School of Life Sciences

University of Essex

Date of Submission: January 2021

A Thesis Submitted for the Degree of Master of Science (by Dissertation) in Molecular Medicine

Exploring the Molecular Mechanisms of Human Cytoglobin and its Role in Cancer

Name: Georgia Louise Scobioala

Supervisors: Dr Brandon Reeder and Dr Greg Brooke

Final Word Count: 29,479

Statement of Originality:

I clarify that this Master's thesis is the result of my own work and solely my own work. All sources used have been referenced and acknowledged.

Statement of Impact of COVID-19 on the Project:

During the course of the project my work was significantly affected by the global pandemic caused by the COVID-19 virus; primarily during the months of March to mid-August where I had no access to the laboratories and then for around six weeks after this where I only had part-time access. Since the masters qualification was a wet-lab research project, I was very limited on collecting data during this time.

Abstract

Cytoglobin (Cygb), a recently discovered globin, is found in low levels in most human tissues. It currently has undefined functions, but theories include scavenging free radicals and redox regulation. It is implicated in cancer and evidence suggests the protein has bimodal functions, acting as a tumour suppressor under normoxic conditions and an oncogene under hypoxic conditions. Additionally, Cygb is proposed to be involved in cancer therapy resistance by protecting tumour cells from harsh environments including high free radical levels generated with chemotherapeutics. The effects of different conditions and interactions were tested on the protein to better understand the protein's reaction to environments including NiR (nitrite reductase), lipid and liposome interactions and NOD (nitric oxide dioxygenase). The WT protein was compared to results gathered from two mutants: H81A and L46W. H81 mutation was selected as it affects NiR activity in neuroglobin and L46 affects NOD activity in myoglobin (Mb) and haemoglobin (Hb). In addition to tests on the three proteins, investigations were undertaken on CYGB (WT and mutant) transfected MCF7 and HEK293S cell lines whereby reactions to stimuli such as H_2O_2 and chemotherapeutics were observed. Cell proliferation and survival assays including crystal violet were used to assess the extent of the effect.

Results indicated significant levels of protection afforded to cells transfected with CYGB when exposed to H_2O_2 whereby these cells had much higher EC_{50} than the control cells. When exposed to chemotherapeutics, L46W appeared to reduce the cytoprotection. The oxidation rate of liposomes was significantly improved in the H81A mutant and slightly decreased in L46W where it mirrored the absorbance patterns of Mb. Lipid binding did not occur in L46W. Further investigation of the in-cell work conducted in this study, replicated in stable-transfected cell lines, would improve understanding of the role of cytoglobin in protecting cells from oxidative stress.

Acknowledgments

Firstly, I would like to say thank you to my two supervisors: Dr Greg Brooke and Dr Brandon Reeder for allowing me the opportunity to do this Masters. I am beyond grateful that you selected me for this studentship as financially supporting myself during this year would have been tremendously difficult. I thank you both for the support you have given me and for making me the independent scientist I now feel I am, prepared for the working world.

I am also incredibly grateful to the family of Christine Desty for creating these studentships in her honour, it has allowed me to pursue my research interests without financial restraint and meant the world to me.

I would like to thank Dr Elizabeth Welbourn and Amanda Clements for their laboratory assistance and kind words during some incredibly challenging times this year, your support and compassion was incredibly appreciated.

Next, I want to thank my family, in particular my parents, for their love and support throughout a difficult year. You have both been the absolute constants in my life and I am so grateful to have you.

In addition, I would like to thank my friends: Alex, Callum, Shiv, Kirsty, Amy, Kara and several others for your love and motivation.

I would also like to thank Shakira for keeping me motivated during the write up of this project.

Lastly, I would like to make my biggest thank you to my husband, best friend and teammate. This year would not have been possible without your support. From driving me to and from university after my accident, being by my side through every hospital appointment, supporting me through the lockdown. You have always been my biggest supporter and I am so grateful and lucky to have you. You really are one in a million.

Abbreviations

- 5-FU- Fluorouracil
- ALA- Alpha Linolenic Acid
- ATG5- Autophagy Related 5
- BCL-2- B-Cell Lymphoma 2
- BL21 D3- *E. coli* cell line
- BRCA- Breast Cancer Susceptibility Gene
- CAP- Cold Atmospheric Plasma
- CI- Confidence Interval
- CO- Carbon Monoxide
- CO_2- Carbon Dioxide
- CpG- 5'-C-phosphate-G-3'
- CSC- Cancer Stem Cells
- CYGB- Cytoglobin Gene
- Cygb- Cytoglobin Protein
- Da- Dalton
- DMEM- Dulbecco's Modified Eagle's Medium Mixture
- DNA- Deoxyribonucleic Acid
- DU145- Human Prostate Cancer Cell Line
- EC_{50}- Concentration Of Drug That Gives ½ Maximal Response
- EDTA- Ethylenediaminetetraacetic Acid
- EMT- Epithelial Mesenchymal Transition
- ER- Endoplasmic Reticulum
- FASTA- Text-Based Format For Representing Nucleotide Sequences of Amino Acid Sequences
- FBS- Foetal Bovine Serum
- FE- Iron

- H_2O- Water
- H_2O_2- Hydrogen Peroxide
- H2DCF- Redox-sensitive dye
- H64L- Missense Mutation At Point 81 Changing A Histidine To An Leucine
- H81A- Missense Mutation At Point 81 Changing A Histidine To An Alanine
- HB- Haemoglobin Gene
- Hb- Haemoglobin Protein
- HBS- Sickle Cell Haemoglobin
- HEK293S- Foetal Embryonic Kidney Cells
- hCPC- Human Cardiac Progenitor Cell
- HIF- Hypoxia Inducing Factor
- iNOS- Inducible Nitric Oxide Synthase
- IPTG- Isopropyl beta-d-1-thiogalactopyranoside
- Kb- Kilobyte
- Kd- Binding Constant
- L46W- Missense mutation at point 46 changing a leucine to a tryptophan
- LB- Luria Broth
- MB- Myoglobin Gene
- Mb- Myoglobin Protein
- MCF7- Human Breast Cancer Cell Line
- NaCl- Sodium Chloride
- NaPi- Sodium Phosphate Buffer
- NO- Nitric Oxide
- NOD- Nitric Oxide Dioxygenase
- NE- Neuroendrocrine Differentiated Cells
- NGB- Neuroglobin Gene
- Ngb- Neuroglobin Protein

- NiR- Nitrite Reductase
- ORF- Open Reading Frame
- PBS- Phosphate-Buffered Saline
- pCMV- Protein Expression Vector
- PET- Protein Expression Vector
- PFA- Paraformaldehyde
- PSA- Prostate Specific Antigen
- PT- Total Protein Concentration
- RAS- Family of genes that are essential for cellular function; word Ras comes from a contraction of Rat Sarcoma
- Rb- Retinoblastoma
- RCF- Relative Centrifugal Force
- RNA- Ribonucleic Acid
- RO- Reverse Osmosis
- ROS- Reactive Oxygen Species
- RSCB- Romanian Society for Cell Biology
- S- Substrate Concentration
- SCS- Spinal Cord Stimulation
- SOD- Superoxide Dismutase
- TAE- Tris-Acetate-Ethylenediaminetetraacetic
- Tbf- Buffer used to grow up competent cells
- VEGF- Vascular Endothelial Growth Factor
- WT- Wild-type

Table of Contents

List of figures

List of tables

This project is based around the potential functions of globin proteins, in particular Cygb. What follows in this report is an examination of the current research on the relationship between Cygb and cancer, particularly the potential involvement with chemoresistance and reactive oxygen species (1.1-1.1.6); globin function (1.2-1.2.6.2.6) and how these two are linked (1.2.7 and 1.2.8).

Chapter 1: Introduction

1.1 Introduction to cancer

1.1.1 History of cancer

The oldest known written record of cancer was on Egyptian papyrus and describes a breast cancer tumour in a human patient in around 3000 BC. Egyptian mummies have also been recovered containing identifiable tumours, some of which had extensively spread across the body. This ancient civilisation also had a base understanding of metastasis with records documenting conditions in which the cancer was unable to be treated once it had spread over the breast (Faguet, 2014). In addition to this, tumours have been found in well-preserved remains of many extinct animals which pre-dated the existence of modern humans. From this point, our understanding of cancer increased significantly with the beginning of human dissections and then the observation that different types of tumours exist within the body. Despite our increased understanding, the number of cancer diagnoses are still high though thankfully our ability to identify and treat cancers mean that the survival rates have greatly improved.

1.1.2 Hallmarks of cancer

Cancer is broadly defined by the six ‘hallmarks of cancer’ that were theorised by Hanahan and Weinberg in 2000; these are biological capabilities that cells acquire as they transition from healthy cells to cancerous cells. The hallmarks are limitless replicative potential, avoidance of cell death and apoptosis, sustained angiogenesis, self-sufficiency in growth signals, metastasis and an insensitivity to anti-growth signals (Hanahan and Weinberg, 2000).

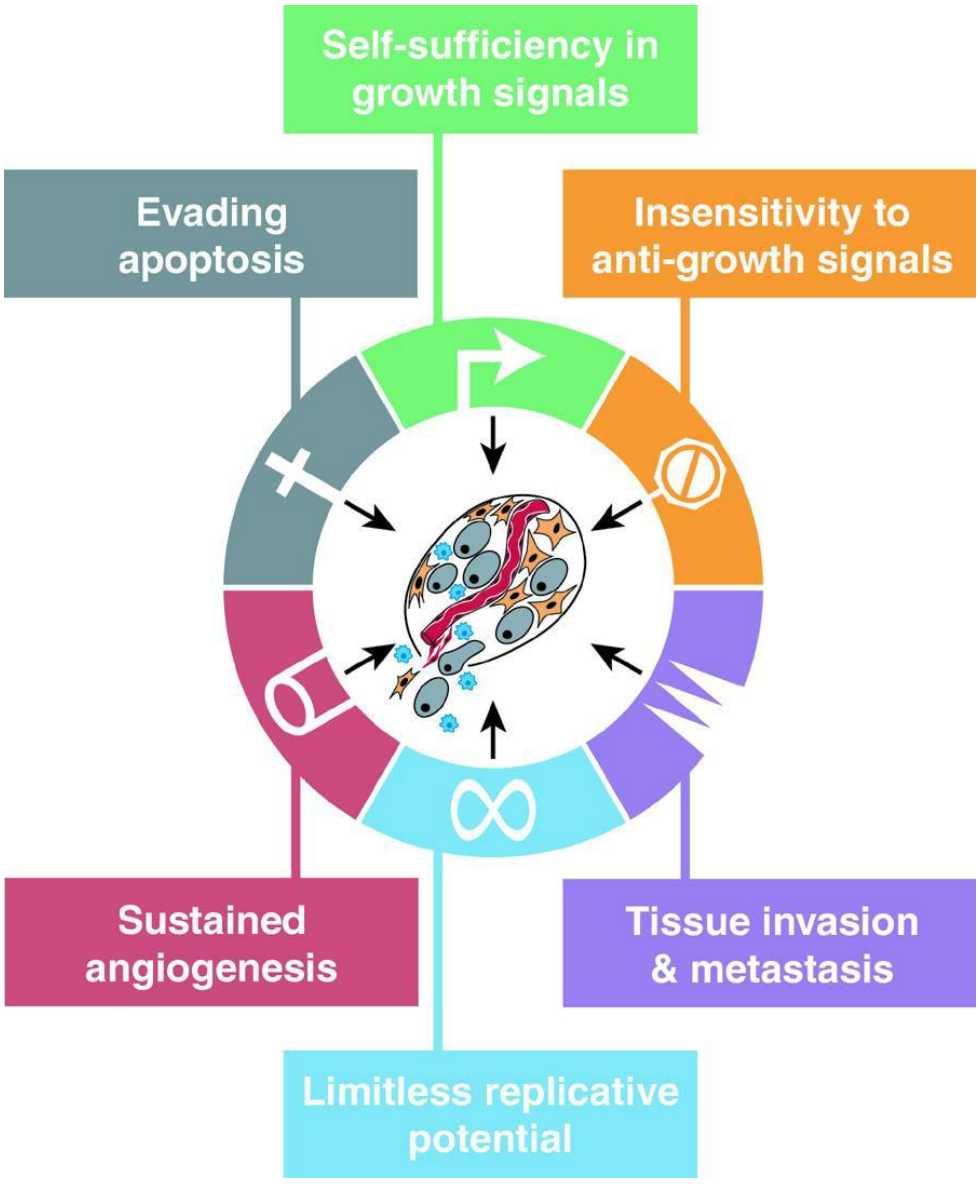


Figure 1.1.2.1- Acquired capabilities of cancer cells as they transition from healthy cells. Schematic showing the 6 hallmarks of cancer believed to collectively dictate malignant growth in almost all tumour types (Hanahan and Weinberg, 2000).

The first of these, limitless replicative potential, occurs when cells develop the ability to avoid the constraints of the cell cycle's many control and check points and divide at an accelerated and uncontrolled rate. It is widely accepted that telomerase, a DNA polymerase responsible for adding telomeric repeats to telomeres, aids this process. As healthy cells replicate, the telomeres shorten with each passage and before they can shorten so much that they degrade the chromosome, senescence is generally induced, after which the cell population enters a crisis phase, and the majority of cells die. Any that survive this become 'immortal' and divide uncontrollably. Cancer cells do not enter a crisis phase and senescence is not observed (Hanahan and Weinberg, 2011). Within cancer cells, telomerase is expressed at much more significant levels compared to normal cells and this continuously adds repeats to the telomere ends so that they cannot degrade enough to be detected and therefore cancer cells continue to replicate without constraint (Hanahan and Weinberg, 2011).

Another hallmark, avoidance of cell death and apoptosis, is a phenomenon in which, despite the cell's genetic code being mutated, the cells avoid death and continue replicating. Apoptosis is generally defined as regulated cell death triggered in response to developmental cues or cellular stress (Campbell and Tait, 2018). Apoptosis occurs in one of two pathways: the extrinsic pathway, activated by ligands that interact with death receptors situated on the cells surface, or intrinsic pathway which occurs via the mitochondria. The intrinsic pathway uses a family of oncoproteins called BCL-2 (B-cell lymphoma 2) to regulate its activation in response to factors such as oncogene activation or DNA damage (Campbell and Tait, 2018). The family contains members that are either anti-apoptotic or pro-apoptotic and these work in conjunction to regulate the levels of mitochondria-mediated apoptosis (Warren *et al.,* 2019). The ability of cancer cells to avoid apoptosis is a desirable one and therefore it is logical that these cells have adapted to increase the expression of the anti-apoptotic proteins; this has been seen in a variety of different cancers such as follicular lymphoma. The upregulation is not a single-step process and can result from a range of different mechanisms such as gene amplification, chromosomal translocation and even increased gene expression. Increased gene copy

numbers are one of the most common mechanisms and an increase in BCL2, MCL1 and Bclx are all predispositions to lymphoma progression (Campbell and Tait, 2018).

Angiogenesis is an essential characteristic that cancer cells require in order to provide enough nutrients and oxygen to the tumour to support excessive cell growth. Throughout angiogenesis, angiogenic factors such as VEGF (vascular endothelial growth factor) are overexpressed while factors which negatively inhibit the growth of blood vessels are down regulated. The process of angiogenesis is a four-step process (Nishida *et al.,* 2006); firstly, the basement membrane in tissues obtains a small local injury causing hypoxic damage. Angiogenic factors then activate the vascular endothelial cells, causing them to migrate. These then proliferate and produce a stable colony which forms blood vessels. The tumour uses these blood vessels to grow extensively, forming a primary tumour. After some time, there becomes limited room for cell proliferation and expansion, so cancer cells develop the ability to 'metastasise' or fracture part of the tumour off and recolonise it in another location. They are able to degrade the extracellular matrix of the blood vessels surrounding them and through a process known as intravasation, enter the blood stream and spread or metastasise to new locations. As shown in Figure 1.1.2.2, this new tumour can leave the blood stream via extravasation and colonise a new area, forming a metastatic tumour which is much harder to treat (Peng *et al.,* 2019). Recent work has shown a possible link between levels of angiogenic factors within the tissues and cancer severity - it was observed that the higher the levels of VEGF, the more aggressive the spread of the tumour. (Nishida *et al.,* 2006). It also correlated to poor prognosis due to a reduced accessibility to cancer treatment drugs such as chemotherapy.

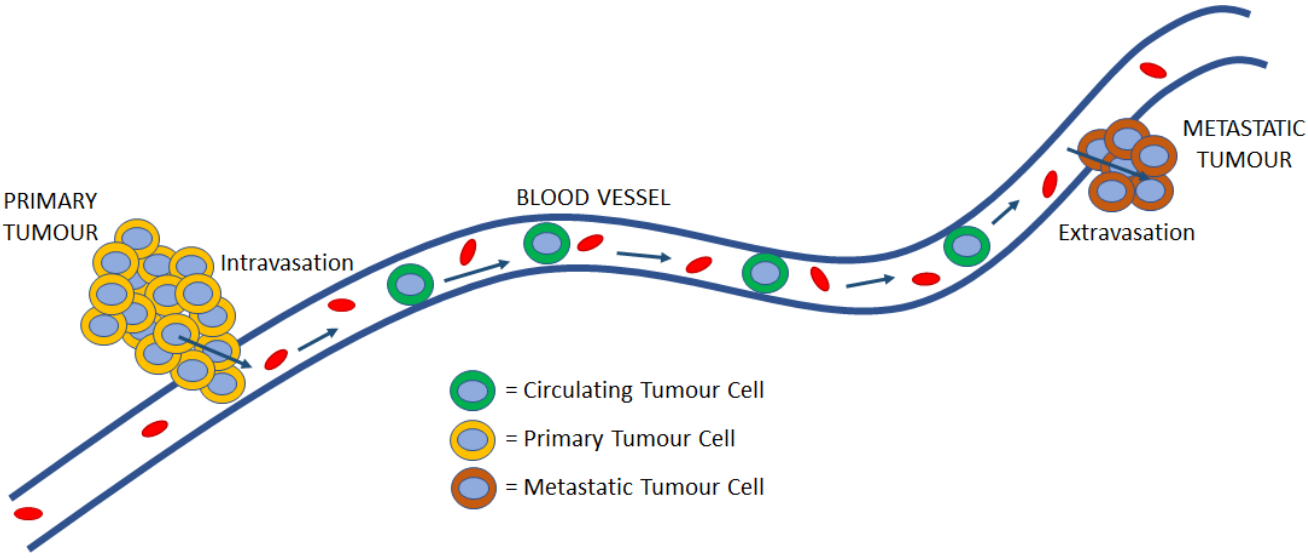


Figure 1.1.2.2- The process by which cancer cells metastasise and colonise new sites. The development of primary tumour cells to circulating tumour cells and finally to metastatic tumour cells.

Self-sufficiency in growth signals are facilitated by oncogenes such as Ras, allowing cells to grow beyond their provisional microenvironments (Giancotti, 2014). Oncogenes including Ras begin as proto-oncogenes which are vital for the cell and are involved in cell growth and regulation of the cell cycle; they become oncogenes once a mutation is present. Self-sufficiency in growth signals is a dual mechanism which partly relies upon the cells producing growth factor ligands of their own but also in the stimulation of healthy cells to provide growth factors for the cancer cells. Normal cells use growth signals such as receptor tyrosine kinases to change from a quiescent state to one which is actively proliferating (Giancotti, 2014). The tyrosine residues on the receptor tyrosine kinases (RTK) are phosphorylated and along with Shc and Hrb2, activate Ras. Oncogenic mutations such as those in *HER2* cause the activation of receptor tyrosine kinase cascades to be deregulated. Additional growth factors come from sources such as neuroendocrine differentiated cells (NE) which are found in many tissues however the populations expand rapidly from malignant precursor cells. This creates peptides such as neurotensin and bombesin related peptides that aid in cancer cells no longer relying solely on the regular channels to grow and proliferate. In addition to increasing levels of mitogens and growth factors, cancer cells develop a heightened sensitivity towards them. This occurs through the production of an intrinsic potential for uncontrolled proliferation; the altered expression of the ion channels and membrane receptors that receive external signals (Prevarskaya *et al.,* 2010).

Metastasis is another key hallmark of cancer. Metastasis is the spread of cancerous cells from the primary tumour site to surrounding tissues (as previously explained above in the angiogenesis section, and displayed in Figure 1.1.2.2) or further, to other organs in the body (Seyfried and Huysentruyt, 2014). Cancer cells have developed an intricate system that enables the cells to detach from the initial tumour and travel throughout the blood stream; this is known as the metastatic cascade. Intravasation is the process of the cancer cells entering the bloodstream and for this to occur, the endothelial barrier must be breached. Once this occurs, neoplastic cells are released into the bloodstream and distributed around the body. In

the bloodstream, many of the circulating tumour cells are killed by the immune system, die as a result of cellular stresses or become stuck in capillary beds. Once it reaches its destination or becomes lodged elsewhere, it can undergo extravasation and form a secondary tumour. The process of extravasation requires the cells to exit the blood vessel lumens and enter the surrounding stromal tissue where it can replicate and grow (Chiang *et al.,* 2016). Once the tumours metastasise, the severity of the cancer generally increases as it makes the disease harder to treat.

The last hallmark of cancer, insensitivity to anti-growth signals, allows the tumour cells to replicate without limit and avoid factors such as senescence, apoptosis and growth arrest. These usually occur through the actions of tumour suppressors such as p53 which, in healthy cells, ensure that only healthy cells pass through the cycle, preventing the formation of tumours. This occurs through elimination of damaged cells or those with the potential to become cancerous. The tumour suppressors become damaged through environmental stress, chemicals or radiation causing genetic damage or mutations. This prevents damaged cells from being repaired and these mutated cells are able to proliferate intensively and without control (Amin *et al.,* 2015). An important example is the retinoblastoma (Rb) gene which was the first tumour suppressor gene to be described. The retinoblastoma protein is important in integrating signals from intracellular and extracellular sources, however in cancers, there are frequent mutations seen in p16 which regulates retinoblastoma. By inhibiting the cyclin-dependent kinases that phosphorylate retinoblastoma and therefore allowing E2F to express genes involved in the beginning of the cell cycle, it is no longer affected by the anti-growth signals produced by Rb.

1.1.3 Cancer and current statistics

Cancer is currently the second highest cause of death around the world, accounting for approximately 1 in every 6 deaths (World Health Organisation, 2018). In 2019, it is forecast that 1,762,450 people will be diagnosed with cancer in the United States alone, which is approximately 4800 new cases every day (Siegel *et al.,* 2019). The probability during a

person's lifetime of being diagnosed with cancer is slightly elevated in men compared to women but by less than 2% and these probabilities are affected by age. Increased age is a known factor in the likelihood of being diagnosed with an invasive cancer (Table 1). In both males and females, the probability of being diagnosed with cancer increases with age, with the largest probabilities seen in the over 70 years category.

Table 1.1.3.1- Probabilities of obtaining cancer within a lifetime, determined by age. The higher the age, the greater the chance of getting cancer with males having an overall higher chance (Siegel *et al.,* 2019).

Gender	**Age (years)**				
	BIRTH- 49	**50-59**	**60-69**	**≥70**	**BIRTH-DEATH**
Male	3.4	6.1	13.2	31.9	39.3
female	5.6	6.2	10.0	26.0	37.7

Though 1,762,450 people are predicted to be diagnosed with cancer within the United States during 2019, of these, 606,880 are calculated to die from this disease- approximately 1700 deaths per day (Siegel *et al.,* 2019). The cancers making up the majority of these deaths are breast, lung, colorectal and prostate. Of the 891,480 estimated new cases in females in 2019, 268,600 of those are predicted to be breast cancer, accounting for 30% of cancer diagnoses within women and making this site the leading cancer among women. Similarly, of the 870,970 cases predicted in the same year within males, 174,650 are estimated to be diagnosed with prostate cancer (almost 1 in 5 new diagnoses), meaning these two types have some of the highest incidence rates among cancer (Siegel *et al.,* 2019).

1.1.4 Prostate cancer and breast cancer

1.1.4.1 Prostate cancer

The prostate is a gland within the lower pelvis around the urethral origin whose primary function is part of the male reproductive system. It secretes a fluid high in pH which forms part of the semen ejaculated from the body that provides the necessary nutrients for the sperm and allows easier movement. The prostate itself is comprised of four ‘zones’ as shown in Figure 1.1.4.1.1 which are as follows: the transition, peripheral, fibromuscular and central zones. The most common place for prostate cancer to develop is peripheral zone where 75% originate. This and more than 95% of prostate cancers are known as adenocarcinomas due to their glandular cell origins (Dunn and Kazer, 2011).

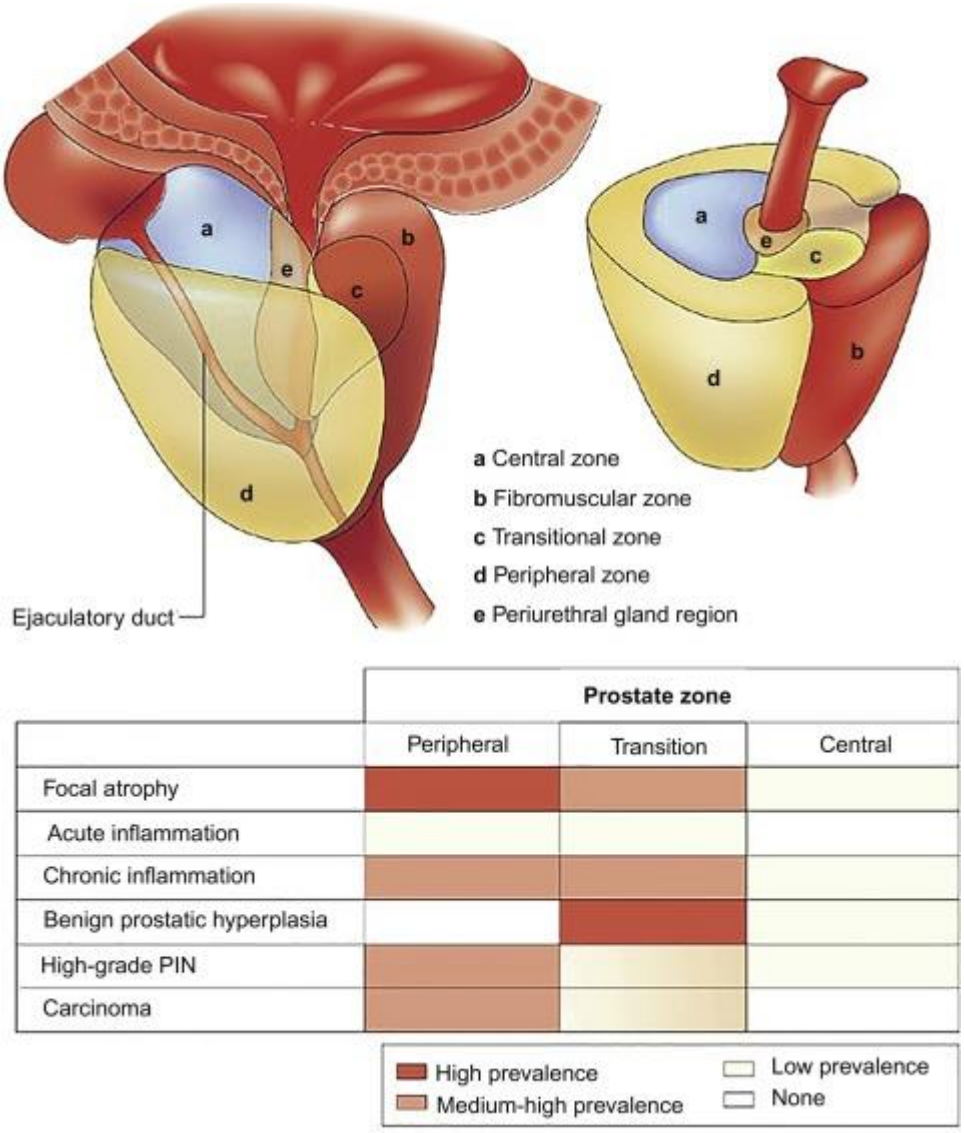


Figure 1.1.4.1.1- The zones of the prostate showing that the peripheral zone is the most common area for cancer. The diagram shows the five different areas of the prostate and examines the likelihood of different diseases and conditions for the three main areas: peripheral, transition and central (Hedayat and Lapraz, 2019).

Prostate cancer can be difficult to diagnose as many patients can be asymptomatic for a long time or present symptoms which are alike to many other conditions. These symptoms include but are not limited to increased urinary frequency, increase in prostate size, difficulty in urinating and erectile dysfunction which can be similar to issues such as a urinary tract infection or impotence. When these symptoms are present, a physical prostate exam will be carried out to determine whether the prostate is enlarged, a urine sample will be taken, and blood will be collected to test for a prostate specific antigen (PSA) level in the serum. The size of a healthy prostate can vary between people but is usually between 29-47 cm^3. PSA is a protease produced in the prostate that provides a good early diagnosis marker for prostate cancer. A normal PSA range is between 1-4 ng/ml with those higher than 4 requiring further investigation; though this test is not perfect, and several factors can increase this value other than the presence of prostate cancer such as infections, recent ejaculation and certain traumas. Further investigation will likely include a biopsy of the prostate tissue to determine whether or not cancerous cells are present and if so, how aggressive the cancer is which is measured against the Gleason scoring system (Hogle, 2009).

As mentioned above, increased age is a significant factor in the incidence rate of prostate cancer with three quarters of the diagnoses seen in men over the age of 65 and a fraction seen in those below the age of 40. Though this is not the only factor; race is another prominent factor- those from an African-american background are at the highest risk, with a 60% increase in the number of cases from this ethnicity compared to Caucasian men within 2009. As well as this, family history is also a notable factor with a much greater increase for men with a close male relative such as a father or brother with the disease; this risk is even greater if the affected relative was diagnosed before the age of 60.

Treatment can include radiotherapy, hormone therapy, radical prostatectomy and chemotherapy with a range of different side effects present for each treatment. Active surveillance is often preferred since prostate cancer is often very slow developing and the

risks of surgery on patients such as those who are elderly, can outweigh the risks of the cancer progression (Haas and Sakr, 1997).

1.1.4.2 Breast cancer

Breasts are layers of tissue which lay over the pectoral muscles of the chest, they are composed of a combination of fatty tissues and (in women) a glandular tissue which is responsible for the production of milk (Pandya and Moore, 2011). The tissues are connected to a vast network of arteries which supply it with the nutrients it requires, these are the axillary, intercostal and internal thoracic arteries as well as many associated veins. Internal thoracic and axillary lymph nodes are key features of the breast involved with lymphatic drainage; approximately three quarters of all lymphatic drainage goes to the axillary nodes (Ellis and Mahadevan, 2013).

Breast cancer is one of the most common forms of cancer in women and contributes to a significant number of cancer related deaths, with around one in every eight women in 2010 developing it during their lifetime (George, 2010). It is not only women who suffer from breast cancer, men are also susceptible. Several risk factors, like those seen in prostate cancer, increase the chances of developing this cancer such as smoking. The number of cigarettes smoked, the duration of time and when the smoking began can influence this, with smoking beginning in the teenage years increasing risk most significantly (Passarelli *et al.,* 2016). In addition to smoking, drinking alcohol also increases the risk of getting breast cancer, particularly in menopausal women by 7.1% for each 4 ounces of wine drunk per day (Chen *et al.,* 2012*)*. There are known heritable genetic mutations: BRCA1 or BRCA2 which put some women at a higher risk of getting breast cancer than others which can be tested if there is a strong family history of the condition. Hormonal exposure for long periods of time such as hormone replacement therapy also increases the risk of developing the disease though physical activity lowered the chances and improved prognosis further down the line (Birrer *et al.,* 2018).

There are several different types of breast cancer, but the two main types are classified as invasive or non-invasive. The prognosis for invasive breast cancer largely depends on the type. They are also characterised by the hormone receptors oestrogen and progesterone and whether these are classed as positive or negative. Patients with tumours positive for these receptors currently tend to have a better prognosis and survivability than their negative counterparts. Despite this, triple negative breast cancer has a good prognosis if given the correct treatment, despite its aggressiveness. Much like prostate cancer, the breast cancer aggressiveness is graded on a scale established by Richardson and Bloom and is based on histology. This looks at the size of the tumour, the presence of lymph node infiltration and whether the cancer has metastasised to other areas of the body. Prognosis is often predicted by examining whether the axillary lymph nodes have become affected, of the 20-30, if four or more are diseased, there is a therapy failure rate of up to 70% over a 10-year period. Due to the number of different types of breast cancers, a variety of different therapies are required to treat them. Though surgery is still a good method of removing the tumour cells from the body, radiotherapy and chemotherapy are also widely used as both neoadjuvant and adjuvant treatments to prevent recurrences (Downs-Holmes and Silverman, 2011).

1.1.5 Current cancer treatments

1.1.5.1 Radiation in the treatment of cancer

Radiation or ionising radiation, deposits energy in cells as its passes through tissues and is either of a high enough energy that it kills the cancer cells or it causes genetic damage to the cells, creating mutations. These mutations will prevent the cells from proliferating normally and therefore limits tumour growth (Ahmad *et al.,* 2012). It can either be given to the patient using an external beam of radiation fired from a machine which aims as accurately at the tumour as possible or internally using catheters to take it directly to the tumour site if the tumour is surrounded by delicate structures which are unavoidable (Song *et al.,* 2017). This method, although highly effective at killing cancer cells, can also kill normal healthy cells which are hit by the radiation or absorb it from neighbouring cells. Radiation therapy is more effective on

some cancer types compared to others; for example, it can cure early prostate carcinomas and early lymphomas. Unfortunately for many cancers such as breast carcinomas or more advanced cancers, this treatment alone is not sufficient. Therefore, it is often using in conjunction with other treatment options such as chemotherapy or surgery in order to maximise the survival chances; for instance, it can be used to shrink tumours before surgery to make the surgery less invasive or can be used afterwards to ensure that none of the tumour was left behind (Baskar *et al.,* 2012).

1.1.5.2 Chemotherapy in the treatment of cancer

Chemotherapy is the application of different chemicals or drugs which aim to directly kill cancer cells (Chabner and Roberts, 2005). These anticancer drugs are characterised into several different groups: antibiotics which prevent DNA replication by disrupting the enzymes necessary for this process; antimetabolites that provide competition for RNA and DNA components; alkylating agents that produce damage to the DNA and create mutations; mitotic inhibitors which prevent the cells dividing and replicating and topoisomerase inhibitors that block the enzymes required for DNA to unwind during transcription (Espinosa *et al.,* 2003).

Currently, cancer treatments contain a platinum agent based doublet such as carboplatin or most commonly cisplatin; these are highly effective against solid tumours. They work by forcibly inducing the cells into cell cycle arrest and apoptosis and are effective in a huge variety of different cell types. Unfortunately, patients which take this drug often relapse and the cancers become resistant to chemotherapy treatment (Huang *et al.,* 2017). Another treatment option is taxanes such as paclitaxel, which target the cytoskeleton. Enhanced polymerisation in the microtubules blocks the degradation of the spindle apparatus by binding to a tubulin subunit and the cell cycle is therefore permanently held in the G2 or M phase (Abal *et al.,* 2003).

1.1.5.3 Surgery in the treatment of cancer

The aim of surgery is to remove as much tumour and as many cancer cells from the site as possible, thereby limiting recurrence and metastatic spread. This treatment is best used before the cancer has metastasised and is still localised to one area of the body. In prostate cancer for example, a radical prostatectomy removes the whole prostate and therefore the localised cancer however it generally does little damage while localised, so it is usually recommended to monitor the cancer until it becomes aggressive or metastasises. It is also offered only to men who are below the age of 70 or have an estimated life expectancy greater than 10 years as the surgery has many side effects such as erectile dysfunction, urinary incontinence, and the risk of infection (Lepor, 2000).

1.1.5.4 Hormone therapy in the treatment of cancer

Some cancers are sensitive to, and exacerbated by, hormone levels in the body, these include breast, prostate, endometrial and ovarian cancer. Therefore, this treatment is designed to block or reduce the amount of hormone in the body in order to prevent the growth of the cancer. The most common drug on the market is tamoxifen, used to treat breast cancer that is sensitive to oestrogen; it is a successful treatment for primary breast cancer and is usually taken for five years (Jordan *et al.,* 2003; Cuzick *et al.,* 2010). Hormone therapy is most successful when used in conjunction with other treatments such as reducing the tumour size before surgery and mitigating the chances that the cancer will reoccur after chemotherapy or radiotherapy (Hu *et al.,* 2015).

1.1.6 Cancer therapy resistance via reaction oxygen species

Of the four main treatment methods stated above, all of them have various side effects which must be carefully weighed up when being given to a patient, however one of the largest issues is the ability of tumours to become resistant to the treatment. For chemotherapy this is known as chemoresistance. Some tumours can be resistant to drug treatment before the process

begins while others, despite being initially sensitive to chemotherapy, can acquire resistance throughout the treatment course (Wilson *et al.,* 2006).

The main roles of reactive oxygen species (ROS) are within cellular signalling pathways in that they aid in the regulation of redox homeostasis which is a careful balance between the scavenging and production of ROS. These signal pathways can be induced by various stimuli which in turn generates ROS and the levels are then maintained by antioxidant proteins such as catalase and superoxide dismutase (SOD). Elevated levels of ROS can contribute to oxidative stress and cause damage to DNA, leading to mutations as well as damaging lipids and proteins resulting in cellular damage and apoptosis. Chemotherapy uses this idea of increasing the levels of ROS within the cell to damage the cancer cells and induce ROS-mediated apoptosis (Kim *et al.,* 2019). Recent studies indicate that altered levels of reactive oxygen species, change in redox balance and a deregulation of redox signalling within cells are indicators of cancer development as well as a resistance to drug treatments like chemotherapy (Kumari *et al.,* 2018). To evade the increased levels of ROS, many cancer cells utilise the protective nature of antioxidants by upregulating these in order to protect themselves from chemotherapy. The mechanisms cancer cells use to avoid elevated ROS levels and become chemoresistant are shown in Figure 1.1.6.1.

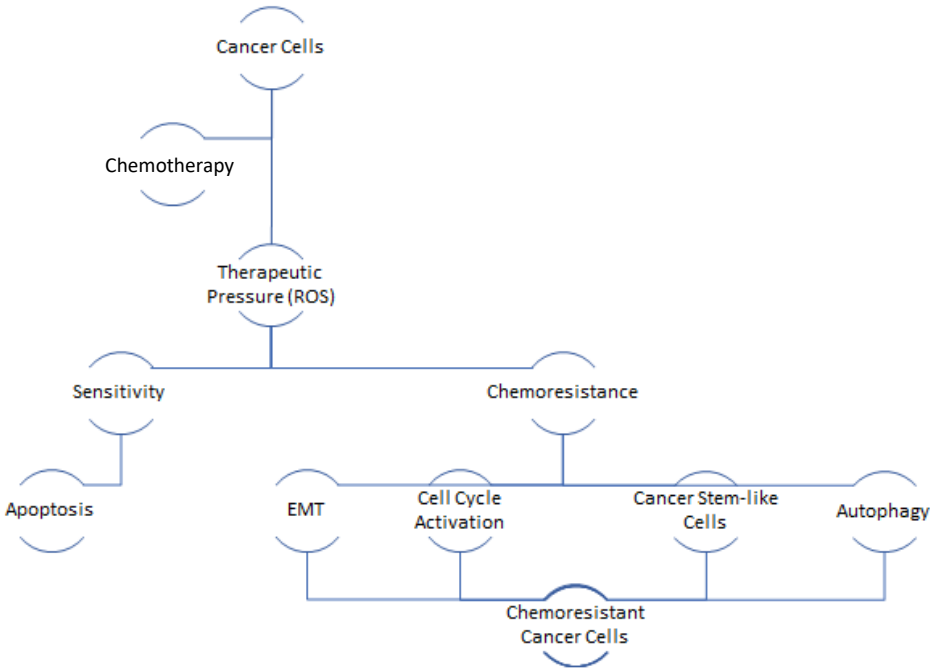


Figure 1.1.6.1- Methods by which cancer cells become chemoresistant and evade elevated ROS levels. These are: endoplasmic reticulum stress and autophagy, cell cycle disruption to elude cell cycle arrest, conversion of epithelial cells to mesenchymal cells and the reprogramming into cancer stem cells.

The first method of resistance is for the cancer cells to become tolerant towards endoplasmic reticulum (ER) stress-induced cell death which is controlled by kinases, enzymes and transcription factors. Due to the mutations within the tumour suppressor genes or the activation of oncogenes, chemotherapy causes the ER to activate stress response related factors which in turn cause tumorigenesis. This ER stress as well as ROS stimulus activates autophagy which degrades organelles and protein aggregation in addition to aiding in the acquisition of chemoresistance. For example, cancer cells resistant to 5-FU show an overexpression of the key components required for autophagy such as ATG5 and increased levels of autophagy in the cells (Li *et al.,* 2020). Next, the cancer cells high levels of reactive oxygen species cause dysregulation and phosphorylation of cell cycle regulatory factors and upregulation of cyclin which allows an accelerated progression through the cell cycle and avoidance of cell cycle arrest. A change in ROS levels also promotes epithelial-mesenchymal transitions (EMT) in cancer cells resistant to drugs such as chemotherapy (Rasad *et al.,* 2017). Chemoresistance in colon cancer, for example, results in cells with stronger invasion and migration properties. Pathways within epithelial-mesenchymal cells are highly linked to renewal and maintenance of cancer stem cells and many drug-resistant cancer cells show properties of both cancer stem cells (CSC) and EMTs (Kim *et al.,* 2019).

1.2 The globin family

Examples of redox-regulatory proteins include a protein discovered in 2001 called Cytoglobin (Cygb). This is one of the most recently characterised additions to the globin family.

1.2.1 Globin family introduction

The globin family are a group of small globular metalloproteins which all share a common haem prosthetic group used for a range of different important biological functions within the cell (Estarellas *et al.,* 2016). They are generally around 150 amino acids in size and in major eukaryotic groups they are often composed of eight alpha-helical segments that have a common three-over-three alpha-helical sandwich conformation. This identifiable feature is

often known as the globin fold though some members of the family only contain four alpha-helices (two-over-two sandwich fold) and various bacterial and plant globins only contain six helices.

Some globins have a singular histidine side chain that coordinates an axial side on the haem iron to aid in the stability of the prosthetic group, leaving the additional axial site open for the binding of exogenous ligands such as oxygen. Reversible binding of these ligands only occurs when the protein is in the ferrous oxidation state. These are known as penta-coordinate. Some globins, rather than having a free site for binding, have two histidine side chains coordinated with their haem domains instead. This means that one of the histidine side chains is capable of reversibly associating and dissociating to allow the exogenous ligands to bind in its place; these are hexacoordinate (Kakar, 2010). The naming system refers to the number and positions of the histidine molecules which coordinate the haem iron; those with a single histidine are known as penta-coordinate and those with two (a proximal and a distal histidine) are hexa-coordinated. Figure 1.2.1.1 shows that in A, the distal histidine is not connected to the haem unlike in B. Examples of hexa-coordinated globins include Cygb, Ngb and Adgb whereas penta-coordinated globins include Mb and Hb.

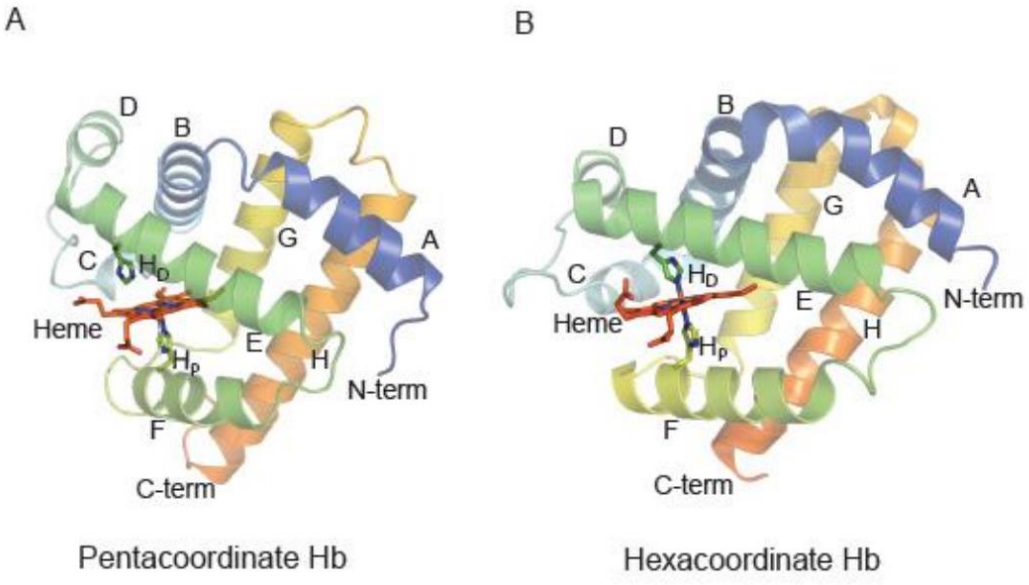


Figure 1.2.1.1- a) Pentacoordinate whale Mb haem, b) Hexacoordinate Ngb haem. The alpha helices are labelled A to H and key residues such as the proximal and distal histidine's as well as the N and C termini are labelled. Protein shown as a ribbon, haem as a stick (Kakar, S., 2010).

Though the various globin proteins have been labelled as either pentacoordinate or hexacoordinate thanks to spectroscopic analyses and protein structure characterisation, some are harder to characterise. Neuroglobin is tightly hexacoordinate, showing all of the characteristics for this. Cygb on the other hand is often described as penta-coordinate-like since it shows characteristics of both for example Cygb's oxygen affinity comparable to that of the penta-coordinate Mb (Trent and Hargove, 2002; Springer *et al.,* 1994).

Additionally, the iron atom within the centre can vary in its oxidation state and can be in either a ferrous (Fe^{2+}), ferric (Fe^{3+}) or ferryl state (Fe^{4+}) but the state restricts oxygen binding capabilities as it can only bind oxygen in the former state.

