The role of the domain linker in the stability of Glutaredoxin-2

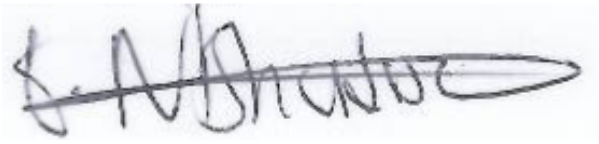
OBAKENG MOLEBOGENG NTSHUDISANE

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science.

Johannesburg, 2012.

DECLARATION

I declare that this dissertation is my own, unaided work. It is being submitted for the degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



Obakeng Molebogeng Ntshudisane

6th September 2012

DEDICATION

I dedicate this work to my mother (Mary Mastiene Ntshudisane) whom despite all the contrary evidence that I have provided,

Thinks I am an angel.

"No great discovery was ever made without a bold guess."

By Sir Isaac Newton (1642-1727)

ABSTRACT

The three dimensional native structure of multi domain proteins is only achieved when the adjacent domains recognise each other through the domain-domain interface. The domain-domain interface of the Glutathione *S*-transferase (GST) family has been studied extensively; however, no studies have been conducted on the role of the linker regions in the domain-domain interactions. Glutaredoxin 2 (Grx2) protein, from the GST family was chosen as model to investigate the possible role of linkers in protein stability by mutational analysis. Bioinformatics data revealed a conserved residue within the linker region (Leu78 in Grx2). A Grx2 mutant was created by replacing the conserved residue (Leu78) within the linker region with an alanine. This mutation (Leu to Ala) was performed in order to assess the role of the conserved residue leucine; whilst maintaining Grx2 function. A previous Grx2 mutant (Grx2 Y58W) was utilised because it incorporates tryptophan into domain 1; therefore it was possible to follow tertiary structural changes in this domain. Grx2 Y58W was compared against the mutant created within the linker Grx2 Y58W/L78A. Far-UV CD spectrum indicated that there was an increase of (~30 %) in ellipticity of Grx2 Y58W/L78A protein whereas; tryptophan fluorescence probes indicated no change in tertiary structure. Conformational stability studies showed a decrease of $\Delta\Delta G\ (H_2O)$ = 3.8 kcal.mol^{-1} due to the impact of the Y58W/L78A mutation. The *m*-value which is indicative of the co-operativity between the two domains has decreased slightly by ~0.4 kcal.mol^{-1} M^{-1}. This reduction in the *m*-value suggested the formation of intermediate however; it was not evident when using ANS as a probe. This study indicates that replacing a leucine with an alanine in the linker region causes a reduction in domain co-operativity. Therefore, the linker region in addition to separating the two domains plays a role in interdomain co-operativity.

ACKNOWLEDGEMENTS

I would like to thank the following people:

Prof Heini Dirr for his enduring patience, constructive criticism, guidance and motivation. Thank you for allowing me to be part of your lab. Your passion for science is an inspiration. I hope some of it has rubbed off to me.

My co-supervisor Dr. Samantha Gildenhuys, words are simply not enough. Thank you, thank you and thank you.

A big THANKS to Stoyan Stoychev and Sindisiwe Buthelezi for all the mass spectroscopy work analysed at CSIR. My project would not be complete without you guys.

All the current members of Protein-Structure-Function Research Unit, University of the Witwatersrand for their encouragement and assistance.

The University of the Witwatersrand for the opportunity and resources to conduct my study and the National Research Foundation for financial support.

My friends, family and “Las Vegas” crew you guys rock!!!!!

Above all, my creator, my God who taught me that nothing is impossible through HIM.

TABLE OF CONTENTS

LIST OF ABBREAVIATIONS

ΔASA	change in (solvent) accessible surface area
Δ*S*	the change in entropy
Å	Ångström
A_{280}	absorbance at 280 nm
ACN	acetonitrile
ANS	8-anilino-1-naphthalene sulphonate
ArsA	Arsernate A
˚C	degrees celsius
AUFs	absorbance units full scale
CD	circular dichroism
Clic	chloride intracellular channel
*C*m	the denaturant concentration at the midpoint of the unfolding curve
D	dielectric constant
Da	dalton
ds	double-stranded
ExPASy	Expert Protein Analysis System
FA	fomic acid
Far-UV CD	far-ultraviolet circular dichroism
ΔΔG(H2O)	change in Gibbs free energy of unfolding in the absence of denaturant
Grx2	glutaredoxin 2
GSH	reduced glutathione
G-site	glutathione-binding site
GSTs	glutathione *S* transferase
Δ*H*	the change in enthalpy
hGST A1-1	human Alpha class glutathione *S*-transferase
hGST M2-2	human Mu class glutathione *S*-transferase
H-site	hydrophobic electrophile binding site
IPTG	isopropyl-β-D-thiogalactopyranoside
K_{eq}	equilibrium constant
LB	luria-Bertani
M	molar
M17A	met17 replaced with Alanine
mdeg	millidegrees
m-value	the dependence of the free energy of unfolding on denaturant concentration
N	native protein
OD_{600}	optical density at 600 nm
PDB	Protein Data Bank
pI	isoelectric point
POU	Its derived from three transcription factors

	Pit-1, Oct 1 and neural Unc-86
q	electrical charge
R	universal gas constant = 8.3145 J/mol.K
SDS - PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
*Sj*GST	*schistosoma japonicum* glutathione *S*-transferase
SspA	stringent starvation protein A
TEMED	*N*, *N*, *N*', *N*'-tetramethylethylenediamine
U	Unfolded protein
Ure2p	yeast prion protein
UV	ultraviolet
Y58W	tyr58 replaced with Tryptophan
Y58W/L78A	tyr58 replaced with Tryptophan and Leu78 replaced with Alanine
YT	yeast tryptone
ε	molar extinction coefficient

The IUPAC-IUBMB three and one letter codes for the amino acids were used.

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In all aspects of life there is a hierarchy level that must be followed and proteins are not above this way of life. There are four levels of protein organisation; primary, secondary, tertiary and quaternary structure. Primary structure refers to the linear sequence of covalently linked amino acids in a polypeptide chain. Anfinsen showed that all the information needed for proteins to fold into a three dimensional structure and perform their function is contained within this linear sequence (Anfinsen, 1973). Secondary structure refers to the spatial arrangement of the polypeptide backbone into local structured regions such as α-helices, β-strands, β-turns. Tertiary structure which is the native protein is the third level, refers to the folded three dimensional state. Quaternary structure is the last level which refers only to multi-subunit proteins and the interactions within. The fundamental unit of a protein structure is the domain (Doolittle, 1995). Recent studies focus on multi-domain proteins probably because two thirds of proteins, in prokaryotes and four fifths in eukaryotes are multi-domain proteins (Teichmann *et al*., 1998; Apic *et al*., 2001; Chothia *et al*., 1975).

1.1. Protein domains

Over the past five decades there have been several definitions of domains based on two concepts; they can either be defined according to structure or evolution. A protein domain can be defined as polypeptide chain that forms compact globular units which are loosely connected and fold autonomously (Wetlaufer, 1973). Having said that it should however, be stated that not all protein domains are autonomous folding units. Several experiments have shown that individual domains in multi-domain proteins fold independently when the interactions between the domains are weak (Han *et al*., 2007). Although domains are defined as 'independent folding units', there are instances where there is lot of 'communication' between domains and a high degree of co-operativity for unfolding of proteins (Batey *et al*., 2005). The individual domains of chicken brain α-spectrin (R16 and R17) fold cooperatively and reversibly to a stable native structure (reviewed by Batey *et al*., 2005). As the R16 domain unfolds, the R17 domain losses its stability and also unfolds. Hence, they have also been defined in terms of folding as cooperative thermodynamic units, which are detected as distinct folding/unfolding transitions (Privalov, 1979). Another definition for domains is

'topological entities' which have more pronounced interactions within the structural unit than with any other parts of the polypeptide chain (Janin and Wodak, 1983). Some of the other definitions for domains are based on the interactions within the domains and the polypeptide chain (Jaenicke, 1999). From an evolutionary concept, domains are defined as "independently evolutionary units" that can form either a single domain or be part of a multi-domain protein (Englander, 2007, Vogel *et al*., 2004). This is achieved through domain swapping, adding and 'stealing'.

Evolution has allowed domain folding primarily because it enhances the rate of protein folding. How? By synchronous folding at multiple sites along the nascent polypeptide chain and this mechanism is known as "folding-by-parts" (Jaenicke, 1999). This mechanism has been conserved probably because of three reasons (Jaenicke, 1999); (i) it is an efficient way of excluding wrong intra-molecular interactions especially in large proteins. (ii) It protects the nascent polypeptide from proteolysis and (iii) it is a simple mechanism to proceed from monomeric to multimeric protein (Bennet *et al*., 1995). Domains represent the folded modular topology of folded protein and they can have independent function or contribute to the function of a multi-domain protein, with the help of other domains. Multi-domain proteins are speculated to have risen from duplication and combination of domains (Apic *et al*., 2003). The function of multi-domain proteins is dependent on the interactions of domains.

1.2 Domain-domain association and interactions (interface) in multi-domain proteins

Before domains in multi-domain proteins can interact, they first have to come together and pair. This is vital since most of the multi-domain proteins have a shared active site between domains. Domains are also connected sequentially by the linker region, which co-ordinates the coupling of the active site. Domain association is the last step in protein docking and is considered the rate determining step (Garel *et al*., 1976). This is probably because domains have to pair in the correct complementary interface which is exceedingly time consuming (Jaenicke, 1999). According to Chothia and Janin (1975) the basis for domain association is that the interface must be highly complementary. Geometric complementarity is also essential at the domain interfaces. How are domains able to recognise each other? Specific complementarity is required for accurate molecular recognition (Chothia and Janin, 1975; Duquerroy *et al*., 1991). Shape complementarity plays a vital role in molecular recognition

because the two domains have to physically fit between two surfaces (Argos, 1988). The recognition of domains is assisted by some physical and chemical characters of the interface residues (Argos, 1988).They are also able to achieve recognition through specific interactions found within the domain interface(s). These interactions are absent in single domain proteins and are vital for protein stability and folding. Domain interfaces are described as the surface area buried upon domain association as well as the contacts formed between an interacting pair of domains (Stoychev, 2008). Most domain interfaces are preformed and rigid; hence they are able to come together as a lock-in-key. However, there are exceptions to this rule, as in the case of induced fit (Wilson and Standfield, 1993). Although domain interface are rigid, there is still some flexibility in the following; surface side chain, surface loops and hinges. Loops are less constrained than the body of globular proteins and are able to move and change shape to fit a binding interface (Wilson and Stanfield, 1993). Domains can also move at hinges. Even though the domains are still rigid, they are connected by flexible hinges (linkers) (George and Heringa, 2002). To understands the role of domain interfaces in stability one needs to understand fully the interactions that occurs there.

1.3 Forces stabilising and destabilising domain-domain interactions

Forces that stabilise the domain-domain interactions are as a resultant of a balance between interactions that favour domain association and interactions that favour domain dissociation. The difference is represented by the change in Gibbs free energy ($\Delta G°$) upon folding. Under constant pressure, ($\Delta G°$) is made up of two contributions:

$$\Delta G° = \Delta H° - T\Delta S° \quad (1)$$

where $\Delta H°$ is the enthalpic (bond formation) and $\Delta S°$ is the entropic (freedom of a system to explore conformational space) contribution to $\Delta G°$. The overall balance between stabilizing and destabilizing forces involved in protein domains is estimated in the range of 5 - 20 kcal/mol (Pace *et al*., 1996). These amounts of energy are relatively low and it is assumed that there is an evolutionary advantage to that. According to Becktel and Schellman, the energy must be low enough for a protein domain to maintain its native conformation but also not so low that it cannot be flexible and perform its function (Becktel and Schellman, 1987). The most dominant forces that contribute towards $\Delta H°$ are electrostatic interactions (which includes ionic and van der Waals interactions), hydrophobic effects and hydrogen bonds. The

main destabilising force that contributes towards domain dissociation is the loss of conformational entropy.

1.3.1 van der Waals interactions and hydrogen bonding in proteins

van der Waals interactions, which are also known as the London dispersion forces, are as a result of induced dipole-dipole interactions between non bonded atoms. Relative to other forces involved in domain association, van der Waals forces are weak. Hence they only function at short range distances and their association energy is proportional to their r^{-6} (Pace, 2001). However, because they occur in large numbers they contribute significantly towards domain-domain association. The tight packing of the hydrophobic groups and the steric complementarity within the proteins interior ensures inter-atomic contacts, which make van der Waals forces vital for maintaining protein stability (Liang and Dill, 2001).

Hydrogen bonding as the name suggests involves hydrogen atom shared between two electronegative atoms such as oxygen, nitrogen or sulphur. The strength of a hydrogen bond is highly dependent on the electronegativity and orientation (linear or directional) of the bonding atoms and varies from 2 - 14 kcal/mol (Hagler *et al*., 1979; Dauber and Hagler, 1980; Dill, 1990; Privalov and Makhatadze, 1993). Since proteins are made up of α-helices and β-sheets, hydrogen bonding has a major influence on protein structure. In globular proteins 68 % of hydrogen bonds occur between the carbonyl (C═O) and amide (NH) groups of the peptide backbone and the polar amino acid side chains (Stickle *et al*., 1992). The polar amino acid side chains are able to hydrogen bond due to the presence of amine, carbonyl, thiol and hydroxyl groups. Hydrogen bonding is able to satisfy all the polypeptide backbone requirements for the formation of α-helices and β-sheet. But how do proteins fold in such a way that most of its hydrogen bonding potential is maximised? It was illustrated by Dill and George through the use of protein X-ray structure that most hydrogen bonds involved in proteins are local (Dill and George, 1999). Local, meaning that they only involve donors and acceptors that are close together in sequence and hence they can readily find each other. Their individual contribution to protein stability is relatively low; but because they are able to form a co-operative network where a donor or acceptor participates in a number of hydrogen bonds their contribution is enhanced (Stickle *et al*., 1992). Although most hydrogen bond contacts are local, it is possible to have non-local contacts as in the case of domain

interface(s). These contacts are possible due to the polar nature of the amino acid residues. Local and non-local contacts emphasise the importance of correct packing in the interior and the domain interface(s). In the 1930's the school of thought was that hydrogen bond was the most important factor in protein stability (Mirsky and Pauling, 1936). However, it was later proven that the hydrophobic effect is the most dominant force in protein folding (Kauzmann, 1959). Recent studies have also suggested that the van der Waals attraction resulting from the tight packing of the hydrophobic groups in the protein interior makes their stability contribution comparable to that of the classic hydrophobic effect (Chen and Stites, 2001).

1.3.2 Hydrophobic effect in proteins

The contribution of hydrophobic interaction in protein folding was first pointed out by Kauzmann (1959). It is believed that hydrophobic effect is the largest force behind protein folding and stability (Rose *et al*., 1985; Dill, 1990). Let us first understand how this effect comes about; this is due to the influence that causes non-polar substance to minimise their contact with water because they are hydrophobic (Dill, 1990). This effect is observed in proteins whereby 81 % of non-polar amino acid residues are buried within the core (Lesser and Rose, 1990). The burial of hydrophobic side chains in the folding reaction is energetically favourable (Dill, 1990) and the process (at room temperature) is entropically driven since the addition of non-polar molecules to water disrupts the hydrogen bonded structure of water (Dill, 1990). This results in water molecules ordering themselves around non-polar molecules to maximise contacts with each other because like attracts like and minimise contact with non-polar molecules (Geiger *et al*., 1979; Stillinger, 1980). The contribution of the hydrophobic effect towards protein stability is estimated to be 60 kcal.mol^{-1} (Dill *et al*., 1989).

1.3.3 Electrostatic interactions in proteins

Electrostatic interactions within proteins are governed by Coulombs' law:

$$F = (k \times q_1 \times q_2) / D \times r^2 \qquad (2)$$

where F is the force between the two electrical charges, q_1 and q_2 separated by distance r. k is the proportionality constant (k = 2.14x10^9 cal.m.C^{-2}) and D is the dielectric constant of the medium. From the equation (**2**) one can see that the dielectric constant (D) is inversely proportional to the strength of the electrostatic interactions. Therefore, a non-polar environment; such as the protein interior, will strengthen the electrostatic interactions. Electrostatic interactions have been classified in two ways; classical effects and specific charge

interactions (Dill, 1990). Classical effects refer to nonspecific repulsion that arises between surface residues with the same charge that destabilises proteins (Dill, 1990). Classical interactions are affected by the pH and the ionic strength. Since the electrostatic free energy is depended on the square of the net charge (see equation **2**), no electrostatic contribution is expected near the isoelectric point (net charge = zero). Therefore, by increasing ionic strength (by adding salts) in protein's environment results in a better shielding of opposite charges, hence a decrease in repulsion which leads to a stabilised protein. Specific charge interactions also known as ion pair or salt bridges arise from oppositely charged residues in close spatial proximity (Dill, 1990). Salt bridges are formed between acidic/negatively charged and basic/positively charged residues. The energy contributions of salt bridges varies between 5 - 15 kcal/mol/ion pair according to their geometry, location in the protein, whether they are isolated or networked, if they are hydrogen bonded or not (Kumar and Nussinov, 1999).

1.3.4. Entropic effect in proteins

The main force opposing domain association is the entropic effect. This effect is as a result of different components such as translational, rotational, vibrational and conformational entropies (Dill, 1990). As proteins unfold the degrees of freedom available to the unfolding molecule increases (side chains and peptide groups become exposed to the solvent). This is because less constraint (rotation around the peptide bond becomes freer) is placed on the unfolding molecule which provides a strong entropic force for protein unfolding.

1.4 Connectivity of domains

Domains are connected sequentially by short stretches of polypeptide residues. For years, domain linkers have been thought of as simple connectors and/or separators of domains. But recently it has been discovered that they might play more than just a physical role. Linkers are also known as hinge regions, and these regions 'experience' localised change in the torsion angles which allows domains to pivot. The main feature about hinge regions is that they have few packing constraints on their main chain atoms (Wrigglers *et al.*, 2005). Linkers are able to move independently of one another while maintaining the three dimensional shape of the domains. The geometry and the flexibility of the linkers enable them to move linked domains to and from close spatial proximity (Wrigglers *et al.*, 2005). Linkers must be able to connect to domains in the native conformation without causing too much strain. There are also different types of domain linkers.

1.4.1 Types of linkers

Helical and non-helical linkers are the two types of domain linkers that have been classified to date (George and Heringa, 2002). A linker is classified as helical if over 33 % of the linker residues are in helical structures; otherwise they are classified as non-helical. Non-helical linkers are rich in prolines, which results in structural rigidity to the domains whereas helical linkers act as spacers or modular rulers (George and Heringa, 2002). Glycine rich linkers confer flexibility to the domains (Briggs *et al*., 1999). Several factors pertaining to linkers have been shown to be vital in the stability and function of domains, two of which are flexibility and hydrophylicity (Wrigglers *et al*., 2005).

1.4.2 Sequence-depended linker properties

The average length of a domain linker is 5.15 residues and it normally varies between 2 to 18 amino acids (Wrigglers *et al*., 2005). The length and the composition of the linker also affect the stability and folding of multi-domain proteins. In the work carried out by van Leeuwen *et al*. (1997), the binding specificity and affinity of the DNA-binding domains were affected when the length of the linker region (23 amino acids) was altered to have between 2 and 37 amino acids in the Pit-1, Oct 1 and neural Unc-86 (POU) domain and the conserved residue (a glutamate) within the linker was mutated. Many studies have shown that linkers lack regular secondary structure and are rich in Ala, Pro and charged residues (Dieckmann *et al*., 1999; Packman and Perham, 1987; Radford *et al*., 1989). Gly rich linkers are generally more flexible because the absence of a β-carbon permits the polypeptide backbone to access dihedral angles that energetically forbidden for other amino acids (Ramachandran and Sasisekharan, 1968). Pro is the most preferred terminal linker residue, followed by Arg, Phe and Thr (Wrigglers *et al*., 2005). Proline is favored over other amino acids in linking domains probably because of its inability to donate hydrogen bonds which ensures rigid separation of the domains (Wrigglers *et al*., 2005). Even though it is the most preferred residue in the linker, prolines can form tight turns which affects domain independence (Wrigglers *et al.*, 2005).

1.4.3 Role of the domain linker in multi domain proteins

Domain linkers have been shown to play a functional role in the structure and activity of proteins. The linker plays an important role in domain co-operativity (Macdonald *et al*., 2004). This property was demonstrated when a seven Ala insert was introduced into the linker region in spectrin (between the R17 and R16 domain) which resulted in the loss of

cooperativity between the two domains (Batey *et al*., 2005). The deletions of four amino acid residues from the linker region connecting the sub-domains of phosphorylated smooth-muscle myosin lead to the termination of the translocating activity (Ikebe *et al*., 1998). Other studies have also shown that the linker region is required for the catalytic function and the conformation of the nucleotide binding domains (Jia and Kaur, 2001). This was demonstrated when the entire linker region in ArsA was deleted, resulting in the deletion mutant not being able to confer resistance towards arsenate proving that linker region is required for the catalytic domains (Jia and Kaur, 2001). Folding and stability in multi-domain proteins is impacted to some degree by the length of the linker region and the nature of the domain interface (van Leeuwen *et al*., 1997). Linkers have also been implicated in maintaining the orientation of two domains necessary for the normal function of proteins (Wong *et al*., 1986). The role of the linker region in most proteins is still being questioned, as to whether it is just a flexible hinge that brings/holds the domains together and if that is true then surely the length of the linker region would be more vital than the nature of the residues themselves. The linker lengths have been shown to affect protein stability (Robinson and Sauer, 1998). These effects have been attributed to changes in the folding and unfolding rates. Apparently if the linker region is shortened, the folding rate increases because the denatured domains are constrained to a smaller conformational regions thus, requiring less random sampling (Robinson and Sauer, 1998). Also if the linker becomes too short then it will not be able to connect to the domain, thus, leading to failure of a protein to fold correctly because of steric hindrance.

1.5. Classification of GSTs

The glutathione *S*-transferases (GSTs) (EC 2.5.1.18) is a group of multi domain proteins that are involved in the detoxification of xenobiotics. They achieve this by conjugating the tripeptide glutathione (GSH) to electrophillic compounds. By this act, they render toxins more water soluble; hence, they can be easily exported out of the cell. GSTs are categorised into three main subfamilies; soluble or cytosolic GSTs (cGSTs) (otherwise known as the canonical GST), microsomal GSTs and the plasmid-encoded bacterial fosfomycin-resistance GST (Frova, 2006). The focus of this study will be on the canonical GSTs. The cGSTs have been identified in plants (Frova, 2005; Soranzo *et al*, 2004), bacteria (Vuilleumier and Pagni, 2002), insects (Ranson *et al*., 1998) and mammals (Hayes *et al*., 2005; Frova, 2006). The seven cGSTs

identified in mammalian are Alpha (A), Mu (M), Omega (O), Pi (P), Sigma (S), Theta (T) and Zeta (Z). The non-mammalian are Beta (B), Delta (D), Phi and Tau (U). There are eleven independent gene classes (Armstrong, 1997) and newer members of the GSTs have also been discovered such as the chloride intracellular channel (Clic) (Harrop *et al*., 2001), glutaredoxin 2 (Grx2) (Xia *et al*., 2001) and Ure2p (Bousset *et al*., 2001). Yeast elongation factor 1-gamma is another family of proteins that has joined the GSTs family (Jepperson *et al*., 2002). Most of the GSTs are active as dimers of either identical homodimers or heterodimers of subunits 23-30 KDa in size and an average length of 200-250 amino acid (Sheehan *et al*., 2001; Frova *et al*., 2006). Each subunit has two distinct domains (Figure 1), separated by a short linker and has an active site which is independent of the other subunit (Mannervik and Daneilson, 1988). Dimerisation is only achieved by specific residues which are located between domain 1 of subunit 1 and domain 2 of subunit 2. The active site has two functional regions; the hydrophilic G-site for binding its natural substrate glutathione and adjacent to it the hydrophobic H-site (Figure 1) for binding electrophillic substrates (Dirr *et al*., 1994). Recognition of the substrate (glutathione) is dependent on the interactions of conserved residues of both domains. Geometric complementarily and orientation is vital at the dimer interface. Although most GSTs are dimers there are some active monomers such as Grx2 and Clic enzymes but they lack the GSH-dependent conjugating activities with standard substrates (Hayes *et al.*, 2005). The sequence homology across the different classes of GSTs is quite low (< 25%) but despite that, structures show that they all share the same GST fold (Sheehan *et al.*, 2001; Hayes *et al*., 2005).

