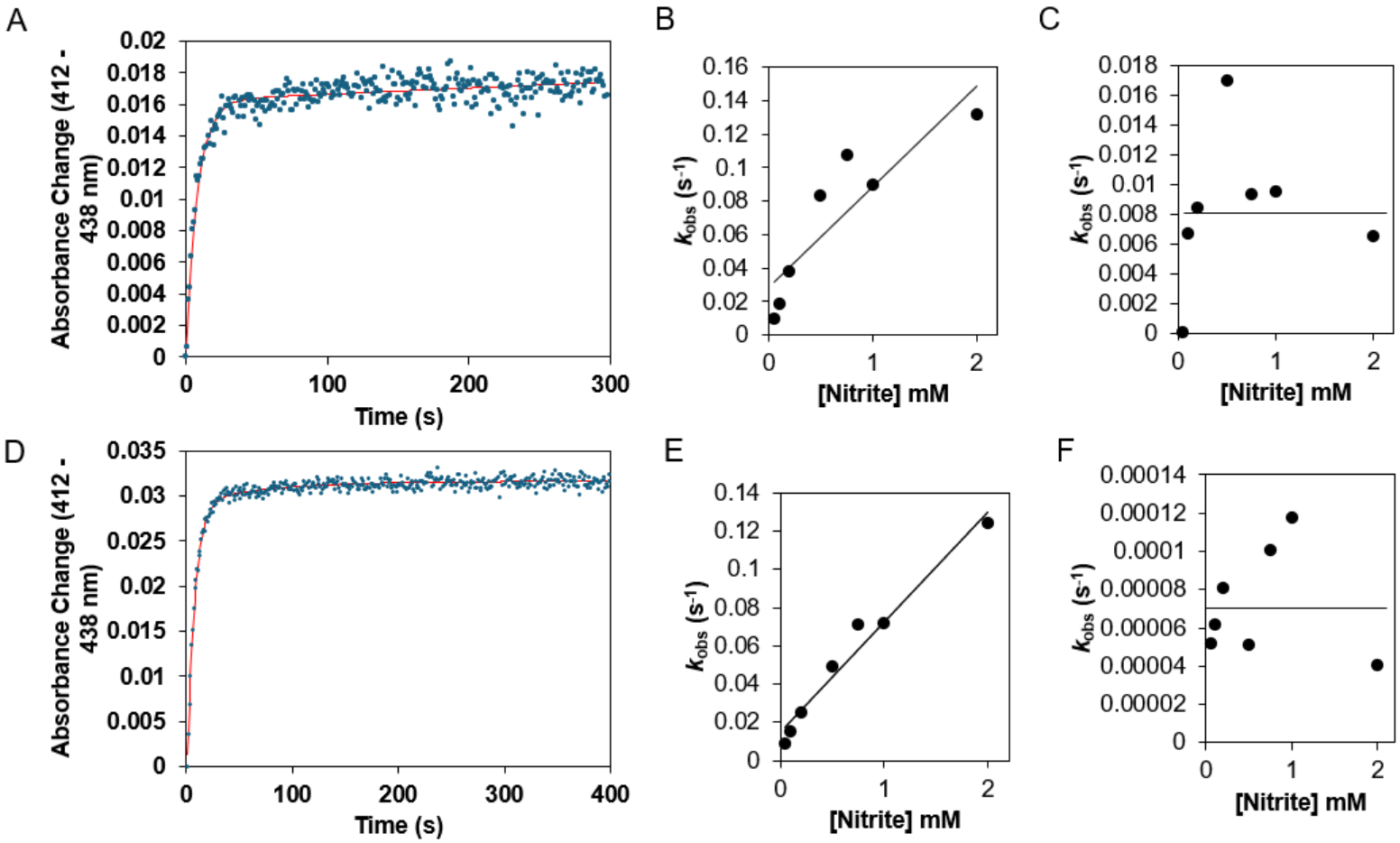


Figure 4.3.4.2. Kinetic traces of the nitrite reaction with Human deoxyAdgb-GD Mutants.

(A) Q792G: Rate constants were obtained by fitting the time course (412 – 438 nm) to a double exponential function (2 mM nitrite). Colour code: fit = red, dataset = blue dots. (B) The second-order bimolecular rate constant obtained from the linear fit of both data where fast phase: 60.2 $M^{-1}s^{-1}$ and k_{off} = 0.029 s^{-1} and (C) Slow phase: zero order reaction. (D) Q792H: Fit using similar condition in A. (E) Fast phase: 58 $M^{-1}s^{-1}$ and k_{off} = 0.014 s^{-1} and (F) Slow phase: zero order reaction.

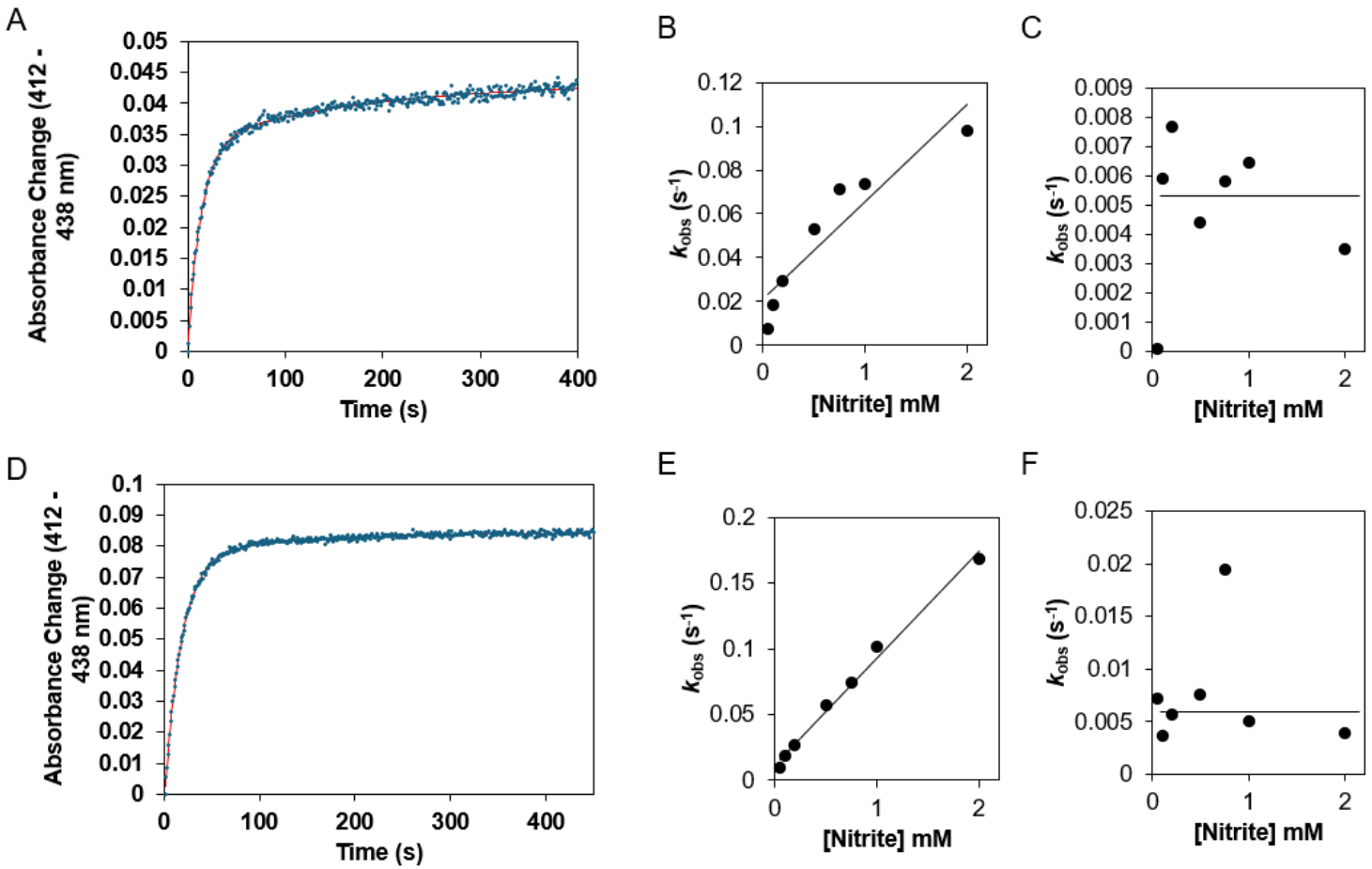


Figure 4.3.4.3. Kinetic traces of the nitrite reaction with Human deoxyAdgb-GD Mutants.

(A) H824C: Rate constants were obtained by fitting the time course (412 – 438 nm) to a double exponential function (2 mM nitrite). Colour code: fit = red, dataset = blue dots. (B) The second-order bimolecular rate constant obtained from the linear fit of both data where fast phase: 44.5 $M^{-1}s^{-1}$ and k_{off} = 0.021 s^{-1} and (C) Slow phase: zero order reaction. (D) H824Y: Fit using similar condition in A (0.5 mM nitrite). (E) Fast phase: 81.2 $M^{-1}s^{-1}$ and k_{off} = 0.012 s^{-1} and (F) Slow phase: zero order reaction.

4.3.5 Kinetics of NO binding to WT androglobin globin domain

To determine the binding of NO to Adgb-GD, deoxyferrous Adgb WT was rapidly mixed with known concentration of NO in degassed NaPi buffer pH 7.4. Optical transition was monitored and is displayed in Figure 4.3.5.1. The presence of isosbestic points in the spectrum at 406 nm, 453 nm, 520 nm and 571 nm (Figure 4.3.5.1A and B) indicate that these transitions are identical in terms of a single start and final spectrum with no optically distinct intermediates. Additionally, moving from a higher NO concentration to a lower NO concentration does not seem to affect the spectra as it can be observed on the same figure below (Figure 4.3.5.1A vs 4.3.5.1B). The time courses and corresponding fits at 421 nm and 434 nm (Figures 4.3.5.1C and D) followed a double exponential pattern, indicating that the reaction kinetics consist of two phases.

As mentioned earlier, WT Adgb-GD readily binds to NO and thus the optical transition in Figure 4.3.5.1A and 4.3.5.1B. The optical transition is identical to the optical transition observed in Figure 4.3.4.1A, thus indicating the formation of Adgb-GD-NO complex. A similar trend was observed for Mb. Optical transition indicative of the formation Mb-NO complex as previously shown in Figure 4.3.4.1A was observed in Figure 4.3.5.2A. Global fitting analysis revealed the reaction kinetics for NO binding to Mb is biphasic and the rate constant vs concentration plot for the two phases are given in Figure 4.3.5.2B and C, where B = fast phase and C = slow phase. This contrasts with the typical monophasic behaviour observed for Mb in previous studies (e.g., Moore and Gibson, 1976). The biphasic binding of NO to deoxyMb produced a second-order rate constant of 1.14×10^6 $M^{-1}s^{-1}$ for the fast phase and 6.45×10^4 $M^{-1}s^{-1}$ for the slow phase. The fast phase observed for Mb in this study differs significantly from the literature value, being approximately 15 times slower than the reported second-order rate constant for NO binding to Mb, 1.7×10^7 M^{-1} s^{-1} (Moore and Gibson, 1976). The slow phase is

comparable to the more recently reported value of second-order rate constant reported: 2.5 ± 0.6 × 10^4 M^{-1} s^{-1} (Laverman *et al*., 2001).

Figure 4.3.5.3 show the plots of the global fit of rate constant vs concentration for Adgb-GD. The plot revealed a linear relationship between the rate constant and NO concentration in both phases, indicative of pseudo-first-order kinetics. The second-order rate constant for the fast phase and slow phase were 1.39 x 10^6 M^{-1} s^{-1} and 1.32 x 10^4 M^{-1} s^{-1} respectively (Figure 4.3.5.3A and 4.3.5.3B). The high y axis intercept also suggests a high NO dissociation rate. This characteristic more closely resembles that of Mb than Ngb. The recent study that reported Mb with a second-order rate constant of 2.5 ± 0.6 × 10^4 M^{-1} s^{-1} also reported its dissociation rate constant (k_{off}) to be 25 ± 6 s^{-1} (Laverman *et al*., 2001), closely matching the k_{off} of 24 s^{-1} observed for Adgb-GD. In contrast, Ngb displays significantly lower second-order rate constants of 21 ± 1 M^{-1} s^{-1} and 2.9 ± 0.4 M^{-1} s^{-1} for its fast and slow phases, respectively (Herold *et al*., 2004).

The amplitudes of Adgb-GD binding to NO were independent of NO concentration, showing that NO binding to the heme was saturated at all concentrations used. Their amplitudes at 421 nm and 434 nm are shown in Figure 4.3.5.3C and D respectively. The global fitting profile used in this analysis is shown in Figure 4.3.5.4, illustrating two ferrous forms of Adgb-GD converging into a single final species, denoted as Form B. This model supports the hypothesis of an alternative binding site due to the protein heterogenous nature (section 4.3.4).

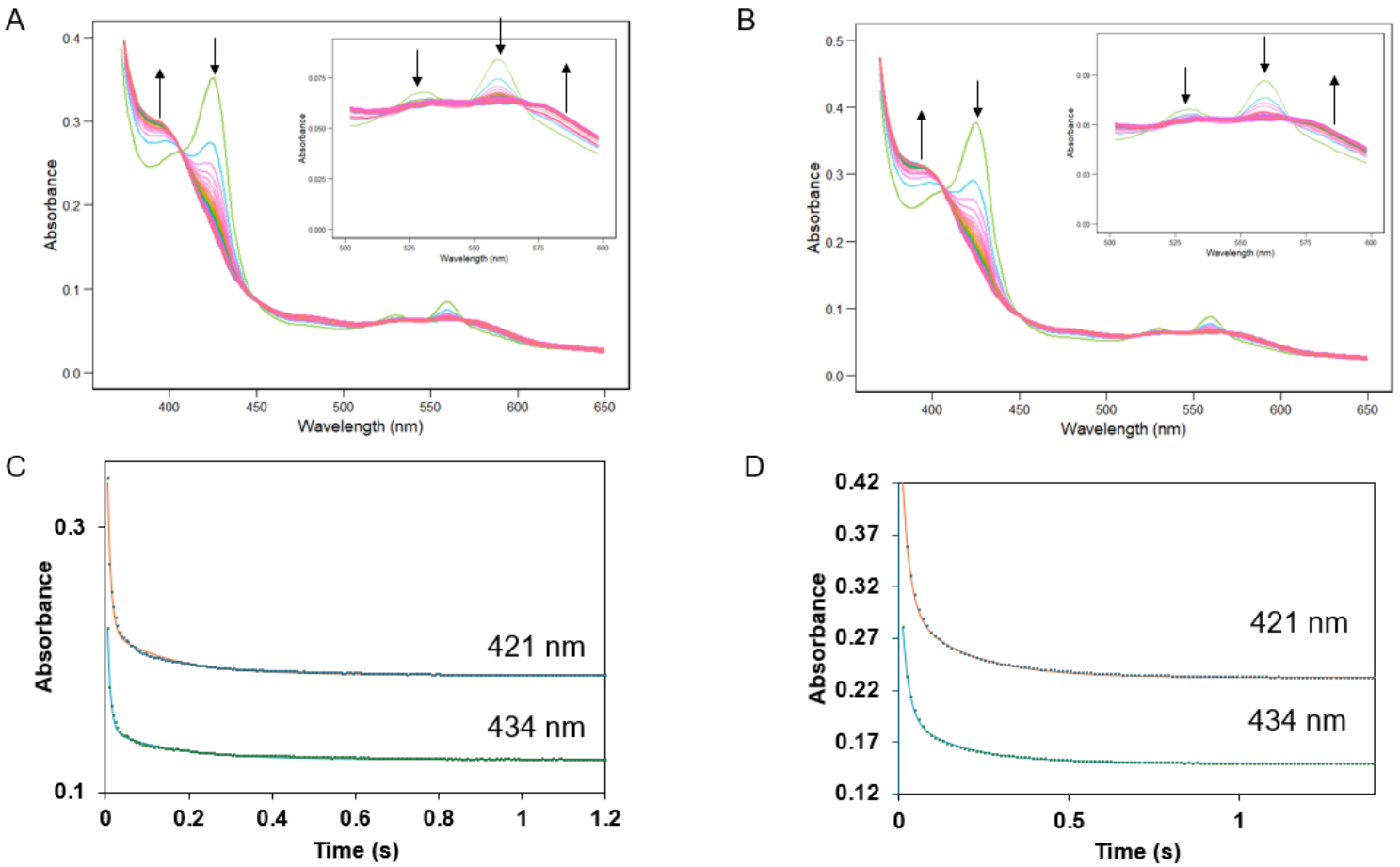


Figure 4.3.5.1. The optical spectra of NO bound to WT Adgb-GD (5 µM) at different NO concentrations.

(A) 72 µM NO concentration and (B) 60 µM NO concentration, visualised using RStudio. The time dependent spectrum exhibits isosbestic points therefore suggesting that only two optical species are present, namely the ferrous species and the ferrous-NO bond complex formed. The time course at different wavelengths for different NO concentration (C) 72 µM NO concentration and (D) 60 µM NO concentration and their fits are also illustrated. Colour code: Fits = orange and blue. Dataset = black dots.

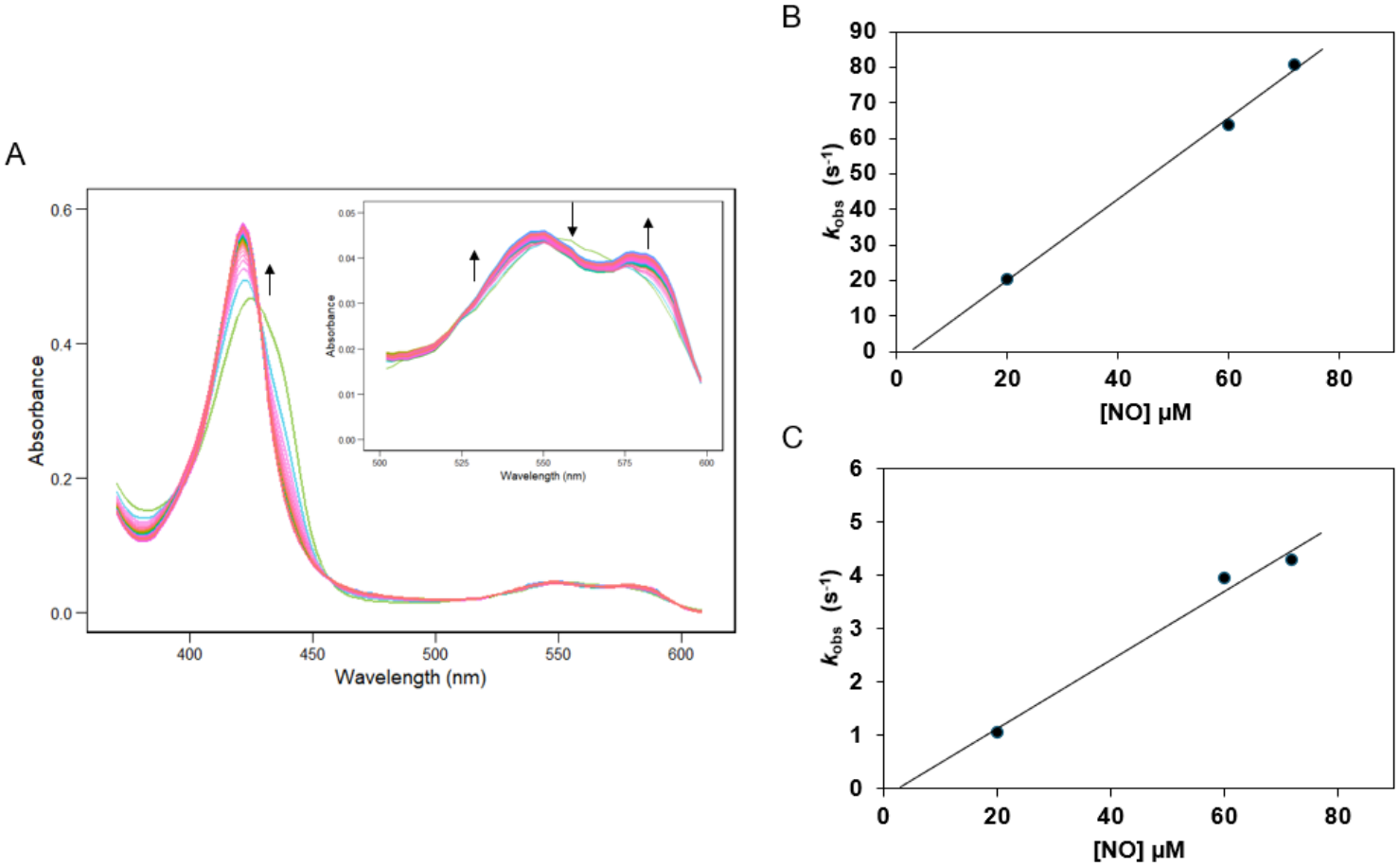


Figure 4.3.5.3. The optical spectra of NO bound to Equine Mb (5 µM), visualised using RStudio.

(A) Time dependent spectra of NO bound to deoxyMb (5 µM) at 72 µM NO concentration. (B) Global fitting of deoxyMb binding kinetics with three NO concentrations. Concentration at 40 µM NO was readjusted using 1.7 x 10^7 M^{-1} s^{-1} (Moore and Gibson, 1976) due to suspected oxygen contamination in the system. Fast phase k_1, has corrected second-order rate constant of 1.14 x 10^6 M^{-1} S^{-1}. (C) Slow phase k_2 with corrected second-order rate constant of 6.45 x 10^4 M^{-1} s^{-1}.

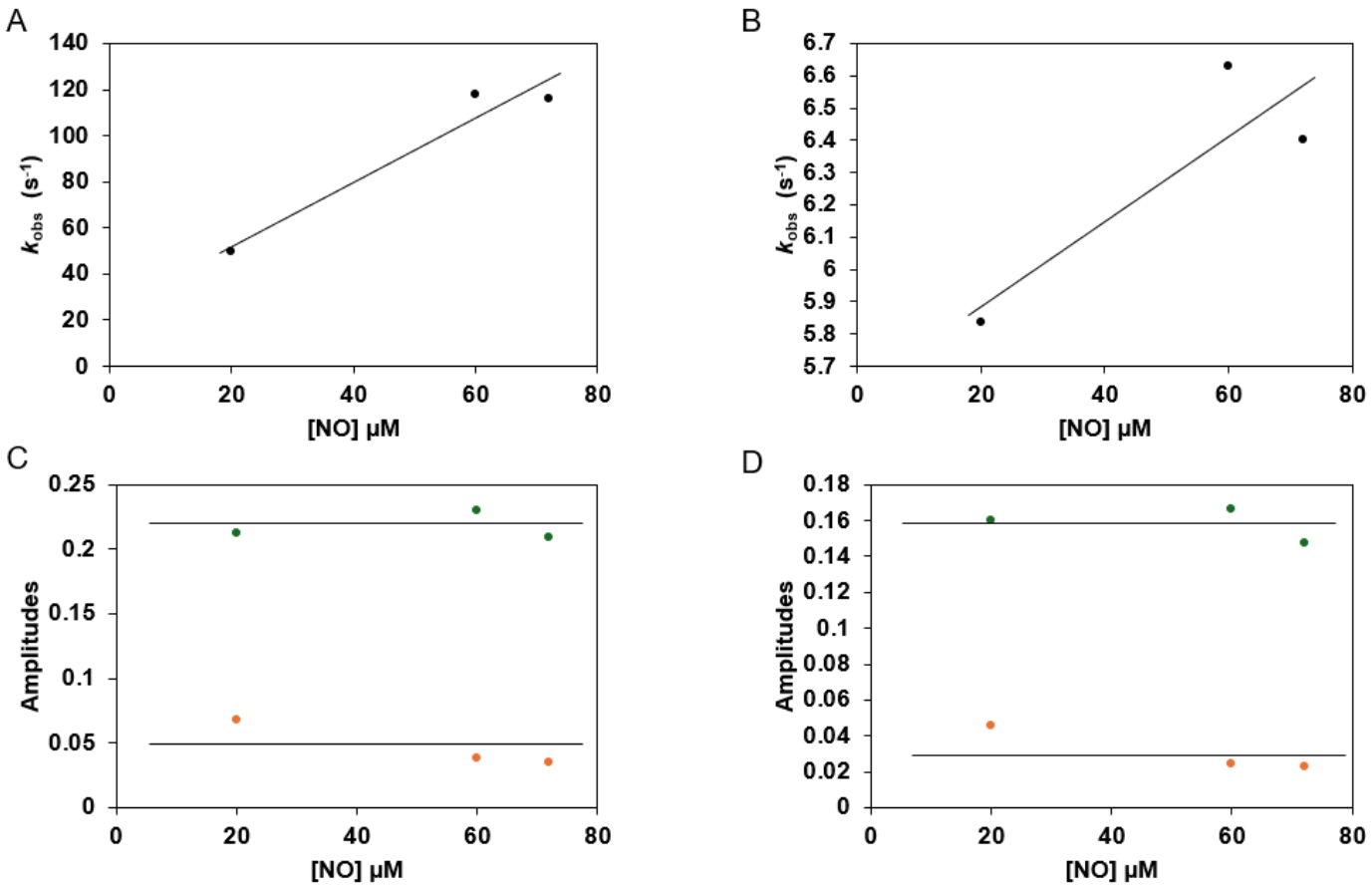


Figure 4.3.5.3 Global fitting of WT Adgb-GD binding kinetics with three NO concentrations revealed a linear dependence of rate constants on NO concentration.

(A) Rate constant, k_1, of ferrous WT reduced with excess dithionite reacting with NO in the stopped flow chamber in a degassed 0.1M sodium phosphate pH 7.4 buffer. The second-order rate constant is 1.39 x 10^6 $M^{-1}s^{-1}$ and k_{off} = 24.1 s^{-1}. (B) Rate constant of k_2 with second-order rate constant is 1.32 x 10^4 $M^{-1}s^{-1}$ and k_{off} = 5.6 s^{-1}. (C) Amplitude of WT Adgb-GD binding with NO at 421 nm and (D) 434 nm. Green dots = k_1, orange = k_2.

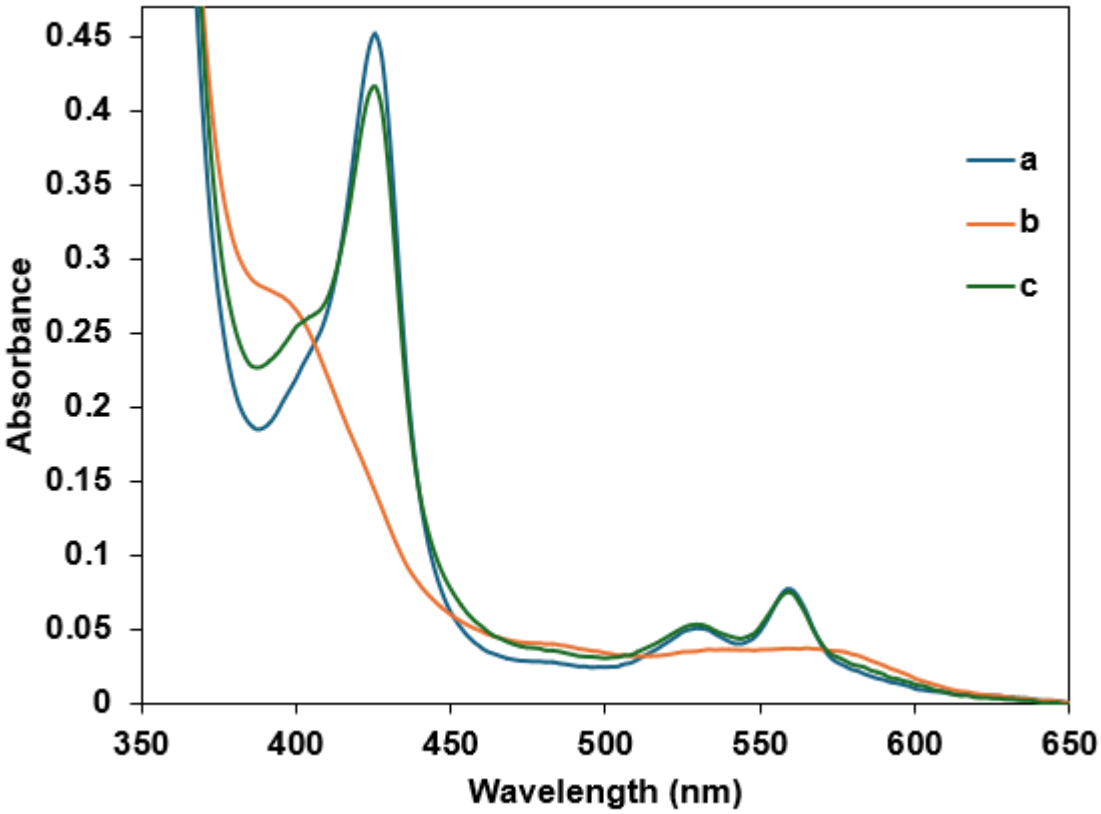
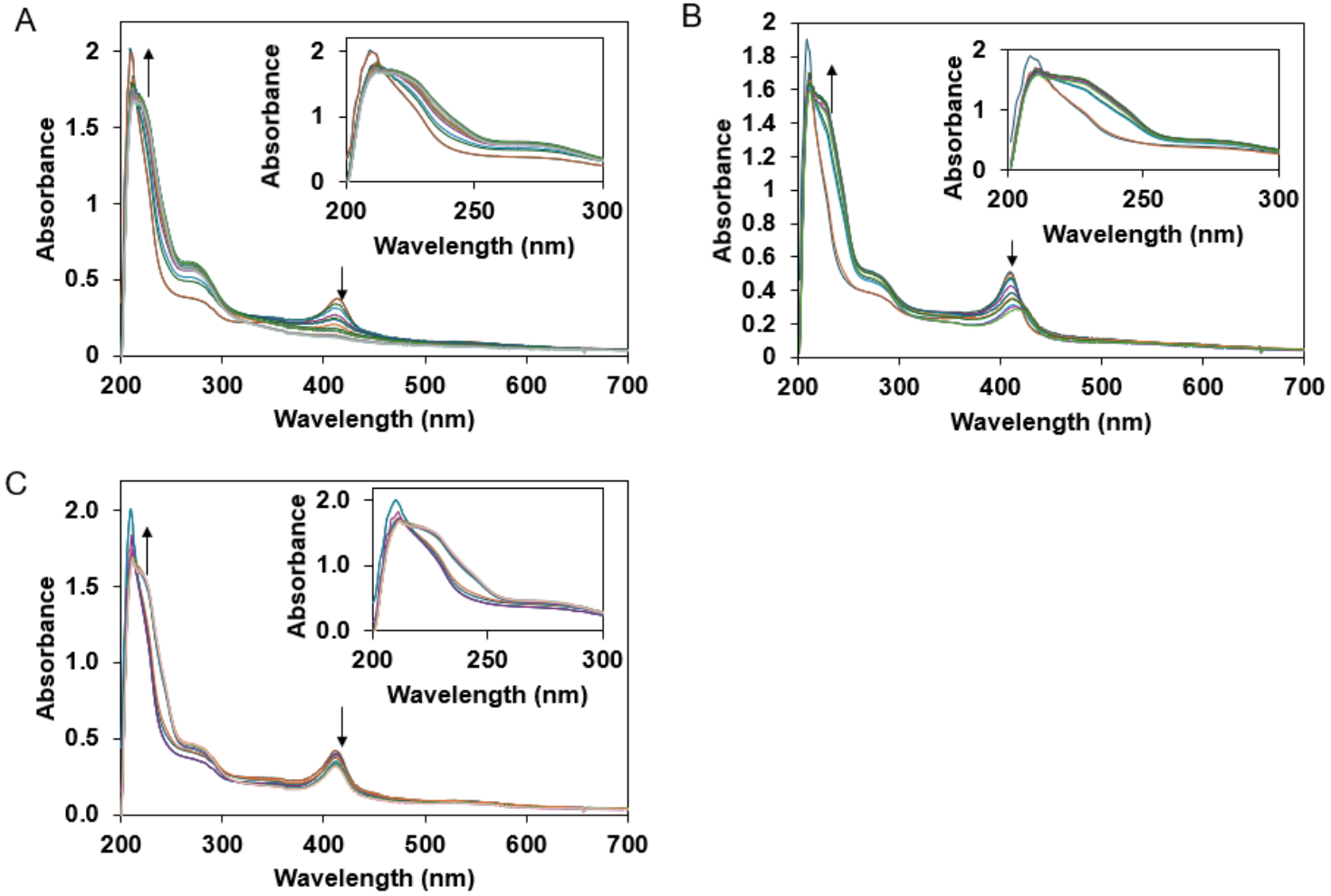


Figure 4.3.5.4. Global analysis fit to an A → B ← C sequential mechanism between 350 - 650 nm at NO concentration of 72 µM.