1.2.2 Haemoglobin

Hb is a two-way respiratory carrier that delivers oxygen from respiratory surfaces including lungs to organs such as the heart through the circulatory system and then facilitating the return transport of carbon dioxide. This affinity for oxygen is affected by properties such as pH, temperature and carbon dioxide levels. It also has enzymatic activity and, when oxygenated, can scavenge toxic nitrous oxide and allow this to be converted to nitrate; when deoxygenated, it creates nitrous oxide from nitrite. These form the basis of nitric oxide metabolism which causes blood vessel dilation (Burmester and Hankeln, 2014).

In terms of its structure Hb is made up of four subunits with each one comprising of a haem group and a polypeptide chain of 141 and 146 amino acid residues for alpha and beta chains respectively (Marengo-Rowe, 2006). They have a molecular mass of 64,458 Da and has a complex quaternary structure (Thomas and Lumb, 2012). The haem prosthetic group, as seen in Figure 1.2.2.1, is comprised of a protoporphyrin ring made up of four individual pyrrole rings that are connected via methine bridges that create a tetrapyrrole ring which is seen in biological molecules such as chlorophyll and cytochrome. These all surround a central iron atom. The central iron has four nitrogen atoms from the porphyrin ring bound to it in addition to the proximal histidine, HisF8 which is found in the F helix (Yang *et al.,* 1995). When reduced,

oxygen molecules bind to the available site of the iron, the sixth coordination (Burmester and Hankeln, 2009). It has cavities which act as temporary ligand docking stations for the small gaseous ligands including oxygen, carbon monoxide and nitrous oxide (Sanctis *et al.,* 2004).

There are also a number of different side chains attached to this structure including two propionate groups, four methyl groups and two vinyl groups (Ponka, 1999).

Figure 1.2.2.1: Crystal structure of human Hb alpha in the carbonmonoxy state showing the haem domain. Created on deepview and pymol, structure obtained from NCBI, 1IRD. Protein is shown as a 3D ribbon and the haem in stick format.

There are many different types of Hb, one of the most famous discovered in 1949 by Pauling and his team who characterised sickle cell haemoglobin (HbS) which differs from standard adult Hb by a singular amino acid. There are now over 200 different variants of Hb, many of which are abnormal (Marengo-Rowe, A.J., 2006).

1.2.3 Myoglobin

Mb is a globular protein of around 153 amino acids in length and also contains the characteristic proto-haem seen in Hb (Austin *et al.,* 1975). This protein had its crystal structure deciphered in 1958 by John Kendrew, making it the first protein to have a three-dimensional structure characterised (Kendrew *et al.,* 1960). It is a globin found within striated muscles and cardiac muscles that contract for extended periods of time. It is produced within the sarcoplasm in response to the mitochondria not receiving enough oxygen and is transported from the sarcolemma through Mb-facilitated oxygen diffusion (Wittenberg and Wittenberg, 2003). Like Hb, Mb is also pentacoordinate, as shown in Figure 1.2.3.1.

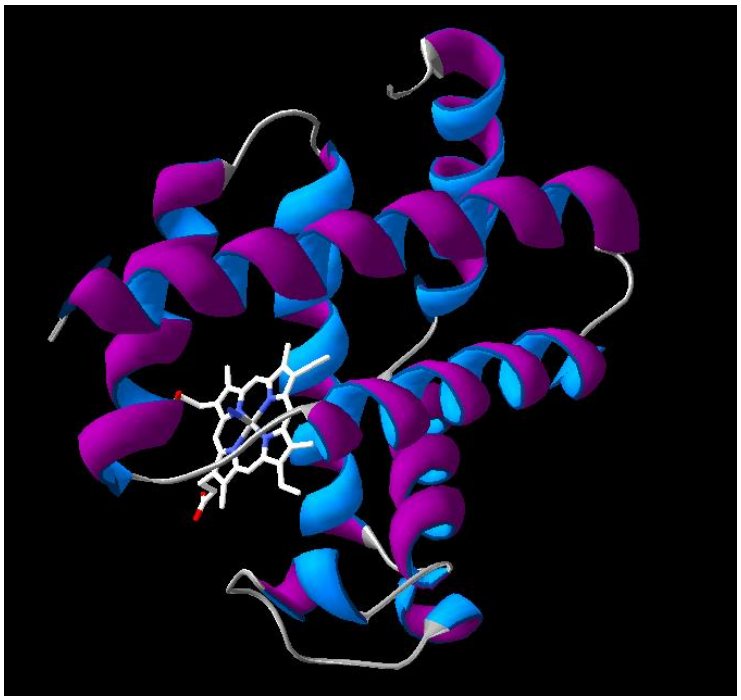


Figure 1.2.3.1: Single crystal structure of human Mb. The protein is displayed as a ribbon and the haem domain as a stick. Image created using deepview and pymol; structure obtained from NCBI, 3RGK, which contained a mutant (K45R) which was corrected in this structure.

It reversibly binds oxygen using the complex created by the haem residue and porphyrin ring containing iron. Its main function is to buffer the intracellular oxygen concentration when there is an increase in muscle activity and to aid in intracellular gas diffusion. In addition, it is widely accepted that Mb can scavenge nitric oxide, but theories are also present that suggest it could act as a ROS (Ordway and Garry, 2004), similar to Cygb and is explained further in section 2.6.2.7. It is able to store oxygen within the muscle in order to release it during periods of low oxygen or hypoxia, though the ability of this is dependent on the type of mammal. Whales have a high capacity for this whereas in humans this is limited.

1.2.4 Androglobin

Adgb is the newest member of the globin family and was discovered in 2012. It is much larger than the other globins at approximately 1667 amino acids in length, with the next biggest, CYGB, at 190 amino acids. In terms of predicted structure, it has the characteristic globin fold embedded within a conserved multi-domain protein and is reported to be hexacoordinated in *E. coli* lysate. Additionally, it contains an IQ calmodulin binding domain and an N-terminal calpain-like domain. However, unlike other globins, the globin domain of Adgb is circularly permuted. Such circular permuted proteins are very rare in nature and unique amongst globins. Adgb is found in greatest abundance in the testes but is also found in lower levels in the lung and in very low levels in other tissues. During testis development there is a strong increase in Adgb levels which then remains through adulthood. Definitive functions are yet to be determined, as is the crystal structure, though it has been postulated that functions could include redox-regulated signalling functions, autoxidation potential or alternatively mediation of oxygen level-dependent protein activity (Hoogewijs *et al.,* 2012).

1.2.5 Neuroglobin

Discovered in 2000 during genome sequence data mining, a globin highly expressed in the nervous system was discovered by Burmester and Hankeln and named Ngb (Burmester and Hankeln, 2009). Much like Adgb and Cygb, functions of this protein remain elusive although

most agree that the primary role is NO scavenging and protecting neuron mitochondria from NO toxicity (Jin *et al.,* 2008). It has been seen to improve cell viability under different oxidative stresses including hypoxia without being upregulated. Due to its positive relationship with both oxidative metabolism as well as mitochondria, it has been hypothesised that it could be involved with oxygen supply. Conversely, detoxification of ROS and nitric oxide have been proposed alongside its participation in signalling mechanisms for a range of functions such as apoptosis inhibition and transmission of redox states (Raychaudhuri *et al.,* 2010).

The protein itself is similar in structure to Mb as shown in Figure 1.2.5.1 and is a monomer of around 150 amino acids in length. As with the other globins, the typical globin fold is present though Ngb is a hexacoordinate globin in the absence of exogenous ligands in both ferrous and ferric oxidation states, which differs from the penta-coordinate forms of Hb or Mb. This means that external ligands such as O_2, NO and CO must compete with the histidine when binding to the central iron (Burmester and Hankeln, 2009).

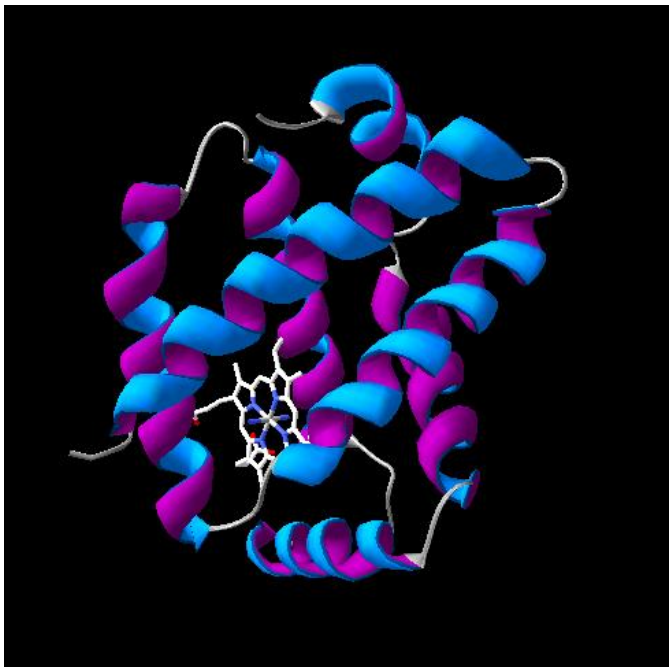


Figure 1.2.5.1: Crystal structure of murine Ngb monomer bound to CO. Created using deepview and pymol. PDB file obtained from NCBI, code 6I3T. Protein in ribbon and haem in stick.

1.2.6 Cytoglobin structure and functions

1.2.6.1 Cytoglobin discovery and structure

Cygb was first discovered in hepatic stellate cells by Katsutoshi Yoshizato and his team in 2001 while investigating tissue repair regulators in damaged livers (Yoshizato *et al.,* 2016). They performed a comparative proteomic analysis of normal and activated hepatic stellate cells within rats during which they uncovered a protein which at the time was named the stellate cell activation-association protein. It was then renamed Cygb due to its high similarity to members of the globin family (Yoshizato *et al.,* 2016) but has also been published under the name Histoglobin (Grozdanic *et al.,* 2004). Since then, Cygb has been found in almost all human tissues and is localised to nuclear speckles as well as the cytosol. A possible factor which allowed this protein to remain hidden for so long is that Cygb has relatively low expression levels within cells (Fago *et al.,* 2004). Structurally, Cygb is composed of 190 amino acid residues which makes it roughly average in size compared to the very large Adgb and the smaller Ngb. Figure 1.2.6.1.1 demonstrates the globin's varying sizes.

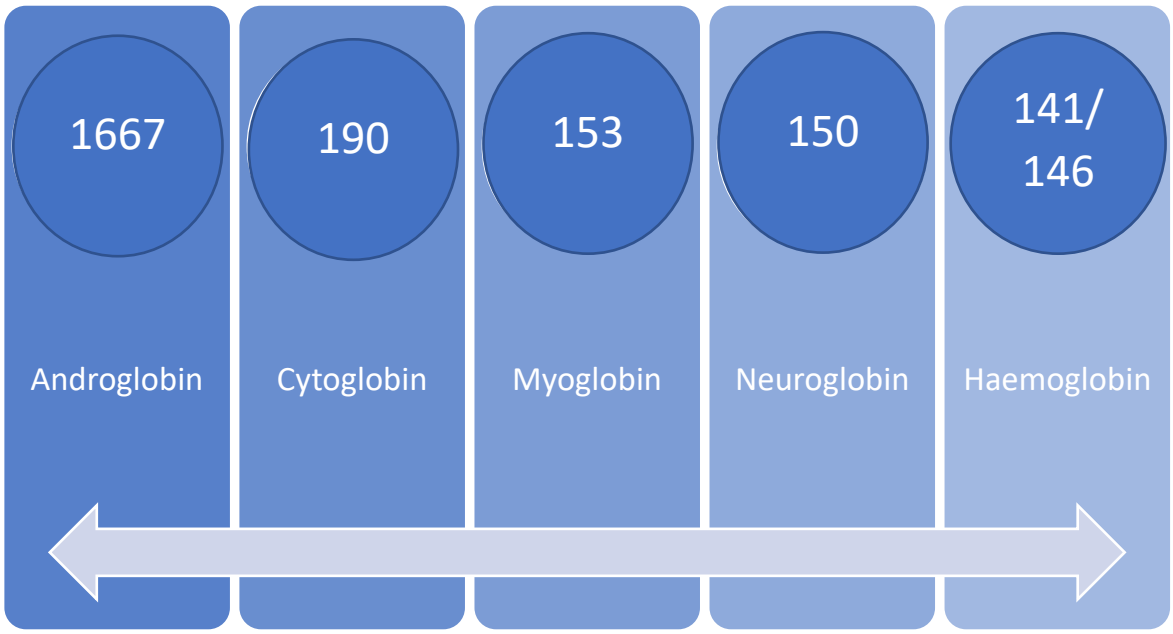
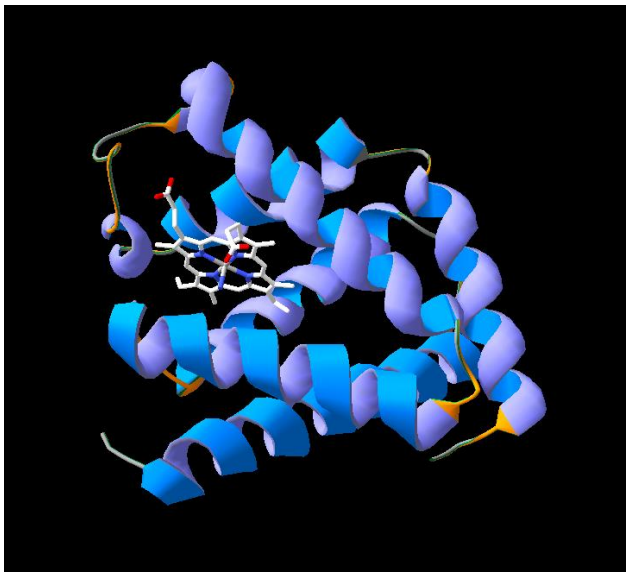


Figure 1.2.6.1.1: Length of human globins in amino acid residues from largest to smallest. All proteins in monomer form (for Hb 141 is alpha and 146 is beta).

Sequence alignments show that the protein core of Cygb is very similar to both Mb and Hb and suggests a shared ancestry. Despite their similar structure, Cygb is dimeric and monomeric comparative to the solely monomeric structures of both Mb and Ngb and the tetrameric structure of Hb. The change in state between monomer and dimer within Cygb is due to the position and integrity of the disulfide bonds within the structure. The forms include a monomeric protein with an intramolecular disulfide bond between Cys38 and Cys83, two cysteine resides; a dimeric form which has two intermolecular disulfide bonds and a monomeric form with an absence of intramolecular disulfide bond through reduced thiols. The intramolecular disulfide bond changes the dissociation rate constant for the intrinsic histidine, H81, which is postulated to manage the binding of exogenous ligands.

Cygb also has the characteristic three-over-three alpha helical globin fold shown in Figure 1.2.6.1.2. When not bound to oxygen, the distal histidine on the E helix is coordinated to the iron atom within the haem domain resulting in a lower ligand affinity, much like Ngb, Cygb is hexa-coordinated with the distal and proximal histidine's providing the fifth and sixth coordination's (Zhang *et al.* 2014). Cygb has an additional 20 residues at both its C terminus and N terminus which are distinctive of this globin though the purpose of these is unclear and are generally thought to be disordered due to the lack of crystal structure information. It also has a large protein matrix cavity which is apolar that allows a haem ligand diffusion pathway.

a)



b)

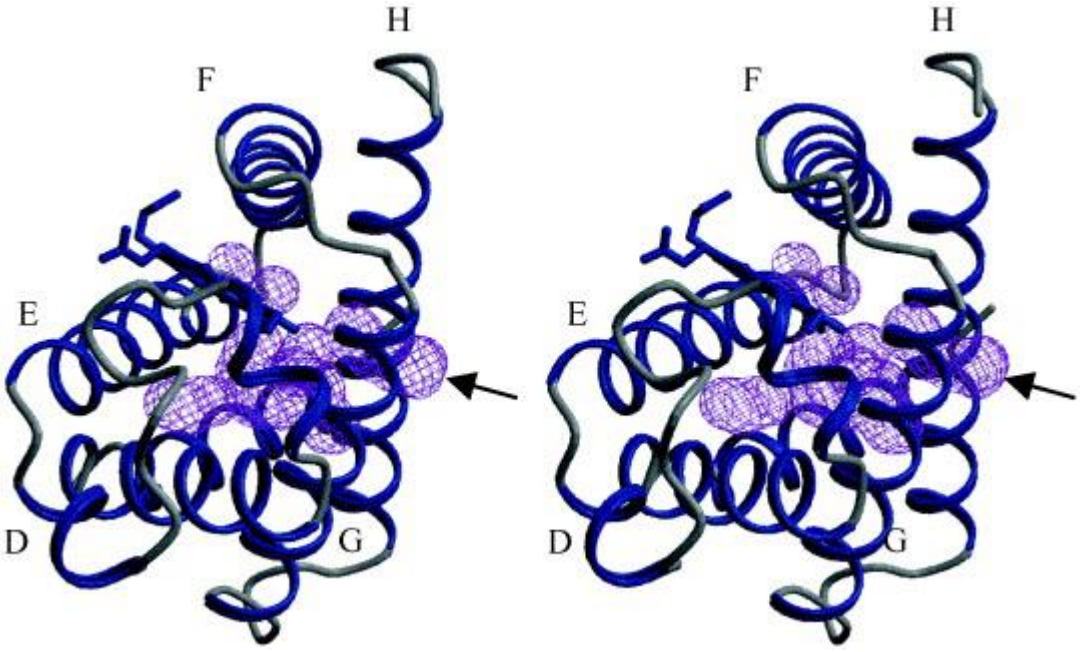


Figure 1.2.6.1.2: a) Single crystal structure of wild type Cygb. a) Created using deepview and povray. Structure obtained from NCBI, 2dc3 homo sapiens. Protein displayed as a ribbon, haem with a stick. **b) Cygb matrix cavity system.** b) Tunnel and cavity surface defined by a 1.4 Armstrong radius probe with the black arrows indicating a narrow connection between the inner cavity region and external solvent space. Images drawn with MOLSCRIPT and Raster3D (Sanctis *et al.,* 2004).

1.2.6.2 Potential functions of cytoglobin

Due to its many similarities to other globins, Cygb was initially assumed to be a respiratory protein comparable to Hb or Mb however recent investigation has shown that it plays a much less specific role than other globins. The functional roles of Cygb still remain unclear though many theories have been proposed. While Hb is found within red blood cells, Mb in muscle cells and Ngb within the nervous system; Cygb is found within almost all human tissues which makes it more difficult to isolate a function. Cygb does have some interesting characteristics in response to oxidative stress and hypoxia and potential roles have been listed as ROS scavenging, oxygen storage and antifibrotic activities.

1.2.6.2.1 Oxygen storage

The oxygen storage hypothesis has arisen due to shared common structural elements with Hb and like Mb, has a high affinity for oxygen- Mb is around 1 Torr, Cygb is around 0.2 Torr (McRonald *et al.,* 2012) and adult Hb is 26 Torr, though this aptitude for binding ligands such as oxygen is dependent upon pH. The oxygen binding to Cygb has also been reported to be cooperative (Fago *et al.,* 2004). Cygb binds to a histidine on the E helix when not bound to oxygen or other ligands and so when reduced, the affinity for these gases decreases and makes it harder to bind further molecules (Bholah *et al.,* 2015).

1.2.6.2.2 Nitric oxide metabolism

Cygb, as mentioned above, is very similar to Mb in terms of structure and evolutionary lineage. Cygb levels in vascular smooth muscle are elevated compared to Mb; this suggests that Cygb has a unique function in muscles (Liu *et al.,* 2017). Cygb has a faster reduction rate than Mb with $Cygb^{3+}$ having a 415-fold greater initial reduction rate than Mb (Liu *et al.,* 2014). It also has a higher rate of NO consumption; NO mediates vascular relaxation through binding to and subsequently activating soluble guanylate cyclase in the smooth muscles of vessels. The levels in these vessels are maintained through a balance between NO generation and NO metabolism. NO degradation is thought to be mediated by NO dioxygenase or NOD which

oxidised NO to nitrate (Liu *et al.,* 2017). Cygb, like most globins, has a high nitric oxide dioxygenase activity (NOD) that is able to be reduced rapidly by the reducing systems which infers a function in oxygen-dependent metabolism of NO.

Equation 1.2.6.2.2.1: $Fe^{2+} - O_2 + NO \rightleftharpoons Fe^{3+} - ONOO^- \rightarrow Fe^{3+} + NO_3$ Equation 1

Equation 1.2.6.2.2.2: $Fe^{2+} + NO_2 - + H^+ \rightarrow Fe^{3+} + OH^- + NO$ Equation 2

The above equations (Reeder and Ukeri, 2018) explain the concept that under normoxic conditions, Cygb functions as an NOD to scavenge NO which is the concept of equation 1 and in equation 2, under hypoxic conditions, the protein generated NO to function as a nitrite reductase. Since ferric Cygb is produced, this needs to be re-reduced before the process begins again, hence the fast reduction. This is illustrated in Figure 1.2.6.2.2.1 which depicts the processes of Cygb-mediated NO homeostasis.

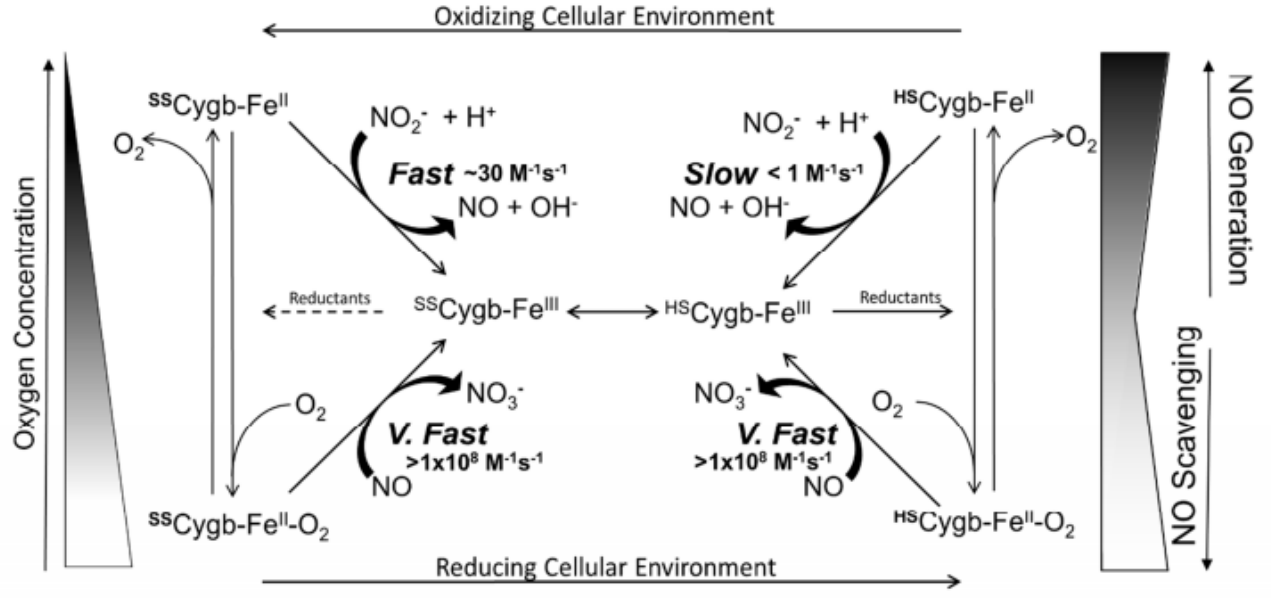


Figure 1.2.6.2.2.1 Cygb-mediated control of NO homeostasis showing the generation and scavenging of NO. These two processes occur under different conditions: normoxic and hypoxic. (Reeder and Ukeri, 2018).

Using knockdown studies, it was found that the amount of endothelium-dependent and independent relaxation was increased with a greater than 20-fold shift in the vasodilation dose-response curves. Additionally, without Cygb, the diffusion flux of NO across the aortic wall was more than six times greater (Liu *et al.,* 2017).

1.2.6.2.3 Cytoglobin and nitrite reductase

Numerous reports have proposed that haem-proteins such as Ngb, Mb and Hb reduce nitrite to nitric oxide in mammalian cells and blood when under anaerobic conditions. Nitric oxide is an important molecule that is an essential component in the pathogenesis of cellular injury (Li *et al.,* 2012). Therefore, it was suggested whether Cygb would perform the same task; a study conducted by Li (2012) and his team found that Cygb reduced nitrite with NO generation and is facilitated by acidic conditions. They saw that the oxidation rate of Cygb increased by approximately 15 times between pH 7.0 and 5.5. This is due to the need for H^+ as shown in Figure 1.2.6.2.2.1 so all the globins react faster in acidic conditions (Kim-Shapiro and Gladwin, 2014).

Since the observation that Cygb does indeed have nitrite reductase properties, further studies were used to compare the reaction kinetics and speeds. It was initially reported that Cygb, like all globins, has rapid NO dioxygenase activity but seemingly slow rates of nitrite reductase activity. It was later found that the reductase activity was strongly dependent on the oxidation state of the two cysteine residues, which are exposed on the surface of the protein and are found at positions 38 and 83 on the amino acid chain. These two amino acids are important since they can form the intramolecular disulfide bond though the oxidation state *in vivo* is not known. This bond formation accelerates the nitrite activity by 50 times when compared to the monomer and significantly more (140 times) when compared with the dimer displaying intermolecular disulfide bonds (Reeder and Ukeri, 2018). This is the highest nitrite reductase activity reported for animal globins, though some plants have measured higher. Studies on Ngb have shown that mutation of the distal histidine increases the nitrite reductase activity by around 2000-fold (H64L) compared to the wild type protein (Tiso *et al.,* 2011).

1.2.6.2.4 Cytoglobin and H_2O_2

Though the functions of Cygb are still contested, evidence for interactions with H_2O_2 are gaining momentum. Within other globins such as Mb or Hb, reactions with H_2O_2 are well characterised. Both globins form the ferryl species when the ferrous or ferric protein interacts with peroxides (King and Winfield, 1963).

A 2007 study found that Cygb was selectively upregulated with H_2O_2 treatment. After treating them with 250 μM H_2O_2 for a range of different times (0, 3, 6, 12 and 24 hours), it showed that the Cygb expression levels increased by around three-fold, though this was not tested using different concentrations of peroxide. They proposed that Cygb might be sensitive to H_2O_2 but not hypoxia. Additionally, it was shown that the Cygb expression knockdown in cell lines created with siRNA aggravated H_2O_2-induced cell death. These results indicate a critical role for Cygb in protecting cells from oxidative injury (Li *et al.,* 2007).

1.2.6.2.5 Antifibrotic effects of Cygb

Cygb is expressed by fibroblasts in many different organs. Overexpression of Cygb in transgenic rats improved renal function, lessened the effects of histological injury and ameliorated fibrosis (Nishi *et al.,* 2011). This showed that Cygb can inhibit collagen synthesis and in addition, when the haem was disrupted in an area known for antioxidant properties, the result was a loss of antifibrotic effects (Mimura *et al.,* 2010). This allowed a tentative conclusion that Cygb is able to reduce fibrosis through a radical scavenging function. The group also showed that Cygb is essential in the protection of kidneys against fibrosis through the amelioration of oxidative stress in both *in vitro* and *in vivo* studies. It is considered that it could be a potential target for some kidney diseases (Mimura *et al.,* 2010).

1.2.6.2.6 Antioxidant functions of Cygb

One of Cygb's most reported claims is the ability to scavenge ROS produced by cells and therefore act as an antioxidant. Studies showed that when ROS levels were assessed by measuring the oxidation of the redox sensitive dye (H2DCF-DA), this dye is trapped within the cell and then oxidised by ROS to create a fluorescent DCF molecule. The level of fluorescence

is then proportional to the amount of oxidative stress in cells. The oxidative stress is then promoted artificially by inhibiting the rate-limiting step in glutathione synthesis (McRonald *et al.,* 2012). Evidence showed that overexpression of Cygb in cancer cell lines reduced the levels of chemically induced ROS within the cells (McRonald *et al.,* 2012). Additionally, the same laboratory showed that Cygb affords protections from pro-oxidant induced injury, namely lipid peroxidation and the breaking of DNA strands due to oxidation. This has led to the belief that Cygb has peroxidase activity (McRonald *et al.,* 2012).

Other studies showed that Cygb was seen to protect hCPCs against cell death induced by oxidative stress, providing potential therapeutic target for enhancing effectiveness of cardiac stem/progenitor cell therapy for ischemic heart disease. One plausible hypothesis is that Cygb functions as a pro-survival factor in response to oxidative stress which is associated with the upregulation of primary antioxidant systems e.g., haem oxygenase-1 and anti-apoptotic factors such as MCL1 (Zhang *et al.,* 2017).

1.2.7 Cygb and cancer

Recently, Cygb has been implicated in cancer though it is still unclear as to what extent this globin is involved. It is found in the chromosomal region of 17q25 which is often lost in multiple malignancies (Shivapurkar *et al.,* 2008).

Cygb has been shown to be downregulated by hypermethylation in most cancer cells and its exogenous over-expression reduced proliferation of cancer cells so it is a possible tumour suppressor (Bholah *et al.,* 2015). Genetically, Cygb's gene contains a long CpG island, a long stretch of DNA with unusually elevated levels of CpG dinucleotides (Jeziorska *et al.,* 2017). DNA methylation modifications result in additional methyl group being added to cytosine bases and in animals, this methylation mainly occurs at CO or CpG dinucleotides. The process of DNA methylation has become an essential evolutionary mechanism that is involved in systems such as expression of endogenous genes and X-chromosome inactivation (Xin *et al.,* 2011). This is found upstream and has many transcription sites with unusual elements such as those

sensitive to hypoxia (hypoxia responsive elements) and hypoxia-inducible protein binding sites such as HIF1, SP1 and AP1. When overexpressed in hypoxic conditions, Cygb promotes cell survival whereas low expression resulted in higher levels of apoptosis in control cells under the same conditions (Bholah *et al.,* 2015). This short CpG-rich segment in the promoter region that is correlated with gene silencing is the first evidence of gene silence and Cygb methylation being linked. An investigation of a correlation between tumour specificity and Cygb methylation in a number of cancers was determined by targeting this short sequence in many different cancers including lung cancer, though this did not provide any conclusive evidence. In conclusion to this, the study hypothesised that Cygb methylation may be an early change in the tumour's pathogenesis (Shivapurkar *et al.,* 2008). It also possesses a slower mutation rate in its amino acid sequences compared to other members of the family which proposes a potentially essential function that has been highly conserved over an incredibly extended period of time (Fago *et al.,* 2004).

When Cygb was knocked down in a lung cancer cell line cell using RNAi, it displayed a higher level of colony formation than the WT. Additionally, when transfecting the protein cDNA into a cancer cell line as well as into a cancer cell line that showed no Cygb expression, they both exhibited a reduction in tumour cell growth. A further study showed a change in the growth of the tumour cells- this time growing on top of each other, as a result of p53 knockdown. All of these studies provided evidence that Cygb acts as a tumour suppressor (Shivapurkar *et al.,* 2008). Further evidence to support the idea that Cygb is a tumour suppressor comes from the Head and Neck Surgery Unit in Liverpool. Cygb is the only gene completely contained within tylosis that suffers frequent deletions in sporadic oesophageal cancer. Tylosis is a rare condition which causes focal thickening of the skin on both hands and feet and causes a very high risk of developing squamous cell carcinoma in the oesophagus, particularly in people over the age of 65 (Ellis *et al.,* 2015). The team found that the Cygb promoter was found to be significantly methylated in around two thirds of the head, neck and squamous cell carcinoma samples and significantly more than the controls (Shaw *et al.,* 2006).

Despite the considerable evidence that Cygb performs as a tumour suppressor, data from different studies is conflicting. The upregulation of Cygb is associated with tumour hypoxia and aggressiveness. However, overexpression of Cygb in cancer cell lines under normoxic conditions was found to reduce cell migration, invasion and anchorage-independent growth; it also hindered cell proliferation in lung cancer lines. However, once H_2O_2 treatment was given, the cell viability, migratory potential and anchorage-independence was not reduced, and these processes were protected. In hypoxic conditions, Cygb overexpression caused a decrease in cell viability, migration and anchorage-independence which displays oncogene functions (Oleksiewicz *et al.,* 2013). This appears to show that in normoxic conditions Cygb behaves as a tumour suppressor but under hypoxic conditions, it performs the role of an oncogene.

Cygb has been heavily implicated in cancer therapy resistance via the protection of the tumour cells from harsh environments created by chemotherapeutics, such as oxidative stress. A study showed that Cygb was able to protect a human neuroblastoma cell line from oxidative-induced cell death under certain anoxic conditions which deprived the cells of oxygen and glucose (Fordel *et al.,* 2006). This was created through the introduction of antisense oligodeoxynucleotides against the Cygb gene but could also be achieved by overexpressing Cygb in the cells since it is believed to play a part in ROS regulation. The protective properties included limiting the induction of intracellular ROS and pro-oxidant-induced damage to the DNA (Lv *et al.,* 2008).

1.2.8 Current advances in cancer treatment using Cygb

Following the discovery that Cygb may have oncogenic properties, the potential use of Cygb in cancer therapies has been widely investigated. One study has found that the highest Cygb expression was in melanocytes, the melanin-producing cells in the epidermis. In addition, they associated DNA methylation induced reduction of Cygb expression with melanocyte to melanoma transition. Melanoma only makes up a small fraction of total skin cancer cases, however it is a very aggressive form with a high death rate.

The viability of a technique known as cold atmospheric plasma (CAP) in destroying cancer cells was investigated by one group (Backer *et al.,* 2018). Since it is widely published that Cygb acts as a ROS scavenger and therefore regulator; when studying CAP on melanoma cell lines, they also measured the ROS concentration intracellularly in the cell lines. The results found a correlation between the levels of Cygb expression, the sensitivity of the cancer cell line towards the CAP treatment in addition to the intracellular concentration of ROS (Backer *et al.,* 2018). Though the mechanisms are still unclear, it is evident that Cygb has an effect on the effectiveness of the treatment.

It has additionally been speculated that it could be a candidate for cancer gene therapy, though could also play a part in the treatment of fibrosis and diabetes (Lv *et al.,* 2008). Another potential avenue of investigation is the use of Cygb as a biomarker for cancer and some other diseases since the induced overexpression of the protein causes a decrease in the ability of cancer cells to proliferate (Bholah *et al.,* 2015).

1.3 Aims of the project

1.3.1 Characterisation of the structure and function of cytoglobin

This project will aim to further characterise Cygb and understand its potential functions, with emphasis on its redox sensitivity. These investigations will compare results obtained from the wild type Cygb protein as well as two mutants: H81A and L46W. The effects of the following interactions and conditions will be tested on the protein:

- Nitrite Reductase- The H81 mutation that will be used in this investigation has been mapped in Ngb and significantly increased the rate of nitrite reductase. The aim is to observe whether this is consistent within Cygb.
- NO Dioxygenase- The L46 mutation selected for experimental work decreased NO dioxygenase activity in both Mb and Hb so again, the aim is to see whether results are comparable in Cygb.
- Lipid Interactions- Previous studies have shown that upon binding to a lipid molecule the protein changes from a hexacoordinate to a pentacoordinate state. One of the mutations that will be used fixes the protein in a pentacoordinate state so this will give an insight into the binding of the lipid and the protein.
- Liposome Interactions- Interactions with liposomes containing phospholipids will allow peroxidase activity to be measured since this has been thought to be relevant to cytoprotection of cells.
- Ascorbate- Ascorbate has been seen to reduce the protein and in Mb, produce ascorbyl radicals. This will be investigated in Cygb and in particular, the two mutants.

1.3.2. Protective abilities of cytoglobin when under oxidative stress

The project also aims to investigate the protective effect of Cygb (WT and the two mutant varieties when subjected to conditions which create oxidative stress or damage to the cells. These include the chemotherapy drugs: paclitaxel and docetaxel as well as H_2O_2. They will mainly be tested on breast cancer cell lines (MCF7) and human embryonic kidney cell lines (HEK). This is due to the high incidence level of this cancer within the population and the difficulty in treating breast cancer, but also in HEK cells due to their very fast replication time and ease of transfection.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 All reagents and their suppliers

Table 2.1.1.1: Reagents and suppliers

Reagent	Supplier
1 Kb DNA ladder	ThermoFisher Scientific
6x DNA loading dye	ThermoFisher Scientific
Agarose	Fisher
Ammonolevulinic acid	Fluka
Ampicillin	Sigma-Aldrich
Carbon monoxide	BOC
Crystal violet stain	Pro-Lab Diagnostics
DMEM F12	Life Technologies Inc.
Docetaxel	Sigma
EDTA disodium salt	Fisher
Ethanol	Fisher
Ethidium bromide	Sigma
FBS	Sigma
Fe Citrate	Sigma
FuGene HD	Promega
G418 sulfate	ThermoFisher Scientific
Gentamycin sulfate	Corning
Glacial ethanoic acid	Fisher
Hydrogen peroxide	Sigma
Imidazole	Fisher Scientific
IPTG	Melford
Kanamycin sulfate	Sigma-Aldrich
Lipofectamine 3000	Thermofisher
Luria broth	Melford
L-α- phosphatidylcholine	Sigma-Aldrich
Optimem	Invitrogen
Paclitaxel	Sigma
PBS	Genzyme
PEI transfection reagent	Sigma-Aldrich
PFA	Merck
Proli NoNoate	Cayman Chemical Company
Sodium ascorbate	Sigma
Sodium chloride	Sigma-Aldrich
Sodium dithionite	Fisher
Sodium nitrite	Sigma-Aldrich
Sodium oleate	Sigma-Aldrich
Sodium phosphate	Aldrich

Sodium tetraborate	Sigma
STU1 restriction enzyme	New England BioLabs
Tris base	Fisher
Trypsin with versene	Lonza

2.1.2 DNA purification and extraction kits

Table: 2.1.2.1 Kits

Kit	Supplier
PCR clean up kit	New England BioLabs
Qiagen mini/midi-prep kit	Qiagen
Qiaquick gel extraction kit	Qiagen

2.1.3 Buffers, media, reagents and solutions

2.1.3.1 Agarose gel electrophoresis

Table 2.1.3.1.1: Reagents for agarose gel electrophoresis

Name of reagent	Components
1% agarose gel	0.3 g agarose in 30 ml 1x TAE buffer Agarose dissolved by boiling
1 Kb DNA ladder	N/A
Ethidium bromide	N/A
1x TAE buffer	100 ml TAE (48.4 g Tris Base, 11.4 ml glacial ethanoic acid,3.7 g EDTA disodium salt), 900 ml ultra-pure milliQ water
10x FastDigest green buffer	

2.1.3.2 Bacterial cloning

Table 2.1.3.2.1 Reagents used in bacterial cloning

Solution	Reagents	Sterilisation method
Ampicillin	1 g ampicillin made to 10 ml with 18.2 MilliQ water	N/A
Kanamycin	0.5 g kanamycin made to 10 ml with 18.2 MilliQ water	N/A
Luria broth (LB)	4 g Luria broth in 200 ml 18.2 MilliQ water	Autoclave
LB ampicillin	100 µl ampicillin in 100 ml LB	N/A
LB kanamycin	100 µl kanamycin in 100 ml LB	N/A

2.1.3.3 Protein Transformation

Table 2.1.3.3.1 Reagents used in protein transformation

Solution	Reagents
Ampicillin	1 g ampicillin made to 10 ml with 18.2 MilliQ water
Kanamycin	0.5 g kanamycin made to 10 ml with 18.2 MilliQ water
Luria broth (LB)	4 g Luria broth in 200 ml 18.2 MilliQ water
LB ampicillin	100 µl ampicillin in 100 ml LB
LB kanamycin	100 µl kanamycin in 100 ml LB
LB agar	5 g LB, 3 g agar made up in 200 ml RO MilliQ water.

2.1.3.4 Protein Expression

Table 2.1.3.4.1 Reagents used in protein expression

Solution	Reagents
Kanamycin	0.5 g kanamycin made to 10 ml with 18.2 MilliQ water
Luria broth (LB)	4 g Luria broth in 200 ml 18.2 RO water
ALA	0.66 g in 50 ml MilliQ water
IPTG	2.38 g in 50 ml MilliQ water
Fe citrate	1.5 g in 50 ml MilliQ water
CO	Excess
100 mM Sodium phosphate buffer	15.6 g NaH_2PO4 and 35.8 g Na_2HPO4, each made up to 1 L with RO water, added together until pH 7.4
160 mM Sodium phosphate with 4 M NaCl	233.76 g NaCl and 22.7 Sodium phosphate, made up with 500 ml RO water to make pH 7.4.
20 mM Phosphate buffer with 500 mM NaCl and 20 mM imidazole	62.5 ml 160 mM sodium phosphate with 4 M NaCl buffer added to 437.5 ml 160 mM imidazole.
20 mM Phosphate buffer with 500 mM NaCl and 500 mM imidazole	62.5 ml 160 mM sodium phosphate with 4 M NaCl buffer added to 437.5 ml 2M imidazole.

2.1.4 Cell cultures and treatments

2.1.4.1 Cell line information

Table 2.1.4.1.1 Cell line information

Cell line	Cell type	Derived from	Media	Source
DU145	Metastatic prostate cancer	Metastatic site of 69-year-old Caucasian man	DMEM	
HEK293S	Human embryonic kidney	Human embryo kidney cells transformed with fragments of adenovirus type 5 DNA	DMEM	
MCF7	Breast adenocarcinoma	Metastatic site of 69-year-old Caucasian woman	DMEM	
BL21DE3	Competent *E. coli* cells	Non-T7 expression *E. coli* strain	LB	New England BioLabs
XL1-Blue	QuickChange ZL1 Supercompetent cells	*E. coli* strain with mutant alleles for blue-white colour screening for recombinant plasmids	LB	Agilent Technologies

2.1.4.2 Media and other cell culture reagents

Table 2.1.4.2.1 Media and other cell culture reagents

Media or Reagent	Supplier	Additives
Dulbecco's Modified Eagle Medium (DMEM)	Life Technologies Inc.	50 ml Foetal Bovine Serum (10%) 2.5 ml gentamycin
Hormone and phenol-red free Dulbecco's Modified Eagle Medium	Life Technologies Inc.	50 ml Foetal Bovine Serum (10%) 2.5 ml gentamycin (5%)

2.1.4.3 Primers and their sequences

Table 2.1.4.3.1: Primers and their sequences

Primer Name	Sequence	Source
pBR3	TCCCCATCGGTGATGTC	Eurofins
CMV-F	CGCAAATGGGCGGTAGGCGTG	Eurofins

2.2 Methods

2.2.1 Protein creation and biophysical characterisation of the wild-type and mutant cytoglobin

2.2.1.1 Protein Transformation, expression, and purification

WT gene contained in a pET28a vector (2 µl of approximately 30-100 ng/ul) was transformed into 100 µl BL21 De3 competent *E. coli* cells by first leaving on ice for 30 minutes, followed by a heat shocking at 42 °C for 45 seconds in a water bath and then finally chilling on ice for 2 minutes. 1 ml LB media (see Table 2.1.3.3.1) was added and incubated for 50 minutes at 37 °C, 160 rpm. Agar plates (see Table 2.1.3.3.1) containing kanamycin sulfate were aseptically spread with 100 µl cells and incubated in a Thermoscientific MaxQ 6000 incubator for 10 minutes at 37 °C before being flipped over and incubated for a further 18 hours.

One colony was aseptically isolated using a pipette tip before being transferred to media (100 ml LB broth containing 50 µg/ml kanamycin sulfate) and placed into the incubator at 37 °C, 160 rpm overnight. This 100 ml culture was split evenly between several (4-8) 1.4 L LB media (and kanamycin) flasks. The cell growth monitored on a Thermoscientific NanoDrop 2000C spectrophotometer every hour. When the OD (600 nm) reached between 0.6-0.8, the following was added to each flask containing 1.4 L of LB media (plus 50 µg/ml kanamycin sulfate): 3.5 ml of 100 mM ALA (final concentration of 250 µM), 3.5 ml of 200 mM IPTG (final concentration of 500 µM), 0.7 µl of 100mM Fe citrate (50 µM final concentration) and excess CO was bubbled through the solution for ~30 seconds. The flask was sealed using a rubber bung and incubated overnight at 37 °C, 160 rpm. The media was transferred into 1 L centrifuge tubes and spun in a Sorvall Lynx 6000 centrifuge for 20 minutes at 8,000 g, 4 °C until a pink pellet was produced. The media was decanted, and the pellet extracted into a 50 ml falcon tube, diluted in 10 ml water to break up the pellet and frozen at -20 °C.

The cells were lysed using an Emulsiflex C3 machine by first priming the system with water, then cells suspended in water were filtered using a tea strainer and passed through the emulsiflex 2-3 times, 15,000-20,000 psi. Lysate was transferred into 50 ml oakridge centrifuge

tubes and spun in a Sorvall Lynx 6000 centrifuge at 22,000 g for 20 minutes at 4 °C. Volume was determined from measuring the weight of the lysate assuming 1g/ml. As the protein contained an N-terminal HisTag, the following single stage method was used to purify the protein. To the lysate, 12.5 ml 20 mM phosphate buffer with 500 mM NaCl per 100 mls of supernatant, plus 1 ml of 2 M imidazole per 100 ml supernatant was added. This was then passed through a HisTrap Chelating HP column using a Pharmacia Biotech Peristalsis P-1 pump. Once all of the coloured protein was bound to the column, the column was then washed with ~5 column volumes of 20 ml 20 mM phosphate buffer with 500 mM NaCl and 20 mM imidazole, followed by elution of the protein using 20 mM phosphate buffer with 500 mM NaCl and 500 mM imidazole. Eluant was collected and transferred into dialysis tubing, (8000 molecular weight cut off (MWCO)), and sealed at both ends. The tubing was then placed into 5 L of 1 mM sodium tetraborate, ~pH 10 and stirred slowly using a magnetic stirrer. The buffer was replaced approximately every two hours, three times and left overnight before collection and either stored at – 80 °C or further concentrated before storage. To concentrate, the protein was transferred into a 3k molecular weight cut off filter and spun in an ALC benchtop centrifuge at 4000 rpm at 10 °C. Once solution reached between 5-6 ml, this was aliquoted into 1.5 ml tubes and the concentration of one was measured on the Agilent Cary 50 UV-visible spectrophotometer.