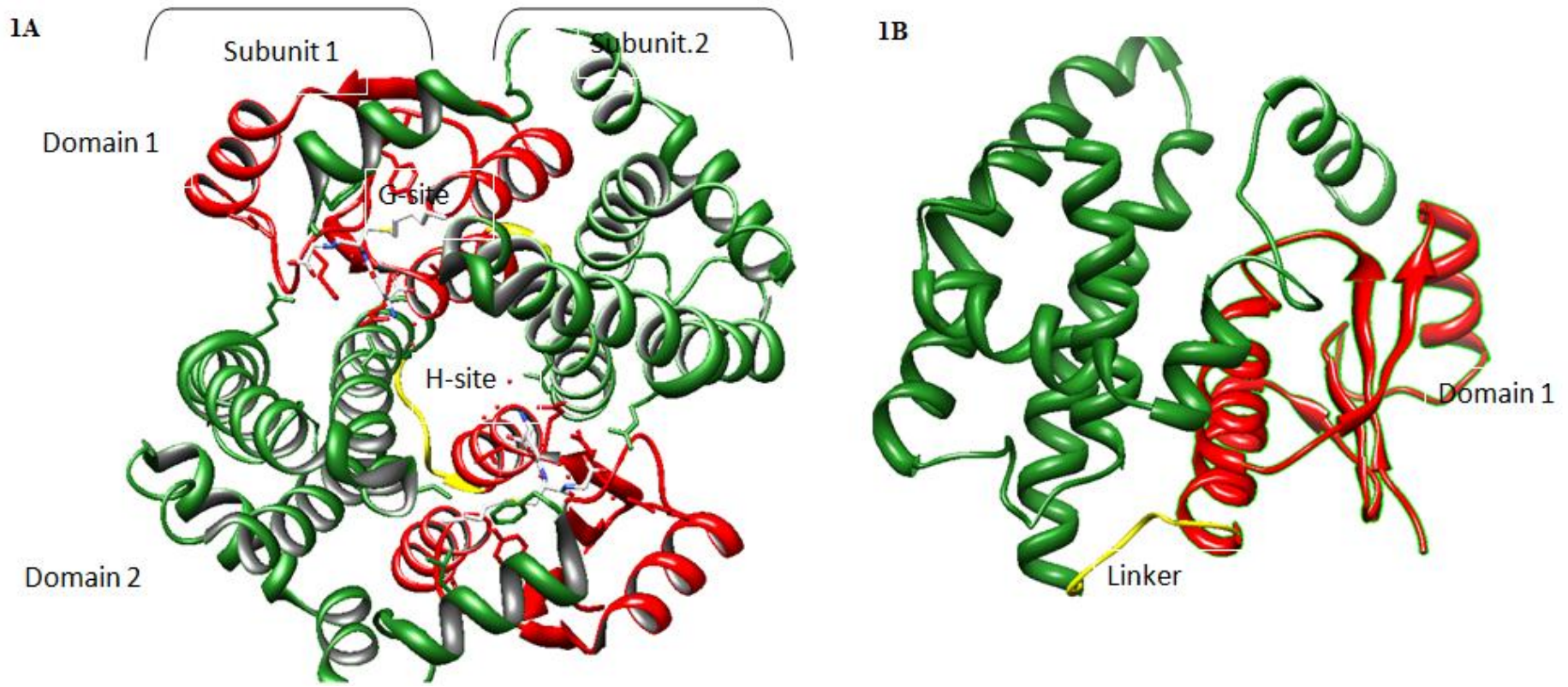


Figure 1. Ribbon representation of the GST in dimeric and monomeric form. A) dimer and b) subunit. The conserved residues involved in the dimer interface are shown as sticks. The two domains are represented by the different colours, with the N-terminal domain shown in red and C-terminal domain in green. The linker regions that connect each domain are shown in yellow. The G-site is occupied by *S*-hexyl glutathione while the H-site has the hydrophobic benzyl ring. This figure was generated from PDF 1K3L (Le Trong *et al*., 2002) using CHIMERA (Pettersen *et al*., 2004).

1.5.1 Glutathione transferase fold

Glutathione *S*-transferase (GST) fold is characterised into a large family of monomeric and dimeric proteins (http://scop.mrc-lmb.cam.ac.uk/scop-1.69/data/scop.b.d.gf.b.html, Murzin *et al.*, 1995). The topology of GST fold consists of a thioredoxin/glutaredoxins fold in the N terminal domain (domain 1) and a larger all-α-helical C terminal domain (domain 2) as illustrated by the structure of hGSTA1-1 (Figure 1). The thioredoxin/glutaredoxins fold is made up of four mixed β strands and three α helices (βαβαββα) (Sheehan *et al.*, 2001). This thioredoxin/glutaredoxin fold is conserved throughout the thioredoxin super-family which includes enzymes such as DsbA, NrdH redoxin, protein disulfide isomerase, chaperons, GSTs, and glutathione peroxidase (http://scop.mrc-lmb.cam.ac.uk/scop-1.69/data/scop.b.d.gf.b.html, Murzin *et al.*, 1995). The active site is positioned along the interfaces between the two domains. Interactions that occur at the interfaces assist significantly towards stabilising the tertiary structures of individual subunits (Dirr, 2001). The C terminal domain (domain 2) has a varying number of helices across the different classes of GSTs and it is speculated that the difference accounts for the different substrate specificity (Wilce and Parker, 1994). It had been proposed that the GST-fold played some undefined role in cell entry (Morris *et al.*, 2009). GST P1-1 is the most prevalent enzyme expressed in high levels in cancer cells (Tew, W., 2007; Tew, W., 1994). It is also implicated in the development of anticancer drug resistance and in tumorigenesis. Recent studies have revealed that the C terminal domain in hGST M2-2 specifically the hydrophobic helix-6 (α6) is responsible for cell translocation (Morris *et al.*, 2010). This property is also shared by the pore forming proteins such as the CLIC family. The preservation of the GST fold highlights the need of conservation of structure within proteins. The C-terminal domain of GST P1-1 has also been implicated.

1.5.2 Unfolding and stability of GSTs

Equilibrium unfolding studies for hGSTA1-1 displayed two-transition state between the native dimer and the unfolded monomers and a three state kinetic pathway for the unfolding of this protein was proposed (Wallace *et al.*, 1998). Wallace and colleagues also investigated the role of domain packing on the stability and function of hGSTA1- 1 (Wallace *et al.*, 2000). This was performed by mutating the conserved residue Trp20 located at the domain interface. This residue (Trp20) protrudes from the N-terminal domain into the

hydrophobic pocket of C- terminal domain where it is completely buried (Wallace *et al*., 2000). The mutation from tryptophan to alanine (cavity forming mutation) was both disruptive and destabilising and the equilibrium unfolding data indicated the accumulation of one or more intermediates (Wallace *et al*., 2000).

Another study showed that the stability of the domain interface contributes towards the catalytic function of hGSTA1-1 (Balchin *et al*., 2010). This was demonstrated by also mutating the conserved residue at the domain interface (Trp21Ala in hGSTA1- 1) and performing enzyme kinetic studies, which revealed that the stability of the domain-domain interface plays a role in mediating the catalytic functionality (Balchin *et al*., 2010). These results highlight the importance of domain-domain contacts and their correct packing contribution in maintaining protein stability and function. Analogous to hGSTA1-1, both monomeric proteins, Clic1 and Grx2 have a conserved residue (Met32 and Met17, respectively) that protrudes from the N-terminal towards the hydrophobic pocket of the C-terminal domain (Figure 1) (Stoychev *et al*., 2009; Parbhoo *et al*., 2011). Equilibrium unfolding studies for Clic1, showed the accumulation of one or more stable intermediates (Stoychev, 2008), similarly in Grx2 equilibrium unfolding data a stable intermediate is observed (Parbhoo, 2010). A two-state unfolding was suggested for GST P1 (Dirr and Reinemer, 1991) however there was another study that proposed three-state equilibrium based on structural data (Aceto *et al.*, 1992). Gildenhuys *et al.* (2010) re-examined the equilibrium unfolding of GST P1-1 and proposed a three-state model. The unfolding/refolding of Sj26GST is a two-state transition which is indicative of cooperativity between the native dimer and unfolded monomer (Kaplan *et al*., 1997). The stability of this protein is highly depended on its concentration. The unfolding of the class mu (GSTM1-1 and GSTM2-2) (Hornby *et al*., 2000) and Sigma (Stevens *et al*., 1998) have multi-step equilibrium pathways which involves stable intermediates. The yeast prion protein 2 has only a single sigmoidal transition state (Perret *et al*., 1999) meaning that the native dimer and the unfolded monomer are not linked. The following classes alpha/pi/Sj26 are stabilised by inter-subunit interactions (Aceto *et al.*, 1992; Gildenhuys *et al*., 2010; Dirr and Reinemer, 1991, and Kaplan *et al*., 1997). The effect of inter-subunit interactions (Dirr *et al*., 2001, Neet *et al*., 1994) on the stability and folding of GST proteins has been studied at great detail

however, relatively little is known about the impact of domain connectivity on the mechanism of protein folding.

1.5.3 Glutaredoxins

The glutaredoxin system was first discovered in 1976 as a dithiol hydrogen donor system for ribonucleotide reductase (Fernandes and Holmegren, 2004). Monothiol glutaredoxins (Grxs) play important roles in maintaining redox homeostasis in living cells and are conserved across species (Fernandes and Holmegren, 2004). Studies have also indicated that the glutaredoxin family serve as hydrogen donors for the reduction of arsenate by ArsC (Shi *et al*., 1999). The glutaredoxin family has diversified over the years and is classified into three categories (Vlamis-Gardikas *et al*., 2002). The first class is the classical glutaredoxin distinguished by their small sizes (10 kDa); the second class is structurally related to the GSTs. The last class is defined by having a monothiol active site. There are three known members of *Escherichia coli* Glutaredoxins (Grx) namely Grx1, Grx2 and Grx3 (Aslund *et al*., 1994). Both Grx1 and Grx3 share sequence identity of over 33 % (Aslund *et al*., 1994) and belong to the first category. For this study, the focus will be on the second category, and Grx2 falls under this category. Grx2 differs from the other glutaredoxins by its large size (24.3 kDa) and its low amino acid sequence homology (Holmgren *et al.*, 1995). Grx2 varies from the other glutaredoxin because it does not show any activity as hydrogen donor for ribonucleotide reductase (Aslund *et al*., 1994). It has also been shown that Grx2 plays a central role in the response of mitochondria to both redox signals and oxidative stress by facilitating the interplay between the mitochondrial glutathione pool and protein thiols (Beer *et al*., 2004).

Grx2 has two domains, the N-terminal domain (residues 1-72) and a larger C-terminal domain (residues 84-215) as shown in Figure 2. Helix 1 (α1) is parallel to α3 which is connected to the linker region that connects to the C-terminal domain through α4. The active site is located at the end of turn of α1 at the domain interface and has a conserved sequence of C-P-Y-C (Xia *et al*., 2001). The C-terminal domain of Grx2 consists of α helices and 2 short 3_{10} helices joined by loops (Xia *et al.,* 2001) whereas GSTs are mainly α helical, with helices joined by short connections (Sinning *et al*., 1993). It also differs from the GSTs in that the active site is located in the N-terminal domain. An observation of the structure of

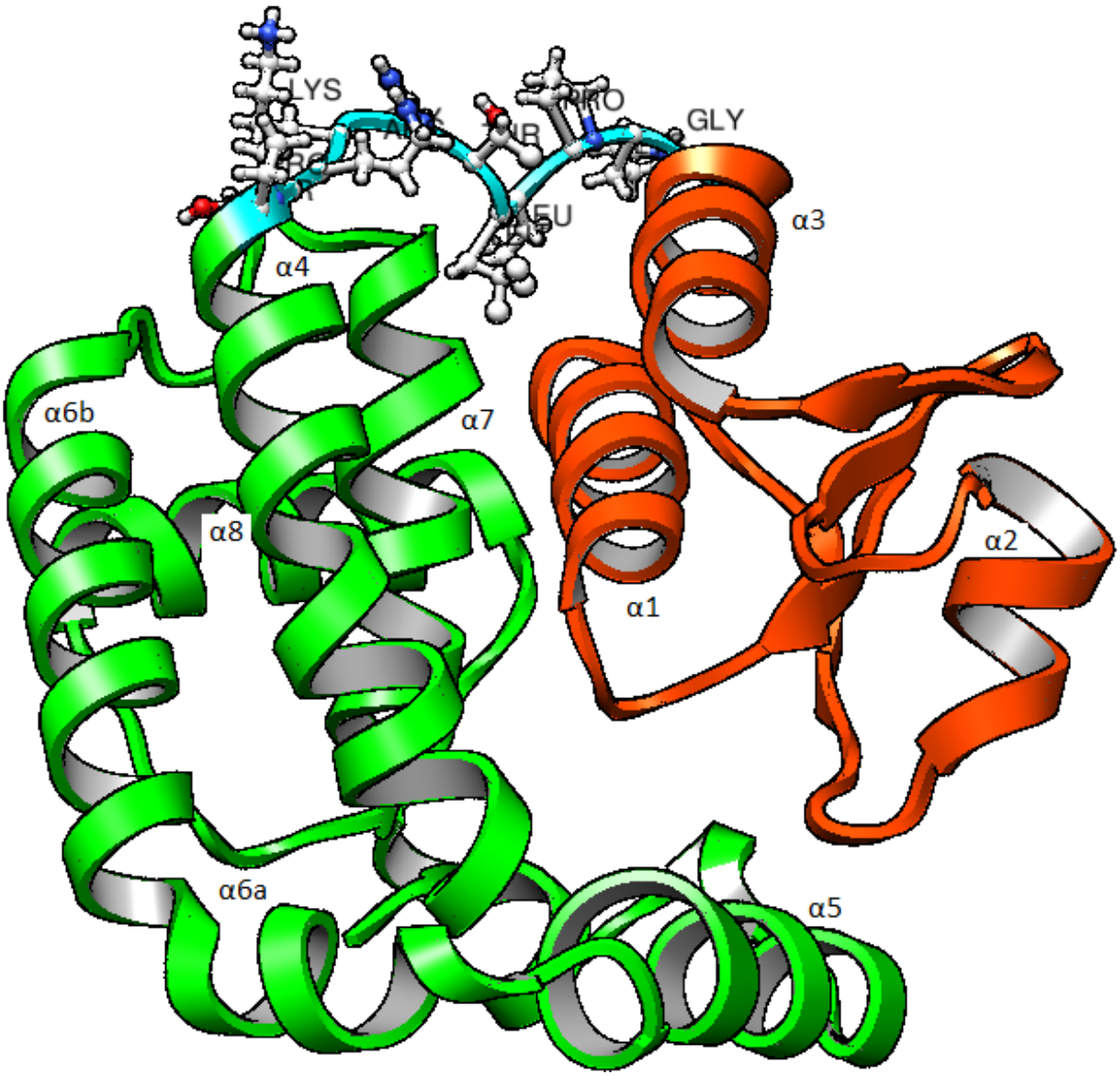


Figure 2. The ribbon structure of Grx2 resolved by NMR (Xia *et al*., 2001). The N-terminal domain is shown in orange-red with the linker region shown in cyan (with the different residues represented in ball and stick) and the C terminal is shown in green. The figure was produced using CHIMERA (Pettersen *et al*., 2004), PDF 1G7O.

Grx2 (Figure 2) shows that the linker region (represented in cyan colour) does not form any regular secondary structure. The linker sequence is made up of the following residues: Asp-Gly-Lys-Pro-Leu-Leu-Thr-Gly-Lys-Arg-Ser. Thus, the linker region is composed of hydrophobic residues (four) and contains three charged and polar residues. No studies have been conducted on the role of the linker region in the stability of protein within the GSTs.

1.5.4 Stability and folding of glutaredoxin2

Previous studies had implied that since the N-terminal domain is less stable than the C-terminal domain in GST proteins, the unfolding could be non-cooperative (Dragani *et al*., 1998, Thompson *et al*., 2006). However, at equilibrium, the unfolding pathway for Grx2 has been shown to be a two state process meaning that the two domains unfold and refold as a cooperative unit (Gildenhuys *et al*., 2008). Interestingly, kinetics data speculate a more complex relationship (Gildenhuys *et al*., 2008). These findings indicate the presence of two unfolding reactions that occur in parallel: major fast phase and a minor slow phase. The following model was proposed for unfolding of Grx2 protein: major track, N_f (80%) $\rightarrow$ U_{cis} (50%) $\leftrightarrow$ U_{trans} (30%); minor track, N_s (20%) U_f (20%), where N_f and N_s represent two native species that unfold in parallel on a fast track (f subscript) and slow track (s subscript), respectively. U_{cis} and U_{trans} are unfolded Grx2 with *cis*- and *trans*-Pro49.

1.6 Objective and aims

The GSTs and GST-like protein family in addition to their catalytic role in detoxification of xenobiotics have been implicated in drug interactions and drug resistance against cancer chemotherapeutic agents, herbicides and microbial antibiotics. GST structural stability is one of the main factors contributing to these effects. The domain-domain interactions play a pivotal role in the stability of GSTs and have been studied relatively well (Wallace *et al*., 2000; Luo *et al*., 2002; Stoychev *et al*., 2009; Parbhoo *et al*., 2011; Balchin *et al.*, 2010). However, the role of the linker domain in the folding and stability of the N- and C- domains in the GSTs and GST-like family has not been evaluated to date. The objective of this study is to investigate the possible role that the linker region might play in the stability of Grx2 which is a monomeric, GST-like protein. This endeavour was conducted in the following manner:

- A conserved residue within the linker domain was mutated in order to assess its impact on the overall stability of Grx2
- The mutant was characterised in terms of structure and function using circular dichroism, fluorescence spectroscopy and ANS binding assay respectively.
- Urea-induced equilibrium studies were conducted using structural probes such as tryptophan fluorescence and circular dichroism.

CHAPTER 2

EXPERIMENTAL PROCEDUREs

2.1 Materials

The cDNA encoding for wild type Grx2 protein has been cloned into the *Escherichia coli* plasmid pET24a (Xia *et al*., 2001) and it was a generous gift from Dr. J. Dyson (The Scripps Institute, California, USA). Ultra pure urea (99.5 %) was purchased from Merck laboratory supplies. DEAE-Sepharose was purchased GE Healthcare Life Sciences (Uppsala, Sweden) and 8-anilino-1-naphthalene sulphonate (ANS) was obtained from Sigma-Aldrich (St. Louis, MO, USA). Isopropyl-1-thio-β-D-galactopyranoside (IPTG) and dithiothretol (DTT) were supplied by Fermentas Life Sciences (St. Leon-Rot, Germany). SDS-PAGE molecular markers were purchased from Fermentas, Inqaba Biotech™ (Pretoria, South Africa). All other reagents were of standard analytical grade. Sequencing to confirm identity of plasmids was conducted at Inqaba Biotech™ (Pretoria, South Africa). The identity of the unknown protein was analysed using mass spectrometer at CSIR (Pretoria, South Africa).

2.2 Methods

2.2.1 Bioinformatics of the GST family

There are sixteen classes of the GST family, eleven of which are: Pi, Mu, Alpha, Theta, Sigma, Omega, Zeta, Delta, Tau, Phi and Beta. The other five classes are *Plasmodium falciparum* GST, the yeast prion protein 2, Grx2, GST-like domain of elongation factor 1-Gamma, and chloride intra-cellular channel 1 (Clic1) and all the GST classes were identified from SCOP)(http://scop.mrc-lmb.cam.ac.uk/scop1.69/data/scop.b.d.gf.b.f.html, Murzin *et al*., 1995). Protein data codes were obtained from (http://www.rcsb.org/pdb/home/home.do).Non- redundant sequences were obtained from http://scop.mrc-lmb.cam.ac.uk/scop1.69/data/scop.b.d.gf.b.f.html , Murzin *et al*., 1995) and a structural/sequence alignment of the GST family linker region was obtained using the UCSF Chimera: a visualization system for exploratory research and analysis (Petterson *et al*., 2004). The results of the alignment were used to 'identify' certain characteristic features within the linker region of all the GST family such as conservation of specific residue(s), variation in length and the flexibility/rigidity of the linker.

2.2.2 Construction of Grx2 Y58W/L78A using site-directed mutagenesis

Two Grx2 variants were used for this study. The first variant was Grx2 Y58W, which had been previously constructed (Gildenhuys *et al*., 2008): tryptophan had been introduced in domain 1 so that tertiary structural changes can be monitored. The second variant Grx2 Y58W/L78A was constructed by introducing a L78A mutation into the Grx2 coding region [which already had another mutation Y58W] of the pET24a construct. This was derived from cloning Grx2 gene into pET-24a with *Ndel* and *BamH1* restriction endonuclease site using the QuikChange™ Site-directed Mutagenesis kit (Papworth *et al*., 1996) from Stratagene (La Jolla, USA). This technique is used to make point mutations and it does so, by utilizing a double stranded plasmid DNA which contains insert of interest and two oligonucleotide primers (both containing the desired mutation) which are complementary to opposite strands of the plasmid. The primers are then extended during thermal cycling by P*fuUltra* HF DNA Polymerase followed by digestion of the parental plasmid with *Dpn 1* endonuclease (Nelson and McClelland, 1992). Oligonucleotide primers for the mutagenesis were designed in accordance with the mutagenesis kit instructions, with the aid of web based program Primer X™ (http://bioinformatics.org/primerx/) and the Gene Runner™ computer software package v3.05. (Hastings Software Inc., NY, USA). Inqaba Biotech™ (Pretoria, South Africa) synthesized the primers:

L78A forward: 5' GATGGCAAACCTTTAGCGACCGGCAAACGTTC 3'

L78A reverse: 5' GAACGTTTGCCGGTCGCTAAAGGTTTGCCATC 3'

The nucleotides highlighted in red represent the mutation that generates the Leu to Ala (GCG) substitution. The thermo-cycling and post thermo-cycling were carried out in accordance to the QuikChange™ Site-directed Mutagenesis kit (Papworth *et al*., 1996) from Stratagene (La Jolla, USA). The construct pET-24a encoding Grx2 Y58W (Gildenhuys *et al*., 2008) was used as a template DNA. The reaction mixture had a final volume of 51µl consisting of 5 µl (10x) reaction buffer, 1 µl (50 ng) double stranded DNA template, 1 µl (125 ng) forward primer, 1 µl (125 ng) reverse primer, 1 µl (10 mM) dNTP mix, 41 µl milli-Q water and 1 µl (2.5 U/µl) Pfu DNA polymerase. The product was generated through sixteen amplification cycles of 30 seconds at 95 ˚C to denature the Y58W DNA, 60 seconds at 55 ˚C to anneal the mutant primers and 60 seconds at 68 ˚C for DNA extension. The reaction products were treated with *DpnI* (10U/µL) (Stratagene, USA) for one hour at 37 ˚C to digest the methylated template. The reaction product was then used to transform *Escherichia coli*

XL1-blue supercompetent cells (Chung *et al*., 1989) which were supplied with the mutagenesis kit. Cells were plated onto the LB agar plates which had been supplemented with kanamycin (30 μg.ml^{-1}) and were incubated for 12-16 hours at 37 ˚C. An overnight culture of a selected colony was grown in LB medium made up of 1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 1.0 % (w/v) NaCl.