4.3.6 Liposome oxidation in Adgb

Many globins have a pronounced ability to initiate lipid oxidation reactions through redox cycling of the ferric-ferryl oxidation states (Reeder *et al*., 2011). The ability of Adgb-GD to oxidise lipids was tested by examining the oxidation of multilamellar liposomes compared to Mb and Ngb. Adgb, like Mb was found to initiate lipid oxidation, as demonstrated in Figure 4.3.6.1A as it showed optical transition similar to other heme proteins, such as Mb and Ngb, shown in Figures 4.3.6.1B and C, respectively. Mb shows a change in wavelength due to ferryl formation, however, ferryl is unstable for Ngb and Adgb (Reeder *et al*., 2024) and hence this shift is not observed. However, significant bleaching of the heme in Adgb was observed from radical damage, much more that the other proteins. The time course of the three heme proteins was compared in Figure 4.3.6.2A. All three heme proteins exhibited a characteristic sigmoidal shape curve (Beckerson *et al*., 2014). The process includes an initial lag phase, followed by a rapid phase driven by a cascade of lipid oxidation occurring between 3300 – 7300 s for Mb, 4000 – 6000 s for Adgb, and 6300 – 12000 s for Ngb, and concluding with a termination phase (Figure 4.3.6.2A) (Beckerson *et al*., 2014). Adgb and Mb exhibited similar pro-oxidant activity and their rate of activity are both greater than that of Ngb, with their approximate maximal lipid oxidation rates measured at 7.82 nM s^{-1} for Adgb, 7.42 nM s^{-1} for Mb, and 3.71 nM s^{-1} for Ngb, as illustrated in the bar chart (Figure 4.3.6.2B). This shows that like Mb, Adgb-GD has the ability to lead to lipid oxidation under mild oxidative conditions.



4.3.6.1 The optical spectra of the oxidation of lipids and heme proteins (2.5 μM) with different heme proteins.

(A) Adgb (B) Mb and (C) Ngb. It is showing the rate at which the conjugated diene is forming. The spectrum was recorded with an Agilent Cary diode ray and shows the bleaching of the heme in all the protein. The formation of conjugated dienes was monitored at ~234 nm.

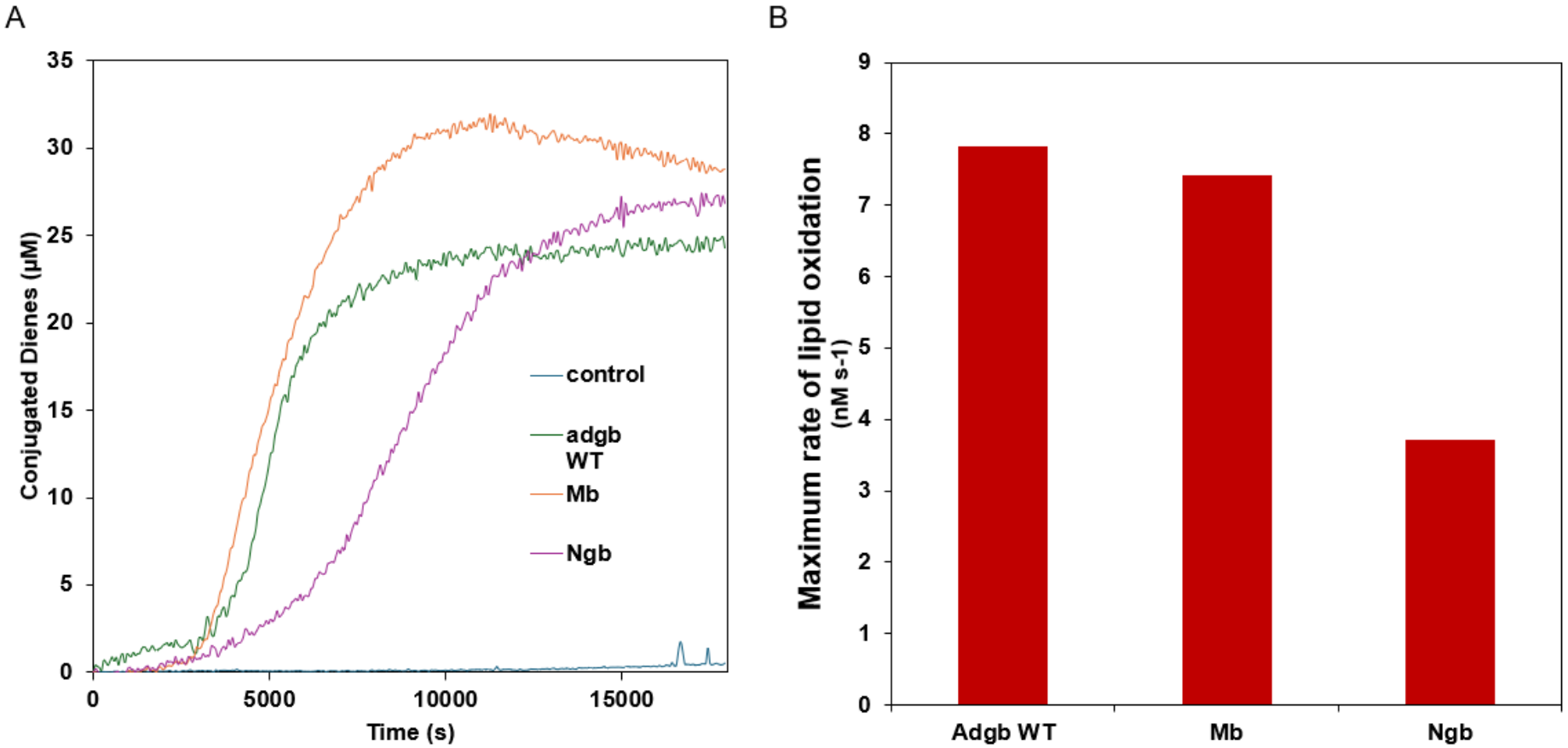


Figure 4.3.6.2 Comparison of the oxidation of liposomes by Adgb-GD, Mb and Ngb.

The lipid oxidation was monitored at 234, 220 and 250 nm and the concentration of conjugated diene was calculated by dividing the 3-point baseline drop absorbance using extinction coefficient 11000 M^{-1} cm^{-1}. (A) Time course of the oxidation of all proteins (2.5 μM) in 0.1 M NaPi buffer pH 7.4. The oxidation of the liposomes follows the characteristic lag period, where oxidation of the lipids shows a slow initiation phase, before a cascade of lipid oxidation. (B) Bar chart illustrating the maximum liposome lipid oxidation rate catalysed by the various proteins.

4.4 Discussion

4.4.1 Effect of pH on the EPR spectrum of Androglobin Globin Domain

Earlier in this work, it was shown that the ferric WT Adgb-GD is primarily hexacoordinated at room temperature. However, it also exhibits pentacoordinate-like properties which is consistent with the optical and kinetic behaviour of Adgb-GD observed in this study. Adgb exhibits a weaker hexacoordinate character in its ferric state compared to other hexacoordinate globins like Ngb, as evidenced by a higher proportion of HS signals. This may relate to a rapid distal endogenous ligand off rate or a slow on rate (or both). These optical features and spin states are similar to the partially pentacoordinate nature observed in ferric Cygb (Beckerson *et al*., 2014). For Cygb, the difference in heme geometry between the optical and EPR spectra was attributed to a freezing artifact. The dissociation of the endogenous ligand in Cygb at the distal site was likely influenced by a temperature-dependent equilibrium between conformers, where electrostatic interactions within the protein became stronger and imposed strain on the Fe-ligand bond (Reeder *et al*., 2011). This may partly explain the discrepancy between the optical and EPR spectra of Adgb. However, it does not fully account for the observed behaviour, as both proteins are hexacoordinate yet exhibit markedly different properties. For example, Adgb displays a very high autooxidation rate (Reeder *et al*., 2024), whereas Cygb shows a much slower autooxidation rate (Ukeri, 2020).

The variation in the ratio of HS to LS signals across different pH levels in Adgb, observed in this study may be functionally relevant, depending on the physiological context in which it occurs. One of the proposed functions of Adgb, described earlier in the introduction section 1.4, was that it may have an important role in the late phases of spermatogenesis (Bracke *et al*., 2018). Spermatogenesis is a complex process that produces functional sperm capable of fertilizing the female gamete and initiating embryonic development. A key

requirement for this is the maintenance of an optimal luminal environment throughout the male reproductive tract, particularly in terms of pH balance (Bernardino *et al*., 2019). Luminal fluid begins forming within the seminiferous tubules and undergoes compositional and pH changes along the tract, from testis to epididymis to support successive stages: spermatogenesis, sperm maturation, capacitation, and fertilization (Bernardino *et al*., 2019). This means that there are fluctuations to this ionic and metabolic homeostasis, especially pH, at different stages. Disruptions to this ionic and metabolic homeostasis, especially pH, can impair sperm development and fertility (Bernardino *et al*., 2019).

Adgb-GD may function similarly to known sperm-protective proteins such as albumin, clusterin and serotransferrin. Given that this study confirms the pH-dependent variation in Adgb's HS/LS signal ratio, it is likely that the globin behaves differently across pH environments, possibly to stabilise its structure or perform its function (Adgb pKa in this study: pK_{a1} = 8.9, pK_{a2} = 10.5, Figure 4.4.1.2). Notably, these studies are with the globin domain in isolation to the other domains, nonetheless it has been proposed that all regions except the globin domain are lost in mature spermatozoa, leaving only the globin domain. Thus, the observed properties of Adgb-GD discovered in this study may enable it to carry out its sperm-protective and redox-regulatory function(s) (Bracke *et al*., 2018). Therefore, this indicates that it could remain active in the epididymis, where the fluid is acidic (Levine & Kelly, 1978). Additionally, Adgb has also been reported to be significantly expressed in the female reproductive tract (Koay *et al*., 2021), reinforcing its potential role across multiple pH environments during reproduction. This is particularly relevant given that the vaginal environment is typically acidic (Bernardino *et al*., 2019).

Overall, the results from the EPR analysis of the mutants show that the proximal and distal mutations do not significantly affect EPR signals. However, both distal and proximal mutants influence the variation in the HS to LS signal ratio across different pH levels, with the

effect being more pronounced with the distal mutants (Figure 4.3.1.5). A hypothesis for this may be as a result of local structural flexibility introduced by the substitution of the distal glutamine. Residues at the B10 position are known to influence ligand binding at the distal site. For example, in Ngb, the B10 phenylalanine contributes to stable oxygen coordination (Ishikawa *et al*., 2007). In the absence of the distal glutamine, another residue within the distal pocket could therefore be considered as a potential functional substitute. One possible logical candidate is the conserved B10 leucine residue (Leu962) (Figure 4.4.1.1), which is also found at the B10 position in other vertebrate heme proteins such as Mb and Hb (Ishikawa *et al*., 2007). However, because leucine lacks an ionisable side chain, it is unlikely to participate in proton transfer or protonation-mediated stabilisation, limiting its ability to functionally replace glutamine in this context.

This hypothesis on the local structural flexibility is supported by the observation that both distal mutants exhibit similar pK values despite involving different substitutions, as shown in Figure 4.4.1.2. This suggests that the mutations are structurally tolerated and may be functionally compensated for. Furthermore, the EPR spectra for the WT HS and LS states were found to be similar in shape at different pH values (Figure 4.3.1.4), indicating that the observed results likely stem from changes in the local heme pocket environment.

Therefore, the variations observed in the signal intensities of the HS and LS spectra (Figure 4.3.1.5), suggest protonation of at least two amino acid residues: one with a pK_a below 7 and another with a pK_a above 7. Although the identities of these residues remain unknown, this behaviour is consistent with the acid–alkaline transitions observed in the distal glycine and distal histidine mutants (Figure 4.4.1.2). While the low-pH transition is only partially evident in the WT, this does not necessarily indicate its absence. The amino acids that could be involved include Cys, His, Tyr, Lys, Arg, Asp and Glu as they all have ionisable side chains.

Reduction of the disulfide bridge in Adgb leads to notable changes in the EPR spectrum, reflecting changes in the heme pocket environment. Specifically, the appearance of LS signal at g = 2.43, suggests a shift in the HS to LS equilibrium toward a more hexacoordinate state across pH conditions. This is in contrast to Ngb, where the reduction of disulfide bond led to the disappearance of the second low-spin signal (Vinck *et al*., 2004). This implies that in Adgb, the reduced form favours stronger axial coordination, likely due to improved accessibility of the distal glutamine to the heme iron. In the oxidised form, the disulfide bridge may restrict glutamine's positioning, limiting its ability to coordinate effectively. Therefore, Adgb may contain a structural linkage between the disulfide bond, the distal pocket, and the positioning of a protonatable residue discussed earlier that could act as a functional substitute.

This can be similar to the interplay observed in Ngb between the disulfide bond, the distal pocket and the positioning of the PheB10 (Doorslaer *et al*., 2014). In Ngb, PheB10 plays a critical role in transmitting structural information from the disulfide bridge to the heme pocket, with the dynamics of this residue playing a gating role. Substituting this phenylalanine with a smaller leucine residue disrupts the van der Waals interactions between PheB10 and the bound CO molecule, leading to altered ligand binding behaviour (Doorslaer *et al*., 2014).

When glutamine is mutated in Adgb-GD, the absence of the glutamine may allow this protonatable residue to occupy the distal coordination site, potentially even competing with glutamine for binding. This interplay may explain the observed optical shifts in the WT and mutants and supports a model in which distal site coordination is dynamically modulated by both redox state and local structural elements. Therefore, the presence of the disulfide bridge in Adgb should modulate its ligand-binding behaviour in a manner distinct from Ngb. In Ngb, the endogenous distal HisE7 competes with exogenous ligands for coordination to the heme iron, resulting in slower ligand binding (Doorslaer *et al*., 2014). In contrast, the disulfide bridge in Adgb may introduce steric hindrance that limits access of the distal glutamine to the heme,

thereby reducing competition and allowing faster binding of small exogenous ligands. This differential behaviour is explored throughout this chapter. Alternatively, the HS signal observed in both the reduced and non-reduced form could also be attributed to the pentacoordinate HisF8–Fe(III) species or to the aquomet form of the protein.

While a crystal structure would be ideal to confirm or refute these hypotheses in future studies, in its absence, resonance Raman spectroscopy might provide details of the coordination environment.

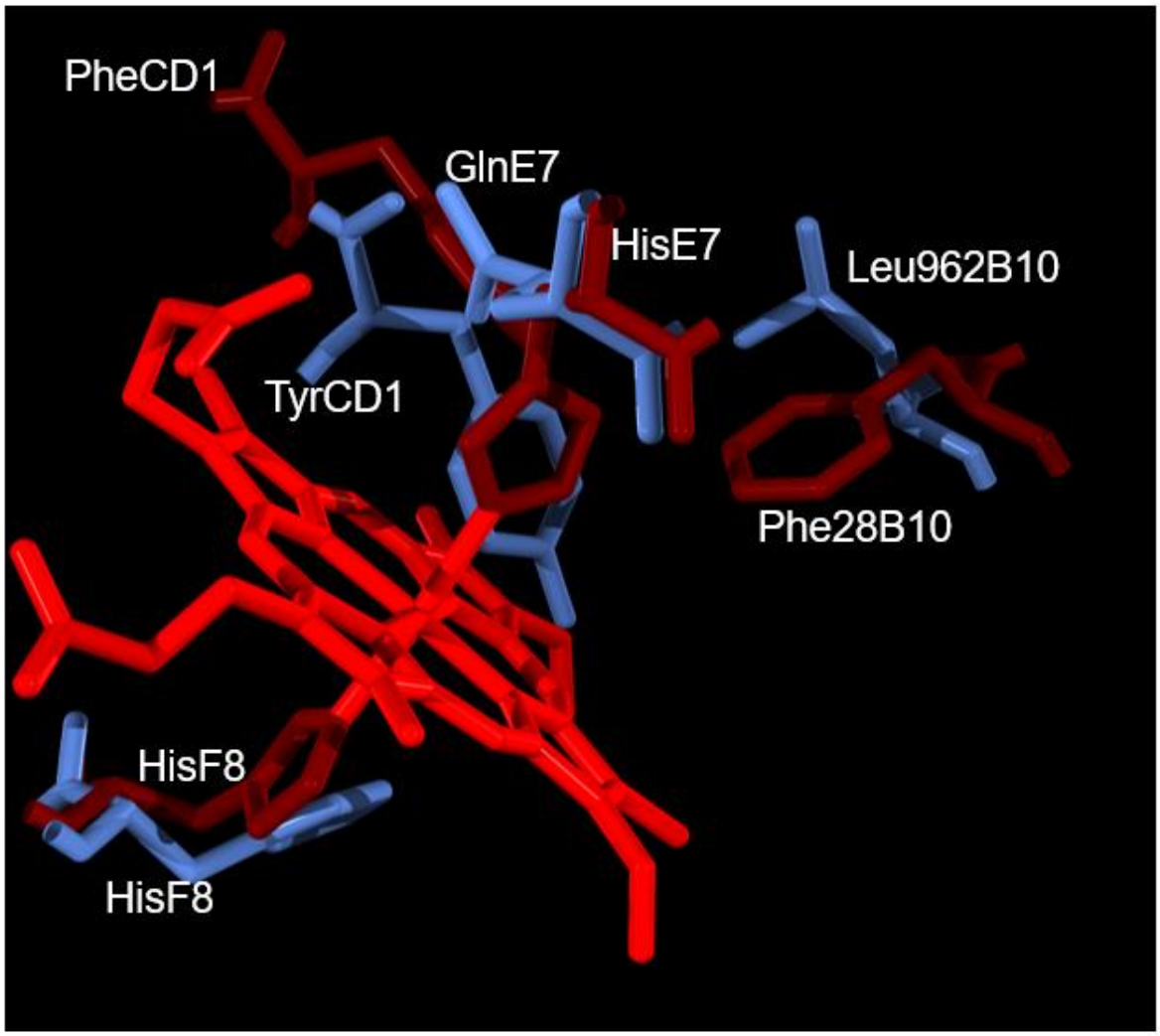


Figure 4.4.1.1 Show the key residues of Adgb including the B10 residue from the revised sequence (Reeder *et al*., 2024) in blue, compared with hNgb (PDB ID: 1OJ6) in dark red.

The B10 residue of both Adgb and Ngb are Leu962 and Phe28 respectively.

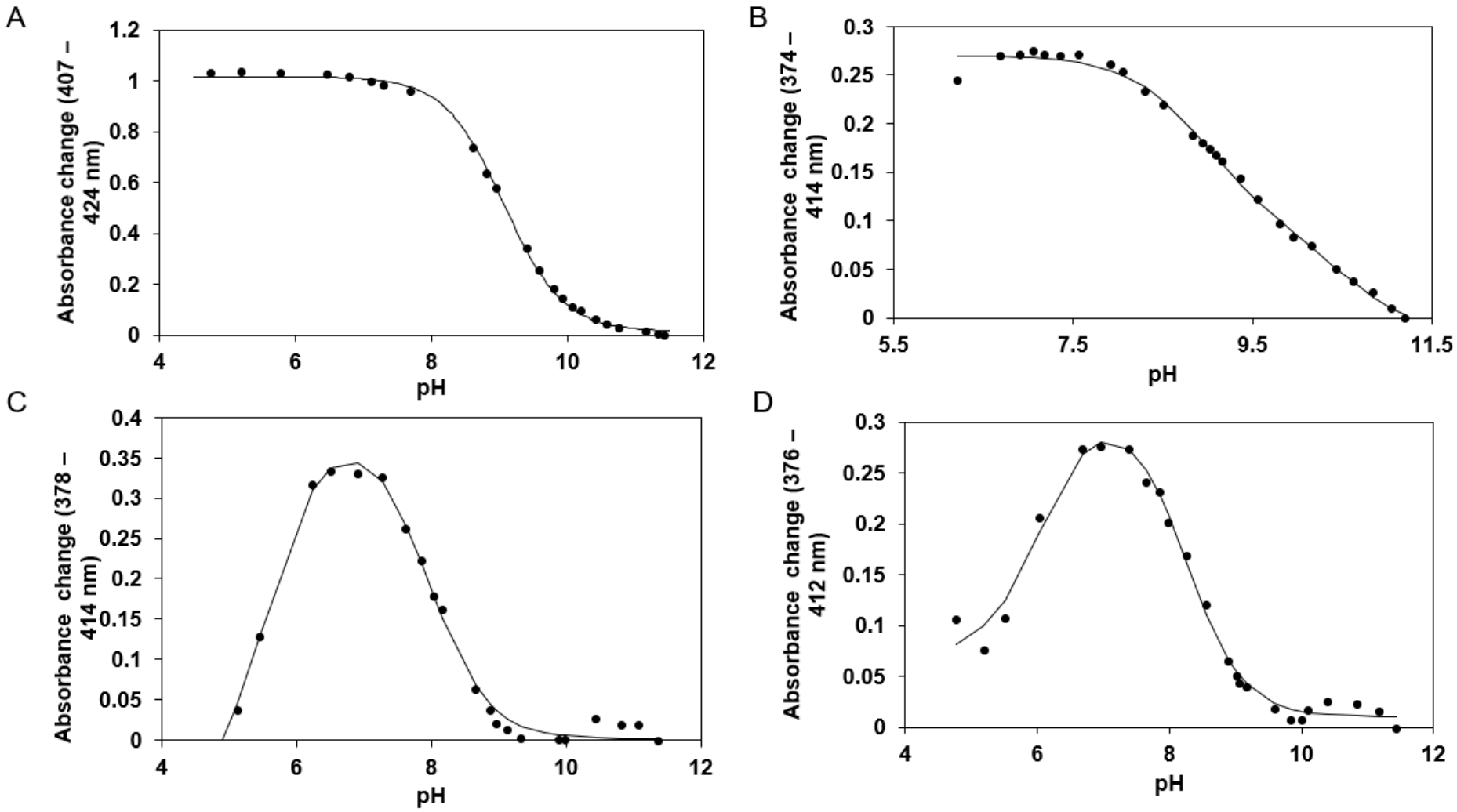


Figure 4.4.1.2 The acid–alkaline transition of ferric heme proteins (5 μM).

Acid (HCl 0.05 – 5 M) was titrated into the solution containing ferric proteins that has been adjusted to approximately pH > 11. (A) Mb: pk = 9.06. (B) Adgb: pK_{a1} = 8.95, pK_{a2} = 10.52. (C) Adgb mutant Q792G: pK_{a1} = 7.95, pK_{a2} = 5.50. (D) Adgb mutant Q792H: pK_{a1} = 8.24, pK_{a2} = 6.04.

4.4.2 Ligand Binding

The rapid bimolecular phase observed for both CO and CN binding in Adgb WT corresponds to a competitive binding of CO or CN with the distal residue. The CO binding behaviour between the WT and disulfide reduced proteins align with the expected binding profile. In the WT protein, for the fraction that binds glutamine, the return to the final CO-bound state involves a faster dissociation of glutamine from the distal site, thereby allowing quicker CO association. This observation supports the previously proposed hypothesis that the disulfide bridge may hinder the interaction of the distal residue with the heme iron, facilitating a more rapid dissociation of the distal residue. Notably, after reduction of the disulfide bond using TCEP, a significant change was observed in Adgb: the dissociation of distal glutamine became slower, leading to delayed ligand association. This result underscores the role of the disulfide bridge in modulating ligand binding dynamics.

As shown in Section 1.3.2, Figure 1.3.2.2, two cysteine residues are structurally positioned to form an intramolecular disulfide bridge, which may influence the structure and dynamics of the CD region of the protein. This region is in close proximity to the heme distal site and could thereby affect exogenous ligand-binding activity. They are Cys 978 on α-helices A to C and Cys 787 on the D to H α-helices (Reeder *et al*., 2024), with predicted distances of 2.71 Å when the axial positions of the sulphides on the Cys were shifted. The reduction of the disulfide bond also revealed the presence of distinct conformers in Adgb. The disulfide bridge might be important in stabilising the heme pocket structure through forming an ‘artificial’ CD loop, normally absent because of the circular permutation of the globin domain for Adgb (Reeder *et al*., 2024). Therefore, the absence of this disulfide bond means that there are conformational variants that are optically indistinguishable, in which residues nearby amino acids are protonated and may replace the distal GlnE7, contributing to the distinct slow kinetic phases observed. These alternative conformations may also be present in the WT protein, but

likely at lower abundance due to the stabilising effect of the disulfide bond on the heme pocket. This indicates that the disulfide bond serves a bifunctional role: it stabilises the heme pocket architecture while also restricting the accessibility or positioning of GlnE7. This is not the first instance of heterogeneity reported in hexacoordinate globin samples; a similar phenomenon has been observed in Ngb, both in the presence and absence of the disulfide bond (Hamdane *et al*., 2003).

The small difference in the rate of CN binding for Adgb compared to Mb, relative to the large difference in Ngb (Table 4.3.2.1), indicates weak distal GlnE7 binding to the iron within Adgb. Similar results were also observed for the mutants and may be attributed to the nearby protonated residues described earlier in Section 4.4.1. These residues are capable of compensating for the Gln at the distal site especially if the amino acid residue is as small as glycine. Similar results were observed when HisE7 in human Hb was replaced by Gly, with no significant change in the free energy of the overall transition state for isonitrile binding (Mathews *et al*., 1989). Therefore, this suggests that the barrier to isonitrile entering and exiting the distal pocket is either physically removed from the immediate vicinity of the iron atom and the E7/E11 residues, or involves a more global structural rearrangement such as an expansion of the distal pocket accompanied by movement of residues like LeuB10 and/or PheCD1 (Mathews *et al*., 1989).

Nonetheless, several studies have shown that the B10 amino acid can influence both the geometry and ligand-binding characteristics of the distal heme pocket. For example, in *Ascaris* Hb, a strong hydrogen bond between TyrB10 and the distal oxygen of the ligand, coupled with a weaker hydrogen bond between GlnE7 and the proximal oxygen, serves to stabilise the ligand within the binding pocket (Yang *et al*., 1995).

The results of this study raise the question of whether the distal glutamine plays a less significant role than previously thought, and whether the weak hydrogen bond to the ligand or

the hydrogen bond between E7 and B10 residues is more critical in stabilising the heme pocket. Moreover, the results could indicate that the mixed coordination described for Adgb could be attributed to Adgb binding their ligand in a pentacoordinate manner, driven by the presence of the disulfide bond, rather than representing a true mixed-ligand species.

4.4.3 Implications for *in vivo* functional roles for Adgb

One of the physiological roles frequently attributed to many globins is their involvement in NO redox functions. The findings of this study support this role for Adgb, indicating a strong likelihood that Adgb contributes to both the release and regulation of NO levels *in vivo*. It has been suggested that Adgb-GD could be the sperm oxidase responsible for the production of superoxide anion or nitric oxide during capacitation in eukaryotes (Bracke *et al*., 2018), see also section 1.4. The high efficiency of Adgb as a nitrite reductase supports this hypothesis, as it appears to exhibit greater nitrite reductase activity than even Mb. However, further investigation is required to confirm this theory. For example, by examining whether the Adgb globin domain is expressed during the capacitance phase of spermatogenesis. This work also showed that Adgb could also work as a NO scavenger in the reproductive tract, analogous to the NO-scavenging role of Mb in the heart. This is supported by its comparable NO binding rate to both Mb and Hb (Hankeln *et al*., 2008). The tight regulation of NO during spermatogenesis further supports this role, as evidenced by the presence of four nitric oxide synthase (NOS) enzymes in the testes instead of the normal three (endothelial NOS (eNOS), inducible NOS (iNOS), and neuronal NOS (nNOS) and the additional testis-specific NOS (tnNOS)) (Lee and Cheng, 2009).

The precise mechanism underlying Adgb's nitrite reductase activity remains unclear e.g. whether NO binds on the distal or proximal site. However, mutations at both the proximal

and distal sites did not significantly alter the nitrite reductase behaviour of Adgb-GD. This observation may be attributed to the influence of the protonated residues within the heme pocket, suggesting that other amino acid residues contribute to maintaining the functional integrity of the active site. The unexpected preservation of activity in the Q792H and H824Y mutants remains puzzling. It is possible that the overall structure of the protein remains sufficiently robust to sustain enzymatic function despite local structural alterations in the heme pocket. To better understand the behaviour, future studies should investigate the impact of disulfide bond reduction on nitrite reductase activity. This would be valuable, particularly in determining whether it impairs redox regulation and enzymatic function or has negligible effects.