The optical spectra of a suitably diluted protein (diluted in 0.1 M NaPi, pH 7.4 buffer) was measured in a 1 ml quartz cuvette following baselining with buffer. The UV-visible spectrum was recorded between 200 to 800 nm to give the ferric spectrum. A few grains of sodium dithionite were added and gently mixed to make the deoxyferrous form of the protein. The Soret peak absorbance was used to calculate the concentration using the extinction coefficient of 165,000 $M^{-1}cm^{-1}$ at 428 nm (Beckerson *et al.,* 2015).

2.2.1.2 DNA sequencing to verify mutation positions and DNA amplification

Two mutants; L46W and H81A were created in-house using site-directed mutagenesis. To confirm these mutations were correct, the DNA was diluted to 30 ng/ml in milliQ water and

sequenced (Eurofins). Those with genes contained in pET vectors were sequenced using pBR3 primers and those Cygb genes in a pCMV vector used CMV-F primer (see Table 2.1.4.3.1 for sequences) (Eurofins, 2019). The sequence data was compared to the WT for both the pCMV and pET vectors using SnapGene. The DNA sequence data was translated to a protein sequence using open reading frames (ORF) finder (NCBI, 2020) to identify and confirm the intended mutations.

The WT and mutant DNA was each transformed into XL1B cells using the transformation method as described in section 2.2.1.1. Transformed XL1B cells (100 µl) were spread aseptically onto agar plates containing antibiotics (50 µg/ml kanamycin for pET vectors and 100 µg/ml ampicillin for pCMV) and incubated for 18 hours as described in section 2.2.1.1. One colony was aseptically isolated and then incubated overnight with the appropriate antibiotic. This was then placed into a falcon tube and centrifuged for 20 minutes, 4 °C at 4000 rpm in an ALC benchtop centrifuge before plasmid DNA was extracted using the Qiagen Mini-Prep kit. The DNA concentration was measured using the Thermoscientific Nanodrop 2000C Spectrophotometer and the DNA frozen at -80 °C.

The pET28a plasmid for the two mutants was transformed into BL21 De3 cells and mutant protein expressed and purified, as described in section 2.2.1.1.

2.2.1.3 DNA linearization and purification to improve productive integration in transfection

WT pCMV-GFP, H81A pCMV and L46W pCMV were digested using restriction enzyme S*tuI* following the manufacturer's instructions (1U in 20 µl of 1 µg/ml DNA) and agarose gel electrophoresis (1% agarose gel) used to confirm linearization. The DNA was purified using a PCR clean up kit (see Table 2.1.2.1) . The DNA concentration and purity were measured on a Thermo Scientific Nanodrop 2000C Spectrophotometer.

2.2.1.4 Measuring optical properties of wild-type and mutant cytoglobin

Protein concentration (WT, L46W, H81A, each 10 µl) was recorded in a 3 ml quartz cuvette on an Agilent Cary 50 UV-visible spectrophotometer, baselined with sodium phosphate buffer

(0.1 M NaPi pH 7.4). The spectra were recorded between 200-800 nm with baseline correction. Optical spectra for ferric, deoxyferrous and ferrous-CO spectra were measured as described above. Oxyferrous spectra were also measured following reduction by 2 mM sodium ascorbate for five minutes as well as immediately following the addition of 800 µM NO, from a 400 µM addition of Proli NONOate (40 mM stock solution in 25 mM NaOH).

2.2.1.5 Nitrite reductase activity of cytoglobin and cytoglobin mutants

Kinetic measurements were obtained on an Agilent 8453 Diode Array spectrophotometer with a multi-cell carriage or an Applied Photophysics SX20 stopped-flow spectrophotometer with a diode array attachment. To measure the kinetics of NiR activity 10 µM protein (1400 µl) was placed into a 3 ml plastic cuvette with a few grains of sodium dithionite. Buffer, and nitrite of varying concentrations (0.5-5 mM final concentrations) were placed into a 1.5 ml micro-centrifuge tube with a few grains of sodium dithionite. Protein and nitrite solutions were rapidly mixed in a 1:1 volume ratio using a Pasteur pipette and spectra recorded. The difference in absorbance at 416 nm and 430 nm for WT Cygb (419 nm and 436 nm for L46W Cygb) was used to generate a time trace to measure the rate constant of the NiR kinetics.

For each nitrite concentration, the time trace data was fitted to a single exponential function using the Microsoft solver function.

For the H81A mutation, unlike the WT and L46W mutant, the reaction was too fast to be obtained from the Agilent 8453 diode array spectrophotometer, so a stopped flow (like that shown in Figure 2.2.1.5.1) was used to observe the reaction.

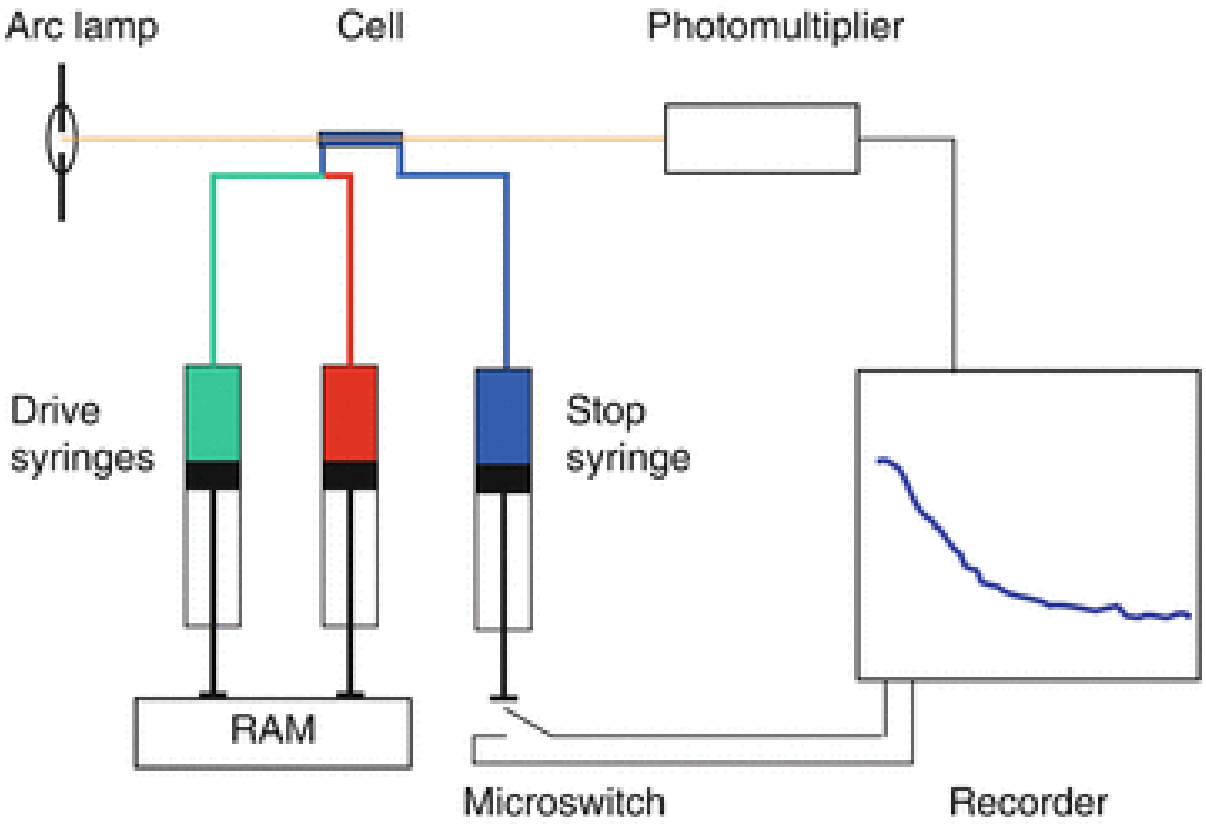


Figure 2.2.1.5.1 Schematic diagram of a stopped-flow spectrophotometer (Bagshaw, 2013). Equipment configuration includes pneumatic driven syringes, low volume turbulent flow mixer, low volume 10 mm optical cell, stopped syringe and micro-switch linked to a computer for recording. The nitrite and protein were placed into separate syringes, with small amounts of sodium dithionite added to each. Following rapid mixing in a 1:1 ratio, the spectra were monitored using a diode array system. Temperature was maintained at 25°C using a Peltier controlled water bath. Each nitrite concentration (0.5-5 mM) was measured a minimum of three times.

This data was analysed using the Applied Photophysics Pro-Kineticist analysis software, fitted to an exponential time course function by global analysis.

2.2.1.6 Binding of sodium oleate lipids of cytoglobin and cytoglobin mutants

Lipid binding was measured using oleate titrations for both the wild type CYGB protein and two mutants: H81A and L46W. The absorbances were measured on an Agilent Cary 5000 spectrophotometer. The optical spectrum of the ferric Cygb WT protein (5 µM) in sodium phosphate buffer (100 mM, pH 7.4) was recorded between 350-750 nm. A solution of fresh 10 mM of sodium oleate was made and then diluted to 300 µM oleate solution (0.5 µl) was added to the protein, mixed and the spectrum recorded. Further additions of oleate were added to the cuvette in 0.5 µl increments until fully saturated. This was then repeated for the L46W and H81A Cygb mutants. The absorbances and wavelength data was analysed and a wavelength difference for the peak and the trough of the Soret peak was calculated. These were plotted in Microsoft Excel and the Kd value and total protein concentration values were calculated to determine the stoichiometry using the fractional saturation binding curve equation (equation 2.2.1.6.1)

Equation 2.2.1.6.1:

$$y = \frac{([P_T] + [S] + \frac{1}{Kd}) - \sqrt{([P_{T]} + [S] + \frac{1}{Kd})^2 - 4([P_T][S])}}{2[P_T]}$$

Where $[P_T]$ was the total protein concentration, [S] was the substrate concentration (oleate) and Kd was the binding constant (Reeder et al., 2011). The data was fitted to the equation using the least-squares method, utilising Microsoft Excel's solver program. Error of the fit was obtained using Kaleidograph (version 4.0) software.

2.2.1.7 Cytoglobin and cytoglobin mutant proteins oxidised with L-α-phosphatidyl-choline liposomes

L-α-phosphatidyl-choline (50 mg) and 10 ml NaPi buffer (100 mM, pH 7.4) was placed in a falcon tube. The L-α-phosphatidyl-choline suspension was sonicated in a sonicator water bath to break up particles into multi-lamellar liposomes. Small, uni-lamellar liposomes (~90 nm in diameter) were generated using a Northern Lipids stainless steel extruder with a 25mm Whatman nucleopore membrane (0.1 µm pore size) and a drain disc. The contents of the falcon tube extruded once under pressure (5-20 bar) whereupon the membranes were replaced (adding an extra nucleopore membrane) and the liposomes extruded a further nine times. The now uniform and single lamellar liposomes were placed in the fridge in a falcon tube and used within 2 hours of preparation. Liposomes (200 µg/ml) diluted in 0.1 M sodium phosphate buffer, pH 7.4, were reacted with Mb, WT Cygb, L46W Cygb or H81A Cygb (all 2 µM) at 25 °C in a 1 ml quartz cuvette and run on the Agilent 8453 diode array spectrophotometer, measuring lipid oxidation by an increase in conjugated lipids at 234 nm (E_{234nm} =25,000 $M^{-1}cm^{-1}$), (Beckerson *et al.,* 2015). A 3-point baseline drop (234 nm – 220 and 250 nm) was used to correct for any effects in light scattering and the maximal rates of these were calculated from the maximum slope by calculating the first derivative.

2.2.1.8 Reduction of cytoglobin by sodium ascorbate

CYGB reduction from ferric to ferrous oxidation state was measured using addition of ascorbate for both the wild type CYGB protein and H81A and L46W mutants. The absorbances were measured on an Agilent Cary 5000 spectrophotometer. Ferric Cygb (5 µM) in 0.1 M sodium phosphate buffer, pH 7.4 was reduced to oxyferrous Cygb through addition of sodium ascorbate (5 mM) and the spectra recorded between 350-750 nm until reduction was complete. This was then repeated for the two mutants. The absorbances were measured against time and the data transferred into Microsoft Excel for data analysis.

2.2.2 Examining the potential protective function of cytoglobin within tissue cultures when exposed to conditions which create oxidative stress on cells

2.2.2.1 Cell culture

MCF-7, DU145 and HEK293S cells were grown in Dulbecco's modified Eagle's medium mixture (DMEM) F12 supplemented with 10% foetal bovine serum and 0.25 mg/ml (final concentration in DMEM) gentamicin sulfate. Cells were cultured in a humidified atmosphere with 5% CO_2 at 37°C. Stable transfected cell lines (CYGB knock-out) and GFP-tagged CYGB knock-in were generated by Brandon Reeder.

2.2.2.2 Bacterial transformation into *E. coli* cells

Plasmid DNA containing CYGB gene (~50-100 ng/µl) was added to 100 µl chemically competent cells and transformed as mentioned in section 2.2.1.1, the only exception being the antibiotic used: ampicillin (100 µg/ml) as opposed to kanamycin. 1 ml LB was added and incubated for 50 minutes at 37 °C, 160 rpm. Agar plates containing 100 µg/ml ampicillin were aseptically spread and incubated for 10 minutes, 37 °C before being flipped over and incubated for a further 24 hours. 50 µg/ml ampicillin was added to 100 ml LB broth and a single colony grown up in LB. The cultures were incubated at 37 °C, 180 rpm for 24 hours.

The cultures were centrifuged for 10 minutes, 4 °C at 4000 rpm to pellet the cells before plasmids were purified with a Qiagen Mini-Prep Kit. The supernatant was discarded, and the pellet resuspended in 250 µl resuspension solution and then placed into microcentrifuge tubes. 250 µl lysis solution was added to each tube and then inverted 4-6 times, followed by 350 µl neutralisation solution and immediately inverted 4-6 times. This was centrifuged for 5 min at 12,000 rcf. The supernatant was transferred to a Thermo Scientific GeneJET Spin column and centrifuged for 1 min. 500 µl wash solution was added to the spin column and centrifuged for 1 min, the flow-through discarded and then repeated. The empty column was then centrifuged for 1 min. The column was added to a clean microcentrifuge tube and 50 µl elution buffer was added to the middle of the column and incubated at room temperature for

2 min. It was centrifuged for 2 min and eluate collected. The DNA concentration was measured on the Thermo Scientific Nanodrop 2000C Spectrophotometer.

2.2.2.3 Optimisation of stable transfected cell lines - geneticin concentration

To identify the optimal geneticin concentration for stable cell line generation, MCF7 cells were seeded at a density of ~0.3 $x10^6$ cells per well in 6-well plates. They were then exposed to geneticin (0-1500 mg/ml) and incubated for 5 days. Cells were fixed with 2% PFA and quantified with a crystal violet assay.

2.2.2.4 Transient transfection

MCF7 and HEK293S cells were plated into 96 well plates at a density of 0.4 $x10^5$ cells per well and left to grow for 24 hours in DMEM media at 37 °C, 5% CO_2. The DNA was diluted to 2 µg per 100 µl optimem media and added to FuGene HD transfection reagent, the volume recommended by the manufacturer's protocol was added to each well and incubated for 48-72 hours. After this time, the cells were checked using a fluorescent microscope to investigate transfection efficiency.

2.2.2.5 Effect of H_2O_2 on cytoglobin-transfected cells (wild-type and mutant)

2.2.2.5.1 H_2O_2 optimisation

HEK293S were incubated in DMEM containing no phenol or glutamine, since the phenol red increases the total antioxidant capacity of the cell culture media (Lewinska *et al.,* 2007), supplemented with FBS and gentamicin sulfate. The cultured cells were exposed to H_2O_2 (0-23 mM). Cells were incubated for four days in a humidified atmosphere with 5% CO_2 at 37 °C. After this time, the cells were fixed with crystal violet and quantified with a crystal violet assay (see section 2.2.2.7).

2.2.2.5.2 Effect of H_2O_2 on cytoglobin-transfected cells

After identifying the optimum H_2O_2 concentrations required to generate a dose curve; HEK293S cells transfected with Cygb (WT, H81A and L46W) had their media changed 24-48

hours post transfection to non-phenol/glutamine media and the cultured cells were exposed to H_2O_2 in a humidified atmosphere between 0-42 nM for four or seven days. Cells were fixed, and proliferation was quantified using a crystal violet assay (see section 2.2.2.7). Dose response curves were modelled using an adapted version of the following equations:

Equation 2.2.2.5.2.1:

	linear	log.
IC50	$y = \max - (max - min)\frac{IC50^n}{IC50^n + x^n}$	$y = min + \frac{max - min}{1 + 10^{n(\log x - \log IC50)}}$
EC50	$y = min + (max - min)\frac{EC50^n}{EC50^n + x^n}$	$y = min + \frac{max - min}{1 + 10^{n(\log EC50 - \log x)}}$

This equation (Nagy, 2020) allowed a curve that could be modelled to the data. Once the data was modelled to an equation, EC_{50} values were calculated.

2.2.2.6 Effect of chemotherapy drugs on cytoglobin-transfected cells (wild-type and L46W/H81A mutants)

MCF7 and HEK293S cells were transfected with DNA (WT CYGB, H81A and L46W pCMV) using FuGene HD transfection reagent (as described in section 2.2.2.4) and incubated for 72 hours. After this time, the cells were checked on a fluorescent microscope to confirm successful transfection. The cells were then exposed for seven days to a range of different concentrations of both docetaxel and paclitaxel: 0-1000 nM. Cell proliferation was quantified using a crystal violet assay (as described in section 2.2.2.7).

2.2.2.7 Crystal violet assay

Following the incubation with peroxides and chemotherapy agents for 4/7 days, cells were fixed with 2% paraformaldehyde (PFA) for 1 hour and then washed 3 times with H_2O. Once dried, cells were stained with 5.52% crystal violet for 1 hour. Wells were again washed 3 times with H_2O, dried and the crystal violet was resolubilised with 99% methanol (with shaking) for

1 hour. The plates were subsequently read on a Tecan Infinite M200 PRO plate reader (570 nm with a reference of 750 nm).

2.2.3 Examining the potential functions of cytoglobin through bioinformatics investigations

2.2.3.1 Protein modelling to examine the effect of different mutations in the cytoglobin structure

The WT protein structures obtained from RCSB protein data base (RCSB PDB, 2020) were mutated *in silico* in Swiss PDB viewer/ Deepview (Guex *et al.,* 2019).

The structural alignment of Cygb onto Mb, Hb1 and Ngb was created using the 'magic fit' function of Swiss PDB viewer/Deepview.

2.2.3.2 Globin sequence alignment and conservation

Using NCBI (NCBI, 2020), CYGB was searched and narrowed down by selecting only animals and restricting the number of residues to between 140 and 220 and downloaded in FASTA format. The fragments and partial sequences were removed, and the sequence files were aligned using Clustal X2. The positions of the key residues (38, 46, 60, 81, 83 and 113) were identified and frequencies of amino acid variants noted and displayed as pie charts.

WebLogo Berkeley plots were created to show the variations in amino acid sequences in surrounding residues. Approximately 20 residues with the residue of interest in the centre were selected and copied into the 'WebLogo Berkeley', a web-based application (Crooks *et al.,* 2004). Key residues were surrounded by a black box for easier viewing.

Clustal X2 were also used to create a phylogenetic tree. For this, only a sample of the total animal species studied were included as it made the display quality poorer. The order of animals selected was random but the selection criteria for the phylogenetic tree was simply those which were closest to human within the list (not closest in terms of homology).

2.2.3.3 Presence and effects of cytoglobin in cancer

2.2.3.3.1 Cytoglobin mutations in cancer: positions, frequencies, and cancer types

Data obtained from cBioPortal cancer genomics database- all studies in all different tissues were selected which came to a total of 116,641 samples and were queried by the CYGB gene. In total there were 84740 patients from 278 studies. The data showed different mutations associated with each cancer in amino acid order, removing duplicates. This data was also used to create a schematic to show the positions of the amino acids on the globin structure (created in PowerPoint). In addition to this, when queried by cancer type, the incidence frequency of CYGB mutations in different cancer patients was calculated and displayed graphically.

2.2.3.3.2 Cytoglobin staining in tissues in healthy patients versus cancer patients

Samples obtained from The Human Protein Atlas (Human Protein Atlas, 2020) of breast cancer tissues were examined. Three samples with either no staining detected, weak or medium staining present were used. Additionally, all images available on the database were individually examined and the staining presence and intensity was tabulated to calculate the percentage of healthy and cancerous patients with CYGB staining in both breast and prostate cancer.

2.2.3.3.3 Cytoglobin, Myoglobin and Neuroglobin RNA expression levels and distribution across various tissues

Data was gathered from GTExPortal software after selecting by the globin genes, the data was displayed by tissue type and listed alphabetically, and expression levels (transcripts per million) are colour coded with a key.

2.2.3.3.4 Survivability of cancer patients with high cytoglobin expression versus low expression

Kaplan Meier Plots were generated using the 'Kaplan Meier Plotter' (Nagy *et al.,* 2018). Data for mRNA gene chip CYGB was collected for breast, ovarian, lung and gastric cancer as well as the mRNA gene chip data collected for CYGB, MB, NGB and HB1 for breast cancer (using the multiple genes function). The four different plots for each investigation were compared against each other and assigned a hazard ratio value, CI interval and upper quartile survival value.

Chapter 3: Results- *In silico* analysis of CYGB with respect to the presence of two mutations in relation to cancer

3.1 Tissue distribution of cytoglobin, Neuroglobin and Myoglobin

CYGB is reported to be present in most human tissues, though like NGB, is expressed in very low levels, generally around the micromolar range (Fago *et al.,* 2004). This is potentially a contributing factor as to why the globin took until 2001 to be discovered, compared to those proteins with higher expression levels. This investigation looked at the presence of CYGB within different tissues, as well as the level of RNA expression. This was analysed using GTExPortal software (Broad Institute, 2020).

The expression of CYGB, NGB and MB was compared across a range of human tissues, per million RNA molecules in the RNA-sequencing sample (Figure 3.1). For NGB, out of the 52 different tissue categories, the vast majority had zero transcripts for every million RNA molecules in the RNA-sequencing sample including adipose and the heart. The highest expressing NGB tissues were the brain cortex and hypothalamus. MB only had 10 different tissues with zero transcripts, compared to NGB's 37. However, MB had substantially more tissues with a mid-range expression with only one highly expressive tissue: the minor salivary gland. Unlike the other two globins, MB has three very highly expressed tissues with $>8.5x10^3$

per million RNA molecules, these are the left ventricle, atrial appendage of the heart and the skeletal muscle. CYGB has the fewest tissues with no expression and these tissues were the EBV-transformed lymphocytic cells and whole blood.

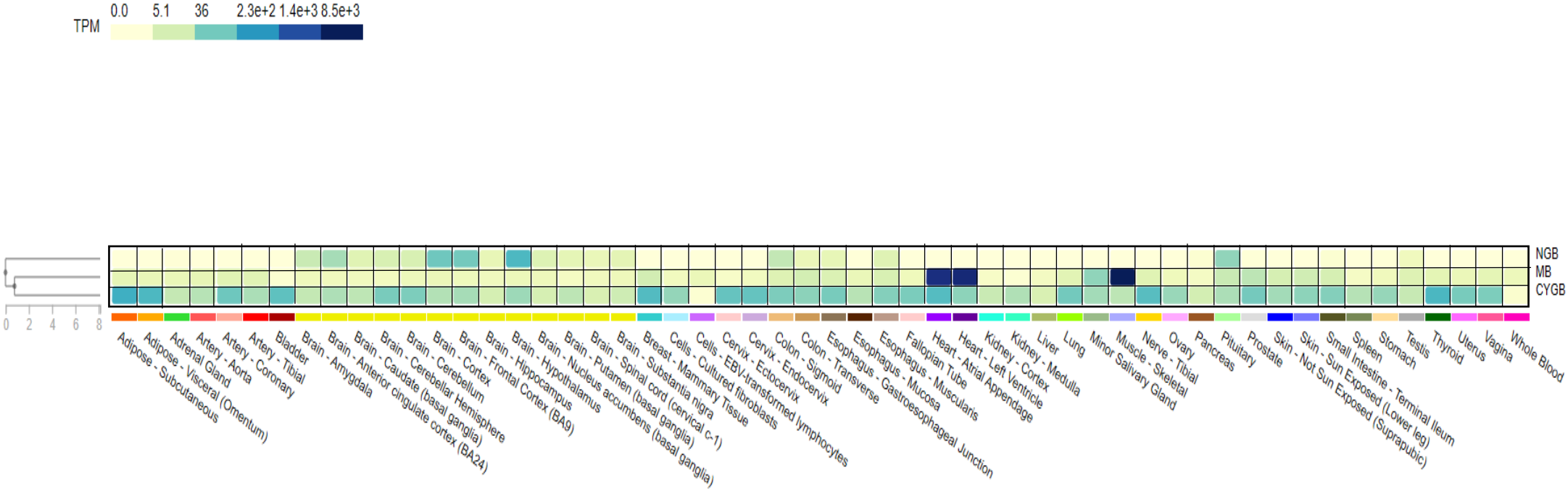


Figure 3.1.1- Lower levels of RNA expression in NGB compared with CYGB and MB in human tissue. Normalised method for RNA-sequencing data for CYGB, NGB and MB genes in different human tissues (listed alphabetically) using GTExPortal software. Expression values in TPM (Transcripts per million).

3.2 Conservation and alignment of key residues in the cytoglobin structure

The widespread distribution of CYGB expression in most tissues (albeit low-level expression) indicates the importance of the gene. Another method to determine functional importance is to see how conserved the protein is in other species.

In order to investigate this, six key amino acid residues which are considered to be important in the structure, or of interest in research, were selected for comparison for the reasons displayed in Table 3.2.1.

Table 3.2.1- A consolidated list of the locations and descriptions of key Cygb amino acids. Key amino acid locations (human) obtained from UniProtKB (UniProt, 2020).

Key Amino Acid	Description
C38	Cysteine that forms disulfide bonds. (UniProt, 2020)
L46	Mutants affect NOD activity in other globins e.g., Mb and Hb (Tiso *et al.,* 2011)
F60	Highly conserved amino acid in all globins; stabilises binding of haem to protein (Geuens *et al.,* 2010)
H81	Distal haem ligand; mutants known to affect nitrite reductase in other globins e.g., Mb and Ngb (Yi *et al.,* 2009)
C83	Cysteine that forms disulfide bonds. (UniProt, 2020)
H113	Proximal haem ligand. (UniProt, 2020)

379 animal Cygb sequences were extracted from the NCBI database (NCBI, 2020) and aligned using Clustal X (Larkin *et al.,* 2007). This total number includes those that are theoretical, fragments and isoforms that were consequently removed.

Two mutants were studied: H81A (distal histidine) and L46W, in addition to H113 (proximal histidine), two cysteine residues which form intermolecular disulfide bonds: C38 and C83 *in vitro* (*in vivo* state is still unknown) and F60 which is believed to stabilise the binding of haem to protein.

As shown in Figure 3.2.1, three of the amino acids: F60, H81 and H113 showed 100% conservation across all species. The most diverse was C83 which had 10 different possible amino acids, the most common being cysteine and arginine. C38 and L46 also have a very

high percentage of cysteine and leucine respectively which indicates a high level of conservation across species. This low-level change could be the result of a misalignment.

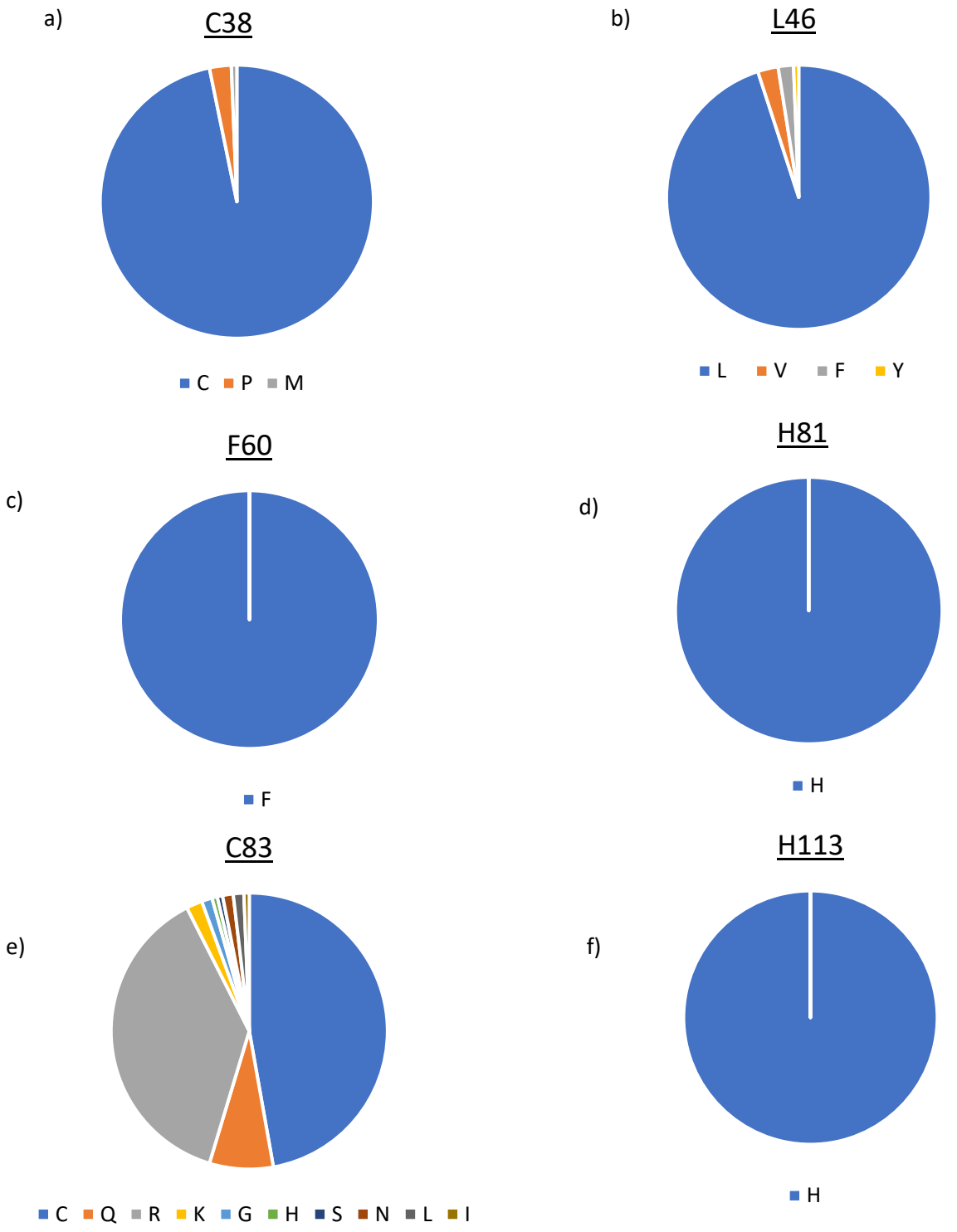


Figure 3.2.1- High animal sequence homology in key Cygb amino acids with the exception of C83 which is responsible for the intramolecular disulfide bond. Frequency levels of different base substitutions in animal Cygb sequences showing substitutions at positions a) 38, b) 46, c) 60, d) 81, e) 83 and f) 113. Sequences obtained from NCBI and analysed using Clustal X (Version 2.1).

Figure 3.2.1 shows the amino acid conservations for the key amino acids (highlighted with a black box) as well as the surrounding amino acids, in the form of a WebLogo Berkeley plot (Crooks *et al.,* 2004). Those with a single letter for each position have a conservation of 100%. For those with multiple different amino acids, the size of the letters portrayed indicates the frequency of that amino acid's presence.

The H81 amino acid is 100% conserved as stated earlier, additionally the entire region is likewise highly conserved with the four amino acids that precede it also having 100% conservation. L46 does not have 100% conservation and instead consists of a few possible variations; the same being true for the surrounding amino acids which also show variability. This is also seen in C38 which had a few variations though cysteine was the dominant amino acid present. The amino acids either side also showed a single more dominant variant accompanied by a few less frequent amino acids. C83 is very different to its counterpart (C38) in that it, and the region it is in, is highly variable with a range of different amino acids present at that particular location across the animal kingdom. H113 and F60 are very similar in that they both have 100% conservation, and surrounding amino acids are a combination of both 100% conserved amino acids and some where two variations were seen.

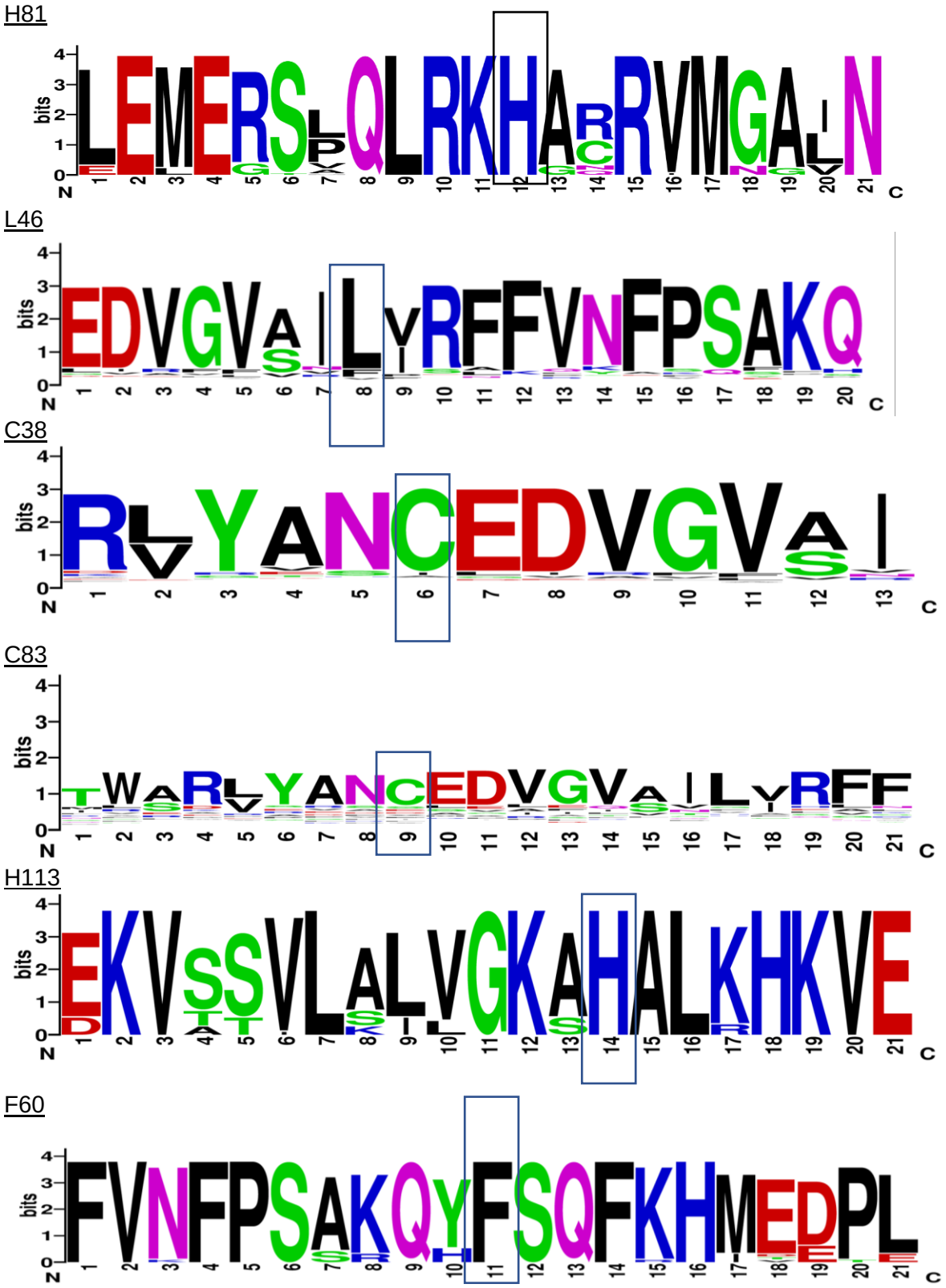


Figure 3.2.2- Key amino acids with high animal sequence homology are more likely to have increased homology in neighbouring residues. The images show the frequency levels of six different amino acids in both human and other animal species. Created using WebLogo Berkeley. Blue: histidine, lysine and arginine; red: glutamic acid and aspartic acid; black: leucine, methionine, valine and phenylalanine; green: serine, glycine and tyrosine; pink: glutamine and asparagine.

Though different parts of the Cygb protein varied in terms of the conservation percentages, the relatedness of the animals' Cygb proteins were also investigated in Figure 3.2.3 in order to observe which species Cygb were the most similar.

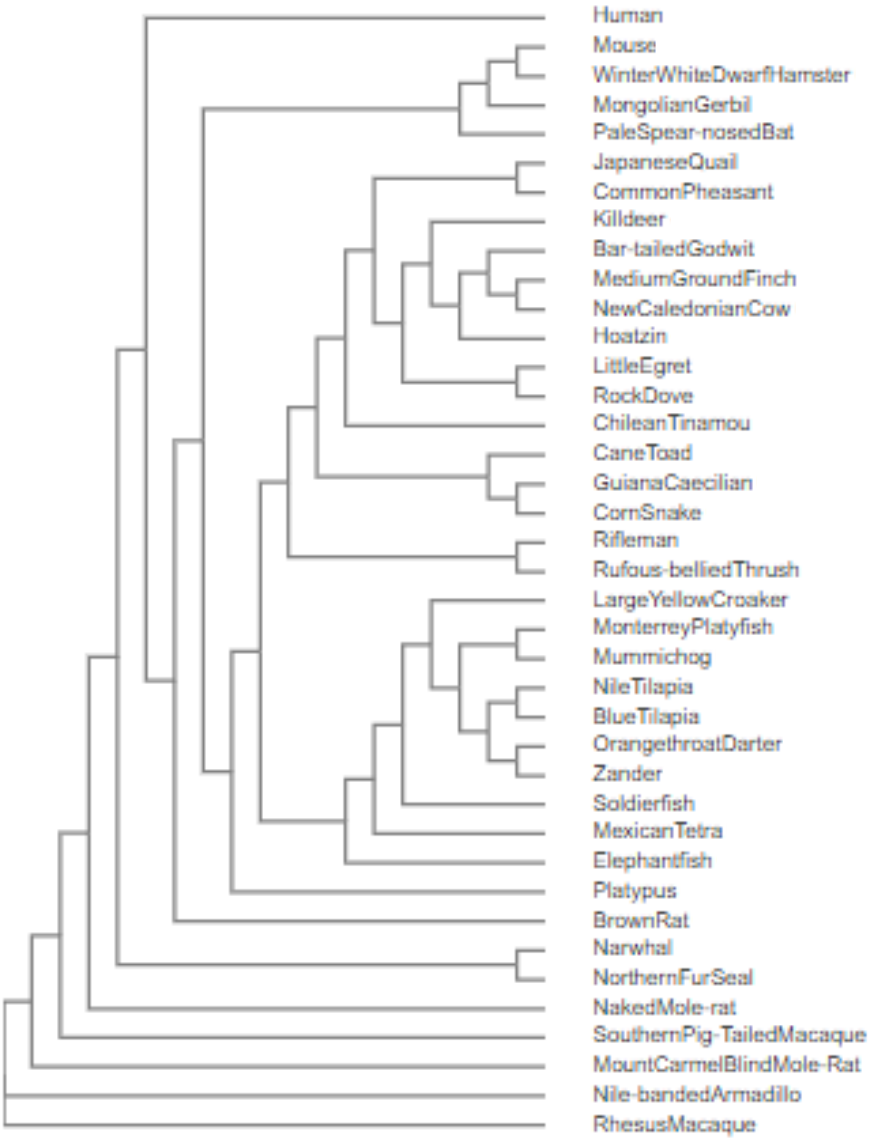


Figure 3.2.3- Closest animal relations have the most similar Cygb sequence homology. Phylogenetic tree showing the relatedness of several animals' Cygb, created on ClustalW2 using neighbour-joining clustering method. Sequences obtained from NCBI and aligned using Clustal Omega.

The phylogenetic tree in Figure 3.2.3 represents the relatedness of the Cygb protein present in various species. The branch points indicate a 'divergence event' whereby through a variety of mechanisms the Cygb protein became altered or changed. At this branch point lies the most recent common ancestor. For instance, the little egret and rock dove shared a last common ancestor with the same Cygb protein, though these two have comparably different proteins themselves now. The number of branch points between species suggests how related the Cygb protein is between them, and this often correlated to the relatedness of the species themselves. For example, the Japanese quail and the common pheasant are only one branch point apart from each other, whereas the Japanese Quail and the Elephant fish are seven branch points apart. With both the quail and pheasant being fowl native to Great Britain, they are not too distantly related in terms of species and this is reflected in Figure 3.2.3 by how closely related they are in terms of their Cygb.

Another example is the brown rat and humans which are only 2 branch points apart, which makes the rat such a viable option for drug testing prior to human trials as their physiology does not differ too much from of a human.

3.3 Effect of point mutations on the 3D structure of the cytoglobin protein

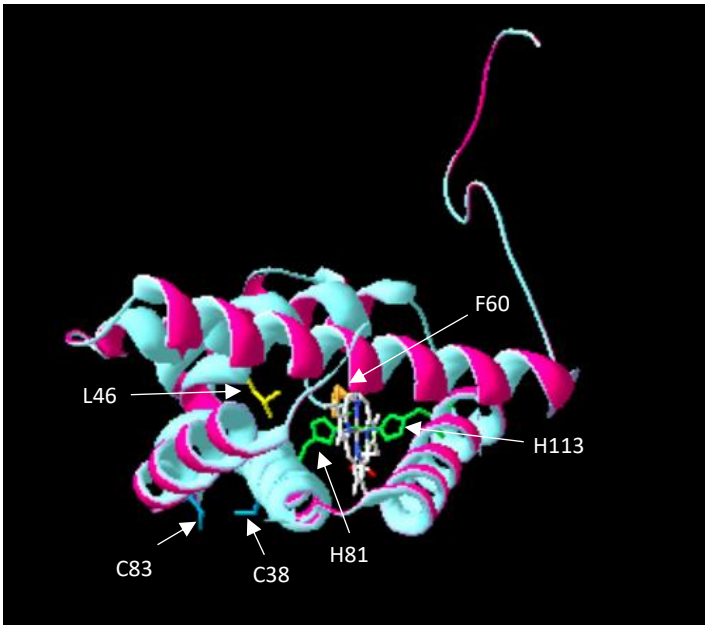
The high level of conservation in five out of the six key residues investigated, shows the importance of these in the structure and potentially the function. The effect of the mutations on the quaternary structure was consequently examined. To observe potential disruptions to the protein structure; the mutations, in addition to some key amino acids, were compared to the WT type using Swiss-pdb viewer (version 4.0), with the wild type structures obtained from the RSCB protein data base (Berman *et al.,* 2000).

Figure 3.3.1 shows there is a large difference in size between the amino acids at positions 46 and 81 in the WT and in their respective mutant varieties. There is a difference in the space between the haem and the distal histidine. The H81A mutation shows the alanine methyl side chain is much further away from the haem iron, suggesting that this mutant would have no

distal ligation to the haem iron (i.e., pentacoordinate). The L46W mutant shows that the tryptophan, now present in the structure, occupies much of the space between the haem and the back of the haem pocket due to the presence of the indole ring.