2.2.3 Sequencing of Grx2 mutants (Y58W and Y58W/L78A)

Plasmid DNA was extracted from the overnight culture cells, using the Strataprep plasmid miniprep kit from Strategene (La Jolla, USA). The plasmid DNA (10 μl) was sent for sequencing at Inqaba Biotech (Pretoria, South Africa) using the T7 terminator primer. This was done in order to confirm that only the desired mutation (Tyr58-Trp and Tyr58-Trp/ Leu78-Ala) had occurred and also to affirm the absence of any miscellaneous mutations. The resulting nucleotide sequence was aligned with the known mRNA nucleotide sequence encoding *Escherichia coli* Grx2 (Gene ID: X92076), using the alignment tool BLASTN in the NCBI BLAST tools (Altschul *et al*., 1990). The nucleotide sequence was also translated using the TRANSLATE Tool on the ExPASY server, and the resulting peptide sequence was aligned with the known wild-type Grx2 peptide sequence to confirm the presence of Y58W/L78A mutant. Similarly the above mentioned protocol was applied in confirming Grx2 Y58W (Gildenhuys, 2006).

2.2.4 Transformation of Grx2 mutants (Y58W and Y58W/L78A)

Plasmids encoding both Grx2 (Y58W and Y58W/L78A) inserts were used to transform competent cell, *Escherichia coli* BL21 (DE3)/pLysS cells (Lucigen, Middleton, WI, USA) and *Escherichia coli* T7 Express I^q (New England Bio-labs Inc., Ipswich, MA, USA), respectively. The cells were transformed using a one step transformation method described by (Chung *et al*., 1989). About (15 μl) of *Escherichia coli* BL21 (DE3)/pLysS cells and *Escherichia coli* T7 Express I^q competent cells, were thawed on ice for 15 minutes. This step was followed by the addition of 1 μl of plasmid DNA (100 ng.μl^{-1}) and the reaction mixture was incubated for 30 minutes on ice. The cells were then heat-shocked (on a heating block) for 45 seconds at 42 ˚C and then immediately placed on ice for 2 seconds. About 950 μl of fresh sterile 2xYT (yeast tryptone) medium made up of (1.6 % (w/v) tryptone, 1.0 % (w/v) yeast extract, 0.5 % (w/v) NaCl per volume of water) was added to the reaction mixture and incubated on a shaking incubator at 37 ˚C for 90 minutes. The cells were then plated on LB agar plates that

had been made with [1% (w/v) tryptone, 0.5 % (w/v) yeast extract, 1.0 % (w/v) NaCl, 1.5 % (w/v) agar] supplemented with the antibiotics (kanamycin and chloramphenicol) 30 $\mu g.ml^{-1}$. Transformed cells were grown on 30 µg/ml kanamycin and chlorophenicol containing plates because both plasmids confer resistance to these antibiotics (pLysS confers resistance to chlorophenicol and pET24a to kanamycin). The plates were incubated for 12-16 hours at 37 ˚C.

2.2.5 Over-expression and purification of Grx2 mutants (Y58W and Y58W/L78A)

Escherichia coli BL21 (DE3)/pLysS and *Escherichia coli* T7 Express I^q cells (1 single colony) transformed with pET24a plasmid that contains (Y58W and Y58W/L78A) gene inserts, respectively were added to a fresh sterile 2xYT media (1.6 % (w/v) tryptone, 1.0 % (w/v) yeast extract, 0.5 % (w/v) NaCl per volume of water) supplemented with 30 $\mu g.ml^{-1}$ of kanamycin and chlorophenicol. The cells were grown overnight at 37 ˚C with shaking at 250 rpm for 12-16 hours of which a 20x dilution was then used to inoculate into a fresh sterile 2xYT media(1.6 % (w/v) tryptone, 1.0 % (w/v) yeast extract, 0.5 % (w/v) NaCl per volume of water) containing 30 $\mu g.ml^{-1}$ of (kanamycin and chlorophenicol). The cells were then grown at 37 ˚C for *Escherichia coli* BL21 (DE3)/pLysS and 20 ˚C for *Escherichia coli* T7 Express I^q *competent* cells*;* induction trials were conducted for this strain and indicated that this is the best temperature for this cells till they reached OD_{600} (optical density) ~ 600 (2 L). This OD_{600} is considered mid-log phase for cell growth and probably the best time to induce cell growth. *Escherichia coli* BL21 (DE3)/pLysS cells contain pLysS which expresses the T7 lysozyme (Amann *et al*., 1983). The T7 lysozyme reduces the basal expression of target genes by inhibiting T7 RNA polymerase therefore there is a tight control of T7 RNA polymerase. Whereas *Escherichia coli* T7 Express I^q cell is an enhanced derivative of *Escherichia coli* BL21 (DE3)/pLysS, and has the T7 RNA polymerase gene inserted into the lac operon. Once the cells reached OD_{600}, over-expression of cells were induced by the addition of 1 mM IPTG and allowed to grow further for 10 - 16 hours at appropriate temperatures in order to achieve optimum protein expression (37 ˚C for Grx2 Y58W and 20 ˚C for Grx2 Y58W/L78A). The cells were then harvested by centrifugation (4200 x *g*, 25 minutes) using a Sorval model RC5C (SLA-3000) at 4 ˚C. This step was followed by re-suspending the pellet in 35 ml re-suspension buffer (20 mM Tris-HCl, 1 mM EDTA, 2 mM $MgCl_2$, 0.02 (w/v) % NaN_3, pH 10.0). Approximately (10 $mg.ml^{-1}$) of DNase and lysozyme (cautionary step because both

cell lines have lysozyme activity) were added to the cell culture to initiate cell lysis and the cell suspension was rotated at 4 ˚C for 30 minutes. They were then lysed by sonification on ice for 4 cycles on 30 seconds pulsed (intensity 4) on a Sonicator^{TM} Ultrasonic Processor XL (Misonix Inc.). The lysed cells were then centrifuged at (1600 x *g*, 20 minutes) at 4˚ C. Samples of the whole cell extract; soluble and insoluble fractions were collected from the supernatant and a pellet of the lysed cells in order to verify the success of the lyses steps by analysing on a SDS-PAGE gel (see section **2.2.6**).

Grx2 mutants (Y58W and Y58W/L78A) were purified using anion exchange chromatography based on the protocol described previously (Aslund *et al*., 1994; Vlamis-Gardikas *et al*., 1997; Xia *et al.*, 2001) and optimised by (Gildenhuys, 2006; Parbhoo *et al.,* 2011). Depending on the pH of a protein, it may carry either a positive or negative net charge. This depends on the pI (isoelectric point) of a protein. At above or below its pI it may either carry a negative or positive charge, respectively. The pI of Grx2 is (7.7) as determined by using ProtParam tool on the ExPASy proteomics server (Gasteiger *et al*., 2005) therefore above its pI it will be negatively charged. Purification of wild-type Grx2 is based on the principle that at pH 10, the protein (Grx2) is negatively charged (Gasteiger *et al*., 2005). The 30 ml DEAE (diethylaminoethyl) sepharose column (GE Healthcare Life Sciences, Uppsala, Sweden) used for purification is positively charged within a narrow pH range. Therefore the column will bind negatively charged molecules within that pH range. The column was pre-equilibrated with buffer A (20 mM Tris-HCl buffer, pH 10, 0.02 % (w/v) NaN_3) at 20 ˚C using an ÄKTAprime protein purification system attached to a computer with PrimeView software (GE Healthcare Life Sciences, Uppsala, Sweden). This step was followed by injecting the soluble fraction after centrifugation onto a DEAE-Sepharose column. The column was then washed with 10 column volumes of buffer A, followed by 4 column volumes of buffer B (20 mM Tris-HCl buffer, pH 9, 0.02% (w/v) NaN_3). Bound Grx2 Y58W was eluted off the column using a linear pH gradient (10 column volumes) from pH 9 to pH 8. This pH gradient was produced by mixing buffer B with 50 mM Tris-HCl buffer, pH 8, 0.02% (w/v) NaN_3. Although Grx2 Y58W was able to bind the column, Grx2 Y58W/L78A did not. Since Grx2 Y58W/L78A pI (8.84) had slightly increased, the concentration of buffer A and buffer B were increased gradually by 10 mM increments till 50 mM while still maintaining the same pH. However, Grx2 Y58W/L78A was still not purified. Purified Grx2 Y58W/L78A was only noticed on the

column flow through only when the pH 10 of buffer A (20 mM Tris-HCl buffer, pH 10.34, 0.02 % (w/v) NaN_3) was greater than 10, while the other contaminants were bound to the column. The column was regenerated using 1 M NaCl. Different samples were collected during and after purification in order to ensure efficient binding of Grx2 Y58W and to assess purity and molecular mass later on using SDS-PAGE (see section **2.2.6**). The purified Grx2 Y58W or Y58W/L78A protein was then dialysed against three changes of Grx2 storage buffer (50 mM sodium phosphate buffer, pH 7.0, containing 50 mM NaCl, 1 mM DTT and 0.02% (w/v) NaN_3), lasting 4 hours each, after which it was snap frozen and stored at -80˚ C. Before each usage, the frozen protein was thawed at 4˚ C and then dialysed with storage buffers. Unless stated all experiments for Grx2 Y58W or Y58W/L78A were conducted in the Grx2 storage buffer.

2.2.6 SDS-PAGE

Grx2 mutants were separated and characterised using SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis). SDS is an anionic detergent that binds to a protein in the ratio of 1.4 g of SDS/g protein and is able to mask the intrinsic protein charge, whereas electrophoresis refers to the migrations of ions in an electric-field. The mobility of ions is depended on the following factors size, shape and charge. Since the charge is the same because of SDS, molecules in an electric field will separate according to size rather than charge. This protocol was first described by Laemmli (1970) and is a discontinuous system (is termed such because both the acrylamide stacking and separating gels have different ionic strength and pH). This discontinuous system consists of a 15 % (w/v) separating gel acrylamide/bisacrylamide, 0.1 % (w/v) SDS, 0.1 % (w/v) ammonium persulfate, 0.1 % (w/v) TEMED and 0.375 M Tris-HCl buffer, pH 8.8) and a 4 % (w/v) stacking gel acrylamide/bisacrylamide, 0.1 % (w/v) SDS, 0.1 % (w/v) ammonium persulfate 0.1 % (w/v) TEMED and 0.125 M Tris-HCl buffer, pH 6.8). Protein samples collected during over-expression and purification process were diluted 1:1 dilution with a sample buffer [10 % (w/v) glycerol, 2 % (w/v) SDS, 5 % (w/v) β mercaptoethanol, 0.05 % (w/v) bromophenol blue and 0.0625 M Tris-HCl buffer, pH 6.8]. This step was followed by boiling the samples for 2 minutes to make certain that the proteins are denatured. The electrode buffer used was made up of 1 % (w/v) SDS, 0.192 M glycine and 0.025 M Tris, pH 8.5. The protein samples (15 µl) were run using a Mini VE vertical electrophoresis system (Hoefer, San Francisco USA)

with power pack for approximately 1 hour at 160 V to allow for complete separation of proteins. The molecular weight markers used (Fermentas Inqaba Biotech^{TM}: Pretoria; South Africa) contained a mixture of seven proteins namely: β-galactosidase (116 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), lactate dehydrogenase (35kDa), restriction endonuclease Bsp98I (25 kDa), β-lactoglobulin (18.4 kDa) and lysozyme (14.4 kDa). Gels were then stained in 2 % (w/v) Coomassie Blue R250 staining solution containing 13.5 % (v/v) glacial acetic acid and 18.75 % (v/v) ethanol for an hour. This was followed by destaining gels with 40 % (v/v) ethanol and 10 % (v/v) glacial acetic acid overnight until the background was clear.

2.2.7 Peptide sequencing of Grx2 Y58W/L78A using mass spectroscopy

Mass spectroscopy is a technique used in identifying and probing covalent structures of proteins by determining their mass-to-charge ratio of the whole protein (Domon and Aebersold, 2006). Mass spectroscopy function by converting individual molecule into small ions, followed by manipulating those ions using external and magnetic field. This technique achieves its function by having three components.

- The ion source (where samples are ionised usually by a loss of an electron).
- The mass analyser (the ions will be separated according to their mass and charge).
- The detector (separated ions will be detected and displayed on a charge).

Ionisation of molecules can be accomplished in many different ways but the most common procedure is to bombard a sample with a high energy beam electron (this process is known as electron impact). Separation of ions is achieved by focusing and accelerating the ions in a beam, which results in the bending of the magnetic field. When the ion beam experiences a strong magnetic field perpendicular to its direction of motion, ions are deflected in arc whose radius is inversely proportional to the mass of the ion. Lighter ions will be deflected more in comparison to heavier ones and a detector is used to detect the separated ions. Proteins samples analysed on the mass spectroscopy are prepared by digesting the whole protein into small peptides, using digestive enzymes such as trypsin. Digestion is then followed by separating the peptides using HPLC (high pressure liquid chromatography) that is directly coupled to the mass spectrometer. The peptides are then eluted from the column by increasing organic content (separated due to hydrophobicity). Grx2 Y58W/L78A protein was sent for sequencing at CSIR (Pretoria, South Africa) to confirm the identity of the

protein. Protein sample was first separated by using SDS-PAGE (see section **2.2.6**). The gel-protein bands of interest were cut into cubes and sent for sequencing at CSIR (Pretoria, South Africa). The Schevchenko *et al*. (2007) protocol was followed in treating the sample. The gel bands of interest were first destained using 50 mM NH_4HCO_3 (ammonium bicarbonate) and 50 % (v/v) MeOH. This step was followed by the reduction in gel-protein using (50 mM DTT in 25 mM NH_4HCO_3) and alkylation using 55 mM iodoacetamide in 25 mM NH_4HCO_3. Protein were then digested over night at 37 °C using freshly prepared 10 ng/µl tryspin solution in 50 mM NH_4HCO_3. Peptides were extracted using 50 % (v/v) ACN (acetonitrile)/5 % (v/v) FA (formic acid) and vacuum dried. Peptides were purified before loading on a mass spectrometer. This was conducted using protocol by (Rappsilber *et al*., 2003), the peptides are re-suspended in 5 % (v/v) FA and loaded on pre-equilibrated StageTip column, washed then eluted using 50 % (v/v) MeOH/ 5 % (v/v) FA. Purified samples are then loaded on a Proxeon (ES 380) NanoES capillary and ionized using cm IonSpray voltage of 900V. A TOF-MS spectrum was acquired in positive ion mode using the range of 400 – 1600 m/z. Product ion spectra were generated using the Information Dependent Acquisition (IDA) function of the Analyst QS 2.0 software where multiply charged parent ions were selected for fragmentation by collision induced dissociation (CID), using nitrogen as collision gas. Protein identification was performed using NCBI's msdb database and the Paraghon™ algorithm *Thorough* search in Protein Pilot v3.0 (Shilov *et al*., 2007).

2.2.8 Protein concentration determination

The concentration of Grx2 mutants were determined using two different methods. The first one was determined spectrophotometrically using a Jasco V-630 UV-VIS spectrophotometer and by applying the Beer-Lambert law where:

$$A = \varepsilon_\lambda c l \qquad (3)$$

A is the absorbance at 280 nm, ε is the extinction coefficient of a protein at a wavelength, c is the concentration in M and l is the path length of light in centimetres. The wild-type Grx2 extinction coefficient at 280 nm is 21680 $M^{-1}.cm^{-1}$ (Vlamis-Gardikas *et al*., 1997). The extinction coefficient of Grx2 mutants were calculated at 280 nm using the formula by Perkins (1986). This formula by (Perkins, 1986) calculates the average values of the extinction coefficient of tryptophan (5550 $M^{-1}. cm^{-1}$), tyrosine (1340 $M^{-1}. cm^{-1}$) and cysteine (150 $M^{-1}. cm^{-1}$).

$$\varepsilon_{280}\ M^{-1}.\ cm^{-1} = [\ 5550\ x\ (\sum \text{Trp residues})] + [\ \sum \text{Tyr residues})] + [\ 150\ x\ (\sum \text{Cys residues})\ ] \quad (4)$$

For both Grx2 mutants (Y58W and Y58W/L78A) the extinction coefficient at 280 nm was calculated using Perkins (1986).

$$\varepsilon_{280}\ M^{-1}.\ cm^{-1} = 5550 \sum (Trp) + 1340 \sum (Tyr) + 150 \sum (Cys)$$
$$= 5550\ (3) + 1340\ (8) + (2)$$
$$= 27670\ M^{-1}.\ cm^{-1}$$

The concentration was determined by fitting a linear regression to 5 or more points from a serial dilution (measurements were conducted in triplicates). All readings were buffer corrected with the appropriate buffer used for the concentration determination. The following equation was used for buffer correcting to eliminate any possible aggregates and light scattering interference at 340 nm:

$$A_{280\ (corrected)} = (A_{280\ (protein)} - A_{280\ (buffer)}) - (A_{340\ (protein)} - A_{340\ (buffer)}) \quad (5)$$

A quartz cuvette of 1 cm path length was used for all readings.

The second method used in determining Grx2 mutants (Y58W and Y58W/L78A) protein concentration was using the Bradford assay (Bradford, 1976). This assay is based on the fact that Coomasie Blue G-250 binds to proteins (Bradford, 1976) and under acid condition the dye is fully protonated and becomes red in colour, but when bound to a protein it is converted to a stable unprotonated blue form which can be measured spectroscopically at 595 nm. A known BSA (bovine serum albumin) concentration was mixed with the Bradford assay; measured at 595 nm and the concentrations of both Grx2 mutants were extrapolated from the standard curve. A blank sample consisted of the buffer and the Bradford assay.

2.2.9 Circular dichroism spectroscopy

Circular dichroism spectroscopy is technique that is used to study proteins and is based on the differential absorption of left- or right- handed circularly polarised light by optically active molecules. The ability of proteins to absorb light arises from chromophores (peptide bond, aromatic rings, disulfide groups and prosthetic groups (Woody, 1995). These chromophores have characteristic absorption bands in specific areas of visible and ultraviolet wavelength. The most abundant chromophore group is found in the peptide linking amino acid in proteins (Woody, 1995). The secondary structures adopted by proteins have a distinct CD spectrum in the far ultraviolet regions (Woody, 1995) and hence it is used to assess the secondary structure of proteins. Disulfide and aromatic rings are absorbed

predominantly in the near-UV region. The signal for near-UV will depend on the number of each aromatic amino acid (Phe, Tyr and Trp) present, the nature of their environment and their spatial disposition (Kelly *et al*., 2005).

2.2.9.1 Far-UV spectroscopy

The far-UV spectra (190 nm – 250 nm) were recorded for both Grx2 mutants (Y58W 758W/L78A) using 5 µM of protein. Protein samples were prepared in storage buffer (50 mM sodium phosphate buffer, pH 7.0, containing 50 mM NaCl, 1 mM DTT and 0.02 % (w/v) NaN_3) and all measurements were taken using a Jasco J-810 spectropolarimeter with Spectra Manager software v1.5.00 (Jasco Inc., Tokyo, Japan) using a pathlength of 2 mm. All spectra measurements were recorded at a temperature of 20 ˚ C, bandwidth of 5 nm, data pitch 0.2 nm and the scan speed of 200 nm/min. Buffer contributions were subtracted from all data collected. The spectra were normalised by calculating the mean residue ellipticity [θ] deg.cm^2.dmol^{-1}.residue^{-1} using the following equation (Woody, 1995):

$$[\theta] = (100\ \theta)/cnl \qquad (6)$$

where (θ) is the ellipticity signal in mdeg, c is the protein concentration in mM, n is the total number of residues in the protein chain and l is the path length in cm.

2.2.9.2 Near-UV CD

The near-UV CD spectra of Grx2 mutants were recorded in the regions 250-300 nm using a Jasco model J-180 to characterise tertiary structures. As mentioned earlier in section **2.2.9** that proteins have chromophores that absorb and have characteristics bands in the UV and visible region (Woody, 1995). Near-UV is able to provide structural detail of the tertiary structures. About 20 µM of both Grx2 mutants at 5˚ C in storage buffer were used to obtain near-UV CD spectra. All near-UV CD spectra were recorded at a scanning speed of 100 nm.min^{-1}, using a 1 cm path length cuvette. The sensitivity was set to 35 high (10 mdeg), data pitch was 0.05 nm, response 1 sec, bandwidth 0.5 nm, and each spectrum was the result of 10 accumulations.

2.2.10 Fluorescence spectroscopy

Fluorescence is the emission of light that results when molecules are excited (absorb light) at certain wavelength and then emit light at a longer wavelength. The tertiary structures of proteins can be characterised by fluorescence spectroscopy because of the aromatic residues (Phe, Tyr and Trp) (Lacowicz, 1999). The fluorescence of tryptophan is much higher

than that of other residues. This is due to the indole ring of the tryptophan which is sensitive to structural alterations and can be used as a probe to provide information on its local environment (Lacowicks, 1999). Tryptophan's residues experience a red shift from the native state emission maxima to denatured state emission maximum of 355 nm in the presence of a denaturant (Teale, 1960). All the fluorescence measurements were recorded using the Jasco 6300 in a quartz cuvette with 10 mm path length.

2.2.10.1 Intrinsic fluorescence due to tryptophan

Wild-type Grx2 has only two tryptophan residues (Trp89 + Trp190) located in domain 2 (Xia *et al*., 2001). Both the Grx2 mutants (Y58W and Y58W/L78A) have three tryptophan (on the account that tyrosine 58 has been mutated to tryptophan), one in domain 1 and two in domain 2 this will provide the local environment surrounding the tryptophan's. The tryptophan residues were selectively selected at a wavelength of 295 nm. The fluorescence emission spectra for Grx2 mutants were recorded using 5 µM of protein in the range of 280 – 450 nm, both the excitation and emission slid width used were 5 nm. The measurements were taken at a temperature of 20˚ C; with the scan speed 200 $nm.min^{1}$. The buffer contributions were subtracted.

2.2.10.2 Extrinsic Fluorescence: binding of 8-Anilino-1-naphthalene sulphonate (ANS)

8-Anilino-1-naphthalene sulphonate (ANS) is an amphipathic dye that is able to bind to exposed hydrophobic clusters on a protein (Semisotnov, 1991). Hence it is used as a probe for diagnosing conformational changes induced during protein unfolding and folding reactions (Ptitsyn *et al*., 1990). Normally when ANS in water is excited at a wavelength of 390 nm, it emits at wavelength of 540 nm. However, when ANS is bound to the exposed non-polar patches of a protein, the emission wavelength is usually shifted to lower wavelengths around 460 nm, depending on the hydrophobicity of the ANS binding site on the protein. The increase in fluorescence quantum yield is due to the exposure of hydrophobic core regions that are inaccessible to the dye in the native structure (Semisotnov, 1991). A stock of ANS (2 mM) was prepared in Grx2 storage buffer. The concentration was confirmed spectrophotometrically using a series of dilution of stock ANS and an extinction coefficient of ε_{350} = 4 950 $M^{-1}cm^{-1}$ (Weber and Young, 1964). Equilibrium unfolding in the presence of ANS was prepared by allowing 2 µM of protein (Grx2 Y58W/L78A) to unfold in 0 – 8 M urea for 1 hour as described (see section **2.2.11.2**). About

200 µM ANS was added to all the samples and incubated at room temperature (20˚ C) for at least an hour (to allow binding of ANS to any exposed regions) before further experiments were performed. A series of blank samples were also generated containing all the above mentioned except for the protein (Grx2 Y58W/L78A). This was performed in order to correct for the increasing free ANS emission signal. Each sample was excited at 390 nm and the emission was recorded from 390 - 600 nm. The scan speed used was 400 $nm.min^{-1}$, excitation and emission slit width 5 nm, respectively. Fluorescence emission intensities at 500 nm were plotted as a function of urea concentration.