NO binding to Adgb had been proposed to resemble that of cytochrome c prime (cyt c′) (Reeder *et al*., 2024). Where the key optical feature is the appearance of an optical band at ~390 nm in the NO-bound ferrous form, indicative of a five-coordinate NO-bound species as observed in cyt c′ (Reeder *et al*., 2024). Interestingly in cyt c′, NO can only bind to the distal iron location of cyt c′ following a conformational change involving an occlusion in the heme pocket (by a Phe residue in the case of cyt c′) (Manole *et al*., 2015). These conformations are optically indistinguishable (Manole *et al*., 2015). This behaviour closely parallels our hypothesis regarding Adgb, wherein structural elements such as the disulfide bridge, GlnE7, and nearby (protonated) residues may create steric hindrance that influences ligand access, particularly to the distal E7 site. This suggests a regulatory mechanism potentially mediated by conformational gating (Reeder *et al*., 2024). Crystal structures of the NO bound protein or further experiments are required to definitively determine whether NO binds at the proximal or distal site in Adgb.

The mechanism of NO scavenging by Adgb *in vivo* is also unclear. Bacterial flavohemoglobins (Hmp) are the best-characterised NO-detoxifying enzymes (Shepherd *et al*.,

2011), operating via two distinct catalytic pathways. Under aerobic conditions, Hmp functions primarily as a NO dioxygenase (NOD), using NADH as an electron source (Gardner *et al*., 1998). In this pathway, ferrous heme binds O_2 to form an oxy-ferrous (Fe^{2+}–O_2) complex, which reacts rapidly with NO to yield nitrate (NO_3^-) as the major product (Hausladen *et al*., 1998). Under low O_2 and high NO conditions, Hmp can instead operate via a denitrosylase (reductive) pathway, where NO binds to ferrous heme to form a Fe^{2+}–NO complex that is subsequently reduced via NADH-dependent electron transfer, regenerating free heme and producing reduced nitrogen species, including N_2O under some conditions (Kim *et al*., 1999).

In vivo, the pathway utilised by Adgb is likely to depend on oxygen and NO availability, as well as ligand-binding affinities. Evidence from related globins supports a dynamic shift between mechanisms. For example, in *Campylobacter jejuni* globin (Cgb), NO detoxification is proposed to proceed via a dioxygenase mechanism during early nitrosative stress, when O_2 is available, and shift toward reductive pathways during prolonged NO exposure as O_2 becomes limiting (Shepherd *et al*., 2011). Further insight into potential NO reactivity comes from mammalian Hb, where multiple competing NO fates have been characterised. Under physiological conditions, most Hb is oxygenated, so NO binds primarily to a small fraction of vacant heme sites, promoting both dioxygenation to nitrate and S-nitrosylation of cysteine residues to form S-nitrosohaemoglobin (Gow *et al*., 1999). While dioxygenation represents the dominant route of NO clearance *in vivo* (Wennmalm *et al*., 1992; Kelm, 1999), the formation of nitrosyl and S-nitrosothiol species demonstrates that globins can partition NO between detoxification and storage pathways.

Together, these observations suggest that, unlike the specialised detoxification role of Hmp, Adgb may support multiple NO fates *in vivo*, with its dominant mechanism determined by cellular O_2 and NO conditions.

The liposome oxidation study showed that Adgb could potentially act as a lipid oxidation enzyme, potentially leading to the production of powerful signalling modules. However, it is slow compared to Cygb (Reeder *et al*., 2011) and is similar to Mb that we know only induced lipid oxidation under pathological conditions (Wilson and Reeder, 2022). So, although such a physiological role cannot be ruled out, it seems unlikely. Given that Adgb-GD exhibits a tendency to aggregate, future experiments could explore the effect of varying NaCl concentrations on its kinetic behaviour. Ideally, conducting kinetic analyses using the full-length Adgb protein, or the globin domain's natural binding partner (if one exists) would be the most appropriate approach to better reflect physiological conditions.

CHAPTER 5: Exploring molecular structure and gene expression of Androglobin Globin and N-terminal domains

5.1 Introduction

Proteins are essential to life through the reactions they perform, the conformations they adopt, and the interactions they make in solution. This is why it is important to have detailed knowledge of the three-dimensional tertiary or quaternary configuration of protein in order to fully understand their behaviour. For proteins such as Adgb, whose function is largely unknown and for which current *in silico* models have low confidence, a better understanding of their structure can significantly improve insights into their function or guide experiments to explore potential physiological roles.

X-ray crystallography is one of the most powerful and versatile method used for determining the tertiary and quaternary structure of proteins. It employs coherent X-ray scattering, where X-rays are used to examine the arrangement of atoms in crystals through non-destructive methods (Nachiappan *et al*., 2021). Some X-ray damage to the crystal can occur, depending on the energy used. Modern facilities like the Diamond Light Source use extremely high-energy X-rays, which allow image capture in milliseconds. However, the downside is that radiation can destroy the crystal within seconds. To manage this, the crystal's quality is continuously monitored, allowing exclusion of any images taken after deterioration (Garman, 2010). The crystalline atoms diffract the X-rays in specific directions i.e. it's the electrons of the atoms which scatter the X-rays. The intensity and angles of these diffracted beams produce an electron density map. This map reveals the average positions of the atoms in the crystal, their chemical bonds, and any structural disorder (Nachiappan *et al*., 2021). The smallest repeating unit of a crystal is known as the unit cell, and its shape and size are determined by the lengths of three axes (a, b, and c) and the angles between them (α, β, and γ) in three-dimensional space. The molecules in X-ray crystallography are highly organised within a crystal lattice thus their diffraction pattern is sharp, and their peaks are well-defined.

For proteins such as Adgb, where crystal structures are currently not available and have proven challenging to obtain, small-angle X-ray scattering (SAXS) may provide another pathway to obtain structural information. SAXS is a powerful technique which can provide structural information on proteins in solution with a resolution range from roughly 50 Å to 10 Å (Putnam *et al*., 2007). SAXS can quickly characterise the shape and conformation of macromolecules in solution, from small to very large systems, filling the gap between Nuclear Magnetic Resonance (NMR) and Electron microscopy (EM) methods. This technique provides additional information e.g. conformational dynamics that is useful both before and after high-resolution structures are obtained (Putnam *et al*., 2007). Additionally, SAXS can determine low-resolution structures without requiring supplementary experimental data. This is a significant advantage for our study because it allows us to analyse the overall shape and conformational flexibility of Adgb N-terminal and globin domain in solution, even in the absence of high-resolution structural information or well-diffracting crystals. It's also an advantage in this study due to its ability to handle samples that are difficult to crystallise or are conformationally heterogeneous. Similar to X-ray crystallography, SAXS also employs coherent X-ray scattering, and thus are fundamentally similar techniques. Consequently, both techniques can utilise the same equipment for generating, preparing, and detecting X-rays. However, they differ in the arrangement of target molecules during data collection. The arrangement in X-ray crystallography has been described above to be highly organised within a crystal lattice. In contrast, scattering is continuous and radially symmetric in SAXS as the signals from all orientations of the target molecules, relative to one another and the experimental apparatus, are averaged together (Putnam *et al*., 2007).

X-ray crystallography provides high signal strength and atomic level structural detail, surpassing the resolution achievable with SAXS. However, the crystal lattice environment can impose non-biological packing interactions that may lock proteins like Adgb into non-native

conformations (Putnam *et al*., 2007). Since proteins are inherently dynamic, these forced static conformations may not accurately reflect their functional states *in vivo*. In contrast, SAXS enables structural analysis in solution, preserving the native conformational ensemble of the protein. This makes SAXS particularly advantageous for capturing the dynamic features relevant to Adgb's overall domain structure that is currently lacking in *in silico* models of Adgb such as AlphaFold.

5.2 Methodology

5.2.1 Sample preparation

The Adgb N-terminal and Adgb-GD (holo and apo) proteins were prepared following the procedures outlined in chapter 2.2.2 of this study. After purification, both samples were concentrated to 10 – 20 mg/ml in 1 mM borax buffer at pH 10 (post-dialysis) for their respective SAXS and crystallography experiments. All Adgb globin domain proteins are all purified using the method described in chapter 2.2.3.

5.2.2 Low resolution structure analysis of the N-terminal and globin domains of Androglobin using SAXS

SAXS was performed at Diamond Light Source Ltd, Beamline B21, Harwell Science and Innovation Campus Didcot, Oxfordshire. The technique used was size exclusion chromatography small-angle X-ray scattering (SEC-SAXS). The column used was the Shodex KW-402.5 column with size range between 5000 – 100000 and flow rate 0.16 ml/min in the HPLC.

Protein solution scattering was measured from fresh samples of Adgb N-terminal (20 mg/ml), Apo Adgb-GD (10 mg/ml) and Adgb-GD (10 mg/ml and 20 mg/ml) in 0.1 M NaPi 7.4 (final concentration). Background scattering was measured from the size exclusion column equilibrated in the buffer before the protein was eluted. Bovine serum albumin (BSA) was used as control and reference samples used to correct for background scattering.

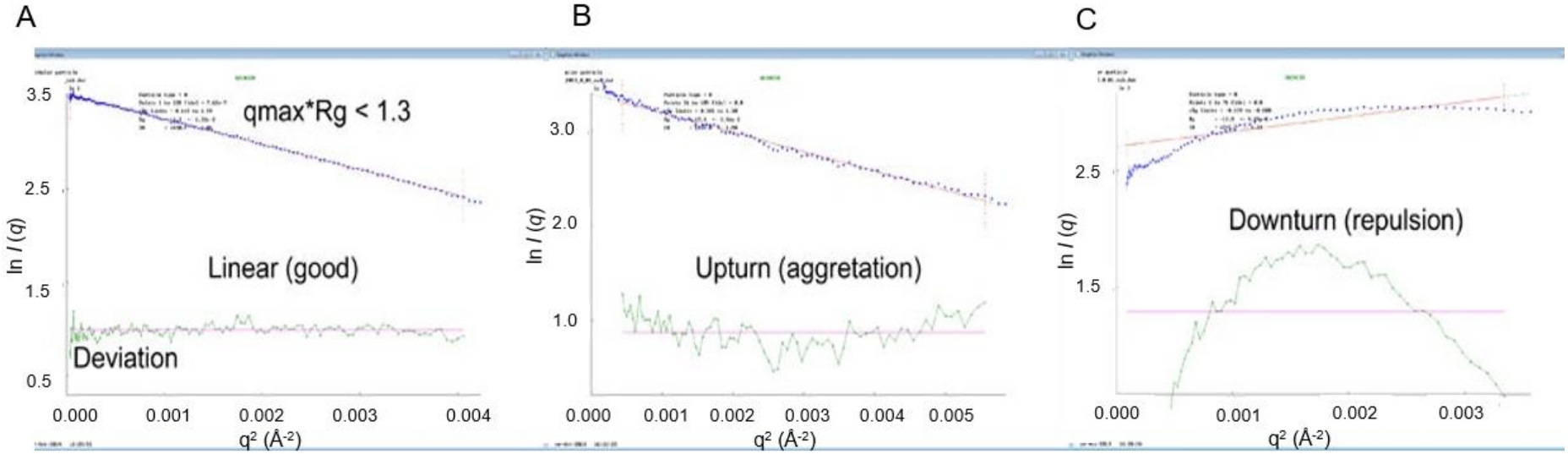
Following data collection, appropriate scattering of the buffer was subtracted from the sample data to isolate the scattering from the particle of interest. This was done by inspecting the buffer and sample data for any mismatch. Guinier plot was used to analyse the SAXS

experimental data via ATSAS including Bayesian inference. Guinier plot gives an estimation of the radius of gyration (Rg) and the forward scattered intensity (I(o)) thus determining the overall size the protein. The Guinier plot also suggests the behaviour of the molecule (Figure 5.2.2.1), where (*Data Analysis Primer | Small Angle X-ray Scattering/Diffraction*, October 2024):

Straight line = proteins are behaving well under the X-ray beam.

If a straight line is not obtained, there are two possible scenarios:

- Upturn shape at low q (q = scattering vector) would be aggregation, radiation damage, inter-particle interactions and/or high polydispersity of the sample protein.
- Downturn shape at low angle would be inter-particle repulsion of the sample.



5.2.2.1. This image shows a Guinier plot of how the scattering of proteins would look like after Guinier analysis using lysozyme as an example.

(A) Good linear scattering suggests good data, (B) Upturn scattering suggests aggregation and finally (C) Downturn scattering suggests repulsion. This image is reproduced from the Stanford Synchrotron Radiation Lightsource's Small Angle X-ray Scattering/Diffraction Data Analysis Primer (October 2024).

5.2.3 Protein crystallography of the N-terminal and globin domains of Androglobin

Following preparation of Adgb N-terminal and Adgb-GD (holo and apo), the samples were then concentrated to approximately 20 mg/ml using a Millipore spin concentrator with 5 kDa MWCO (section 5.2.1). The crystallisation method used to grow crystals for these proteins is the sitting drop vapour diffusion method i.e. Mrc 2-drop plates set up using an Art Robbins Instruments (ARI) Gryphon robot. Briefly, 0.2 µl of protein solution at 20 mg/ml was mixed with an equal volume of pre-made precipitant screening solutions (PEG ion (from Hampton Research), NeXtal pH clear, basic crystal screen (BCS) 1, BCS 2, and Structure screen JCSG-Plus crystallisation screens from Molecular Dimensions). Each screen included 96 variations of precipitants. The reservoir chamber contained 60 µl of precipitant. 96 well plates were from Molecular Dimensions and contained 2 well chambers per reservoir chamber. The plates were sealed and initially left overnight in a cold room and then observed daily for 10 days and periodically thereafter for up to 8 weeks to monitor crystal formation. Images were taken using a Leica microscope fitted with a digital camera.

5.3 Results

5.3.1 Crystallisation of Androglobin domains

To explore the structure of Adgb, specific sections of this large protein were isolated and expressed in order to attempt to crystallise these domains. This was done by the sitting drop vapour diffusion method facilitated by an ARI Gryphon robot (see methods, section 5.2.3). Of the five crystallisation screening kits tested (comprising 480 conditions in total), only two yielded promising results after multiple trials involving variations in protein concentration and purification.

Figures 5.3.1.1 to 5.3.1.3 display images of the crystals formed. Although these appear to be crystals, they may in fact be salt deposits given the extended incubation period of 8 weeks. Figure 5.3.1.1 shows a crystal formed by mixing holo Adgb-GD (20 mg/ml) with a pH-clear solution (0.1 M BICINE buffer, 65% (v/v) MPD, final pH 9.0) in equal volumes (0.2 µL). Figure 5.3.1.2 and Figure 5.3.1.3 depict what appears to be microcrystals of holo Adgb-GD obtained by mixing Adgb-GD (20 mg/ml) with PEG ion solutions (salt: 0.2 M Lithium chloride, 20% (w/v) PEG 3350, and salt: 0.2 M Lithium acetate, 20% (w/v) PEG 3350 respectively). However, as the crystals are colourless in Figure 5.3.1.2 it is possible that the protein is in the apo form or that these are largely salt crystals. Further optimisation of the crystallisation conditions is necessary to confirm the presence of microcrystals using these precipitant solutions. No crystals were observed for the N-terminal domain or apo-Adgb GD.

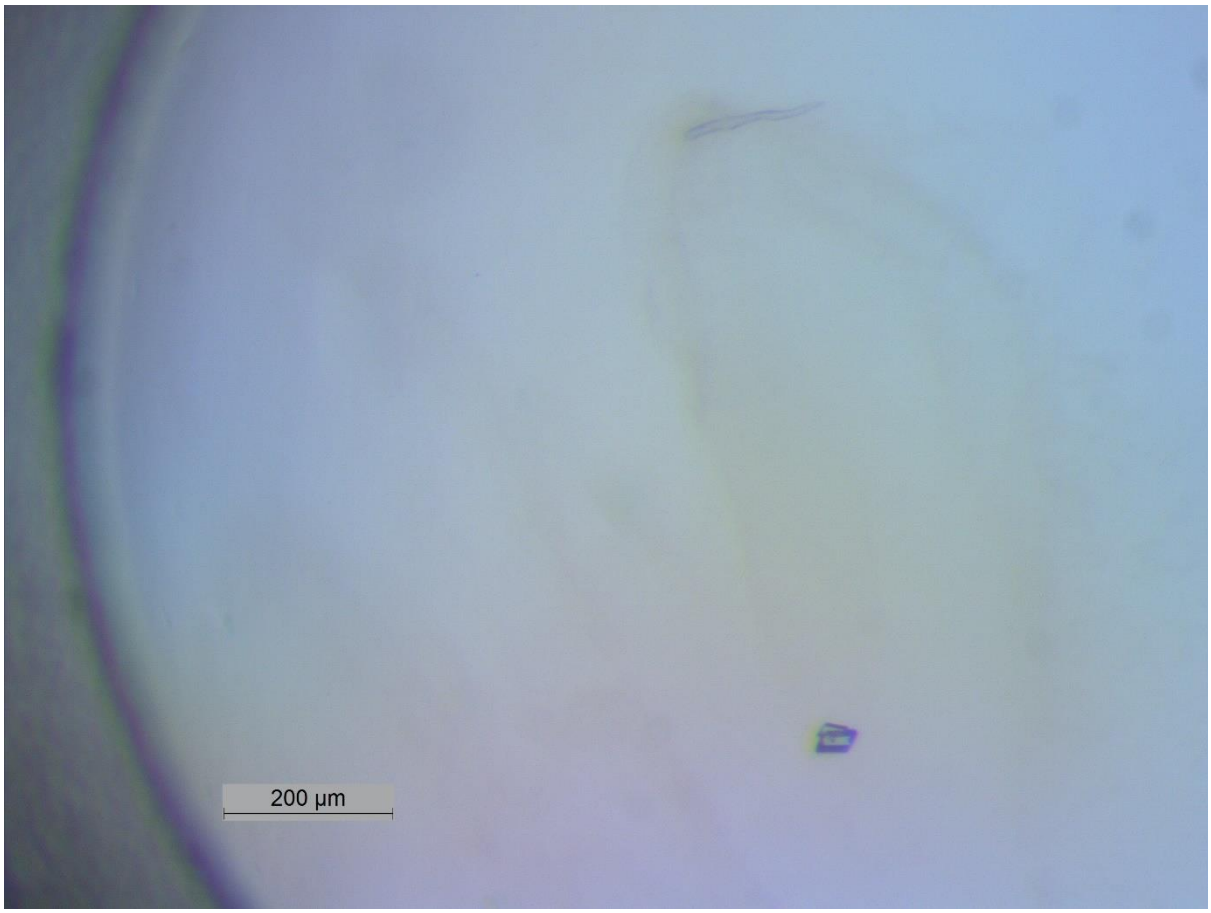


Figure 5.3.1.1 Image of a crystal taken with Leica LAS EZ software after 8 weeks. Generally crystal grow from a few hours to 2-3 weeks. This length of time suggests salt crystals. This crystallisation plate contains holo Adgb-GD mixed with precipitant solution pH clear (0.1 M BICINE buffer, 65% (v/v) MPD, final pH 9.0) in equal volume.

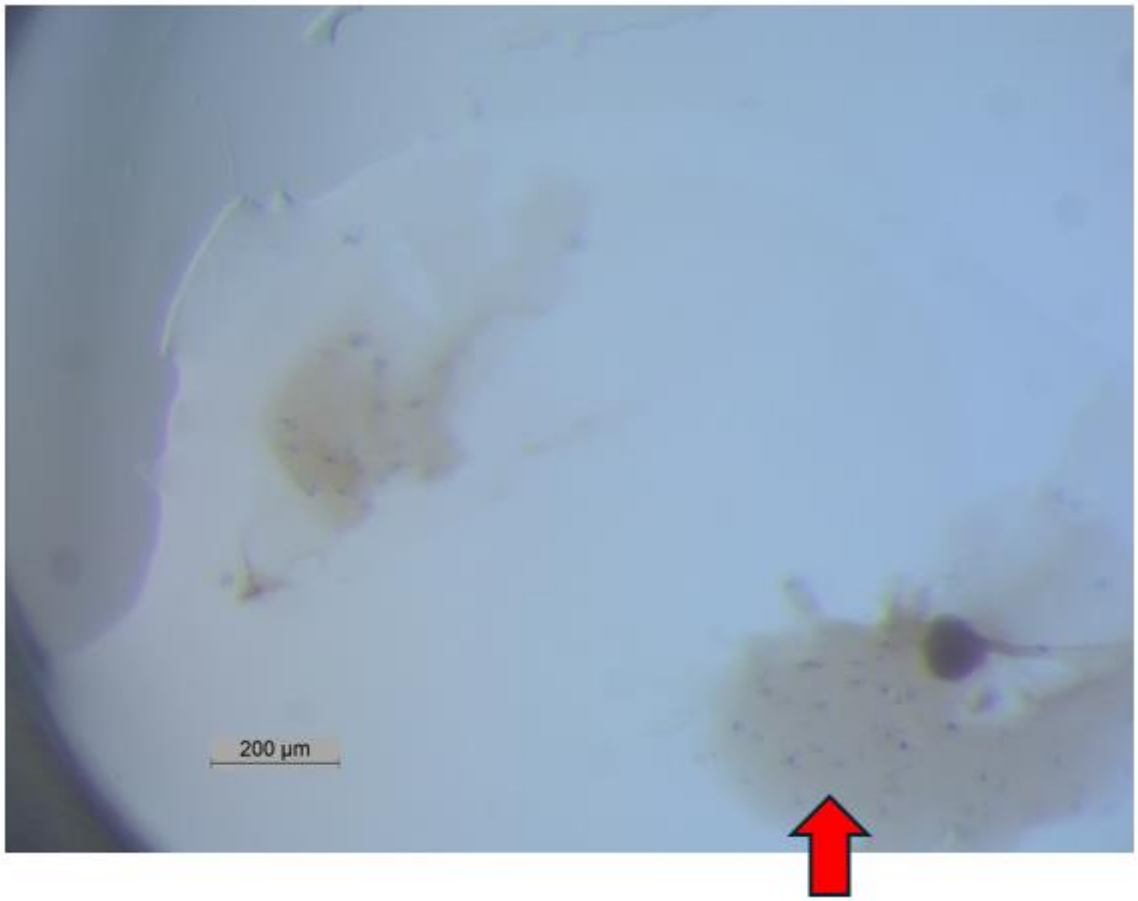


Figure 5.3.1.2 Image of microcrystals in the precipitate taken with Leica LAS EZ software after 8 weeks. This crystallisation plate contains holo Adgb-GD mixed with precipitant solution PEG ion (0.2 M Lithium chloride, 20% (w/v) PEG 3350) in equal volume.

Figure 5.3.1.3 Image of amorphous precipitate which might be crystalline in nature taken with Leica LAS EZ software after 8 weeks. This crystallisation plate contains holo Adgb-GD mixed with precipitant solution PEG ion (0.2 M Lithium acetate, 20% (w/v) PEG 3350) in equal volume.

5.3.2 Exploring the oligomeric state of Androglobin N-terminal domain and Globin domain in solution using small-angle X-ray scattering

Crystallography attempts were unsuccessful, prompting further investigation of the proteins (N-terminal and Adgb-GD) in solution using SAXS. This analysis revealed that both proteins were aggregated in solution, likely explaining the failure to obtain suitable crystals for X-ray diffraction.

To determine the oligomeric state of Adgb domains in solution and potential low resolution structural information, SAXS data were collected at a range of concentration for Adgb domains (10 – 20 mg/ml) in 0.1 M NaPi 7.4. Figure 5.3.2.1 show the running time of BSA. The CHROMIX-SAXS profile in Figure 5.3.2.1 show multiple peaks which indicate heterogeneity in the BSA sample used. The experimental BSA SAXS data also gave a larger radius of gyration (Rg) 4.768 ± 1.60 nm than its normal range for monomeric BSA 3.2 – 3.8 as shown in Figure 5.3.2.1. This strongly suggests that the BSA sample used is not in a monomeric state, probably similar to the BSA sample in Figure 2.3.4.1. Moreover, a study by Zhang *et al*., 2006 indicate that BSA can have varying Rg as their calculated Rg for BSA was 2.76 nm. Additionally, the Guinier analysis of the experimental SAXS data showed equal distribution (straight line) which indicate that the data is a good Guinier fit for the scattering profile of BSA (Figure 5.3.2.1). The experimental SAXS data gave 63.8 kDa as BSA Mw which is in close agreement with the literature value of BSA according to the Small Angle Scattering Biological Data Bank (SASBDB) which is 66 kDa (70 kDa for SEC-SANS) (Morishima *et al*., 2025). Overall, the experimental BSA SAXS data from the BSA sample indicate that the SEC-SAXS technique performed as intended.

The CHROMIX-SAXS profile for the experimental data of 10 mg/ml Adgb-GD data show a picture that suggests the sample is homogenous (Figure 5.3.2.2). The Rg for Adgb-GD

is 14.256 ± 0.2 nm which is much larger than typical values for most monomeric small and medium size proteins, including globular proteins like BSA (Figure 5.3.2.1) or Mb (~ 1.7 nm; Levantino *et al*., 2015). The value obtained for Adgb-GD likely reflects large protein complexes or a multimer species rather than a monomeric form, explaining the high Rg of the protein. Additionally, the Guinier analysis of the Adgb-GD experimental SAXS data, shown in Figure 5.3.2.2, exhibited an upturn shape rather than a straight line (see Figure 5.2.2.1). This indicated the presence of inter-particle interactions or protein aggregation which hindered further analysis. This is further supported by the predicted Mw of the protein (Bayesian Inference) which is ~873.05 kDa (Figure 5.3.2.2), compared to the predicted monomeric Adgb Mw of 25.9 kDa (Expasy). Similar results were obtained at a higher Adgb-GD concentration of 20 mg/ml, as evident in the CHROMIX-SAXS profile shown in Figure 5.3.2.3A. Unequal distribution with an upturn shape was also observed with Rg of 17.66 ± 0.08 nm (figure 5.3.2.3A) which suggests inter-particle interactions or aggregation of the protein. The results from the experimental apo-Adgb-GD SAXS data are shown in Figures 5.3.2.3B with Rg of 16.64 ± 0.26 nm. The Rg is also similar to the Rg of the normal WT Adgb-GD given in the Figures 5.3.2.2 and 5.3.2.3A. The Guinier analysis also indicated inter-particle interactions or aggregation in the protein. Further analysis revealed that both the apo-Adgb-GD (10 mg/ml) and Adgb-GD (20 mg/ml) exhibited predicted molecular weights similar to the WT Adgb-GD at the lower concentration (10 mg/ml) i.e. ~873.05 kDa. Overall, both Adgb-GD holo and apo displayed comparable behaviours.

Figure 5.3.2.4 show the CHROMIX-SAXS profile of Adgb N-terminal domain. The plot shows multiple peaks similarly to that of BSA which indicate the heterogeneity of the sample. However, the analyses were more similar to Adgb-GD. The experimental data produced high Rg of 11.445 ± 0.22 nm (Figure 5.3.2.4) and thus indicating that it is also a heterogeneous aggregated protein. This is supported by the Guinier analysis in figure 5.3.2.4

that show unequal distribution as well as an upturn shape in the plot. Bayesian inference was unable to predict the Mw of the protein. In conclusion, both Adgb-GD and Adgb N-terminal showed inter-particle effects, highlighting the need for further refinement of their purification protocols and buffer conditions.

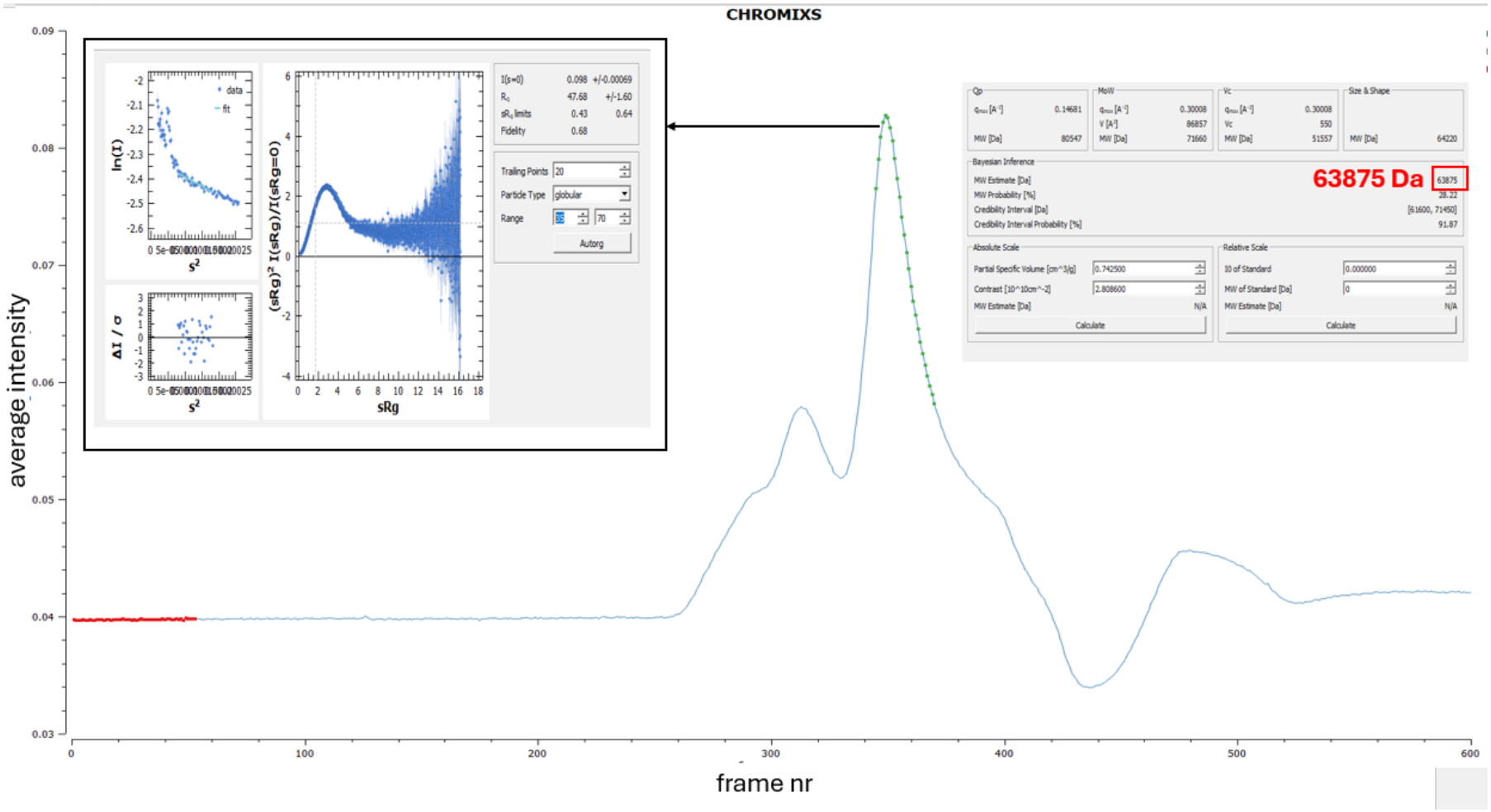


Figure. 5.3.2.1 Size exclusion chromatography small-angle X-ray scattering (SEC-SAXS) profile for bovine serum albumin (BSA).

