

https://dspace.org/

The N-subdomain of the thioredoxin fold of glutathione transferase is stabilised by topologically conserved leucine residue

Khoza, Thandeka Ntokozo

2013-04-30
http://hdl.handle.net/10539/12700

Downloaded from a DSpace repository. https://dspace.org/

THE N-SUBDOMAIN OF THE THIOREDOXIN FOLD OF GLUTATHIONE TRANSFERASE IS STABILISED BY TOPOLOGICALLY CONSERVED LEUCINE RESIDUES

Thandeka Ntokozo Khoza

A thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of Doctor of Philosophy.

Johannesburg, 2012

Declaration

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



8 November 2012

Abstract

The thioredoxin-like (Trx-like) fold is preserved in various protein families with diverse functions despite their low sequence identity. Glutathione transferases (GSTs) are characterised by a conserved N-terminal domain with a thioredoxin–like βαβαββα secondary structure topology and an all alpha-helical domain. GSTs are the principal phase II enzymes involved in protecting cellular macromolecules from a wide variety of reactive electrophilic compounds. It catalyses the conjugation of reduced glutathione (GSH) to an electrophilic substrate to form a hydrophilic and non-toxic compound. The binding site for GSH (G-site) is located in the N-terminal domain of GSTs. The sequence identity within members of the Trx-like superfamily is low; however, the members of this family fold into a conserved βαβαββα topology. It, therefore, seems reasonable that there are topologically conserved residues within this fold whose main role is to drive folding and/or maintain the structural integrity of the Trx-like fold. Structural alignments of the N-subdomain (βαβ motif) of the GST family shows that Leu7 in β1 and Leu23 in α1 are topologically conserved residues. The Leu7 side chain is involved in the packing of α1β1α2 and α3, whilst Leu23 is mainly involved in van der Waals interactions with residues in α1 and the loop region connecting α1 and β2. Taking into account the types of interaction that both Leu7 and Leu23 are involved in, as well their location in close proximity to the G-site, it was postulated that both these residues may play a role in the structure, function and stability of the GST family of proteins. Leu7 and Leu23 are not directly involved in the binding of GSH but they could be important in maintaining the G-site in a functional conformation via correct packing of the N-subdomain.

The homodimeric human class Alpha of GST (hGSTA1-1) was used as the representative of the GST family to test this hypothesis. The bulky side chains of Leu7 and Leu23 were replaced with a less bulky alanine residue to prevent altering the hydrophobicity of the βαβ motif. The effect of the mutation on the structure, function and stability of hGSTA1-1 was, therefore, studied in comparison with the wild-type using spectroscopic tools, X-ray crystallography, functional assays and conformational stability studies.

The impact of the mutations on the structure of the enzyme was determined using spectroscopic tools and X-ray crystallography. The X-ray structures of the L7A and L23A mutants were resolved at 1.79 Å and 2.2 Å, respectively. Analysis of both X-ray structures

shows that the mutation did not significantly perturb the global structure of the protein, which correlates with far-UV CD and intrinsic fluorescence spectroscopic data. In addition, structural alignments using the C-alpha gave root mean square deviation (r.m.s.d) values of 0.63 Å (L7A) and 0.67 Å (L23A) between the wild-type and mutant structures. However, both the L7A and L23A structures showed the presence of a cavity within the local environment of each mutation. The functional properties of the mutants were also similar to those of the wild-type as determined by specific activity and 8-anilino-1-naphthalene sulfonate (ANS)-binding, indicating that Leu7 and Leu23 are not involved in the function of hGSTA1-1.

The conformational stability of L7A and L23A proteins was probed using thermal-induced unfolding, pulse proteolysis and urea-induced equilibrium unfolding studies. The thermal stability of L7A and L23A hGSTA1-1 was reduced in comparison to the wild-type protein. This was consistent with proteolytic susceptibility of L7A and L23A proteins which indicates that both mutants are more prone to thermolysin digestion when compared to wild-type hGSTA1-1. This also correlates with urea-induced equilibrium studies. The $\Delta G(H_2O)$ value (23.88 kcal.mol^{-1}) for the wild-type protein was reduced to 12.6 and 10.49 kcal.mol^{-1} in L7A and L23A hGSTA1-l, respectively. Furthermore, the *m*-values obtained for the L7A and L23A proteins were 1.46 and 1.06 kcal.mol^{-1}.M^{-1} urea, respectively; these were much lower than that obtained for the wild-type protein (4.06 kcal.mol^{-1}.M^{-1} urea). The low *m*-values obtained for the mutant proteins indicated that the cooperativity of hGSTA1-1 unfolding was significantly diminished in both mutations. The results obtained in this study indicate that the topologically conserved Leu7 and Leu23 in the N-subdomain of hGSTA1-1 play a crucial role in maintaining the structural stability of the thioredoxin-like domain and are not involved in the function of the enzyme.

This is for my THREE boys

My husband, Duncan Sibiya: There is no me without you, My Love!!

Sakhanya: Mommy is finally done with her homework. Thank you for being a good Boy.

Lethokuhle: from now onwards, I will read you more stories my Pumkins

'It is not the brightest who succeed… Nor is success simply the sum of the decisions and efforts we make on our own behalf. It is, rather, a gift. Outliers are those who have been given opportunities — and who have had the strength and presence of mind to seize them.'

Malcom Gladwell

Acknowledgements

I would like to thank the following people:

Professor Heini Dirr for his guidance, motivation and support over the course of my studies. Thank you for giving me opportunity to be part of your lab. It has been a wonderful experience and lots of lessons learnt!

All past and current members of the Protein Structure-Function Research Unit for all their assistance and friendship.

Dr Ikechukwu Achilonu, for your positive influence, helpful discussions, motivation and the friendship.

Dr Fritha Hennessy for giving up your time to edit and proof read my work.

Dr Sylvia Fanucchi, Dr Ikechukwu Achilonu and Dr Manuel Fernandes for their assistance on L7A and L23A crystal structures.

My friends at the CSIR: Dr Nobalanda Mokeona, Nqobile Ngubane and Lindiwe Nkabinde: your support certainly got me this far!

The University of the Witwatersrand, Professor H.W. Dirr and the South African National Research Foundation for financial support.