2.2.11 Urea-induced equilibrium unfolding

2.2.11.1 Determination of reversibility of unfolding in urea

Reversibility is an obligatory prerequisite for the interpretation of equilibrium unfolding studies because equilibrium between the native and the unfolded state can only occur if the reaction is reversible. The reversibility of protein unfolding can be established by allowing the reaction to reach equilibrium at a denaturant concentration which allows complete unfolding. The shift towards the unfolded state in the presence of a denaturant will give an indication of thermodynamic parameters that define the stability of proteins. The reversibility of unfolding events was determined for Grx2 Y58W/L78A mutant. The Grx2 Y58W/L78A mutant was allowed to unfold in the presence of 8 M urea for 1 hour at 20 ˚C, after which refolding was initiated by a 10-fold dilution with Grx2 storage buffer for an hour at 20 ˚C. The refolded state was assessed using structural probes; far-ultraviolet circular dichroism (see section **2.2.9.1**) and intrinsic tryptophan fluorescence (see section **2.2.10.1**). All of the urea solutions used for experiments were prepared using the method of Pace (Pace, 1986). The stock urea solution was filtered, pH adjusted to 7 and the concentration of the stock urea solution was confirmed using an Atago R5000 refractometer (Tokyo, Japan) and the method of Pace (1986).

2.2.11.2 Urea-induced equilibrium unfolding studies

Ultra-pure urea in Grx2 storage buffer (50 mM sodium phosphate buffer, pH 7.0, containing 50 mM NaCl, 1 mM DTT and 0.02 % (w/v) NaN_3) was used as a denaturant because urea is able to bind to the amino acids side chains or peptide group and as the protein unfolds, more peptide groups are exposed. Consequently urea is able to shift the equilibrium between the native and folded towards the unfolded state (Pace, 1986). The idea was to

construct unfolding curve to monitor the structural changes induced by urea. When one needs to determine the stability of proteins, denaturation curves (unfolding curves) are generally used (Pace, 1986). Parameters such as the equilibrium constants (K_{eq}), ΔG° (H_2O) and *m-value* can be determined from the denaturation curves. Triplicates of 1 ml samples of 2 µM Grx2 Y58W/L78A in urea concentration ranging from 0 – 8 M, in 0.2 M intervals were prepared by adding denaturant to protein, and mixing gently. The different samples were allowed to unfold for an hour at 20˚ C to allow for equilibrium to be reached. Following that process the samples were monitored using the above mentioned probes.

2.2.11.3 Equilibrium unfolded data-fitting

The far-UV CD and fluorescence monitored transitions of Grx2 Y58W/L78A were generated after making certain that these two criteria were met:

1. That the reaction was reversible.
2. The reaction was at equilibrium before measurements were taken (Pace, 1986).

The CD data of Grx2 Y58W/L78A were plotted as the average of three data sets of ellipticity at 222 nm as a function of urea. The urea-induced equilibrium unfolding data were analysed according to a monomeric two-state unfolding process (Pace, 1986) because of the following reasons:

1. The unfolding curve was monophasic, showing no shoulders or inflections in the transition state.
2. The curves generated by the techniques (far-UV CD and fluorescence) were superimposable.

In a two-state monomer unfolding transitions, only two species exist: the native (N) and the unfolded forms (U).

$$N \leftrightarrow U \qquad (7)$$

A two-state unfolding curve has three regions (Pace and Stoltz, 1997):

1. The pre-transition region, in which the native state predominates.
2. The transition region, where native and unfolding species exist in varying concentrations, and cooperative unfolding occurs
3. The post-transition region, in which the unfolded state predominates.

In a two-state unfolding transition, only two species are predominant at significant concentration, which are the native and the unfolded states (Pace, 1986). Therefore, for a two-state mechanism:

$$f_N + f_U = 1 \quad \therefore f_N = 1 - f_U \tag{8}$$

where f_N is the fraction of folded protein and f_U is the fraction of unfolded protein. At any point during unfolding, there is a contribution to the signal from the concentration of both species:

$$y = y_N f_N + y_U f_U \tag{9}$$

where y is the signal obtained from using circular dichroism or fluorescence as probes. y_N and y_U represent the measured properties of the folded and unfolded state, respectively. In addition y_N and y_U can be calculated by linear extrapolation of the pre- and post- transition regions. Both equation (8) and (9) can be combined to calculate the fraction of unfolded protein.

$$\therefore y = y_N (1 - f_U) + y_U f_U$$

$$\therefore y = y_N - y_N f_U + y_U f_U$$

$$\therefore y - y_N = f_U (y_U - y_N)$$

$$\therefore f_U = (y_N - y)/(y_N - y_U) \tag{10}$$

Similarly the fraction of folded protein or native protein can be obtained

$$f_N = (y_U - y)/ y_N - y_U) \tag{11}$$

The conformational stability of a protein i.e. the free energy upon unfolding *(ΔG°)* can be calculated from denaturation curves. The equilibrium constant (K_{eq}) can be calculated by the equation:

$$K_{eq} = f_U/f_N \tag{12}$$

Equation 10 and 11 can be substituted into equation 12 to give:

$$K_{eq} = (y_N - y)/(y - y_U)$$

and

$$\Delta G° = - RT \ln K_{eq}$$

$$\therefore \Delta G° = - RT \ln K_{eq} \tag{13}$$

$$= - RT \ln[(y_N - y)/(y - y_U)]$$

where *(ΔG°)* is the free energy upon unfolding, *R* is the gas constant (1.987 cal. mol^{-1}. K^{-1}) and T is the temperature constant in Kelvin.

In order to obtain equilibrium parameters [(ΔG° (H_2O) and *m*-value)] linear extrapolation methods were used, previously described by (Green and Pace, 1973). The method used assumes that ΔG° values have a linear dependence on the denaturant concentration. So in order to obtain ΔG° in the absence of urea ΔG° (H_2O), the results are extrapolated back to zero urea concentrations according to the following equation:

$$\Delta G° = \Delta G° (H_2O) - m\,[denaturant] \qquad (14)$$

The *m*-value indicates the measure of dependence of the free energy on denaturant concentration (Shirley, 1995); it is also a good indicator of co-operativity and is related to the amount of surface area exposed upon denaturation.

Combining equation 12, 13 and 14 gives

$$y = [y_N + y_U * e^{-(\Delta G (H2O) - m\,[denaturant])/RT}] / [1 + e^{-(\Delta G (H2O) - m[denaturant])/RT}] \qquad (15)$$

The equilibrium data obtained were fitted to equation 15 using SigmaPlot version 11.0 (Systat Software Inc., Chicago, IL, USA) and the parameters ΔG° (H_2O) and *m*-value were obtained.

CHAPTER 3

RESULTS

3.1 Structural alignment of GST linkers

A multiple structural-based sequence alignment of the linker regions belonging to the selected members of the GST and GST-like protein family was performed using CHIMERA (Pettersen *et al*., 2004). Table 1 shows a list of the different classes of GST and GST-like protein family that were selected in this particular study to obtain the multiple linker alignment. A leucine residue located within the linker regions appeared at equivalent position in over 60 % of the selected members of the GST family (Figure 3A and Figure 3B). However, other residues namely Asn, Gln, Met, Ser, Ile and Try were observed in other GST family (Figure 3A). Despite the occurance of the above mentioned amino acids, 100 % of the GST family structurally aligned with leucine located at an equivalent position (Figure 3B). The frequency and structural conservation of this amino acids at this position suggest that leucine might play a significant role in GST function. The length of the linker region within the GSTs was found to be in the range of 10 - 15 amino acids. It would be interesting for future studies to investigate the length of the linker in Grx2.

Table 1. List of GST proteins and GST like protein used to obtain the multiple linker alignment.

GST class	PDB code	Reference
Alpha (hGST A1-1)	1K3L	Le Trong *et al*., 2002
Beta (bGST B1)	1PMT	Rossjohn *et al*., 1998
Clic1 (GST-like)	1KOM	Harrop *et al*., 2001
Delta (Ad-GST D1-3)	1JLV	Oakley *et al*., 2001
Elongation factor1 gamma (GST-like)	1NHY	Jeppesen *et al*.,2002
Grx2 (GST-like)	1G7O	Xia *et al*., 2001
Mu (cGST M1-1)	1C72	Chern *et al*., 2000
Omega (hGST O1-1)	1EEM	Board *et al*., 2000
Phi (AtGST F1)	1GNW	Reinemer *et al*., 1996
Pi (Ov-GST P2-2)	1TU7	Perbandt *et al*., 2005
PfGST	1OKT	Fritz-Wolf *et al*., 2003
Sigma (lGST S1)	2WB9	Line *et al*., 2009
Stringent starvation protein A (SspA) (GST-like)	3LYK	Ramogopal *et al*., 2010
Tau (TaGST U4-4)	1GWC	Thom *et al*., 2002
Theta (hGST T1-1)	2C3N	Tars *et al*., 2006
Yeast prion protein (Ure2)	1G6W	Bousset *et al*., 2001
Zeta (hGST Z1-1)	1E6B	Thom *et al*., 2001

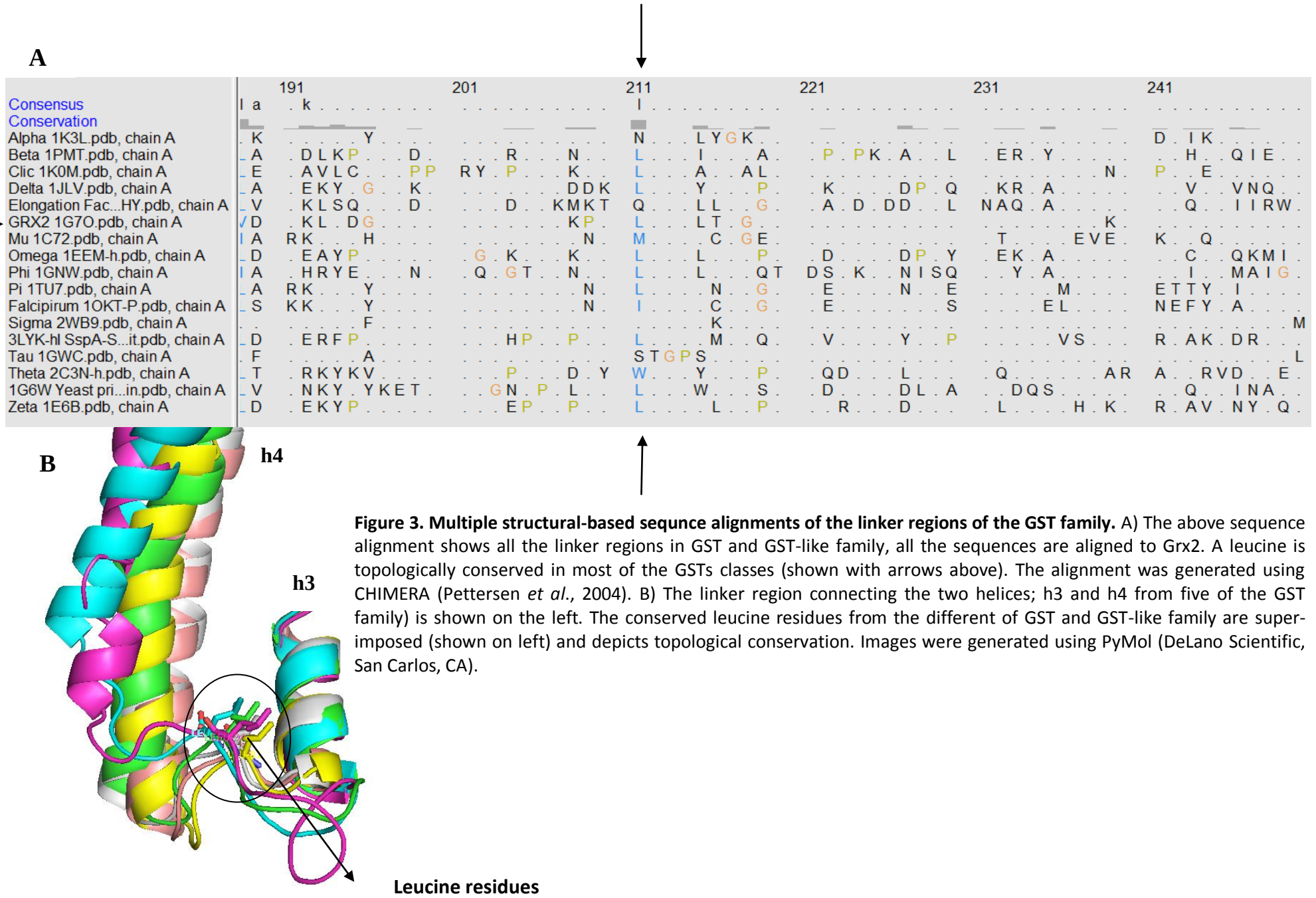


Figure 3. Multiple structural-based sequnce alignments of the linker regions of the GST family. A) The above sequence alignment shows all the linker regions in GST and GST-like family, all the sequences are aligned to Grx2. A leucine is topologically conserved in most of the GSTs classes (shown with arrows above). The alignment was generated using CHIMERA (Pettersen *et al.*, 2004). B) The linker region connecting the two helices; h3 and h4 from five of the GST family) is shown on the left. The conserved leucine residues from the different of GST and GST-like family are super-imposed (shown on left) and depicts topological conservation. Images were generated using PyMol (DeLano Scientific, San Carlos, CA).

3.2 Plasmid verification

Using Grx2 Y58W plasmid DNA as a template, the L78A mutant was created using site-directed mutagenesis. The sequencing results of the mutated plasmid DNA (Figure 4A) indicated that no other mutations besides Grx2 Y58W (Figure 4B-A) and Grx2 Y58W/L78A (Figure 4B-B) were incorporated during mutagenesis.

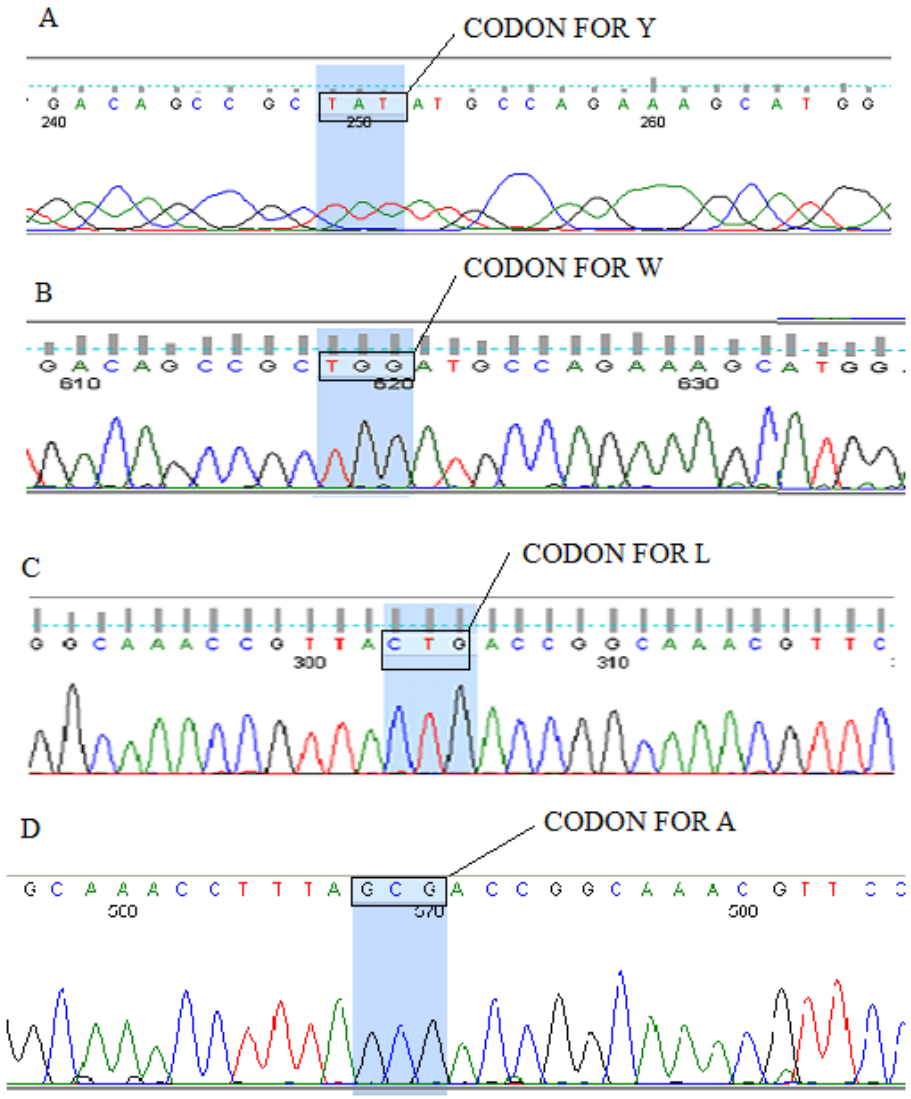


Figure 4A. The sequencing results of plasmid Grx2 are presented above. A) The wild-type TAT codon for a tyrosine 58 (Y), this is the wild-type sequence for Grx2. B) The wild-type sequence that codes for tyrosine 58 has been replaced with a TGG codon for tryptophan (W) . Once the plasmid DNA had been confirmed that it had been incorporated the Y58W mutation it was used to generate a second mutation. C) The CTG codon for leucine (L) residue in Y58W plasmid DNA has been mutated to a GCG codon for alanine (A) residue as illustrated in (D). The sequencing results were viewed using Finch TV version 1.4.0 (http://www.geospiza.com/FinchTV: Geospiza Inc). No additional mutations occurred during mutagenesis.

A

```
>lcl|8367 unnamed protein product
Length=215

 Score =  430 bits (1106),  Expect = 1e-125
 Identities = 214/215 (99%), Positives = 215/215 (100%), Gaps = 0/215 (0%)
 Frame = +1

Query  94   MKLYIYDHCPYCLKARMIFGLKNIPVELHVLLNDDAETPTRMVGQKQVPILQKDDSRWMP  273
            MKLYIYDHCPYCLKARMIFGLKNIPVELHVLLNDDAETPTRMVGQKQVPILQKDDSR+MP
Sbjct  1    MKLYIYDHCPYCLKARMIFGLKNIPVELHVLLNDDAETPTRMVGQKQVPILQKDDSRYMP  60

Query  274  ESMDIVHYVDKLDGKPLLTGKRSPAIEEWLRKVNGYANKLLLPRFAKSAFDEFSTPAARK  453
            ESMDIVHYVDKLDGKPLLTGKRSPAIEEWLRKVNGYANKLLLPRFAKSAFDEFSTPAARK
Sbjct  61   ESMDIVHYVDKLDGKPLLTGKRSPAIEEWLRKVNGYANKLLLPRFAKSAFDEFSTPAARK  120

Query  454  YFVDKKEASAGNFADLLAHSDGLIKNISDDLRALDKLIVKPNAVNGELSEDDIQLFPLLR  633
            YFVDKKEASAGNFADLLAHSDGLIKNISDDLRALDKLIVKPNAVNGELSEDDIQLFPLLR
Sbjct  121  YFVDKKEASAGNFADLLAHSDGLIKNISDDLRALDKLIVKPNAVNGELSEDDIQLFPLLR  180

Query  634  NLTLVAGINWPSRVADYRDNMAKQTQINLLSSMAI  738
            NLTLVAGINWPSRVADYRDNMAKQTQINLLSSMAI
Sbjct  181  NLTLVAGINWPSRVADYRDNMAKQTQINLLSSMAI  215
```

B

```
>lcl|18863 unnamed protein product
Length=215

 Score =  428 bits (1101),  Expect = 5e-125
 Identities = 213/215 (99%), Positives = 214/215 (99%), Gaps = 0/215 (0%)
 Frame = -1

Query  751  MKLYIYDHCPYCLKARMIFGLKNIPVELHVLLNDDAETPTRMVGQKQVPILQKDDSRWMP  572
            MKLYIYDHCPYCLKARMIFGLKNIPVELHVLLNDDAETPTRMVGQKQVPILQKDDSR+MP
Sbjct  1    MKLYIYDHCPYCLKARMIFGLKNIPVELHVLLNDDAETPTRMVGQKQVPILQKDDSRYMP  60

Query  571  ESMDIVHYVDKLDGKPLATGKRSPAIEEWLRKVNGYANKLLLPRFAKSAFDEFSTPAARK  392
            ESMDIVHYVDKLDGKPL TGKRSPAIEEWLRKVNGYANKLLLPRFAKSAFDEFSTPAARK
Sbjct  61   ESMDIVHYVDKLDGKPLLTGKRSPAIEEWLRKVNGYANKLLLPRFAKSAFDEFSTPAARK  120

Query  391  YFVDKKEASAGNFADLLAHSDGLIKNISDDLRALDKLIVKPNAVNGELSEDDIQLFPLLR  212
            YFVDKKEASAGNFADLLAHSDGLIKNISDDLRALDKLIVKPNAVNGELSEDDIQLFPLLR
Sbjct  121  YFVDKKEASAGNFADLLAHSDGLIKNISDDLRALDKLIVKPNAVNGELSEDDIQLFPLLR  180

Query  211  NLTLVAGINWPSRVADYRDNMAKQTQINLLSSMAI  107
            NLTLVAGINWPSRVADYRDNMAKQTQINLLSSMAI
Sbjct  181  NLTLVAGINWPSRVADYRDNMAKQTQINLLSSMAI  215
```

Figure 4B. The predicted amino acid sequence for Grx2 Y58W and Grx2 Y58W/L78A. This results show the alignment of the predicted protein sequence versus the wild-type for Grx2, query is the sequence that has been mutated while subject is the wild-type sequence. A) Grx2 Y58W, only tyrosine has been changed to tryptophan as depicted by the arrow. B) The above sequence Grx2 Y58W/L78A incorporates both mutations, tryptophan →tyrosine and leucine→alanine. The resuts indicate that no other miscellaneuos mutations had been incorporated. The results were generated using TRANSLATE from ExPASY server.