Additionally, figure 3.3.2 shows the backbone of the WT Cygb which is fitted to Ngb, Hb and Mb with the mutations shown to see the effect on functions *in vitro* for these other globins mapped onto Cygb *in silico.* In each case the Cygb protein is shown in red. The mapping of Cygb and Ngb were the most different in terms of the key amino acids and Cygb and Mb seemed the most similar.

a)



b)

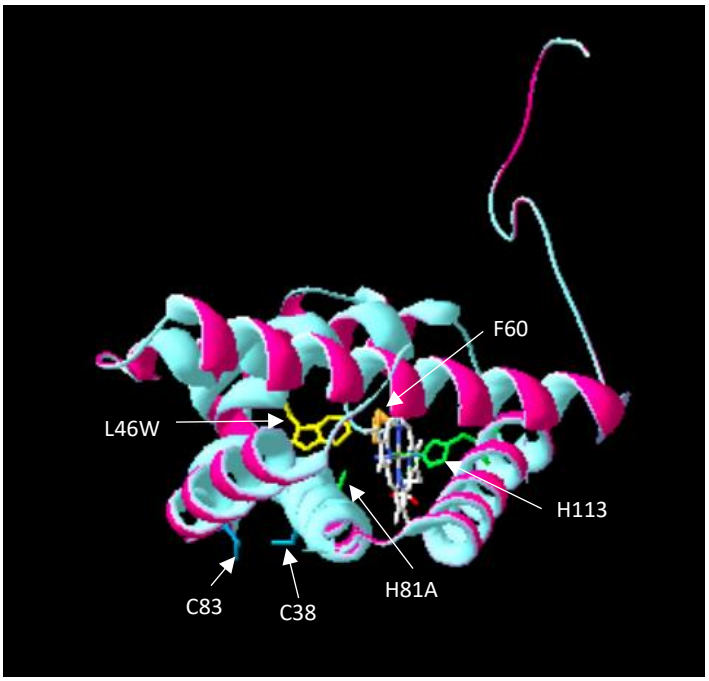
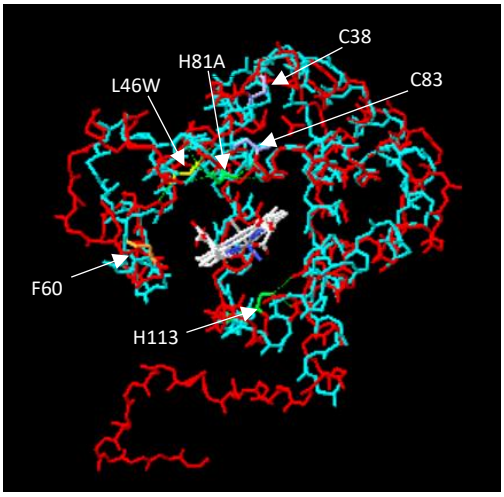
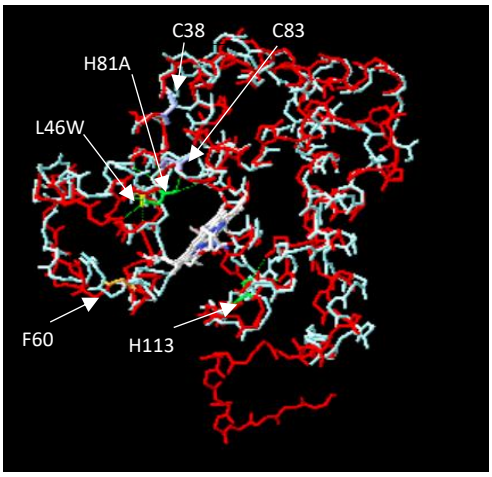


Figure 3.3.1- Mutation of two key amino acids in the primary structure changes the stability of the haem domain. a) WT Cygb highlighting key features and amino acids of the protein. The haem is shown in a combination of white, blue and red, the two histidines (including H81) in green, the cysteines in blue, the L46 in yellow and the F60 orange. b) Cygb highlighting key features and amino acids of the protein including the mutants. The haem is shown in a combination of white, blue and red, the proximal histidine and the mutated distal histidine (now an alanine) in green, the cysteines in blue, the L46W in yellow and the F60 in orange. Crystal structures obtained from NCBI (Cygb 2dc3, Hb alpha 1IRD, Mb 3RGK and Ngb 6I3T).

a)



b)



c)

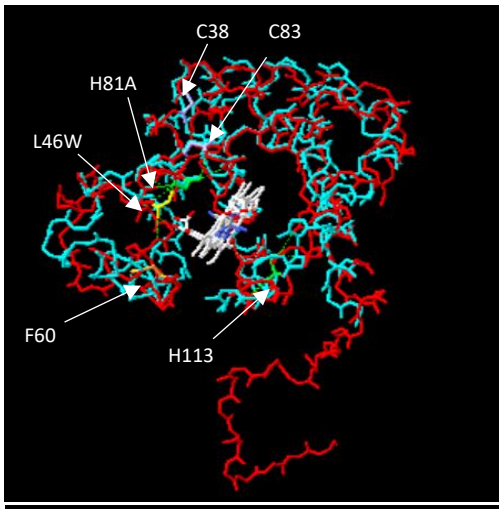


Figure 3.3.2- Mapping of mutated Cygb onto human Hb alpha, Ngb and Mb shows structural similarities, particularly in Mb. a) Cygb (2dc3) and human Hb alpha (4hhb), b) Cygb and Mb (1mbn), c) Cygb and Ngb (4mpm). In all cases Cygb is displayed in red and the contrasting globin in blue. The Cygb haem is shown in a combination of white, red and blue, the two histidines (including H81A) in green, the cysteines in dark blue, the L46W in yellow and the F60 in orange. Crystal structures obtained from NCBI (Cygb 2dc3, Hb alpha 1IRD, Mb 3RGK and Ngb 6I3T).

3.4 Cytoglobin and cancer

3.4.1 Naturally occurring cytoglobin mutations in cancer

As these mutations in other globins are known to affect key functional properties, it is hypothesised that these will have a significant effect on the Cygb protein. Should this mutation affect an area responsible for cell proliferation or control of the cell cycle then it is possible that these mutations are found naturally in diseases such as cancer. The different mutations in CYGB found in various cancers are shown in Table 3.4.1.

From the known variants, only one out of the six key residues previously investigated is found with a natural mutation in cancer: C38Y. This was the cysteine with the higher level of homology and conservation, compared to C83 (see Figure 3.2). The majority of the CYGB mutations within cancer are missense and the copy number diploid.

In order to identify whether there are areas where the mutations are more common or 'hotspots', the mutations were translocated onto the globin primary structure. The data in Figure 3.4.1 shows that the mutations occur at an array of different areas within the amino acid chain, starting at amino acid 11 and finishing at 190 (190 being the last amino acid in the sequence). There are no common clusters present.

Table 3.4.1.1- The cancerous mutations associated with the CYGB gene. Data collated from 84740 patients/cases across 278 different studies (cBioPortal, 2020)

Cancer Type	Protein Change	Mutation Type	Copy
Mixed Cancer Types	E1K	Missense	/
Bladder Urothelial Carcinoma	E11Q	Missense	Diploid
Lung Adenocarcinoma	R13S	Missense	/
Lung Squamous Cell Carcinoma	R13S	Missense	/
Cutaneous Melanoma	S16Rfs*32	FS del	Diploid
Cervical Squamous Cell Carcinoma	S20A	Missense	Diploid
Colorectal Adenocarcinoma	A22T	Missense	Diploid
Mixed Cancer Types	A22T	Missense	Diploid
Breast Invasive Ductal Carcinoma	E23D	Missense	/
Lung Squamous Cell Carcinoma	R24S	Missense	Diploid
Colon Adenocarcinoma	V27G	Missense	/
Mixed Cancer Types	M30t	Missense	Diploid
Bladder Urothelial Carcinoma	R33W	Missense	Gain
Mixed Cancer Types	C38Y	Missense	Diploid
Lung Adenocarcinoma	G42V	Missense	Diploid
Colorectal Adenocarcinoma	V43Wfs*5	FS del	/
Cutaneous Melanoma	X48_splice	Splice	/
Mixed Cancer Types	P54S	Missense	Diploid
Small Cell Lung Cancer	S61I	Missense	/
Mixed Cancer Types	M72I	Missense	Diploid
Mixed Cancer Types	R74Q	Missense	Diploid
High-Grade Serous Ovarian Cancer	S75R	Missense	ShallowDel
Head and Neck Squamous Cell Carcinoma	Q77	Nonsense	Diploid
Rectal Adenocarcinoma	R79W	Missense	Diploid
Bladder Urothelial Carcinoma	A82T	Missense	Amp
Cervical Squamous Cell Carcinoma	A82T	Missense	Diploid
Mixed Cancer Types	R84*	Nonsense	Diploid
Cutaneous Melanoma	G87V	Missense	Diploid
Mucinous Adenocarcinoma	V93M	Missense	Diploid
Glioblastoma Multiforme	E94K	Missense	Diploid
Cutaneous Melanoma	P99L	Missense	/
Bladder Urothelial Carcinoma	D100N	Missense	Diploid
Prostate Adenocarcinoma	S104F	Missense	Amp
Renal Clear Cell Carcinoma	S104T	Missense	Diploid
Oesophageal Carcinoma	A114T	Missense	Diploid
Hepatocellular Carcinoma	L115P	Missense	Gain
Colorectal Adenocarcinoma	K116E	Missense	/
Diffuse Type Stomach Adenocarcinoma	K116T	Missense	Diploid
Prostate Adenocarcinoma	V119M	Missense	Diploid
Serous Ovarian Cancer	Y123C	Missense	Gain

Skin Cancer, Non-Melanoma	X126_Splice	Splice	/
Mixed Cancer Types	S128C	Missense	/
Stomach Adenocarcinoma	A136T	Missense	Diploid
Acute Myeloid Leukaemias	Q138Afs*135	FS ins	/
Bladder Urothelial Cancer	E146K	Missense	Gain
Early T-Cell Precursor Lymphoblastoma	T147M	Missense	Diploid
Colon Adenocarcinoma	R155C	Missense	Gain
Bladder Urothelial Carcinoma	S160R	Missense	/
Lung Squamous Cell Carcinoma	V162M	Missense	Gain
Mixed Cancer Types	A178T	Missense	Diploid
Germinal Centre B-Cell Type	T179I	Missense	/
Lung Squamous Cell Carcinoma	X180_splice	Splice	Diploid
Adenoid Cystic Carcinoma	P182L	Missense	/
Mucinous Adenocarcinoma	P182L	Missense	Diploid
Oesophageal Adenocarcinoma	S188L	Missense	Diploid
Cutaneous Squamous Cell Carcinoma	G189R	Missense	/
Small Cell Lung Cancer	P190L	Missense	/

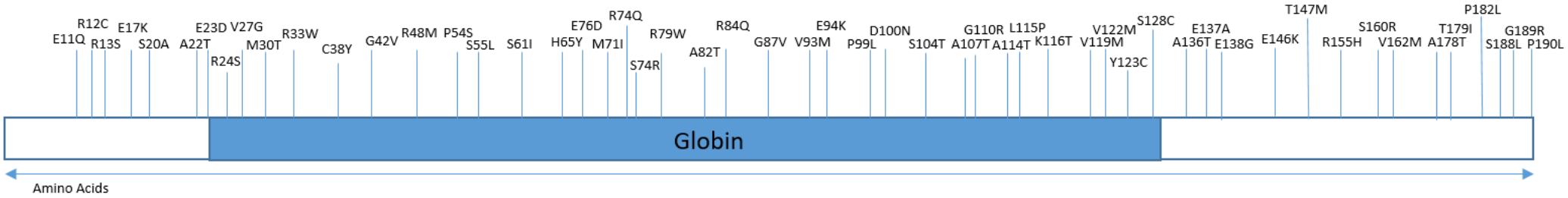


Figure 3.4.1.1- Globin mutations span entire length of genome and show a lack of mutation clusters. The positions of different mutations on the Cygb primary amino acid sequence obtained from cBioPortal, 2020 and displayed using Microsoft Powerpoint.

Despite having no clear pattern as to the distribution of mutations in the structure, there were some cancers where mutations of CYGB were more frequent. These frequencies of CYGB mutations within different cancers are shown in Figure 3.4.1.2.

The two most common types of cancer associated with CYGB mutations are nerve sheath tumours and cancers of unknown primary origin, which had incidence frequencies of 20% and approximately 15%, respectively.

The frequency of mutations in the other cancers typically ranged from 5% and under.

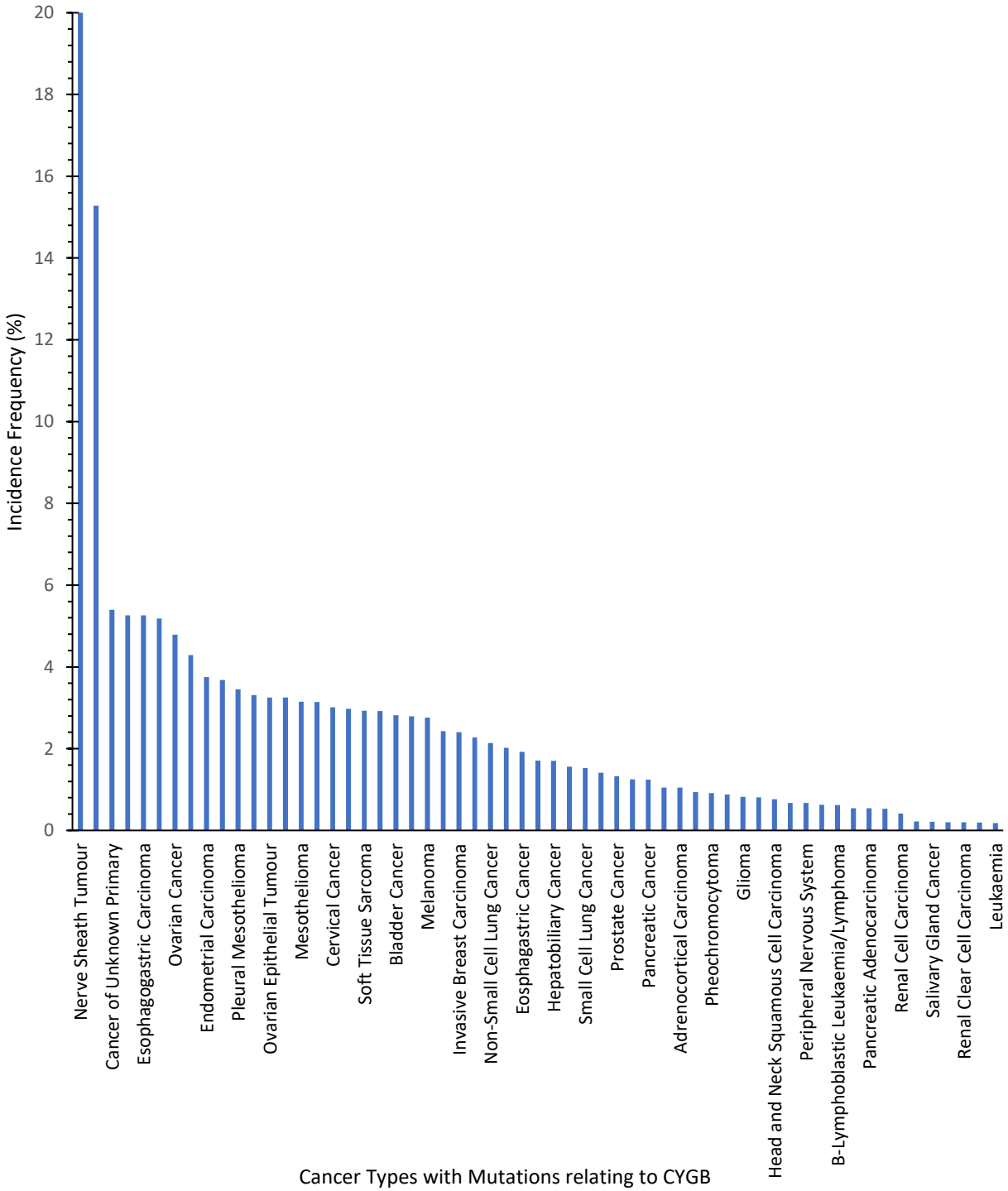


Figure 3.4.1.2- Elevated frequency of CYGB mutations in certain cancers including Nerve Sheath Tumours. The incidence frequencies of different cancer types with mutations relating to CYGB. Data obtained from cBioPortal (Version 3.4.17, 2020).

3.4.2 Cytoglobin tissue staining as a measure of cancer presence

It is evident that CYGB is implicated in a variety of different cancers (to varying degrees) however this investigation focussed on prostate and breast cancer due to how common these two cancers are.

In addition to analysing the mutations of CYGB in cancers, CYGB expression in both healthy tissues and tumours (immunohistochemistry images from The Human Protein Atlas) was observed to see whether alterations in CYGB levels could be used diagnostically.

Figure 3.4.2.1 shows the staining intensities (none detected, weak and medium) in three different breast cancer patients as well as healthy breast tissue. In prostate cancer, all healthy patients had no CYGB staining visible in their tissues (Figure 3.4.2.2). All cancer patients had some staining, with an equal division between low and medium staining. Breast cancer also showed that all healthy patients had no CYGB staining in their tissues. Unlike prostate cancer, some breast cancer patients had no staining, and this made up a larger number than the medium staining in breast cancer patients. Nevertheless, the majority of breast cancer patients had some staining, either in low or medium intensity.

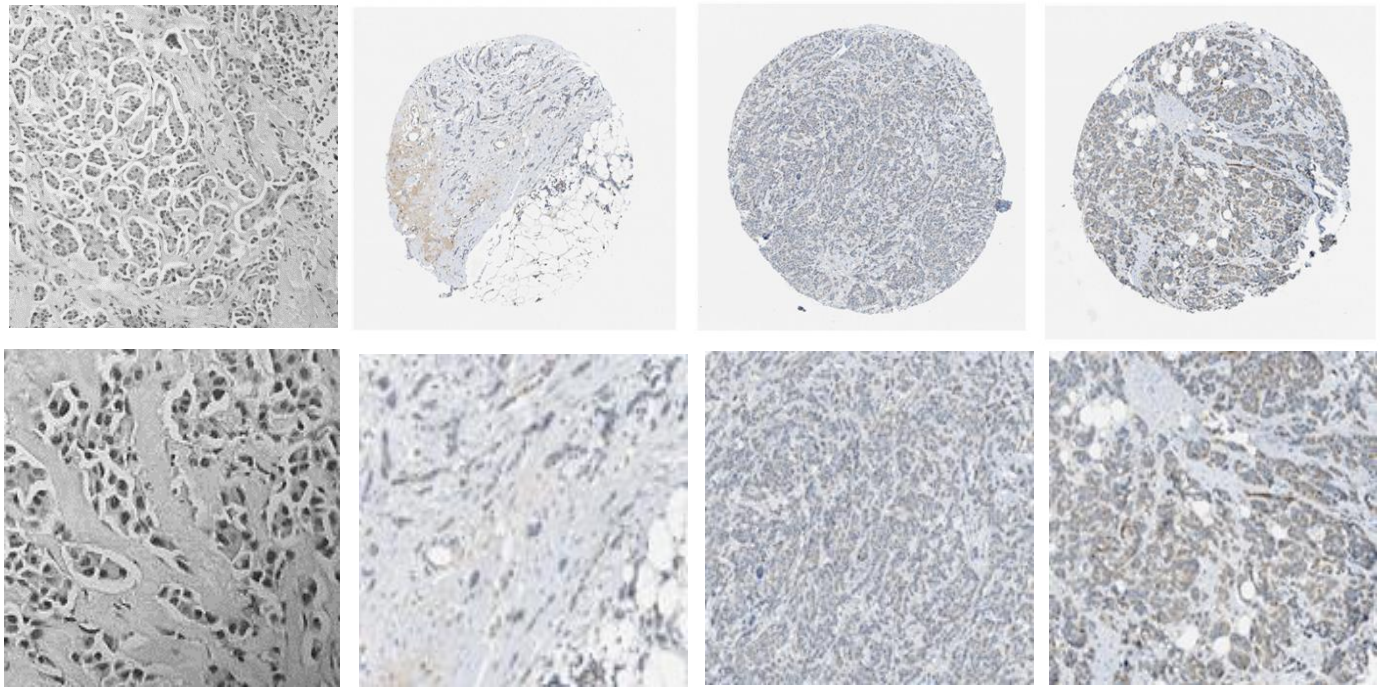


Figure 3.4.2.1- CYGB staining in different breast cancer patient tissues compared to healthy breast tissue. The images show the CYGB staining intensity in three breast cancer patients as well as a healthy breast tissue sample, from left to right: healthy tissue, no staining detected, weak and medium staining (images underneath are magnified versions of those above). Data obtained from The Protein Atlas (2020) and John Hopkins Medicine Pathology (2020).

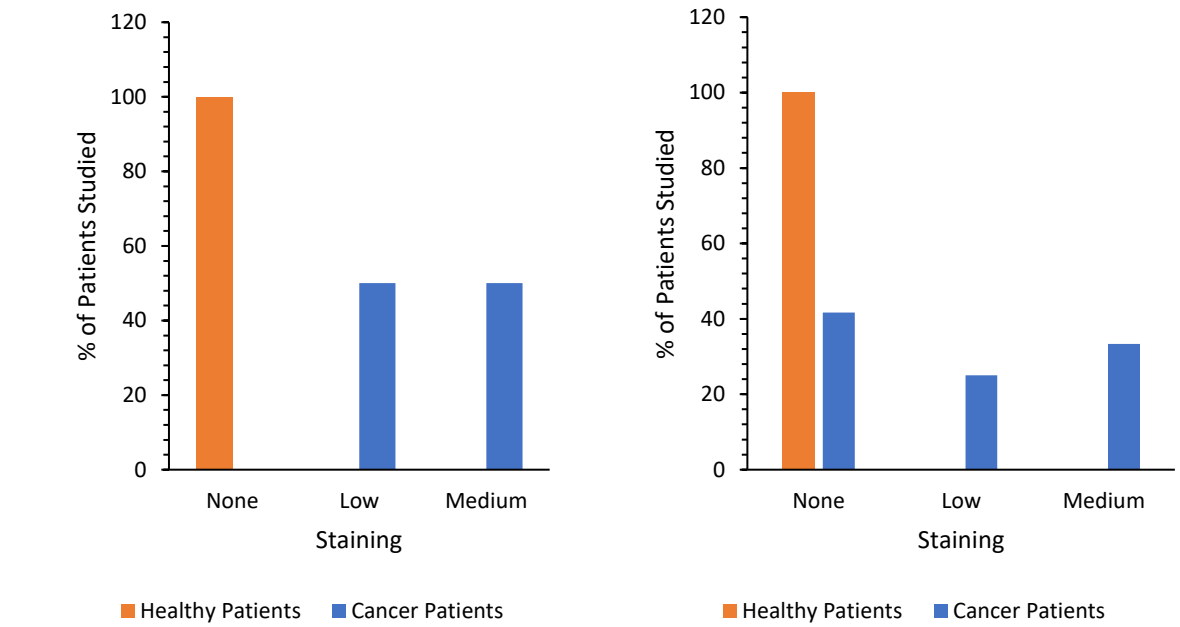


Figure 3.4.2.2- Absence of CYGB staining in healthy patients but presence in prostate and breast cancer patients. The presence of CYGB and the intensity of staining in a) prostate cancer and b) breast cancer in human tissues. Total number of patients in the prostate and breast cancer studies are 15 and 14 respectively; data obtained from The Human Protein Atlas (2020).

3.4.3 Cancer patient survivability depending on the level of cytoglobin expression

With the knowledge that CYGB expression appears to be increased in breast and prostate cancer, it was next investigated whether CYGB expression levels correlate with patient survival. Using microarray data, four different cancers (breast, ovarian, lung and gastric) were analysed and the survival probabilities for high and low expression compared over 150 months (Figure 3.4.3.1).

In breast cancer, patients with low CYGB expression had a lower survival probability than patients with high CYGB expression. The hazard ratio (HR) is 0.85 and so the high CYGB expression group is 15% less likely to die than those with a low CYGB expression and the study has 95% confidence that the true value lies between the confidence interval (CI) of 0.72-0.99 so those with a high expression level are 1-28% less likely to die than those with low expression. The *P-value* shows that this is a significant difference. In comparison, ovarian cancer patients with a low CYGB expression began with a higher survival probability than those with a high CYGB expression; at a time of roughly 100 months these appear to plateau out to a survival of ~0.15 for both expression types. Overall, there was no significant difference. The high and low CYGB cohorts within lung cancer mirror each other very closely until approximately 100 months. After this time, the high expression levels begin to have higher survival than their counterparts with the exception of the dramatic dip at the end which does not have any corresponding data points in the low expression. The HR is 0.96 which means the high CYGB expression group has a 4% higher survival though the CI is 0.81-1.13 which is significant. Finally, although the expression levels of both gastric cancer groups begin equal, very quickly the survival probability for the high expression group drops far below the low expression group by approximately 30%. The HR is 1.94 which means that the survivability for high expression patients is almost half as good as those with low CYGB expression. Gastric cancer had a remarkable *P-value* of 2.3×10^{-7} which shows a very significant difference in the patient survivability.

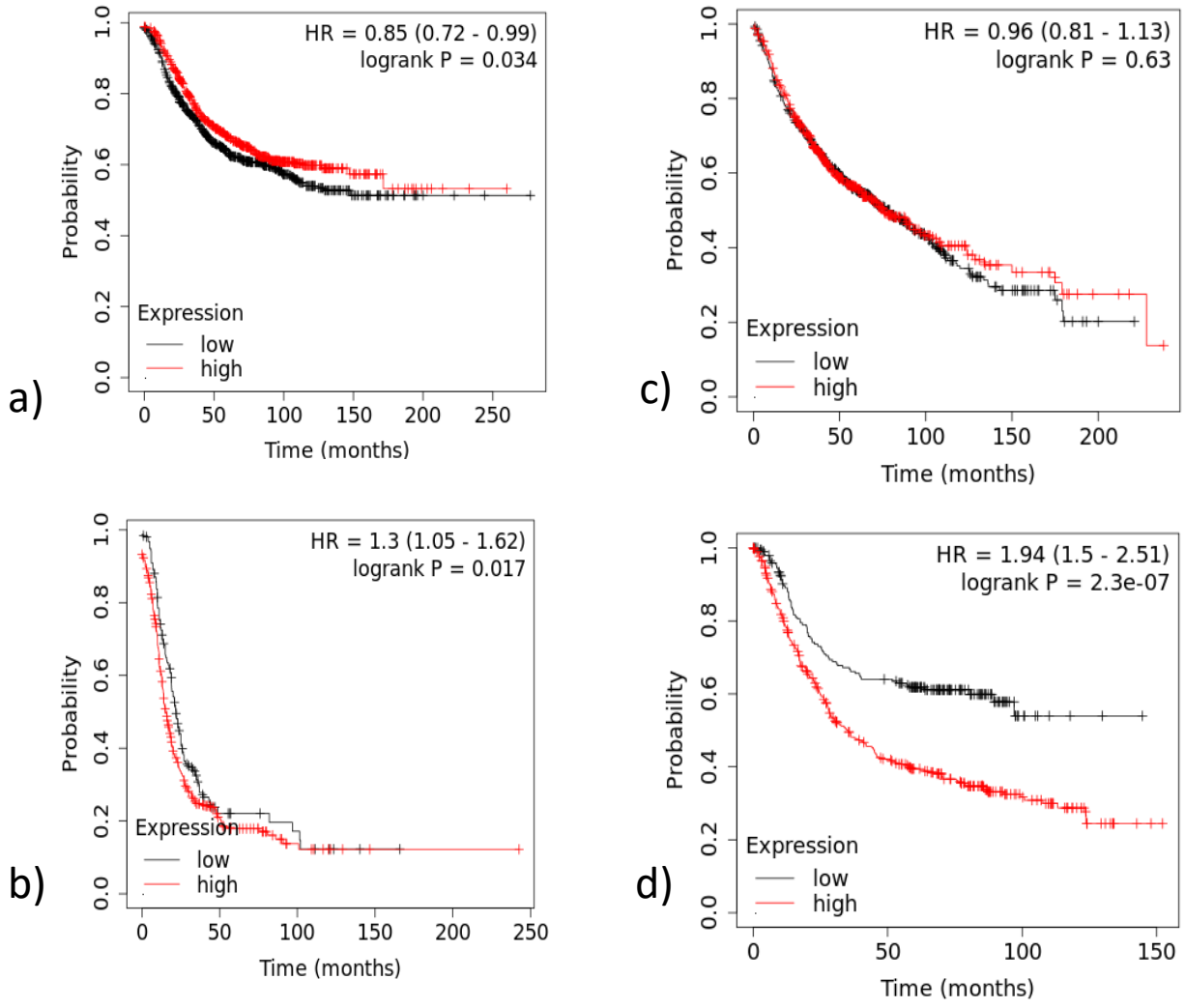


Figure 3.4.3.1- High CYGB expression results in higher survivability in cancers such as breast and lung but drastically lower survivability in those with gastric cancer. Kaplan Meier plots using mRNA gene chip data for the following cancers: a) breast cancer, b) ovarian cancer, c) lung cancer and d) gastric cancer. The patient sizes for each study were 1764, 614, 1144 and 613, respectively. The data was generated using the Kaplan Meier Plotter (Nagy *et al.,* 2018). Hazard ratio has 95% confidence intervals to compare the high and low CYGB expression groups.

In addition to investigating patient survival with high or low CYGB expression in different cancers; the survival of patients with high or low expression in breast cancer for four different globins was investigated to see whether this was something unique to CYGB or a trend across multiple globin proteins. The globins compared against CYGB were NGB, HB1 and MB.

The four Kaplan Meier plots show the correlation between high or low expression of four different globins on the survivability of breast cancer patients. CYGB and HB both have a significant difference in their patient survivability whereby a higher expression was associated with higher survival. MB has a better survival for low expression patients up until around 200 months after which the survival rate of the low expression group drops dramatically. Due to this inconsistency, the overall data is not significant when looking at the *P-value*.

As mentioned above, the hazard ratio for CYGB is 0.85 and so the high CYGB expression group is 15% less likely to die than those with a low CYGB expression. Conversely, MB has a HR value of 1.08 which means that there is almost an equal survival rate between the high and low MB expression groups and there is no significant difference. Similarly, NGB is also close to the value of 1 at 0.91 meaning there is also no significant difference in the survival of patients with high and low NGB expressions. The CI borders both sides but is heavily biased towards the high CYGB expression group surviving better than the low expression group. The HB1 group CI is 0.64-0.8 which means that the high expression group has the higher survival rate when compared to the low expression group, this data had a *P-value* of 3.1×10^{-9}.

When looking at similar studies on Kaplan Meier Plotter, liver cancer also showed that for three out of four globins (CYGB, MB and NGB), the high expression group had the better survival rates when compared to the low expression group with NGB and MB having a marked difference between the two (56% and 42% respectively). Pancreatic cancer also showed that three out of the four globins (MB, NGB and HB1) correlated with longer patient survival in higher expression groups than low expression groups.

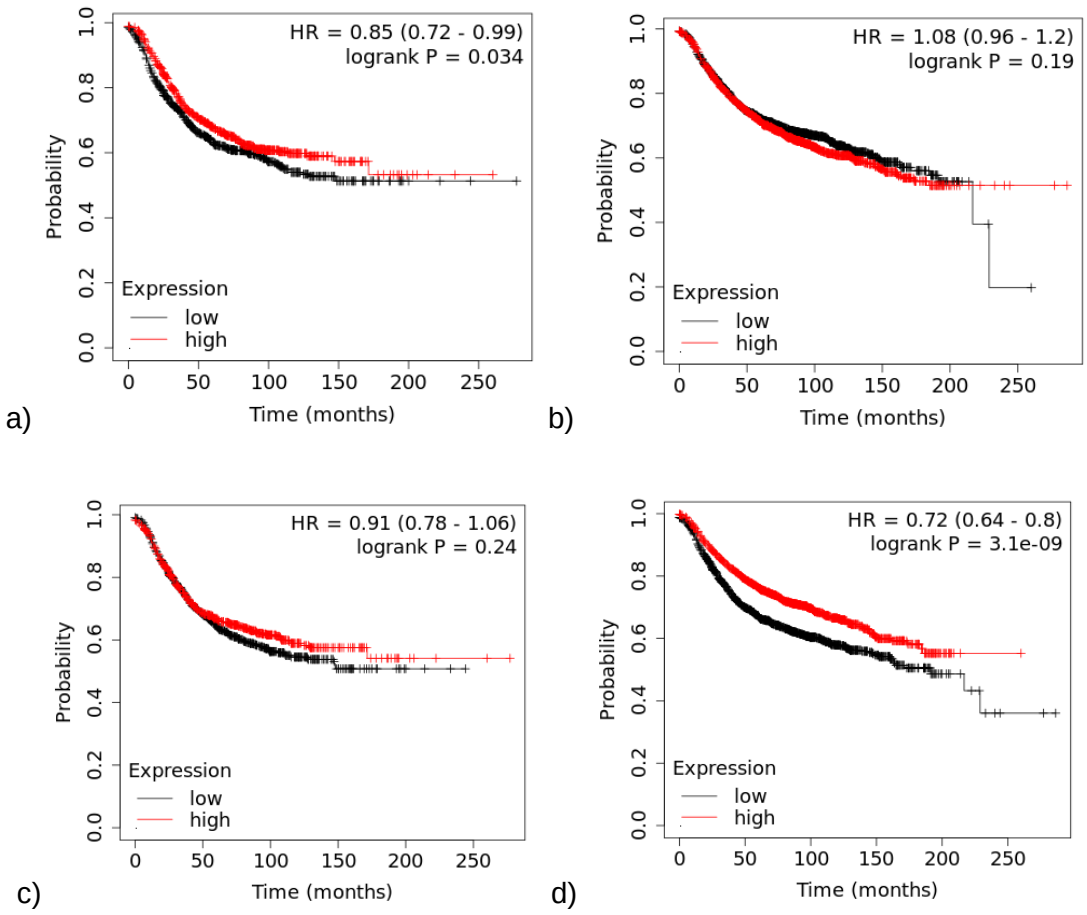


Figure 3.4.3.2- Survivability increased with higher expression in breast cancer for different globin proteins. Kaplan Meier plots using mRNA gene chip data for breast cancer in the following genes: a) CYGB, b) MB, c) NGB and d) HB1. The data was generated using the Kaplan Meier Plotter (Nagy *et al.,* 2018). Hazard ratio has 95% confidence intervals to compare the high and low expression groups for each globin.

Chapter 4: Results- Investigating the effects of different stress-inducing environments on CYGB-transfected cells.

4.1 Confirmation of mutations present through DNA sequencing

By utilising the power of bioinformatics, a picture has emerged which shows a potential relationship between CYGB and cancer.

The study will investigate the wild-type protein and two mutants known to affect the protein's functions of nitrite reductase and nitric oxide dioxygenase in other globins such as Ngb and Mb.

To confirm the presence of the mutations in both the pET and pCMV vectors, the DNA samples were sequenced by Eurofins.

The chromatograms confirm the presence of the two mutants compared to the WT (Figure 4.1) and the sequencing also displays that there are no other additional mutants present which could cause an effect on the protein and affect the results gathered (Appendix 1). Due to poor sequencing past certain points the data sets are not entirely complete but sufficient to show mutations; there were no frame shifts or stop codons present. All differences from the WT will be purely down to the individual mutations only.

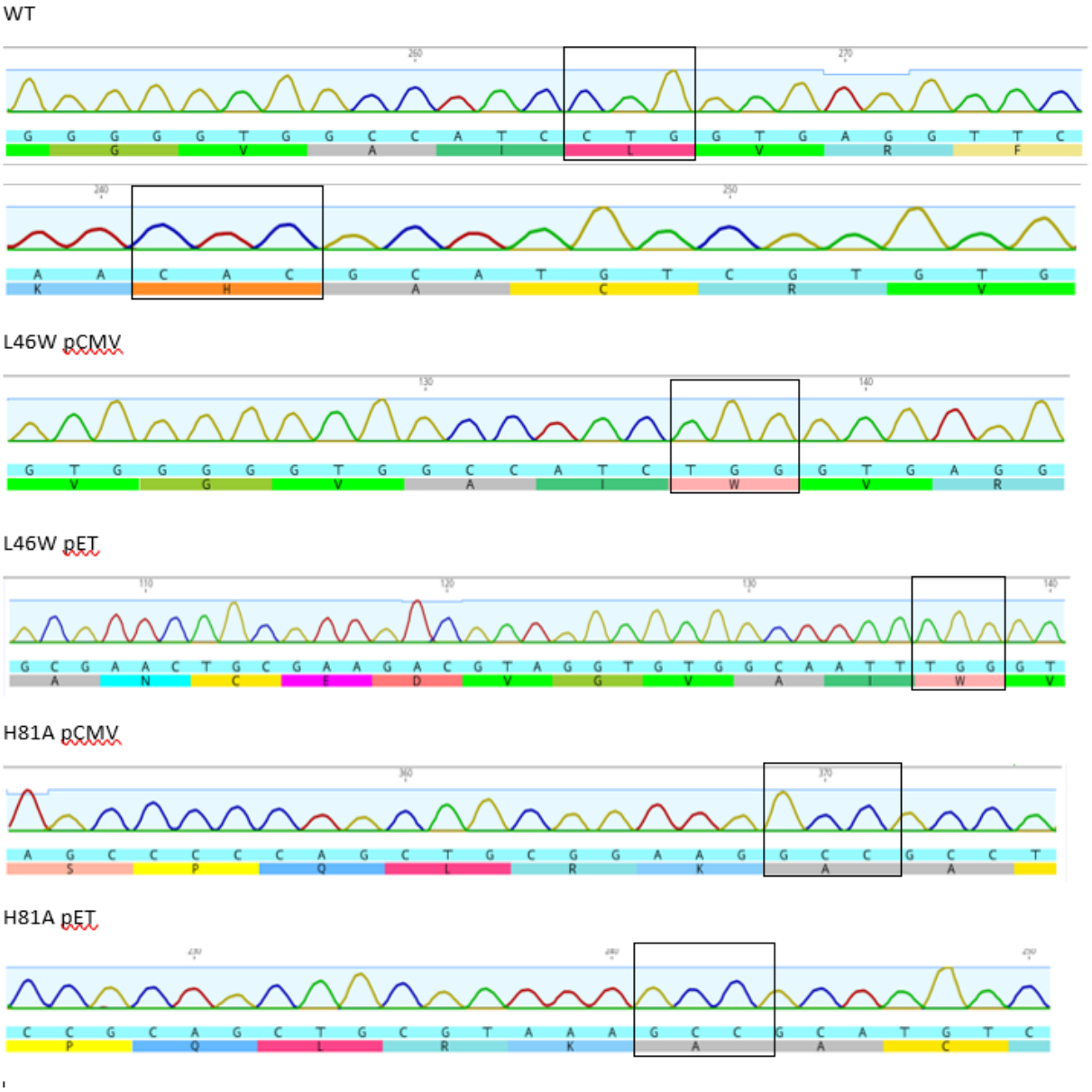


Figure 4.1.1- Chromatogram confirmation of CYGB mutations in both pCMV and pET vectors. DNA sequence alignments showing the presence of the two singular mutations H81A and L46W in both the pCMV and pET vectors. Sequenced by Eurofins, data analysed using Geneious Biologics (2021).

4.2 Gel electrophoresis to confirm correctly linearised plasmids

The three DNA samples (WT, H81A and L46W) were singly digested with the *Stu1* restriction enzyme to linearise them prior to transfection. A small sample was run on an electrophoresis gel against their respective controls to ensure the samples were linearised.

Figure 4.2 showed an electrophoresis gel containing single restriction enzyme digests, and undigested controls for the three proteins: WT CYGB, H81A CYGB and L46W CYGB. The gel indicates that the samples containing the restriction enzymes successfully digested most of the plasmids into a linearised form, highlighted by the distinct bands at approximately 7100 bp which matches the size of the CYGB plasmid. The controls do have some supercoiled DNA which produces the lower band on the gel; this is due to supercoiled DNA travelling much faster through the gel compared to circular single-stranded due to its conformation.

The purpose of the linearised plasmids was to transfect into cells to create stable-transfected cells lines though these were not used in the final experiments, instead transient-transfected cell lines were used with circular DNA plasmids instead.

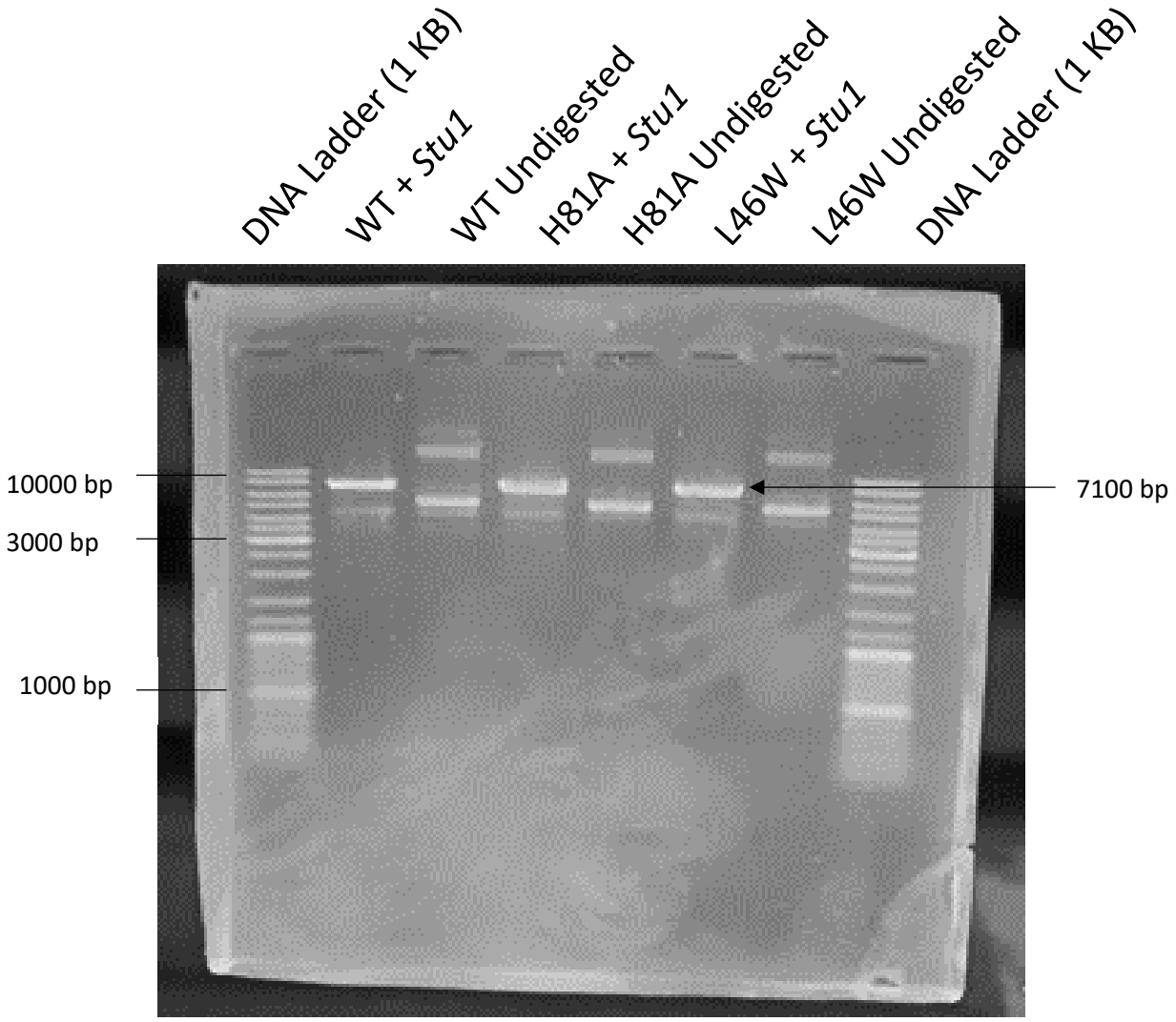


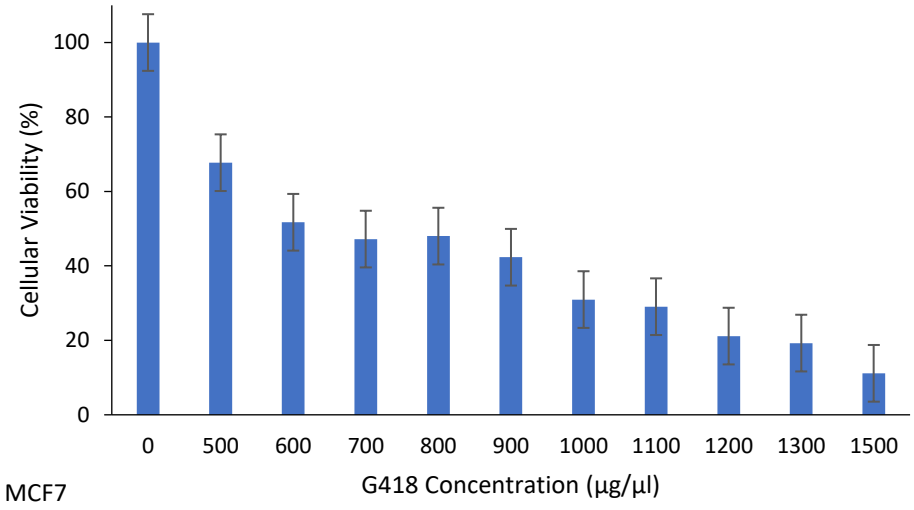
Figure 4.2.1- Restriction digest confirming plasmid linearisation. Electrophoresis gel of the single restriction digest performed on the WT, H81A and L46W DNA (pCMV vector) using *Stu1* restriction enzyme. DNA ladder (1 KB) obtained from Thermofisher.

4.3 Transfection optimisation- geneticin (G418) dose curves

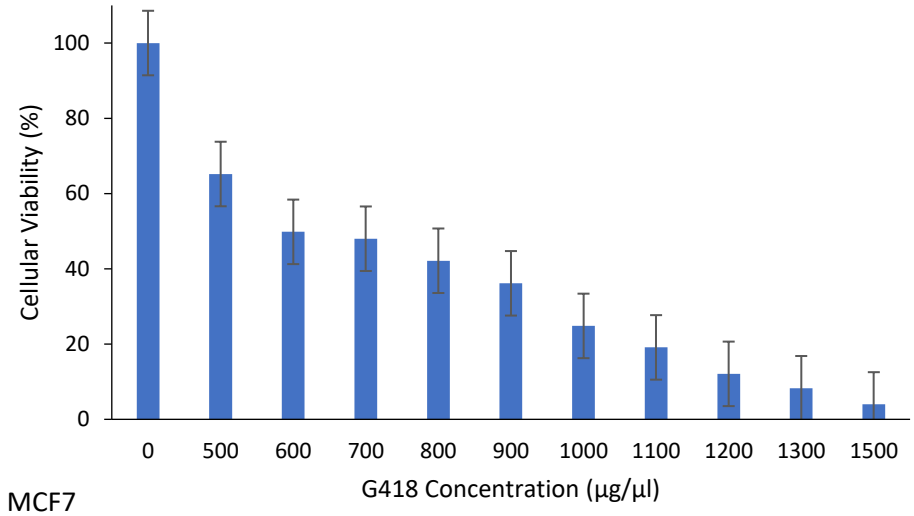
Geneticin (G418) is often used to select for stable-transfected cells. During the transfection optimisation, the optimal concentration of G418 was obtained and determined using a crystal violet assay.

As shown in Figure 4.3a, increasing G418 concentrations led to a reduction in MCF7 numbers, however some cells remained even at the highest concentration of the drug. Due to the slow proliferation speed of the MCF7 cells, the experiment was repeated for an extended period of time (Figure 4.3.1b). The results were similar but there was a lower percentage of live cells with increasing concentrations of G418 when left for a longer period of time. Irrespective of time, the optimal concentration to use was 1500 µg/µl as this killed off the highest proportion of cells.