My family: Dad and Mom for always supporting me in my decisions and giving me strength to work towards my aspiration. Life is worth living knowing that I have you on my back!!! My brothers: Muzi, Mhlengi and my little ones Sandile and Sanele for ever creating a lovely and crazy atmosphere at home. My sisters: Thule, the best big sister you can ever wish for and my baby sister Siyathokoza, it feels good to have a baby sister that you can count on and for stepping in to give my Boys some love when I am working against time.

Table of Contents

Declaration i
Abstract ii
Acknowledgements v
Table of Contents vi
List of Figures viii
List of Tables x
List of Abbreviations xi
Research Outcomes xiii
1. Introduction 1
1.1. Glutathione transferase as a member of Trx-like superfamily 1
1.2. Structure of GST 6
1.3. Evolution of the Trx-like fold 10
1.4. Evolution across the GST family 14
1.5. Conformational stability of Trx-like fold 17
1.6. Objective 23
2. Material and Methods 24
2.1. Materials 24
2.2. Identification of conserved residues in the Trx-like domain of GST 24
2.3. Expression and purification of wild-type, L7A and L23A hGSTA1-1 25
2.3.1. Verification of plasmid DNA 25
2.3.2. Expression of wild-type, L7A and L23A hGSTA1-1 25
2.3.3. Purification of wild-type, L7A and L23A hGSTA1-1 26
2.4. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) 27
2.5. Protein concentration determination 28
2.6. Size exclusion-high pressure liquid chromatography (SE-HPLC) 28
2.7. Spectroscopic studies 28
2.7.1. Far-UV circular dichroism (CD) 28
2.7.2. Intrinsic fluorescence 29
2.8. Functional assays of wild-type, L7A and L23A hGSTA1-1 30
2.8.1. Specific activity 30
2.8.2. ANS binding properties 31
2.9. Crystallisation of L7A and L23A 32

2.10. Conformational stability 32
2.10.1. Thermal unfolding 32
2.10.2. Proteolysis 33
2.10.3. Urea-induced equilibrium studies 33
2.10.4. Fitting of data 34
3. Results 39
3.1. Multiple structural alignment 39
3.2. Over expression and purification of wild-type, L7A and L23A hGSTA1-1 44
3.3. Spectroscopic properties of L7A, L23A and wild-type hGSTA1-1 proteins 48
3.4. Far-UV CD 48
3.5. Fluorescence spectroscopy 52
3.6. Crystal structure of L7A and L23A proteins 52
3.6.1. Crystallisation of L7A and L23A mutants 52
3.6.2. Refinement of L7A and L23A models 55
3.6.3. Effects of L7A mutation on hGSTA1-1 structure 60
3.6.4. Effects of L23A mutation on hGSTA1-1 structure 60
3.7. Functional properties of wild-type, L7A and L23A hGSTA1-1 proteins 64
3.7.1. Specific activity 64
3.7.2. Ligand binding properties 67
3.8. Conformational stability of wild-type, L7A and L23A hGSTA1-1 68
3.8.1. Thermal-induced unfolding 69
3.8.2. Proteolytic susceptibility of wild-type, L7A and L23A hGSTA1-1 to thermolysin digestion 71
3.8.3. Urea-induced equilibrium unfolding studies of wild-type, L7A and L23A hGSTA1-1 proteins 74
4. Discussion 87
4.1. hGSTA1-1 structure tolerates L7A and L23A mutation in the βαβ subdomain 88
4.2. The packing of βαβ motif is optimised for functional properties of hGSTA1-1 96
4.3. L7A and L23A play a role in the stability of hGSTA1-1 97
4.4. Conclusions 109
5. References 110

List of Figures

Figure 1: Ribbon representation of the basic structure of the Trx-like superfamily 5

Figure 2: Ribbon presentation of the dimeric structure of hGSTA1-1 in complex with S-hexyl glutathione 7

Figure 3: Structural differences in the GST classes 9

Figure 4: Modification of the Trx-like fold 11

Figure 5: Evolution of cytosolic GSTs 15

Figure 6: The key interactions that initiate and consolidate the folding of Trx-like fold 20

Figure 7: Structure-based sequence alignment of the proteins in the GST family corresponding to βαβ motif of Trx-like fold 40

Figure 8: The environment of Leu7 in hGSTA1-1 42

Figure 9: The βαβ of hGSTA1-1 depicting the environment of Leu23 43

Figure 10: A portion of nucleotide sequencing chromatograms encoding for wild-type, L7A and L23A hGSTA1-1 45

Figure 11: Expression of wild-type, L7A and L23A hGSTA1-1 46

Figure 12: Purification of L7A hGSTA1-1 47

Figure 13: SDS-PAGE analysis of purified wild-type, L7A and L23A hGSTA1-1 49

Figure 14: SE-HPLC elution profile of hGSTA1-1 50

Figure 15: Far-UV CD spectra of wild-type, L7A and L23A hGSTA1-1 51

Figure 16: Fluorescence emission spectra for the wild-type, L7A and L23A hGSTA1-1 53

Figure 17: L7A and L23A hGSTA1-1 crystals 54

Figure 18: Fit of the L7A and L23A into the $2F_0$-F_c electron density map 56

Figure 19: Structural alignment of wild-type and its variants 58

Figure 20: Ramachandran plots of L7A and L23A structural models 59

Figure 21: Effect of L7A mutation on the Trx-like domain of hGSTA1-1 61

Figure 22: Phe30 side chain adopts a new conformation as a result of the L7A mutation 62

Figure 23: Effect of L23A mutation on the local environment of Leu23 63

Figure 24: The location of Leu7 and Leu23 residues with regards to the active site location. 65

Figure 25: Specific activity of the wild-type, L7A, L23A hGSTA1-1 using the standard CDNB conjugation assay 66

Figure 26: Fluorescence emission spectra of ANS bound to wild-type, L7A and L23A hGSTA1-1 68

Figure 27: Thermal-induced unfolding curves for wild-type L7A, L23A hGSTA1-1 70

Figure 28: The proteolytic susceptibility of wild-type, L7A and L23A hGSTA1-1 under native conditions 72

Figure 29: Proteolytic susceptibility of wild-type, L7A and L23A hGSTA1-1 to thermolysin digestion in the presence of urea..... 73

Figure 30: Recovery of wild-type, L7A and L23A hGSTA1-1 unfolding..... 75

Figure 31: Reversibility of urea-induced equilibrium unfolding of wild-type, L7A and L23A hGSTA1-1 76