CHROMIXS profile shows SAXS curve used for analysis in the Guinier analysis in the boxes. Where green shows sample data selected for Guinier analysis and red shows data selected for buffer reference. The first 30-odd data points were removed as this X-rays was not fully blocked by the backstop. The line shows the fit to the data for the Guinier and the residual show no significant up/downturn. The estimated Mw from Bayesian inference was 63.8 kDa. The literature value according to the SASBDB lists BSA with a monomeric Mw of 66 kDa.

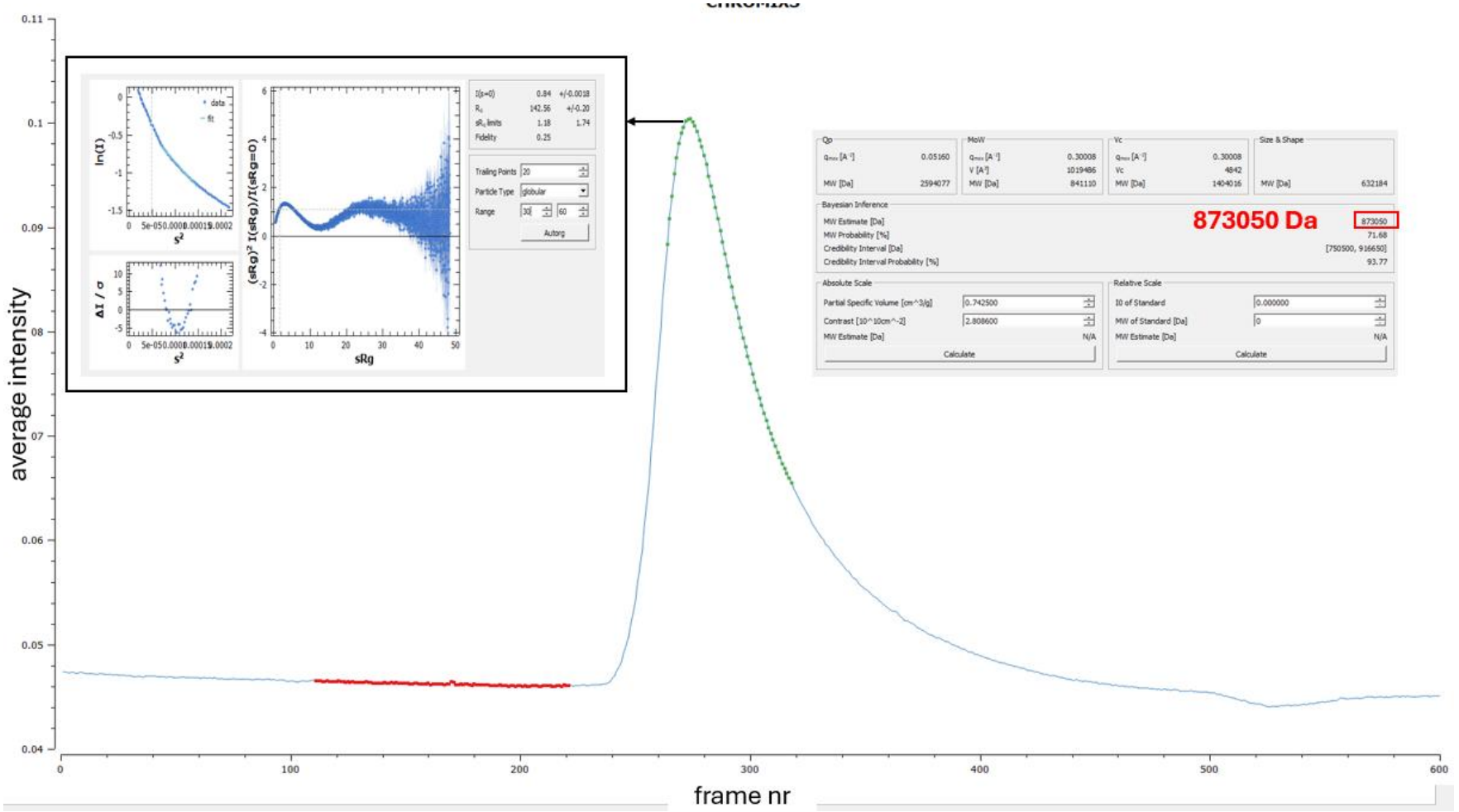


Fig. 5.3.2.2. Size exclusion chromatography small-angle X-ray scattering (SEC-SAXS) profile for Adgb globin domain (10 mg/ml).

Where the CHROMIXS profile is showing the SAXS curve (green) used for analysis in the Guinier analysis (box). The Guinier analysis show a sub-optimal Guinier fit for the scattering profile of Adgb globin domain and the residual show significant up turn. Therefore, this scattering could be characteristic of multimer formation or aggregation in the protein. The estimated Mw from Bayesian inference was 873.05 kDa compared to the theoretical Mw 25.9 kDa from Expasy server (https://web.expasy.org/compute_pi/).

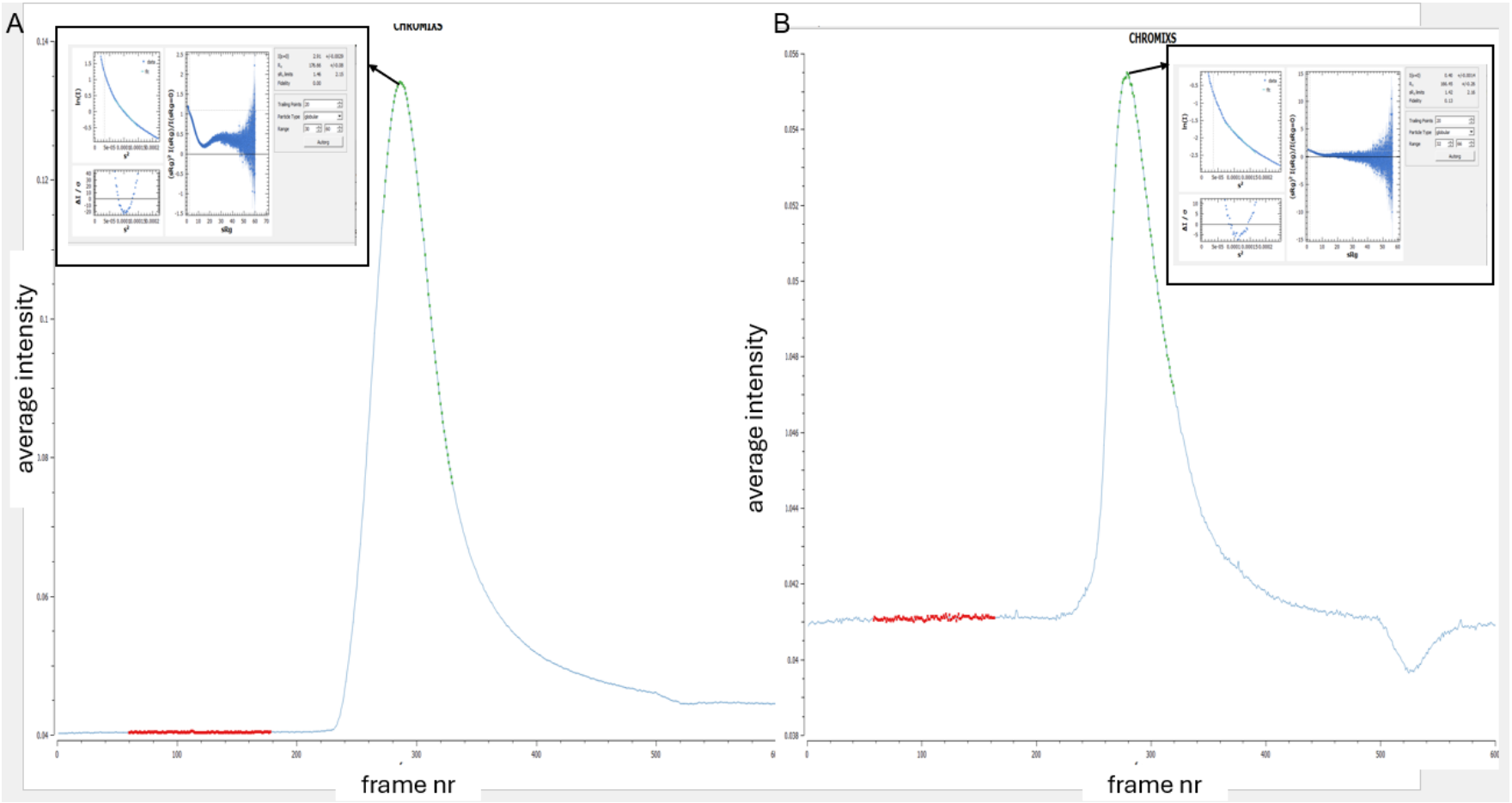


Fig. 5.3.2.3. The size exclusion chromatography small-angle X-ray scattering (SEC-SAXS) profile was recorded for the Apo form of the Adgb globin domain at 10 mg/ml and for the WT Adgb globin domain at a higher concentration of 20 mg/ml.

(A) CHROMIXS profile of Adgb globin domain show similar Guinier fit for the scattering profile as WT Adgb globin domain (10 mg/ml) even at higher concentration. (B) CHROMIXS profile of Apo form of Adgb show a bad Guinier fit for the scattering profile of Apo form of Adgb and the residual show significant up turn. Estimated Mw predicted was approximately similar with Mw as Figure 5.3.2. The scattering in the Guinier analysis of both proteins could be a characteristic of aggregation in the protein.

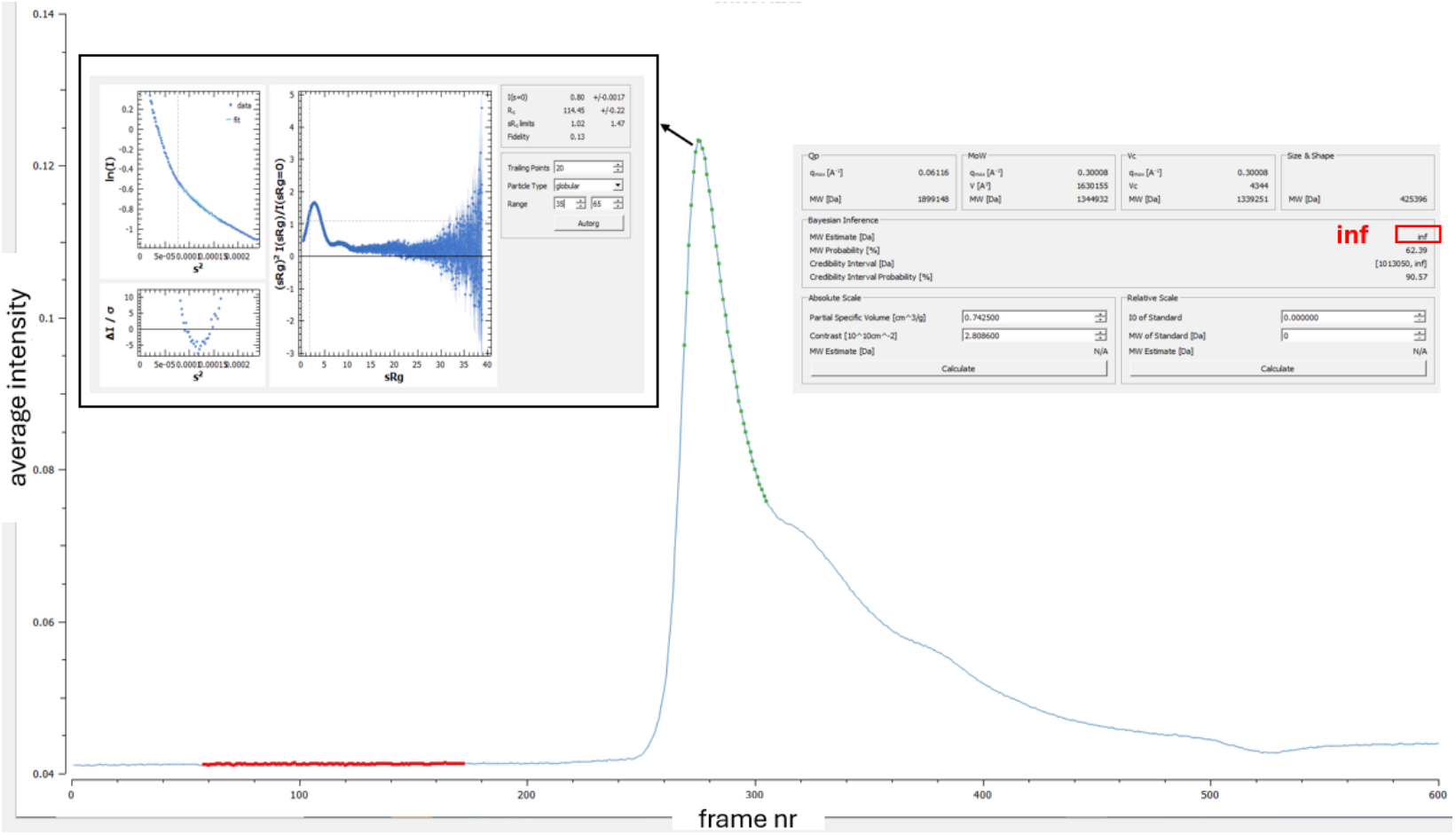


Fig. 5.3.2.4. Size exclusion chromatography small-angle X-ray scattering (SEC-SAXS) profile for Adgb N-terminal domain (20 mg/ml).

Where (A) CHROMIXS showing the SAXS curve used for analysis in the Guinier analysis (box). Guinier analysis shows a bad Guinier fit for the scattering profile of for Adgb N-terminal domain because it showed an upturn shape. Therefore, this scattering is more likely attributed to inter-particle interactions and/or high polydispersity of the protein sample as well as protein aggregation. This conclusion is supported by previous results, which indicated a range of particle sizes within the protein sample. Bayesian inference was not able to predict the molecular size as shown in the figure. This is probably due to the heterogeneous nature of the N-terminal protein.

5.4 Discussion

The crystallography attempts were unsuccessful, likely due to protein heterogeneity through aggregation, prompting further investigation of the N-terminal and Adgb-GD proteins in solution using SAXS. SAXS is capable of detecting inhomogeneities (variations) in electron density within a sample i.e. where regions scatter X-rays differently due to differences in density, across length scales ranging from 10 to 100 Å (Firer-Sherwood *et al*., 2011). This range corresponds to the size of large molecules or molecular complexes, such as proteins, protein complexes, or nucleic acids in solution.

While it does not yield high-resolution structural models (50 to 10 Å compared to the < 2 Å of X-ray crystallography) (Putnam *et al*., 2007), it is particularly valuable for studying proteins that are difficult to crystallise, offering important insights into their structure in solution (Firer-Sherwood *et al*., 2011). The radius of gyration (Rg) is a key parameter in SAXS that provide insight into the overall size, shape, and structural properties of macromolecules or particles in solution where smaller Rg indicates a compact structure and a larger Rg indicate an extended or larger assembly. Rg is derived from the Guinier region ($qRg < 1.3$ for globular protein) of the scattering curve (Putnam *et al*., 2007). It can be combined with other data to provide clues about the shape of the particle. Globular proteins typically have smaller Rg values due to their compact nature while elongated and aggregated proteins will display much higher Rg suggesting clustering of particles. As demonstrated in the study investigating the molecular mechanisms by which *S. oneidensis* interacts with and reduces mineral surfaces using SAXS and X-ray crystallography (Edwards *et al*., 2014), the Rg value of *S. oneidensis* OmcA, one of the four major clades of the *Shewanella* family of outer membrane multi-heme cytochromes (OMMCs), was determined to be 3.06 ± 0.2 nm. Complementary SAXS parameters, such as the maximum particle dimension (D_{max}) of 96 Å, provided insights into the molecular shape

of OmcA, revealing that it behaves as a monomer in solution with weak dimerization tendencies (Edwards *et al*., 2014).

Exploring the molecular structure of Adgb proved to be more complex than OMMCs. Although overall, the data indicate pronounced interparticle interactions consistent with protein aggregation. This observation is consistent with the requirement for the stabilisation measures employed during expression, including purification with SDS to maintain the protein's solubility. The removal of SDS during the dialysis of the protein could promote protein–protein interactions that drive aggregate assembly in a non-homogenous manner, potentially explaining the unusually high predicted molecular weight (Mw) of Adgb-GD (Figure 5.3.2.2). The size exclusion analysis results presented in Chapter 2.3.4 support the hypothesis that SDS removal contributes to the aggregation tendencies of the globin domain, as the protein appears to exist as a hexamer both in a high-salt buffer and a native buffer solution i.e. 0.1 mM NaPi at pH 7.4. While SDS is widely used to solubilise proteins during purification, its removal can pose challenges in maintaining protein solubility and stability. A detailed study investigating the impact of SDS on protein structure, particularly how it induces protein unfolding and how proteins can be refolded after surfactant removal, revealed the complexities associated with refolding SDS-denatured proteins (Nedergaard Pedersen *et al*., 2020). The study demonstrated that during the refolding of β-lactoglobulin (bLG), SDS binds rapidly to bLG, leading to the formation of large complexes of bLG and SDS. Additionally, refolding tends to promote the formation of dimers (Nedergaard *et al*., 2020).

In conclusion, the SAXS data for Adgb-GD suggest two possible explanations: either the protein is misfolded, or it aggregates due to salt bridges and/or hydrophobic patches. Misfolding is unlikely, as the optical spectra of the holo protein with reconstituted heme was typical of that of a correctly folded globin. Additionally, SEC-HPLC performed at lower protein concentrations showed a significantly lower molecular weight, therefore supporting

aggregation through non-covalent interactions (see section 2.3.4). However, ionic interactions have little effect, as similar oligomeric states (e.g., hexamers) were observed in both high-salt and native buffers (see Section 2.3.4). This points to hydrophobic interactions as the more likely cause of aggregation. Future experiments using surfactants such as deoxycholate or optimising the SDS condition may help reduce aggregation and further confirm this hypothesis.

Future studies could potentially avoid the issue of aggregation due to hydrophobicity by employing the method described in the purification study of membrane proteins using SDS or other detergents and subsequent renaturation by Hjertén *et al*. (1988). The study suggested that adding an efficient, non-ionic, and non-denaturing detergent, such as heptaoxyethylene lauryl ether (G-3707), at a high concentration to SDS-solubilised proteins increases the likelihood of proper protein folding. This is achieved because detergents like G-3707 form mixed micelles with free SDS molecules, reducing their concentration. Consequently, SDS micelles associated with the proteins are rapidly dissociated due to the high concentration of G-3707. This process allows membrane proteins to be freed from SDS and fold correctly while being surrounded by the non-denaturing, lipid-like hydrophobic tails of the G-3707 molecules, thereby preventing protein aggregation (Hjertén *et al*., 1988). The recommended protocol involves adding a 20% solution G-3707 to the SDS-solubilised protein to achieve a final concentration of 5% – 10%. Another alternative is to use detergents that are typically used to solubilise membrane and trans proteins. This includes triton-x100, tween 20 or deoxycholate used in cytochrome c oxidase purification (Smith and Stotz, 1954).

The heterogeneity of the N-terminal region could be attributed to the presence of multiple fragments in the solution, as observed in Figure 3.3.2.1. Additionally, the weak dimerization or mixture of higher oligomeric states (polydispersity) in the sample could further complicate the analysis. These factors may account for the lack of comprehensive structural information about the N-terminal region, including an accurate determination of the protein's

molecular weight. It is also possible that the N-terminal region shares similar ionic or hydrophobic patches characteristics with Adgb-GD. Further experiments will be needed to confirm this.

Preferably, isolating the full-length protein in the absence of auto-proteolytic fragmentation would be the key to stabilising both the globin and N-terminal domains, as their structural integrity likely depends on interactions with other regions of the protein. Previous studies have reported interactions between Adgb and proteins such as Piwil1, Tdrd6, and Spata20, based on co-immunoprecipitation (co-IP) and fluorescence resonance energy transfer (FRET) analyses (Bracke *et al*., 2018). These potential interacting partners may also contribute to the structural stability of Adgb, particularly since Tdrd6 belongs to the TUDOR domain-containing protein family, which is known to function as a molecular adaptor, facilitating protein–protein interactions and the assembly of macromolecular complexes (Pek *et al*., 2012).

CHAPTER 6: General Discussion

More than a decade has passed since the discovery of Adgb, yet it remains a novel protein with limited functional and structural information. To date, the full-length protein has not been successfully expressed, although expression of its globin domain was recently reported (Nie *et al*., 2024; Reeder *et al*., 2024). In this study, various purification strategies for the globin domain were evaluated, revealing that SDS-based purification is the most effective method for obtaining high yields of the protein. While purification using urea also produced a stable form of the protein, its yield was notably lower compared to SDS-based purification.

Expression of the stable form of Adgb resulted in the production of apo-protein (heme-free). This phenomenon also occurred in the study by Bracke *et al*. (2018), although this was likely due to the purification of the protein using urea. The underlying cause of the heme loss remains unclear; however, it may be related to the unusual heme pocket geometry with a disulfide substituting for the normal CD loop (Reeder *et al*., 2024). The absence of the CD loop in globins significantly affects their ligand binding, heme coordination, electron transfer, and protein stability (Vallone *et al*., 2004; Hamdane *et al*., 2005; Singh *et al*., 2014). For example, it has been reported that the absence of the CD loop in Ngb disrupted its native structure and altered the ligand characteristics (Exertier *et al*., 2022). The absence of the CD loop in Ngb was found to contribute to a more rigid protein core. Although Ngb retains its ferric hexacoordinate state, the absence of the CD loop, which normally interacts with one of the heme propionates via Tyr44, may reduce the barrier to distal His64 swinging outward (Exertier *et al*., 2022). This was supported by resonance Raman spectroscopy of the CO-bound form, which revealed a predominantly open heme cavity (Exertier *et al*., 2022).

Additionally, the CD1 Phe, normally stabilising the heme pocket, is replaced in Adgb with Tyr. The absence of the CD1 Phe in Adgb could lead to reduced heme affinity and increased loss during purification. A similar observation was reported in Hb Hammersmith (Akiyama *et al*., 2005), where a heterozygous mutation replaced β42(CD1) Phe with Ser,

leading to weakened heme–globin interactions and enhanced heme dissociation. This heme loss triggered protein denaturation and precipitation, ultimately causing membrane damage and premature red blood cell destruction. However, it should also be noted that numerous bacterial globins possess a CD1 Tyr without reported heme stability issues (Ouellet *et al*., 2007).

Overall, this study demonstrates the successful expression and purification of the Adgb globin domain, with a molecular weight of ~23 kDa, that is predominantly hexacoordinate in its deoxyferrous and ferric forms, with the hexacoordinated ferric form observed directly *in vivo* in *E.coli* and HEK cells. However, the ferric form shows pentacoordinate-like characteristics like that of Cygb. Two possible explanations have been proposed for this observation: (i) Adgb may genuinely exist as a mixed-coordinate species. Given that the protein can adopt multiple conformations, one of these may favour a pentacoordinate state. Some globins exhibit this behaviour; for example, Cygb adopts this pentacoordinate state when interacting with lipids (Reeder *et al*., 2011). The 'initiator or trigger' for Adgb to do this has not yet been identified. (ii) Alternatively, the observed mixed coordination may not reflect a true mixed-ligand species but rather arise from Adgb binding to its endogenous ligand in a pentacoordinate manner, potentially influenced by the presence of a disulfide bond. Additionally, the extinction coefficient for ferric Adgb has been reanalysed and revised from 106 $mM^{-1}cm^{-1}$ at 412 nm (Reeder *et al*., 2024) to 127.6 $mM^{-1}cm^{-1}$ at 413 nm. The previously reported lower value is now believed to have resulted from partial reduction of the protein in the original expression batch due to unforeseen factors.

One of the most interesting discoveries from this study is the inherent tendency of the Adgb globin domain to bind either to itself or to other proteins. This suggests that the globin domain may not be capable of existing as a monomer in isolation under native conditions, as it readily aggregates. This behaviour implies that Adgb may require interaction with partner proteins or other Adgb domains to perform its physiological function. This propensity is likely

driven by the globin domain's significant surface electrostatic potential and an unusually high localised density of hydrophobic regions, features not typically observed in monomeric globins such as Mb (see Figure 2.4.3.1 and 2.4.3.2). Consequently, this characteristic increases its likelihood of aggregation or nonspecific interactions in the absence of its partner protein(s).

These findings point toward a potential physiological relationship between the N-terminal region and the globin domain. It is plausible that, following proteolytic processing by the calpain domain, a portion of the N-terminal region remains associated with the globin domain possibly to stabilise it or regulate its interactions. The N-terminal region of Adgb is capable of autolysis due to the presence of a calpain-like proteolytic domain, a finding supported by evidence presented in this study. While further investigation is needed to fully characterise this region, initial results indicate that the N-terminal region displays a strong propensity to bind other proteins, in this case endogenous *E. coli* proteins, as evidenced by the appearance of multiple bands on SDS-PAGE (Figure 3.3.3.3). This behaviour may be attributed to several factors, including the possibility that the catalytic core of the proteolytic domain is more exposed and less structurally shielded compared to that of calpain-7. Therefore, molecules other than the intended substrate may bind, potentially leading to off-target interactions and unintended effects. To fully validate the behaviour of this region to bind to other proteins, it is essential to purify this region under denaturation conditions comparable to those used for Adgb-GD, to check if it reduces nonspecific interactions.

Sequence realignment analysis and mutagenesis studies revealed that the cysteine nucleophilic site is more likely located at either Cys181 or Cys206, rather than at Cys132 as previously proposed by Hoogewijs *et al*. (2012). Mutagenesis data strongly suggest that Cys206 is the most probable nucleophile, as its mutation significantly delayed autolysis and enabled the expression of the first full-length N-terminal region observed in this study. These findings support the classification of the protease as a potential cysteine protease. However,

the identities and locations of the other residues that form the complete catalytic site remain elusive.

It was proposed that the N-terminal region of Adgb may contain multiple catalytic sites. However, no calpain family member has been reported to possess more than one active site within a single polypeptide chain. While conventional serine proteases (e.g., trypsin, chymotrypsin, thrombin) typically feature a single catalytic site per enzyme chain, there are rare, non-canonical cases in which serine proteases exhibit multiple active sites. For instance, the viral serine protease from human cytomegalovirus forms a dimeric active structure, resulting in two separate catalytic sites with relatively low catalytic efficiency. Therefore, allowing the protease to cleave substrates at two distinct locations simultaneously (Chen *et al*., 1996; Darke *et al*., 1996; Chan *et al*., 2002).

SAXS data indicate that the N-terminal region of Adgb exists as a heterogeneous, aggregated species, suggesting a tendency to dimerise or oligomerise. This behaviour supports the hypothesis that multiple catalytic sites may arise through dimerization, rather than within a single polypeptide. Furthermore, the catalytic core of Adgb appears to be more exposed than that of calpain-7, potentially enabling simultaneous cleavage at multiple sites and allowing for more complex mechanisms, including unique interactions with the globin domain. However, no definitive conclusions can be drawn until it is confirmed that the multiple bands observed are not due to the co-elution of endogenous proteins.

This means that two possible conclusions may be drawn from the proposed future experiment. If the multiple bands on their western blot persist despite purification under denaturing conditions, it may indicate the presence of multiple active sites. Conversely, if the bands disappear, it will support the conclusion that the N-terminal region has a strong tendency to bind to other proteins.

As previously discussed, this study revealed that the heme pocket of the globin domain of Adgb can adopt multiple conformations, influenced by the structural relationship between the disulfide bond, the distal heme pocket, and the protonation of nearby residues. Overall, it was found that the reduced form of the disulfide bond slows down protein activity, likely due to the globin domain adopting a more rigid hexacoordinate configuration. In this state, the bond between the distal site amino acid and the heme iron becomes tighter, resulting in a slower rate of ligand binding. In contrast, the WT form with an intact disulfide bridge exhibited the opposite effect. This was attributed to the disulfide bond influencing the positioning of the distal glutamine, thereby limiting its ability to coordinate effectively with the iron centre. Nearby (protonated) residues could compete with the distal glutamine for coordination, potentially occupying the distal site in certain conformations. This supports the hypothesis that Adgb-GD does not exist as a single, uniform structure but rather as a dynamic ensemble of closely related conformations with similar, though not identical, optical spectra. All converging toward the same final state.

Finally, this study confirmed that Adgb is able to play a role in both the release and regulation of NO levels *in vivo*. This indicates that Adgb may function as a sperm oxidase (Keppner *et al*., 2022), contributing to the production of NO during the capacitation phase in eukaryotes, a role previously described as elusive in Section 1.4.

6.1 Future perspectives

This study has demonstrated a successful strategy for expressing and purifying the Adgb protein, at least its individual domains, enabling *in vitro* characterization. However, it also raises several important questions that warrant further investigation:

For the N-terminal region, the precise location of the proteolytic active site(s) remains to be determined. In addition, further work is needed to confirm whether the multiple bands observed are due to co-elution with endogenous proteins. This can be addressed by purifying the N-terminal region under denaturing conditions using SDS or urea, following the same approach employed for the globin domain in this study.

For the globin domain, further investigation is required to identify the residue whose protonation alters the local heme environment. Subsequent studies, such as site-directed mutagenesis, could then be used to determine how these changes affect the interaction between the distal pocket and the disulfide bridge. Clarifying this relationship would help elucidate the mechanism of ligand binding in Adgb-GD. Additionally, measuring if Adgb-GD can generate superoxide would provide further insight into its potential physiological functions, particularly its role in redox biology.

Lastly, the crystal structure of the full-length Adgb protein or any of its individual domains has not yet been resolved. Therefore, more structural studies using techniques such as SAXS or X-ray crystallography is necessary to fully characterise the architecture of this protein.

APPENDIX

A1. Effect of Androglobin Overexpression in HEK293T Human Cell line on mRNA Expression

To observe the effect of Adgb overexpression in HEK293T human cell line on mRNA expression, the cells were prepared and Adgb gene expression was induced by transfection as previously described in chapter 2.2.12. Following an incubation at 37 °C, 5 % CO_2, the cells were harvested by gently pipetting up and down in 1x PBS, allowing them to be resuspended in the well without the use of trypsin. After harvesting, the RNA in the cells expressing the full-length Adgb, Adgb globin domain (both contained in pcDNA3.1+ plasmids) and the control cells transfected with empty pcDNA3.1+ plasmid were isolated from their cells using RNA extraction kit (RNeasy, Qiagen, UK) according to the manufacturer instructions. The RNA concentrations were quantified using a Nanodrop 2000c (> 100 ng/µl and A260/A280 between 1.8 – 2.1) and they were sequenced by Novogene (Cambridge, UK) for analysis. The read counts obtained from this analysis were normalized to the mRNA levels of the empty plasmid control. The results were visualised using a volcano plot generated in Microsoft Excel.