As mentioned previously, the stable-cell lines were unsuccessful due to time constraints and therefore not used within the main experiments so G418 was no longer required.



a)



b)

Figure 4.3.1- Dose curve of G418 showing minimum of 1500 µg/µl is required to kill most cells. Kill curve of G418 concentration on WT MCF7 cells, seeded at a density of ~0.3x10^6 cells per well in 6-well plates and grown in culture medium consisting of Dulbecco's modified Eagle's medium mixture (DMEM) F-12 supplemented with 50 ml 5% foetal bovine serum and 2.5 ml 50 mg/ml gentamicin sulfate. Incubated with G418 for a) 5 days and b) 7 days before being fixed with PFA and analysed using a crystal violet assay. **a)** IC_{50}- 539.3 µg/µl and 95% CI 406.3-703.7 µg/µl, **b)** IC50- 425.9 µg/µl and 95% CI 279.9-616.3 µg/µl. Statistical analysis performed in GraphPadPrism 9.0.0 (121). Error bars show the standard error of the mean of 2 independent repeats.

4.4 Effect of H_2O_2 on cytoglobin transiently transfected cells

4.4.1 H_2O_2 assay optimisation

A wide range of literature has cited the possibility of CYGB having a protectant property on cancer cells and contributing to chemoresistance (Bholah *et al.,* 2015; Oleksiewicz *et al.,* 2013; Shaw *et al.,* 2009). H_2O_2 is a cell damaging agent which leads to oxidative stress in cells so was used in this investigation to induce stress.

In order to calculate the optimal concentrations of H_2O_2 to use in future experiments, a dose response curve was created using un-transfected HEK293S cells.

As shown in Figure 4.4.1.1, the doses of H_2O_2 had a significant effect on the cellular viability of the HEK293S cells. After adding 1 mM, the viability dropped and then after increasing the amounts it decreased by a smaller degree. None of the concentrations had a 100% efficacy so in future work higher concentrations were used.

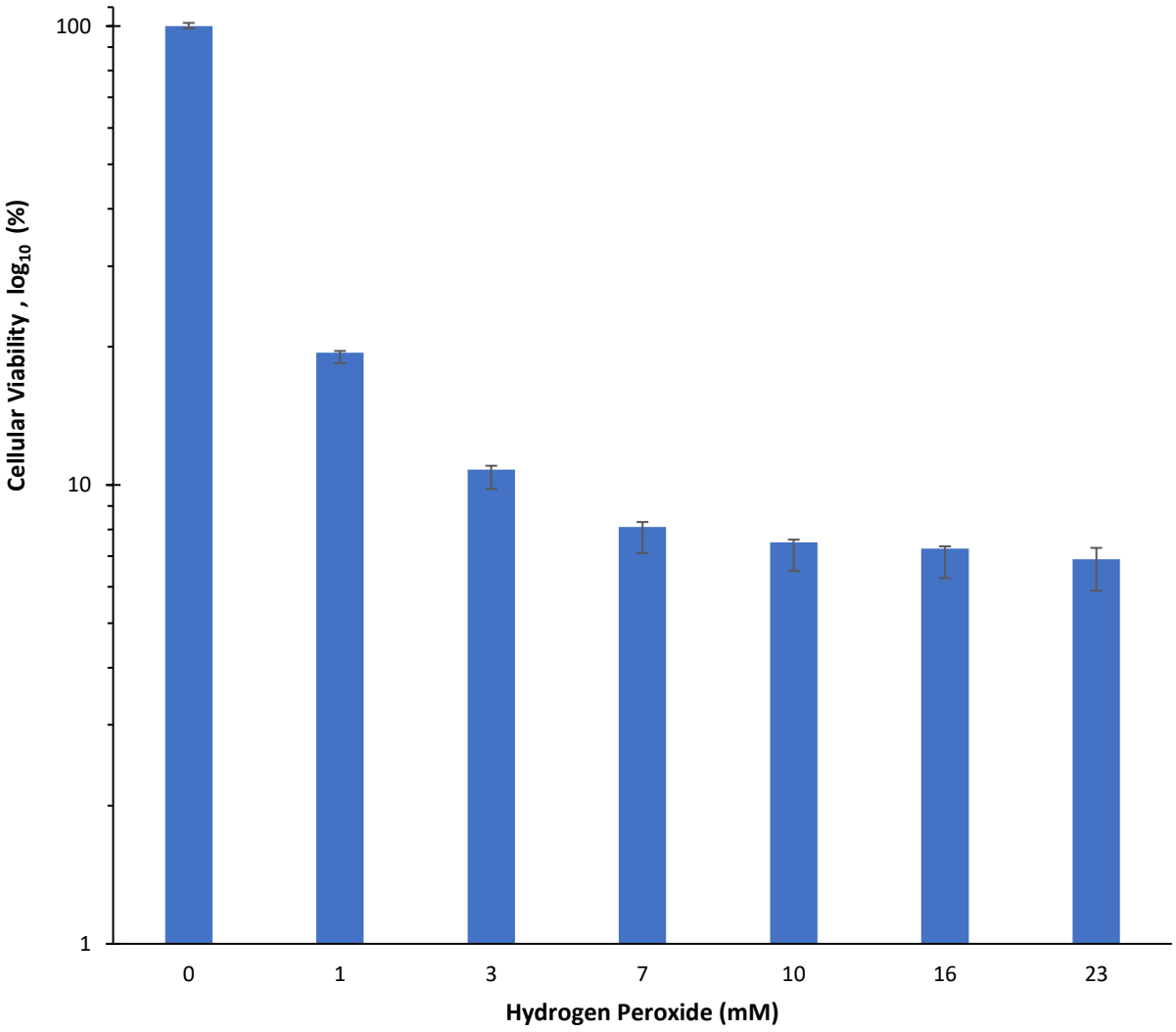


Figure 4.4.1.1- Concentrations higher than 23 mM are required to produce a successful dose curve. The cellular viability of HEK293S cells when exposed to varying concentrations of H_2O_2, for four days in a humidified atmosphere with 5% CO_2 at 37ºC. Error bars show the standard error of the mean of three repeats.

4.4.2 Effect of H_2O_2 on cytoglobin-transfected cells after four-day incubation

Following optimisation of this assay, the same time frame was used on HEK293S cells transfected with WT CYGB, H81A CYGB and L46W CYGB, before being subjected to varying concentrations of H_2O_2.

The aim of the experiment was to observe whether the presence of CYGB had any protective properties when compared to mock-transfected HEK293S cells. As shown in Figure 4.4.2.1, when cellular viability was measured against concentration of H_2O_2, all the CYGB variants displayed a sigmoidal dose-response curve. The curves were modelled and fitted to the data using equation 2.2.2.5.2. The greatest rate of cell death when exposed to H_2O_2 occurred at ~21 mM for the mock-transfected cells (no DNA) as opposed to ~25 mM for cells transfected with wild type/mutant CYGB. The viability of the mock-transfected cells decreased from 84% to 13% between 17 mM and 21 mM. For WT CYGB, H81A and L46W the viability drops from 63% to 16%, 84% to 18% and 84 to 23% respectively, however these occurred between 21 mM and 25 mM instead. This shows a greater tolerance towards the H_2O_2 from the transfected cells when compared to the mock-transfected HEK293S cells. Cellular viability eventually reaches 0% with increasing concentration of H_2O_2. The WT-transfected cells offered similar protection from the reactive oxygen species (ROS) than the two mutant CYGB variants.

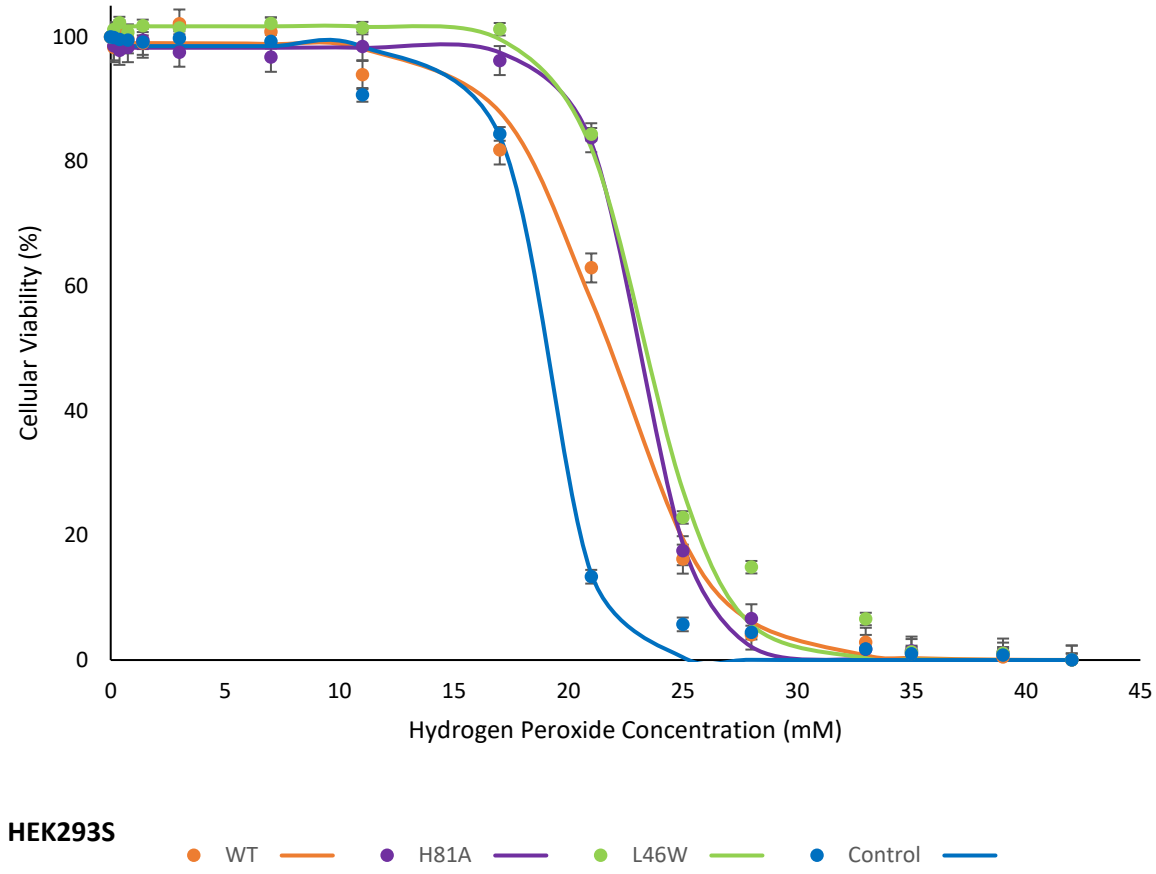


Figure 4.4.2.1 – Transient CYGB expression protects cells from damage by reactive oxygen species. Cell viability of mock-transfected HEK293S cells (control) compared to HEK293S cells transiently transfected with WT CYGB and two mutant CYGB variants; H81A and L46W, when exposed to varying concentrations of H_2O_2. The cells were seeded in a 96-well plate and left for 24 hours prior to transfection with WT CYGB, H81A CYGB and L46W CYGB using a 3:1 ratio of FuGene HD transfection reagent to DNA. After 48 hours the cells were incubated with the H_2O_2 for 4 days and then fixed with PFA stained with crystal violet. Mean of 3 repeats used.

Figure 4.4.2.2 displays the EC_{50} values of the four different cell types in addition to the errors of the fit from the graphical models which allowed for the calculation of the EC_{50} values. The lowest EC_{50} value is the HEK293S cells used as a control, they had EC_{50} values of 18.9 mM and WT CYGB, H81A and L46W had values of 21.8, 23.1 and 23.2 mM. The EC_{50} values obtained show that cells transfected with any of the CYGB variants compared to the standard HEK293S cells were statistically significant, since the *P-values* were below the threshold of 0.05 (WT 0.000579, H81A 0.0042, L46W 0.0069) and both mutants were significant to the transfected WT (H81A 0.0042 and L46W 0.0069).

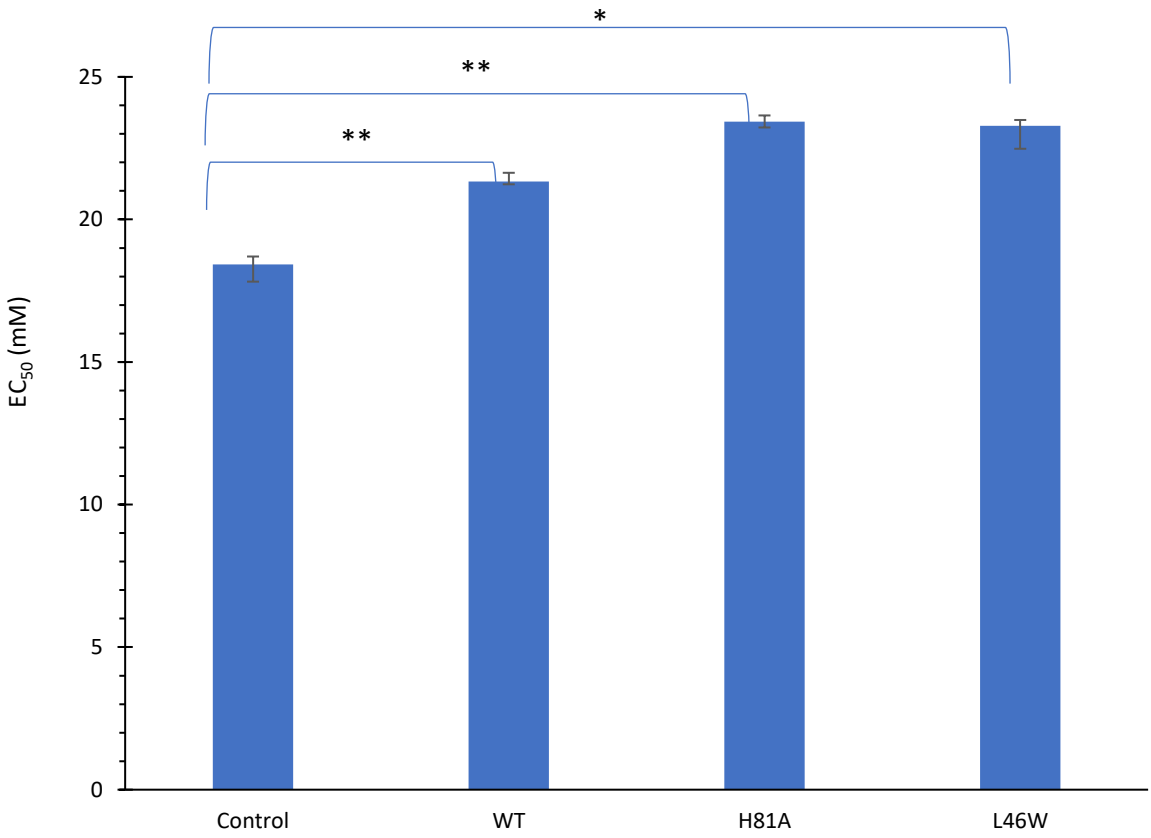


Figure 4.4.2.2- Significant enhancement of reactive oxygen species protection in CYGB-transfected HEK293S cells. Graph demonstrating EC_{50} values of cells transiently transfected with either WT CYGB or two mutant CYGB varieties: H81A and L46W compared to mock-transfected HEK293S cells (control) incubated with varying concentrations of H_2O_2. Cellular viability measured using a crystal violet assay and EC_{50} values calculated by modelling and fitting the data using Equation 2.2.2.5.2. t-test values displayed using: *P<0.05, **P<0.005 when compared to the mock-transfected control cells, the following parameters used for the test: paired two-tailed t-test. Results are mean of three experiments + SD.

4.4.3 Effect of H_2O_2 on cytoglobin-transfected cells after seven day incubation

To see if CYGB can protect cells from ROS damage over longer time periods, the experiment was repeated for a duration of 7 days (Figure 4.4.3.1).

As before, HEK293S cells were transfected with WT CYGB, H81A CYGB and L46W CYGB before being exposed to varying concentrations of H_2O_2. The data shown in Figure 4.4.2.2 was very similar to the data shown in Figure 4.4.3.1 with both having a classic sigmoidal dose-response curve once modelled and fitted in the same manner as previously. It is clear that the data sets for the control are different to the CYGB-transfected data sets which have a similar pattern. The drop in cellular viability of the control cells when exposed to H_2O_2 occurred after 10 mM whereas the three CYGB variants were relatively stable until after 17 mM. The control cells viability decreased by 28.3% between 11-17 mM of H_2O_2 whereas the CYGB variants in the same concentration range decreased by 4.6%, 4.1% and 1.2% (WT, H81A and L46W respectively). The control cells hit ~0% cell viability at 28 mM whereas a higher concentration of H_2O_2 was needed to reduce viability to 0% in the CYGB-transfected cells. The data shows a clear improved tolerance to the ROS-inducing H_2O_2 of cells transfected with CYGB when compared to the control cells and an even greater impact than the cells incubated for four days (4.4.2.2). Again, the WT CYGB offered similar protection to ROS as the two CYGB variants.

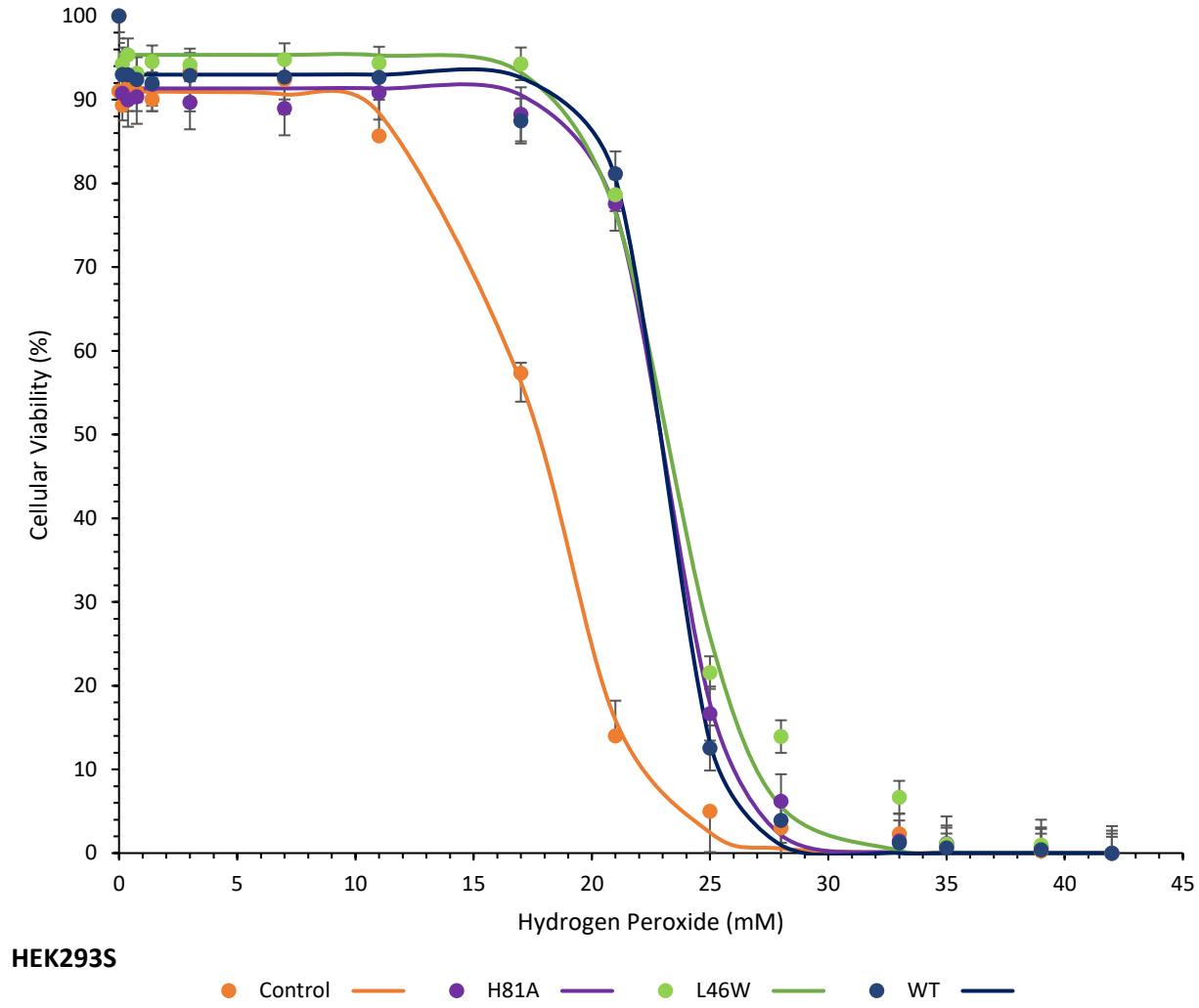


Figure 4.4.3.1- Transient CYGB expression protects cells from reactive oxygen species damage after seven days exposure to H_2O_2. Cell viability of HEK293S cells (control) compared to HEK293S cells transiently transfected with WT CYGB and two mutant CYGB variants; H81A and L46W, when exposed to varying concentrations of H_2O_2. The cells were seeded in a 96 well plate and left for 24 hours prior to transfection with WT CYGB, H81A CYGB and L46W CYGB using a 3:1 ratio of FuGene HD transfection reagent to DNA. After 48 hours the cells were incubated with the H_2O_2 for seven days and then fixed with Paraformaldehyde and stained with Crystal violet. Mean of 3 repeats used.

As before, the EC_{50} values were calculated and a t-test performed.

Figure 4.4.3.2 displays the EC_{50} values of the three different variants in addition to the errors of the fit. The EC_{50} value of the control cells was the lowest at 17.9 mM compared to WT, H81A and L46W which had EC_{50} values of 23.0, 23.1 and 23.3 mM respectively. The EC_{50} values show that all of the CYGB-transfected cells were more resistant to the effects of higher concentrations of H_2O_2 and therefore ROS, than the mock-transfected control cells.

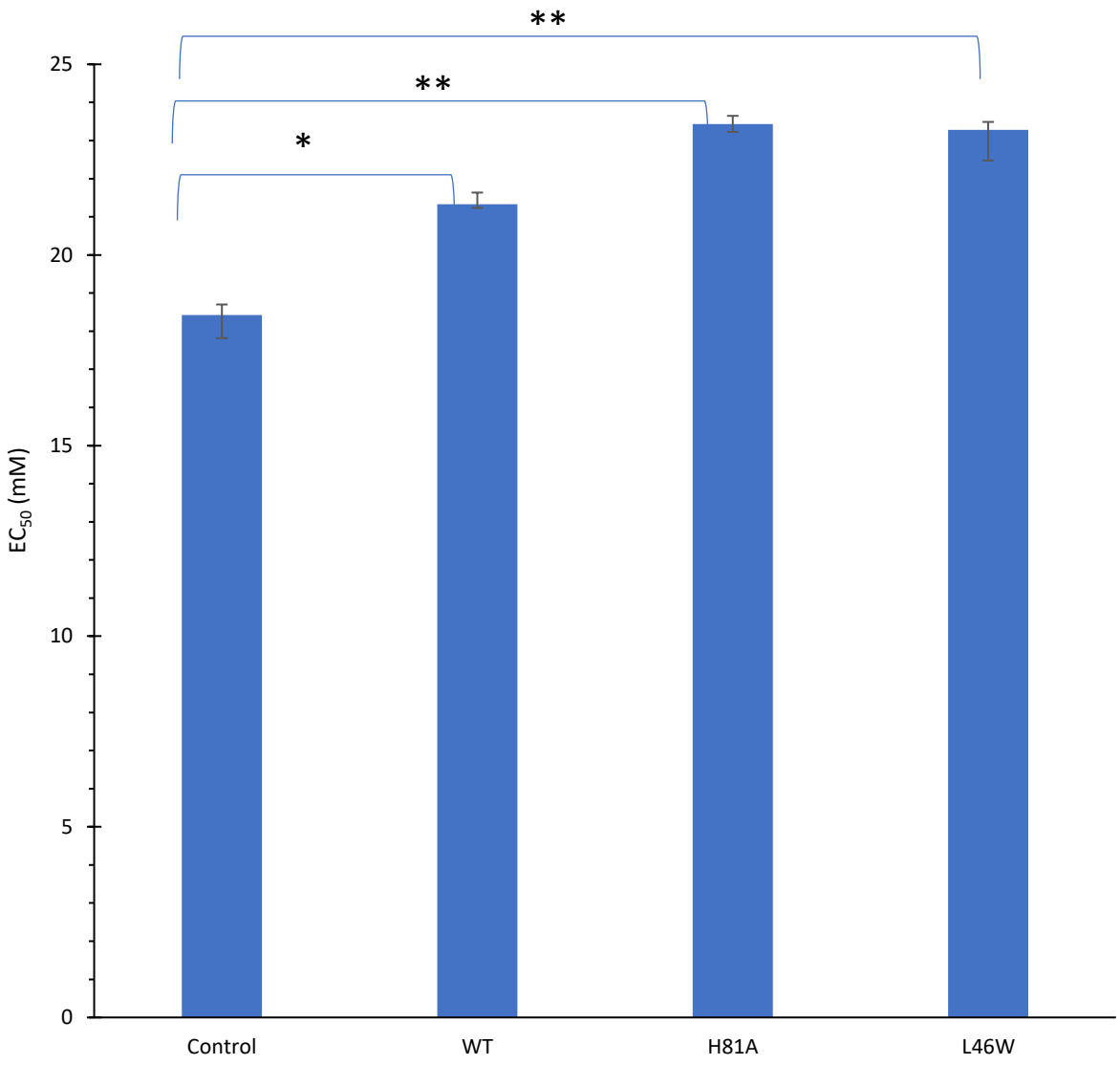


Figure 4.4.3.2- Significant enhancement of reactive oxygen species protection in CYGB-transfected HEK293S cells when incubated with H_2O_2 for an extended period of time. Graph shows EC_{50} values of cells transiently transfected with WT CYGB and two mutant variants: H81A and L46W, compared to mock-transfected HEK293S control cells incubated with varying concentrations of H_2O_2 for seven days. Cellular viability measured using a crystal violet assay and EC_{50} values calculated by modelling and fitting the data using Equation 2.2.2.5.2. t-test values displayed using: *P<0.05, **P<0.005 when compared to the mock-transfected cells, test parameters were two-tailed distribution and paired test. All three transfected variants were significant against the control cells (WT 0.028, H81A 0.0015 and L46W 0.0012) but neither mutant was significant against the transfected WT (H81A 0.92 and L46W 0.76). Results are mean of three experiments.

4.5 Effect of chemotherapy drugs on the viability of cytoglobin-transfected HEK293S and MCF7 cells

In order to evaluate the effect of CYGB's potential chemoprotective function, cells transfected with WT CYGB, H81A CYGB and L46W CYGB were exposed to two different chemotherapy drugs: paclitaxel and docetaxel, both of which are not ROS-inducing chemotherapeutics.

In HEK293S cells, both paclitaxel and docetaxel reduced cell viability to approximately 50% at 100 nM, in MCF7 cells treated with paclitaxel, the distribution and reduction in viability was similar to the HEK293S cells. When exposed to docetaxel, MCF7 control cells have little difference before 1 nM since the cellular viability is still at 100%; the others are all below 80% with L46W at the lowest at ~60%.

The WT and mutant varieties did not appear to offer any protection to HEK293S cells in response to the chemotherapeutics used, however MCF7 cells showed slight improvement in cellular viability. L46W appeared to decrease the cellular viability, particularly in MCF7 cells.

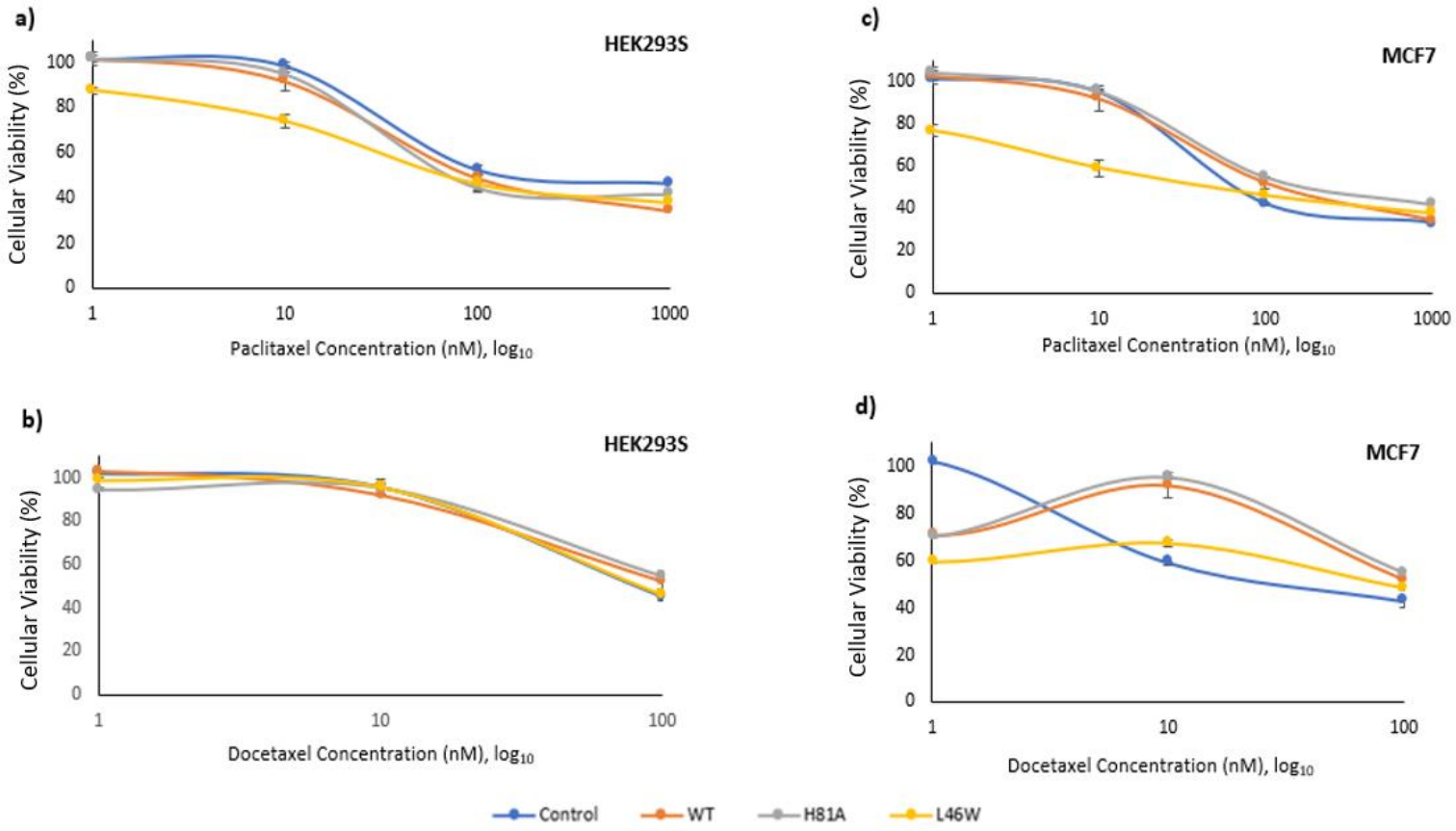


Figure 4.5.1- Decreased protection from reactive oxygen species in mutant L46W CYGB-transfected cells. Graphs to show cellular viability in mock-transfected HEK293S and MCF7 cells as well as HEK293S and MCF7 cells transfected with WT CYGB, H81A and L46W CYGB mutants when subjected to varying concentrations of the chemotherapy drugs paclitaxel and docetaxel. The cells were seeded in a 96 well plate and left for 24 hours prior to transfection with WT CYGB, H81A CYGB and L46W CYGB using a 3:1 ratio of FuGene HD transfection reagent to DNA. After 48 hours the cells were incubated with the chemotherapy drugs for seven days and then fixed with Paraformaldehyde and stained with Crystal violet. Errors of the fit are mean of two experiments + SD.

EC_{50} values were derived from this data and Figure 4.4.2 shows the concentration values at which the cellular viability was halved.

The data shows that there was no significant difference between the mock-transfected control cells and those transfected with WT CYGB. There was a significant difference between the L46W-transfected MCF7 and HEK293S cells when treated with paclitaxel and a significant difference in docetaxel-treated MCF7 cells when compared to all other variants. These docetaxel- treated MCF7 cells also showed a significant difference between the WT and H81A variant as well as the control and H81A.

This data shows that the cells transfected with L46W have a significantly lower cell viability than both the controls and other CYGB variants, which implies that the presence of this mutation could afford the cells a decreased tolerance to these drugs.

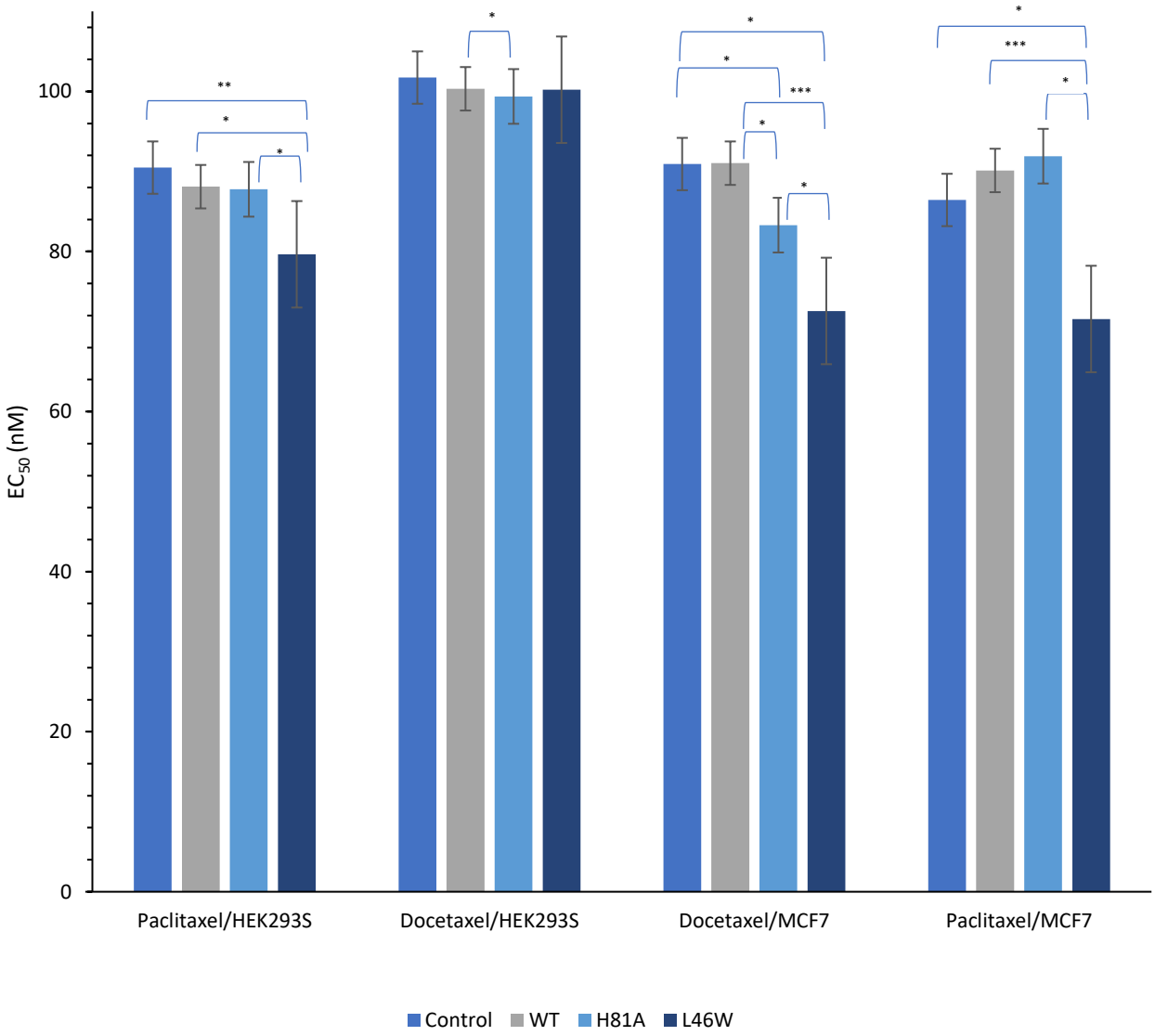


Figure 4.5.2- EC_{50} decrease in transfected CYGB L46W cells when incubated with chemotherapy drugs. Graph shows EC_{50} values of HEK293S and MCF7 cells transiently transfected with WT CYGB and two mutant variants: H81A and L46W compared to mock-transfected HEK293S and MCF7 control cells, incubated with varying concentrations of docetaxel and paclitaxel for seven days. Cellular viability measured using a crystal violet assay and EC_{50} values calculated using Microsoft Excel to calculate the concentration at which the viability reaches 50%. t-test parameters were two-tailed distribution and paired test. t-test results values displayed using: *P<0.05, **P<0.005, ***P<0.005. Mean of two repeats.

In addition to the cellular viability and the EC_{50} data, microscopic images also showed a noticeable difference in the cells treated with two different chemotherapy drugs: docetaxel and paclitaxel (Figure 4.5.3).

The images show the effect of 100 nM of docetaxel and paclitaxel on MCF7 cells when incubated for seven days. In the absence of the drug, cells are small, confluent and densely populated. After incubation with 100 nM paclitaxel, the cells are much fewer and more spread out. They are much more irregular in shape and there is evidence of cell death. Similarly, with the docetaxel, there are very few cells, even fewer than the paclitaxel at the same concentration. The cells which are present do not look healthy, they are tightly bound, and some appear to have ruptured. The difference between the control and those with chemotherapy drugs is evident in the number of the cells, the distribution and the quality of the remaining cells.

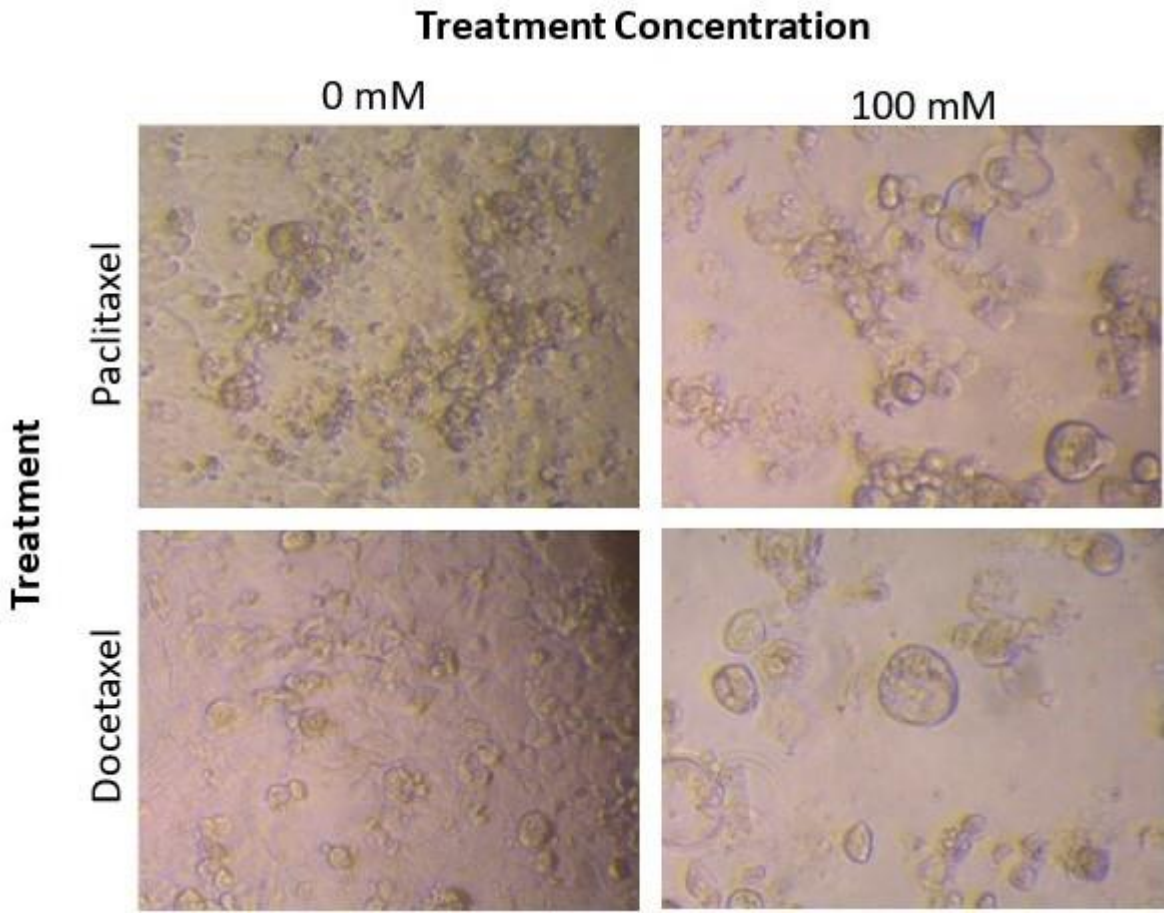


Fig 4.5.3 – Chemotherapy treatments cause cell proliferation loss in MCF7 cells. Images of MCF7 cells after incubation for 7 days with either 0 or 100 nM of the two chemotherapy drugs: docetaxel and paclitaxel. Images taken on the Optika Brightfield microscope with a Carl Zeiss MicroImaging GmbH AxioCam ERc 5s camera attachment on x20 magnification using the Optika Vision Lite software.

Chapter 5: Results- Biophysical characterisation of cytoglobin protein wild type and mutants designed to interfere with potential physiological and pathological functions

5.1 Optical characterisation of the recombinantly expressed mutants compared to wild type protein

5.1.1 Ferric, deoxyferrous and carbonmonoxy-ferrous forms

Figure 5.1.1.1 shows the spectra of the three Cygb variants firstly in the ferric form, then when exposed to dithionite to generate the deoxyferrous form of the protein and finally following addition of CO. The ferric form of the WT Cygb (*a*) shows a curve with two relatively flat peaks at approximately 535 nm and 570 nm. Conversely, ferric H81A CYGB (*b*) has three peaks, with the additional peak at approximately 630 nm. This is typical of ferric pentacoordinate globins such as Mb and Hb and so indicates that the H81A mutation changes the protein from primarily hexacoordinate to a pentacoordinate conformation.

The addition of the dithionite to form the deoxyferrous protein results in a broad peak at 555 nm. The binding of CO showed three separate peaks; the two peaks (535 and 570 nm) are identical to that of the WT ferrous-CO protein. However, a third peak at approximately 635 nm is sharper and more distinct than seen in the ferric protein. The spectra of ferric L46W are more similar to the ferric WT, suggesting that the L46W protein is hexacoordinate, with only two peaks presents and at similar approximate wavelengths. There are minor differences in peak amplitude. In terms of the Soret peaks, the WT ferric protein had the smallest peak, with a large jump from 1.3 to 1.9 to the peak of the protein combined with CO. The deoxyferrous Soret peak was the largest with an absorbance of approximately 2.07. The ferric protein saw a hyperchromic and bathochromic shift in its Soret peak. Interestingly, H81A did not follow this pattern and showed that the deoxyferrous protein had the lowest peak with an absorbance of approximately 1.56 and was unusually broad. The ferric Soret peak was around 2.42 and then when reduced and bound to CO, rose to 2.36. The ferric protein saw a bathochromic and hypochromic Soret shift when forming the deoxyferrous protein but saw a bathochromic and hyperchromic with the binding of CO. This pattern was also mirrored by the L46W mutant.

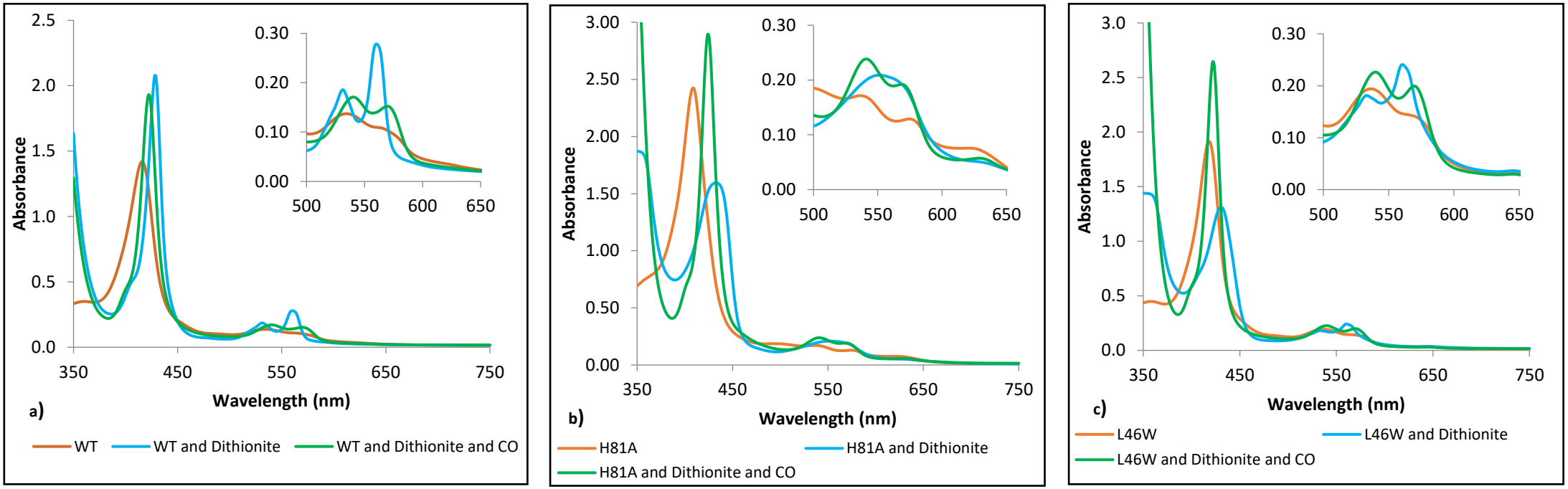


Figure 5.1.1.1- Optical spectra showing the addition of dithionite and CO to a) WT Cygb, b) H81A Cygb and c) L46W Cygb protein. The absorbance of 5 μM protein was measured using a Cary 5000 Spectrophotometer between 350-750 nm with baseline correction. After the initial spectra of the ferric protein (orange lines), few grains of dithionite were added to each protein to form the deoxyferrous protein (blue line). Finally, an excess of CO was bubbled into the cuvette to produce the ferrous CO form (green line). Inset shows a magnified version of the visible region of the spectra.