Figure 32: Rayleigh scatter of wild-type, L7A and L23A hGSTA1-1 during urea-induced equilibrium unfolding..... 77

Figure 33: Urea-induced equilibrium unfolding of wild-type, L7A and L23A hGSTA1-1 ... 79

Figure 34: Comparison of urea-induced equilibrium unfolding of wild-type, L7A and L23A hGSTA1-1 81

Figure 35: Residual plot of the data fitted to a two-state model..... 83

Figure 36: Protein-concentration dependence of urea-induced equilibrium unfolding of wild-type, L7A and L23A hGSTA1-1 86

Figure 37: Differences in the βαβ motif of Alpha and Delta class GST..... 92

Figure 38: The βαβ motif of Theta class GST..... 94

Figure 39: Structural alignment of hGSTA1-1 (1K3L) and Trx (2TRX) 105

List of Tables

Table 1: The biological function of thioredoxin-like (Trx-like) proteins 2

Table 2: Hydrophobicities of the 20 naturally occurring amino acids..... 41

Table 3: Crystallographic X-ray data collection and refinement statistics L7A and L23A hGSTA1-1 57

Table 4: Urea-induced equilibrium unfolding thermodynamic parameters of wild-type, L7A and L23A hGSTA1-1 84

Table 5: Comparison of amino acid residues that are with 4 Å of Leu7, Ala21 and Ala19 side chain of class Alpha, Theta and Delta class GST, respectively..... 93

List of Abbreviations

A_{280}	Absorbance at 280 nm
Amp	ampicillin
ANS	anilino-1-naphthalene sulfonate
ArsC	arsenate reductase
BLAST	Basic Local Alignment Search Tool
CD	circular dichroism
CDNB	1-chloro-2,4-dinitrobenzene
CLIC	chloride intracellular channel
CM	carboxymethyl sepharose
C_m	urea-induced midpoint of unfolding
DHAR	dehydro-ascorbate reductase
DsbA	disulfide oxidoreductase
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetic acid
eEF1	eukaryotic elongation factor
GPX	glutathione peroxidase
Grx	glutaredoxin
GSH	reduced glutathione
G-site	glutathione binding site
GST	glutathione transferase
hGSTA1-1	human class alpha glutathione transferase with two type one subunits
H-site	hydrophobic electrophilic binding site
I	intermediate
IPTG	isopropyl-β-D-thio-galactoside
k_{cat}	catalytic constant
kDa	kilo dalton
K_M	Michaelis-Menten constant
L23A	replacement of wild-type leucine (L) with alanine (A) at position 23
L7A	replacement of wild-type leucine (L) with alanine (A) at position 7
LB	Luria-Bertani
MAPEG	membrane-associated proteins involved in eicosanoid and glutathione metabolism
Mr	relative molecular mass

m-value	susceptibility of protein to denaturant
N	native folded protein
NaCl	sodium chloride
PDB	Protein Data Bank
PDI	protein disulfide isomerase
PDO	protein disulfide oxidoreductase
*pf*GST	*Plasmodium falciparum* GST
Prx	peroxiredoxin
QSO	flavoprotein quiescin-sulfhydryl oxidase
r.m.s.d.	root mean square deviation
ScoI	copper binding proteins
SCOP	Structural Classification of Proteins database
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE-HPLC	size exclusion-high pressure liquid chromatography
SOC	super optimal broth with catabolite repression
SSPA	stringent starvation protein A
T_m	melting temperature
Trx	thioredoxin
U	unfolded protein
Ure2p	yeast prion protein
UV	ultraviolet
v/v	volume by volume
$\Delta G(H_2O)$	Gibbs free energy change of unfolding in the absence of denaturant

The IUPAC-IUBMB one and three letter codes for amino acids are used

Enzyme: Glutathione Transferase (EC 2.5.1.18)

Research Outcomes

Conferences:

The N-subdomain of glutathione transferases is stabilised by topologically conserved leucine residues. **Thandeka N. Khoza**, Ikechukwu Achilonu and Heini W Dirr. MBRT Research Day-Poster presentation (15/12/2012)

Protein data bank depositions:

[3U6V] Crystal Structure Analysis of the L23A Mutant of the Apo Form of Human Alpha Class Glutathione Transferase. Fanucchi, S., Achilonu, I.A., **Khoza, T.N**., Fernandes, M.A., Gildenhuys, S., Dirr, H.W. Release Date: 2011-10-26

[3Q74] Crystal Structure Analysis of the L7A Mutant of the Apo Form of Human Alpha Class Glutathione Transferase. Fanucchi, S., Achilonu, I.A., **Khoza, T.N**., Fernandes, M.A., Gildenhuys, S., Dirr, H.W. Release Date: 2011-01-12

1. Introduction

1.1. Glutathione transferase as a member of Trx-like superfamily

The GSTs (EC 2.5.1.18) are a superfamily made up of multifunctional enzymes that play a key role in the cellular detoxification process. Detoxification enzymes are grouped into three phases. The primary role of phase I enzymes is to convert toxic compounds into electrophiles and this reaction is mainly catalysed by cytochrome P450 (Furge and Guengerich, 2006). GSTs are the principal phase II enzymes and they play the main role in protecting cellular macromolecules from a wide variety of reactive electrophilic compounds. They catalyse the conjugation of glutathione (GSH, γ-glutamyl-cysteinyl-glycine) to an electrophilic substrate to form more soluble and non-toxic compounds (Armstrong, 1991; Wilce and Parker, 1994; Sheehan *et al.*, 2001; Frova, 2006). The excretion of these compounds from the cell is catalysed by phase III enzymes such as P-glycoprotein, multidrug resistance-associated proteins and organic anion transporting polypeptide 2 (Xu *et al.*, 2005).

GST catalyses the conjugation of GSH to a wide range of electrophilic compounds. The general scheme for the GST catalysed reaction is shown by the following equation:

$$\mathbf{GSH + RX \rightarrow GSR + HX} \qquad \textbf{Equation 1}$$

where R is the electrophilic substrate and X is the leaving group of the electrophilic compound.