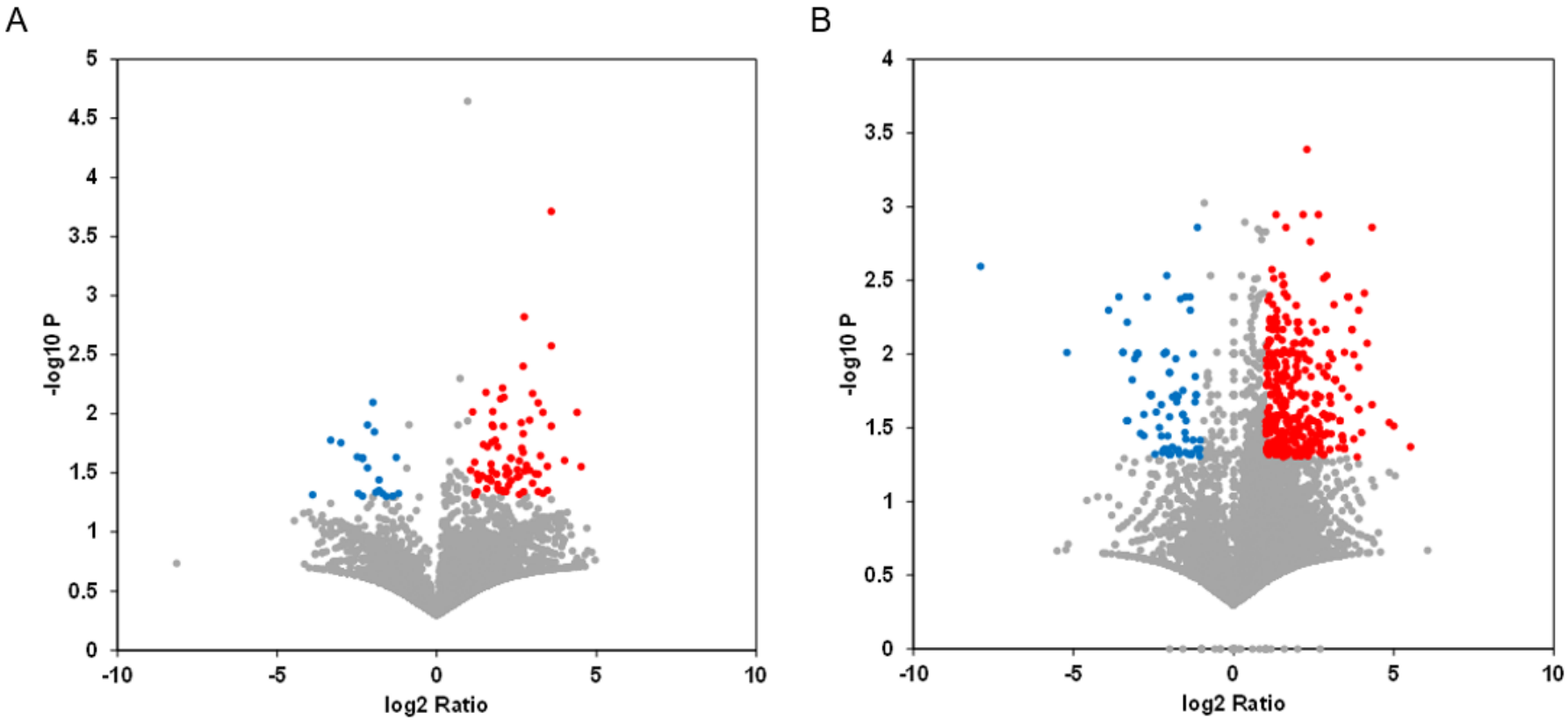


Figure A1. Volcano plot showing the MRNA analysis the (A) full-length Adgb and (B) Adgb-GD.

Bracke *et al*. (2018) provided an overview of genetic defects that have been shown to significantly impact male fertility. The key genes implicated, as identified by Bracke *et al*. (2018), include: *TEX11, ZMYND15, TAF4B, SYCE1, MCM8,* and *TEX15*, all of which are associated with non-obstructive azoospermia (NOA), a form of infertility commonly linked to testicular failure. This condition is typically characterised by reduced testicular volume and elevated levels of luteinising hormone and follicle-stimulating hormone (Wosnitzer *et al*., 2014).

Azoospermia, defined as the complete absence of spermatozoa in the ejaculate, is classified into two main types: obstructive azoospermia, which results from physical blockages in the male excurrent ductal system, and non-obstructive azoospermia, which stems from intrinsic testicular dysfunction, as described above (Bracke *et al*., 2018).

There are also the *AURKC, DPY19L2, SPATA16* and *DNAH1* genes implicated in teratozoospermia, as identified by Bracke *et al*. (2018). Teratozoospermia is a condition of male infertility characterized by the presence of abnormal sperm morphology in over 96% of spermatozoa within a semen sample (Menkveld, 2010). Specific genetic defects are associated with distinct subtypes of teratozoospermia. For example, macrozoospermia, linked to mutations in *AURKC*, is characterized by spermatozoa with abnormally large and irregularly shaped heads and multiple flagella. Globozoospermia, associated with *DPY19L2* and *SPATA16* mutations, presents as round-headed spermatozoa that lack an acrosome. Additionally, multiple morphological anomalies of the flagella (MMAF), which result from *DNAH1* mutations, are defined by spermatozoa with absent, short, bent, coiled, or irregularly shaped flagella (Bracke *et al*., 2018).

Other genes identified by Bracke *et al*. (2018) includes *CATSPER1, CATSPER2, SEPT12*, and *SLC26A8*. These genes are implicated in asthenozoospermia, a form of male infertility characterized by reduced or absent sperm motility despite normal sperm morphology

and count (Curi *et al*., 2003; Menkveld, 2010). Additionally, mutations in the *PLCZ1* gene have been associated with normozoospermic infertility, a condition in which standard semen parameters appear normal, yet infertility persists suggesting an underlying mechanism with an unclear or idiopathic aetiology.

All of the genes mentioned above were specifically examined following mRNA analysis of both full-length Adgb and Adgb-GD. None of these genes were found to be significantly upregulated or downregulated in HEK293T human cells after Adgb overexpression, as shown in Table A1. In contrast, other genes such as *SLC2A6, SLC25A16, SPATA12, ZMYND19P1*, and *PLCH1-AS1* are examples of the genes identified as differentially expressed in response to Adgb overexpression (Table A1).

SLC2A6 and *SLC25A16* are members of the solute carrier (SLC) family, similar to *SLC26A8*, which has previously been associated with male infertility. *SLC2A6*, which facilitates hexose transport into mammalian cells, was found to be upregulated in response to overexpression of the full-length Adgb protein. In contrast, *SLC25A16* a gene with a possible link to autoimmune conditions such as Graves' disease was downregulated following Adgb-GD overexpression.

ZMYND19P1, like *ZMYND15*, encodes a MYND-type zinc finger domain-containing protein. It has been shown to bind to the C-terminus of melanin-concentrating hormone receptor 1 (Bachner *et al*., 2002) and to the N-termini of both alpha- and beta-tubulin (Franke *et al*., 2005). It is broadly expressed in the testis and was found to be upregulated in the Adgb-GD overexpression condition.

PLCH1-AS1, although not a member of the phospholipase C (PLC) family like *PLCZ1*, is an antisense RNA gene to *PLCH1* and is classified as a long non-coding RNA (lncRNA). It

has been associated with glioma susceptibility 1 (Zhi *et al*., 2015) and was also upregulated following Adgb-GD expression.

Among all the genes identified, *SPATA12* is the only one known to play a direct role in testicular development and spermatogenesis. It was also upregulated in the Adgb-GD condition. *SPATA12* is primarily expressed in the testis and has demonstrated antioxidant properties, potentially protecting cells from oxidative stress and apoptosis induced by hydrogen peroxide (Wu *et al*., 2018). Its upregulation in oxidative conditions suggests a role in cellular redox defence. Like *SPATA16*, and *SPATA20* discussed previously in Section 1.4, *SPATA12* belongs to a group of genes functionally linked to spermatogenesis. This property, along with *SPATA12* antioxidant function, supports a potential *in vivo* role for Adgb-GD in redox regulation and testicular function.

In the volcano plot, blue dots represent downregulated genes, while red dots indicate upregulated genes. This study represents a preliminary attempt to explore the effects of Adgb gene expression on mRNA levels in cells. However, these findings are exploratory and require further validation. The current analysis is based on a rough estimation, with inconsistencies such as the use of different transfection reagents across samples and insufficient biological replicates. For instance, only n = 2 for the Adgb-GD condition. Therefore, additional studies involving experimental optimisation and more robust statistical analyses are needed before drawing definitive conclusions.

Table S1. The mRNA genes downregulated and upregulated after the full-length Adgb and Adgb-GD gene expression.

Protein	Significantly downregulated GeneID (p<0.05)	Significantly upregulated GeneID (p<0.05)	
Full-length Adgb protein	LINC02458	CCT6P1	WFDC12
	MX2	CBR4-DT	ENSG00000253189
	OR2A1-AS1	ENSG00000250948	ENSG00000279466
	ENSG00000255320	ENSG00000225884	EIF5-DT
	ENSG00000287278	ENSG00000267620	NEBL-AS1
	ENSG00000288046	FAM78B	NLRP2B
	RN7SL806P	LINC01320	EXTL1
	ENSG00000259563	ENSG00000224301	TMEM40
	RNU6-998P	MTND1P7	ENSG00000232470
	RNU4-5P	HS3ST2	ENSG00000285081
	MSLN	HCAR2	ENSG00000225057
	RPL9P29	ENSG00000272969	ACMSD
	ENSG00000226984	LINC01967	ENSG00000254380
	LCN9	FAM86B3P	ENSG00000286825
	GRIA4	PEAK3	ENSG00000287348
	ENSG00000279422	RNU6-1143P	LAMTOR3P2
	EIPR1-IT1	CNN2P9	ENSG00000252473
	MIR588	ENSG00000288863	ASH1L-IT1
	ENSG00000261187	MICOS10-DT	DBIL5P2
	ENSG00000253358	LIPK	ENSG00000269068

	ENSG00000242282		SLC2A6 ENSG00000198358 ENSG00000266498 HELT BLACAT1 PSCA ENSG00000231128 SPDYC RNU6-529P ENSG00000259066 CYP7A1 HMGN1P8 FOXN3-AS1 ENSG00000286428 ENSG00000272009 HHIPL1 ENSG00000290734 RPL13AP3 EPM2AIP1 ENSG00000285492	ENSG00000270690 ENSG00000237329 ENSG00000274670 DOK6 RHPN1-AS1 ENSG00000230929 DDX18P1 ENSG00000220960 SNORA15B-1 RERE-AS1 KCNH6 LINC02724 CD69 TXN2P1 ENSG00000289474 ENSG00000273212 ENSG00000288004
Adgb-GD	CYP2A7P2 RNU6-163P	IGLV3-7 LINC01934	ENSG00000255210 ENSG00000285492	RBM11P1 ENSG00000229347

	HNRNPUL1 IL13RA1 CPNE1 ENSG00000258210 ENSG00000251675 LINC01145 RNU6-369P TEKT4P3 ENSG00000250039 ACTR3BP4 DMXL1-DT ENSG00000285941 TRBC2 RPSAP10 EGLN3P1 ADAM6 RN7SL199P CHRNE LINC01949 ADCY7 SNHG30 STAG3L2	RNA5SP105 CEBPA-DT ENSG00000196395 ENSG00000236450 RN7SL817P PPIHP1 ENSG00000243818 ENSG00000283692 ENSG00000285936 POGZ ENSG00000254839 ENSG00000275286 NAMPTP3 ENSG00000272444 ENSG00000234699 RNU6-781P OR2A1-AS1 FAM90A17 RN7SL164P MEIKIN CCT4P1 ENSG00000279591	ENSG00000289936 ENSG00000290686 FASTKD2 ENSG00000279379 CYP27C1 ENSG00000224612 WDR76 RN7SKP189 SPON1-AS1 AK3P4 ENSG00000261938 EVPLL ENSG00000227484 ENSG00000277715 UNC5C NUP54 ENSG00000278231 OASL2P	UPF3AP2 MIR4509-3 ENSG00000287755 PYCR1-AS1 MTCO1P55 ENSG00000266805 LINC00452 LINC00613 PRDM6 RNU4ATAC3P LAMTOR3P2 RNA5SP383 IGHV3OR16-12 ENSG00000293216 ENSG00000268531 LINC02932 ENSG00000254787 IGKV1-17

	DKK4	ENSG00000260152	STX16	SMIM10L2B-AS1
	CNPY2-AS1	FAM149A	TFCP2L1	LINC02024
	RPS24P7	ENSG00000217557	BCAS3	ENSG00000272330
	RPSAP68	OR51M1	GSDME	RPL7P35
	MARK2P8	ENSG00000255317	ISM2	AFDN-DT
	PSMD7	ENSG00000274718	PPIL2	KCNA10
	ABCC6P2	RAX	RPL37A-DT	ENSG00000286573
	ATP13A4-AS1	ENSG00000253358	ENSG00000285539	LINC00355
	PDZRN3-AS1	ENSG00000276407	ENSG00000237716	POTEM
	CSKMT	ENSG00000270802	H4C13	ATP1A3
	PLAAT3	LINC01162	VARS1	RPS7P11
	ENSG00000287008	EMB	VN1R35P	ROR1-AS1
	NPW	TRAPPC2B	ENSG00000287610	ENSG00000228162
	ENSG00000225632	IGHD4-17	ENSG00000231515	ENSG00000270829
	AGR2	TMEM256P2	DEFB106A	CSAG3
	CCL3	LHX3	ENSG00000273360	ROCK1
	LINC01282	ENSG00000289601	LHX5-AS1	PARP1
	ENSG00000280400	RNU7-125P	ENSG00000282997	MIR8085
	PGAM1P9	ENSG00000234210		
	MIP	ENSG00000276521		
	ENSG00000257824	NOS2P2		
	RPL7P40	COX6CP3		

	ENSG00000277124 MIR6785 ENSG00000233139 YBX1P2 RNU6-53P ENSG00000253195 MOB1B ENSG00000255320 KRTAP8-1	SLC25A16 ENSG00000286251 KRTAP2-2 LSM6P2 ENSG00000240086 RNU7-45P OR13C9	ENSG00000280234 ENSG00000273804 ENSG00000261144 SPEM1 RN7SKP182 MTL3P ZDHHC22 C6orf47 ENSG00000275392 ENSG00000231977 EFEMP2 DOCK2 LINC03049 ENSG00000241269 DTX2P1 ENSG00000265982 ELMO2P1 MMRN1	PPM1M IGLV3-12 ENSG00000229188 RPL27AP8 RPL7AP32 RBMXL3 ENSG00000293448 ENSG00000248540 ENSG00000229983 HTR1D SLC16A1-AS1 HMGB1P22 HSP90AB1 ENSG00000250060 COL18A1-AS1 ENSG00000253275 ENSG00000233978 RNU6-391P

			TMPOP1	RPL21P35
			ZNF37A	ARL2BPP2
			RNU6-394P	RPL41P1
			ENSG00000251461	ENSG00000260058
			BLOC1S3	MIR3192
			IFRD1	TIMM8B
			UBE2F	SPATA12
			ENSG00000276406	ENSG00000279813
			TCEAL9	ENSG00000255260
			TPT1P13	RNU6-790P
			MIR222HG	RNU6-195P
			LINC01078	RPL8P2
			ENSG00000278367	ADAD2
			SNORA15B-1	LPAR5
			ENSG00000289918	LINC01348
			ENSG00000282022	RNU5D-2P
			LINC01661	ENSG00000233739
			MIR1282	FUNDC2P1

			ENSG00000200991	DUS2
			ENSG00000230821	ENSG00000291000
			RPL14P5	AHNAK
			IGHV1-24	ENSG00000278147
			KRT8P26	KAAG1
			ST8SIA6	POTEH
			RNA5SP182	MIR6843
			ENSG00000290668	RNU6-256P
			HNRNPCL1	FSD1
			MIR3662	TEK
			ENSG00000236601	MIR1247
			OTUD7A	ENSG00000277170
			SETP15	MGAT1
			ENSG00000234584	ENSG00000278863
			LINC01623	OR2T6
			ENSG00000257766	ENSG00000259514
			MIR4703	ENSG00000285694
			ENSG00000263494	IL1RL1

			PTPN14	RNU7-105P
			ENSG00000227531	TYMS
			ENSG00000272969	TBC1D19
			ABHD14B	ENSG00000286829
			ENSG00000290523	ENSG00000288690
			ADI1P3	ENSG00000276842
			ENSG00000258764	LONRF2P1
			MAPK1IP1L	KRT8P10
			ENSG00000290608	HNRNPA3P11
			ZMYND19P1	LRRC40
			HELT	NCOR1P2
			ENSG00000259066	FAM185A
			DOC2A	PFN1P12
			PRR13P3	SDHAP1
			LINC01309	NBPF5P
			AADACL4	SCAF4
			RPL21P121	RPS18P6
			ENSG00000286274	IGHV3-52

			RPL7AP64	LRPAP1
			ENSG00000286421	NELFB
			GABRA4	ENSG00000289239
			ENSG00000255158	BRWD1P3
			GP6	LINC02560
			CILP2	MIR5197
			TPSG1	ENSG00000270380
			OBP2A	ENSG00000226218
			ENSG00000278754	ENSG00000270335
			YRDC	ENSG00000286099
			ENSG00000286906	RNU2-53P
			ENSG00000248763	LINC02148
			ENSG00000239719	ENSG00000264869
			RNU4-91P	SNX5
			ENSG00000235242	ENSG00000248469
			ENSG00000261673	KRT26
			CALM2	ENSG00000289142
			INPP1	ENSG00000288976

			ERVK-28	ENSG00000227090
			SYPL1	ENSG00000281741
			LINC01967	ENSG00000286513
			TVP23CP2	Metazoa_SRP
			FAM157C	ENSG00000287369
			LINC02930	KIR3DP1
			RN7SL183P	ENSG00000224556
			ENSG00000257113	NEBL-AS1
			KLK4	H3P17
			ENSG00000286845	KCNB1
			LINC00544	MIR4422HG
			ENSG00000257194	ENSG00000279446
			ENSG00000289524	ACOT11
			ENSG00000232985	DYNLT3
			TERLR1	PCED1CP
			ENSG00000258205	PLIN3
			ENSG00000237759	ENSG00000235147
			ENSG00000274322	ENSG00000259909

			ENSG00000249685	ENSG00000290780
			ENSG00000256226	COX6B1P4
			MIR3666	ENSG00000279985
			ENSG00000251099	ENSG00000274446
			COMT	RN7SL643P
			ENSG00000227554	ENSG00000287146
			ENSG00000285081	ENSG00000261512
			OR7E101P	KCNJ9
			IRF2	ENSG00000279149
			MRPS21P3	ENSG00000286044
			RN7SKP263	ENSG00000282863
			C16orf87	ELP3
			YBX2P2	LINC01705
			MAST4-AS1	ENSG00000250573
			ENSG00000288437	ENSG00000288807
			TRAIP	ENSG00000278424
			ENSG00000285527	LINC00534
			EDEM2	RPS3P1

			ENSG00000228771	LINC01550
			CYP4F44P	EEF1A1P14
			SEZ6L2	ABCB5
			GABPB2	ENSG00000289368
			GCA	TBC1D3H
			MIR1290	ENSG00000259149
			MIR6724-2	CPS1-IT1
			ENSG00000226208	ENSG00000236837
			CYCSP55	CCDC33
			LRRC37A9P	RPS3AP50
			ENSG00000200706	LINC01423
			ENSG00000270859	UTS2B
			ENSG00000285876	RPL23P8
			CUX2P1	RNU2-7P
			ENSG00000279698	RN7SL253P
			RPS26P24	ENSG00000290096
			ENSG00000258593	ENSG00000291018
			ENSG00000290667	BRD9P1

			UBE2D3P4	PRRC1
			MTND4LP18	OSGEP
			ENSG00000237987	EPS8L2
			ENSG00000270719	RHBDF2
			ENSG00000248518	LINC02571
			RN7SL565P	FZD3
			ENSG00000263729	ECM1P2
			SLITRK3	NPAP1P9
			RPL23AP94	ENSG00000235243
			EEF1A1P1	EYA1
			ENSG00000287887	LRWD1
			IER5L-AS1	ENSG00000287267
			SH3TC2	MIR5694
			ASH1L-IT1	CELF2-AS1
			ENSG00000227355	FAM138B
			IGHV4OR15-8	ENSG00000225505
			PCAT14	OR4A5
			MIR4698	COL12A1

			SLX1A	PEAK3
			CDC37P2	ENSG00000269068
			ENSG00000280364	ENSG00000270204
			ENSG00000231495	ENSG00000232227
			ARL14	SERPINA11
			CCDC192	ENSG00000286190
			ENSG00000234460	RNU6-444P
			HNRNPA1P13	ENSG00000286968
			BRI3P2	ENSG00000256185
			MIR18A	LINC01736
			TMEM17	RNA5SP442
			BCAP31	ENSG00000278972
			ATP5F1EP1	ACER2
			RNU6-84P	RNA5SP416
			ENSG00000235386	TCP1
			SUCLA2P3	NUS1P1
			ENSG00000273059	FCHO2-DT
			ENSG00000248664	ELFN2

			ENSG00000291241 MOXD1 MMP2 RNU7-20P PLCH1-AS1 ENSG00000278901 PITPNM2-AS1 RNU6-796P BIRC7 MRTO4	OR51B5 ALDH7A1P3

A2. Amino Acid Sequences of Androglobin

Figure A2. Comprehensive annotated sequence of human Adgb.

DNA FASTA sequences were downloaded from Uniprot ID: Q8N7X0-1. Arrows indicate the start position of each Adgb domain along the amino acid sequence. In the N-terminal region, blue and turquoise highlight the locations of forward and reverse primers (R1, R2, R3, F1, F2, and F3 respectively) used to construct the nine truncation fragments listed in Table 3.3.2.2B. Red-highlighted text marks the four highly conserved cysteine residues within the N-terminal region, with arrows (in keys) denoting their positions along the sequence as you down. The stop codon is shown in Gray. The calpain-like domain in the N-terminal region is color-coded orange.

The globin domain is shaded Gray. In the C-terminal region, red highlights indicate predicted disordered regions of the protein (calculated using PrDOS), corresponding to the residues located above the threshold line in Figure 1.3.3.1.

A2.1 Amino Acid Sequences of Androglobin N-Terminal Fragments following In-Frame Deletions

In Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F1: the thick arrow precisely marks the positions of the sequences that were removed i.e. Q5 to K172. Key Cys: Cys132 was also part of the sequences deleted.

In Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F2: the thick arrow precisely marks the positions of the sequences that were removed i.e. Q5 to Y231. Key Cys: Cys132, Cys181 and Cys206 were also part of the sequences deleted.

In Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F3: the thick arrow precisely marks the positions of the sequences that were removed i.e. Q5 to P344. Key Cys: Cys132, Cys181 and Cys206 were also part of the sequences deleted.

In Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F1: the thick arrow precisely marks the positions of the sequences that were removed. i.e. K75 to K172. Key Cys: Cys132 were also part of the sequences deleted.

In Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F2: the thick arrow precisely marks the positions of the sequences that were removed i.e. K75 to Y231. Key Cys: Cys132, Cys181 and Cys206 were also part of the sequences deleted.

In Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F3: the thick arrow precisely marks the positions of the sequences that were removed. i.e. K75 to P344. Key Cys: Cys132, Cys181 and Cys206 were also part of the sequences deleted.

In Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F1: the thick arrow precisely marks the positions of the sequences that were removed i.e. N157 to K172.

In Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F2: the thick arrow precisely marks the positions of the sequences that were removed. i.e. N157 to Y231. Key Cys: Cys181 and Cys206 were also part of the sequences deleted.

In Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F3: the thick arrow precisely marks the positions of the sequences that were removed. i.e. N157 to P344. Key Cys: Cys181 and Cys206 were also part of the sequences deleted.

In Adgb_Nterm stop_E395*: The figure present successful substitution of the glycine residue to a stop codon (E395*), which is denoted by the asterisk. The sequence of the forward (top) and reverse (bottom) strands is shown, confirming the presence of the mutation at the highlighted position.

The sequence of the forward (top) and reverse (bottom) strands are also shown, confirming the absence of the frameshifts in the gene.

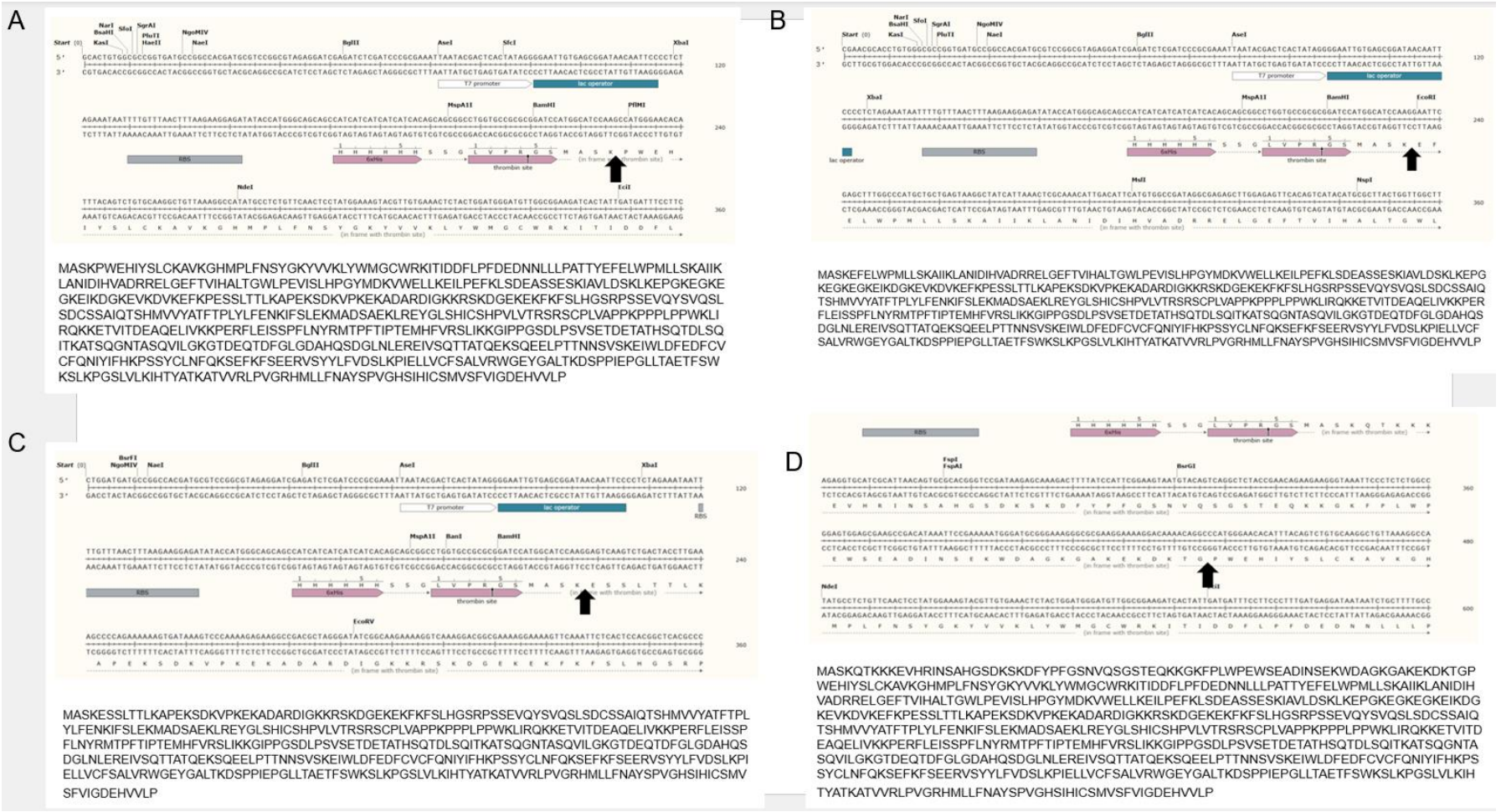


Figure A2.1.1 DNA sequences of amplified PCR product from Adgb primers after Sanger sequencing.

(A) Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F1. (B) Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F2. (C) Adgb_Nterm_trunc_R1 + Adgb_Nterm_trunc_F3. (D) Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F1.

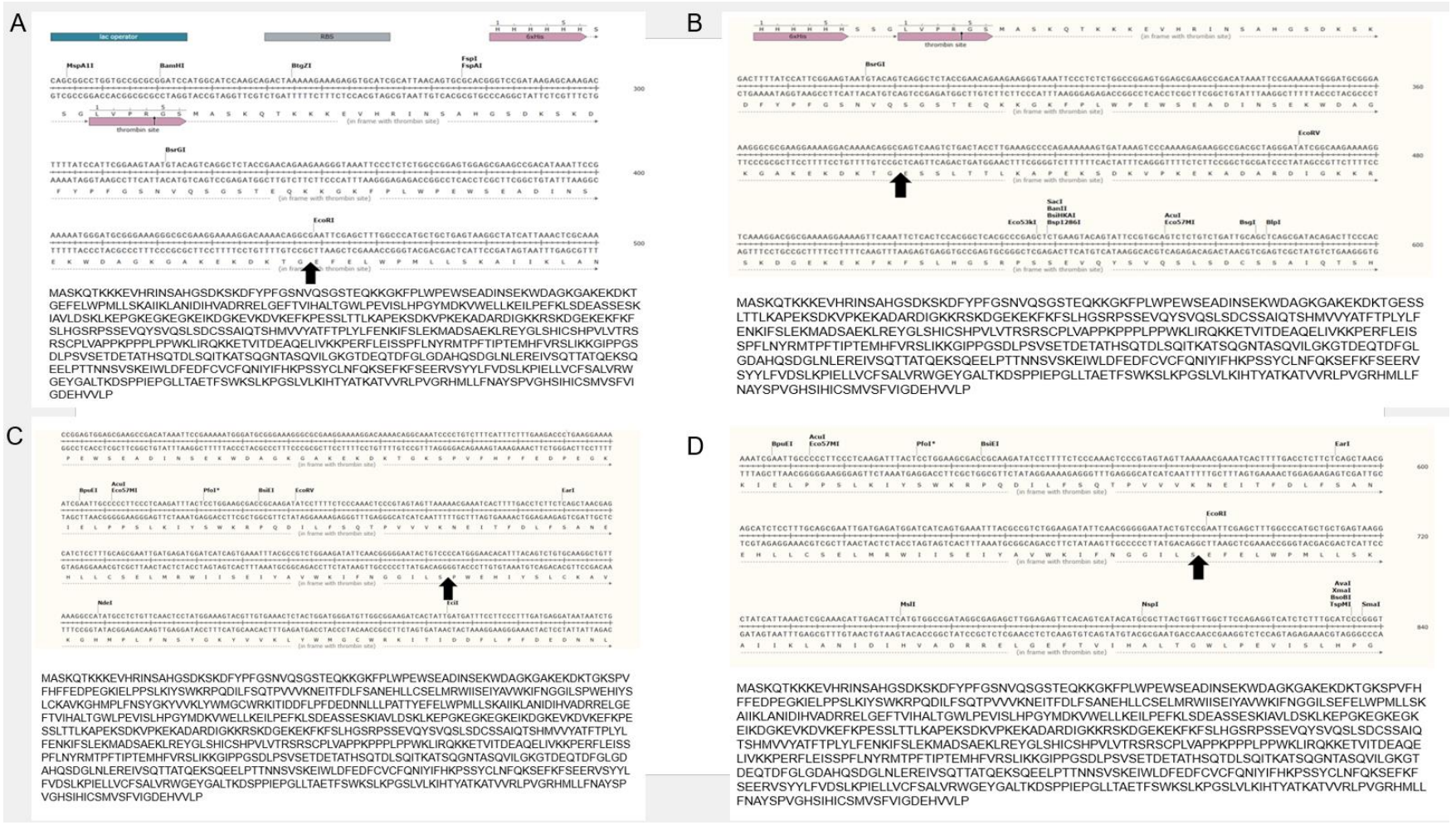


Figure A2.1.2 DNA sequences of amplified PCR product from Adgb primers after Sanger sequencing.