3.3 Over expression of Grx2 mutants

The expression system initially used for both Grx2 mutants consisted of *Escherichia coli* BL21 (DE3)/pLysS cells transformed with pET-24a+ plasmid containing the Grx2 Y58W and Grx2 Y58W/L78A insert. The recombinant Grx2 Y58W protein was soluble when over-expressed in *Escherichia coli* BL21 (DE3)/pLysS at 37˚C (Gildenhuys *et al*., 2008). However, under these conditions the system produced minute soluble protein for Grx2 Y58W/L78A (Figure 5A – lane 3, shown as arrow) whilst it was apparent that most of the protein was located in the pellet (Figure 5A – lane 2, shown as red circle). Could this be due to insufficient cell lyses since lysozyme was added at 4 ˚ C instead of the recommended 37˚ C? However, both cell lines contain lysozyme activity (Amann *et al*., 1983) and sonification was used as well to lyse the cells; therefore adding lysozyme was merely a cautionary step. Thus, the cell line was then changed to the derivative of *Escherichia coli* BL21 (DE3)/pLysS, T7 Express I^q. SDS-PAGE gel showed that the over-expression of Grx2 Y58W/L78A protein using T7 Express I^q resulted in most of the protein being soluble at 20˚ C as depicted in Figure 5B (lane 3, shown as arrow). The difference between the two cell lines is explained in section (**2.2.5**) which made it possible for the system to produce soluble Grx2 Y58W/L78A protein. These observations suggested that the mutation Y58W/L78A might have an effect on the correct folding of Grx2 in the cytoplasm of the *Escherichia coli.*

3.4 Purification and purity determination of Grx2 mutants

Both Grx2 mutants were initially purified using a previous protocol (Vlamis-Gardikas *et al*., 1997; Xia *et al.*, 2001) which had been optimised (Gildenhuys, 2006; Parbhoo *et al*., 2011). The Grx2 Y58W protein was purified using DEAE chromatography and in addition a linear pH gradient was used which resulted in a single peak (Figure 6A). Different samples collected during the Grx2 Y58W purification process are also shown on the SDS-PAGE (Figure 6A). Fractions 38-52 from the observed single peak were collected to assess purity using SDS-PAGE. Grx2 Y58W had no contaminants as shown on SDS-PAGE intense single band Figure 6B (lane 1 and 3). Once Grx2 Y58W purity had been validated, fraction 38-52 were pooled together and concentrated. The size of Grx2 Y58W protein was determined to be 25 kDa when compared to the molecular mass protein markers (Figure 6C). The distinct 25 kDa band corresponds to the published molecular mass data for wild type Grx2 (Gildenhuys *et al*., 2008; Parbhoo *et al*., 2011; Xia *et al*., 2001; Vlamis-Gardikas *et al*., 1997). However, Grx2

Y58W/L78A mutant showed a different behaviour from the wild-type during protein purification: it did not bind to Sepharose column. The following conditions were conducted to optimise the purification process of Grx2 Y58W/L78A:

- A new column was tried but it also produced no recombinant protein. The rationale behind the new column was perhaps the old column was not properly regenerated when using 1 M NaOH and might still contain degraded DNA and RNA (although DNAse and RNAse were added to the cell culture) which could have disrupted the binding capabilities of the DEAE column by degrading Grx2 Y58W/L78A protein.
- The concentrations of the different buffers buffer A (20 mM Tris-HCl buffer, pH 10, 0.02 % (w/v) NaN_3) and buffer B (20 mM Tris-HCl buffer, pH 9, 0.02 % (w/v) NaN_3) used to elute Grx2 were increased each by 10 mM increments till 50 mM but still no changes were observed.
- The pH was also changed because the pI for the mutant Grx2 Y58W/L78A had shifted from 7.7 to 8.84. However the binding of the protein to the column was still not efficient.

It was only after buffer A (50 mM Tris-HCl buffer, pH 10.4, 0.02 % (w/v) NaN_3) pH was changed to 10.4 that a peculiar behaviour was observed. Figure 7A illustrates how an unidentified protein (referred to as unknown protein) is eluting at a pH 10.4 buffer wash (Figure 7A) and not being bound to the column as expected, shown by the two peaks in Figure 7A. It was later observed that unknown protein was pure and the size corresponds to the molecular weight of Grx2 as shown by the intense band Figure 7B lane 2**.** To confirm that the band shown on the gel (Figure 7B, lane 2) was indeed Grx2 Y58W/L78A protein further verification studies had to be taken.

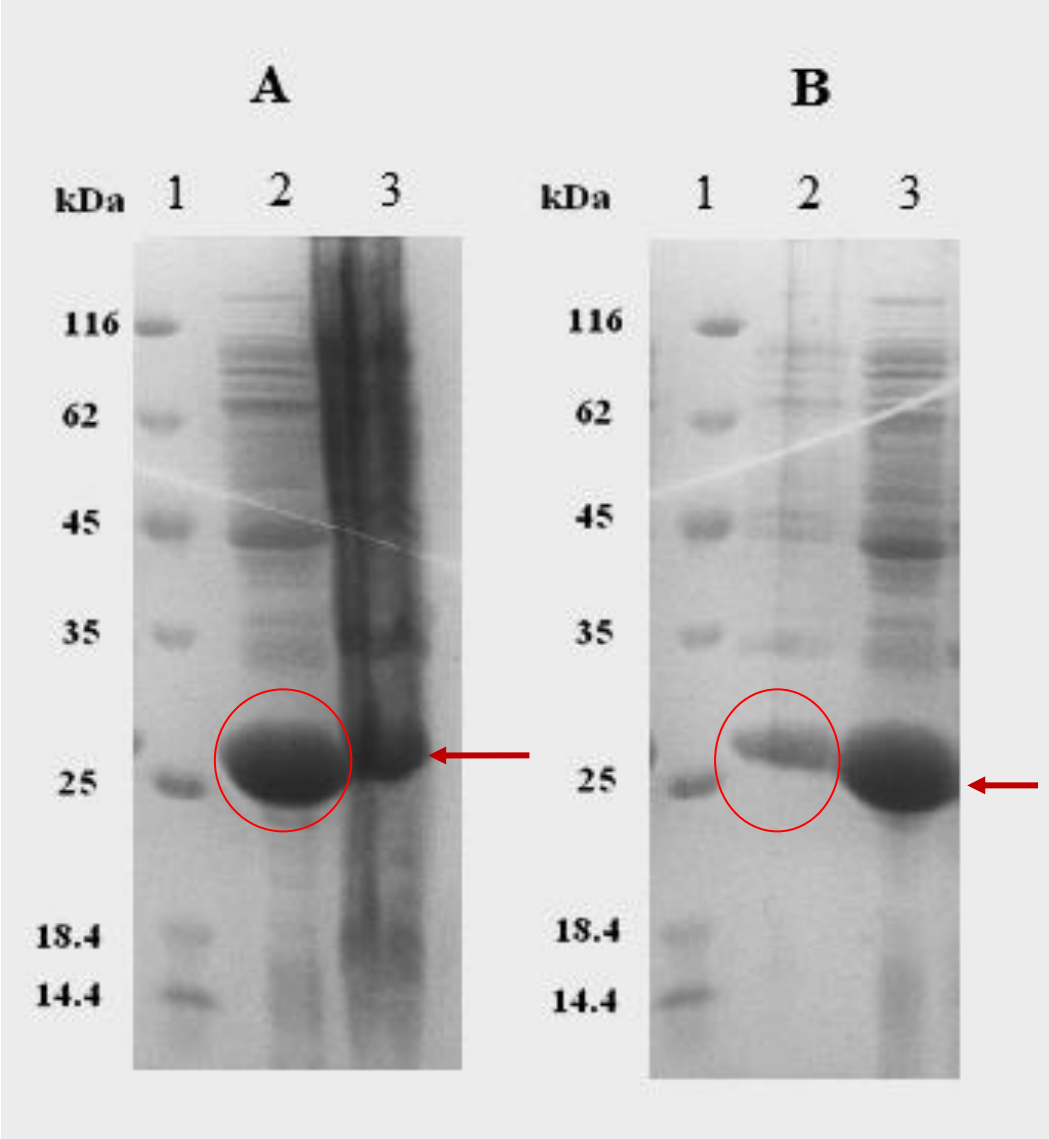


Figure 5. Over-expression studies of Grx2 Y58W/L78A using two different expression systems. 15 % polyacrylamide reducing SDS-PAGE gels showing Grx2 Y58W/L78A expressed in two different systems A) *Escherichia coli* BL21 (DE3)/pLysS cells and B) *Escherichia coli* T7 Express *Iq* cells. Lane 1-3 depicts marker, pellet/insoluble fraction and supernatant/soluble fraction, respectively. The sizes in kDa of the markers are indicated on the left side of the gels. The arrows indicate the over-expressed Grx2 Y58W/L78A protein.

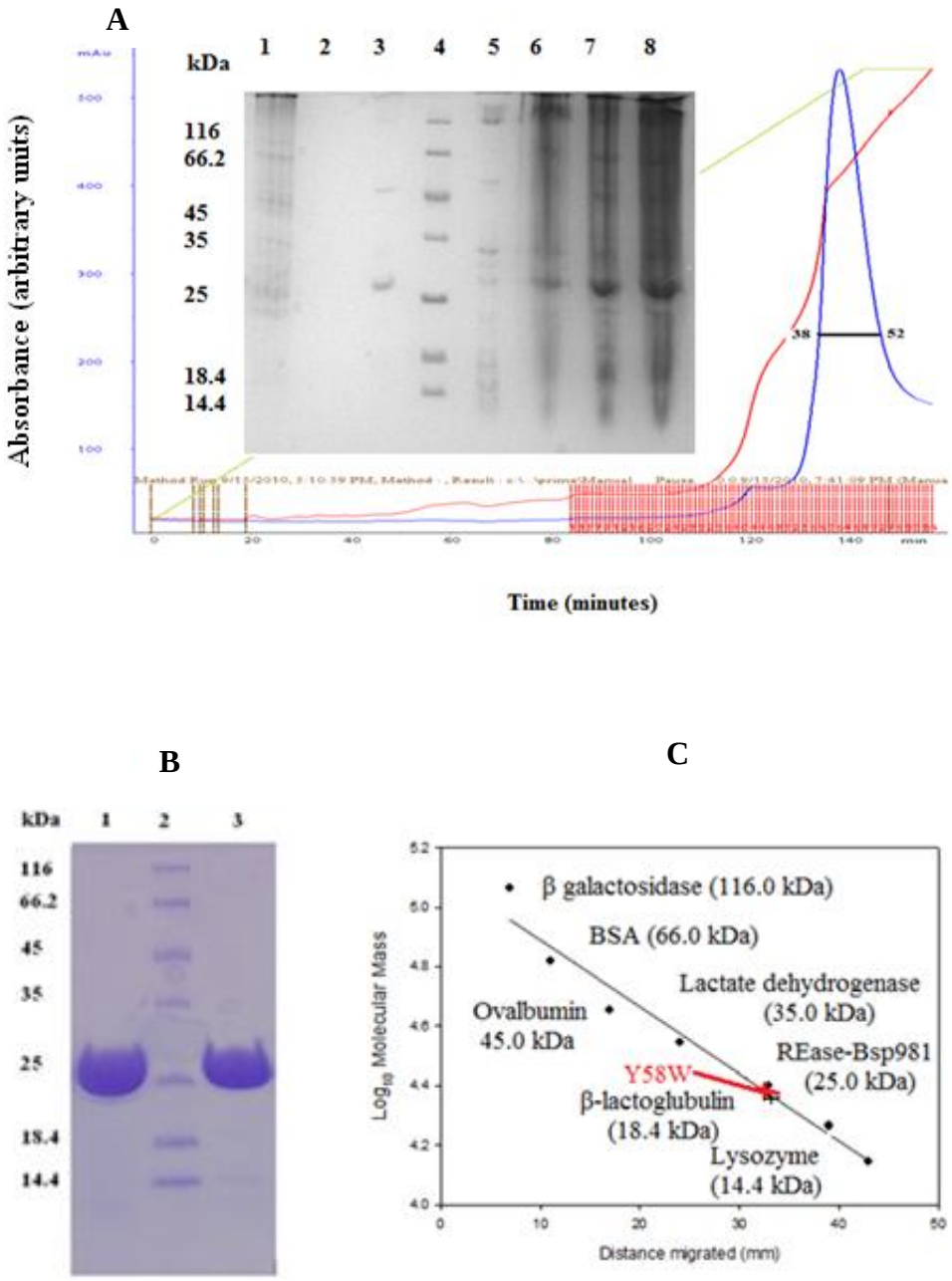


Figure 6. Elution profile (using DEAE ion exchange chromatography), purity and size determination of Grx2 Y58W. The profile was obtained from the purification using DEAE ion exchange chromatography. A) The blue line indicates $A_{280,}$ conductivity shown in red and the pH is represented in gradient green. Different samples were collected during and after purification in order to ensure efficient binding of Grx2 Y58W and assessed for purity on SDS-PAGE as shown by lanes: 1 = regeneration buffer, 2 = pH 9 wash out, 3 pH 10 wash out, 4 = marker, 5 = flow through, 6 = pellet, 7 = supernatant and 8 = whole cell. Fractions 38-52 were pooled together and concentrated and run on SDS-PAGE. B) 15 % polyacrylamide of reducing SDS-PAGE gel depicting the successful purification of Grx2 Y58W as shown by the single band of 25 KDa in lanes 1 and 3. Lane 2 shows protein markers and sizes, which are indicated on the left. C) The calibration curve showing the names and sizes of the markers used to determine the molecular weight of Grx2 Y58W (shown in red).

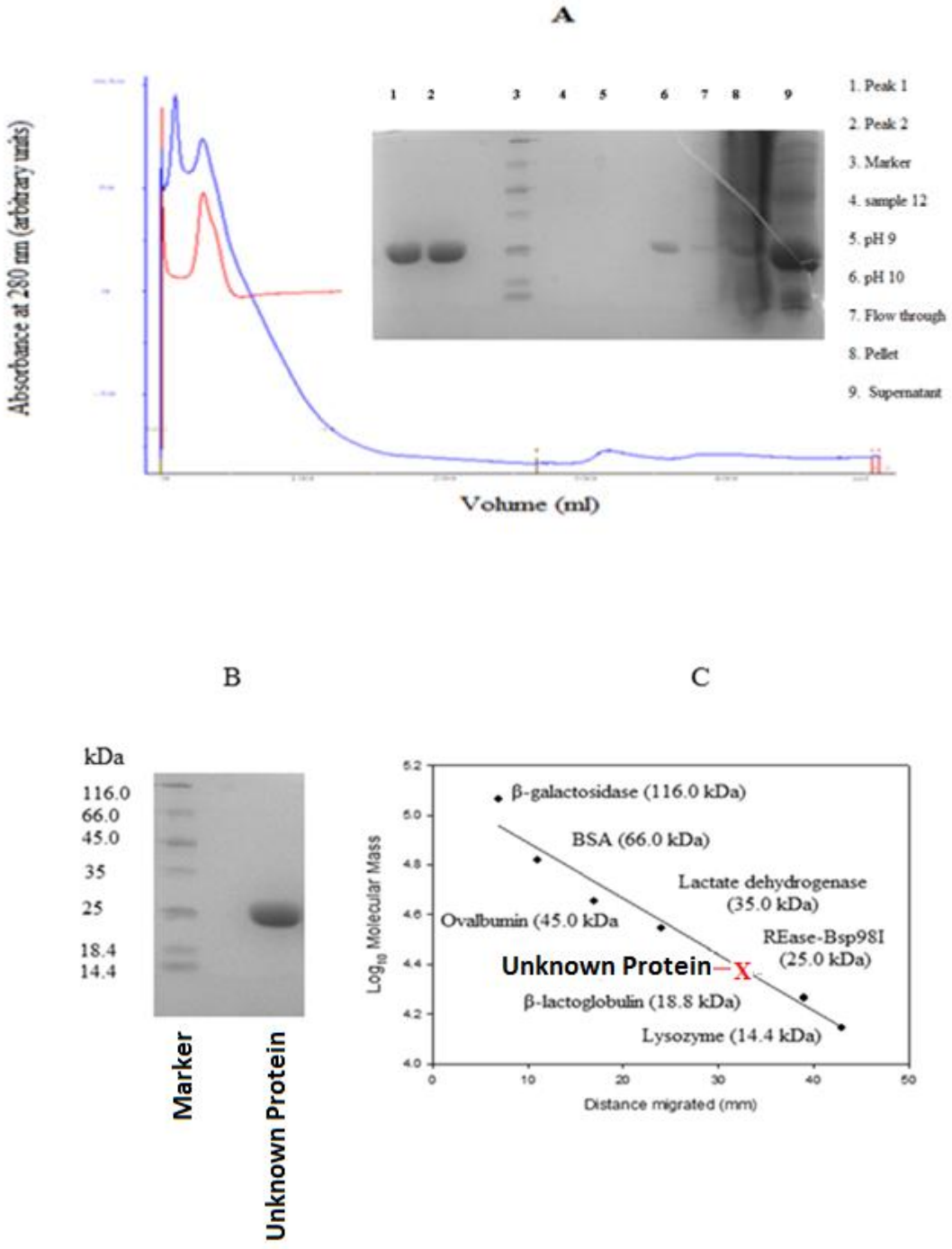


Figure 7. Elution profile (using DEAE ion exchange chromatography) and size determination of Grx2 Y58W/L78A labelled protein X. A) The profile of presumed Grx2 Y58W/L78A which is referred to as unknown protein with SDS PAGE of the different samples collected during purification (the names are indicated on the right). The blue line indicates A_{280} and the red is the conductivity. B) SDS-PAGE analysis of unknown protein (peak 1 and peak 2 pooled together from Figure 7A) which shows a single band indicative of 100% purity. Protein x migrated at a distance which corresponds to 25 kDa. C) The markers lane corresponds to the sizes and names that are shown on the calibration curve. The position of unknown protein is shown on the calibration curve marked with an (X).

3.5 Identification of an unknown protein from a DEAE ion exchange matrix

The purification protocol of Grx2 Y58W/L78A was not comparable to that of wild-type. The statement rises from the fact that Grx2 Y58W/L78A was unable to bind to DEAE column as observed with wild-type (Gildenhuys, 2006; Parbhoo *et al*., 2011). This was attributed to the pI change in Grx2 Y58W/L78A protein. However, the protein was labelled unknown protein until its authenticity had been verified using mass spectroscopy. Unknown protein was identified as recombinant protein Grx2 Y58W/L78A. A total of 18 peptides with > 95% confidence matched to the sequence of native Grx2. Confidence in this study mean that 95 % of the residues matched to the native sequence of Grx2 as shown in Figure 9A. The confidence level is presented according to the different colours as shown in Figure 9A with green = high confidence peptides, yellow = medium confidence peptides; red = low confidence peptides; grey = no peptides detected for this region. Eighteen peptides were matched to the sequence of this protein but only two are shown in (Figure 9B and Figure 9C). These two peptides were selected because they had incorporated the desired mutation for this particular study as shown in Figure (9B) WMPESMDIVHYVDK. From this peptide WMPESMDIVHYVDK; W has been identified with high confidence, however peptide LDGLPLATGK (Figure 9C) which incorporates the L78A mutation has been identified with low confidence. However, the fact that this technique was able to identify that Leu had been mutated to Ala was reassuring and that the overall confidence was > 95 % was more than reliable. From these results, we were able to conclude that the unknown protein is indeed Grx2 Y58W/L78A.

3.6 Characterisation of the Grx2 mutants

3.6.1 Secondary structure characterisation f Grx2 mutants

The secondary structural content of Grx2 Y58W and Grx2 Y58W/L78A were measured using far-UV circular spectra (190-250 nm). This is possible because the amide bond within proteins is a chromophore which is able to absorb circularly polarised light (Woody, 1995). The far-UV CD spectra for Grx2 Y58W and Grx2 Y58W/L78A are shown below (Figure 9). Both proteins have troughs at 208 and 222 nm which are typical for proteins that are predominately α helical (Woody, 1995). This is in agreement with the NMR structure for wild type Grx2 protein which has 56 % α helical content and β sheets in domain 2 (Xia *et al*., 2001). The CD spectrum for Grx2 Y58W/L78A had a significantly higher ellipticity signal

when compared to Grx2 Y58W. The ellipticity signal for both proteins were found to be different because signal for Grx2 Y58W/L78A spectrum increased by 30 % (Figure 9). Since 121 residues make up the eight helices of Grx2, 30 % increase translates into 36 residues, involved in helical structure being folded. These results suggest that the Y58W/L78A mutation has modified the secondary structure of Grx2.

A

MKLYIYDHCPYCLKARMIFGLKNIPVELHVLLNDDAETPTRMVGQKQVPILQKDDSRYMPESMDIVHYVDKLDGKPLLTGKRSPAIEEWLRKVNGYANKLLLPRFAK
SAFDEFSTPAARKYFVDKKEASAGNFADLLAHSDGLIKNISDDLRALDKLIVKPNAVNGELSEDDIQLFPLLRNLTLVAGINWPSRVADYRDNMAKQTQINLLSSMAI

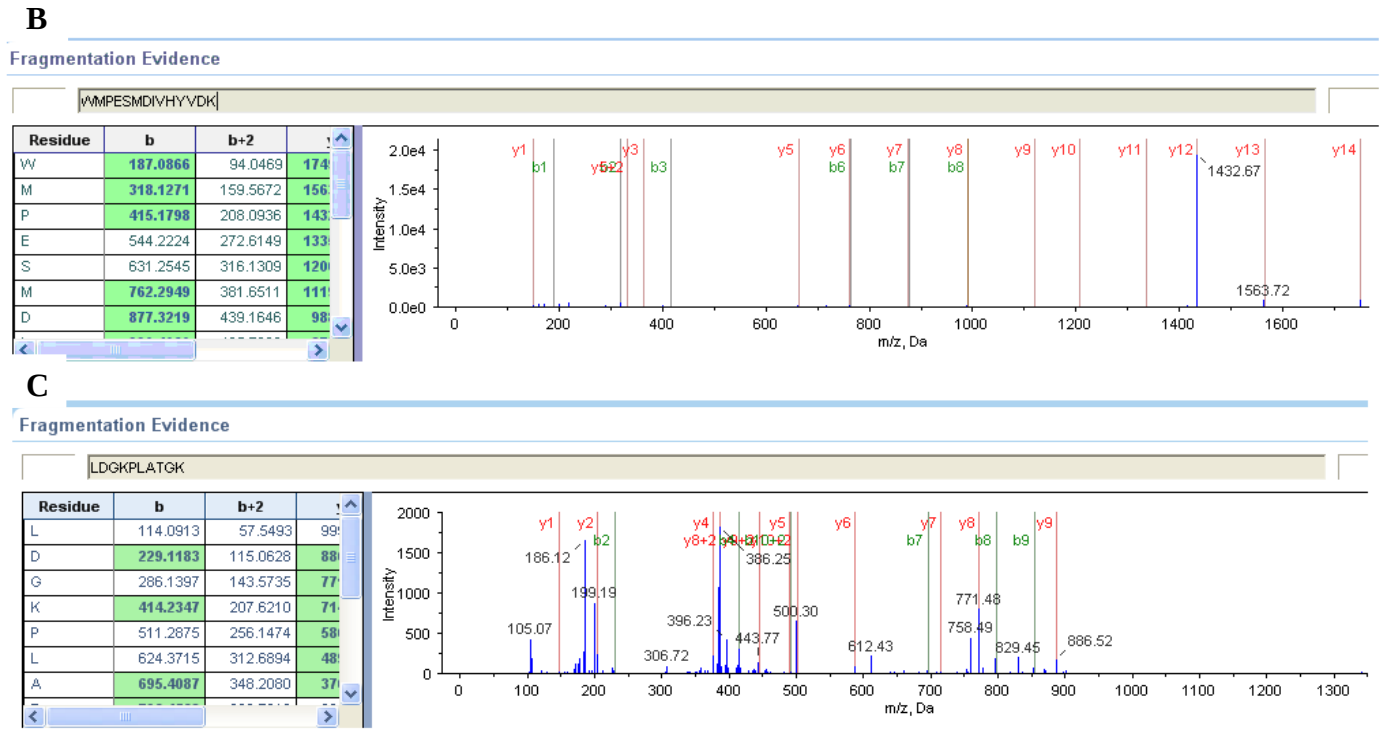


Figure 8. Sequence coverage and fragment ion spectra determined from mass spectroscopy identifying the unknown protein as Grx2 Y58W/L78A. A) The sequence corresponds to Grx2 wild type protein from *Escherichia coli* with different colours indicating the confidence in which the peptides were identified using tandem MS fragmentation (green = high confidence peptides, yellow = medium confidence peptides; red = low confidence peptides; grey = no peptides detected for this region). (B) and (C) Shows the short peptide WMPESMDIVHYVDK and LDGLPLATGK and its corresponding intensity versus m/z mass spectrometry plot. The b, b+2 and y denote the ions formed as a result of the cleavage of a peptide bond. The two short peptides incorporate the mutations tyrosine to tryptophan in (B) and leucine to alanine in (C). The fragmentation results were viewed using ProteinPilotv3.0 (Shilov *et al*., 2007**).**