The role of GST is not limited to detoxifying enzymes; they are also involved in other cellular processes as shown in Table 1. The GST superfamily has been grouped into four subfamilies based on immunological properties, amino acid sequence similarity, and their overall structural and functional properties (Mannervik *et al.*, 1992; Sheehan *et al.*, 2001; Frova, 2006). The four subfamilies are (a) soluble or cytosolic GSTs, (b) mitochondrial GSTs, (c) microsomal GSTs known as membrane-associated proteins involved in eicosanoid and glutathione metabolism (MAPEG) and (d) the plasmid–encoded bacterial fosfomycin-resistance GSTs (Sheehan *et al.*, 2001; Frova, 2006; Oakley, 2011). Sequence identity between the classes is reported to be as low as 20% (Sheehan *et al.*, 2001; Hayes *et al.*, 2005; Frova, 2006; Oakley, 2011).

Table 1: The biological function of thioredoxin-like (Trx-like) proteins

Name	**Biological Function**	**Reference**
ArsC	Catalyses reduction of arsenate to arsenite.	(Ji and Silver, 1992; Martin *et al.*, 2001; Messens *et al.*, 2002)
CLIC	Functions as chloride channel ions in membranes.	(Harrop *et al.*, 2001)
DsbA	Catalyses disulfide bond formation in secreted proteins.	(Martin *et al.*, 1993; Inaba, 2009)
GPX	Catalyses the reduction of H_2O_2 or organic hydroperoxides to water or corresponding alcohols using reduced glutathione. Modulates cellular oxidant stress and redox mediated signalling process.	(Handy *et al.*, 2009; Lubos *et al.*, 2011)
Grx	Plays a role as hydrogen donor for ribonucleotide reductase. It maintains reduced thiols in cellular proteins through glutathione disulfide redox reactions. Play a role as facilitators of protein folding. Play a role in iron regulation in the nucleus.	(Ojeda *et al.*, 2006; Berndt *et al.*, 2008; Meyer *et al.*, 2009)
GST	It catalyses conjugation of glutathione to a wide variety of electrophilic substrate. GSTs are also involved in a broad range of diverse cellular such as stress kinase regulation, prostaglandin synthesis and peroxidase activity.	(Board *et al.*, 1997; Desmots *et al.*, 2005; Eklund *et al.*, 2007)
PDI	It catalyses the reduction, oxidation, and isomerisation of disulfide bonds. Also functions as a molecular chaperone.	(Gruber *et al.*, 2006; Appenzeller-Herzog and Ellgaard, 2008)
Prxs	Catalyses the reduction of hydrogen peroxide, peroxinitrite and organic peroxide. Plays role in regulating H_2O_2- mediated cell signalling. Function as molecular chaperones.	(Chae *et al.*, 1994; Schroder *et al.*, 2000; Jang *et al.*, 2004; Hall *et al.*, 2009)
QSOX	Catalyses oxidation of sulfhydryl groups to disulfides with the reduction of oxygen to hydrogen peroxide.	(Thorpe *et al.*, 2002; Thorpe and Coppock, 2007; Heckler *et al.*, 2008)
ScoI	Plays a role in mitochondrial redox sensing and signalling. Functions as copper transporter.	(Williams *et al.*, 2005)
Trx	Acts as a hydrogen donor for ribonucleotide reductase. Plays main role in thiol-disulfide exchange reaction to maintain the redox state of the cell.	(Martin, 1995; Collet and Messens, 2010)

Each subfamily is further divided into classes and the cytosolic GSTs from a number of species have been well studied and are grouped into sixteen classes. The members of the same class will have slight differences in substrate recognition (Mannervik *et al.*, 1985; Hayes and Pulford, 1995; Sheehan *et al.*, 2001; Frova, 2006) and share very high sequence similarity (greater than 90%) (Sheehan *et al.*, 2001; Hayes *et al.*, 2005; Frova, 2006) Eight of the sixteen cytosolic GST classes are mammalian, including classes Alpha (A), Mu (M), Pi (P), Sigma (S), Theta (T), Omega (O), Kappa (K) and Zeta (Z) (Mannervik *et al.*, 1985; Board *et al.*, 1992; Buetler and Eaton, 1992; Sinning *et al.*, 1993; Board *et al.*, 1997, 2000). Human and rodent cytosolic GSTs are reported to share more than 40% sequence identity for isoenzymes within the same gene class, and less than 25% sequence identity for isoenzymes between classes (Sheehan *et al.*, 2001; Hayes *et al.*, 2005).

The GST nomenclature is based on structural and functional properties of the enzyme. The classes are designated by the names of the Greek letters such Alpha, Mu and Pi, abbreviated in Roman capitals: A, M and P. The class members are distinguished by Arabic numerals and the native dimeric protein structures are named according to their subunit composition. The organism from which the enzyme was isolated is represented by lower case letters and is included in the nomenclature (Mannervik *et al.*, 1992, 2005). Therefore, the homodimeric class Alpha enzyme that is composed of two copies of subunit 1 which was isolated from human is termed hGSTA1-1.

The current subfamilies and classes of GSTs are most likely to change. This is due to the increasing evidence of some members of MAPEG being more similar to Alpha class GST in terms of primary and tertiary structure. A comprehensive study of cytosolic GST sequence and structural similarity resulted in cytosolic GSTs being further divided into two subgroups depending on the catalytic residue GST uses in the interaction with GSH (Atkinson and Babbitt, 2009b). The two subgroups are tyrosine type GSTs (Y-GST) and serine or cysteine type GSTs (S/C-GST) which uses tyrosine, serine or cysteine in their interaction with GSH, respectively. The Y-GSTs includes Alpha, Mu, Pi and Sigma class whereas S/C-GSTs includes Beta, Delta, Omega, Phi, Tau, Theta, Zeta (Frova, 2006; Atkinson and Babbitt, 2009b). Recently, a new class of GST was identified from *Agrobacterium tumefaciens* and classified as Eta class; this furthers infers the likeliness of current GST classification to change (Skopelitou *et al.*, 2012a, 2012b).