5.1.2 Reduction of ferric cytoglobin to oxyferrous using ascorbate

Sodium ascorbate was added to the ferric protein since this reduces the protein.

Figure 5.1.2.1 showed the three spectra all have very uniformly shaped double peaks at the same wavelengths with the only distinction being that the amplitude of the WT peak is much lower than the two mutants. These are all typical of oxyferrous forms of the protein with a dioxygen in the sixth coordination site. A more in-depth view of the kinetics takes place in Section 5.1.8.

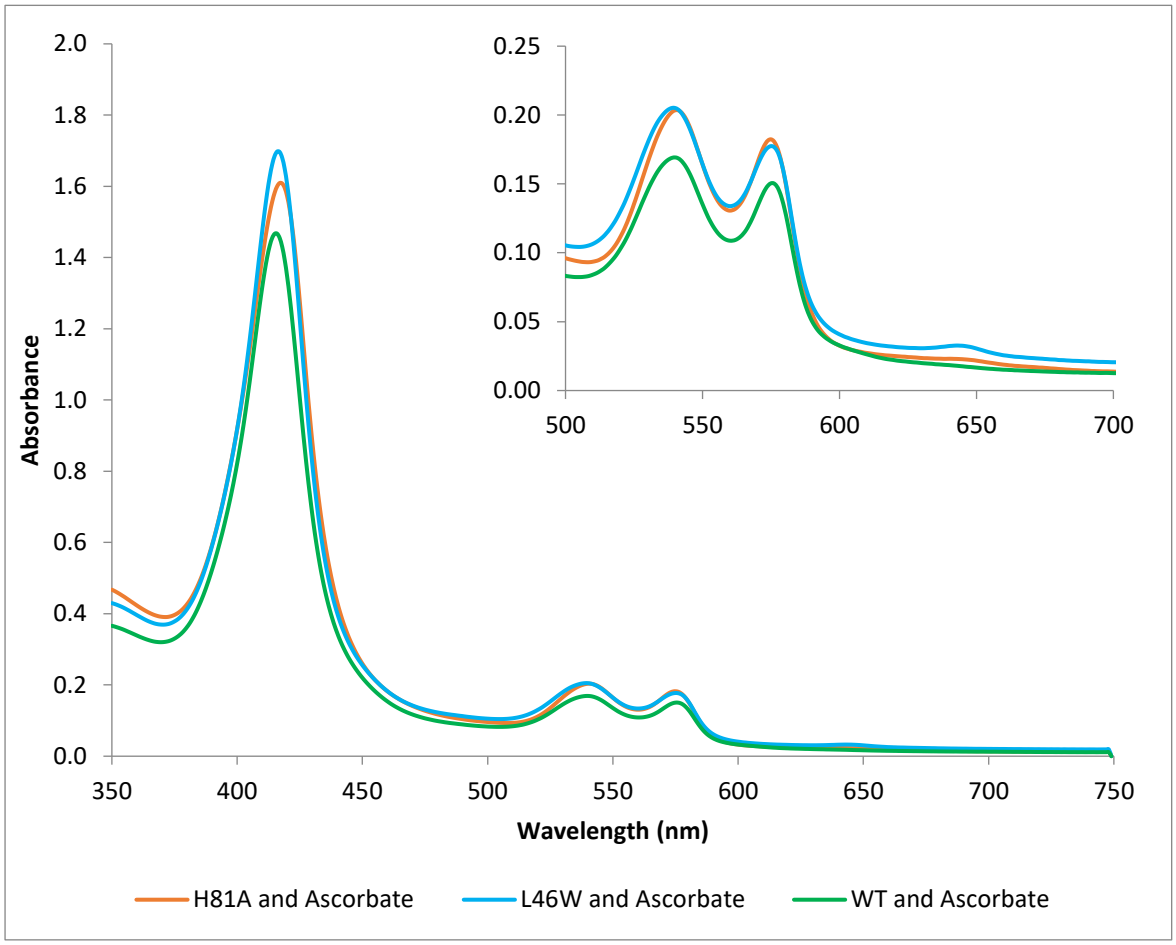


Figure 5.1.2.1- Oxy-ferrous protein reduced with sodium ascorbate. UV-Visible spectra of WT Cygb as well as two mutants: H81A and L46W measured between 350-750 nm on a Cary 5000 spectrophotometer. Brown: H81A protein, blue: L46W protein, green: WT protein. Inset shows magnified version of the visible region of the spectra.

5.1.3 Optical spectra of nitric oxide bound ferrous cytoglobin protein variants

In addition to the conditions investigated already, the protein was exposed to Proli-NONOate to form the ferrous-NO species.

Figure 5.1.3.1 shows that all the peaks of the NO-bound ferrous mutants look very different from that of the WT protein. The NO-bound WT protein has a peak at around 530 nm, then a much higher intensity peak at 560 nm and finally a very low intensity peak at approximately 585 nm, though this peak is very flat compared to the others. H81A has two uniform β and α peaks at approximately 545 and 580 nm respectively. It is difficult to determine the exact status of the peaks in L46W and whether it is a single, very broad αβ peak or whether they are distinguishable as separate peaks since there is a slight increase in the centre.

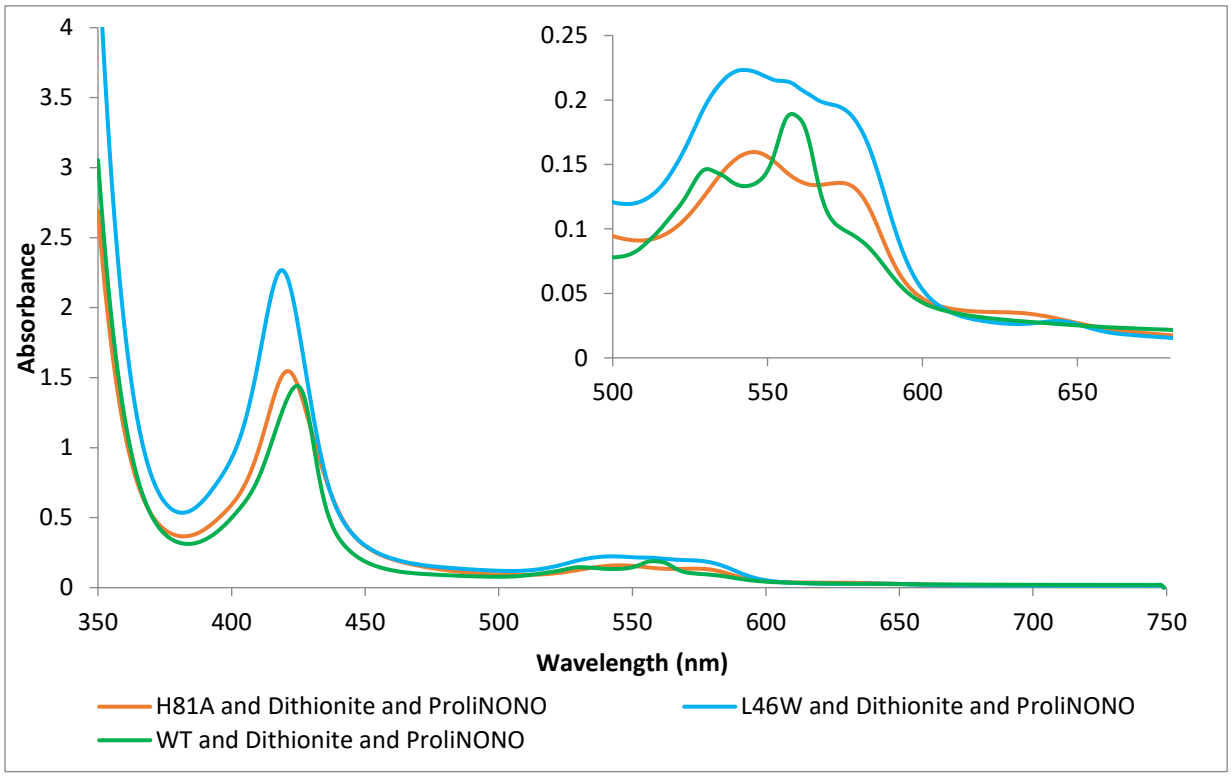


Figure 5.1.3.1- Ferrous-NO protein spectra after exposure to nitric oxide generated by Proli-NONOate. UV-Visible spectra of ferrous-NO WT Cygb as well as two mutants: H81A and L46W measured between 350-750 nm on a Cary 5000 spectrophotometer. Brown: H81A protein, blue: L46W protein, green: WT protein. Inset is magnified version of the visible region of the spectra.

5.1.4 Binding of lipids to wild type cytoglobin and mutant variants

Cygb has been reported to bind lipids, with the protein transitioning from a hexacoordinate to a pentacoordinate conformation when in the ferric oxidation state. This makes it a potential candidate for redox-linked cell signalling mechanisms (Reeder *et al*., 2011).

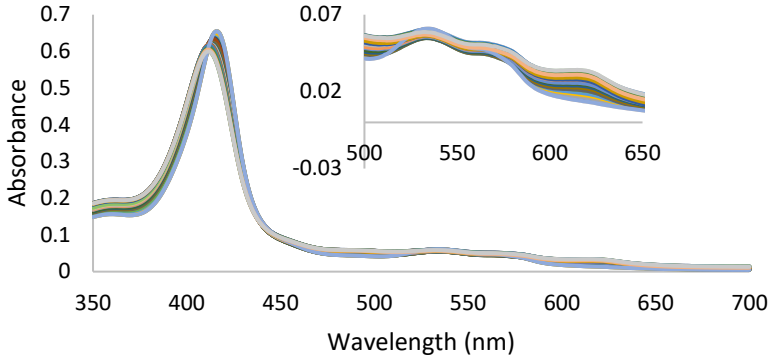
Figure 5.1.4.1 shows the optical spectra produced during the binding of oleate lipids to the three different Cygb variants: WT, H81A and L46W.

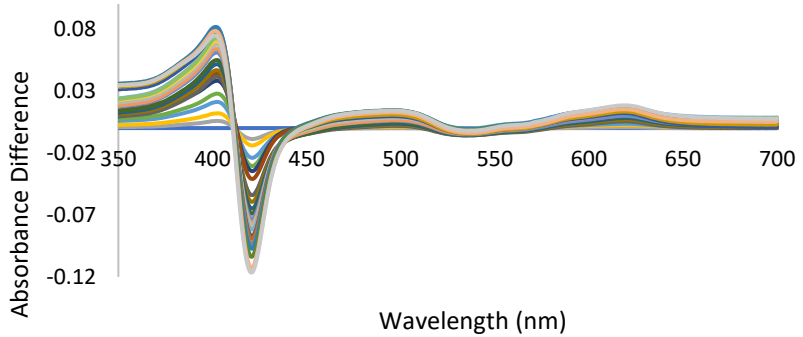
The absorbance difference spectra show that there is an isosbestic point at 411 nm which means that the peak is not a result of dilution and not a true change in absorbance from one state to another (more complex transitions do not show isosbestic points). The most distinct change is around 620 nm which is the appearance of the third peak not usually present. This addition of a third peak following oleate binding shows the transition from hexacoordinate to a pentacoordinate conformation resulting in peaks at wavelengths of 535, 570 and 620 nm.

H81A is already in the pentacoordinate confirmation, having been fixed there by the presence of the mutation. The optical spectra show the same three peaks seen in the WT after oleate addition. Unexpectedly, the Soret peak shows changes upon oleate addition which indicates that there is potentially a conformational change taking place. There is also a small increase in the absorbance of the other three peaks present. Both the WT and the H81A have a similar increase in the absorbance in the difference spectra around 620 nm.

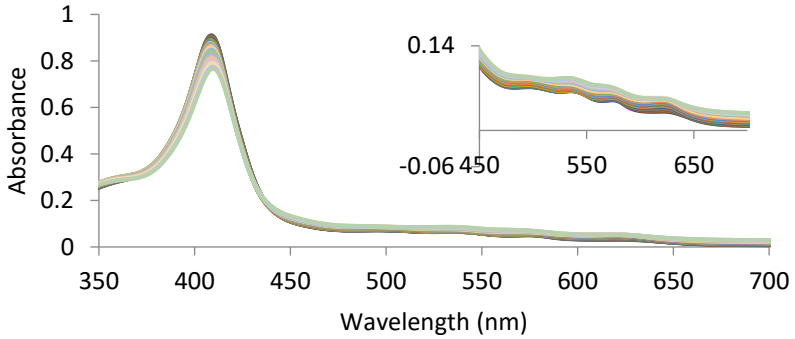
L46W has a different spectrum to the other two Cygb variants who have virtually no difference in the optical spectra. The protein is still in the hexacoordinate confirmation as can be seen by the lack of a 620 nm peak and it does not appear to have bound any lipids. Again unexpectedly, there is no change in the absorbance past 600 nm unlike the other two variants. Additionally, although there is a negative change in absorbance at the Soret peak, the absorbance scale on this protein is much smaller than the other two and can largely be a result of protein dilution.

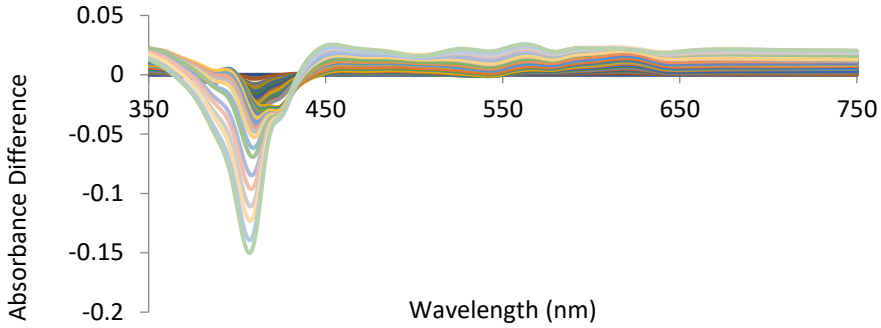
a) WT Cygb



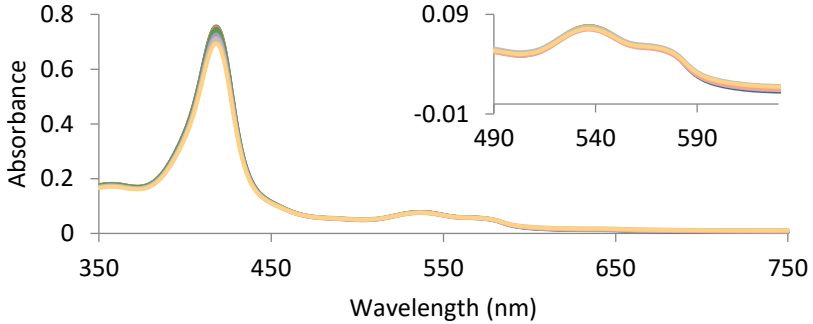


b) H81A Cygb





c) L46W Cygb



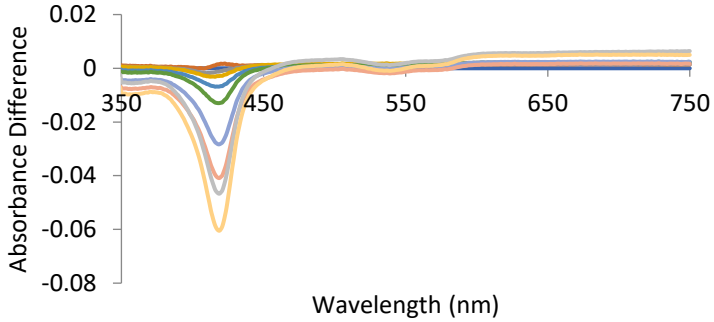


Figure 5.1.4.1- Left: Optical changes of Cygb following the titration of oleate (lipids) and right: a difference spectrum to account for dilution. 0.5 µM oleate was added to 10 µM protein in 5 µl increments until saturated. Measured on a Cary 5000 spectrophotometer.

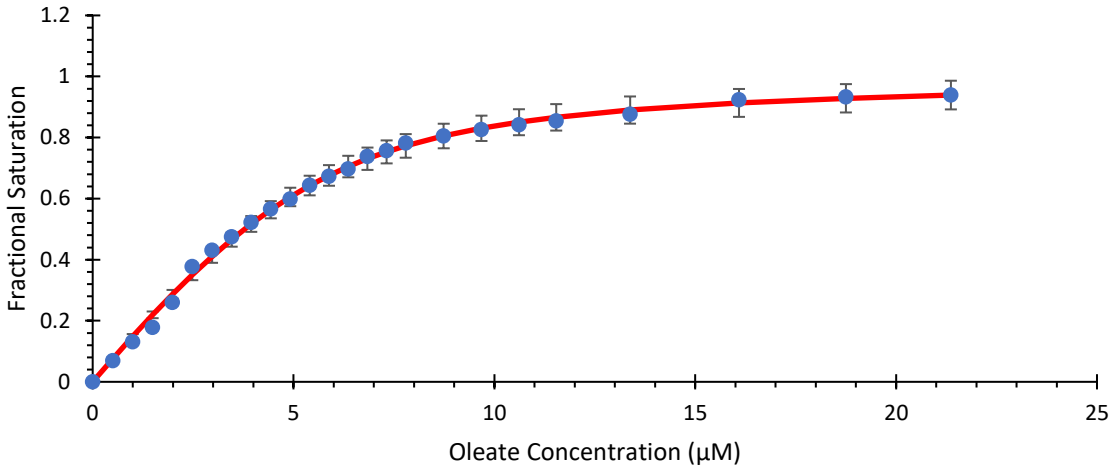
In addition to the optical spectra, a titration curve was created showing the fractional saturation of lipids against the lipid concentration.

As can be seen in Figure 5.1.4.2 there is a significant difference between the WT, H81A and the L46W.

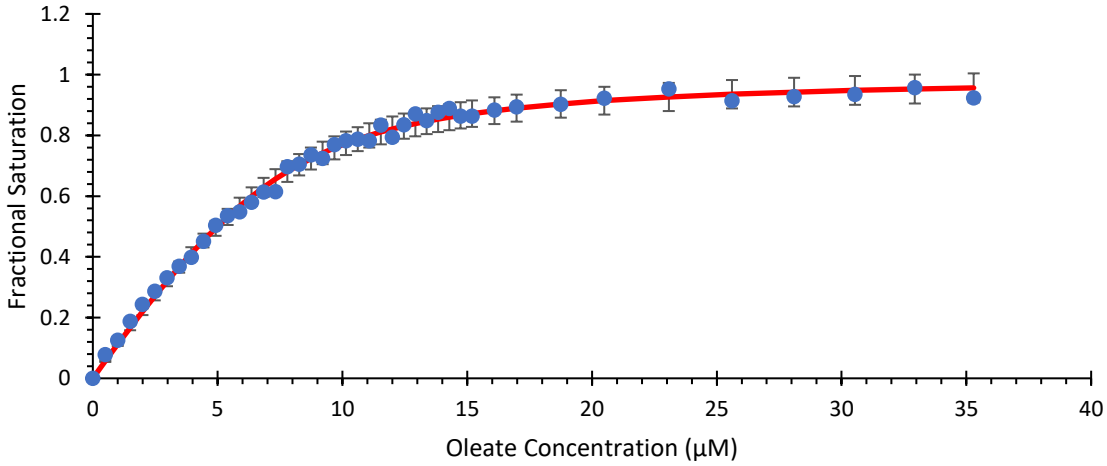
It shows that with increasing amounts of oleate, both the WT and H81A mutant display a standard classical fractional saturation binding curve which flattens off once fully saturated by the lipids and binds to the protein in an approximately 1:1 stoichiometry. L46W on the other hand shows a very linear relationship between the saturation and concentration.

The data has been adjusted for sample dilution and has been normalised to 1 for full saturation. The WT has a binding constant of 1.05 uM at a total protein concentration of 5.45 µM as the protein was 5 µM this gives a stoichiometry of 1:1.09 (protein: lipid); H81A has a binding constant of 1.29 µM at a total protein concentration of 7.41 µM giving a stoichiometry of 1:1.48 (protein: lipid) and finally L46W has a binding constant of 118.85 µM at a total protein concentration of 149.55 µM, giving a stoichiometry of 1:29.91 (protein:lipid).

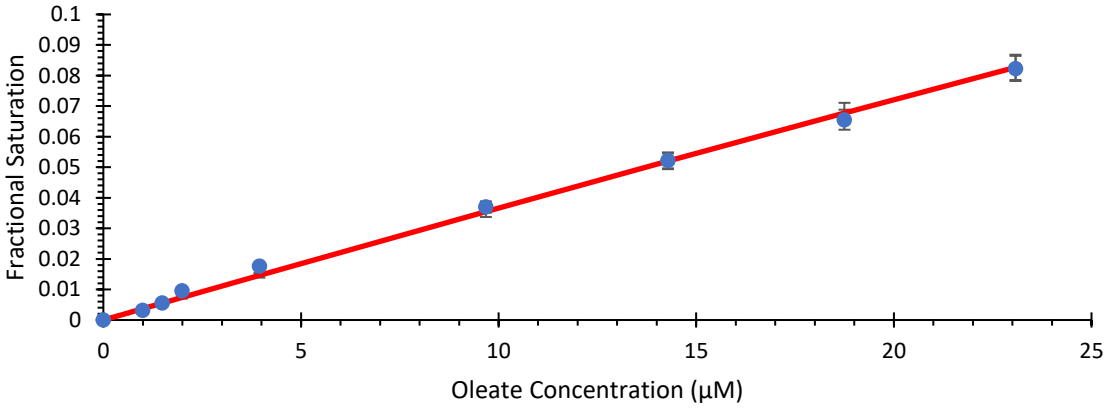
The binding of H81A to the lipids is unexpected since it looks very similar to the WT binding curve despite the fact that H81A is fixed in the penta-coordinate state and so should in theory not produce the same curve or optical spectra. L46W is designed to partially block the back of the haem pocket, due to the indole ring on the tryptophan slowing down NO ligand binding. This should not interfere with oleate binding, however no structural transitions are observed upon lipid addition. There is a slight increase in fractional saturation at high oleate levels, but this could be due to the lipids acting as a detergent at higher concentrations and unravelling the protein, which has previously been seen in Ngb (Reeder *et al.,* 2011). Saturation did not occur under the conditions of the experiment and full saturation levels were estimated from the optical changes observed from the WT protein.



a)



b)



c)

Figure 5.1.4.2- Lipid binding optical spectra showing the addition of oleate to protein. a) WT Cygb, b) H81A mutant Cygb and c) L46W Cygb. Oleate was added to 5 µM protein in 0.5 µM increments until saturated. Measured on a Cary 5000 spectrophotometer. Mean of 3 repeats.

5.1.5 Liposome oxidation by cytoglobin wild-type, mutants (L46W/H81A) and myoglobin

Having examined the binding of lipids to Cygb, the relationship between Cygb and multiple lipids was considered since it is known that Cygb can readily oxidise phospholipids *in vitro* (Reeder *et al.,* 2011) and that this would be more realistic *in vivo*. Given that liposomes are vesicular structures composed of lipids such as bilayers (Monteiro *et al.,* 2014), it is of interest to observe the interaction between Cygb and liposomes as a model for cells, where the oxidation of the model cells can be observed directly from changes in the lipid conjugation at 234 nm. The liposomes used within this investigation were a mix containing ~14-29% L-α-phosphatidyl choline and used an extruder to produce uniform, single lamellar versions.

Figure 5.1.5.1 shows the optical changes for the reactions of Mb along with both WT Cygb and two mutant variants of Cygb. The reactions of Mb and all three Cygb types show an increase in the absorbance at approximately 234 nm which signifies the oxidation of the liposome lipids. Mb is used as a comparison as this is a well-known oxidant of lipids under pathological conditions (Baron and Anderson, 2002; Faustman *et al.,* 2010). Mb reacts with the lipids in the liposomes and after a lag period, results in a cascade of lipid oxidisation forming various lipid radical species. When these are oxidised, some polyunsaturated lipids rearrange to form a conjugated diene (-C=C-C=C-), however some dienes are destroyed or generate dienes conjugated with a ketone, or trienes. The diene can be directly measured at ~234 nm, with the diene-ketone or triene visible at 280 nm (Reeder *et al.,* 2011). In these experiments, formation of conjugated diene outweighs the destruction. The control shows a very small increase in absorbance though this can be attributed to the small amount of peroxides or from oxidation, by metal salts in the buffer which lead to a small, but measurable amount of liposome oxidation. This is similar to what is seen with Cygb however due to the instability of the ferryl Cygb, the result is a lower overall formation of conjugated dienes. Bleaching of the haem from radical damage is seen in Cygb, most with the H81A, less with the WT and very little with L46W.

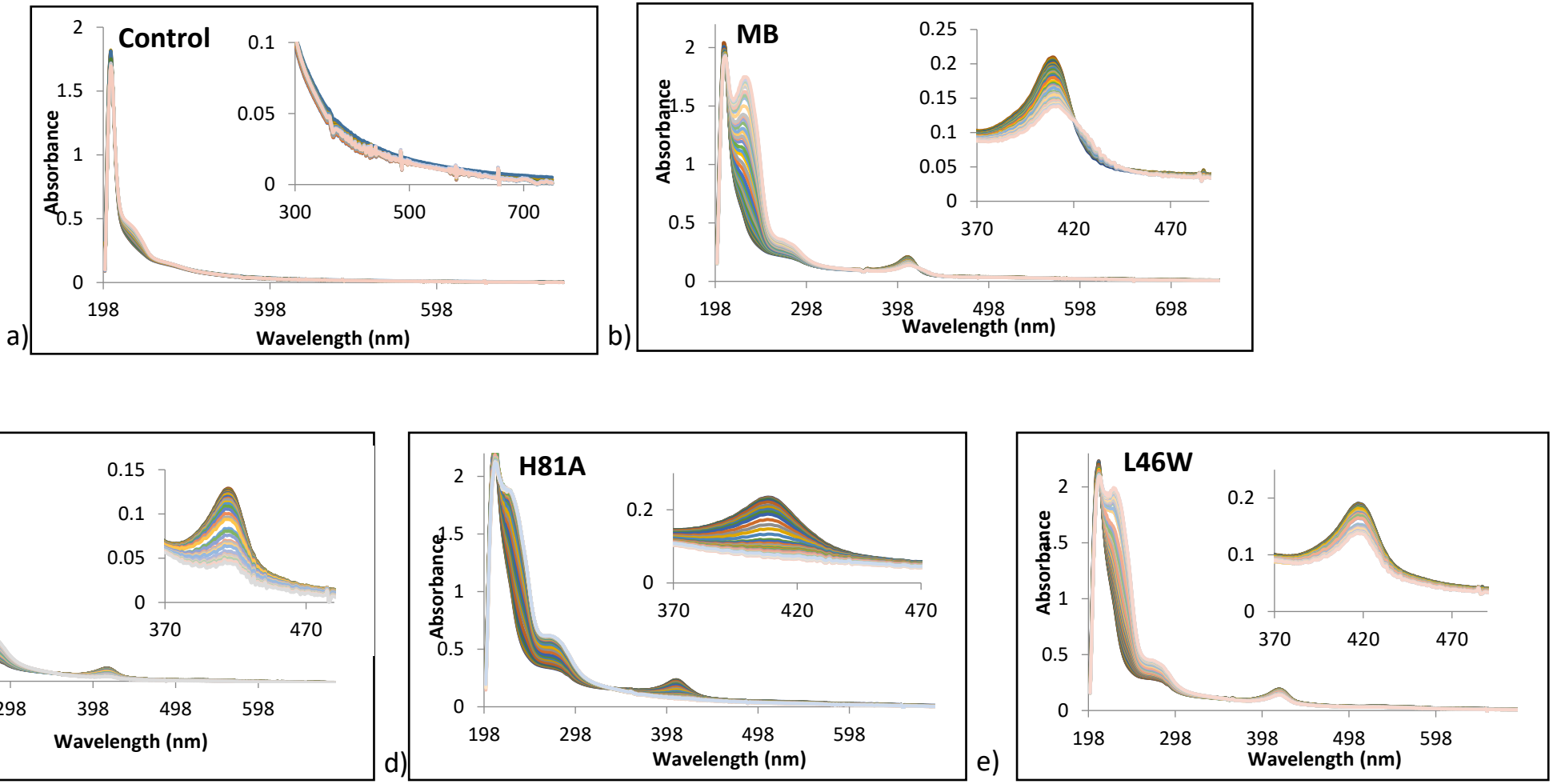


Figure 5.1.5.1- Optical spectra of the oxidation of liposomes by WT Cygb, Cygb mutants and Mb. 1 ml of 2 µM protein was added to small unilamellar liposomes (40 µl, 50 mg in 460 µl 0.1 M phosphate buffer, pH 7.4). Optical changes following addition of a) control, b) Mb, c) WT Cygb, d) H81A and e) L46W. Run for 5 hours, spectra taken every minute though only spectra every 3 minutes were displayed graphically. Inset shows magnification of the visible region of the spectra.

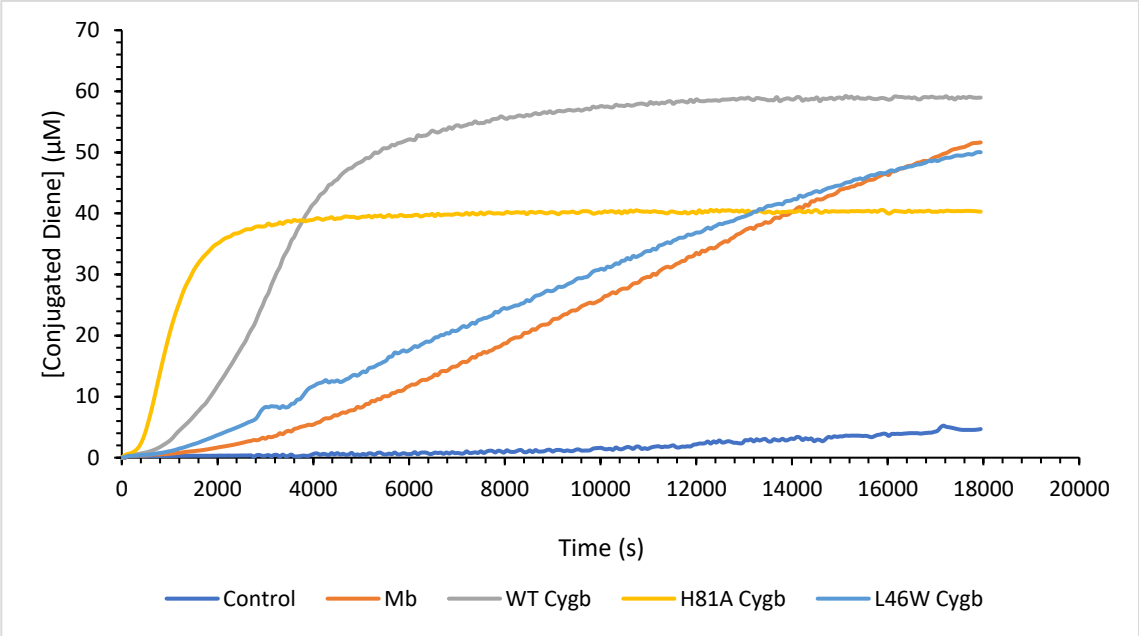
From the kinetic traces of the liposome oxidation, the maximal rates were calculated.

Figure 5.1.5.2 shows the effects of different proteins or protein variants on the oxidation of single lamellar liposomes over time. The data confirms that there is very limited interaction occurring in the control which can be attributed to trace amounts of metal salts in the buffer oxidising the liposomes. The data also shows two distinct patterns occurring, the first shown by L46W and Mb and the other by WT Cygb and H81A. Mb is relatively slow at oxidising liposomes in comparison to WT Cygb and has a relatively linear relationship between absorbance and time. Rhabdomyolysis (pathology caused by muscle injury resulting in acute renal failure) is speculated to involve the direct interaction between Mb and mitochondria, resulting in iron ions release from the haem, promoting peroxidation of mitochondrial membranes (Plotnikov *et al.,* 2009). WT Cygb has a more rapid rate and oxidises the liposomes faster than Mb but then once the reaction is complete, the levels plateau off, this occurs at 3.9 h (14,000 s). Both the control and L46W do not have a complete reaction and do not plateau off in the timeframe of this experiment. H81A has a much faster reaction than any of the other proteins tested and around 33 min (2000 s) the reaction is complete, and the absorbance plateaus off. Despite being much faster than WT Cygb, H81A does not have the highest absorbance. L46W provided very different results to the results of both of the other Cygb varieties and appears to mimic similar patterns to Mb.

The maximal rates of the four proteins, along with the controls, was obtained. The results confirm the data displayed in a); H81A has a significantly faster maximal rate than any of the other types of Cygb and much faster than Mb at 34.9 nMs^{-1}. The control has negligible rates. L46W is 11.7, WT is 15.2 and Mb is 3.5 nMs^{-1}.

Overall, any variety of Cygb has a higher maximal rate than Mb. The highest of the Cygb types is H81A and has a maximal rate greater than the WT (next highest) by 19.7 nMs^{-1}. The lowest is L46W but still 3.3-fold faster than Mb.

a)



b)

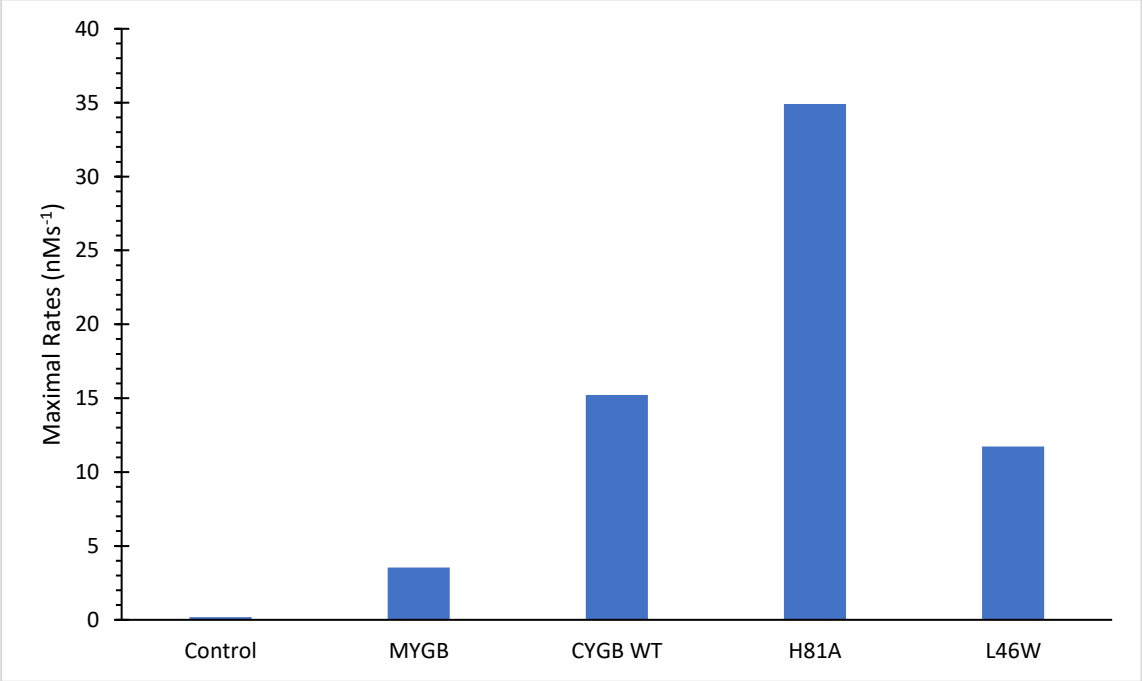


Figure 5.1.5.2- Oxidation rate of liposomes by Mb and Cygb variants shows increase in H81A and decrease in L46W. a) Graph showing the interactions between the different globins and globin variants with liposomes over time and b) the maximal rates of the proteins. Control is buffer and liposomes but no protein present. Absorbance was measured using a diode array spectrophotometer. Maximal rates calculated using their derivatives on Microsoft Excel.

5.1.6 Effects of the two mutants: L46W and H81A on nitrite reductase activity

NO-bound ferrous Cygb was produced by combining Cygb with nitrite in the presence of sodium dithionite in a pseudo-first order reaction (where the substrate concentration is in excess of the enzyme concentration and the enzyme-substrate complex is always full occupied). Figure 5.1.6.1 shows the results for a) WT Cygb, b) H81A Cygb and c) L46W Cygb. The WT optical changes show a hypsochromic shift of the Soret peak from ~429 nm to 422 nm, consistent with previously published data (Reeder and Ukeri, 2018). These spectra also saw a decrease in the alpha and beta peaks and an increase in absorbance after 570 nm. H81A saw a bathochromic shift of the Soret peak from 423 to 429 nm with a slight increase in the general spectra but less pronounced than the WT and the peaks not as affected. L46W also had a bathochromic shift from 425 nm to 431 nm and a small change in absorbance in the alpha peak.

Figure 5.1.6.2 shows the rates of the three different proteins, a) WT Cygb, b) H81A Cygb and c) L46W Cygb, with varying concentrations of nitrite. The Cygb variants ranged from having one second order rate constant to two since some were single exponentials and others double. The WT had a faster second order rate constant that was around 24.54 $M^{-1}s^{-1}$ and a secondary slower rate that was around 0.17 $M^{-1}s^{-1}$. H81A, the mutation which fixed the protein into a pentacoordinate confirmation (without exogenous ligands), had a fast second order rate constant of 578.6 $M^{-1}s^{-1}$, considerably faster than the WT and a second slow rate constant of around 96.5 $M^{-1}s^{-1}$. L46W conversely was much slower than the other two and only had one slow measurable second order rate constant which was ~ 0.44 $M^{-1}s^{-1}$.

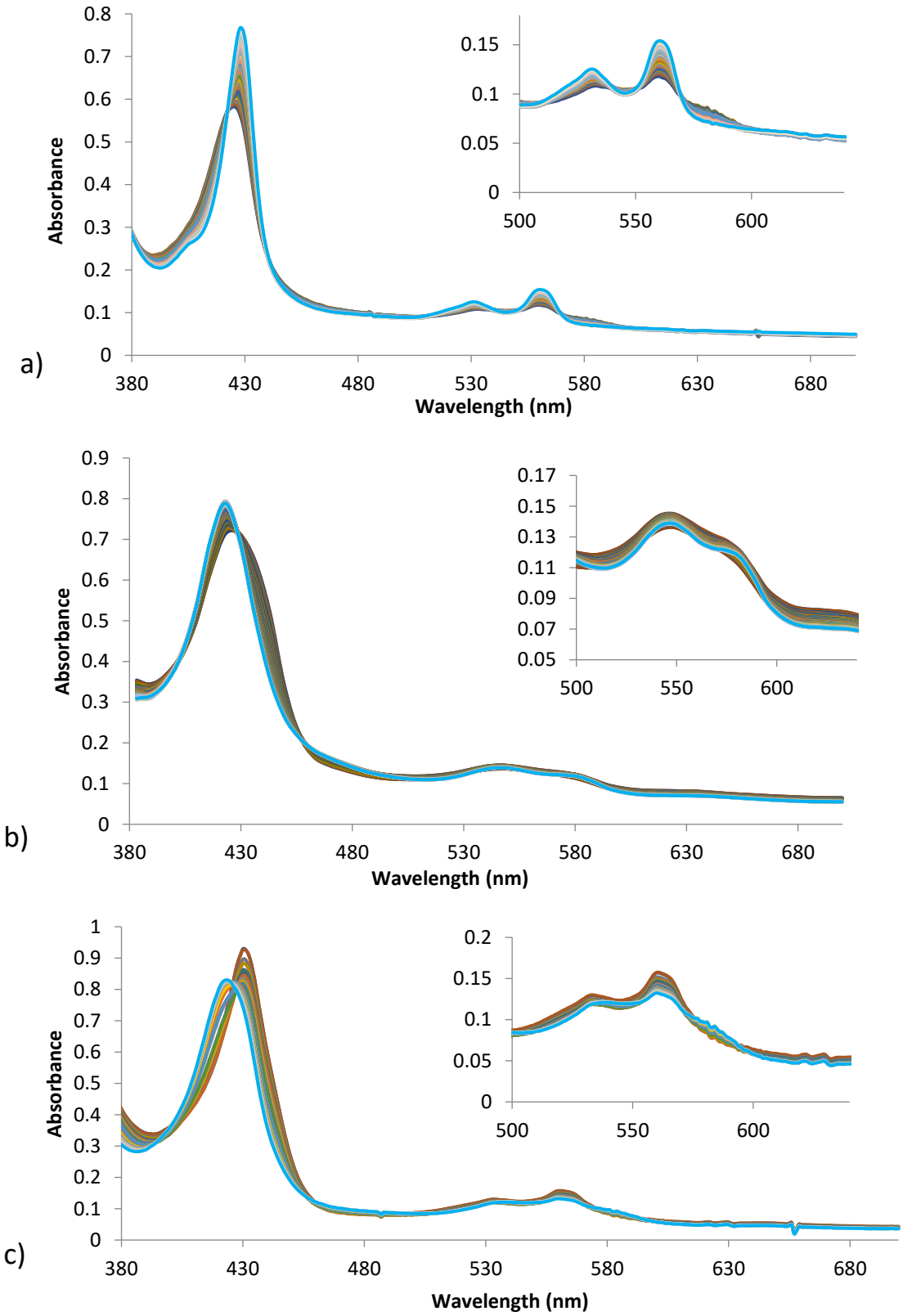
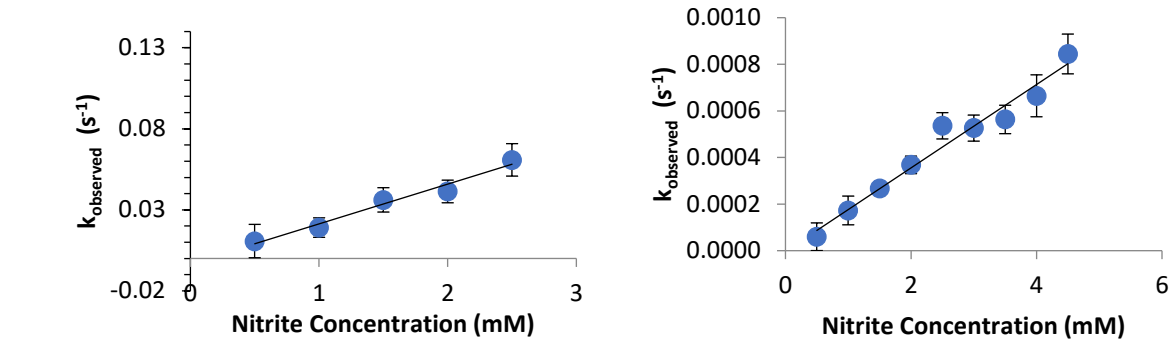
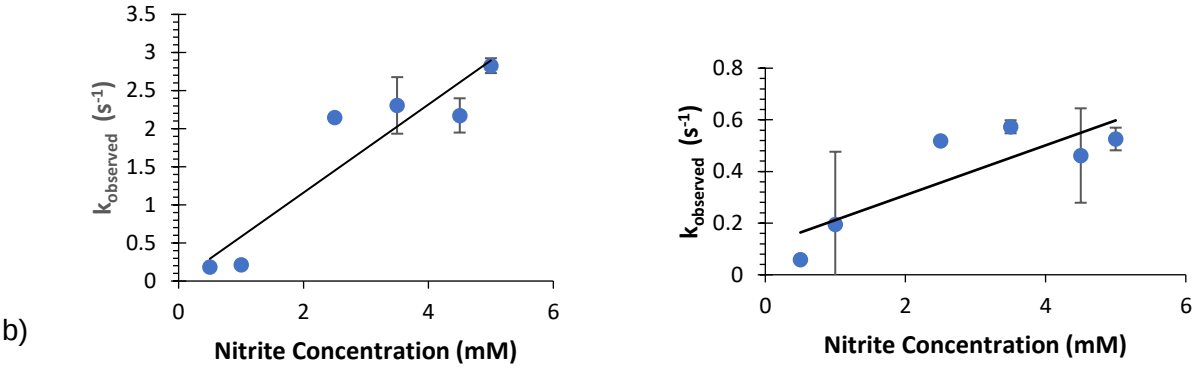


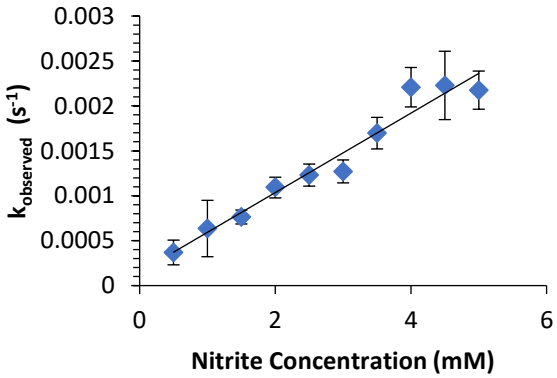
Figure 5.1.6.1- Optical changes of the spectra of a) WT Cygb, b) H81A Cygb and c) L46W deoxy Cygb. Spectra measured under anaerobic conditions (addition of sodium dithionite) on a diode array spectrophotometer and taken every second and then every 5 seconds after 60 seconds. Optical spectra taken with addition of 3 mM nitrite. Inset shows magnified visible region of the spectra.



a)



b)



c)

Figure 5.1.6.2- Cygb mutant H81A increases NiR rate and L46W decreases it, comparable to the WT. a) WT Cygb, b) H81A Cygb and c) L46W Cygb, the left-hand figures show the first fast reaction, and the right-hand side shows the second slower reaction. 5 µM protein was added to varying concentrations of nitrite under anaerobic conditions (in the presence of sodium dithionite) and measured on a diode array, spectra taken every second and then every 5s after 60s. Mean of 3 repeats.

Table 5.1.6.2- Comparison of the second order rate constants of the three cytoglobin variants compared to other globins. Both the first and second reaction rates were included in this data set. Equine Mb and Ngb information from Li *et al.,* 2012 and Tiso *et al.,* 2011 respectively.