(A) Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F2. (B) Adgb_Nterm_trunc_R2 + Adgb_Nterm_trunc_F3. (C) Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F1. (D) Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F2.

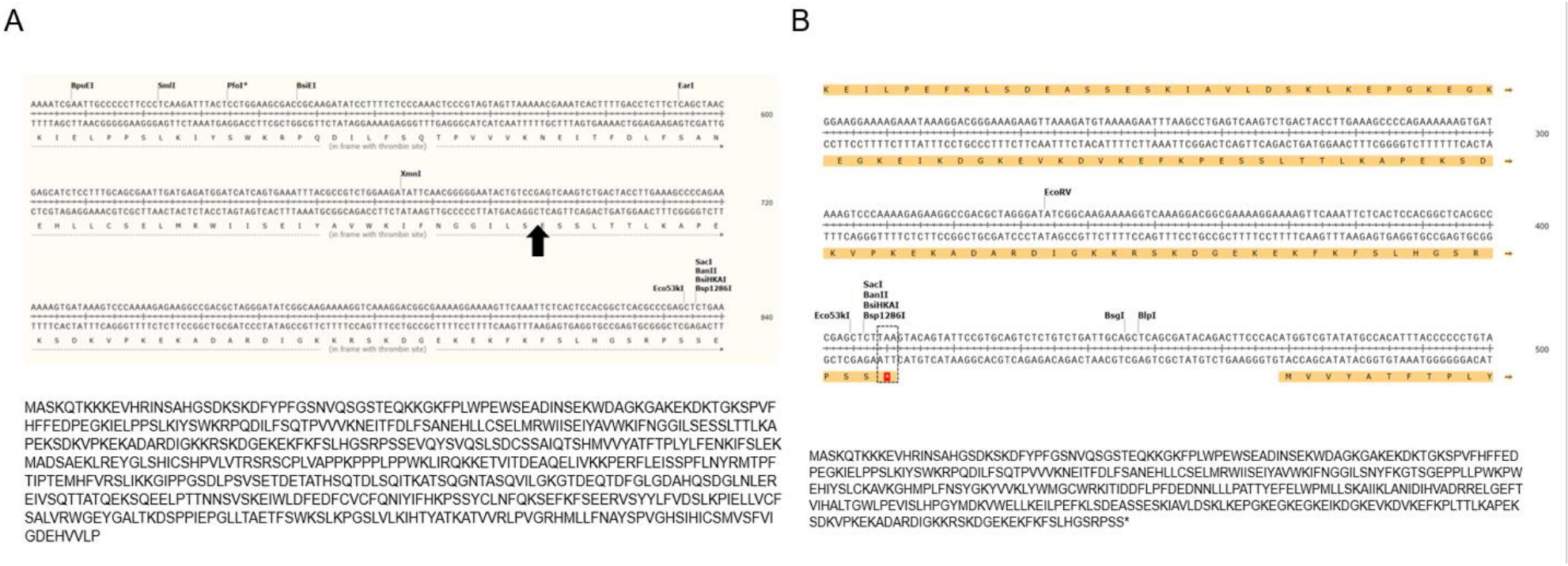


Figure A2.1.3 DNA sequences of amplified PCR product from Adgb primers after Sanger sequencing.

(A) Adgb_Nterm_trunc_R3 + Adgb_Nterm_trunc_F3. (B) Adgb_Nterm stop_E395*.

A3. Densitometric analysis of Androglobin N-terminal mutants C181G and C206G

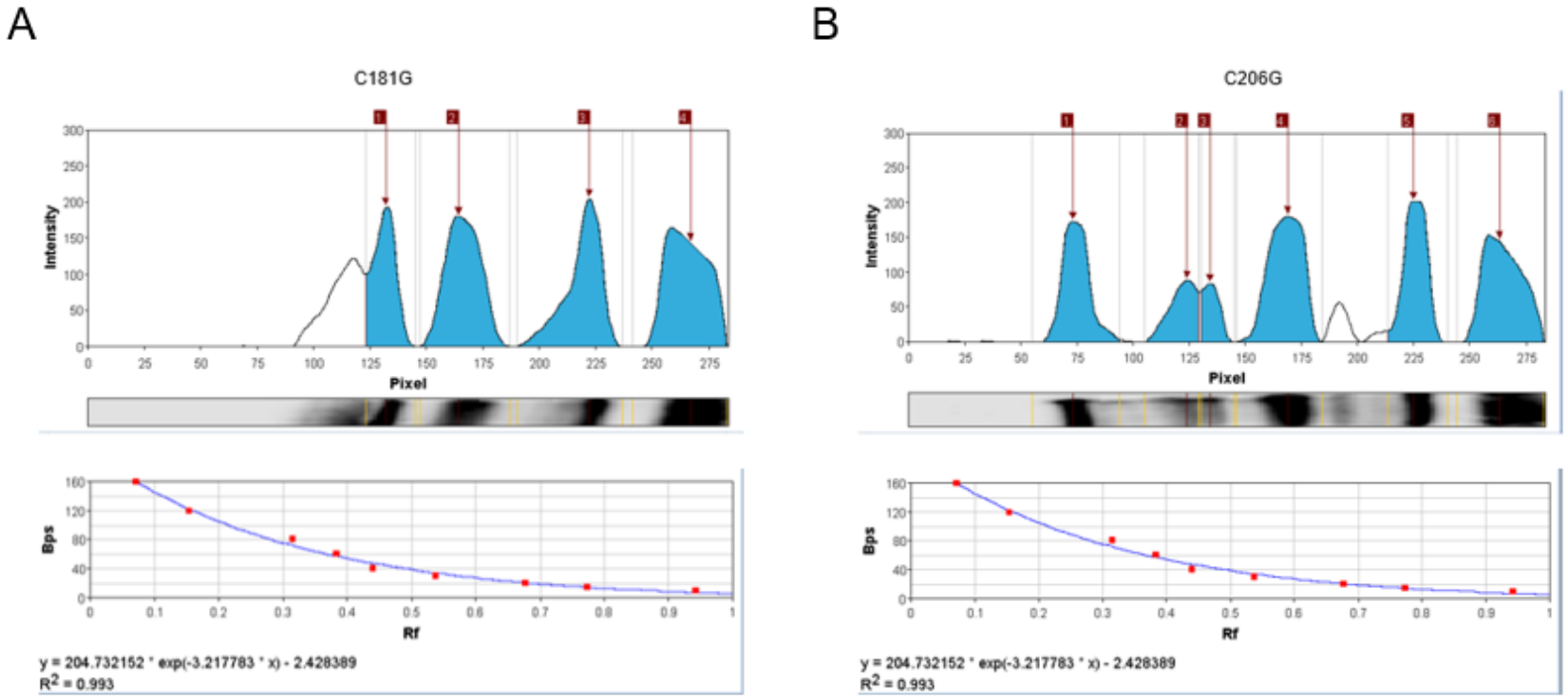
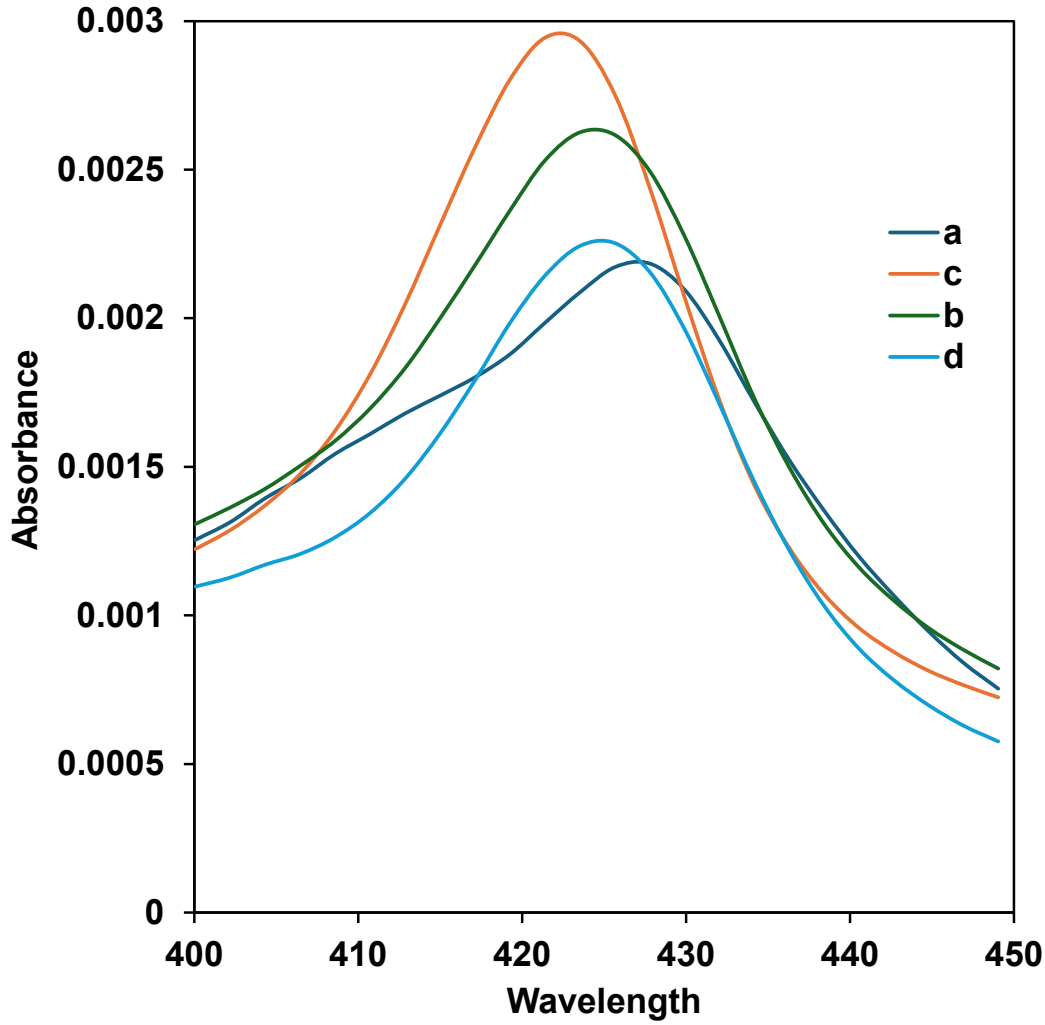


Figure A3. Densitometric analysis of Androglobin N-terminal mutants C181G and C206G.

The result was obtained after increasing the antibody concentration to 1:2000. Densitometric analysis was carried out using *GelAnalyzer v23.1.1*. Three distinct bands were observed in C181G according to gel analyzer as shown in (A). They have Mw (kDa) of 43, 30 and 14. The fourth band may be a smear. (B) In contrast, there are 5 bands (including 4 distinct bands) in C206G with Mw (kDa) of 87, 48, 42, 28 and 14. From this result it can be concluded that C206G mutation partially expressed the full-length N-terminal protein.

A4. Global analysis fit of WT Adgb-GD reaction with CO



A → C ← B
↑
D

Figure A4. Global analysis fit of the Adgb-GD WT to a sequential model between 400 – 450 nm at CO concentration of 100 µM.

A5. Kinetics of CN binding to Myoglobin and Neuroglobin

Following the rapid mixing of ferric Mb or ferric Ngb (5 μM) with solution containing cyanide (KCN) in NaPi buffer pH 7.4, optical transitions were observed showing that CN binds to these heme proteins. This is shown in Figure A5.1 and A5.2. The kinetic traces and the fits to their time courses showed CN binding to Mb comprises of a single kinetic phase (Figure A5.1C). In contrast, the kinetic traces and the fits of CN binding to Ngb showed that this process fitted to two exponential functions (Figure A5.2C).

Mb readily binds CN and showed consistent optical spectra across a range of concentrations, from higher to lower. However, Ngb showed minimal binding to CN across a range of cyanide concentrations, from higher to lower. It does not produce a distinct isosbestic point (Figure A5.2A), consistent with observations reported in previous research (Herold *et al*., 2004). At lower cyanide concentrations, binding to Ngb is less evident and it did not show binding at 1mM CN concentration. This is unsurprising, as ferric Ngb is known to exhibit low affinity for CN, largely due to its hexacoordinated state in which the ferric heme iron has a strong binding preference for the endogenous histidine ligand (Kiger *et al*., 2004).

The plots of rate constants vs cyanide concentration for both Ngb and Mb are given in Figure A5.3.

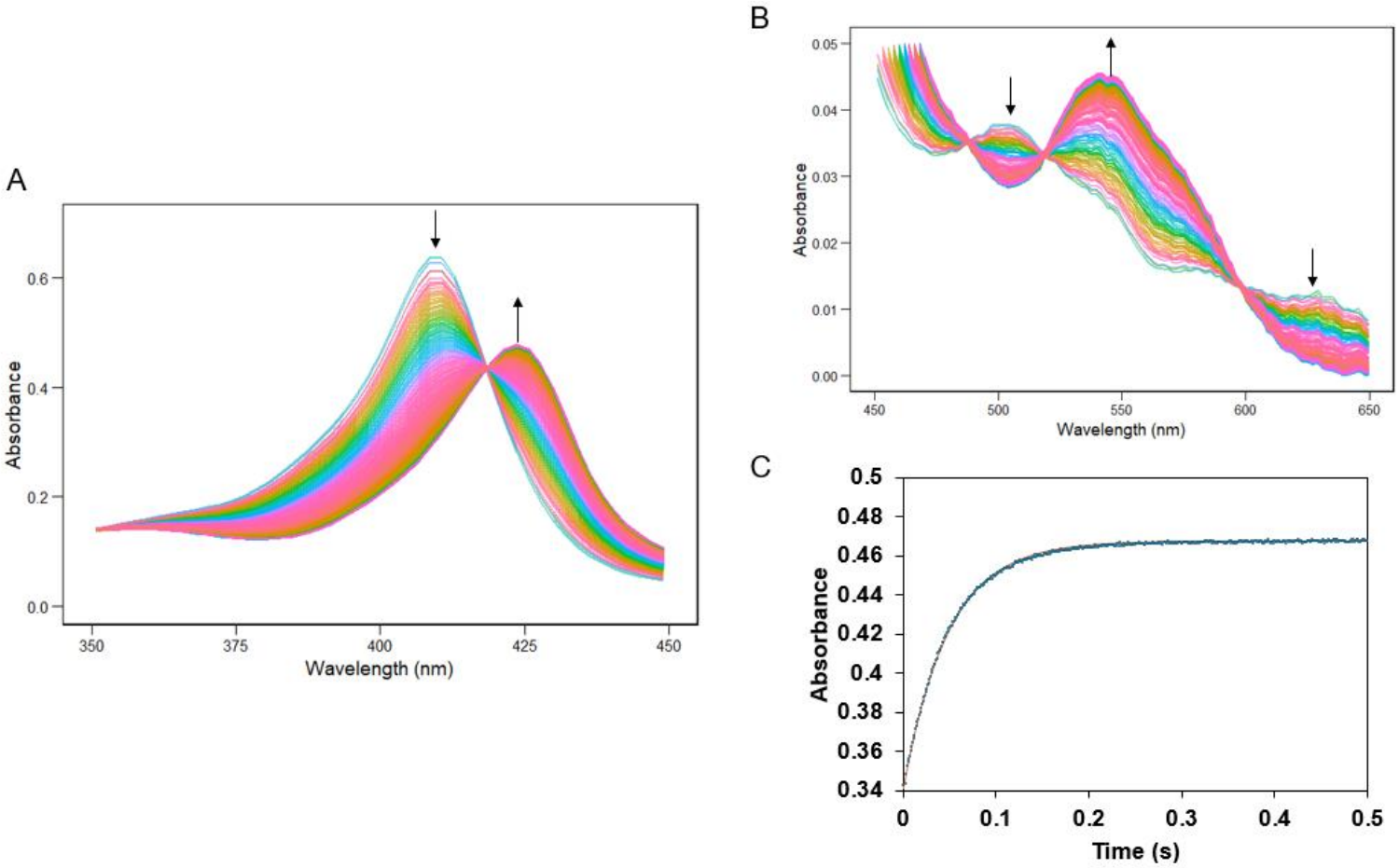


Figure A5.1 The time dependent spectra of CN binding to equine Mb (5 μM) at 20 mM CN concentration.

(A) Optical spectra of equine Mb visualised using RStudio. The time dependent spectrum exhibits a set of isosbestic points therefore suggesting that only two optical species are present, namely the ferric species and the ferric-CN bond complex formed. (B) Highlight the visible region. (C) The time course for Mb at 421 nm and the fit. Colour code: fit = orange line, dataset = blue dots. The arrow indicates the direction in which the optical spectra shift.

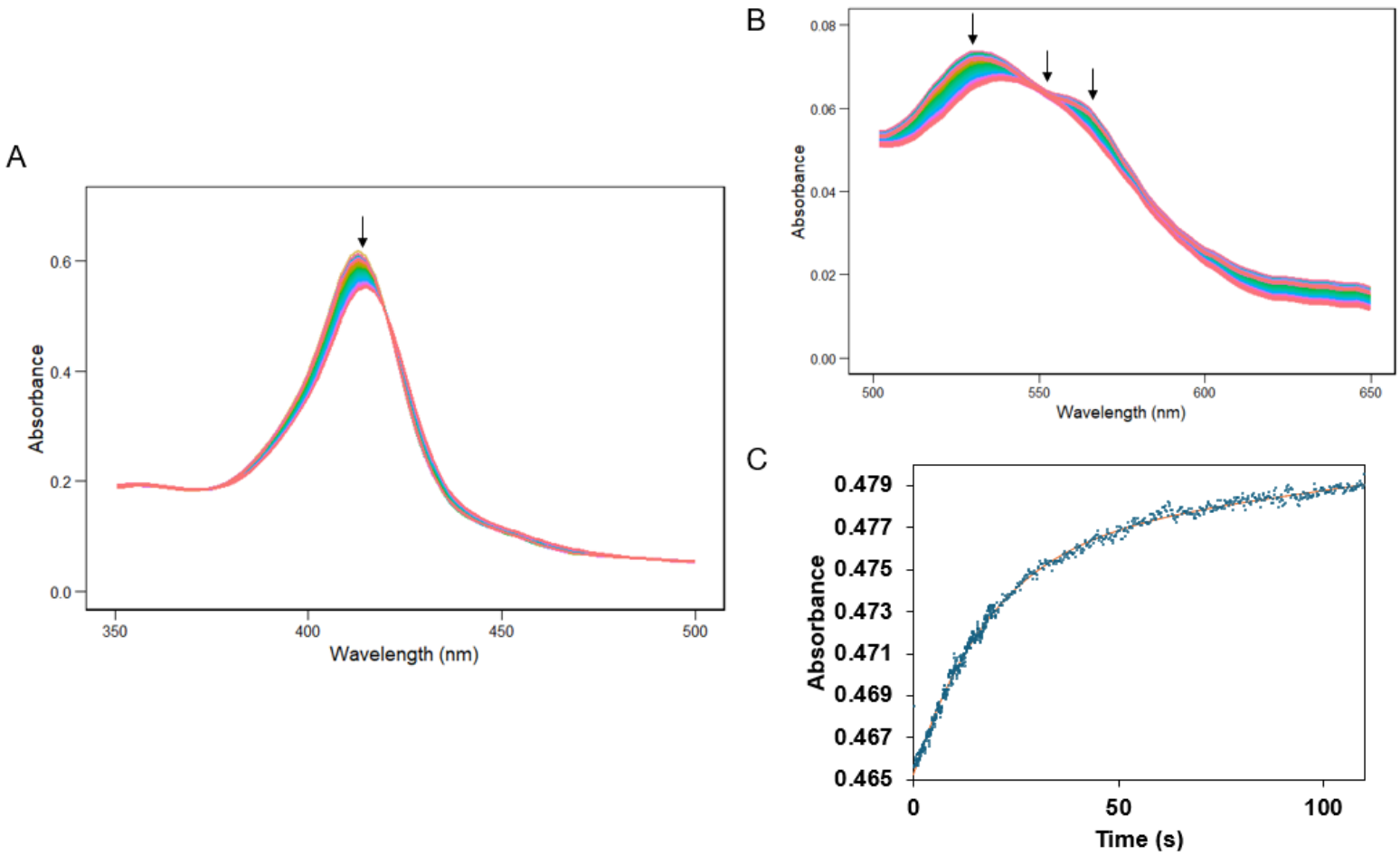


Figure A5.2 The time dependent spectra of CN binding to human Ngb (5 µM) at 20 mM CN concentration.

(A) Optical spectra of Ngb visualised using RStudio. The ferric species and the ferric-CN bond complex formed. (B) Highlight the visible region. (C) The time course for Ngb and the fit at 421 nm. Colour code: fit = orange line, dataset = blue dots. The arrow indicates the direction in which the optical spectra shift.

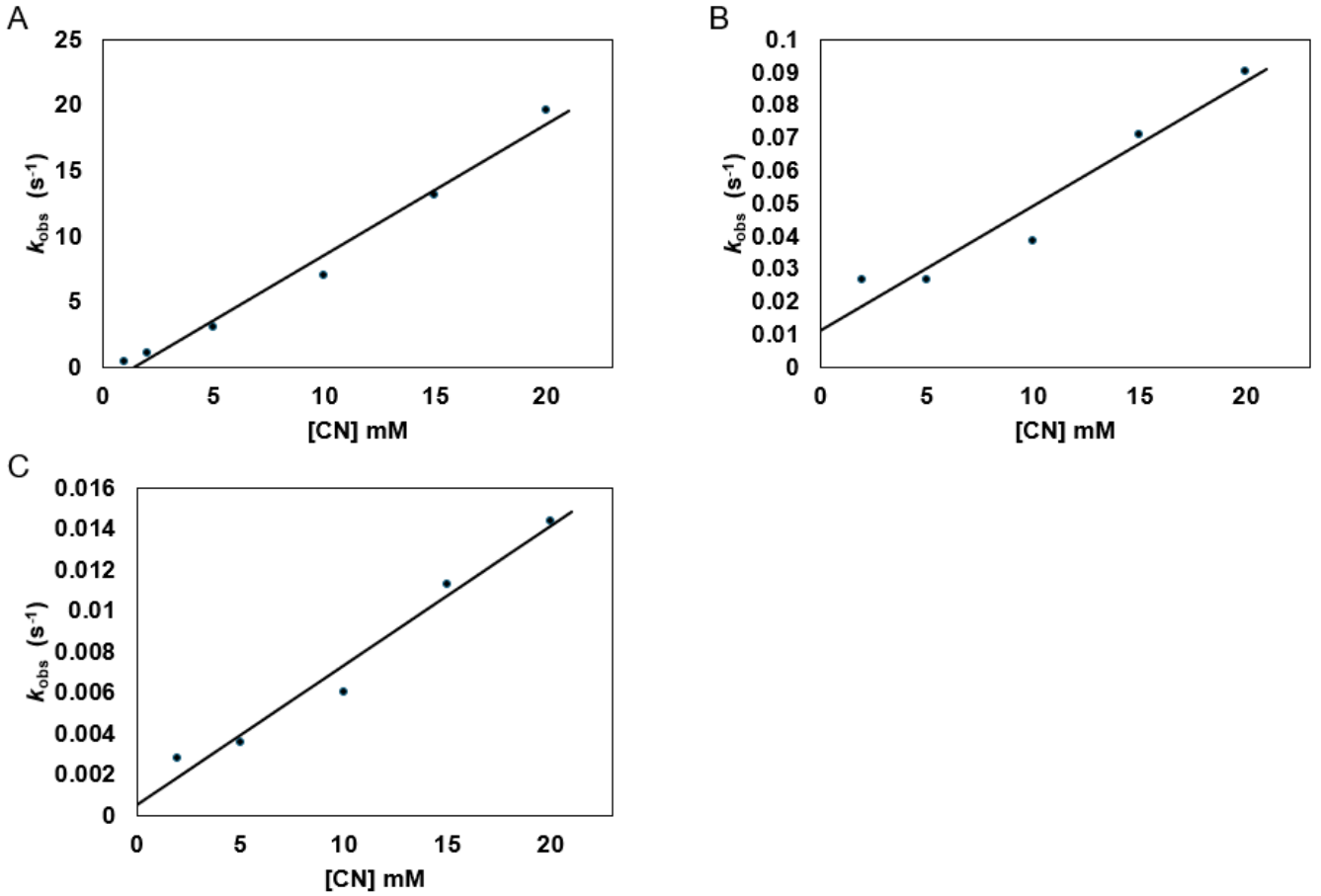


Figure A5.3 Global fitting of Mb and Ngb binding kinetics with CN revealed a linear dependence of rate constants on CN concentration.

(A) Fast phase k_1; of Mb binding with CN in the stopped flow chamber in a 0.1 M sodium phosphate pH 7.4 buffer. k_{on} = 9.97 x 10^2 M^{-1} s^{-1}. (B) Fast phase k_1; of Ngb binding with CN in the stopped flow chamber in a 0.1 M sodium phosphate pH 7.4 buffer. k_{on} = 3.8 M^{-1} s^{-1}, k_{off} = 0.011 s^{-1}. (C) Slow phase k_2; of Ngb binding with CN. k_{on} = 0.681 M^{-1} s^{-1}, k_{off} = 0.00056 s^{-1}.

A6. Nitrite Reductase activity of Myoglobin and Neuroglobin

The formation of the Mb-NO complex led to a shift in the Soret peak from 434 to 420 nm, along with Q-band shifts from a single peak at 560 nm to a beta and alpha band at 547 nm and 580 nm respectively, characteristic of NO-bound ferrous protein (Figure A6.1A) (Shiva *et al*., 2007). The time course (Figure A6.1B) at the two wavelengths (419 nm and 438 nm) of the optical spectrum was fitted to a single exponential function to obtain the rate constants. The observed rate constants were plotted against nitrite concentration (Figure A6.1B), revealing a linear trend.

When deoxyNgb reacted with nitrite in the presence of sodium dithionite under identical conditions to those used for Mb NiR reaction, its optical spectrum also undergoes a transition from deoxyferrous to NO-bound ferrous protein (Figure A6.2A). The formation of the Ngb–NO complex caused a downward shift from the 425 nm deoxy peak as well as the disappearance of the strong beta band at 531 nm and the decreased in intensity of the alpha band at 560 nm (Figure A6.2A), indicating the generation of the iron-nitrosyl species (Fe^{2+}–NO). These optical changes closely resemble those reported in previous studies, particularly the visible α- and β-bands near 550 nm, supporting the presence of a six-coordinate heme configuration in both the ferrous and ferric states of Ngb (Tiso *et al*., 2011).

The time course (Figure A6.2B) at the two wavelengths (413 nm and 426 nm) of the optical spectrum was fitted to a double exponential function to estimate the rate constants. The observed rate constants were plotted against nitrite concentration (Figure A6.2C), revealing a biphasic dependence characterized by fast and slow kinetic phases.

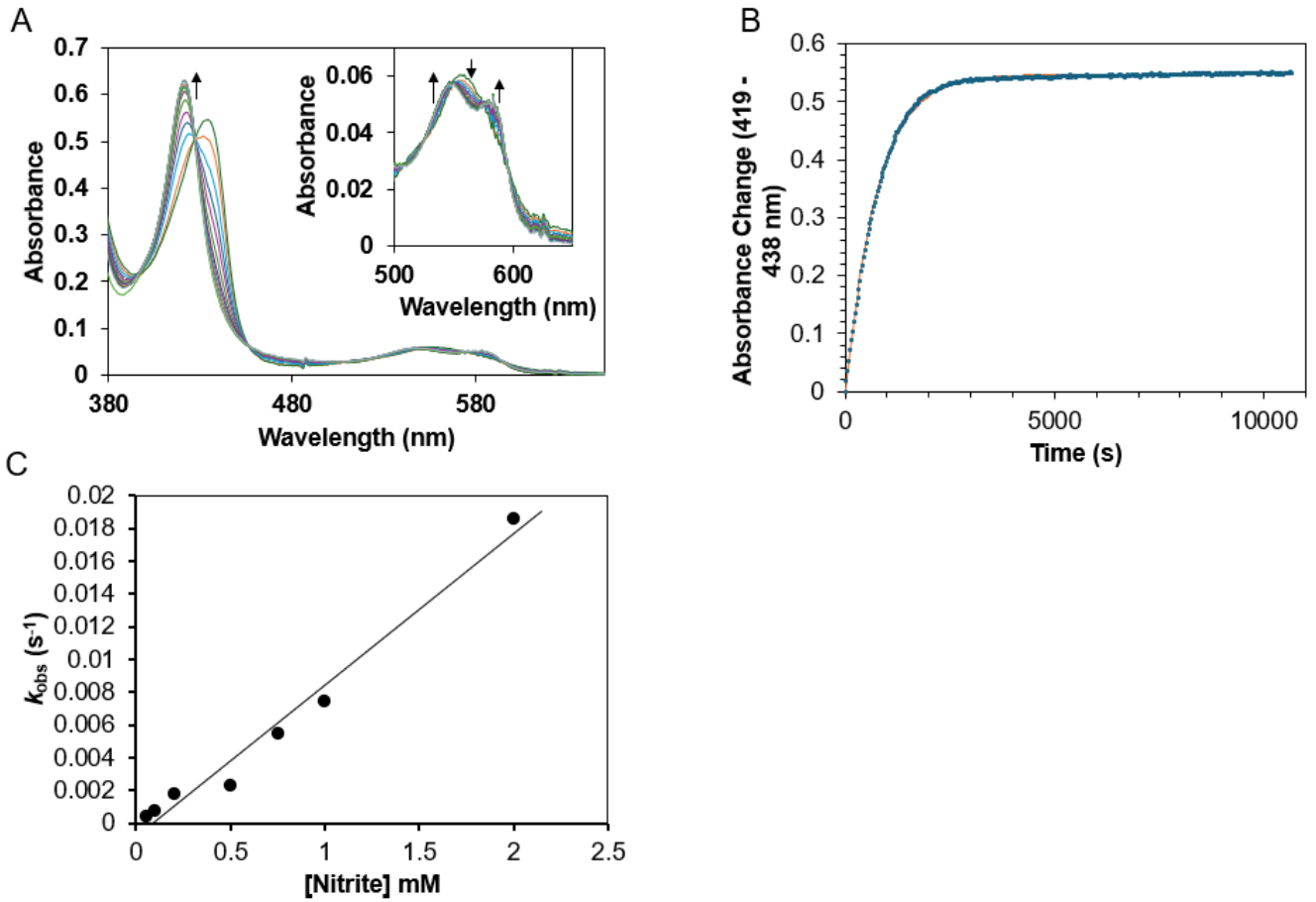


Figure A6.1. Optical changes and kinetics of nitrite reaction with Equine deoxyMb.