The GST family is also classified as a member of the thioredoxin-like (Trx-like) superfamily. The members of this family are characterised by four β-strands sandwiched between two layers of α-helices resulting in an overall βαβαββα topology termed Trx-like fold (Figure 1). This structural feature is present in the N-terminal domain of GST. Thioredoxin proteins (Trx) were the first proteins to be identified with βαβαββα topology (Holmgren *et al.*, 1975). The increasing number of experimentally determined structures in the Protein Data Bank (PDB) has led to the identification of other protein families which are structurally homologous to Trx. These protein families include glutaredoxin (Grx) (Meyer *et al.*, 2009), protein disulfide isomerase (PDI) and its homologs (Gruber *et al.*, 2006; Appenzeller-Herzog and Ellgaard, 2008), disulfide oxidoreductase (DsbA) (Martin *et al.*, 1993; Inaba, 2009), protein disulfide oxidoreductases (PDOs) (Pedone *et al.*, 2010), flavoprotein quiescin-sulfhydryl oxidase (QSOX) (Thorpe *et al.*, 2002; Thorpe and Coppock, 2007; Heckler *et al.*, 2008), arsenate reductase (ArsC) (Ji and Silver, 1992; Martin *et al.*, 2001; Messens *et al.*, 2002), glutathione peroxidase (GPX) (Ladenstein *et al.*, 1979; Margis *et al.*, 2008) and peroxiredoxins (Prx) (Schroder and Ponting, 1998; Wood *et al.*, 2003; Copley *et al.*, 2004; Hall *et al.*, 2009), chloride intracellular channel (CLIC) (Harrop *et al.*, 2001) and copper binding proteins (ScoI) (Williams *et al.*, 2005). The biological functions of these proteins are shown in Table 1.

It has been observed from the aforementioned proteins that they share very low sequence similarity despite having a common fold. For example, the amino acid sequence similarity between Trx and Grx is 32% (Eklund *et al.*, 1984). These proteins were subsequently grouped into a common superfamily (Trx-like) of functionally-related proteins with low sequence similarity and common fold (Eklund *et al.*, 1984). Other structural features of the Trx-like superfamily are: (a) a conserved C-x-x-C catalytic motif, (b) a *cis*-proline residue located at the loop region before β3 which is implicated in substrate binding, (c) a conserved proline located in α1 and (d) charged residues which are near the active site residues that are involved in the proton transfer reaction for the redox mechanism (Martin, 1995; Nathaniel *et al.*, 2003; Copley *et al.*, 2004; Qi and Grishin, 2005; Pan and Bardwell, 2006; Atkinson and Babbitt, 2009a, 2009b; Pedone *et al.*, 2010).

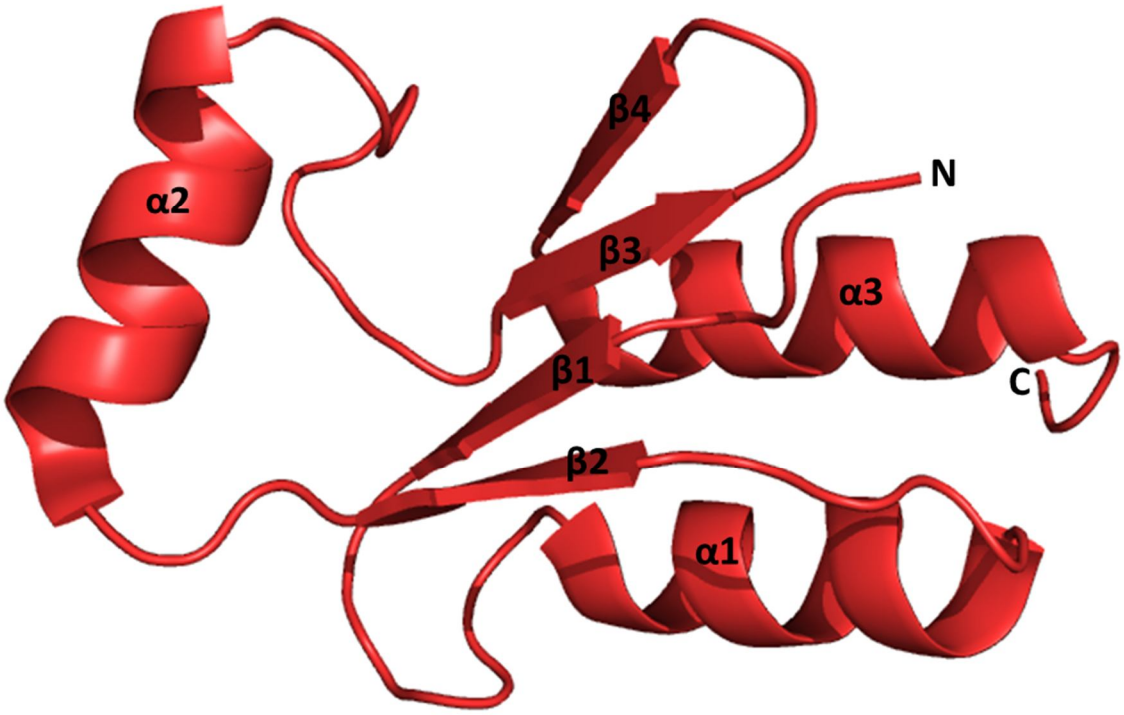


Figure 1: Ribbon representation of the basic structure of the Trx-like superfamily

The basic structure of Trx-like fold is shown using N-terminal domain of hGSTA1-1 (PDB code, 1K3O) which is characterised by βαβαββα topology. The figure was generated using PyMol v0.99 (DeLano Scientific, 2006).

Over the years, the Trx-like fold has been a subject of study at the sequence, structural and functional level mainly because it is present in a group containing some of the largest number of proteins that are believed to have evolved from a common ancestor (Holmgren, 1985; Martin, 1995; Qi and Grishin, 2005; Pan and Bardwell, 2006; Atkinson and Babbitt, 2009a; Pedone *et al.*, 2010). This is consistent with the view that Nature will rather alter or modify existing protein folds to perform new roles rather than to create new folds (Chothia, 1992; Pan and Bardwell, 2006; Bashton and Chothia, 2007).

1.2. Structure of GST

The first GST structure to be determined was from class Pi in 1991 (Reinemer *et al.*, 1991) and currently several structures of GST from all classes are have been determined. The overall structure of GST is similar despite the low sequence similarity that exists at subfamily and class level. The cytosolic GSTs exist as active homo- or hetero-dimers (Dirr *et al.*, 1994a, b; Wilce and Parker, 1994; Armstrong, 1997; Sheehan *et al.*, 2001; Fabrini *et al.*, 2009) and monomeric active forms of GST have been reported in the plant Lambda and DHAR class (Dixon *et al.*, 2002; Frova, 2006). Dimerisation of GST is highly specific, occurring only between subunits within the same gene class and it is consistent with structural data which show high level structural differences between classes (Sheehan *et al.*, 2001; Hayes *et al.*, 2005; Dourado *et al.*, 2008).