Protein Type	First reaction k_{NR} ($m^{-1}s^{-1}$)	Second reaction k_{NR} ($m^{-1}s^{-1}$)
Wild-type Cytoglobin	24.54	0.17
H81A Cytoglobin	578.6	96.5
L46W Cytoglobin	N/A	0.44
Equine Myoglobin	2.9	N/A
Neuroglobin	0.12	0.062

5.1.7 Effects of the different cytoglobin variants on nitric oxide dioxygenase activity

Cygb is known to efficiently regulate the rate of NO consumption in an oxygen-dependent manner, through the metabolism of NO. This system works alongside the cellular reducing system (Liu *et al.,* 2017).

Using a stopped flow as displayed in section 2.2.1.5.1, different amounts of NO (generated from proliNONOate) were added to oxyferrous WT Cygb, L46W and H81A Cygb.

Figure 5.1.7.1 shows the WT Cygb spectra at 40 µM obtained from a global analysis fit to an A to B (or A to B to C) mechanism where the pure spectra of species A and B (and C) are obtained. The spectra observed at 1.2 ms after addition of NO is ferric Cygb, the reaction with NO is already completed by this point and is too fast to be observed on the stopped flow apparatus. The small peak at ~625 nm is typical of some pentacoordinate ferric protein before it relaxes back to hexacoordinate with the distal histidine.

Figure 5.1.7.2 shows the fitted spectra of the L46W Cygb protein at 60 µM NO. The data shows that the reaction of NOD with the protein is clearly visible (unlike WT), and therefore slower in this reaction, since the mutation created is designed to slow down the reaction long enough for it to be observed and measured by the stopped-flow. The first species, *a*, has two peaks at approximately 540 and 580 nm and is characteristic of the oxyferrous species (see section 5.1.2). The final spectrum indicated that the reaction did not go all the way to completion under the conditions of the experiment, so it is possible that NO is being consumed within this reaction by an unknown mechanism.

Figure 5.1.7.3 shows two data sets for the reaction of H81A with different concentrations of NO, a) 40 µM and b) 100 µM. The expected reaction was that, like other pentacoordinate proteins Mb and Hb, the NOD reaction would be too fast or almost too fast for stopped-flow, with the ferric protein being the spectra observed at 1.2 ms. The figure shows that H81A has some unexpected results in that the reactions are different depending on the concentration of NO within the reaction. Both 20 and 40 µM (*a*) showed three distinct different species involved within the reaction.

Form *a* is oxyferrous and has two peaks at ~530 nm and ~570 nm (as seen in section 5.1.2), this then transitions into the ferric pentacoordinate intermediate which is distinguishable with its three peaks at ~540, 580 and 625 nm. The third species formed has two peaks at ~550 and ~580 nm. This appears to be an NO-bound ferric species. On the other hand, the higher concentrations, 60 µM and 100 µM (*b*) went straight from the oxy to the ferric-NO form with no visible intermediate. The oxy form, form *a*, again has two peaks at ~530 and ~570 nm, as does form *c* whereas *b*, the ferric-NO form has two peaks at ~550 and ~580 nm. The presence of the isosbestic points shows one species transforming into another with no optically distinct intermediate species or two different sub-populations going at different rates.

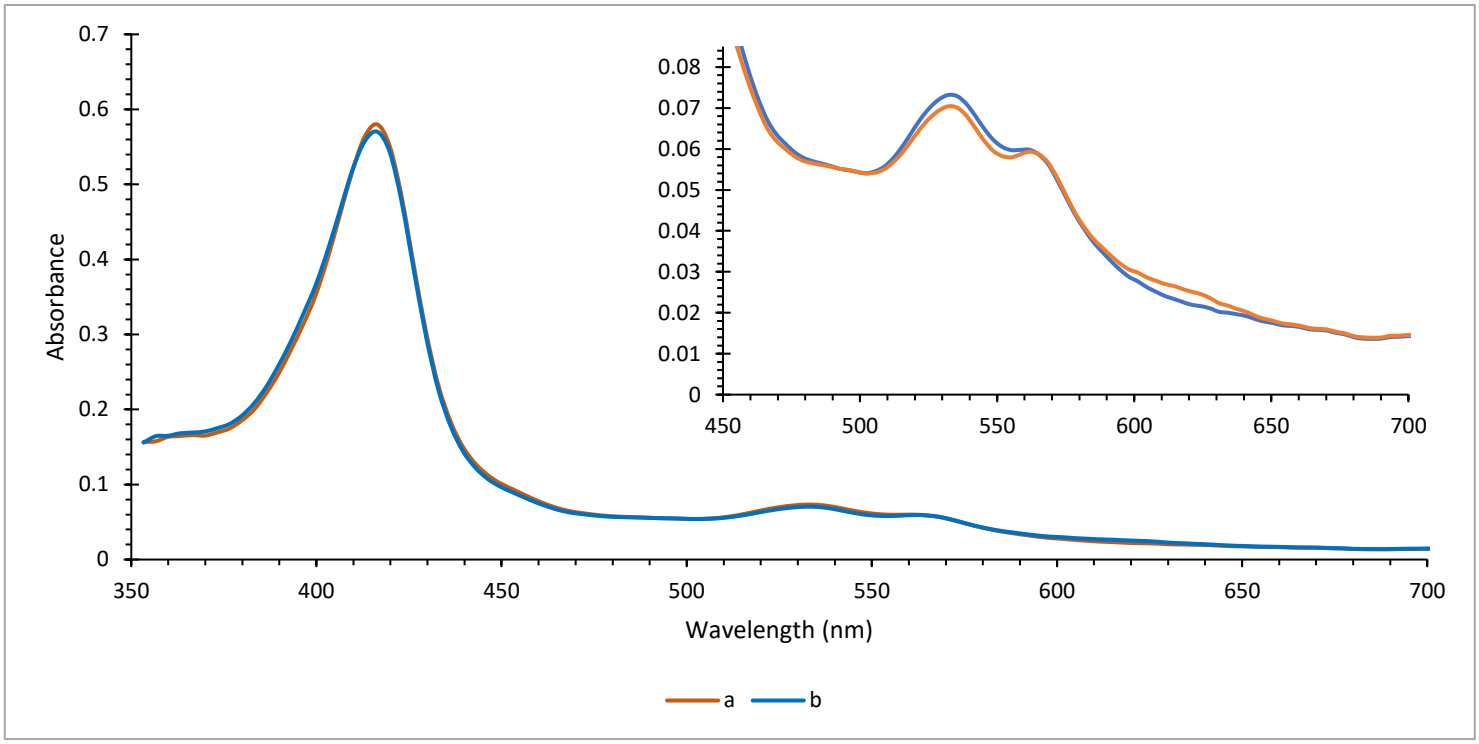


Figure 5.1.7.1- Fitted spectra of the WT Cygb with 40 μM NO. Data modelled using one-step reaction on ProData Kineticist before being exported into Microsoft Excel and the graphs produced. Inset is magnified visible region area.

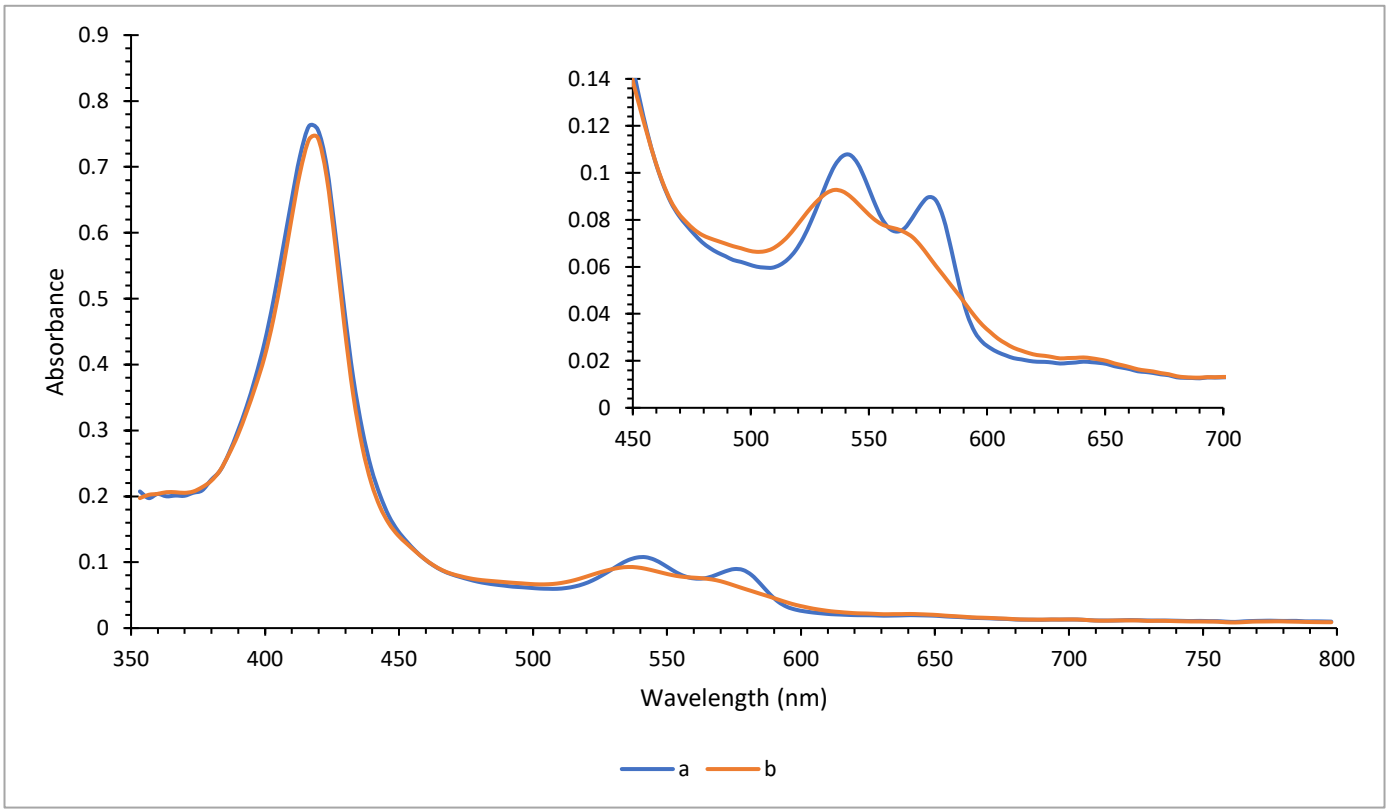
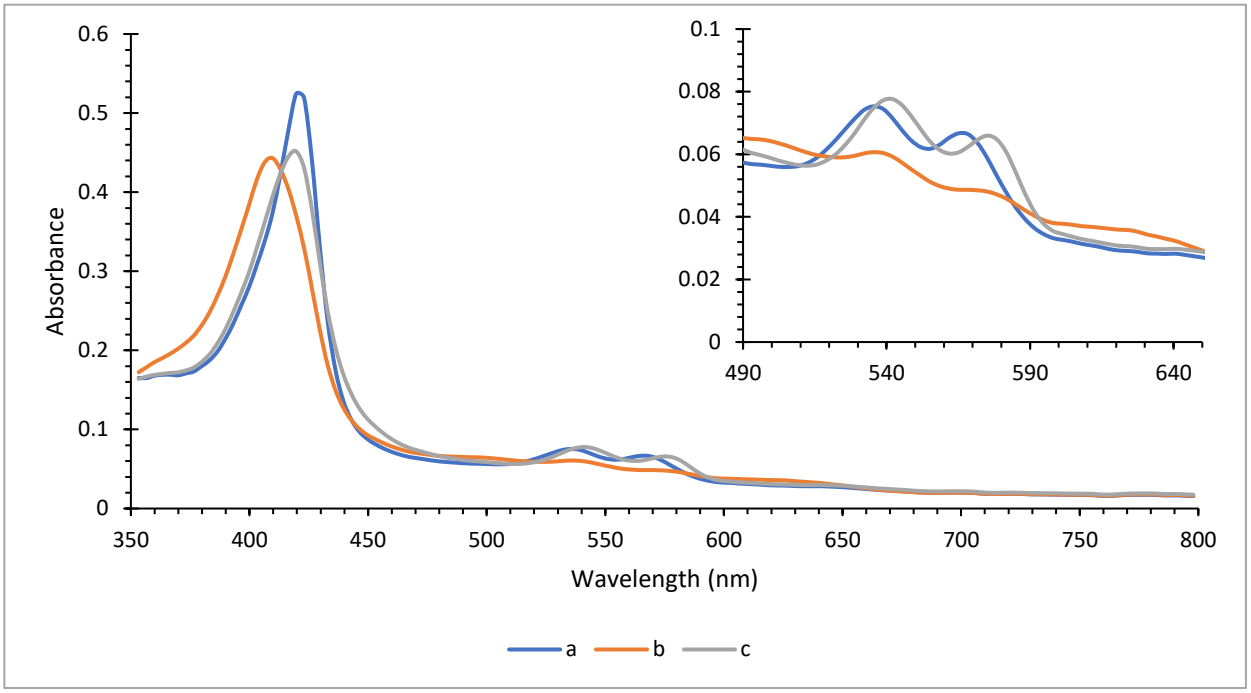


Figure 5.1.7.2- Fitted spectra of the L46W Cygb mutant with 60 μM NO. Data modelled using one-step reaction on ProData Kineticist before being exported into Microsoft Excel and the graphs produced. Measured on a stopped flow. Inset is magnified visible region area.

a)



b)

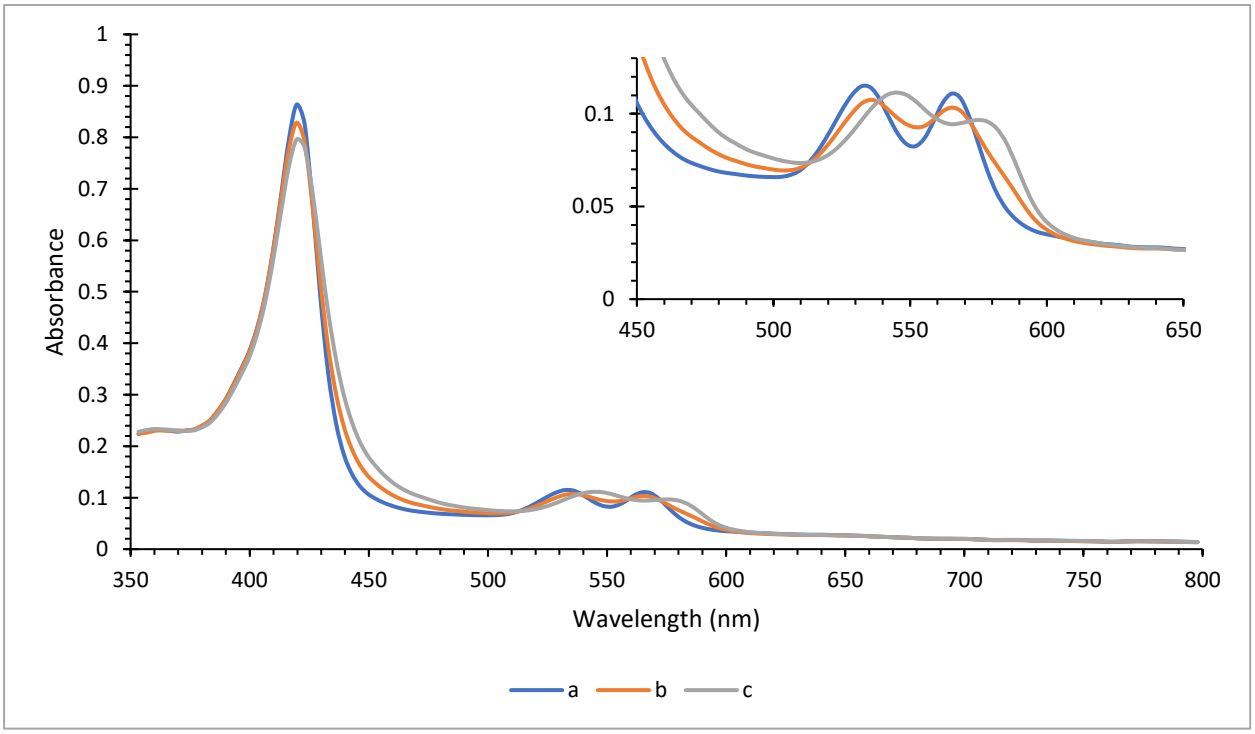


Figure 5.1.7.3- Fitted spectra of the H81A Cygb mutant with a) 40 µM and b) 100 µM NO. Data modelled using two-step reaction on ProData Kineticist before being exported into Microsoft Excel and the graphs produced. Inset is magnified visible region area.

5.1.8 Reduction of cytoglobin variants (wild-type and mutants) by ascorbate

The interaction between Cygb's NOD activity and ascorbate has been previously cited as a supportive one whereby ascorbate reduces oxidised Cygb following reduction of NO to nitrate (Gardner *et al.,* 2010).

Figure 5.1.8.1 shows both the optical traces and the time courses for the addition of ascorbate to the three Cygb protein variants: WT, L46W and H81A. The optical spectra are all similar and all show an α and β peak in addition to the Soret at approximately 544 nm and 580 nm. All three show a similar hyperchromic shift with H81A displaying the highest absorbance and the WT showing the lowest, as also shown in Figure 5.1.2.1.

The time traces show that H81A and L46W have a similar pattern in terms of both having a fast rate and then saturation. The WT has a much steadier and slower start than the mutants and becomes saturated quite quickly.

The data shows that there is a single exponential involved in all of the reactions. In addition, when looking at the absorbances on the time traces, that of L46W is significantly lower than H81A and the WT at around 0.11 s^{-1}. Out of the remaining two Cygb proteins, H81A is the highest at approximately 0.94 s^{-1} and the WT is approximately 0.67 s^{-1}.

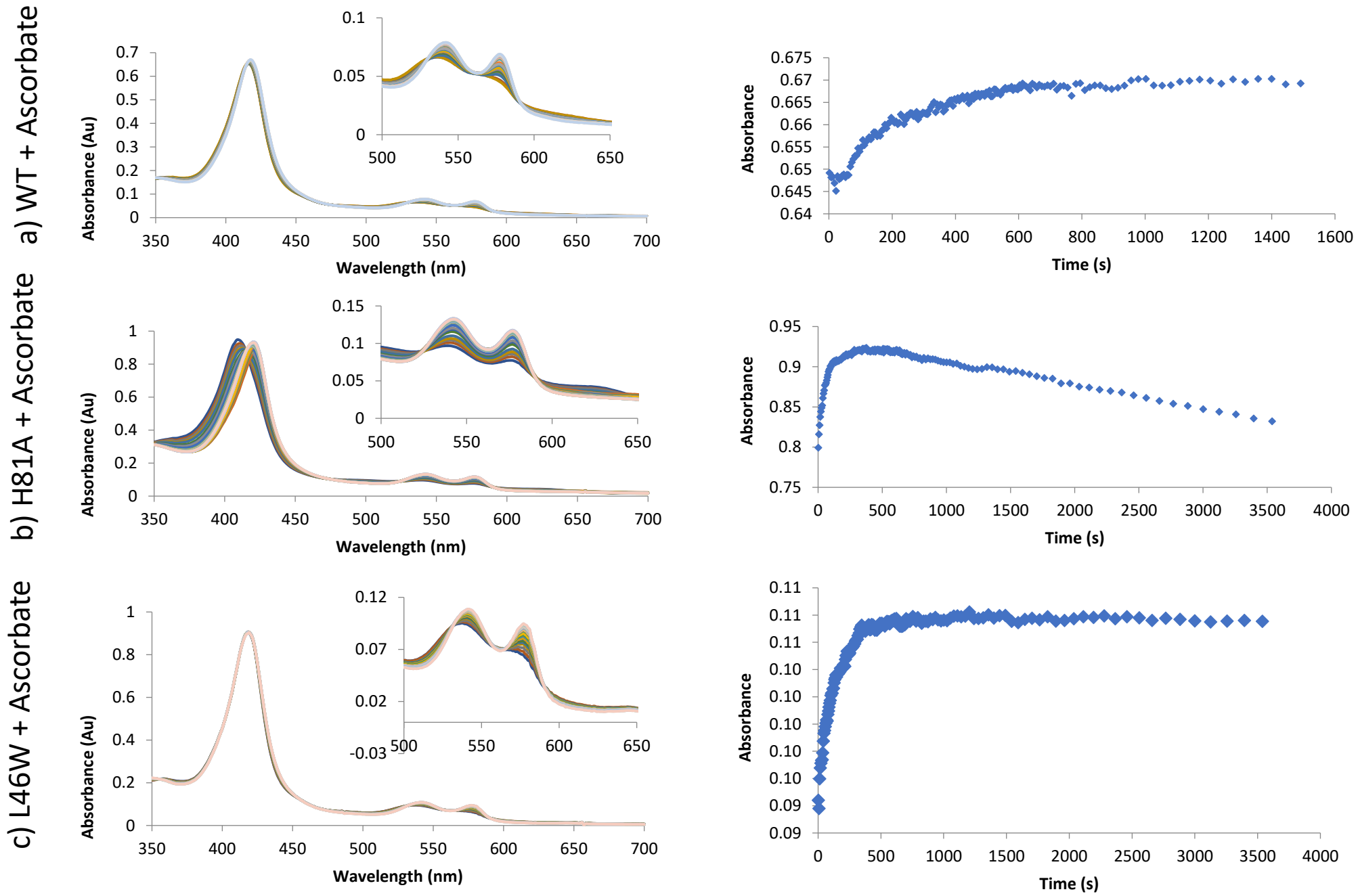


Figure 5.1.8.1- The reduction of WT Cygb and two mutants, by ascorbate. a) WT Cygb, b) H81A Cygb and c) L46W Cygb; left hand-side shows the optical spectra and right-hand side shows the time courses. 5 mM ascorbate was added to 10 µM protein and the optical spectra and time courses taken every 5 seconds and then after 60s, every 10 seconds, using a diode array spectrophotometer. Inset is magnified visible region of the spectra.

Chapter 6: Discussion

The aim of this project was to further characterise the cytoglobin (Cygb) protein and understand its potential functions, with emphasis on redox sensitivity. The study compared results obtained from both the WT as well as two mutants (H81A and L46W). These two mutants were created based on previously published studies on other globins such as Ngb and Mb which produced results that were found to be interesting (Tejeo et al., 2015; Beckerson et al., 2015; Yi et al., 2009). It was hoped that these findings might be replicable in Cygb studies. It investigated nitrite reductase and NOD activity, lipid and liposome interactions as well as reduction by ascorbate. The project additionally aimed to investigate the protective effects of Cygb when subjected to conditions which create oxidative stress or damage to the cells, such as hydrogen peroxide and chemotherapeutics.

This discussion critically analyses the outcomes of these investigations and provides further thought on the impact the findings have. It also addresses the direction in which future work should be aimed at.

6.1 Changes in protein structure as a result of mutation additions

The DNA sequencing confirmed the mutations were present in the correct location and were the only mutations present. Since the primary amino acid structures of the protein (both WT and mutants) were now confirmed, it was important to evaluate how the mutations might affect the overall structure of Cygb and potentially its function. The monomeric WT protein was visualised using Deepview software and the two mutation sites of interest were highlighted (Figure 3.3.1). In addition to this, other key residues found during research to be essential in the structure were also brought to attention, these were positions: C38, F60, C83 and H113. Two of these sites (C38 and C83) are cysteines responsible for the formation of intramolecular disulphide bonds (though they can also form intermolecular disulphide bonds), one is a highly conserved amino acid in all globins which stabilises haem binding (F60) and the other is the proximal haem ligand (H113). Another image was created as a comparison showing the same

six residues highlighted on the structure, only this time mutating the leucine at position 46 to a tryptophan and the histidine at position 81 to an alanine. The most notable distinction is the size difference between leucine and tryptophan. The addition of the large indole side chain moved the amino acid side chain closer to the edge of the haem ring. It was predicted that the mutation would cause a change in the structure, resulting in a decrease in the size of the area at the back of the haem pocket. In the mutation of H81A, the alanine now replacing the distal histidine appears to be sufficiently away from the haem iron that the protein will be pentacoordinate, which would significantly alter the binding of exogenous ligands to the haem iron.

Previous studies on Cygb, where the distal histidine has been mutated, revealed a lower peroxidase activity and enhanced radical-induced degradation although these were on mutations that differed to the one focused on in this study e.g. H81Y and H81V (Beckerson *et al.,* 2015). The evidence provided by the protein structure models in addition to the results found by Beckerson *et al*. show that this distal histidine is likely to modulate the stability of the ferryl oxidation state. Likewise, with the mutation of L46, as mentioned earlier, it was observed that the large change in size between the two amino acids changed the internal cavity of the haem domain due to the partial blockage. This is supported by the findings published by Zhao and Du (2016) who found that mutations on this particular residue affected dynamic properties in the binding of CO ligands to Cygb. It also affected the haem motion, internal cavity rearrangement and fluctuation in the CD-D-E and EF loop regions.

In addition to understanding the influence that the individual mutations have on the protein structure, an investigation was undertaken to examine the likely effect that specific mutations would have on the structure of other globins based on the change in size of the side chain and how this would appear when mapped onto the mutated Cygb. Looking at the image comparisons in Figure 3.3.2, the largest similarity between the two mutated globins was found to be Cygb and Mb. This is not surprising since these have high sequence homology, around

90% in some species (Xi *et al.,* 2007) and are regarded to be the closest relatives in the globin family (Wu *et al.,* 2016).

6.2 The relationship between cytoglobin mutations and cancer- a bioinformatical investigation

6.2.1 Key amino acid mutations and the presence of cancer

Seeing that there is a large functional effect from these mutations of different reactivities, it was considered whether these mutations (H81A and L46W) exist in diseases such as cancer. As in Figures 3.3.1 and 3.3.2, the two mutations were investigated along with four other key residues found to be important in the structure (C38, C83, F60 and H113).

Only one out of a possible 84,840 patients studied had one of the six key residues found to be associated with cancer. In this case the substitution of cysteine 38 to tyrosine and the diagnosis of mixed cancer types. Mixed cancer types occur when cancers develop in two different types of cell from one category or multiple categories for example carcinosarcoma which consists of a mixture of carcinoma and sarcoma (Gotoh *et al.,* 2019). No published research could be found which links this cancer to the cysteine residues or the disulphide bonds.

In terms of the frequencies of the specific mutations in the different cancers, the vast majority appeared singly with five appearing twice each. These were at positions 13, 22, 82, 116 and 182. R13 appears to produce hydrogen bonds with other arginine residues that form the fluorophore binding pocket in the protein (Tangar, A., 2018) and K116 helps to stabilise phosphate groups found in molecules such as PIP3, a second messenger in cellular response systems, (Tejero et al., 2016). The other three have either not been tested yet or there were no significant results to report.

Though no evidence was found to suggest that either of the two mutation sites (H81 and L46) were implicated in cancer, let alone the specific amino acids mutated within the study, the positions on the primary globin chain were analysed. It showed that the distribution of cancer-

associated mutations in Cygb is incredibly diverse. The mutations largely spanned the entire primary structure and were very spread out. Since there were no identifiable areas where there was a higher frequency of mutations present, a current conclusion was derived that there is no identifiable hotspot in the primary amino acid structure for diseases such as cancer.

6.2.2 Frequency of cytoglobin mutations across different cancers

Had there have been an identifiable hotspot, it could have been inferred that this area of the protein was potentially responsible for the presence of the cancers or a contributing factor. Since this did not yield any significant results, the frequency of mutations in specific cancers was observed. The results showed that although most of the cancers had a relatively low incidence frequency, some cancers such as nerve sheath tumour and cancers of unknown primary origin were particularly high. Cancers of unknown primary origin are cancers in which the primary tumour's location is not identifiable (Varadhachary, 2007). The very high level of mutations in nerve sheath tumours is interesting as, although CYGB is expressed exclusively in the cytoplasm in tissues such as supportive tissues; in mouse neurons, CYGB is localised in both the nucleus and cytoplasm (Reuss *et al.,* 2016). This altered localisation of CYGB could be the reason that there is such a high frequency of mutations in this particular cancer. This information can be taken into account as human Cygb and mouse/rat Cygb have a high Cygb homology (Hundahl *et al.,* 2013). Although it is possible that the reason this specific cancer had such a high frequency is because it had a very small sample size and therefore the presence could simply appear amplified. Cancers of unknown primary origin are a fairly unreliable group to analyse since they could all be originating from a similar location or could be distributed very evenly among the other cancer primary locations.

Irrespective of this, the areas where the cancers were located are widely spread around the body. For the majority, there is a low-level appearance of Cygb mutations in a lot of different cancers such as leukaemia and renal clear cell carcinoma. This supports the general low-level expression of Cygb in most human tissues.

6.2.3 The staining of cytoglobin in human tissues of healthy patients and those with cancer

The presence of CYGB can be seen in tissues by an immunopositive response to antibody staining of the cells (Figure 3.4.2.1). In two different cancers: breast and prostate, a comparison between the staining of healthy and cancerous patients was made. For prostate cancer, the results were conclusive in that none of the healthy patients had any immunopositive staining and all of the cancer patients had staining (Figure 3.4.2.2). There was a 1:1 distribution between those with low levels of staining and those with medium levels. In breast cancer, the results were also quite conclusive with all healthy patients having no staining, though some cancer patients also had no staining visible.

Characterising the staining into groups visually is a similar method to that described by Emara *et al.,* whereby two observers noted the overall tissue staining and the presence of positively stained focal regions and then these were assigned a score. The limitations of their method were described as the lack of uniformity in expression throughout the tissue, with some distinct cell types or tissue structures showing strong staining and the remainder with weak or no staining (Emara *et al.,* 2010). The images and data used in this investigation were obtained from The Human Protein Atlas but it was not possible to determine whether or not this was taken into account or whether an average was taken, which would explain why there was no high levels of staining present despite published data citing that the ductal cells of the breast were very strongly stained (Emara *et al.,* 2010). Additionally, CYGB is expressed in very low levels in most tissues so it can be difficult to determine whether there is CYGB present or not. Having considered the subjective nature used to determine whether or not there is staining present and to what degree, it has been decided that there is likely some discrepancy between the staining intensities of the tissues but unlikely in those that do not contain any staining. This still allows a conclusion to be made that CYGB levels were undetectable in healthy prostates, but expression appears to be upregulated in the majority of breast and prostate cancers. This could suggest that CYGB is important in tumour development or progression.

6.2.4 Survivability of cancer patients with high or low cytoglobin expression

Using mRNA gene chip data sourced from KMPlot (Nagy *et al.,* 2018), it was investigated whether CYGB levels correlate with cancer patient survival (Figure 3.4.3.1). The analysis indicated that three of the cancers (breast, ovarian and gastric) showed significant difference in patient survivability. There was no significant difference in lung cancer patients with high or low CYGB expression.

The data for ovarian cancer, though significant, does not appear entirely conclusive. Up until approximately 100 months, low expression showed better survivability, after this time both the high and low CYGB expression levelled out to an approximately even level and then there is a lack of data past ~170 months. This makes it difficult to identify whether the different expression levels are both equal in terms of patient survivability after an extended period of time or whether further events might have occurred.

Out of interest the data for breast cancer with CYGB was compared to data from three other globin members: NGB, HB1 and MB in the same cancer. This found that for two of the globins: CYGB and HB1, higher expression yielded significantly better survival.

The data as a majority shows that those with a high CYGB expression appear to have better survival in cancer. This was supported by published data that showed that low CYGB expression contributed to tumour reoccurrence and lower survivability for patients with glioma (Xu et al., 2013).

This seems to contradict the argument that CYGB cells aid in chemoresistance unless the patients were not being treated with chemotherapy. However, other studies have shown that CYGB can act as a tumour suppressor which would explain the data (Feng et al., 2018).

6.2.5 RNA expression levels in different human tissues compared to other globins

The data found that the difference in expression levels in CYGB influenced the outcome of cancer patients. Therefore, it was investigated which tissues have greater RNA expression for CYGB as well as MB and NGB. It showed that the highest levels of expression were seen in MB and the lowest in NGB (Figure 3.1.1). The tissues in which the highest levels of MB were seen were the left ventricle, atrial appendage of the heart and skeletal muscle which is not unexpected since MB is localised to certain tissues (Kawai *et al.,* 1987). CYGB was also as expected with low level distribution across a range of tissues (Fago *et al.,* 2004). NGB is generally only found in specific tissues such as those in the nervous system and so this explains why most of the tissues show no expression present (Emara *et al.,* 2010).

6.3 Conservation and alignment of key amino acids in the cytoglobin structure across different species

In the previous section, most of the key amino acid residues were not found to be mutated in cancer. Therefore, the conservation of these particular regions in other species was considered. The investigation found that three of the residues: F60, H81 and H113 were 100% conserved across all animal species analysed (Figure 3.2.1). Additionally, C38 and L46 had a high percentage of conservation. C83 was surprisingly diverse considering that C38 was highly conserved so this was unexpected. The residue had a total of 10 different variations in amino acid with the most common being cysteine and arginine. It has been shown that the intramolecular disulphide bond created between C38 and C83 increased the P_{50} of Cygb for oxygen from 0.1 to 1 Torr (Mathai *et al.,* 2020). The authors also modified the position of the distal histidine relative to the haem domain and increased its off rates from the iron (Mathai *et al.,* 2020). This would mean that if O_2 affinity was confirmed to be a function of the protein, those species with a varied version of this residue would require an alternate mechanism to improve the oxygen affinity and also the ligand binding rate. Further investigation did not show any links between those with a variation of the C83 and any particular animal classes for example those which are aquatic, so this is still unknown in this study.

Despite showing similar data to Figure 3.2.1 in terms of allowing the different amino acid residue variations to be seen, Figure 3.2.2 also allows the surrounding residues to be examined to see how highly conserved the entire region is. Due to the essential nature of the distal histidine in allowing exogenous ligand binding, it was unsurprising that the completely conserved distal histidine also had a highly conserved region surrounding it. L46 and C38 are not 100% conserved and neither are those surrounding them, but they all have very few variations with the majority being leucine and cysteine respectively. C83 was the most diverse amino acid residue of the six studied and the region is also poorly conserved with a huge number of different variants. This could be an indication of poor alignment or that many species simply have different mechanisms to support them than humans do.

Having understood the conservation of specific regions of Cygb, it was considered how conserved the whole Cygb gene is across various species (Figure 3.2.3). Not unexpectedly, one of the species with the highest homologies to humans was the mouse with whom we share 85% sequence similarity in coding regions and some genes are up to 98% identical (Batzoglou *et al.,* 2000). This high degree of homology supports confidence that tests being conducted on mice in clinical trials may have similar effects in humans. Figure 3.2.3 is only a small representation of the relationships between humans and other animal species and does not represent all of the species sampled within Figures 3.2.1 and 3.2.2. The decision to take a sample of the species was simply down to display purposes.

6.4 The protective effects of cytoglobin-transfected cells

6.4.1 Cytoglobin transfection optimisation

A dose curve of geneticin (G418) was created in order to find the best concentration for the MCF7 cells used to create stable cell lines (Delrue *et al.,* 2018). The highest concentration was still not enough to kill all of the un-transfected cells. MCF7 cells are a much more slow-growing cell line than cell lines such as HEK293S which usually results in a better kill rate (Delrue *et al.,* 2018) and so it was surprising that these cells were not killed off. Other studies

cite much lower concentrations (800 µg/ml) of G418 killing off all cells compared to the maximum concentration in this study (Ghasemi *et al.,* 2016).

Unfortunately, due to the COVID-19 pandemic, these cells lines were not continued and so the optimal G418 concentrations have been identified for future studies that aim to generate stable clones using these cell lines.

6.4.2 Protective properties of cytoglobin-transfected cells when treated with hydrogen peroxide

It was hypothesised that WT CYGB would provide a protective effect for the cells exposed to harsh environments that cause significant stress. Hydrogen peroxide is an oxidising agent that rapidly produces reactive oxygen species and other radicals that cause oxidative damage (Mahaseth and Kuzminov, 2017). The effect of the H81A and L46W mutants in Cygb was largely unknown before the study. An initial dose response investigation was carried out to determine the best concentrations of hydrogen peroxide to use. Research showed that hydrogen peroxide had a paradoxical effect on some cancer cell proliferation and migration depending on the concentration (Vilema-Enriquez *et al.,* 2016). A study showed that in MCF7 cancer cells, growth was enhanced at 1-10 µM but apoptosis was induced above 200 µM (Chua *et al.,* 2009) and in colon cancer cells, growth was enhanced at 10 µM but apoptosis was induced at 1 mM (Park *et al.,* 2006). This research allowed a good estimate for the best hydrogen peroxide concentrations to use.

The data showed that the presence of CYGB had the predicted effect, the viability of those transfected with CYGB was significantly elevated when compared to the control cells. The WT compared to the two mutants was also significant proving that not only did the presence of CYGB make a difference in the cell's ability to protect themselves from the hydrogen peroxide, but that the positions of the mutations also contributed. The experiment was repeated again with the same concentrations for seven days and the data supported the previous findings in terms of the control cells having a significantly lower EC_{50} than the transfected cells. The EC_{50}

values of the control were lower when incubated for seven days compared to four days although the WT compared to the two mutants no longer yielded significant differences.

Some therapies such as radiotherapy destroy cancer cells by triggering the activation of death signals via the generation of ROS (Kim *et al.,* 2019). These results suggest that cells with higher levels of CYGB could be protected from this therapy, making it potentially ineffective.

6.4.3 Protective properties of cytoglobin-transfected cells when treated with chemotherapy drugs: docetaxel and paclitaxel

In addition to the assays undertaken with hydrogen peroxide, cell viability assays were also carried out with the chemotherapy drugs: docetaxel and paclitaxel. These are not ROS-inducing forms of chemotherapy, so this data provided valuable insights into whether Cygb was just resisting the effects of the ROS or just drugs in general. The crystal violet chemotherapy data (Figure 4.5.1) yielded unusual and unexpected results when it came to one of the transfected mutants in particular: L46W. For many of the studies with the different cell types and different cancer drugs, L46W-transfected cells were significantly different to the other transfected or mock transfected cells. This could suggest that this mutant creates a vulnerability in the cells to the chemotherapy drugs that was not seen in the other transfected cells (WT and H81A) and the control mock-transfected cells.

This mutation occurs at position 46 on the amino acid chain and is buried in the hydrophobic haem pocket. Other studies which observed mutations in this haem pocket region resulted in affected hydrogen bonds between key residues and salts or esters of propionic acid which resulted in decreased haem stability. Mutations changing L46 to an F (phenylalanine) or a V (valine) reduced the standard hydrogen bond distance between carboxyl oxygen atoms in the haem propionate and the protein's CD3 region by around two Angstroms in molecular dynamic simulations (Zhao and Du, 2016). Given the importance of this key residue on the integrity of Cygb's structure, this could suggest that this mutation impedes the stability of the protein's haem pocket though there is no evidence for this at this time.

The overall data showed that there was no significant difference between the control cells and the CYGB-transfected cells, with the exception of L46W. This is due to the fact that this class of chemotherapy drugs do not produce ROS as many others do, but instead interfere with the mitotic spindles. This indicates that CYGB might only affect those that produce ROS and also allows an assumption that patients with high CYGB levels could still use this as a viable treatment option.

This is further reinforced by data showing that CYGB overexpression has been seen to associate with tumour aggressiveness as well as higher chemotherapy and radiotherapy resistance (Oleksiewicz *et al.,* 2013). This would mean that for patients being treated with ROS-generating chemotherapies, the presence of CYGB could be a marker to predict whether these treatments would be more or less effective.

One of the complications with this study was that due to COVID-19 induced time restraints, the transfected cells were not stable cell lines as planned but transiently transfected cell lines. Since most of these studies were investigating short-term expression of genes or gene products as opposed to a long-term study, this type of cell line was still suitable for the experiment. The reason for using HEK293S cells in addition to the MCF7 cancer cells was due to the much quicker replication rate of the HEK293S cells making transfections much faster and allowed the collection of maximum data in a smaller time frame.

The crystal violet assay data was imported into Microsoft Excel and Prism and used to calculate EC_{50} and significance values. Though this was a suitable method for estimating the EC_{50} values, recent articles discuss the ability of these data analysis methods to both fit the data accurately and give potentially misleading approximations of cell proliferation inhibition. A new approach was described by Liu and Crawford (2018) which is based on a mechanistic model of cell division and death by simultaneously analysing apoptosis rates as a function of dose and an estimation of cell division. It is true that during the period of this investigation, results were largely based on the shape of the curve produced and assumptions were made on the replicative qualities of the cells and the inhibition patterns of the chemotherapy drugs

without considering that the inhibition of the cells could happen in physiologically unique ways. Using a more specific chemotherapy-specific method of calculating EC_{50} values would allow more accurate data with fewer assumptions about the parameters and variables.

6.4.4 Microscopic comparison of cells before and after chemotherapy treatment

When observing the cells (mock-transfected MCF7 cells before and after the addition of chemotherapy agents), it was apparent that there was a significant decrease in the number of cells present in the images (Figure 4.5.3) which were a good representation of the whole plate. Both images depicting the cells after chemotherapy drug exposure looked fairly similar since the two drugs used: docetaxel and paclitaxel, are similar in their mechanisms of action in that they both affect the cell's microtubules and therefore the ability to proliferate.

Paclitaxel is a microtubule-stabilising compound that induces mitotic arrest and, in a subset of the arrested population, cell death (Weaver, 2014). Docetaxel inhibits the microtubule depolymerisation and therefore arrests cells in the G(2)M phase of the cell cycle before inducing a signalling cascade that causes apoptosis within the cells (Pienta, 2001). There is evidence of cell apoptosis in addition to fewer cells present in the images.

Additionally, in hindsight, several images would have been taken at different magnifications as opposed to just one, and also daily for a period of seven days so that the effect could be mapped against time. Due to the pandemic, confocal microscopy was not available to examine the cell morphology in more detail. Despite this, it is apparent that both of the chemotherapy drugs have an effect on the MCF7 un-transfected cells which further backs up the crystal violet data shown in Figure 4.5.1.

6.5 Biophysical and biochemical characterisations of wild-type cytoglobin when compared to mutant variants

6.5.1 Optical spectra highlighting the differences between the wild-type and mutants under different conditions

In order to understand how the additions of different reagents or chemicals affected the spectra of the three different Cygb proteins (WT, H81A and L46W); spectra of the ferric, deoxyferrous and the protein following addition of CO were analysed.

The WT showed similar patterns and wavelengths to those reported in published literature for all three spectral types (Beckerson *et al.,* 2015) and L46W showed similar peaks at the same wavelengths. H81A had three peaks (an additional one presenting at 630 nm) which are typical of ferric pentacoordinate globins such as Mb and Hb, indicating that the H81A mutation changes the protein from a hexacoordinate to a pentacoordinate conformation.

Interestingly, another published study looked at the UV-visible spectra of two distal histidine mutations which instead mutated the histidine to a valine (V) and a tyrosine (Y) (Beckerson *et al.,* 2015). The H81V mutant showed only two peaks present rather than the three seen in the H81A. This comparison of the data collected in this investigation and that seen in the data published by Beckerson (2015), shows that the degree of pentacoordination is affected by factors such as the disulphide bond presence. The difference between the pentacoordinate and hexacoordinate is the binding (or not) of the histidine to the haem molecule. It would be useful in terms of future work to create distal histidine mutants with a much larger range of different amino acids so that a conclusion could be more easily attained.

The addition of sodium ascorbate to reduce the protein showed that all the oxyferrous Cygb proteins were the same when looking at the spectral peaks.

Additions of ProliNONOate to form the ferrous-NO species showed a difference in all of the different protein's spectra. The WT has two distinct peaks followed by a smaller flatter elevation in the spectra. H81A has two clearly defined peaks which is very different from its

standard three peaks commonly seen with this mutant. Though the WT and H81A have very clearly defined peaks, L46W is difficult to accurately separate into distinct peaks which makes analysis more difficult. One explanation could be that it is a shoulder, when a smaller peak partially merges with a larger peak (Rambo, 2017).

6.5.2 Binding of cytoglobin and its mutant variants to lipids

The relationship between Cygb and lipids has been previously cited as transforming the WT Cygb from hexacoordinate to pentacoordinate whilst in the ferric oxidation state (Beckerson *et al.,* 2015). This led researchers to speculate the potential role of Cygb as a lipid signalling molecule of oxidative stress (Reeder *et al.,* 2011). Though this work has been performed on the WT type protein, it has not been performed on the two mutant proteins involved in this investigation.

It was found that the result of the WT Cygb protein binding to the lipids was very similar to what was described in the published literature in terms of the 1:1 stoichiometry (Reeder *et al.,* 2011) observed in this investigation and the change from a hexacoordinate to a pentacoordinate state. H81A Cygb had a very similar affinity to WT Cygb.

L46W Cygb was expected to bind much like the WT Cygb but instead did not appear to bind to the lipids at all. It formed a slight linear relationship however the fractional saturation scale on the graph is in such small increments that this cannot be considered a significant change. The slight rise could be due to the denaturing effects of oleate over time which causes the protein to unravel and therefore produce a slight change. The spectra confirm this data, there are no additional peaks in the spectra and the large peak is simply due to the dilution factor.

For the WT Cygb and H81A Cygb, the lipid binding could significantly affect potential functional reactivities which require further exploration (beyond the liposome reactivity below). In the case of L46W Cygb, the lack of change in haem coordination does not necessarily mean that the lipids have not bind and so the effect of reactivities still requires further investigation.

6.5.3 Liposome oxidation- a comparison of myoglobin, wild-type cytoglobin and mutant variants

Liposome oxidation is important given the data obtained from the binding of Cygb and lipids as this is a more realistic *in vivo* scenario. Mb and the three Cygb variants all showed liposome lipids becoming oxidised (Figure 5.1.5.1). The oxidation of liposomes by WT Cygb and Mb has been previously referenced in other publications such as that by Reeder *et al.* (2011) and supports the data found in this investigation since they both had absorbance around 234 nm.