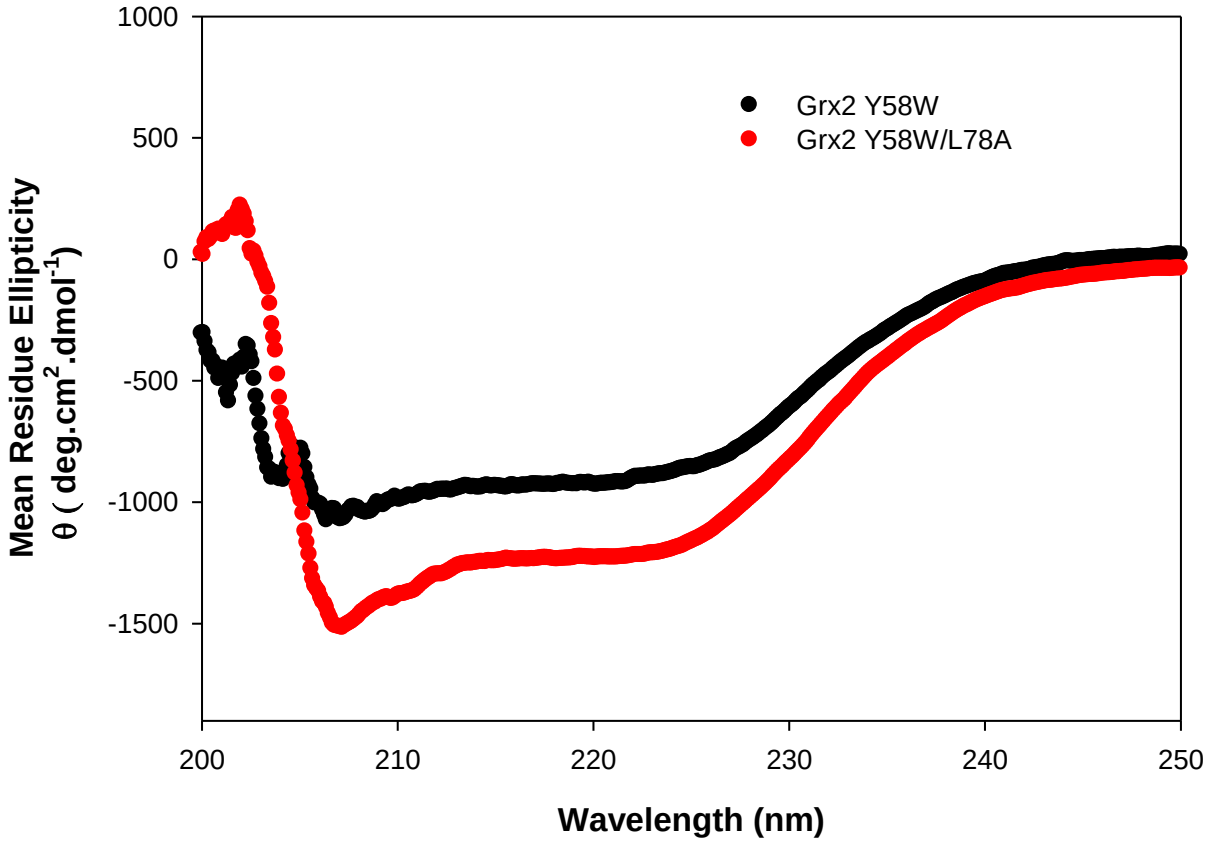


Figure 9. Far-UV CD spectra for Grx2 Y58W and Grx2 Y58W/L78A. Both spectra were obtained using protein concentration of 5 µM in 50 mM phosphate buffer, pH 7.0, 1 mM DTT and 0.02% NaN_3. Grx2 Y58W shown in black (•) and Grx2 Y58W/L78A shown in red (•).

3.6.2 Tertiary structure characterisation

3.6.2.1 Fluorescence Intrinsic spectroscopy of Grx2 mutants

The tertiary structure of Grx2 Y58W and Grx2 Y58W/L78A were analysed using intrinsic fluorescence spectroscopy. This technique measures the local environment of the tryptophan depending on their location and number within the protein (Lakowicks, 1999). It is able to do so, because the indole side chain of the tryptophan residues is sensitive to its microenviroment (Lakowicks, 1999). Wild type Grx2 has two tryptophan residues located in domain 2, Trp-89 and Trp-190 (Xia *et al*., 2001) while Grx2 Y58W and Grx2 Y58W/L78A contain three tryptophan residues because of the additional tryptophan residue in domain 1 (Gildenhuys *et al*., 2008). Both Grx2 Y58W and Grx2 Y58W/L78A exhibit fluorescence emission maxima at 341 nm (Figure 10). This emission maximum is expected for Grx2 protein (Gildenhuys *et al*., 2008; Parbhoo, 2010) because unlike its dimeric GSTs counterpart which have a more buried tryptophan's hence displaying emission maxima at a wavelength of 335 nm (Kaplan *et al*., 1997; Hornby *et a*l., 2000; Hornby *et al*., 2002; Luo *et al*., 2002), Grx2 has an indole ring that is slightly more solvent accessible. The superimposition of the data obtained from the graphs (Figure 10) implies that the mutation within the linker region did not affect the environment of the tryptophan residues in Grx2 protein.

3.6.2.2 Near-UV circular dichroism of Grx2 mutants

Proteins have chromophores (aromatic rings and disulfides group) that absorb light and have characteristics bands in the ultraviolet and visible light (Woody, 1995). For near-UV spectral region (250-350 nm) the signal arises from the environment of each of the aromatic amino acids side chain and disulphide bonds (Myers, 1995). The different bands arise from the contribution of Phe residues in the range 255-270 nm, and between 274-282 nm are contributions from the Tyr and Trp residues. Tryptophan residues contribute the strongest intensity (more predominant) than the other aromatic rings around 288-304 nm. From Figure 11 it can be observed that the Grx2 Y58W and Grx2 Y58W/L78A have a similar fingerprint except for a slight difference in the regions < 260 nm, 288-304

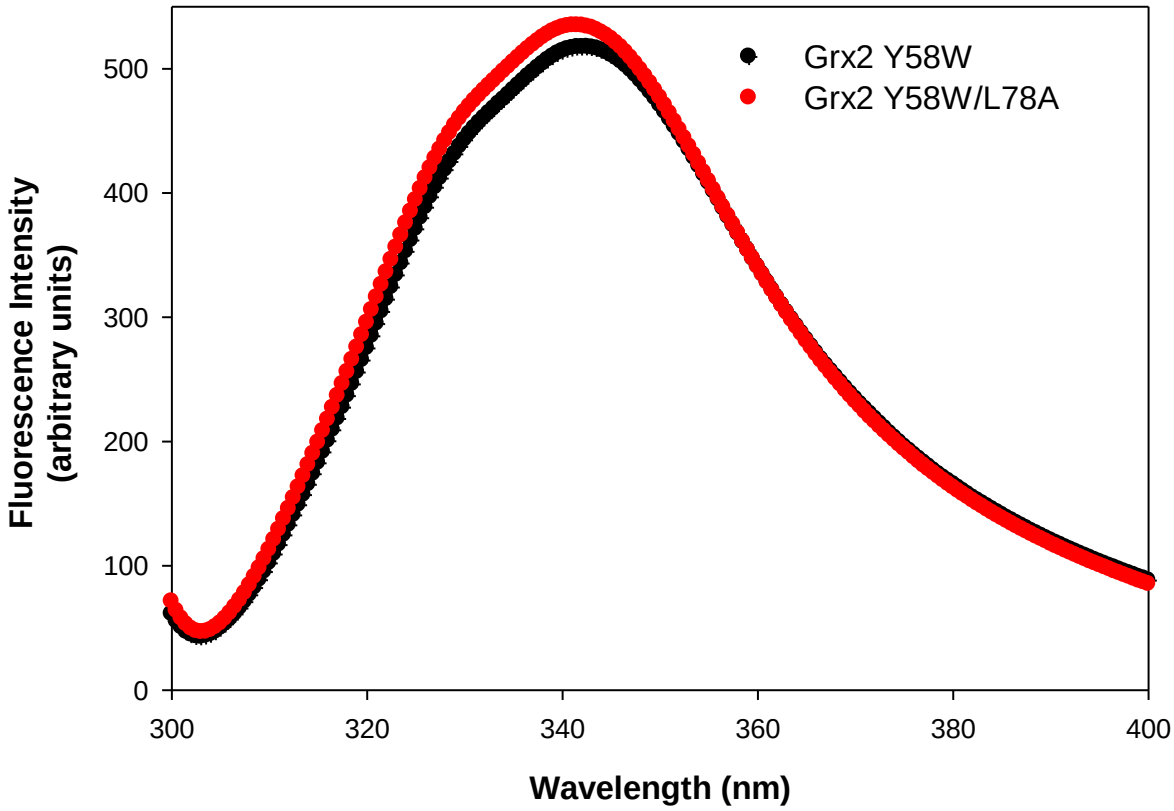


Figure 10. Fluorescence emission spectra for recombinant proteins; Grx2 Y58W and Grx2 Y58W/L78A. Both spectra were obtained using 5 µM of protein in 50 mM phosphate buffer, pH 7.0, 1 mM DTT and 0.02% NaN_3, Grx2 Y58W shown in black (•) and Grx2 Y58W/L78A shown in red (•).

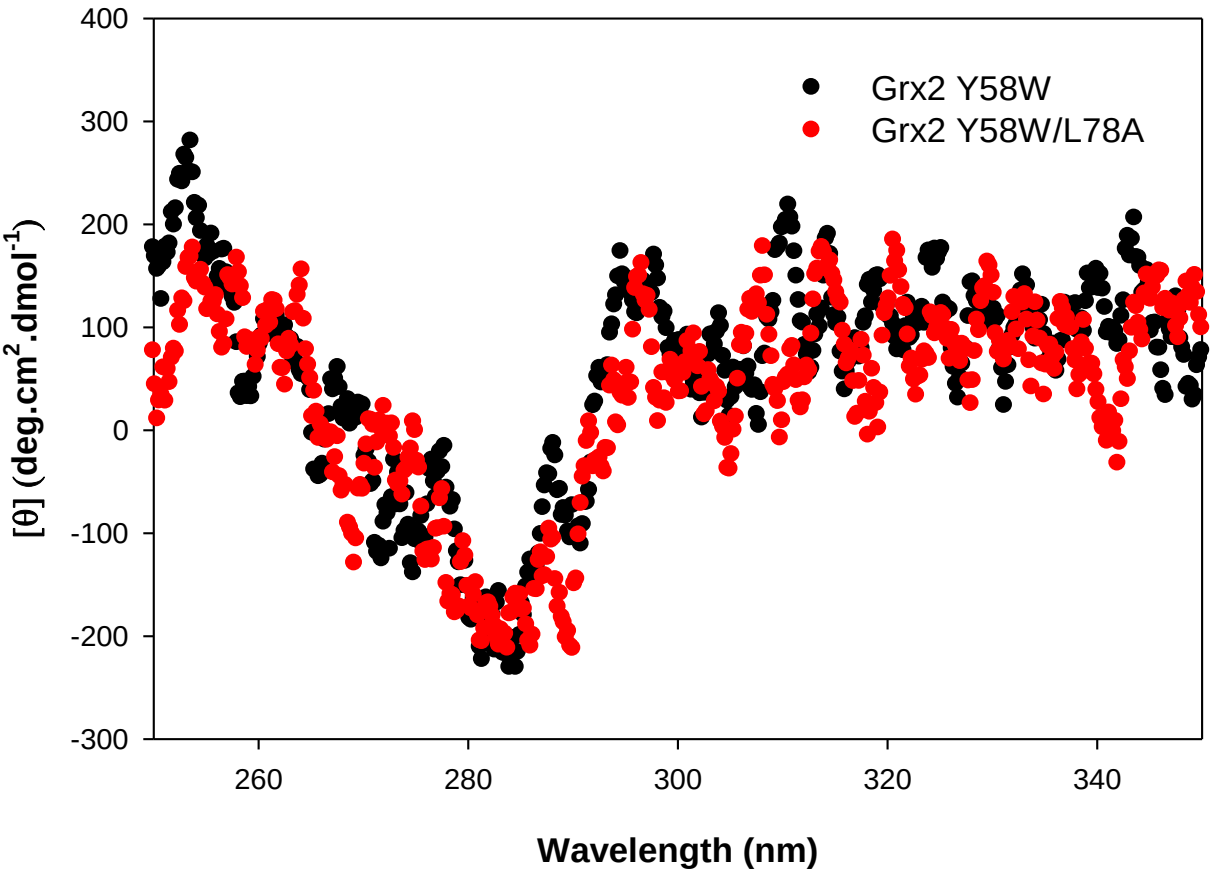


Figure 11. Near-UV CD spectra of Grx2 Y58W and Grx2 Y58W/L78A. Spectra obtained using protein concentration of 20 µM in 50 mM phosphate buffer, pH 7.0, 1 mM DTT and 0.02% NaN_3. Grx2 Y58W shown in black (•) and Grx2 Y58W/L78A shown in red (•).

nm and 310 nm. But overall it can be said that the environment around the chromophore has not changed.

3.7 Conformational stability of Grx2 Y58W/L78A

3.7.1 Reversibility of unfolding

Reversibility is an obligatory prerequisite for the interpretation of equilibrium unfolding studies because equilibrium between the native and the unfolded can only occur if the reaction is reversible and the native fold can be recovered (Pace, 1986). The recovery of Grx2 Y58W/L78A was established using intrinsic tryptophan fluorescence as a probe. Grx2 Y58W/L78A was incubated in the presence of 6 M urea followed by diluting the sample to a lower denaturant concentration (10-fold) where the protein can refold (Pace, 1986). Intrinsic tryptophan fluorescence at 341 nm for Grx2 Y58W/L78A shows 98 % recovery (Figure 12). Previous studies had shown that Grx2 Y58W reaction is reversible and has about 100 % recovery when monitored using both far-UV and fluorescence as probes (Gildenhuys *et al*., 2008).

3.7.2 Urea-induced equilibrium unfolding of Grx2 Y58W/L78A

Urea was used as chemical denaturant to determine the conformational stability of Grx2 Y58W/L78A. This was performed by setting up unfolding reactions from 0 – 6 M urea concentrations. The unfolding reactions were incubated for an hour to reach equilibrium. The urea-induced equilibrium-unfolding reactions were monitored using the following probes; far-UV CD, intrinsic tryptophan fluorescence and extrinsic ANS fluorescence. The probes are used to assess the predominating species present at equilibrium at each urea concentration (Pace, 1986). The ellipticity at 222 nm and the fluorescence emission at 341 nm were plotted as a function of urea concentration and thus used as secondary and tertiary structural probes, respectively (Figure 13). The unfolding curves data for both probes showed a sigmoidal fit (Figure 13) and were fitted to a two-state model (N $\leftrightarrow$U); the parameters obtained are shown in (Table 2). The unfolding curve of Grx2 Y58W/L78A shows that the predominant species for the protein from 0 – 2 M urea concentration is the native state. The transition state begins at 2.4 M urea and ends at 3.8 M urea.

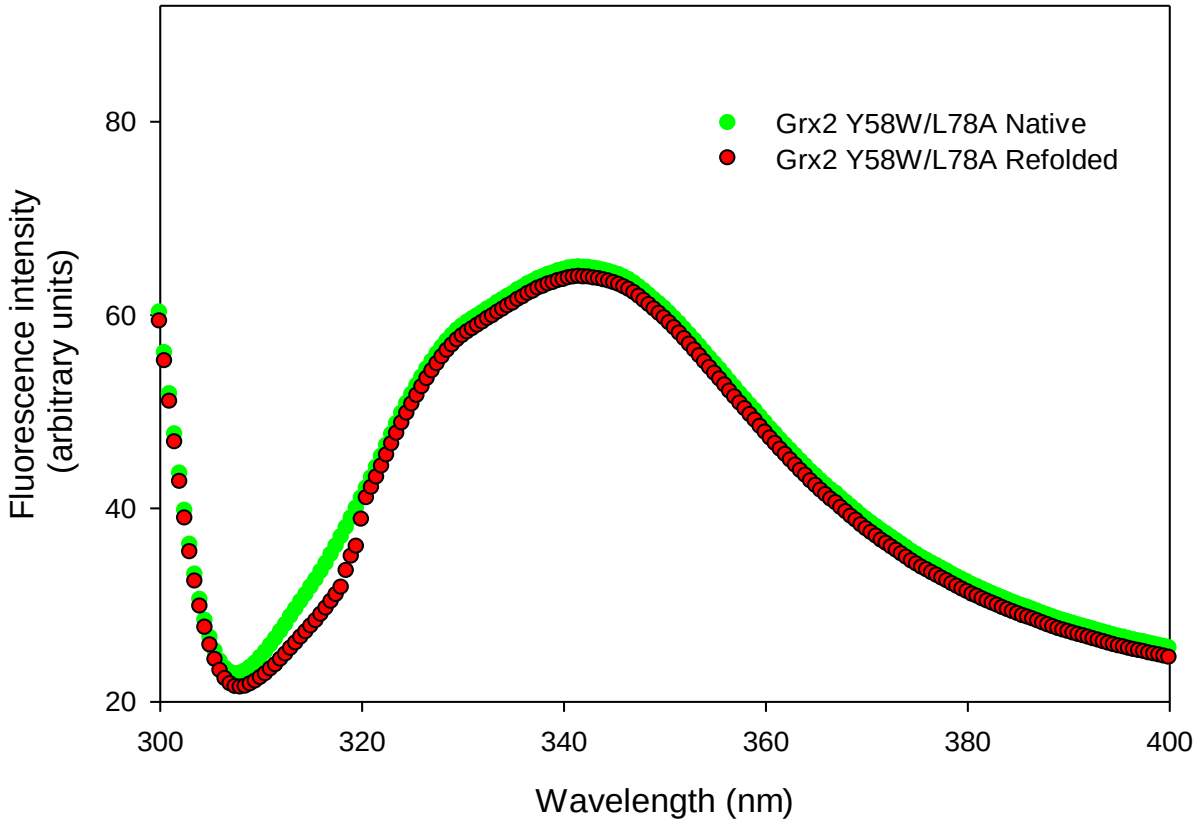


Figure 12. Reversibility of unfolding of Grx2 Y58W/L78A monitored by fluorescence. Fluorescence emission spectra for native Grx2 Y58W/L78A ( ) and refolded Grx2 Y58W/L78A ( ). The concentration for the native recombinant protein was 5 µM in 50 mM phosphate buffer, pH 7.0, 1 mM DTT and 0.02% NaN_3. The residual concentration of urea for the refolded form was 1.0 M.

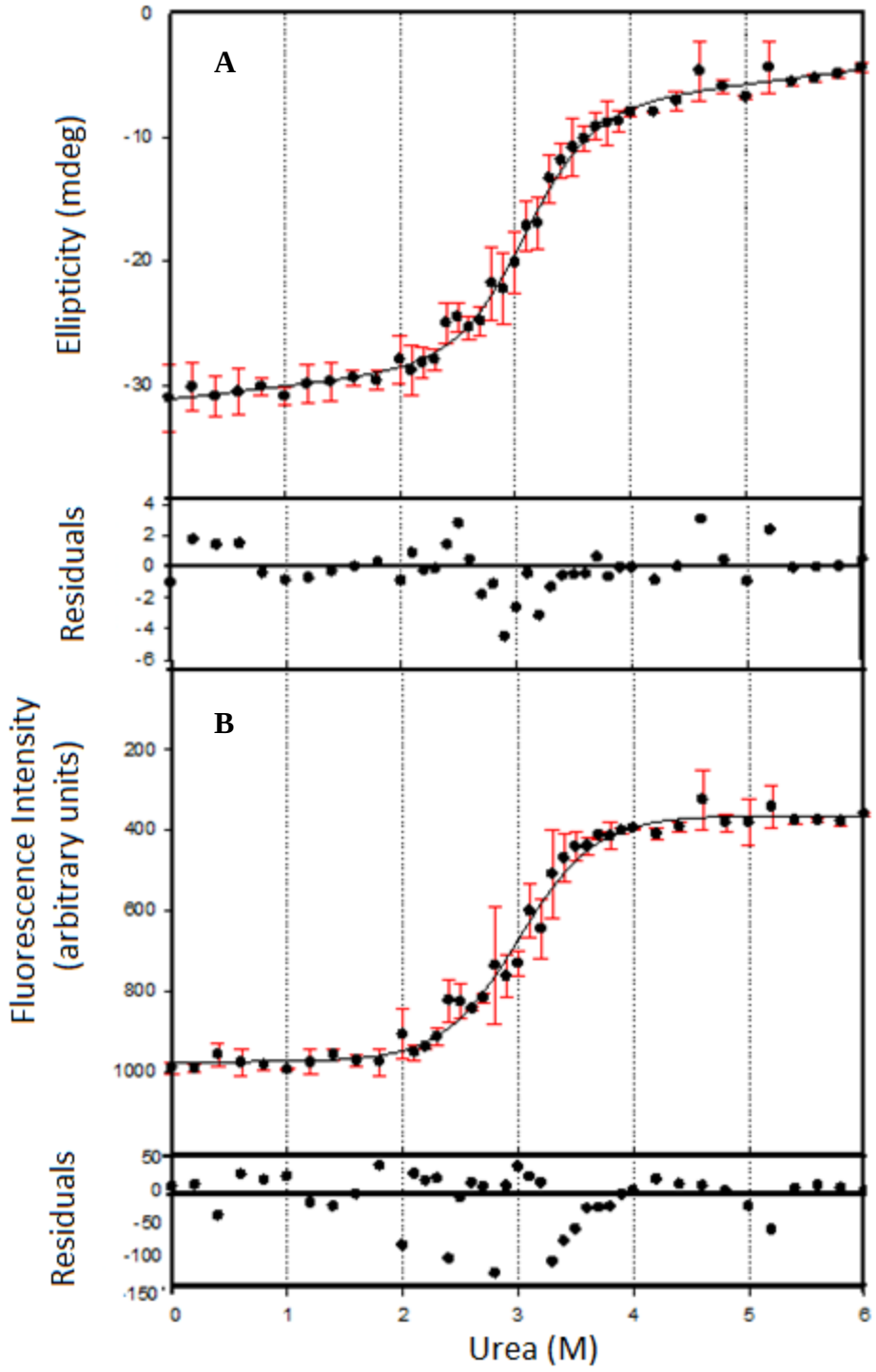


Figure 13. Urea-induced equilibrium unfolding of Grx2 Y58W/L78A. Unfolding curves for Grx2 Y58W/L78A monitored using A) far-UV circular dichroism at 222 nm B) fluorescence intensity at 341 nm. The data were fitted using a two-state model represented by a solid line. Error bars for three replicates are shown. The residuals plots are also shown for each fit and they indicate the goodness of the fit.