The dimeric structure of GST shows that each subunit is made up of two structurally distinct domains, the N-terminal domain (domain 1) and C-terminal domain (domain 2) (Figure 2). The structural feature/s of GST domain 1 has resulted in GST being classified as part of the Trx-like superfamily. The N-terminal domain is highly conserved across classes and species with a Trx-like (βαβαββα) topology whereas the C-terminal domain is variable and contains an all alpha-helical domain (Holmgren *et al.*, 1975; Martin, 1995; Frova, 2006; Pan and Bardwell, 2006). Domain 1 consists of two distinctive structural motifs; the N-subdomain (β1α1β2) and C-subdomain (β3β4α3) which are linked together by α2. Domain 2 is an all α-helical domain and the number of α-helices varies between each class.

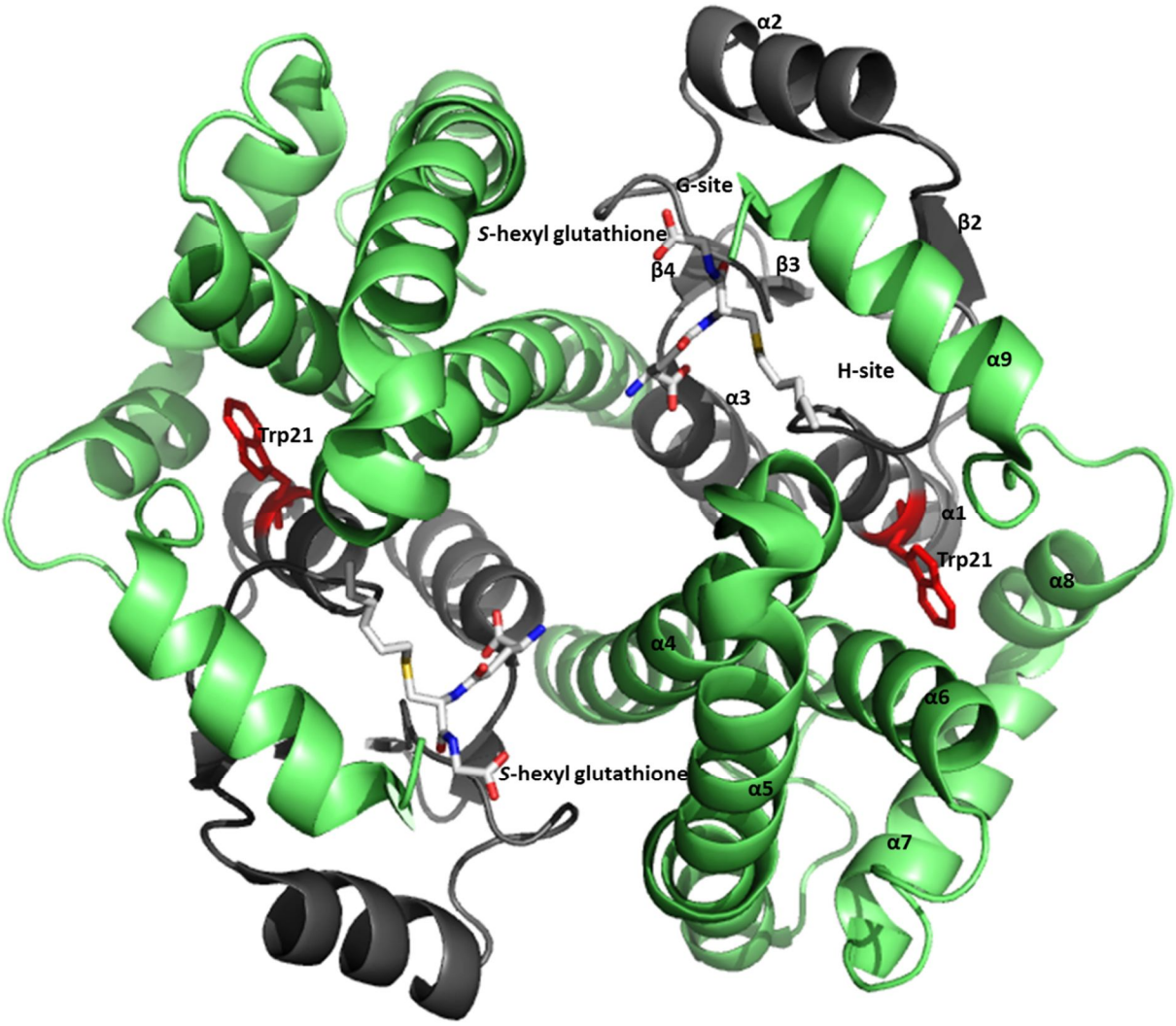


Figure 2: Ribbon presentation of the dimeric structure of hGSTA1-1 in complex with S-hexyl glutathione

The N-terminal domain which is made up of βαβαββα topology is shown in grey and the all α-helical C-terminal domain is shown in green. The ligand *S*-hexyl glutathione is represented as a stick model. The lone tryptophan residue (Trp21) located in the domain interface is shown in red as a stick model. The figure was generated from PDB file 1K3L (Le Trong *et al.*, 2002) using PyMol v0.99 (DeLano Scientific, 2006).

The active site of the enzyme is located at the subunit-subunit and domain-domain interface and, therefore, domain association and dimerisation is essential for functionally active enzyme (Erhardt and Dirr, 1995). Each subunit has its own active site and functions independently from each other. The active site is made up of a glutathione binding site (G-site) and non-specific hydrophobic binding site (H-site) which is located in the N- and C-terminal domains, respectively (Figure 2). The G-site is specific for GSH and subsequently conserved in the GST superfamily. On the other hand, the amino acid composition and topology of the H-site varies between classes thereby modulating substrate specificity between classes (Sinning *et al.*, 1993; Wilce and Parker, 1994; Armstrong, 1997).