The formations of conjugated diene as a product of time are markedly different, depending on the mutation and protein. L46W Cygb most closely resembled Mb but had a slightly faster rate. H81A Cygb was much faster than WT Cygb but had a similar pattern. Both the Mb and L46W Cygb never appear to reach saturation and instead the reaction continues past the end-point of the experiment. Again, it would be useful to repeat this data and carry it out for a much longer amount of time to see if this would reach saturation. The maximal rates (Figure 5.1.5.2) show that there is a significant difference between some of the proteins. The H81A Cygb is much faster than all other proteins investigated by more than double. Conversely, Mb has a slow rate compared to all other CYGB variants.

The fact that H81A Cygb is much faster than the WT Cygb suggests that the pentacoordinate form is more reactive as the lipid data shows that there is no difference in the binding of lipids to Cygb. This may be related to the peroxidatic activity as the haem appears to be bleached much more rapidly and could be investigated further using a guaiacol assay. The L46W Cygb is much lower in activity but still roughly equivalent to Mb. There is a known relationship between Mb lipid oxidation and pathological conditions such as rhabdomyolysis followed by acute renal dysfunctions, and brain haemorrhages (Reeder and Wilson, 2005). Rhabdomyolysis-induced kidney damage involves direct interaction of Mb with mitochondria and potentially results in iron ions release from the Mb haem. This promotes peroxidation of mitochondrial membranes (Plotnikov *et al.,* 2009). Overall, if using these mutations *in vivo*,

they are likely to have marked changes on the ability to generate lipid-based signalling molecules.

6.5.4 Nitrite reductase activity comparison of wild-type cytoglobin, L46W and H81A

Multiple globins such as Mb, Hb and Ngb have been reported to reduce nitrite to nitric oxide (NO) within mammalian cells and therefore it is no surprise that reduction by Cygb is an area which is currently being investigated (Li *et al.,* 2012). The spectra obtained were consistent with those published by Reeder and Ukeri (2018). The rates of reduction were of particular interest; the WT Cygb had both a first reaction that was fast and a secondary reaction which was slower. The rates of these were 24.54 and 0.17 $M^{-1}s^{-1}$ respectively. Comparatively, H81A Cygb had a first rate which was around 24-fold faster than the WT Cygb and even more significantly had a second rate which was 567-fold faster than the comparative WT Cygb. This could be due to the H81A mutant being locked into the pentacoordinate position so the NO can bind straight to the haem without the histidine requiring detachment which occurs in the WT Cygb. L46W Cygb should theoretically not have an impact on the rate since the coordination state remains unaffected but this mutation, which partially blocks the haem pocket, saw a much slower rate than its other variants. This protein saw only one measurable reaction and this single rate was around 0.44 $M^{-1}s^{-1}$ which is considerably slower than the first reaction of the WT Cygb but faster than the slow rate. The reason that some of the protein variants have more than one rate could be due to the fact that there is a fraction of the protein with the disulphide bond and a fraction without. This is corroborated with the published data (Reeder and Ukeri, 2018). Alternatively, it could be that some of the protein is in the hexacoordinate and some in the pentacoordinate conformation which is also related to the disulphide bonds. This seems less likely due to the higher rate constant observed for the H81A.

6.5.5 Nitric oxide dioxygenase activity- an investigation into concentration dependence

Cygb has been reported to have NOD activity which closely resembles that of Mb and Ngb. Cygb scavenges nitric oxide (NO) by means of oxygen-dependent nitric oxide deoxygenation (NOD). In the NOD, oxy-ferrous Cygb interacts with NO to convert it to nitrate, converting the protein to ferric cytoglobin (Tse, 2015). In hypoxic conditions, NO is generated via nitrite reactivity activity whereas under normoxic conditions, NO (signalling molecule and relaxant of the vasculature system) is consumed via the NOD activity (Ukeri, 2020).

By metabolising NO, Cygb is able to regulate NO consumption in an oxygen dependent manner. During this investigation it was discovered that none of the three Cygb types behaved consistently and one was different depending on the NO concentration used.

The WT Cygb showed two species present, with the ferric Cygb being the predominant product since the reaction was so fast that it was already completed by the time the stopped flow was able to produce a spectra ($>1x10^8$ $M^{-1}s^{-1}$). H81A appeared to be concentration dependent with an intermediate being present at lower NO concentrations. The L46W Cygb equivalent mutation was shown to decrease the rate of NOD in other proteins (Tiso *et al.,* 2011) and it was seen that Cygb does indeed result in a marked decrease in NOD. Therefore, the L46W mutation would decrease the NOD activity *in vivo*. If NOD reactivity was a major pathway to protecting the cells, then this mutation would significantly affect the reaction. Of course, other data shows that this mutation has other effects which have to be taken into account.

There were limitations with this experiment in that due to the mechanism of the stopped flow apparatus and accompanying software package, when attempting to obtain the rates, despite ensuring the number of iterations was as low as possible, the *k* values varied enormously on the same samples which made it difficult to determine a true value. Irrespective of this, by analysing only the spectra and time traces, it can be sure that the data is as accurate as possible once the dead time is taken into account.

6.5.6 Reduction of cytoglobin by ascorbate

Ascorbate has previously been described to reduce oxidised Cygb-bound NOD. The data shows that all three proteins show rapid ferric reduction. Some mutants even more than the WT. For the NOD activity, the protein can only function properly as an NOD if there is a system to reduce the protein from ferric to ferrous. For *in vivo* reactions this has been postulated that the reductants are NADH or cyt b5/cyt b5 reductase. Here ascorbate could function to re-reduce the protein (as long as the cells are in a reduced state). Furthermore, the use of the mutants for studying the function of Cygb in cells would not interfere with Cygb re-reduction following either NOD or other oxidative reactions.

6.6 Conclusion

In conclusion, this study (given all the limitations which came with a global pandemic) still brought new insight into a very interesting protein which yields huge potential in so many different areas of biology. It was clear that mutations such as those used within this study could help elucidate the molecular mechanism of Cygb *in vivo* through either stable transfections or using CRISPR in cells and animal models. The data from these would not be simple to interpret since the original hypothesis that a singular mutation affects a singular function was not found to be correct and the mutations had multiple effects. This could still be a part of a multi-vectored approach to determining the physiological or pathological (in cancer) functions of Cygb.

The investigation provided some substantial evidence that Cygb plays a role in protecting cancer cells from the damaging effects of substances such as hydrogen peroxide, and inferred protection against chemotherapeutics that create ROS since CYGB did not appear to have an impact on chemotherapies that altered cancer in other ways. Since treatments such as radiotherapy create ROS in order to destroy cancer cells, the level of CYGB expression in the targeted cells could allow a predictor for the success of the treatment.

6.7 Future work

Cancer, in particular prostate and breast cancer, are devastating diseases which affect a substantial proportion of the world's population: ~1.3 and ~1.7 million people per year respectively (Bray *et al.,* 2018; Aune *et al.,* 2012). The discovery and subsequent development of therapies has meant that these diseases are much more treatable than ever before, but chemotherapy resistance is a common occurrence making disease management complicated. Due to this, it is more important than ever to fully understand the mechanisms involved with cancer cells overcoming the harmful effects of these drugs and protecting themselves against apoptosis.

With this in mind, the most promising results were those of the transfected HEK293S cells treated with hydrogen peroxide which showed favourable evidence for the protective capabilities of CYGB. Since this was carried out on HEK293S cells, it would be more beneficial to carry this out on cancer cell lines such as the MCF7 and DU145 which this investigation originally intended as this cell line would be more physiologically relevant.

Additionally, the aim was to carry the experiments out on stable-transfected cell lines to allow longer term experiments on the role of CYGB in cellular proliferation and therapy response to be undertaken, with all cells expressing CYGB, not a population of cells that were successfully transfected. Therefore, generating data within this line would strengthen evidence shown in this study. Another key area which would be interesting to investigate would be wound assays which were intended to be carried out during the course of this investigation. These involve 'scratching' the adherent cell plates, measuring the diameter of the scratch and timing how long it takes for the cells to migrate back. Preliminary work suggested that there was greater cellular adhesion in cells transfected with CYGB and so quantifiable data would be interesting to observe and important to consider when looking at larger scale studies. A previously unpublished study showed that Intercellular Adhesion Molecule 1 (ICAM-1) expression was not affected by Cygb overexpression, however, other cell adhesion molecules have yet to be fully examined.

Additional studies would include investigating the mechanism of therapy resistance, apoptosis studies to see if cell death is indeed increased, additional therapies to see if CYGB promotes therapy resistance e.g. radiotherapy which generates ROS, as this might support the hydrogen peroxide data.

The Cygb protein itself, despite being discovered in 2001, is still much of a functional mystery. Many hypotheses with convincing evidence have highlighted areas in which Cygb appears to be necessary and this study largely supports and extends data which has been previously published. The addition of the data obtained through the creation of mutant protein allows a greater understanding of the areas in which certain functions take place.

Investigations that would be noteworthy to carry out following this work would be a more detailed investigation onto the effect of the mutations on the peroxidatic activity of the protein, including a peroxidase guaiacol assay and an EPR study to examine the formation of radials. This observes the effects of hydrogen peroxide on protein by measuring the guaiacol oxidation which is the rate determining step in peroxidase activity. Given the remarkable effects of hydrogen peroxide on the cells, this data would be of particular interest. The long-term goal is to set up a database of the effects of specific mutations on their *in vitro* properties and what could be the potential effects on physiological and potential functions. These can be compared to the effects of the mutation on protecting cells against oxidative and nitrative stress to determine the molecular mechanism of this protection and hence provide a clearer understanding of the mechanism of cell protection by Cygb and a pathway for development of new therapies.

References:

1. Abal, M., Andreu, J.M. and Barasoain, I. (2003) 'Taxanes: microtubule and centrosome targets, and cell cycle dependent mechanisms of actions', *Current Cancer Drug Targets*, 3(3), pp. 193-203.
2. Ahmad, S.S., Jena, R., Williams, M.V. and Burnet, N.G. (2012) 'Advances in radiotherapy', *British Medical Journal*, 345, pp. 7765.
3. Alayash, A. I., Patel, R. P. and Cashon, R. E. (2001) 'Redox Reactions of Hemoglobin and Myoglobin: Biological and Toxicological Implications', *Antioxidants and Redox Signalling*, 3(2), pp. 313-327.
4. Amin, A.R.M., Karpowicz, P. A., Carey, T.E., Arbiser, J., Nahta, R., Chen, Z.G., Dong, J., Kucuk, O., Khan, G.N., Huang, G.S., Mi, S., Lee, H., Reichrath, J., Honoki, K., Georgakilas, A.G., Amedei, A., Amin, A., Helferich, B., Boosani, C.S., Ciriolo, M.R., Chen, S., Mohammed, S.I., Azmi, A.S., Keith, W.N., Bhakta, D., Halicka, D., Niccolai, E., Fujii, H., Aquilano, K., Ashraf, S.S., Nowsheen, S., Yang, X., Bilsand, A. and Shin, D.M. (2015) 'Evasion of anti-growth signalling: a key step in tumorigenesis and potential target for treatment and prophylaxis by natural compounds, *Seminars in Cancer Biology*, 35(1), pp. 55-77.
5. Anthoni, U., Christophersen, C. and Nielson, P. H. (2007) 'Naturally Occurring Cyclotryptophans and Cyclotryptamines', *Alkaloids: Chemical and Biological Perspectives*, 13(1), pp. 163-236.
6. Aune, D., Chan, D.S.M., Greenwood, D.C., Vieira, A.R., Rosenblatt, D.A.N., Vieira, R. and Norat, T. (2012) 'Dietary fibre and breast cancer risk: a systematic review and meta-analysis of prospective studies', *Annual of Oncology*, 23(6), pp. 1394-1402.
7. Backer, J.D., Razzokov, J., Hammerschmid, D., Mensch, C., Kumar, N., Bogaerts, A. and Dewilde, S. (2018) 'The Role of Cytoglobin In The Plasma-Treatment of Melanoma', *Clinical Plasma Medicine*, 9, pp. 14

8. Bagshaw, C.R. (2013) ‘Stopped-Flow Techniques’, *Encyclopaedia of Biophysics*, Springer, Berlin, Heidelberg. Available at: doi.org/10.1007/978-3-642-16712-6_59 (accessed 11/12/2020).
9. Baskar, R., Lee, K. A., Yeo, R. and Yeoh, K. (2012) ‘Cancer and Radiation Therapy: Current Advances and Future Directions’, *International Journal of Medical Sciences*, 9(3), pp. 193-199.
10. Batzoglou, S., Pachter, L., Mesirov, J.P., Berger, B. and Lander, E.S. (2000) ‘Human and Mouse Gene Structure: Comparative Analysis and Application to Exon Prediction’, *Genome Research*, 10(7), pp. 950-958.
11. Beckerson, P., Reeder, B. J. and Wilson, M. T. (2015) ‘Coupling of disulphide bond and distal histidine dissociation in human ferrous cytoglobin regulates ligand binding’, *FEBS Letters*, 589(4), pp. 507-512.
12. Beckerson, P., Svistunenko, D. and Reeder, B. (2015) ‘Effect of the distal histidine on the peroxidatic activity of monomeric cytoglobin’, *F1000Res*, 5(87), PMID: 26069730.
13. Beckerson, P., Wilson, M.T., Svistunenko, D.A. and Reeder, B.J. (2015) ‘Cytoglobin ligand binding regulated by changing haem-coordination in response to intramolecular disulphide bond formation and lipid interaction’, Biochemical Journal, 465(1), pp. 127-137.
14. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H. and Shindyalov, I.N., Bourne, P.E. (2000) ‘The Protein Data Bank’, *Nucleic Acids Research*, 28, pp. 235-242.
15. Bholah, T.C., Neergheen-Bhujun, V.S., Hodges, N.J., Dyall, S.D. and Bahorun, T. (2015) ‘Cytoglobin as a Biomarker in Cancer: Potential Perspective for Diagnosis and Management’, *Biomedical Research International*, 2015, PMID: 26339645.
16. Birrer, N., Chinchilla, C., Del Carmen, M., Dizon, D.S. (2018) ‘Is Hormone Replacement Therapy Safe in Women With a BRCA Mutation? A Systematic Review of the Contemporary Literature’, *American Journal of Clinical Oncology*, 41(3), pp. 313-315.

17. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A. (2018) 'Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA Cancer Journal Clinic*, 68(6), pp. 394-424.
18. Broad Institute (2020) 'GTExPortal- Tissue and Sample Statistics', *GTExPortal*, available at: https://gtexportal.org/home/tissueSummaryPage (accessed 10/12/2020).
19. Burmester, T. and Hankeln, T. (2009) 'What is the function of neuroglobin', *The Journal of Experimental Biology*, 212(1), pp. 1423-1428.
20. Campbell, K.J. and Tait, S.W.G. (2018) 'Targeting BCL-2 regulated apoptosis in cancer', *Open Biology*, 8(5), 180002.
21. Chabner, B.A. and Roberts, T.G. (2005) 'Chemotherapy and the war on cancer', *Nature*, 5, pp. 65-72.
22. Chen, W.Y., Rosner, B., Hankinson, S.E., Colditz, G.A. and Willett, W.C. (2012) 'Moderate alcohol consumption during adult life, drinking patterns and breast cancer risk', *Journal of American Medical Association*, 306(17), pp. 1884-1890.
23. Chiang, S. P. H., Cabrera, R. M. and Segall, J. E. (2016) 'Tumor cell intravasation', *American Journal of Physiology*, 311(1), pp. 1-14.
24. Chua, P.J., Yip, W.C. and Bay, B.H. (2009) 'Cell cycle arrest induced by hydrogen peroxide is associated with modulation of oxidative stress regulated genes in breast cancer cells', *Experimental Biology and Medicine*, 234(9), pp. 1086-1094.
25. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) 'WebLogo: A sequence logo generator', *Genome Research*, 14, pp. 1188-1190.
26. Cuzick, J., Sestak, I., Baum, M., Buzdar, A., Howell, A., Dowsett, M. and Forbes, J.F. (2010) 'Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial', *Oncology*, 11(12), pp. 1135-1141.
27. Delrue, I., Pan, Q., Baczmanska, A.K., Callens, B.W. and Verdoodt, L.L.M. (2018) 'Determination of the Selection Capacity of Antibiotics for Gene Selection', *Biotechnology Journal*, 13(8), pp. 700-747.

28. Downs-Holmes, C. and Silverman, P. (2011) ‘Breast cancer: Overview & Updates’, *The Nurse Practitioner*, 36(12), pp. 20-26.
29. Dunn, M. W. and Kazer, M. W. (2011) ‘Prostate Cancer Overview’, *Seminars in Oncology Nursing*, 27(4), pp. 241-250.
30. Ellis, A., Risk, J.M., Maruthappu, T. and Kelsell, D.P. (2015) ‘Tylosis with oesophageal cancer: Diagnosis, management and molecular mechanisms’, *Orphanet Journal of Rare Diseases*, 10(126), PMID: 26419362.
31. Ellis, H. and Mahadevan, V. (2013) ‘Anatomy and Physiology of the breast’, *Surgery*, 31(1), pp. 11-14.
32. Emara, M., Turner, A.R. and Allalunis-Turner, J. (2010) ‘Hypoxic regulation of cytoglobin and neuroglobin expression in human normal and tumor tissues’, *Cancer Cell International*, 10(33), PMID: 20828399.
33. Espinosa, E., Zamora, P., Feliu, J. and Baron, M.G. (2003) ‘Classification of anticancer drugs- a new system based on therapeutic targets’, *Cancer Treatment Reviews*, 29(6), pp. 515-523.
34. Estarellas, C., Capece, L., Seira, C., Bidon-Chanal, A., Estrin, D. A. and Luque, F. J. (2016) ‘Structural Plasticity in Globins: Role of Protein Dynamics in Defining Ligand Migration Pathways’, *Advances in Protein Chemistry and Structural Biology*, 105(1), pp. 59-80.
35. Fago, A., Hundahl, C., Malte, H. and Weber, R.E. (2004) ‘Functional Properties of Neuroglobin and Cytoglobin. Insights into the Ancestral Physiological Roles of Globins’, *Life*, 56(11), pp. 689-696.
36. Faguet, G. B. (2014) ‘A brief history of cancer: Age-old milestones underlying our current knowledge database’, *International Journal of Cancer,* 136(9), pp. 2022-2036.
37. Feng, Y., Wu, M., Li, S., He, X., Tang, J., Peng, W., Zeng, B., Deng, C., Ren, G. and Xiang, T. (2018) ‘The epigenetically downregulated factor CYGB suppresses breast cancer through inhibition of glucose metabolism’, *Journal of Experimental and Clinical Cancer Research*, 37, pp. 313.

38. Fordel, E., Thijs, L., Martinet, W., Lenjou, M., Laufs, T., Bockstaele, D.V., Moens, L. and Dewilde, S. (2006) 'Neuroglobin and cytoglobin overexpression protects human SH-SY5Y neuroblastoma cells against oxidative stress-induced cell death', *Neuroscience Letters*, 410(2), pp. 146-151.
39. Gardner, A.M., Cook, M.R. and Gardner, P.R. (2010) 'Nitric-oxide Dioxygenase Function of Human Cytoglobin with Cellular Reductants and in Rat Hepatocytes', *Journal of Biological Chemistry*, 285(31), pp. 23850-23857.
40. George, S.A. (2010) 'Barriers to breast cancer screening: an integrative review', *Health Care for Women International*, 21(1), pp. 53-65.
41. Ghasemi, F., Rostami, S., Nabavinia, M.S. and Meshkat, Z. (2016) 'Developing Michigan Cancer Foundation 7 with Stable Expression of E7 Gene of Human Papillomavirus Type 16', *Iranian Journal of Pathology*, 11(1), pp. 41-46.
42. Giancotti, F.G. (2014) 'Deregulation of cell signalling in cancer', *FEBS Letters*, 588, pp. 2558-2570.
43. Gotoh, O., Sugiyama, Y., Takazawa, Y., Kato, K., Tanaka, N., Omatsu, K., Takeshima, N., Nomura, H., Hasegawa, K., Fujiwara, K., Taki, M., Matsumura, N., Noda, T. and Mori, S. (2019) 'Clinically relevant molecular subtypes and genomic alteration-independent differentiation in gynaecologic carcinosarcoma', *Nature*, 10(4965).
44. Grozdanic, S.D., Ostojic, J., Syed, N.A., Hargrove, M.S., Trent, J.T., Betts, D.M., Kuehn, M.H., Kwon, Y.H., Hardon, R.H. and Sakaguchi, D.S. (2004) 'Neuroglobin and Histoglobin (cytoglobin)- New hexacoordinate globins in the human eye', *Investigative ophthalmology and visual science*, 45(13), pp. 2586.
45. Guex, N., Diemand, A., Peitsch, M.C. and Schwede, T. (2019) 'Swiss-PdbViewer', version 4.1, available at: https://spdbv.vital-it.ch/ (accessed 06/12/2020).
46. Haas, G. and Sakr, W. (1997) 'Epidemiology of prostate cancer', *CA Cancer Journal,* 47(1), pp. 273-287.
47. Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of Cancer: The Next Generation', *Cell*, 144(5), pp. 646-674.

48. Hanahan, D. and Weinberg, R.A. (2000) 'The Hallmarks of Cancer', *Cell*, 100(1), pp. 57-70.
49. Hankeln, T. and Burmester, T. (2008) 'Neuroglobin and Cytoglobin', *The Smallest Biomolecules: Diatomics and their Interactions with Heme Proteins*, edited by ABHIK GHOSH Department of Chemistry, University of Tromso Norway, pages 597-603.
50. Hedayat, K.M. and Lapraz, J-C. (2019) 'Chapter 7- Disorders of the prostate: Lower urinary tract obstruction and prostatitis', *The Theory of Endobiogeny*, 3, pp. 135-164.
51. Hogle, W. (2009) 'Prostate cancer screening, risk, prevention and prognosis', *Pittsburgh, PA: Oncology Nursing Society*, pp. 19-33.
52. Hoogewijs, D., Ebner, B., Germani, F., Hoffmann, F. G., Fabrizius, A., Moens, L., Burmester, T., Dewilde, S., Storz, J. F. Vinogradov, S. N. and Hankeln, T. (2012) 'Androglobin: a chimeric globin in metazoans that is preferentially expressed in Mammalian testes', *Molecular Biology Evolution*, 29(4), pp. 1105-1114.
53. Hu, J., Xu, H., Zhu, W., Wu, F., Wang, J., Ding, Q. and Jiang, H. (2015) 'Neo-adjuvant hormone therapy for non-metastatic prostate cancer: a systematic review and meta-analysis of 5,194 patients', *World Journal of Surgical Oncology*, 13(73), pp. 1297-1302.
54. Huang, C., Ju, D., Chang, C., Reddy, P. M. and Velmurgugan, K. (2017) 'A review on the effects of current chemotherapy drugs and natural agents in treating non-small cell lung cancer', *BioMedicine*, 7(4), pp. 1-12.
55. Human Protein Atlas (2020) 'The Human Protein Atlas', version 20.0, available at: https://www.proteinatlas.org/ (accessed 06/12/2020).
56. Jeziorksa, D.M., Murray, R.J.S., Gobbi, M.D., Gaentzsch, R., Garrick, D., Ayyub, H., Chen, T., Li, E., Telenius, J., Lynch, M., Graham, B., Smith, A.J.H., Lund, J.N., Hughes, J.R., Higgs, D.R. and Tufarelli, C. (2017) 'DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease', *Proceedings of the National Academy of Sciences of the United States of America*, 114(36), pp. 7526-7537.

57. Jin, K., Mao, X.O., Xie, L., Khan, A.A. and Greenberg, D.A. (2008) 'Neuroglobin Protects Against Nitric Oxide Toxicity', *Neuroscience Letters*, 430(2), pp. 135-137.
58. John Hopkins Medicine (2020) 'Breast Cancer & Breast Pathology', *John Hopkins Medicine- Pathology*, accessible: https://pathology.jhu.edu/breast/ (accessed: 22/12/2021).
59. Jordan, V. C. (2003) 'Tamoxifen: a most unlikely pioneering medicine', *Nature Reviews Drug Discovery*, 2(1), pp. 205-213.
60. Kakar, S. (2010) 'Structure and Reactivity of Hexacoordiante Hemoglobins', *Iowa State University- Graduate Theses and Dissertations*, 11235.
61. Kawai, H., Nishino, H., Nishida, Y., Masuda, K. and Saito, S. (1987) 'Localization of myoglobin in human muscle cells by immunoelectron microscopy', *Muscle Nerve*, 10(2), pp. 144-149.
62. Kendrew, J.C., Dickerson, R.E., Strandverg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C. and Shore, V.C. (1960) 'Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution', *Nature*, 185(4711), pp. 422-427.
63. Kierzkowska-Pawlak, H. (2012) 'Determination of Kinetics in Gas-Liquid Reaction Systems. An overview', *Ecological Chemistry and Engineering*, 19, pp. 175-196.
64. Kim, E., Jang, M., Song, M., Kim, D., Kim, Y. and Jang, H. H. (2019) 'Redox-Mediated Mechanisms of Chemoresistance in Cancer Cells', *Antioxidants*, 8(471), pp. 1-18.
65. Kim, W., Lee, S., Seo, D., Kim, D., Kim, K., Kim, E., Kang, J., Seong, K.M., Youn, H. and Youn, B. (2019) 'Cellular Stress Responses in Radiotherapy', *Cells*, 8(9), pp. 1105.
66. Kim-Shapiro, D.B. and Gladwin, M.T. (2014) 'Mechanisms of Nitrite Bioactivation', *Nitric Oxide*, 38, 58-68.
67. King, N.K. and Winfield, M. (1963) 'The mechanism of metmyoglobin oxidation', *The Journal of Biological Chemistry*, 238, pp. 1520-1528.
68. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and

Higgins, D.G. (2007) ‘Clustal W and Clustal X version 2.0’, *Bioinformatics*, 23, pp. 2947-2948.

69. Lepor, H. (2000) ‘Selecting Candidates for Radical Prostatectomy’, *Reviews in Urology*, 2(3), pp. 182-184, 186-189.
70. Lewinska, A., Wnuk, M., Slota, E. and Bartosz, G. (2007) ‘Total anti-oxidant capacity of cell culture media’, *Clinical and Experimental Pharmacology and Physiology*, 34(8).
71. Li, D., Chen, X.Q., Li, W., Yang, Y., Wang, J. and Yu, A.C.H. (2007) ‘Cytoglobin Up-regulated by Hydrogen Peroxide Plays a Protective Role in Oxidative Stress’, *Neurochemical Research*, 32(8), pp. 1375-1380.
72. Li, H., Hemann, C., Abdelghany, T.M., El-Mahdy, M.A. and Zweier, J.L. (2012) ‘Characterisation of the mechanism and magnitude of Cytoglobin-mediated Nitrite Reduction and Nitric Oxide Generation under Anaerobic Conditions’, *Journal of Biological Chemistry*, 287(43), pp. 36623-36633.
73. Li, X., He, S., Ma, B. (2020) ‘Autophagy and autophagy-related proteins in cancer’, *Molecular Cancer*, 19(1), pp. 12.
74. 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., Kawada, N. and Zweier, J.L. (2017) ‘Cytoglobin regulates blood pressure and vascular tone through nitric oxide metabolism in the vascular wall’, *Nature Communications*, 8, pp. 14807.
75. Liu, X., Tong, J., Zweier, J.R., Follmer, D., Hemann, C., Ismail, R. and Zweier, J.L. (2014) ‘Differences In Oxygen-Dependent Nitric Oxide Metabolism By Cytoglobin And Myoglobin Account For Their Differing Functional Roles’, *Federation of European Biochemical Sciences Journal*, 280(5), pp. 3621-3631.
76. Liu, Y. and Crawford, F.W. (2018) ‘Estimating dose-specific cell division and apoptosis rates from chemo-sensitivity experiments’, *Scientific Reports*, 8(2705), pp. 1038-1047.
77. Liu, Y., Zhang, Y., Sun, C. and Gao, P. (2005) ‘A novel approach to estimate *in vitro* antibacterial potency of Chinese medicine using a concentration-killing curve method’, *American Journal of Chinese Medicine*, 33(4), pp. 671-682.

78. Lv, Y., Wang, Q., Diao, Y. and Xu, R. (2008) 'Cytoglobin: A Novel Potential Gene Medicine for Fibrosis and Cancer Therapy', *Current Gene Therapy*, 8, pp. 287-294.
79. Mahaseth, T. and Kuzminov, A. (2017) 'Potentiation of Hydrogen Peroxide Toxicity', *Mutation Research*, 773, pp. 274-281.
80. Malhotra, S., Alsulami, A.F., Heiyun, Y., Ochoa, B.M., Jubb, H., Forbes, S. and Blundell, T.L. (2019) 'Understanding the impacts of missense mutations on structures and functions of human cancer-related genes: A preliminary computational analysis of the COSMIC Cancer Gene Census', *PLoS One*, 14(7), e0219935.
81. Marengo-Rowe, A.J. (2006) 'Structure-function relations of human hemoglobins', *Baylor University Medical Center Proceedings',* 19(3), pp. 239-245.
82. Mathai, C., Jourd'heuil, F.L., Lopez-Soler, R.I. and Jourd'heuil, D. (2020) 'Emerging perspectives on cytoglobin, beyond NO dioxygenase and peroxidase', *Redox Biology*, 32, PMID: 32087552.
83. McRonald, F.E., Risk, J.M. and Hodges, N.J. (2012) 'Protection from Intracellular Oxidative Stress by Cytoglobin in Normal and Cancerous Oesophageal Cells', *PLoS One*, 7(2), pp. 30587-30591.
84. Mimura, I., Nangaku, M., Nishi, H., Inagi, R., Tanaka, T. and Fujita, T. (2010) 'Cytoglobin, a Novel Globin, Plays an Antifibrotic Role in the Kidney', *American Journal of Physiology, Renal Physiology*, 299(5), pp. 1120-1133.
85. Monteiro, N., Martina, A., Reis, R.L. and Neves, N.M. (2014) 'Liposomes in tissue engineering and regenerative medicine', *Journal of the Royal Society Interface*, 11(101), PMID: 25401172.
86. Moroni, F. (1999) 'Tryptophan metabolism and brain function: focus on kynurenine and other indole metabolites', *European Journal of Pharmacology*, 375(3), pp. 87-100.
87. Nagy, A., Lanczky, A., Menyhart, O. and Gyorffy, B. (2018) 'Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets', *Scientific Reports*, 8, pp. 9227.

88. NCBI (2020) ‘CYGB cytoglobin [*Homo sapiens* (humans)]’ available at: https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=114757 (accessed 06/12/2020).
89. NCBI (2020) ‘Open Reading Frame Finder’ available at: https://www.ncbi.nlm.nih.gov/orffinder/ (accessed 08/07/2020).
90. Nishi, H., Inagi, R., Kawada, N., Yoshizato, K., Mimura, I., Fujita, T. and Nangaku, M. (2011) ‘Cytoglobin, a Novel Member of the Globin Family, Protects Kidney Fibroblasts against Oxidative Stress under Ischemic Conditions’, *The American Journal of Pathology*, 178(1), pp. 128-139.
91. Nishida, N., Yano, H., Nishida, T., Kamura, T. and Kojiro, M. (2006) ‘Angiogenesis in Cancer’, *Vascular Health and Risk Management*, 2(3), pp. 213-219.
92. Oleksiewicz, U., Liloglou, T., Tasopoulou, K-M., Daskoulidou, N., Bryan, J., Gosney, J.R., Field, J.K. and Xinarianos, G. (2013) ‘Cytoglobin has bimodal: tumour suppressor and oncogene functions in lung cancer cell lines’, *Human Molecular Genetics*, 22(16), pp. 3207-3217.
93. Ordway, G. A. and Garry, D. J. (2004) ‘Myoglobin: an essential hemoprotein in striated muscle’, *Journal of Experimental Biology*, 207(20), pp. 3441-34.
94. Pandya, S. and Moore, R. G. (2011) ‘Breast Development and Anatomy’, *Clinical Obstetrics and Gynecology*, 54(1), pp. 91-95.
95. Park, I.J., Hwang, J.T., Young, J.H. and Ock, J.P. (2006) ‘Differential modulation of AMPK signalling pathways by low or high levels of exogenous reactive oxygen species in colon cancer cells’, *Annals of the New York Academy*, 1091(1), pp. 102-109.
96. Passarelli, M.N., Newcomb, P.A., Hampton, J.M., Trentham-Dietz, A., Titus, L.J., Egan, K.M., Baron, J.A. and Willett, W.C. (2016) ‘Cigarette Smoking Before and After Breast Cancer Diagnosis: Mortality From Breast Cancer and Smoking-Related Diseases’, *Journal of Clinical Oncology*, 34(12), pp. 1315-1322.
97. Peng, F., Setyawati, M. I., Tee, J. K., Ding, X., Wang, J., Nga, M. E., Ho, H. K. and Leong, D. T. (2019) ‘Nanoparticles promote *in vivo* breast cancer cell intravasation and

extravasation by inducing endothelial leakiness', *Nature Nanotechnology*, 14(1), pp. 279-286.

98. Pienta, K.J. (2016) 'Preclinical mechanisms of action of docetaxel and docetaxel combinations in prostate cancer', *Seminars of Oncology*, 28(4), pp. 3-7.
99. Plotnikov, E.Y., Chupyrkina, A.A., Pevzner, I.B., Isaev, N.K. and Zorov, D.B. (2009) 'Myoglobin causes oxidative stress, increase of NO production and dysfunction of kidney's mitochondria', *Biochimica et Biophysica Acta*, 1792(8), pp. 796-803.
100. Ponka, P. (1999) 'Cell biology of heme', *The American Journal of the Medical Sciences*, 318(4), pp. 241-256.
101. Prasad, S., Gupta, S.C. and Tyagi, A.K. (2017) 'Reactive oxygen species (ROS) and cancer: Role of antioxidative nutraceuticals', *Cancer Letters*, 387, pp. 95-105.
102. Prevarskaya, N., Skryma, R. and Shuba, Y. (2010) 'Ion Channels and the Hallmarks of Cancer', *Trends in Molecular Medicine*, 16(3), pp. 107-121.
103. Rambo, R.P. (2017) 'Considerations for Sample Preparation Using Size-Exclusion Chromatography for Home and Synchrotron Sources', *Advances in Experimental Medicine and Biology*, 1009, pp. 31-45.
104. Raychaudhuri, S., Skommer, J., Henty, K., Birch, N. and Brittain, T. (2010) 'Neuroglobin protects nerve cells from apoptosis by inhibiting the intrinsic pathway of cell death', *Apoptosis*, 15(4), pp. 401-411.
105. RCSB PDB (2006) 'The Protein Data Bank' available at: https://www.rcsb.org/structure/ (accessed 06/12/2020).
106. Reeder, B. and Ukeri, J. (2018) 'Strong modulation of nitrite reductase activity of cytoglobin by disulphide bond oxidation: Implications for nitric oxide homeostasis', *Nitric Oxide: Biology and Chemistry*, 72, pp. 16-23.
107. Reeder, B.J. and Wilson, M.T. (2005) 'Hemoglobin and myoglobin associated oxidative stress: from molecular mechanisms to disease States', *Current Medicinal Chemistry*, 12(23), pp. 2741-2751.

108. 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', *Biochemical Journal*, 434(3), pp. 483-492.

109. Reuss, S., Wystub, S., Disque-Kaiser, U., Hankeln, T. and Burmester, T. (2016) 'Distribution of Cytoglobin in the Mouse Brain', *Frontiers in Neuroanatomy*, 10(47), PMID: 27199679.

110. Sanctis, D., 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.

111. Seyfried, TN and Huysentruyt, LC. (2014) 'On the Origin of Cancer Metastasis', *Critical Review Oncology*, 18(1), pp. 43-73.

112. Shaw, R.J., Omar, M.M., Rokadiya, S., Kogera, F.A., Lowe, D., Hall, G.L., Woolgar, J.A., Homer, J., Liloglou, T., Field, J.K. and Risk, J.M. (2009) 'Cytoglobin is upregulated by tumour hypoxia and silenced by promoter hypermethylation in head and neck cancer', *British Journal of Cancer*, 101, pp. 139-144.

113. Shivapurkar, N., Stastny, V., Okumura, N., Girard, L., Xie, Y., Prinsen, C., Thunnissen, F.B., Wistuba, I.I., Czerniak, B., Frenkel, E., Roth, J.A., Liloglou, T., Xinarianos, G., Field, J.K., Minna, J.D. and Gazdar, A.F. (2008) 'Cytoglobin, the Newest Member of the Globin Family, Functions as a Tumor Suppressor Gene', *Cancer Research*, 68(18), pp. 7448-7456.

114. Song, G., Cheng, L., Chao, Y., Yang, K. and Liu, Z. (2017) 'Emerging Nanotechnology and Advanced Materials for Cancer Radiation Therapy', *Advanced Materials*, 29(32), PMID: 28643452.

115. Springer, B.A., Slinger, S.G., Olsen, J.S. and Phillips, G.N.J. (1994) 'Mechanisms of ligand recognition in myoglobin', *Chemical Reviews,* 94, pp. 699-714.

116. Standard Primer GATC, eurofins (2019), accessed using: https://www.eurofinsgenomics.eu/media/1610449/standard-primer-gatc_eg-logo.pdf (Accessed 29/11/20).

117. Tangar, A. (2018) 'Structure-Function Relationships in Hexacoordinate Heme Proteins: Mechanism of Cytoglobin Interactions with Exogeneous Ligands', *Florida International University*, DOI: 10.25148, pp. 110.

118. Tejero, J., Kapralov, A.A., Baumgartner, M.P., Sparacino-Watkins, C.E., Anthonymutu, T.S., Vlasova, I.I., Camacho, C.J., Gladwin, M.T., Bayir, H. and Kagan, V.E. (2016) 'Peroxidase Activation of Cytoglobin by Anionic phospholipids: Mechanisms and Consequences', *Biochimica et Biophysica Acta*, 1861(5), pp. 391-401.

119. Tejero, J., Sparacino-Watkins, C.E., Ragireddy, V., Frizzell, S. and Gladwin, M.T. (2015) 'Exploring the Mechanisms of the Reductase Activity of Neuroglobin by Site-Directed Mutagenesis of the Heme Distal Pocket', *Biochemistry*, 54(3), pp. 722-733.

120. Thomas, C. and Lumb, A. B. (2012) 'Physiology of haemoglobin', *Continuing Education in Anasthesia Critical Care and Pain*, 12(5), pp. 251-256.

121. 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', *Journal of Biological Chemistry*, 286(20), pp. 18277-18289.

122. Trent, J.R. and Hargrove, M. (2002) 'A ubiquitously expressed human hexacoordinate hemoglobin', *Journal of Biological Chemistry*, 277, pp. 19538-19545.

123. Tse, S.W. (2015) 'Biochemical and Cellular Studies of Vertebrate Globins', *School of Biological Sciences, University of East Anglia*, accessible via: https://core.ac.uk/download/pdf/41994095.pdf (accessed 17/01/2021).

124. U.S. National Library of Medicine (2020) *Histidine.* Available at: https://pubchem.ncbi.nlm.nih.gov/compound/Histidine (Accessed 07 February 2020).

125. U.S. National Library of Medicine (2020) *Leucine.* Available at: https://pubchem.ncbi.nlm.nih.gov/compound/Leucine (Accessed 07 February 2020).

126. U.S. National Library of Medicine (2020) *Tryptophan.* Available at: https://pubchem.ncbi.nlm.nih.gov/compound/Tryptophan (Accessed 07 February 2020).

127. Ukeri, J. (2020) 'Studies on the effect of cysteine and heme pocket mutations of the nitric oxide dioxygenase and nitrite reductase activities of human Cytoglobin and Androglobin heme domain', *University of Essex*, accessible via: https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.814948 (accessed 16/01/2021).

128. Varadhachary, G.R. (2007) 'Carcinoma of Unknown Primary Origin', *Gastrointestinal Cancer Research*, 1(6), pp. 229-235.

129. Vilema-Enriquez, G., Arroyo, A., Grijalva, M., Amador-Zafra, R.I. and Camacho, J. (2016) 'Molecular and Cellular Effects of Hydrogen Peroxide on Human Lung Cancer Cells: *Potential Therapeutic Implications*', *Oxidative Medicine and Cellular Longevity*, Article ID 1908164.

130. Warren, C.F.A., Wong-Brown, M.W., Bowden, N.A. (2019) 'BCL-2 family isoforms in apoptosis and cancer', *Nature*, 10(3), pp. 177.

131. Weaver, B.A. (2014) 'How Taxol/paclitaxel kills cancer cells', *Molecular Biology of the Cell*, 25(18), pp. 2677-2681.

132. Wittenberg, J. B. and Wittenberg, B. A. (2003) 'Myoglobin function reassessed', *The Journal of Experimental Biology*, 206(1), pp. 2011-2020.

133. World Health Organisation (2018) *Cancer*. Available at: https://www.who.int/news-room/fact-sheets/detail/cancer (Accessed 28 October 2019).

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

135. Xin, Y., O'Donnell, A.H., Ge, Y., Chanrion, B., Milekic, M., Rosoklija, G., Stankov, A., Arango, V., Dwork, A.J., Gingrich, J.A. and Haghighi, F.G. (2011) 'Role of CpG context and content in evolutionary signatures of brain DNA methylation', *Epigenetics*, 6(11), pp. 1308-1318.

136. Xu, H-W., Huang, Y-J., Xie, Z-Y., Lin, L., Guo, Y-C., Zhuang, Z-R., Lin, X-P., Zhou, W., Li, M., Huang, H-H., Wei, X-L., Man, K. and Zhang, G-J. (2013) 'The expression of cytoglobin as a prognostic factor in gliomas: a retrospective analysis of 88 patients', *BMC Cancer*, 247, PMID: 23688241.

137. Yang, J., Kloek, A.P., Goldberg, D.E. and Mathews, 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, pp. 4224-4228.

138. Yi, J., Heinecke, J., Tan, H., Ford, P.C. and Richter-Addo, G.B. (2009) 'The distal pocket histidine residue in horse heart myoglobin directs the O-binding of nitrite to the heme iron', *Journal of American Chemical Society*, 131, pp. 18119-18128.

139. Yoshizato, K., Thuy Le, T. T., Shiota, G and Kawada, N. (2016) 'Discovery of cytoglobin and its roles in physiology and pathology of hepatic stellate cells', *Proceedings of the Japan Academy, Series B, Physical and Biological Sciences*, 92(3), pp. 77-97.

140. Zang, L., Anderson, E.M.E, Khajo, A., Magliozzo, R.S. and Koder, R.L. (2014) 'Dynamic Factors Affecting Gaseous Ligand Binding in an Artificial Oxygen Transport Protein', *Biochemistry*, 52(3), pp. 447-455.

141. Zhang, S., Li, X., Jourd'heuil, F.L., Qu, S., Devejian, N., Bennett, E., Jourd-heuil, D. and Cai, C. (2017) 'Cytoglobin Promotes Cardiac Progenitor Cell Survival Against Oxidative Stress via the Upregulation of the NFKB/iNOS Signal Pathway and Nitric Oxide Production', *Scientific Reports*, 7(1), pp. 1034-1037.

142. Zhao, C. and Du, W. (2016) 'Dynamic features of carboxy cytoglobin distal mutants investigated by molecular dynamics simulations', *Journal of Biological Inorganic Chemistry*, 21, pp. 251-261.

Appendix:

WT

MEKVPGEMEIERRERSEELSEAERKAVQAMWARLYASCEDVGVAILVRFFVNFPSAKQYFSQFKHMEDPL
EMERSPQLRKHACRVMGALNTVVENLHDPDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVVAEEFASD
FPPETQRAWAKLRGLIYSHVTAAYKEVGWVQQVPNATTPPATLPSSGPTRTRPLEMESDESGLPAMEIECR
ITGTLNGVEFELVGGGEGTPEQGRMTNKMKSTKGALTFSPYLLSHVMGYGFYHFGTYPSGYENPFLHAINN
GGYTNTRIEKYEDGGVLHVSFSYRYEAGRVIGDFKVMG

H81A pCMV

MEKVPGEMEIERRERSEELSEAERKAVQAMWARLYASCEDVGVAILVRFFVNFPSAKQYFSQFKHMEDPL
EMERSPQLRKAACRVMGALNTVVENLHDPDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVVAEEFASD
FPPETQRAWAKLRGLIYSHVTAAYKEVGWVQQVPNATTPPATLPSSGPTRTRPLEMESDESGLPAMEIECR
ITGTLNGVEFELVGGGEGTPEQGRMTNKMKSTKGALTFSPYLLSHVMGYGFYHFGTYPSGYENPFLHAINN
GGYTNTRIEKYEDGGVLHVSFSYRYEAGRVIGDFKVMG

H81A pET

MEKVPGEMEIERRERSEELSEAERKAVQAMWARLYANCEDVGVAILVRFFVNFPSAKQYFSQFKHMEDPL
EMERSPQLRKAACRVMGALNTVVENLHDPDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVVAEEFASD
FPPETQRAWAKLRGLIYSHVTAAYKEVGWVQQVPNATTPPATLPSSGP

L46W pCMV

MEKVPGEMEIERRERSEELSEAERKAVQAMWARLYASCEDVGVAIWVRFFVNFPSAKQYFSQFKHMEDPL
EMERSPQLRKHACRVMGALNTVVENLHDPDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVVAEEFASD
FPPETQRAWAKLRGLIYSHVTAAYKEVGWVQQVPNATTPPATLPSSGPTRTRPLEMESDESGLPAMEIECR
ITGTLNGVEFELVGGGEGTPEQGRMTNKMKSTKGALTFSPYLLSHVMGYGFYHFGTYPSGYENPFLHAINN
GGYTNTRIEKYEDGGVLHVSFSYRYEAGRVIGDFKVMG

L46W pET

MEKVPGEMEIERRERSEELSEAERKAVQAMWARLYANCEDVGVAIWVRFFVNFPSAKQYFSQFKHMEDPL
EMERSPQLRKHACRVMGALNTVVENLHDPDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVVAEEFASD
FPPETQRAWAKLRGLIYSHVTAAYKEVGWVQQVPNATTPPATLPSSGP