(A) UV-Vis spectrum analysis of the anaerobic reaction of 5 µM deoxy-Mb with 1 mM nitrite, pH 7.4, at 25 °C in the presence of dithionite. Inset, absorbance changes at the Q band (enlarged). (B) Rate constants were obtained by fitting the time course (419 – 438 nm) to a single exponential function (0.75 mM nitrite). Colour code: fit = orange, dataset = blue dots. (C) Plot of observed rate constants (k_{obs}) versus nitrite concentration; the second-order bimolecular rate constant obtained from the linear fit of the data is 9.2 M^{-1} s^{-1}.

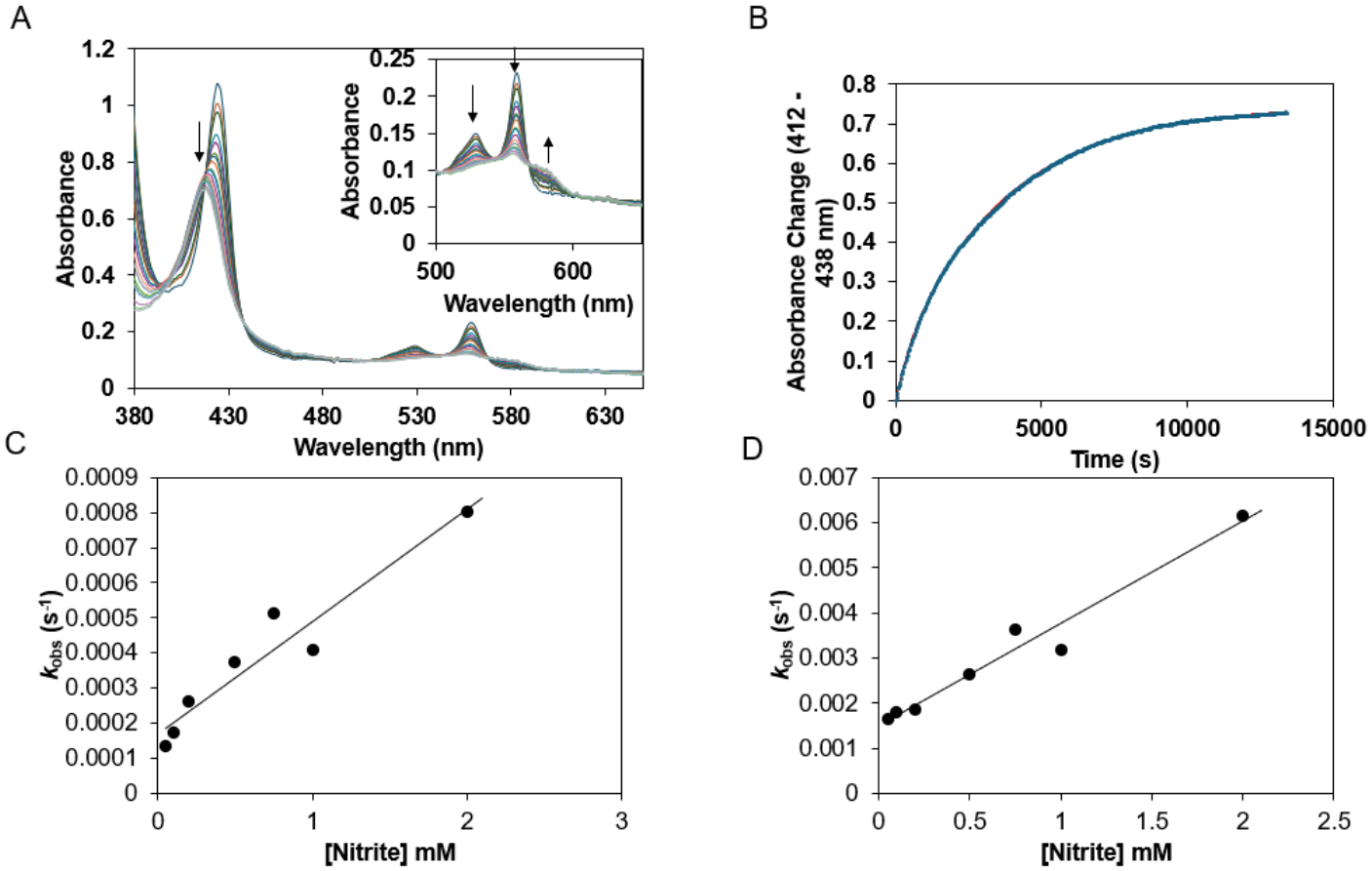


Figure A6.2. Optical changes and kinetics of nitrite reaction with Human deoxyNgb.

(A) UV-Vis spectrum analysis of the anaerobic reaction of 5 μM deoxy-Ngb with 0.75 mM nitrite, pH 7.4, at 25 °C in the presence of dithionite. Inset, absorbance changes at the Q band (enlarged). (B) Rate constants were obtained by fitting the time course (413 – 426 nm) to a double exponential function (0.5 mM nitrite). Colour code: fit = red, dataset = blue dots. (C) Plots of observed rate constants (k_{obs}) versus nitrite concentration; the second-order bimolecular rate constant obtained from the linear fit of both data are 2.27 M^{-1} s^{-1} for the fast phase and (D) 0.32 M^{-1} s^{-1} for the slow phase with k_{off} = 0.0015 s^{-1} and 0.00017 s^{-1} respectively.

REFERENCES

Reference list

Aitken, R., Baker, M. and Nixon, B. (2015). Are sperm capacitation and apoptosis the opposite ends of a continuum driven by oxidative stress? Asian Journal of Andrology, 17(4), p.633. doi:https://doi.org/10.4103/1008-682x.153850.

Aitken, R.J., Jones, K.T. and Robertson, S.A. (2012). Reactive Oxygen Species and Sperm Function--In Sickness and In Health. Journal of Andrology, 33(6), pp.1096–1106. doi:https://doi.org/10.2164/jandrol.112.016535.

Akiyama, M., Murayama, S., Yokoi, K., Yanagisawa, T., Hattori, Y., Yamashiro, Y., Eto, Y. and Fujisawa, K. (2005). Hemoglobin Hammersmith [β 42(CD1) Phe → Ser] causing severe hemolytic anemia in a Japanese girl. Pediatric Blood & Cancer, 47(6), pp.839–841. doi:https://doi.org/10.1002/pbc.20533.

Anikó Lábas, Menyhárd, D.K., Harvey, J.N. and Oláh, J. (2017). First Principles Calculation of the Reaction Rates for Ligand Binding to Myoglobin: The Cases of NO and CO. Chemistry - A European Journal, 24(20), pp.5350–5358. doi:https://doi.org/10.1002/chem.201704867.

AntoniniE and Brunori M. Hemoglobin and myoglobin and their reaction with ligands. In:Frontiers in Biology, edited by Neuberger A and Tatum EL. Amsterdam:North-Holland, 1971, pp. 13–52

Anwised, P., Jangpromma, N., Temsiripong, T., Patramanon, R., Daduang, S., Jitrapakdee, S., Araki, T. and Klaynongsruang, S. (2016). Cloning, Expression, and Characterization of Siamese Crocodile (Crocodylus siamensis) Hemoglobin from Escherichia coli and Pichia pastoris. The Protein Journal, 35(4), pp.256–268. doi:https://doi.org/10.1007/s10930-016-9669-7.

Arthur, J.S.C.,Gauthier, S., & Elce, J.S. (1995). Active site residues in m-calpain:identification by site-directed mutagenesis. FEBS Letters, 362(2),263–266. https://doi.org/10.1016/0014-5793(95)00271-7

Ascenzi, P., di Masi, A., Leboffe, L., Fiocchetti, M., Nuzzo, M.T., Brunori, M. and Marino, M. (2016). Neuroglobin: From structure to function in health and disease. Molecular Aspects of Medicine, 52, pp.1–48. doi:https://doi.org/10.1016/j.mam.2016.10.004.

Baldauf, S.L., Bhattacharya, D., Cockrill, J., P Hugenholtz, Pawlowski, J. and Simpson, A.G.B. (2004). The Tree of Life: An Overview. pp.43–75. doi:https://doi.org/10.1093/oso/9780195172348.003.0005.

Barr, I. and Guo, F. (2015). Pyridine Hemochromagen Assay for Determining the Concentration of Heme in Purified Protein Solutions. BIO-PROTOCOL, 5(18). doi:https://doi.org/10.21769/bioprotoc.1594.

Beckerson, P., Wilson, M., Dimitri A. Svistunenko and Reeder, B. (2014). Cytoglobin 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. doi:https://doi.org/10.1042/bj20140827.

Benchling (2023). Cloud-Based Informatics Platform for Life Sciences R&D. [online] Benchling. Available at: https://www.benchling.com/.

Berg, J.M., Tymoczko, J.L., Gatto, G.J., Stryer, L., Hines, J.K. and Heller, J.B. (2023). Biochemistry. 10th ed. New York: W.H. Freeman & Company, A Macmillan Education Imprint, pp.241–327.

Bernardino, R.L., Carrageta, D.F., Sousa, M., Alves, M.G. and Oliveira, P.F. (2019). pH and male fertility: making sense on pH homeodynamics throughout the male reproductive tract. Cellular and Molecular Life Sciences, 76(19), pp.3783–3800. doi:https://doi.org/10.1007/s00018-019-03170-w.

Bhatwa, A., Wang, W., Hassan, Y.I., Abraham, N., Li, X.Z. and Zhou, T., 2021. Challenges Associated with the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. Frontiers in Bioengineering and Biotechnology, 9, p.65.

Bocahut, A., Derrien, V., Bernad, S., Sebban, P., Sacquin-Mora, S., Guittet, E. and Lescop, E. (2012). Heme orientation modulates histidine dissociation and ligand binding kinetics in the hexacoordinated human neuroglobin. JBIC Journal of Biological Inorganic Chemistry, 18(1), pp.111–122. doi:https://doi.org/10.1007/s00775-012-0956-2.

Bolognesi, M., Bordo, D., Rizzi, M., Tarricone, C. and Ascenzi, P. (1997). Nonvertebrate hemoglobins: Structural bases for reactivity. Progress in Biophysics and Molecular Biology, 68(1), pp.29–68. doi:https://doi.org/10.1016/s0079-6107(97)00017-5.

Bracke, A. (2018). Unraveling the Molecular Function of Androglobin: a testis-specific Globin.

Bracke, A., Van Raemdonck, G., Peeters, K., Punjabi, U. and Dewilde, S., 2018. VI. Expression analysis of androglobin in human spermatozoa and testis tissue. An Bracke, p.141.

Brondyk, W.H. (2009). Chapter 11 Selecting an Appropriate Method for Expressing a Recombinant Protein. Methods in Enzymology, 463, pp.131–147. doi:https://doi.org/10.1016/s0076-6879(09)63011-1.

Brunori, M., Alessandro Giuffrè, Nienhaus, K., Nienhaus, G.U., Scandurra, F.M. and Vallone, B. (2005). Neuroglobin, nitric oxide, and oxygen: Functional pathways and conformational changes. Proceedings of the National Academy of Sciences, 102(24), pp.8483–8488. doi:https://doi.org/10.1073/pnas.0408766102.

Burmester, T. and Hankeln, T., 2014. Functionand evolution of vertebrate globins. Acta Physiologica, 211(3),pp.501-514.

Burmester, T., Weich, B., Reinhardt, S. and Hankeln, T. (2000). A vertebrate globin expressed in the brain. Nature, 407(6803), pp.520–523. doi:https://doi.org/10.1038/35035093.

Bächner, D., Kreienkamp, H.-J. and Richter, D. (2002). MIZIP, a highly conserved, vertebrate specific melanin-concentrating hormone receptor 1 interacting zinc-finger protein1. FEBS Letters, 526(1-3), pp.124–128. doi:https://doi.org/10.1016/s0014-5793(02)03115-0.

Campbell, R.L. and Davies, P.L. (2012). Structure-function relationships in calpains. The Biochemical Journal, [online] 447(3), pp.335–351. doi:https://doi.org/10.1042/BJ20120921.

Carillo, S., Pariat, M., Steff, A.-M., Jariel-Encontre, I., Poulat, F., Berta, P. and Piechaczyk, M. (1996). PEST motifs are not required for rapid calpain-mediated proteolysis of c-fos protein. Biochemical Journal, 313(1), pp.245–251. doi:https://doi.org/10.1042/bj3130245.

Challener, C.A. (2014). Analyzing Proteins Using SEC, MALS, and UHPLC. BioPharm International, [online] 27. Available at: https://www.biopharminternational.com/view/analyzing-proteins-using-sec-mals-and-uhplc? [Accessed 4 Jan. 2025].

Chan, C.-K., Brignole, E.J. and Gibson, W. (2002). Cytomegalovirus Assemblin (pUL80a): Cleavage at Internal Site Not Essential for Virus Growth; Proteinase Absent from Virions.

Journal of Virology, 76(17), pp.8667–8674. doi:https://doi.org/10.1128/jvi.76.17.8667-8674.2002.

Chen, P., Tsuge, H., Almassy, R.J., Gribskov, C.L., Katoh, S., Vanderpool, D.L., Margosiak, S.A., Pinko, C., Matthews, D.A. and Kan, C.-C. (1996). Structure of the Human Cytomegalovirus Protease Catalytic Domain Reveals a Novel Serine Protease Fold and Catalytic Triad. Cell, 86(5), pp.835–843. doi:https://doi.org/10.1016/s0092-8674(00)80157-9.

Cheng, J., Guido Novati, Pan, J., Bycroft, C., Akvilė Žemgulytė, Applebaum, T., Pritzel, A., Lai Hong Wong, Michał Zieliński, Sargeant, T., Schneider, R.G., Senior, A.W., Jumper, J., Demis Hassabis, Kohli, P. and Žiga Avsec (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science, 381(6664). doi:https://doi.org/10.1126/science.adg7492.

Craik, C.S., Largman, C., Fletcher, T., Roczniak, S., Barr, P.J., Fletterick, R. and Rutter, W.J. (1985). Redesigning Trypsin: Alteration of Substrate Specificity. Science, 228(4697), pp.291–297. doi:https://doi.org/10.1126/science.3838593.

Curi, S.M., Ariagno, J.I., Chenlo, P.H., Mendeluk, G.R., Pugliese, M.N., Sardi Segovia, L.M., Repetto, H.E.H. and Blanco, A.M. (2003). Asthenozoospermia: Analysis of a large population. Archives of Andrology, 49(5), pp.343–349. doi:https://doi.org/10.1080/01485010390219656.

Dang, K. (2024). Advantages and Disadvantages of Various Protein Expression Systems. BetaLifeSci. [online] doi:https://doi.org/10.1016/j.biotechadv.2011.09.013%5B2.

Darke, P.L., Cole, J.L., Waxman, L., Hall, D.L., Sardana, M.K. and Kuo, L.C. (1996). Active Human Cytomegalovirus Protease Is a Dimer. Journal of Biological Chemistry, 271(13), pp.7445–7449. doi:https://doi.org/10.1074/jbc.271.13.7445.

Data Analysis Primer | Small Angle X-rayScattering/Diffraction (9th October 2024) www-ssrl.slac.stanford.edu.Available at: https://www-ssrl.slac.stanford.edu/smb-saxs/content/data-analysis-primer.

De Sanctis, D., Dewilde, S., Pesce, A., Ascenzi, P., Burmester, T., Hankeln, T., Moens, L. and Bolognesi, M. (2003). New insight into the haemoglobin superfamily: preliminary crystallographic characterization of human cytoglobin. Acta Crystallographica Section D Biological Crystallography, 59(7), pp.1285–1287. doi:https://doi.org/10.1107/s0907444903009867.

Demon, D., Van Damme, P., Berghe, T.V., Vandekerckhove, J., Declercq, W., Gevaert, K. and Vandenabeele, P. (2009). Caspase substrates: easily caught in deep waters? Trends in Biotechnology, 27(12), pp.680–688. doi:https://doi.org/10.1016/j.tibtech.2009.09.007.

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. doi:https://doi.org/10.1074/jbc.m106438200.

Thermofisher (2022). Tm Calculator - US. [online]www.thermofisher.com. Available at: https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html.

Doorslaer, S.V., Florin Trandafir, Harmer, J.R., Moens, L. and Dewilde, S. (2014). EPR analysis of cyanide complexes of wild-type human neuroglobin and mutants in comparison to horse heart myoglobin. Biophysical Chemistry, 190-191, pp.8–16. doi:https://doi.org/10.1016/j.bpc.2014.03.007.

DuVerle, D.A. and Mamitsuka, H. (2019). CalCleaveMKL: a Tool for Calpain Cleavage Prediction. Methods in Molecular Biology, 1915, pp.121–147. doi:https://doi.org/10.1007/978-1-4939-8988-1_11.

E. Antonini and M. Brunori, in Frontiers in Biology, eds. A.Neuberger and E. L. Tatum, North Holland publishing company, Amsterdam-London,1971, vol. 21, pp. 13-52.

Edwards, M.J., Nanakow Baiden, Johs, A., S.J. Tomanicek, Liang, L., Shi, L., Fredrickson, J.K., Zachara, J.M., Gates, A.J., Butt, J.N., Richardson, D.J. and Clarke, T.A. (2014). The X-ray crystal structure of Shewanella oneidensis OmcA reveals new insight at the microbe–mineral interface. FEBS Letters, 588(10), pp.1886–1890. doi:https://doi.org/10.1016/j.febslet.2014.04.013.

Exertier, C., Sebastiani, F., Freda, I., Gugole, E., Cerutti, G., Parisi, G., Montemiglio, L.C., Becucci, M., Viappiani, C., Bruno, S., Savino, C., Zamparelli, C., Anselmi, M., Abbruzzetti, S., Smulevich, G. and Vallone, B. (2022). Probing the Role of Murine Neuroglobin CDloop–D-Helix Unit in CO Ligand Binding and Structural Dynamics. ACS Chemical Biology, 17(8), pp.2099–2108. doi:https://doi.org/10.1021/acschembio.2c00172.

Ezhevskaya, M., Florin Trandafir, Moens, L., Dewilde, S. and Doorslaer, S.V. (2011). EPR investigation of the role of B10 phenylalanine in neuroglobin — Evidence that B10Phe mediates structural changes in the heme region upon disulfide-bridge formation. Journal of Inorganic Biochemistry, 105(9), pp.1131–1137. doi:https://doi.org/10.1016/j.jinorgbio.2011.05.020.

Fago, A., Hundahl, C., Dewilde, S., Gilany, K., Moens, L. and Weber, R.E. (2004). Allosteric Regulation and Temperature Dependence of Oxygen Binding in Human Neuroglobin and Cytoglobin. Journal of Biological Chemistry, 279(43), pp.44417–44426. doi:https://doi.org/10.1074/jbc.m407126200.

Fago, A., Mathews, A.J., Moens, L., Dewilde, S. and Brittain, T. (2006). The reaction of neuroglobin with potential redox protein partners cytochrome b 5 and cytochrome c. FEBS Letters, 580(20), pp.4884–4888. doi:https://doi.org/10.1016/j.febslet.2006.08.003.

Feelisch, M. and Stamler, J.S. (1996). Methods in Nitric Oxide Research; in The Oxyhemoglobin Assay. 1st ed. Wiley, London, pp.455–478.

Fernández-Montalván, A. *et al*. (2004) 'Electrostaticinteractions of domain III stabilize the inactive conformation of μ-calpain',Biochemical Journal, 382(2), pp. 607–617. doi:10.1042/bj20040731.

Firer-Sherwood, M.A., Ando, N., Drennan, C.L. and Elliott, S.J. (2011). Solution-Based Structural Analysis of the Decaheme Cytochrome, MtrA, by Small-Angle X-ray Scattering and Analytical Ultracentrifugation. The Journal of Physical Chemistry B, 115(38), pp.11208–11214. doi:https://doi.org/10.1021/jp203603r.

Fischer, H. and Zeile, K. (1929). Synthese des Hämatoporphyrins, Protoporphyrins und Hämins. Justus Liebigs Annalen der Chemie, 468(1), pp.98–116. doi:https://doi.org/10.1002/jlac.19294680104.

Francke, F., Buck, F. and Dietmar Bächner (2005). MYND domain specific interaction of the melanin-concentrating hormone receptor 1 interacting zinc-finger protein with α- and β-tubulin. Biochemical and Biophysical Research Communications, 334(4), pp.1292–1298. doi:https://doi.org/10.1016/j.bbrc.2005.07.027.

Freitas, T.A.K., Saito, J.A., Wan, X., Hou, S. and Alam, M., 2008. Protoglobin and globin-coupled sensors. In The smallest biomolecules: diatomics and their interactions with heme proteins (pp. 175-202). Elsevier.

Freitas, T.A.K., Saito, J.A., Wan, X., Hou, S.and Alam, M., 2008. Protoglobin and globin-coupled sensors. In Thesmallest biomolecules: diatomics and their interactions with heme proteins (pp.175-202). Elsevier.

Gal, J., Dufour, A., Young, D., Yang, E.S. and Geddes, J.W. (2025). The Identification of Proteolytic Substrates of Calpain-5 with N-Terminomics. International Journal of Molecular Sciences, 26(13), p.6459. doi:https://doi.org/10.3390/ijms26136459.

Gardner, P.R., Gardner, A.M., Martin, L.A. and Salzman, A.L. (1998). Nitric oxide dioxygenase: An enzymic function for flavohemoglobin. Proceedings of the National Academy of Sciences, 95(18), pp.10378–10383. doi:https://doi.org/10.1073/pnas.95.18.10378.

Garman, E.F. (2010). Radiation damage in macromolecular crystallography: what is it and why should we care? Acta Crystallographica Section D Biological Crystallography, 66(4), pp.339–351. doi:https://doi.org/10.1107/s0907444910008656.

Gasteiger E., Gattiker A., Hoogland C., IvanyiI., Appel R.D., Bairoch A.ExPASy: the proteomics server for in-depthprotein knowledge and analysisNucleic Acids Res. 31:3784-3788(2003).

Geisse, S. and Fux, C. (2009). Chapter 15 Recombinant Protein Production by Transient Gene Transfer into Mammalian Cells. Methods in Enzymology, pp.223–238. doi:https://doi.org/10.1016/s0076-6879(09)63015-9.

George, P. and Irvine, D.H. (1956). A kinetic study of the reaction between ferrimyoglobin and hydrogen peroxide. Journal of Colloid Science, 11(4-5), pp.327–339. doi:https://doi.org/10.1016/0095-8522(56)90151-9.

Gonzalez, M.A., Fry, N.L., Burt, R., Davda, R., Hobbs, A. and Mascharak, P.K. (2011). Designed Iron Carbonyls as Carbon Monoxide (CO) Releasing Molecules: Rapid CO Release and Delivery to Myoglobin in Aqueous Buffer, and Vasorelaxation of Mouse Aorta. Inorganic Chemistry, 50(7), pp.3127–3134. doi:https://doi.org/10.1021/ic2000848.

Gow, A.J., Luchsinger, B.P., Pawloski, J.R., Singel, D.J. and Stamler, J.S. (1999). The oxyhemoglobin reaction of nitric oxide. Proceedings of the National Academy of Sciences, 96(16), pp.9027–9032. doi:https://doi.org/10.1073/pnas.96.16.9027.

Gray, D. (1997). Overview of Protein Expression by Mammalian Cells. Current Protocols in Protein Science, [online] 10(1). doi:https://doi.org/10.1002/0471140864.ps0509s10.

Hamdane, D., Kiger, L., Dewilde, S., Green, B.N., Pesce, A., Uzan, J., Burmester, T., Hankeln, T., Bolognesi, M., Moens, L. and Marden, M.C. (2003). The Redox State of the Cell Regulates the Ligand Binding Affinity of Human Neuroglobin and Cytoglobin. Journal of Biological Chemistry, 278(51), pp.51713–51721. doi:https://doi.org/10.1074/jbc.m309396200.

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. FEBS Journal, [online] 272(8), pp.2076–2084. doi:https://doi.org/10.1111/j.1742-4658.2005.04635.x.

Hankeln, T. and Burmester, T., 2008. Neuroglobin and cytoglobin. In The smallest biomolecules: diatomics and their interactions with heme proteins (pp. 203-218). Elsevier.

Hanna, R.A., Campbell, R.L. and Davies, P.L. (2008). Calcium-bound structure of calpain and its mechanism of inhibition by calpastatin. Nature, [online] 456(7220), pp.409–412. doi:https://doi.org/10.1038/nature07451.

Hargrove, M.S., Krzywda, S., Wilkinson, A.J., Dou, Y., Masao Ikeda-Saito and Olson, J.S. (1994). Stability of Myoglobin: A Model for the Folding of Heme Proteins. Biochemistry, 33(39), pp.11767–11775. doi:https://doi.org/10.1021/bi00205a012.

Hausladen, A., Gow, A.J. and Stamler, J.S. (1998). Nitrosative stress: Metabolic pathway involving the flavohemoglobin. Proceedings of the National Academy of Sciences, 95(24), pp.14100–14105. doi:https://doi.org/10.1073/pnas.95.24.14100.

Herold, S., Fago, A., Weber, R.E., Dewilde, S. and Moens, L. (2004). Reactivity Studies of the Fe(III) and Fe(II)NO Forms of Human Neuroglobin Reveal a Potential Role against Oxidative Stress. Journal of Biological Chemistry, 279(22), pp.22841–22847. doi:https://doi.org/10.1074/jbc.m313732200.

Hjertén, S., Sparrman, M. and Liao, J. (1988). Purification of membrane proteins in SDS and subsequent renaturation. Biochimica Et Biophysica Acta, [online] 939(3), pp.476–484. doi:https://doi.org/10.1016/0005-2736(88)90094-6.

Hoogewijs, D., Ebner, B., Germani, F., Hoffmann, F.G., Fabrizius, A., Moens, L., Burmester, T., Dewilde, S., Storz, J.F., Vinogradov, S.N. & Hankeln, T. 2012. Androglobin: a chimeric globin in metazoans that is preferentially expressed in Mammalian testes. Mol Biol Evol 29, 1105– 1114.

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), p.2179.

Ishikawa, H., Finkelstein, I.J., Kim, S., Kwak, K., Chung, J.K., Wakasugi, K., Massari, A.M. and Fayer, M.D. (2007). Neuroglobin dynamics observed with ultrafast 2D-IR vibrational echo spectroscopy. Proceedings of the National Academy of Sciences, 104(41), pp.16116–16121. doi:https://doi.org/10.1073/pnas.0707718104.

Jeffreys, A.J., Wilson, V., Alain Blanchetot, Weller, P., Kessel, van, Spurr, N., Solomon, E. and Goodfellow, P. (1984). The human myoglobin gene: a third dispersed globin locus in the human genome. Nucleic Acids Research, [online] 12(7), pp.3235–3243. doi:https://doi.org/10.1093/nar/12.7.3235.

Jensen, F.B. (2009). The dual roles of red blood cells in tissue oxygen delivery: oxygen carriers and regulators of local blood flow. Journal of Experimental Biology, 212(21), pp.3387–3393. doi:https://doi.org/10.1242/jeb.023697.

Kalderon, D.,Roberts, B. L., Richardson, W. D., & Smith, A. E. (1984). A short aminoacid sequence able to specify nuclear location. Cell, 39(3 Pt 2),499–509. DOI: 10.1016/0092-8674(84)90457-4

Kelm, M. (1999). Nitric oxide metabolism and breakdown. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1411(2-3), pp.273–289. doi:https://doi.org/10.1016/s0005-2728(99)00020-1.

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. doi:https://doi.org/10.1038/181662a0.

Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C. and Shore, V.C. (1960). Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. Nature, 185(4711), pp.422–427. doi:https://doi.org/10.1038/185422a0.

Keppner, A., Correia, M., Santambrogio, S., Koay, T.W., Maric, D., Osterhof, C., Winter, D.V., Clerc, A., Stumpe, M., Chalmel, F., Dewilde, S., Odermatt, A., Kressler, D., Hankeln, T., Wenger, R.H. and Hoogewijs, D. (2022). Androglobin, a chimeric mammalian globin, is required for male fertility. eLife, [online] 11, p.e72374. doi:https://doi.org/10.7554/eLife.72374.

Kerr, E.A., Yu, N.T., Bartnicki, D.E. and Mizukami, H. (1985). Resonance Raman studies of CO and O2 binding to elephant myoglobin (distal His(E7)—-Gln). Journal of Biological Chemistry, 260(14), pp.8360–8365. doi:https://doi.org/10.1016/s0021-9258(17)39481-4.

Kiger, L., Uzan, J., Dewilde, S., Burmester, T., Hankeln, T., Moens, L., Hamdane, D., Baudin-Creuza, V. and Marden, M. (2004). Neuroglobin Ligand Binding Kinetics. IUBMB Life, 56(11), pp.709–719. doi:https://doi.org/10.1080/15216540500037711.

Koay, T.W., Osterhof, C., Orlando, I.M.C., Keppner, A., Andre, D., Yousefian, S., Suárez Alonso, M., Correia, M., Markworth, R., Schödel, J., Hankeln, T. and Hoogewijs, D. (2021). Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis. Journal of Biological Chemistry, 296, p.100291. doi:https://doi.org/10.1016/j.jbc.2021.100291.

Kuriyan, J., Pahler, A., Simon, R.J., Toshio. Kokubo and Kaiser, E.T. (1988). The x-ray structure of flavohemoglobin. A semisynthetic hydroxylase. Journal of the American Chemical Society, 110(18), pp.6261–6263. doi:https://doi.org/10.1021/ja00226a060.