Dimerisation plays a crucial role in the functioning and stability of GSTs. The overall dimeric structure of cytosolic GSTs is characterised by a V-shaped intersubunit crevice with the exception of the Kappa GST (Figure 3). The dimeric structure of Kappa GST adopts a unique open wing butterfly-like conformation. The class-specific intersubunit interactions lead to different dimeric conformations between GST classes. These class-specific intersubunit interactions also influences the ligand binding and catalytic activity of GST. It is believed that one of the mechanisms that GSTs have used to evolve and diversify their function, particularly through substrate binding is their structural differences between GST classes. The dimeric structures of Alpha, Mu and Pi classes are stabilised by a similar lock-and-key mechanism where two conserved residues, Met51 and Phe52 (numbered according to hGSTA1-1) located in the loop region protrude from one subunit and insert into the hydrophobic pocket between α4 and α5 of the adjacent subunit (Alves *et al.*, 2006). The number of residues involved in this lock-and key varies between classes. Urea-induced equilibrium and functional studies has shown that the lock-and-key between subunits is essential for maintaining dimer stability as well as the conformation of a functional active site (Sayed *et al.*, 2000; Stenberg *et al.*, 2000; Hornby *et al.*, 2002; Alves *et al.*, 2006). This lock-and-key motif at the dimer interface does not to be seem crucial for the GST dimer folding and assembly and this is in agreement with other GST classes such as Delta, Kappa and Sigma which lack this motif.

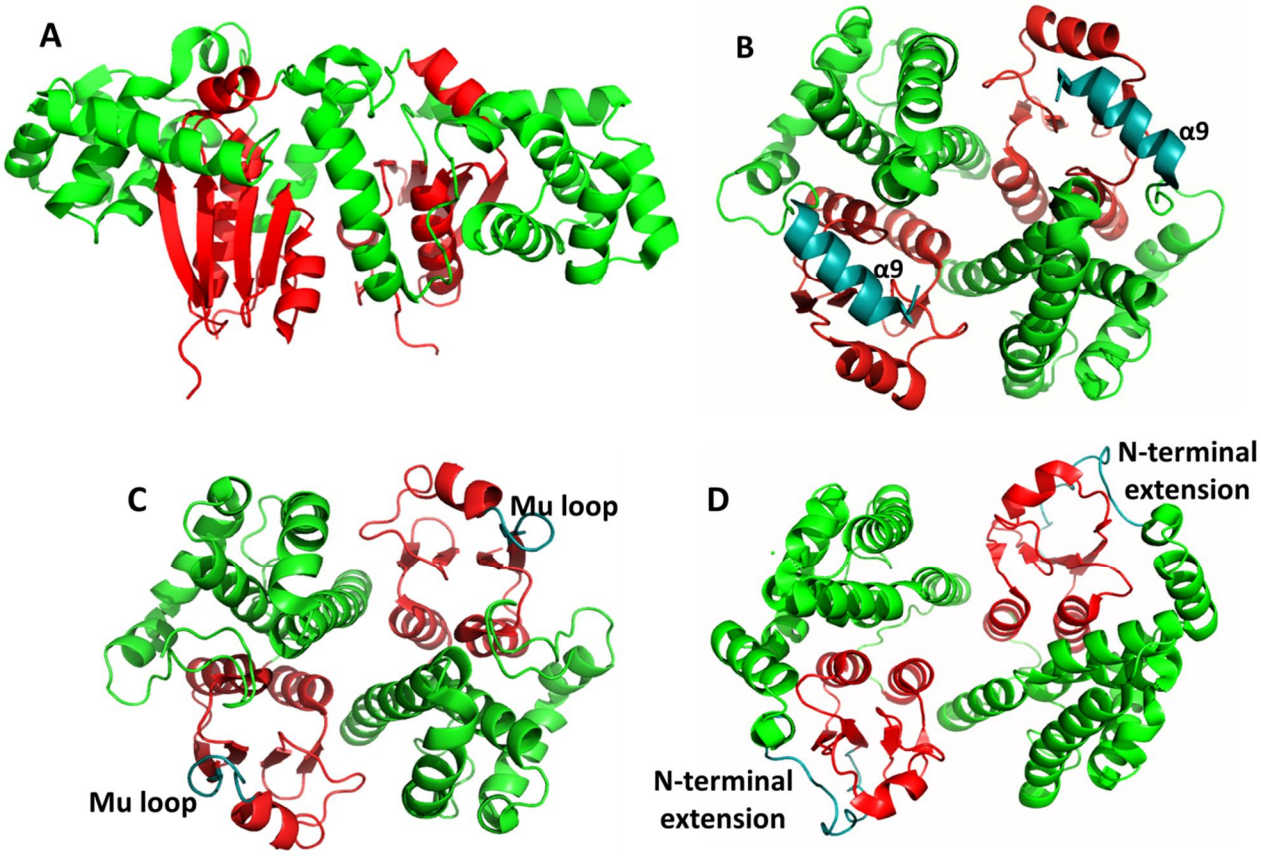


Figure 3: Structural differences in the GST classes

The dimeric structures of Kappa, Alpha, Mu and Omega Class GST are shown in A, B, C, and D, respectively. The N-terminal domain of GST (βαβαββα) is shown in red in all structures. The unique structural feature of Kappa GST is the insertion of an N-terminal domain in the C-terminal domain. In addition, the dimeric structure of Kappa GST (A) adopts a butterfly-like conformation which is different to the usual V-shaped conformation seen in Alpha (B), Mu (C) and Omega (D) GST classes. The structural class-specific features of the Alpha, Mu and Omega structures are shown in cyan. The figure was generated from the following PDB files:1RAW for Kappa class GST (Ladner *et al.*, 2004); 1K3L for Alpha class GST (Le Trong *et al.*, 2002); 1C72 for Mu class GST (Chern *et al.*, 2000) and 3Q18 for Omega class GST (Zhou *et al.*, 2012) using PyMol v0.99 (DeLano Scientific, 2006).

The GST classes which lack the lock-and-key intersubunit interaction are stabilised by a different mechanism. For example, the dimer interface of Sigma class is stabilised by electrostatic interactions (Stevens *et al.*, 2000; Huang *et al.*, 2008). On the other hand, the Delta class uses a lock-and-key 'clasp' motif where a conserved residue (key residue) protrudes and inserts into a hydrophobic pocket (lock) of the adjacent subunit (Wongsantichon and Ketterman, 2006). Once the 'key' residue is inserted into the lock, it then acts as a lock for the other key residue in the other subunit. The 'key' from both subunits 'hook around' each other generating a 'clasp' in the middle of the subunit interface (Wongsantichon and Ketterman, 2006).