Table 2. Parameters obtained for equilibrium unfolding. The equilibrium unfolding of Grx2 Y58W/L78A and Grx2 Y58W was monitored using far-UV circular dichroism and fluorescence. Data obtained from all probes were fitted to a two-state model (N↔U). The data obtained are as result of three triplicates and the data in parentheses represents the standard error. The data shown for Grx2 Y58W was conducted by (Gildenhuys, S., 2006).

Grx2 Y58W/L78A

Parameters	Far-ultraviolet circular dichroism	Intrinsic tryptophan fluorescence
$\Delta G(H_2O)$ (kcal.mol-1)	6.0 (± 1.1)	5.4 (± 0.49)
m (kcal mol^{-1} M^{-1} urea)	1.9 (± 0.38)	1.8 (± 1.1)
C_m (M urea)	3.1	3

Grx2 Y58W

Parameters	Far-ultraviolet circular dichroism	Intrinsic tryptophan fluorescence
$\Delta G(H_2O)$ (kcal.mol-1)	9.0 (± 1.2)	9.9 (± 1.1)
m (kcal mol^{-1} M^{-1} urea)	2.1 (± 0.3)	2.3 (± 1.1)
C_m (M urea)	4.3	4.3

3.7.3 Urea-induced equilibrium unfolding of Grx2 Y58W/L78A in the presence of ANS

8-Anilino-1-naphthalene sulphonate (ANS) is an amphipathic dye that is able to bind to exposed hydrophobic clusters on a protein molecule (Semisotnov *et al*., 1991). Hence it is used as a probe for diagnosing conformational changes induced during protein unfolding and folding reactions (Ptitsyn *et al*., 1990). The effect of urea on the binding of ANS was assessed for Grx2 Y58W/L78A (Figure 14). The linker region is very close to the active site of Grx2 Y58W/L78A, ANS would assist in assessing any changes close around the active site. ANS was found to maximally fluoresce at 500 nm when bound to Grx2 Y58W/L78A protein; the intensities at this wavelength were plotted as a function of urea (Figure 13A). The results indicate no significant binding of Grx2 mutants to ANS dye. The presence of aggregates in unfolding reactions affects their equilibrium unfolding. Therefore light scattering (Figure 13B) is used to detect the presence of aggregates. This is done so by setting the excitation and emission wavelength at 340 nm. The results show that there was negligible aggregate formation.

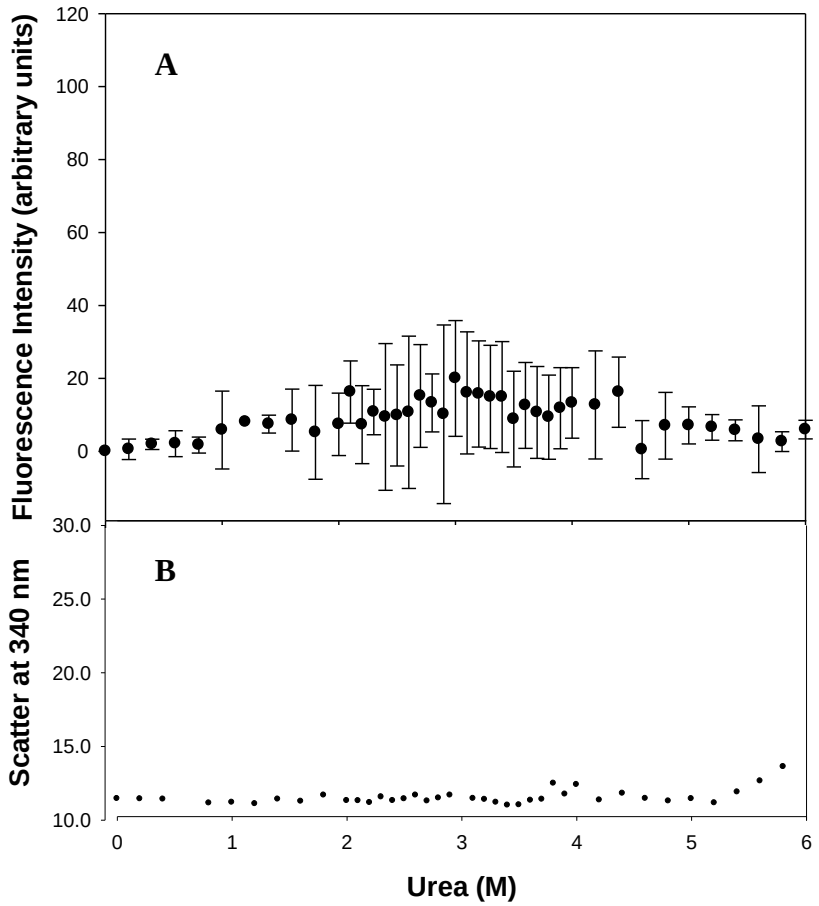


Figure 14. Urea-induced equilibrium unfolding of Grx2 Y58W/L78A in the presence of ANS. A) Fluorescence intensity at 500 nm for Grx2 Y58W/L78A. ANS (200 µM) was added in excess of protein (5 µM). B) Light scattering data used to determine the presence of aggregates by setting the excitation and emission wavelength at 340 nm.

CHAPTER 4

DISCUSSION

Several studies have recognised that in addition to being a physical connector for protein domains, the linker region plays a significant role in domain communication (Gokhale and Khosla, 2000). Therefore, it is vital to understand the exact mechanism of the linker region in order for us to fully comprehend protein folding and function. GSTs are of great interest because in addition to their detoxification role, they have been observed in a range of pathologies such as cancer, cystic fibrosis and neurodegenerative disorders. Since they are dimeric in nature, we have chosen its monomeric homologue Grx2 as a model to investigate the role of the linker region in protein stability. A monomer, Grx2 was chosen in order to investigate the possible roles of domain linker in the absence of any quaternary interactions. Multiple structural-based sequence alignments of the GST and GST-like family depicted that Leu 78 residue is topologically conserved within the linker region (Figure 3A). This was further affirmed by the superimposition of the linker GST on Grx2 (Figure 3B). Conservation of amino acids implies that the residues are vital in protein structure, stability and function. Thus; Leu78 was mutated to an Ala.

4.1 Structural analysis of Y58W/L78A Grx2

Based on the structural data of Grx2 (Xia *et al*., 2001) it is worth noting that the conserved Leu78 interacts with domain 1 through van der Waals forces by curving in and interacting with Ile18 and Leu21 (Figure 15). Asp70 of h3 (helix 3) in the N terminal domain is located within 4 Å of Leu78. Leu78 which is a hydrophobic residue stabilises Grx2 structure by forming van der Waal's interaction with Asp70. The truncation of Leu upon replacing Leu78 with Ala results in the loss of the three rotational bonds (R-group> $CH\text{-}CH_3\text{-}CH_3$), thus making the linker region more flexible. Leu78 also interacts with the neighbouring residues: Pro76, Leu77, Thr79 and Gly80 through van der Waals interactions (Figure 15). Similarly, it interacts with another residue in the C-terminal domain (Glu 170) located in h7. The structural data further affirms the co-operative

nature of van der Waals interactions and how one modification can have an impact in the overall protein.

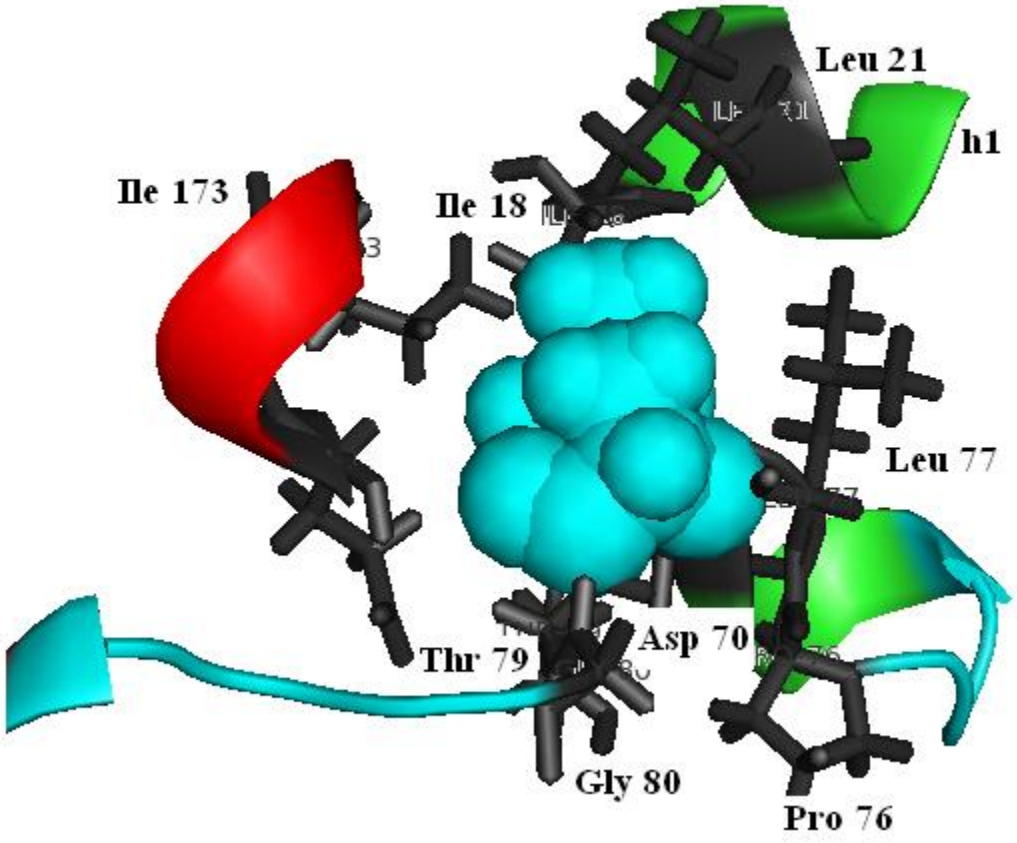


Figure 15. Representation of the residues surrounding Leu78 for Grx2 Y58W/L78A. A) Leu78 (as depicted in spheres coloured cyan) and its surrounding residues are shown as stick representation. B) Leu78 interacts with surrounding residues Pro 76, Leu77, Thr79 and from helix 3 Asp70 (residues are shown in grey).The linker region is coloured cyan and the different helices are shown on the image. All the annotated residues shown are found within 4 Å distance radii of Leu78**.** The image was generated using PyMol (Delano Scientific, San Carlos, CA.)

4.2 Structural integrity of Grx2 Y58W/L78A

Far-UV CD spectrum of Grx2 Y58W/L78A indicates that the removal of Leu 78 seem to affect the secondary structure significantly showing an approximately an increase of (~30 %) in ellipticity (Figure 9), which in turn suggests that Grx2 Y58W/L78A has gained 30 % helical content as compared to Grx2 Y58W. These results may not be a true reflection/misleading because of the following factors:

- The linker has not caused any perturbations in the tertiary structure of Grx2 as depicted by both local probes (Figure 10 and Figure 11).
- The other possible explanation for the increase would be that aromatic residues (provided by the contribution of tryptophan from the mutation) in helical peptides give rise to substantial far-UV induced circular bands (Charkrabartty *et al*., 1992). Therefore, causing significant errors in the estimation of the helix content.

Other studies have also observed an increase in secondary structure due to the mutation within the linker region. One such would be that of Lin *et al*. 2007 who investigated the case of *Rhizopus oryzae* glucoamylase (*Ro*GA). *Ro*GA is an exohydrolase that catalyses the release of β-D-glucose by hydrolyzing α-1,4- and α-1,6-glucosidic linkages at the non-reducing ends of raw or soluble starches and related oligosaccharides (Hiromi *et al*., 1966; Sauers *et al.*, 2000). It is made up of two domains connected by the linker region which is rich in hydroxyl residues. They were able to construct linker mutants of different lengths and they observed an enhanced stability of the secondary structure (Lin *et al*., 2007). Although, several studies also show that mutations within the linker region results in improper positioning between the two domain (N- and C- terminal) which results in less ordered structured species (Feng *et al*., 2007; He *et al*., 2007).

4.3 Role of the linker region of Grx2 Y58W/L78A

Expression studies were the first signs of destabilisation caused by the mutation. This became apparent when the expression of Grx2 L78A/Y58W using *the* same expression system as Grx2 Y58W*,* produced recombinant Grx2 Y58W/L78A protein in inclusion bodies (Figure 5A). Thus, an enhanced expression system, *Escherichia coli* T7 Express I^q

was used in order to obtain Grx2 Y58W/L78A recombinant protein (Figure 5B). Both *Escherichia coli* T7 Express I^q and *Escherichia coli* BL21 (DE3)/pLysS expression systems are under the control of the lac operon. The difference between the two expression systems is that the *Escherichia coli* T7 Express I^q, has the T7 RNA polymerase gene inserted in a different location. This new configuration enabled it to have a more inducible control of transcription genes downstream of the T7 promoter. The mutant (Y58W) was expressed 50 mg per litre whereas Y58W/L78A was 70 mg per litre. Also, the purification protocol for Y58W/L78A protein altered completely from that of wild-type (Figure 7). A new column was tried but it also showed no results, buffer concentration and pH conditions for purification protocol were increased slightly since the pI of Grx2 Y58W/L78A had changed from 7.7 to 8.84 (Gasteiger *et al*., 2005) but the purification could not be optimised to one that yielded protein. However, when SDS-PAGE gel was analysed a pure protein was observed that had come off when washed with pH 10.4 buffer; which had the same size as Grx2 (Figure 7B). It was then send for sequencing to validate its authenticity (Figure 8).

The ΔG (H_2O) indicates the conformational stability of proteins (Pace, 1986). The range of stabilities expected for monomers is between 6 - 14 kcal.mol^{-1} (Neet and Tim, 1994). The ΔG (H_2O) values obtained for Grx2 Y58W/L78A and Grx2 Y58W are 5.7 kcal.mol^{-1}, 9.5 kcal.mol^{-1}, respectively. These values are within the expected range for monomers. However, the $\Delta\Delta G$ (H_2O) for Grx2 Y58W/L78A has decreased ~ 3.8 kcal.mol^{-1}. Previously, Y58W mutation had not altered the stability of Grx2 protein (Gildenhuys, 2006). This was attributed to the fact that the change from a tyrosine to tryptophan residue has a minimal effect on stability, because both residues had similar hydrophobicity values (Wolfenden *et al*., 1979) and average buried volumes (Chonthia, 1975). The destabilisation caused by the mutation could be explained by the following factors:

- We know that the mutation L78A is a cavity forming mutation which can results in the loss of tight packing required for van der Waals interactions (Ratnaparkhi and Varadarajan, 2000).

- Mutational studies have also shown that replacement of a leucine residue for an alanine destabilises a protein structure by 1.19 ± 0.43 per methyl group because of the loss of hydrophobic contacts. The ΔG (H_20) for this study is 5.7 kcal.mol^{-1} (average of both techniques from Table 2) which is almost 2 times greater that the predicted decrease (Kellis, 1988; Pace, 2001).

A comparison of thermodynamic parameters between Grx2 Y58W/L78A (**Table 2**) and Grx2 Y58W (Gildenhuys *et al*., 2008) elucidated a ~0.4 kcal.mol^{-1} M^{-1} decrease in the *m*-values. The *m*-values are an indication of the amount of surface area exposed during unfolding of a protein (Alonso and Dill, 1991; Myers *et al*., 1995). The low *m*-value obtained for Grx2 Y58W/L78A suggests that the protein has a less exposed surface area. The low value obtained may also imply deviation from a two-state equilibrium unfolding mechanism (Myers *et al*., 1995, Pace 1986). But this is not the case in Grx2 Y58W/L78A because of the following factors:

- CD and fluorescence-monitored transitions are coincident; accounting for similar values (Table 2) obtained when using two different two probes. This behaviour is consistent with a two-state unfolding mechanism.
- Rayleigh scatter did not detect any aggregates formed during Grx2 Y58W/L78A equilibrium unfolding (Figure 14B).
- Although they displayed large error bars, ANS binding studies indicated no intermediates present at equilibrium (Figure 14A). This technique has been used quite extensively in studying GSTs (Fanucci *et al*., 2008; Hornby *et al*., 2000; Sayed *et al*., 2000; Stevens *et al*., 1998; Wallace and Dirr, 1999). Although the results indicate no binding, one should also bear in mind the limitations of this technique, which is that the resolution may not be sufficient enough to detect possible intermediates (Soulages, 1998).

Since it is quite apparent that Grx2 Y58W/L78A follows a two-state equilibrium mechanism, it means that the highly co-operative nature that is observed in other GST members (Erhardt and Dirr, 1995; Wallace and Dirr, 1999) and wild-type Grx2 (Gildenhuys *et al*., 2008) is slightly reduced in Grx2 Y58W/L78A (~0.4 kcal.mol^{-1} M^{-1}).

Similarly a reduction in domain co-operativity is observed in M17A Grx2 when the key residue involved in 'key' in the 'lock' domain interfaces is disrupted (Parbhoo, 2010). The reduction in *m*-value suggests that the linker region does to a certain degree play a role in maintaining the co-operativity between the two domains (domain 1 and domain 2). This is not surprising, when one considers that linker domain is located between the two helices (h3 and h4) and both helices enclose helix (h1), which houses the active site (Figure 2). Hence, the hydrophobic contacts between Leu78 and (Ile18 and Leu21) of helix 1 are lost. The C_m values obtained for Grx2 Y58W/L78A (**Table 2**) has shifted from Grx2 Y58W 4.3 M urea → Grx2 Y58W/L78A 3 M urea. The C_m value is the midpoint of the unfolding transition; the results suggest that less urea is required to unfold 50 % of the protein.

4.4 Implication of this study in terms of function and structure of GSTs

This study was conducted in order to study the possible role of the linker region in protein stability. The investigation has shown that in addition to connecting the two domains, linkers in GST and GST-like do play a role in maintaining domain co-operativity. In terms of protein functionality, it would be interesting to further investigate this, since the linker region is not that far from the active site. Other aspects would be to investigate how the linker region affects the rate of folding.

CHAPTER 5

REFERENCES

Aceto, A., Caccuri, A.M., Sacchetta, P., Bucciarelli, T., Dragani, B., Rosato, N., Federici, G., and Di Ilio, C. (1992). *Biochem. J.* ***285***, 241-245.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.

Amann, E., Brosius, J. and Ptashne, M. (1983). Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in *Escherichia coli*. *Gene* **25,** 167-78.

Apic, G., Huber, W. and Teichmann, S. A. (2003). Multi-domain protein families and domain comparison: comparison with known structures and random model of domain recombination. *J. Struct. Funct.Genom*. **4**, 67-68.

Argos, P. (1988). An investigation of protein subunit and domain interfaces. *Protein Eng.* **2**, 101-113.

Armstrong, R.N. Structure, catalytic mechanism, and evolution of the glutathione S-transferase. (1997). *Chem. Res. Toxicol.* **10**, 2-18.

Aslund, F., Ehn, B., Miranda-Vizuette, A., Pueyo, C. and Holmgren, A. (1994). Two additional glutaredoxins exist in *Escherichia coli*: Glutaredoxin 3 is a hydrogen donor for ribonucleotide reductase in a thioredoxin/glutaredoxin1 double mutant. *Proc. Natl. Acad. Sci. USA*. **91**, 9813-9817.

Balchin, D., Fanucchi, S., Achilonu, I., Adamson, R.J., Burke, J., Fernandes, M., Gildenhuys, S. and Dirr, H.W. (2010). Stability of the domain interface contributes towards the catalytic function at the H-site of class alpha glutathione transferase A1-1. *Biochim. Biophys. Acta*. **1804**, 2228-2233.

Batey, S., Randles, L.G., Steward, A., and Clarke, J. (2005). Co-operative folding in a multi-domain protein. *J. Mol. Biol*. **349**, 1045-1059.

Becktel, W.J. and Schellman, J.A. (1987) Protein stability curves. *Biopolymers* **26,** 1859-1877.

Beer, S. M., Taylor, E. R., Brown, S. E., Dahm, C. C., Costa, N.J., Runswick, M. J., and Murphy, M. P. (2004). Glutaredoxin 2 catalyzes the reversible oxidation and glutathionylation of mitochondrial membrane thiol proteins: implications for mitochondrial redox regulation and antioxidant DEFENSE. *J. Biol. Chem.* ***279*****,** 47939-47951.

Bennett, M. J., Schlunegger, M.P. and Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* **4**, 2455-68.

Board, P.G., Coggan, M., Chelvanayagam, G., Easteal, S., Jermiin, L.S., Schulte, G.K., Danley, D.E., Hoth, L.R., Griffor, M.C., Kamath A.V., Rosner, M.H., Chrunyk, B.A., Perregaux, D.E., Gabel, C.A., Geoghegan, K.F. and Pandit, J. (2000). Identification, characterization, and crystal structure of the omega class glutathione transferases. *J. Biol. Chem.* **275**, 24798-24806.

Bousset, L., Belrhali, H., Janin, J., Melki, R. and Morera, S. (2001). Structure of the globular region of the prion protein Ure2 from the yeast *Saccharomyces cerevisiae*. *Structure* **9**, 39-46.

Bradford, M. (1976). "A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding". *Anal. Biochem*. **72,** 248-254.

Briggs, S.D. and Smithgall, T.E. (1999). SH2-Kinase Linker Mutations Release Hck Tyrosine Kinase and Transforming Activities in Rat-2 Fibroblasts. *J. Biol. Chem.* **274**, 26579–26583.

Chakrabartty, A., Kortemme, T., Padmanabhan, S. and Baldwin, R.L. (1993). Aromatic side chain contribution to far-UV circular dichroism of helical peptides and its effect on measurements of helical propenseties. *Biochemistry* **32**, 5560-5565.

Chen, J. and Stites, W.E. (2001). Packing is a key selection factor in the evolution of protein hydrophobic cores. *Biochemistry* **40**, 15280-15289.

Chothia, C. and Janin, J. (1975). Principles of protein-protein recognition. *Nature* **256**, 705-708.

Chung, C.T., Niemela, S.L. and Miller, R.H. (1989). One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA.* **86,** 2172-2175.

Cromer, B.A., Morton, C.J., Board, P.G. and Parker, M.W. (2002). From glutathione transferase to pore in a CLIC. *Eur Biophys J*. **31**, 356-364.

Dauber, P., Hagler, A.T. (1980). Crystal packing, hydrogen bonding, and the effect of crystal forces on molecular conformation. *Accts. Chem. Res*. **13**, 105-112.

Dieckmann, R., Pavela-Vrancic, M., von Dohren, H. and Kleinkauf, H. (1999). Probing the domain structure and ligand-induced conformational changes by limited proteolysis of tyrocidine synthetase. *J. Mol. Biol*. **288,** 129-140.

Dill, K.A. (1990). Dominant forces in protein folding. *Biochemistry* **29**, 7133-55.

Dill, K.A., Alonso, D.O. and Hutchinson, K. (1989). Thermal stabilities of globular proteins. *Biochemistry* **28**, 5439-49.

Dirr, H.W. and Reinemer, P. (1991). Equilibrium unfolding of class pi glutathione S-transferase. *Biochem. Biophys. Res. Commun*. **180**, 294-300.

Dirr, H., Reinemer, P., and Huber, R. (1994). X–ray structures of cytosolic glutathione S–transferases. *Eur. J. Biochem.* **220,** 645-661.

Dirr, H.W. (2001). Folding and assembly of glutathione transferases. *Chem.Biol.Interact*. **133**, 19-23.

Dixon, D.P., Davies, B.G., Edwards, E. (2002). Functional divergence in the glutathione transferase superfamily in plants. *J. Biol. Chem*. **277**, 30859-30869.