Lange, A., Mills, R.E., Lange, C. J., Stewart, M., Devine, S. E., & Corbett, A. H. (2007). Classicalnuclear localization signals: definition, function, and interaction withimportin a. J Biol Chem, 282(8), 5101–5105. DOI:10.1074/jbc.R600026200

Lardinois, O.M., Tomer, K.B., Mason, R.P. and Deterding, L.J. (2008). Identification of Protein Radicals Formed in the Human Neuroglobin−H2O2 Reaction Using Immuno-Spin Trapping and Mass Spectrometry. Biochemistry, 47(39), pp.10440–10448. doi:https://doi.org/10.1021/bi800771k.

Laverman, L.E., Wanat, A., Janusz Oszajca, Grażyna Stochel, Ford, P.C. and Rudi van Eldik (2001). Mechanistic Studies on the Reversible Binding of Nitric Oxide to Metmyoglobin. Journal of the American Chemical Society, 123(2), pp.285–293. doi:https://doi.org/10.1021/ja001696z.

Lazrak, A. *et al*. (2013) 'The silent codon change i507-atc?attcontributes to the severity of the ?F508 CFTR channel dysfunction', The FASEBJournal, 27(11), pp. 4630–4645. doi:10.1096/fj.13-227330.

Lechauve, C., Chauvierre, C., Dewilde, S., Moens, L., Green, B.N., Marden, M.C., Célier, C. and Kiger, L. (2010). Cytoglobin conformations and disulfide bond formation. FEBS Journal, 277(12), pp.2696–2704. doi:https://doi.org/10.1111/j.1742-4658.2010.07686.x.

Lee, N.P.Y. and Cheng, C.Y. (2009). Nitric Oxide and Cyclic Nucleotides: Their Roles in Junction Dynamics and Spermatogenesis. Advances in experimental medicine and biology, vol 636, pp.172–185. doi:https://doi.org/10.1007/978-0-387-09597-4_10.

Lei, R., Butcher, D., Bernad, S., Derrien, V. and Miksovska, J. (2019). Mutation of Residues in CD Loop and Distal Pocket Impact Protein Stability of Human Neuroglobin. Biophysical Journal, 116(3), p.471a. doi:https://doi.org/10.1016/j.bpj.2018.11.2545.

Levantino, M., Schirò, G., Lemke, H.T., Cottone, G., Glownia, J.M., Zhu, D., Chollet, M., Ihee, H., Cupane, A. and Cammarata, M. (2015). Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. Nature Communications, 6(1). doi:https://doi.org/10.1038/ncomms7772.

Levine, N. and Kelly, H. (1978). Measurement of pH in the rat epididymis in vivo. Reproduction, 52(2), pp.333–335. doi:https://doi.org/10.1530/jrf.0.0520333.

Li, F., Chen, Y., Qi, Q., Wang, Y., Yuan, L., Huang, M., Elsemman, I.E., Feizi, A., Kerkhoven, E.J. and Nielsen, J. (2022). Improving recombinant protein production by yeast through genome-scale modeling using proteome constraints. Nature Communications, 13(1). doi:https://doi.org/10.1038/s41467-022-30689-7.

Lilly, B., Dammeyer, K., Marosis, S., McCallinhart, P.E., Trask, A.J., Lowe, M. and Sawant, D., 2018. Endothelial cell-induced cytoglobin expression in vascular smooth muscle cells contributes to modulation of nitric oxide. Vascular pharmacology, 110, pp.7-15.

Liu, L., Martínez, J.L., Liu, Z., Petranovic, D. and Nielsen, J. (2014). Balanced globin protein expression and heme biosynthesis improve production of human hemoglobin in Saccharomyces cerevisiae. Metabolic Engineering, 21, pp.9–16. doi:https://doi.org/10.1016/j.ymben.2013.10.010.

Liu, X., El-Mahdy, M.A., Boslett, J., Varadharaj, S., Hemann, C., Abdelghany, T.M., Ismail, R.S., Little, S.C., Zhou, D., Thuy, L.T.T. and Kawada, N., 2017. Cytoglobin regulates blood pressure and vascular tone through nitric oxide metabolism in the vascular wall. Nature communications, 8(1), pp.1-14.

Lo, W.C. and Lyu, P.C., 2008. CPSARST: an efficient circular permutation search tool applied to the detection of novelprotein structural relationships. Genome biology, 9(1),pp.1-16.

Maemoto, Y., Kiso, S., Shibata, H. and Maki, M. (2013). Analysis of limited proteolytic activity of calpain-7 using non-physiological substrates in mammalian cells. FEBS Journal, 280(11), pp.2594–2607. doi:https://doi.org/10.1111/febs.12243.

Manole, A., Kekilli, D., Svistunenko, D.A., Wilson, M.T., Dobbin, P.S. and Hough, M.A. (2015). Conformational control of the binding of diatomic gases to cytochrome c'. Journal of biological inorganic chemistry: JBIC: a publication of the Society of Biological Inorganic Chemistry, [online] 20(4), pp.675–686. doi:https://doi.org/10.1007/s00775-015-1253-7.

Marcero, J.R., Piel, R.B., Burch, J.S. and Dailey, H.A. (2016). Rapid and sensitive quantitation of heme in hemoglobinized cells. BioTechniques, 61(2), pp.83–91. doi:https://doi.org/10.2144/000114444.

MassSpectrometry. Biochemistry, 47(39), pp.10440–10448.doi:https://doi.org/10.1021/bi800771k

Mathews, A.J., Rohlfs, R.J., Olson, J.S., Tame, J., Renaud, J.P. and Nagai, K. (1989). The effects of E7 and E11 mutations on the kinetics of ligand binding to R state human hemoglobin. Journal of Biological Chemistry, 264(28), pp.16573–16583. doi:https://doi.org/10.1016/s0021-9258(19)84745-2.

Menkveld, R. (2010). Clinical significance of the low normal sperm morphology value as proposed in the fifth edition of the WHO Laboratory Manual for the Examination and Processing of Human Semen. Asian Journal of Andrology, 12(1), pp.47–58. doi:https://doi.org/10.1038/aja.2009.14.

Miller, Y.I. and Shaklai, N. (1999). Kinetics of hemin distribution in plasma reveals its role in lipoprotein oxidation. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1454(2), pp.153–164. doi:https://doi.org/10.1016/s0925-4439(99)00027-7.

Moldoveanu, T., etal. (2002). Structure of calpain reveals that the active site is formed by a flexible interdomain interface. EMBO Journal, 21(13), 3363–3370.https://doi.org/10.1093/emboj/cdf325

Molinari, M., Anagli, J. and Carafoli, E. (1995). PEST Sequences Do Not Influence Substrate Susceptibility to Calpain Proteolysis. Journal of Biological Chemistry, 270(5), pp.2032–2035. doi:https://doi.org/10.1074/jbc.270.5.2032.

Moore, E.G. and Gibson, Q.H. (1976). Cooperativity in the dissociation of nitric oxide from hemoglobin. Journal of Biological Chemistry, 251(9), pp.2788–2794. doi:https://doi.org/10.1016/s0021-9258(17)33557-3.

Morishima, K., Inoue, R., Nakagawa, T., Shimizu, M., Sakamoto, R., Oda, T., Koichi Mayumi and Sugiyama, M. (2025). Size-exclusion chromatography–small-angle neutron scattering system optimized for an instrument with medium neutron flux. Journal of Applied Crystallography, 58(2). doi:https://doi.org/10.1107/s1600576725000779.

MUIRHEAD, H. and PERUTZ, M.F. (1963). Structure Of Hæmoglobin: A Three-Dimensional Fourier Synthesis of Reduced Human Haemoglobin at 5.5 Å Resolution. Nature, 199(4894), pp.633–638. doi:https://doi.org/10.1038/199633a0.

Nachiappan, M. *et al*. (2021) '3D StructuralDetermination of Macromolecules Using X-ray Crystallography Methods', in MolecularDocking for Computer-Aided Drug Design: Fundamentals, Techniques, Resources andApplications. doi: 10.1016/B978-0-12-822312-3.00007-2.

Nedergaard Pedersen, J., Lyngsø, J., Zinn, T., E. Otzen, D. and Skov Pedersen, J. (2020). A complete picture of protein unfolding and refolding in surfactants. Chemical Science, [online] 11(3), pp.699–712. doi:https://doi.org/10.1039/C9SC04831F.

Nicolis, S., Monzani, E., Ciaccio, C., Ascenzi, P., Moens, L. and Casella, L. (2007). Reactivity and endogenous modification by nitrite and hydrogen peroxide: does human neuroglobin act only as a scavenger? Biochemical Journal, 407(1), pp.89–99. doi:https://doi.org/10.1042/bj20070372.

Nie, L.-S., Liu, X.-C., Han, H., Ren, Z.-H., Gao, S.-Q. and Lin, Y.-W. (2024). Calmodulin interacts with androglobin and regulates the nitrite reductase activity. RSC Chemical Biology. doi:https://doi.org/10.1039/d4cb00245h.

Nilsson, T., Jackson, M., & Peterson, P. A. (1989). Retention of disulfide-linked oligomers in the endoplasmic reticulum. Cell,58(4), 707–718. DOI: 10.1016/0092-8674(89)90104-7

Olson, J.S. and Phillips, G.N. (1996). Kinetic Pathways and Barriers for Ligand Binding to Myoglobin. Journal of Biological Chemistry, 271(30), pp.17593–17596. doi:https://doi.org/10.1074/jbc.271.30.17593.

Ono, Y. and Sorimachi, H., 2012. Calpains—Anelaborate proteolytic system. Biochimica et Biophysica Acta(BBA)-Proteins and Proteomics, 1824(1), pp.224-236.

Osako, Y., Maemoto, Y., Tanaka, R., Suzuki, H., Shibata, H. and Maki, M. (2010). Autolytic activity of human calpain 7 is enhanced by ESCRT-III-related protein IST1 through MIT–MIM interaction. FEBS Journal, 277(21), pp.4412–4426. doi:https://doi.org/10.1111/j.1742-4658.2010.07822.x.

Ota, M., Isogai, Y. and Nishikawa, K., 1997. Structural requirement of highly conserved residues in globins. FEBS letters, 415(2), pp.129-133.

Otto, H.-H. and Schirmeister, T. (1997). Cysteine Proteases and Their Inhibitors. Chemical Reviews, 97(1), pp.133–172. doi:https://doi.org/10.1021/cr950025u.

Ouellet, H., Milani, M., LaBarre, M., Martino Bolognesi, Couture, M. and Guertin, M. (2007). The Roles of Tyr(CD1) and Trp(G8) in Mycobacterium tuberculosis Truncated Hemoglobin O in Ligand Binding and on the Heme Distal Site Architecture,. Biochemistry, 46(41), pp.11440–11450. doi:https://doi.org/10.1021/bi7010288.

Pek, J.W., Anand, A. and Kai, T. (2012). Tudor domain proteins in development. Development, 139(13), pp.2255–2266. doi:https://doi.org/10.1242/dev.073304.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G. and North, A.C. (1960). Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature, [online] 185(4711), pp.416–422. doi:https://doi.org/10.1038/185416a0.

Pesce, A., Bolognesi, M., Bocedi, A., Ascenzi, P., Dewilde, S., Moens, L., Hankeln, T. and Burmester, T. (2002). Neuroglobin and cytoglobin. EMBO reports, 3(12), pp.1146–1151. doi:https://doi.org/10.1093/embo-reports/kvf248.

Pesce, A., Dewilde, S., Nardini, M., Moens, L., Ascenzi, P., Hankeln, T., Burmester, T. and Bolognesi, M. (2004). The human brain hexacoordinated neuroglobin three-dimensional structure. Micron, 35(1-2), pp.63–65. doi:https://doi.org/10.1016/j.micron.2003.10.013.

Pesce, A., Nardini, M., Milani, M. and Bolognesi, M. (2007). Protein structure in the truncated (2/2) hemoglobin family. IUBMB Life, 59(8-9), pp.535–541. doi:https://doi.org/10.1080/15216540701225933.

Petersen, M.G., Dewilde, S. and Fago, A. (2008). Reactions of ferrous neuroglobin and cytoglobin with nitrite under anaerobic conditions. Journal of Inorganic Biochemistry, 102(9), pp.1777–1782. doi:https://doi.org/10.1016/j.jinorgbio.2008.05.008.

Pham, P.V. (2018). Medical Biotechnology. Omics Technologies and Bio-Engineering, pp.449–469. doi:https://doi.org/10.1016/b978-0-12-804659-3.00019-1.

Putnam, C.D., Hammel, M., Hura, G.L. and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solutions. Quarterly Reviews of Biophysics, [online] 40(3), pp.191–285. doi:https://doi.org/10.1017/S0033583507004635.

Reeder, B., Deganutti, G., Ukeri, J., Atanasio, S., Svistunenko, D., Ronchetti, C., Mobarec, J.C., Welbourn, E.M., Vos, M., Asaju, J.O., Wilson, M. and Reynolds, C.A. (2024). The circularly permuted globin domain of Androglobin exhibits atypical heme stabilization and nitric oxide interaction. Chemical Science, [online] 15, 6738-6751. doi:https://doi.org/10.1039/D4SC00953C.

Reeder, B.J. and Ukeri, J. (2018). Strong modulation of nitrite reductase activity of cytoglobin by disulfide bond oxidation: Implications for nitric oxide homeostasis. Nitric Oxide, 72, pp.16–23. doi:https://doi.org/10.1016/j.niox.2017.11.004.

Reeder, B.J., 2010. The redox activity of hemoglobins: from physiologic functions to pathologic mechanisms. Antioxidants & redox signaling, 13(7), pp.1087-1123.

Reeder, B.J., Svistunenko, D.A. and Wilson, M.T. (2011). Lipid binding to cytoglobin leads to a change in haem co-ordination: a role for cytoglobin in lipid signalling of oxidative stress. The Biochemical Journal, [online] 434(3), pp.483–492. doi:https://doi.org/10.1042/BJ20101136.

Reeder, B.J., Svistunenko, D.A., Sharpe, M.A. and Wilson, M.T. (2001). Characteristics and Mechanism of Formation of Peroxide-Induced Heme to Protein Cross-Linking in Myoglobin. Biochemistry, 41(1), pp.367–375. doi:https://doi.org/10.1021/bi011335b.

Reeder, J.B. and Wilson, T.M. (1998). Mechanism of reaction of myoglobin with the lipid hydroperoxide hydroperoxyoctadecadienoic acid. Biochemical Journal, 330(3), pp.1317–1323. doi:https://doi.org/10.1042/bj3301317.

Roesner, A., Fuchs, C., Hankeln, T. and Burmester, T. (2004). A Globin Gene of Ancient Evolutionary Origin in Lower Vertebrates: Evidence for Two Distinct Globin Families in Animals. Molecular Biology and Evolution, 22(1), pp.12–20. doi:https://doi.org/10.1093/molbev/msh258.

Rohlfs, R.J., Olson, J.S. and Gibson, Q.H. (1988). A comparison of the geminate recombination kinetics of several monomeric heme proteins. Journal of biological chemistry/□The □Journal of biological chemistry, 263(4), pp.1803–1813. doi:https://doi.org/10.1016/s0021-9258(19)77948-4.

Rosano, G.L. and Ceccarelli, E.A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. Frontiers in Microbiology, [online] 5(172). doi:https://doi.org/10.3389/fmicb.2014.00172.

RStudio Team (2020). RStudio | Open Source & Professional Software for Data Science Teams. [online] rstudio.com. Available at: http://www.rstudio.com/.

Sakashita, G., Kiyoi, H., Naoe, T. and Urano, T. (2018). Analysis of the oligomeric states of nucleophosmin using size exclusion chromatography. Scientific Reports, 8(1). doi:https://doi.org/10.1038/s41598-018-22359-w.

Sanctis, D. de, Dewilde, S., Pesce, A., Moens, L., Ascenzi, P., Hankeln, T., Burmester, T. and Bolognesi, M. (2004). Crystal Structure of Cytoglobin: The Fourth Globin Type Discovered in Man Displays Heme Hexa-coordination. Journal of Molecular Biology, 336(4), pp.917–927. doi:https://doi.org/10.1016/j.jmb.2003.12.063.

Sasha De Henau, Tilleman, L., Vangheel, M., Evi Luyckx, Stanislav Trashin, Pauwels, M., Germani, F., Vlaeminck, C., Vanfleteren, J.R., Bert, W., Pesce, A., Nardini, M., Martino Bolognesi, Karolien De Wael, Luc Moëns, Dewilde, S. and Braeckman, B.P. (2015). A redox signalling globin is essential for reproduction in Caenorhabditis elegans. Nature Communications, 6(1). doi:https://doi.org/10.1038/ncomms9782.

Schulz, G.E. and Schirmer, R.H. (1984). Hemoglobin: Structure, Function, Evolution, and Pathology. Von R. E. Dickerson und I. Geis. Benjamin/Cummings Publishing Co., Menlo Park, CA 1983. 176 S., geb. $ 40.50 - ISBN 0-8053-2411-9. Angewandte Chemie, 96(12), pp.1003–1004. doi:https://doi.org/10.1002/ange.19840961240.

Schwarze,K. & Burmester, T. 2013. Conservation of globin genes in the “livingfossil” Latimeria chalumnae and reconstruction of the evolution of thevertebrate globin family. Biochim Biophys Acta 1834, 1801–1812

Shepherd, M., Bernhardt, P.V. and Poole, R.K. (2011). Globin-mediated nitric oxide detoxification in the foodborne pathogenic bacterium Campylobacter jejuni proceeds via a dioxygenase or denitrosylase mechanism. Nitric Oxide, 25(2), pp.229–233. doi:https://doi.org/10.1016/j.niox.2010.12.006.

Shi, H.-J., Wu, A.Z., Santos, M., Feng, Z.-M., Huang, L., Chen, Y.-M., Zhu, K. and Chen, C.-L.C. (2004). Cloning and Characterization of Rat Spermatid Protein SSP411: A Thioredoxin-Like Protein. Journal of Andrology, 25(4), pp.479–493. doi:https://doi.org/10.1002/j.1939-4640.2004.tb02819.x.

Shinkai-Ouchi, F., Koyama, S., Ono, Y., Hata, S., Ojima, K., Shindo, M., duVerle, D., Ueno, M., Kitamura, F., Doi, N., Takigawa, I., Mamitsuka, H. and Sorimachi, H. (2016). Predictions of Cleavability of Calpain Proteolysis by Quantitative Structure-Activity Relationship Analysis Using Newly Determined Cleavage Sites and Catalytic Efficiencies of an Oligopeptide Array. Molecular & Cellular Proteomics, 15(4), pp.1262–1280. doi:https://doi.org/10.1074/mcp.m115.053413.

Shiva, S., Huang, Z., Grubina, R., Sun, J., Ringwood, L.A., MacArthur, P.H., Xu, X., Murphy, E., Darley-Usmar, V.M. and Gladwin, M.T. (2007). Deoxymyoglobin Is a Nitrite Reductase That Generates Nitric Oxide and Regulates Mitochondrial Respiration. Circulation Research, 100(5), pp.654–661. doi:https://doi.org/10.1161/01.res.0000260171.52224.6b.

Shiva, S., Huang, Z., MacArthur, P.H., Ringwood, L.A. and Gladwin, M.T. (2006). Myoglobin Is a Nitrite Reductase That Generates NO and Regulates Mitochondrial Respiration. Blood, 108(11), pp.1561–1561. doi:https://doi.org/10.1182/blood.v108.11.1561.1561.

Simons, M., Gretton, S., Silkstone, Gary G.A., Rajagopal, Badri S., Allen-Baume, V., Syrett, N., Shaik, T., Leiva-Eriksson, N., Ronda, L., Mozzarelli, A., Strader, Michael B., Alayash, Abdu I., Reeder, Brandon J. and Cooper, Chris E. (2018). Comparison of the oxidative reactivity of recombinant fetal and adult human hemoglobin: implications for the design of hemoglobin-based oxygen carriers. Bioscience Reports, [online] 38(4). doi:https://doi.org/10.1042/BSR20180370.

Singh, S., Thakur, N., Oliveira, A., Petruk, A.A., Hade, M.D., Sethi, D., Axel Bidon-Chanal, Martí, M.A., Datta, H., Raman Parkesh, Estrin, D.A., Luque, F.Javier. and Dikshit, K.L. (2014). Mechanistic Insight into the Enzymatic Reduction of Truncated Hemoglobin N of Mycobacterium tuberculosis. Journal of Biological Chemistry, 289(31), pp.21573–21583. doi:https://doi.org/10.1074/jbc.m114.578187.

Smith, L. and Stotz, E. (1954). PURIFICATION OF CYTOCHROME c OXIDASE. Journal of Biological Chemistry, 209(2), pp.819–828. doi:https://doi.org/10.1016/s0021-9258(18)65511-5.

Sorimachi, H., Mamitsuka, H. and Ono, Y. (2012). Understanding the substrate specificity of conventional calpains. Biological Chemistry, 393(9), pp.853–871. doi:https://doi.org/10.1515/hsz-2012-0143.

Sotomayor-Vivas, C., Hernández-Lemus, E., &Dorantes-Gilardi, R. (2022). Linking protein structural and functional changeto mutation using amino acid networks. PloS one, 17(1), e0261829. https://doi.org/10.1371/journal.pone.0261829

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 humanm-calpain suggests an electrostatic switch mechanism for activation bycalcium. Proceedings of the National Academy of Sciences, 97(2),pp.588-592.

Sun, M., Wang, Y., Zhang, Q., Xia, Y., Ge, W. and Guo, D. (2017). Prediction of reversible disulfide based on features from local structural signatures. BMC Genomics, 18(1). doi:https://doi.org/10.1186/s12864-017-3668-8.

Sung Oog Kim, Yutaka Orii, Lloyd, D., Hughes, M.N. and Poole, R.K. (1999). Anoxic function for the Escherichia coli flavohaemoglobin (Hmp): reversible binding of nitric oxide and reduction to nitrous oxide. FEBS letters, 445(2-3), pp.389–394. doi:https://doi.org/10.1016/s0014-5793(99)00157-x.

Svistunenko, D.A. (2021). EPR spectroscopy of whole blood and blood components: can we diagnose abnormalities? The Journal of Biomedical Research, 35(4), p.294. doi:https://doi.org/10.7555/jbr.35.20210011.

Svistunenko, D.A., Reeder, B.J., Wankasi, M.M., Silaghi-Dumitrescu, R.-L., Cooper, C.E., Rinaldo, S., Cutruzzol?, F. and Wilson, M.T. (2007). Reaction of Aplysia limacina metmyoglobin with hydrogen peroxide. Dalton Transactions, (8), p.840. doi:https://doi.org/10.1039/b615770j.

Svistunenko, D.A., Sharpe, M.A., Nicholls, P., Wilson, M.T. and Cooper, C.E. (2000). A New Method for Quantitation of Spin Concentration by EPR Spectroscopy: Application to Methemoglobin and Metmyoglobin. Journal of Magnetic Resonance, 142(2), pp.266–275. doi:https://doi.org/10.1006/jmre.1999.1935.

Tamás Jávorfi, János Erostyák, József Gál, Buzády, A., László Menczel, Győző Garab and Naqvi, K.R. (2005). Quantitative spectrophotometry using integrating cavities. Journal of Photochemistry and Photobiology B Biology, 82(2), pp.127–131. doi:https://doi.org/10.1016/j.jphotobiol.2005.10.002.

The, M.-J. (1989). Human insulin: DNA technology's first drug. American Journal of Health-System Pharmacy, 46(11_suppl), pp.S9–S11. doi:https://doi.org/10.1093/ajhp/46.11_suppl_2.s9.

Thermofisher (n.d.). Tm Calculator - US. [online] www.thermofisher.com. Available at: https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html.

Thijs, L., Vinck, E., Bolli, A., Trandafir, F., Wan, X., Hoogewijs, D., Coletta, M., Fago, A., Weber, R.E., Van Doorslaer, S., Ascenzi, P., Alam, M., Moens, L. and Dewilde, S. (2007). Characterization of a Globin-coupled Oxygen Sensor with a Gene-regulating Function. Journal of Biological Chemistry, 282(52), pp.37325–37340. doi:https://doi.org/10.1074/jbc.m705541200.

Tiso, M., Tejero, J., Basu, S., Azarov, I., Wang, X., Simplaceanu, V., Frizzell, S., Jayaraman, T., Geary, L., Shapiro, C., Ho, C., Shiva, S., Kim-Shapiro, D.B. and Gladwin, M.T. (2011). Human Neuroglobin Functions as a Redox-regulated Nitrite Reductase. The Journal of Biological Chemistry, [online] 286(20), pp.18277–18289. doi:https://doi.org/10.1074/jbc.M110.159541.

Ukeri, J.U. (2020). Studies of the effect of cysteine and heme pocket mutations of the nitric oxide dioxygenase and nitrite reductase activities of human Cytoglobin and Androglobin heme domain. [online] pp.36–37. Available at: https://repository.essex.ac.uk/28516/1/UkeriJohnThesis.pdf [Accessed 30 Apr. 2026].

Uno, T., Ryu, D., Tsutsumi, H., Yoshikazu Tomisugi, Ishikawa, Y., Wilkinson, A.J., Sato, H. and Hayashi, T. (2004). Residues in the Distal Heme Pocket of Neuroglobin. Journal of Biological Chemistry, 279(7), pp.5886–5893. doi:https://doi.org/10.1074/jbc.m311748200.

Uzan, J., Dewilde, S., Burmester, T., Hankeln, T., Moens, L., Hamdane, D., Marden, M.C. and Kiger, L. (2004). Neuroglobin and Other Hexacoordinated Hemoglobins Show a Weak

Temperature Dependence of Oxygen Binding. Biophysical Journal, 87(2), pp.1196–1204. doi:https://doi.org/10.1529/biophysj.104.042168.

Vallone, B., Nienhaus, K., Brunori, M. and Nienhaus, G.U. (2004). The structure of murine neuroglobin: Novel pathways for ligand migration and binding. Proteins: Structure, Function, and Bioinformatics, 56(1), pp.85–92. doi:https://doi.org/10.1002/prot.20113.

Vickery, L.E., Nozawa, T. and Sauer, K. (1976). Magnetic circular dichroism studies of myoglobin complexes. Correlations with heme spin state and axial ligation. Journal of the American Chemical Society, 98(2), pp.343–350. doi:https://doi.org/10.1021/ja00418a005.

Villalobo, A., María Jesús González-Muñoz and Berchtold, M.W. (2019). Proteins with calmodulin-like domains: structures and functional roles. Cellular and Molecular Life Sciences, 76(12), pp.2299–2328. doi:https://doi.org/10.1007/s00018-019-03062-z.

Villar, I., Rubio, M.C., Calvo-Begueria, L., Pérez-Rontomé, C., Estibaliz Larrainzar, Wilson, M.T., Sandal, N., Mur, L.A., Wang, L., Reeder, B., Deqiang Duanmu, Toshiki Uchiumi, Jens Stougaard and Becana, M. (2021). Three classes of hemoglobins are required for optimal vegetative and reproductive growth of Lotus japonicus: genetic and biochemical characterization of LjGlb2-1. Journal of Experimental Botany, 72(22), pp.7778–7791. doi:https://doi.org/10.1093/jxb/erab376.

Vinck, E., Doorslaer, S.V., Dewilde, S. and Moens, L. (2004). Structural Change of the Heme Pocket Due to Disulfide Bridge Formation Is Significantly Larger for Neuroglobin than for Cytoglobin. Journal of the American Chemical Society, 126(14), pp.4516–4517. doi:https://doi.org/10.1021/ja0383322.

Vuletich, D.A., Lecomte, J.T. A Phylogeneticand Structural Analysis of Truncated Hemoglobins. J Mol Evol 62,196–210 (2006). https://doi.org/10.1007/s00239-005-0077-4

Wajcman, H., Kiger, L. and Marden, M.C. (2009). Structure and function evolution in the superfamily of globins. Comptes Rendus Biologies, 332(2-3), pp.273–282. doi:https://doi.org/10.1016/j.crvi.2008.07.026.

Wang, N., Chen, W., Linsel-Nitschke, P., Martinez, L.O., Agerholm-Larsen, B., Silver, D.L. and Tall, A.R. (2003). A PEST sequence in ABCA1 regulates degradation by calpain protease and stabilization of ABCA1 by apoA-I. Journal of Clinical Investigation, 111(1), pp.99–107. doi:https://doi.org/10.1172/jci200316808.

Wang, Y.-C., Peterson, S.E. and Loring, J.F. (2013). Protein post-translational modifications and regulation of pluripotency in human stem cells. Cell Research, [online] 24(2), pp.143–160. doi:https://doi.org/10.1038/cr.2013.151.

Wilson, M.T. and Reeder, B.J. (2022). The peroxidatic activities of Myoglobin and Hemoglobin, their pathological consequences and possible medical interventions. Molecular aspects of medicine, [online] 84, pp.101045–101045. doi:https://doi.org/10.1016/j.mam.2021.101045.

Wosnitzer, M., Goldstein, M. and Hardy, M.P. (2014). Review of Azoospermia. Spermatogenesis, 4(1), p.e28218. doi:https://doi.org/10.4161/spmg.28218.

Wu, D., Li, D., Liu, Z., Liu, X., Zhou, S. and Duan, H. (2018). Role and underlying mechanism of SPATA12 in oxidative damage. Oncology Letters. doi:https://doi.org/10.3892/ol.2018.7749.

Yang, J., Kloek, A.P., Goldberg, D.E. and Mathews, F.S. (1995). The structure of Ascaris hemoglobin domain I at 2.2 A resolution: molecular features of oxygen avidity. Proceedings of

the National Academy of Sciences, 92(10), pp.4224–4228. doi:https://doi.org/10.1073/pnas.92.10.4224.

Zhang, F., Skoda, M.W.A., Jacobs, R.M.J., Martin, R.A., Martin, C.M. and Schreiber, F. (2006). Protein Interactions Studied by SAXS: Effect of Ionic Strength and Protein Concentration for BSA in Aqueous Solutions. The Journal of Physical Chemistry B, 111(1), pp.251–259. doi:https://doi.org/10.1021/jp0649955.

Zhi, F., Wang, Q., Li, X., Shao, N., Wang, R., Deng, D., Wang, S., Xia, X. and Yang, Y. (2015). The Use of Three Long Non-Coding RNAs as Potential Prognostic Indicators of Astrocytoma. PLOS ONE, 10(8), pp.e0135242–e0135242. doi:https://doi.org/10.1371/journal.pone.0135242.

Å. Wennmalm, Benthin, G. and Petersson, A. -S (1992). Dependence of the metabolism of nitric oxide (NO) in healthy human whole blood on the oxygenation of its red cell haemoglobin. British Journal of Pharmacology, 106(3), pp.507–508. doi:https://doi.org/10.1111/j.1476-5381.1992.tb14365.x.