Doolittle, R.F. (1995). The multiplicity of domains in proteins. *Annu. Rev. Biochem.* **64**, 287-314.

Domon, B. and Aebersold, R. (2006). Mass spectrometry and protein analysis. *Biochemistry* **312**, 212-217.

Dragani, B., Stenberg, G., Melino, S., Petruzzelli, R., Mannervik, B. and Aceto, A. (1997). The conserved N-capping box in the hydrophobic core of glutathione S-transferase P1-1 is essential for refolding of a buried and conserved hydrogen bond important for protein stability. *J. Biol. Chem.* **272**, 25518-25523.

Duquerroy, S., Cherfils, J., and Janin, J. (1991). Protein-protein interaction: an analysis by computer simulation. *Ciba Found. Symp.* **161**, 237-249.

Englander, S.W., Mayne, L. and Krishna, M.M. (2007). Protein folding and misfolding: mechanism and principles. *Q. Rev. Biophys*. **40**, 287-326.

Erhardt, J. and Dirr, H. (1995). Native dimer stabilizes the subunit tertiary structure of porcine class Pi glutathione S-transferase. *Eur. J. Biochem*. **230**, 614-620.

Fanucchi, S., Adamson, R. J. and Dirr, H. W. (2008). Formation of an unfolding intermediate state of soluble chloride intracellular channel protein CLIC1 at acidic pH. *Biochemistry* **47**, 11674-81.

Feng, S., Zhao, T.-J., Zhou, H.-M., & Yan, Y.-B. (2007). Effects of the single point genetic mutation D54G on muscle creatine kinase activity, structure and stability. *J. Biochem. Cell Biol.* ***39*** **(2),** 392-401.

Fernandes, A.P. and Holmgren, A. (2004). Glutaredoxins: Glutathione-dependent redox enzymes with functions far beyond a simple thioredoxin backup system. *Antioxid. Redox Signal*. **6**, 63-74.

Fritz-Wolf, K., Becker, A., Rahlfs, S., Harwaldt, P., Schirmer, R.H., Kabsch, W., and Becker, K. (2003). X-ray structure of glutathione S transferase from the malarial parasite Plasmodium falciparum. *Proc. Natl. Acad. Sci*. **100**, 13821-13826.

Frova, C. (2006). Glutathione transferases in the genomics era: new insights and perspectives. *Biomol. Eng*. **23**, 149-169.

Geiger, A., Rahman, A. and Stillinger, F.H. (1979). Molecular dynamics study of the hydration of Lennard-Jones solutes. *J. Chem. Phys*. **70,** 263-276

George, R.A. and Heringa, J. (2002). An analysis of protein linkers: their classification and role in protein folding. *Protein Engineering* **15**, 871-879.

Gerstein, M. (1998). How the representatives are the known structures of the proteins in a complete genome? A comprehensive structural census. *Folding & Design.* **3**, 497-512.

Gildenhuys, S. (2006). Folding mechanism of Glutaredoxin-2. PhD Thesis. University of the Witwatersrand.
http://wiredspace.wits.ac.za/handle/10539/4845

Gildenhuys, S., Wallace, L.A., Burke, J.P., Balchin, D., Sayed, Y. and Dirr, H.W. (2010). Class Pi Glutathione Transferase Unfolds via a Dimeric and Not Monomeric Intermediate:Functional Implications for an Unstable Monomer. *Biochemistry* , **49**, 5074–5081.

Gildenhuys, S., Wallace, L.A. and Dirr, H.W. (2008). Stability and Unfolding of Reduced Escherichia coli Glutaredoxin: A Monomeric Structural Homologue of the Glutathione Transferase Family. *Biochemistry* **47**, 10801-10808.

Garel, J., Barry, T.N. and Baldwin, R.L. (1976). Guanidine-unfolded state of ribonuclease A contains both fast- and slow-refolding species. *Proc. Natl. Acad. Sci. USA.* **73***,* 1853-1857.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook (Ed. Walker, J.M.), pp. 571-607. Humana Press, Totowna, NH, USA.

Gokhale R.S., Tsuji, S.Y., Cane, D.E. and Khosla, C. (1999**).** Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482-485.

Gokhale, R.S. and Khosla, C. (2000). Role of linkers in communication between protein modules. *Curr. Opin. Chem. Biol*. **4**, 22-27.

Greene, R.F.Jr., and Pace, C.N. (1973). Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α-chymotrypsin and β-lactoglobulin. *J. Biol. Chem.* **249,** 5388-5393.

Hagler, A, T., Dauber, P., Lifson, S. (1979). Consistent force field studies of intermolecular forces in hydrogen bonded crystals. III. The C=O...H−O hydrogen bond and the analysis of the energetics and packing of carboxylic acids. *J. Am. Chem. Soc*. **101,** 5131-41.

Han, J.H., Batey, S., Nickson, A.A., Teichmann, S.A. and Clarke, J. (2007). The folding and evolution of multi-domain proteins. *Nat. Rev. Mol. Cell Biol*. **8**, 319-330.

Harrop, S.J., DeMaere, M. Z., Fairlie, W. D., Reztsova, T., Valenzuela, S.M., Mazzanti, M., Tonini, R., Qiu, M. R., Jankova, L., Warton, K., Bauskin, A.R., Wu, W.M., Pankhurst, S., Campbell, T.J., Breit, S.N. and Curmi, P.M. (2001). Crystal structure of a soluble form of the intracellular chlorideion channel CLIC1 (NCC27) at 1.4 Å resolution. *J. Biol. Chem.* **276,** 44993-45000.

Hayes, J.D., Flanagan, J.U. and Jowsey, I.R. (2005). Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.* **45**, 51-88.

He, H.-W., Feng, S., Pang, M., Zhou, H.-M. and Yan, Y.-B. (2007). Role of the linker between the N- and C-terminal domains in the stability and folding of rabbit muscle creatine kinase. *J. Biochem. Cell Biol.* **39**, 1818-1827.

Hiromi, K., Hamauzu, Z.I., Takahashi, K. and Ono, S. (1966). Kinetic studies on gluc-amylase. II. Competition between two types of substratehaving alpha-1,4 and alpha-1,6 glucosidic linkage**.** *J Biochem.* **59 (4)**, 411-418.

Holmgren and F. Åslund. (1995). Glutaredoxin. *Methods Enzymol.* **252**, 283-292.

Hornby, J.T., Luo, J.K., Stevens, J.M., Wallace, L.A., Kaplan, W., Armstrong, R.N. and Dirr, H.W. (2000). Equilibrium folding of dimeric class mu glutathione transferases involves a stable monomeric intermediate. *Biochemistry* **39**, 12336-12334.

Ikebe, M., Kambara, T., Stafford, W.F., Sata, M., Katayama, E. and Ikebe, R. (1998). *J. Biol. Chem.* **273**, 17702-17707.

Jaenicke, R. (1991). Protein folding: local structures, domains, subunits and assemblies. *Biochemistry* **30**, 3147-3161.

Jaenicke, R. (1999). Stability and folding of domain proteins. *Prog. Biophys. Mol. Biol.* **71**, 155-241.

Janin, J. and Wodak, S.J., (1983). Structural domains in proteins and their role in the dynamics of protein function. *Prog. Biophys. Mol. Biol*. **42**, 21-78.

Jia, H. and Kaur, P**.** (2001). Role of the Linker Region of the Anion-stimulated ATPase ArsA. *J .Biol. Chem.* **276***,* 29582-29587.

Jeppesen, M.G., Ortiz, P., Shepard, W., Kinzy, T.G., Nyborg, J. and Andersen, G.R. (2003). The crystal structure of the glutathione S-transferase-like domain ofelongation factor 1Bgamma from Saccharomyces cerevisiae. *J. Biol Chem.* **278**, 47190-47198.

Kaplan, W., Husler, P., Klump, H., Erhardt, J., Sluis-Cremer, N. & Dirr, H. (1997). Conformational stability of pGEX-expressed *Schistosoma japonicum* glutathione S-transferase: a detoxification enzyme and fusion protein affinity tag. *Protein Sci.* **6**, 399-406.

Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem*. **14,** 1-63.

Kellis, J. T., Jr., Nyberg, K., Sali, D. and Fersht, A. R. (1988). Contribution of hydrophobic interactions to protein stability. *Nature* **333**, 784-6.

Kelly, S.M., Jess, T.J. and Price, N.C. (2005). How to study proteins by circular dichroism *Biochim. Biophys. Acta.* **1751**, 119-139.

Kumar, S. and Nussinov, R. (1999). Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **293**, 1241-1255.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lakowicz, J.R. (1999). Principles of fluorescence spectroscopy, Protein fluorescence, pp 445-465. Plenum Publishers, USA.

Lesser, G.J. and Rose, G.D. (1990). Hydrophobicity of amino acid subgroups in proteins. *Proteins. Struct. Func. Genet*. **8,** 6-13.

Le Trong, I., Stenkamp, R.E., Ibarra, C., Atkins, W.M. and Adman, E.T. (2002). 1.3-Å resolution structure of human glutathione S-transferase with S-hexyl glutathione bound reveals possible extended ligandin binding site. *Proteins. Struct. Funct. Genet.* **48**, 618-627.

Liang, J. and Dill, K.A. (2001). Are proteins well packed? *Biophys. J.* **81**, 751-766.

Lillig, C. H., Berndt, C., Vergnolle, O., Lonn, M. E., Hudemann, C., Bill, E., and Holmgren, A. (2005). *Proc. Natl. Acad. Sci.U S A.* **102**, 8168-8173

Lin, S.-C., Liu, W.-T., Liu, S.-H.,Chou, W.I., Hsiung, B.-K, Lin, I.-P., Sheu, C.-C. and Chang, M. (2007). Role of the linker region in the expression of *Rhizopus oryzae glucoamalyse*. *BMC Biochemistry,* **8 (9)**, 1471-1483.

Macdonald, R.I. and Cummings, J.A. (2004). Stabilities of folding of clustered; two-repeat fragments of spectrin reveal a potential hinge in the human erythroid spectrin tetramer. *Proc. Natl. Acad. Sci. USA.* **101**, 1502-1507.

Mannervic, B. and Danielson, U.H. (1988). Glutathione-S-transferase structure and catalytic activity. *Crit. Rev. Biochem*. **23**, 283-337.

Mirsky, A.E. and Pauling, L. (1936). On the structure of native, denatured, and coagulated proteins. *Proc. Natl. Acad. Sci. USA.* **22,** 439-447.

Morris, M.J., Craig, S.J., Sutherland, T.M., Board, P.G., Casarotto, M.G. (2009). Transport of glutathione transferase-fold structured proteins into living cells. *Biochim. Biophys. Acta.* **1788***,* 676-685.

Murzin A.G., Brenner S.E., Hubbard, T., Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536-540.
Myers, J.K., Pace, C.N. and Scholtz, J.M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138-48.

Neet, K.E. and Timm, D.E. (1994). Conformational stability of dimeric proteins: Quantitative studies by equilibrium denaturation. *Protein Sci.* **3**, 2167-2174.

Nelson, M. and McClelland, M. (1992). Use of DNA methyltransferase/endonuclease enzyme combinations for megabase mapping of chromosomes. *Methods Enzymol***. 216**, 279-303.

Pace, C.N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266-280.

Pace, C.N., Shirley, B.A., McNutt, M. and Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *Faseb J.* **10**, 75-83.

Pace, C.N. and Scholtz, J.M. (1997) Measuring the conformational stability of a protein. In Protein structure, a practical approach (2nd ed.) (Ed. Creighton, T.E.), pp. 299. Oxford University Press, UK.

Pace, C.N. (2001). Polar group burial contributes more to protein stability than nonpolar group burial. *Biochemistry* **40**, 310-313.

Packman, L.C. and Perham, R.N. (1987). *Biochem. J*. **242**, 531-538.

Papworth, C., Bauer, J.C., Braman, J. and Wright, D.A. (1996). Site-directed mutagenesis using double-stranded plasmid DNA templates. *Strategies* **9**, 3-4.

Parbhoo, N. (2010). The role of the conserved interdomain interaction in *Escherichia-coli* glutaredoxin-2. MSc dissertation. University of the Witwatersrand.
http://wiredspace.wits.ac.za/handle/10539/8557

Parbhoo, N., Stoychev, S.H., Fanucchi, S., Achilonu, I., Adamson, R.J., Fernandes, M., Gildenhuys, and Dirr, H.W. (2011). A conserved interdomain interaction is a determinant of folding cooperativity in the GSTfold. *Biochemistry* **50,** 7067-7075**.**

Perkins, S.J. (1986). Protein volumes and hydration effects. *Eur. J. Biochem*. **157**, 169-180.

Perrett, S., Freeman, S.J., Jonathan, P., Butler, G. and Fersht, A.R. (1999). Equilibrium folding properties of the yeast prion protein determination Ure2. *J. Mol. Biol.* **290**, 331-345.

Privalov, P.L. (1979). Stability of proteins: small globular proteins. *Adv Protein Chem.* **33**, 167-241.

Privalov, P.L. and Makhatadze, G.I. (1993). Contribution of hydration to protein folding thermo-dynamics. II. The entropy and Gibbs energy of hydration. *J. Mol. Biol.* **232**, 660-679.

Ptitsyn, O.B., Pain, R.H., Semisotnov, G.V., Zerovnik, E. and Razgulyaev, O.I. (1990). Evidence for a molten globule state as a general intermediate in protein folding. *Biochemistry* **262**, 20-21.

Radford, S.E., Laue, E.D., Perham, R.N., Martin, S.R. and Appella, E. (1989). Conformational flexibility and folding of synthetic peptides representing an interdomain segment of polypeptide chain in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli J. Biol. Chem*. **264**, 767-775.

Ratnaparkhi, S.G., and Varadrajan, R. (2000). Thermodynamics and structural studies of cavity formation in proteins suggests that loss of packing interactions rather than the hydrophobic effect dominates the observed energetics. *Biochemistry* **39**, 12365-12374.

Ramachandran, G.N. and Sasisekharan, V. (1968). Conformation of a polypeptides and proteins. *Adv. Protein Chem.* **23,** 283-437.

Ranson, H., Collin, F., Hemingway, J. (1998). The role of alternative mRNAsplicing in generating heterogeneity within the Anopheles gambiae class Iglutathione S-transferase family. *Proc. Natl. Acad. Sci. USA*. **95**, 14284-14289.

Rappsilber, J., Ishihama, Y. and Mann, M. (2003). Stop and go extraction tips for matrix-assisted desorption/ionization, nanoelectrospray and LC/MS sample pre-treatment in proteomics. *Anal. Chem.* **75**, 663 – 670.

Robinson, C.J. and Sauer, R.T. (1998).Optimising the stability of single-chain proteins by the linker length and composition mutagenesis*. Acad. Sci. USA.* **95**, 5929-5934.

Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H. and Zehfus, M.H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science* **229**, 834-838.

Rowe, E.S. and Tanford, C. (1973). Equilibrium and kinetics of the denaturation of a homogeneous human immunoglobulin light chain. *Biochemistry* **12**, 4822-4827.

Sauer, J., Sigurskjold, B.W., Christensen, U., Frandsen, T.P., Mirgorodskaya, E., Harrison, M., Roepstorff, P. and Svensson, B. (2000). Glucoamylase: structure function relationships, and protein engineering. *Biochim*. *Biophys. Acta*, **1543(2),** 275-293.

Sayed, Y., Wallace, L.A. and Dirr, H.W. (2000). The hydrophobic lock-and-keyintersubunit motif of glutathione transferase A1-1: implications for catalysis, ligandin function and stability. *FEBS Lett.* **465**, 169-172.

Schevchenko, A., Tomas, H., Havliš, J., Olsen, V. J. and Mann, M. (2007). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols* **1**, 2856 - 2860.

Semisotnov, G.V., Rodionova, N.A., Razgulyaev, O.I., Uversky., V.N., Gripas' A.F. and Gilmanshin, R.I. (1991). Study of the 'molten globule' intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31,** 119-128.

Sheehan, D., Meade, G., Foley, V.M. and Dowd, C.A. (2001). Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *J. Biochem.* **360**, 1-16.

Shi, J., Vlamis-Gardikas, A., Áslund, F., Holmgren, A. and Rosen, B. P. (1999). Reactivity of glutaredoxins 1, 2, and 3 from *Escherichia coli* shows that glutaredoxin 2 is the primary hydrogen donor to ArsC-catalyzed arsenate reduction. *J. Biol. Chem.* **274**, 36039-36042.

Shilov, V.I., Seymour, L.S., Patel, A.A., Loboda, A., Tang, H.W., Keating, O.S., Hunter, L.C., Nuwaysir, M.L. and Schaeffer, A.D. (2007). The paragon algorithm, a next generation search engine that uses sequence temperature values and features probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Protem.* **69**, 1638 – 1655.

Shirley, B.A. (1995). Urea and guanidine hydrochloride denaturation curves in Methods in *Molecular Biology: Protein Stability and Folding: Theory and Practice* (Shirley, B.A., Ed.). **40**. pp 177-190. Humana Press Inc. Totowa, New Jersey.

Sinning, I., Kleywegt, G.J., Cowan, S.W., Reinemer, P., Dirr, H.W., Huber, R., Gilliland, G.I., Armstrong, R.N., Ji, X., Board, P.G., Olin, B., Mannevrik, B., and Jones, T.A. (1993). Structure Determination and Refinement of Human Alpha Class Glutathione Transferase A1-1, and a comparison with the Mu and Pi Class Enzymes. *J. Mol. Biol.* **232,** 192-212.

Soranzo, N., Gorla, S., Mizzi, L., De Toma, G., Frova, C. (2004). Organisation and structural evolution of the rice glutathione S-transferase gene family. *Mol. Gen. Genomics*. **271**, 511–521.

Soulages, L.J. (1998). Chemical denaturation: potential impact of undetected intermediates in the free energy of unfolding and *m*-values obtained from a two-state assumption. *Biophys. J.* **75**, 484-492.

Stevens, J.M., Hornby, J.A., Armstrong, R.N. and Dirr, H.W. (1998). Class sigma glutathione transferase unfolds via a dimeric and a monomeric intermediate. Impact of subunit interface on conformational stability in the superfamily. *Biochemistry* **37**, 15534-15541.

Stickle, D.F., Presta, L.G., Dill, K.A., and Rose, G.D. (1992). Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226,** 1143-1159.

Stillinger, F.H. (1980). Water Revisited. *Science* **209**, 451-457.

Stoychev, S. H., Nathaniel, C., Fanucchi, S., Brock, M., Li, S., Asmus, K., Virgil, L. W. Jr., and Dirr, H. W. (2009). Structural Dynamics of Soluble Chloride Intracellular Channel Protein CLIC1 Examined by Amide Hydrogen-Deuterium Exchange Mass Spectrometry. *Biochemistry.* **48,** 8413-8421.

Stoychev, S. (2008). The role of the domain interface in the stability, folding and function of Clic1. PhD thesis, University of the Witwatersrand.
http://wiredspace.wits.ac.za/handle/10539/5604

Teale, F.W. (1960). The ultraviolet fluorescence of proteins in neutral solution. *J. Biochem.* **76,** 381-388.

Teichmann, S.A., Park, J. and Chothia, C. (1998). Structural assignments to the *Mycoplasma genitalium* proteins show extensive gene duplication and domain rearrangement. *Proc. Natl. Acad. Sci. USA.* **95**, 14658-14663.

Tew,K.D. (1994) Glutathione-associated enzymes in anticancer drug resistance. *Cancer Res*. **54**, 4313–4320.

Tew, K. D. (2007) Redox in redux: Emergent roles for glutathione S-transferase P (GSTP) in regulation of cell signaling and S-glutathionylation. *Biochem. Pharmacol*. **73**, 1257–1269.

Thompson, L.C., Walters, J., Burke, J., Parsons, J.F., Armstrong, R.N. and Dirr, H.W. (2006). Double-mutation at the subunit interface of glutathione transferase rGSTM1-1 results in a stable, folded monomer. *Biochemistry* **45,** 2267-2273.

van Leeuwen,H.C., Strating, M.J., Rensen, M., deLaat, W. and van der Vliet , P.C.(1997). *EMBO J.* **16**, 2043–2053.

Vlamis-Gardikas A and Holmgren A. (2002). Thioredoxin and glutaredoxin isoforms. *Methods Enzymol.* **347**, 286–296.

Vlamis-Gardikas, A., Aslund, F., Spyrou, G., Bergman, T. and Holmgren, A. (1997) Cloning, overexpression, and characterization of glutaredoxin 2, an atypical glutaredoxin from *Escherichia coli*. *J. Biol. Chem.* **272**, 11236-11243.

Vogel, C., Bashton, M., Kerrison, C.C., Chothia, C. and Teichmann, S.A. (2004). Structure, function and evolution of multi-domain proteins. *Curr. Op. Struc. Biol.* **14**, 208-216.

Vuilleumier, S. and Pagni, M. (2002). The elusive roles of bacterial glutathione S transferases: new lessons from genomes. *Appl. Microbiol. Biotechnol*. **58**, 138–146.

Wallace, L.A., Sluis-Cremer, N. and Dirr, H.W. (1998). Equilibrium and kinetic unfolding properties of dimeric human glutathione transferase A1-1. *Biochemistry* **37**, 5320-5328.

Wallace, L.A. and Dirr, H.W. (1999). Folding and assembly of dimeric human glutathione transferase A1-1. *Biochemistry* **38**, 16686-16694.

Wallace, L.A., Burke, J. and Dirr, H.W. (2000). Domain-domain interface packing at conserved Trp-20 in class alpha glutathione transferase impacts on protein stability. *Biochim. Biophys. Acta* **1478**, 325-32.

Wang, Y., Geer, L.Y., Chappey, C., Kans, J.A. and Bryant, S.H. (2000). Cn3D: sequence and structure views for Entrez. *Trends Biochem. Sci.* **25**, 300-302.

Weber, G. and Young, L.B. (1964) Fragmentation of bovine serum albumin by pepsin. I. The origin of the acid expansion of the albumin molecule. *J. Biol. Chem.* **239,** 1415-1423.

Wetlaufer, D.B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA* **70,** 697-701.

Wilce, M.C., and Parker, M.W. (1994). Structure and function of glutathione S-transferases. *Biochim. Biophys. Acta.* **1205**, 1-18.

Wilson, I.A and Stanfield, R.L. (1994). Antibody-antigen interactions: new structures and new conformational changes, *Curr. Opin. Struct. Biol.* **4**, 857-867.

Wolfenden, R.V., Cullis, P.M. and Southgate, C.F. (1979). Water, protein folding and the genetic code. *Science* **206**, 575-577.

Woody, R. W. (1995). Circular Dichroism. *Methods Enzymol.* **246**, 34-71.

Wrigglers, W., Chakravarty, S. and Jennings, P.A. (2005). Control of Protein Functional Dynamics by Peptide Linkers. *Biopolymers*, **80**, 736-746.

Xia, B., Chung, J., Vlamis-Gardikas, A., Holmgren, A., Wright, P. E. and Dyson, H. J. (1999). Assignment of 1H, 13C, and 15N resonances of reduced Escherichia coli glutaredoxin 2. *J Bio. MolNMR*. **14**, 197-8.

Xia, B., Vlamis-Gardikas, A., Holmgren, A., Wright, P.E. and Dyson, H.J. (2001). Solution structure of *Escherichia coli* glutaredoxin-2 shows similarity to mammalian glutathione-S-transferases. *J. Mol. Biol.* **310**, 907-918.

Zhu, Z.Y. and Karlin, S. (1996). Clusters of charged residues in protein three dimensional structures. *Proc. Natl. Acad. Sci. USA.* **93**, 8350-8355.