1.3. Evolution of the Trx-like fold

Sequence, structural and functional analyses of the Trx-like superfamily have shown that the Trx-like fold has evolved in various ways to enable different catalytic reactions and yet maintain its structure. The Trx-like fold has evolved via (a) insertion/s, deletions, or even loop extensions and secondary structural elements, (b) combination with other domains, (c) duplication of the Trx-like fold, (d) modification in the C-x-x-C catalytic motif and (e) deletion of the C-x-x-C catalytic motif (Holmgren, 1985; Martin, 1995; Qi and Grishin, 2005; Pan and Bardwell, 2006; Atkinson and Babbitt, 2009a; Pedone *et al.*, 2010). The structural modifications of Trx-like fold are illustrated in Figure 4. The Trx-like fold is tolerant towards insertion of secondary elements in the N- or C- terminus as well as between the core of its elements. Examples of secondary structure insertions in the N-terminus include Trx and PDI which contain an additional α-helix and β-strand at the N-terminus (Martin, 1995; Pan and Bardwell, 2006; Appenzeller-Herzog and Ellgaard, 2008; Atkinson and Babbitt, 2009a; Pedone *et al.*, 2010) as well as PDO which has an extra α-helix. Even though Trx and PDI contain similars insertion at the N-terminus of the Trx-like fold, the location and arrangement of insertions within the fold is different.

The region between the N-subdomain and α2 is regarded as a "hot spot" for insertion. GPX, Prx, ArsC and DsbA have additional secondary structural elements in this region (Martin *et al.*, 1993, 2001; Martin, 1995; Sandalova *et al.*, 2001; Pedone *et al.*, 2010). These insertions can be as large as 70 residues as in the case of DsbA (Martin *et al.*, 1993). Another modification of the Trx-like fold is an inclusion of an α-helical domain as in DsbA and GST to diversify its function (Martin, 1995; Pedone *et al.*, 2010).

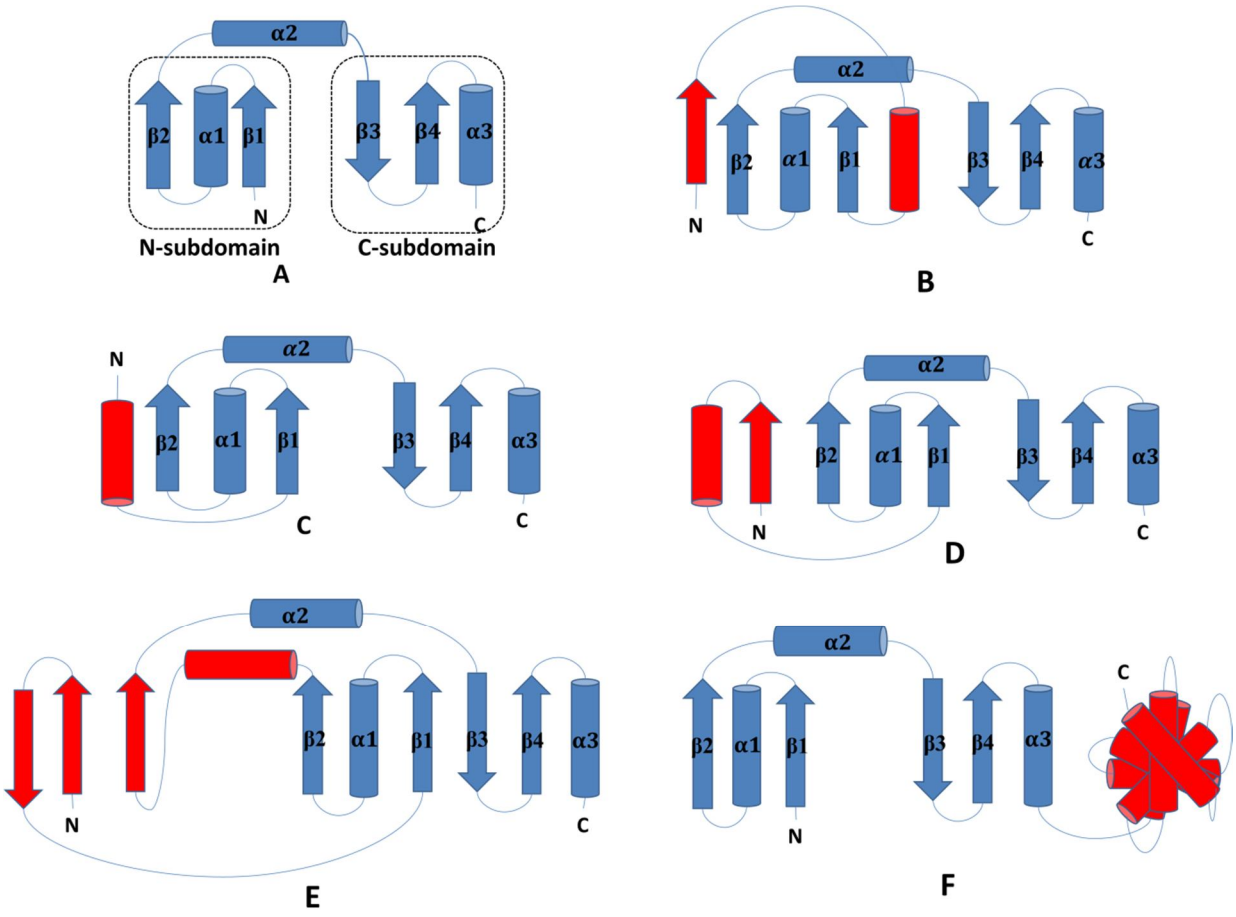


Figure 4: Modification of the Trx-like fold

The basic structure of the Trx-like fold is shown in A using the Grx protein. The α-helices and β-sheets are represented as cylinders and arrows, respectively. In A, the N- and C-subdomain are connected by α2. The modifications that exist in the Trx-like fold are shown by red secondary elements in B to F. Shown in B (Trx), C (PDO), and D (PDI) are different types of insertion in the N-terminus of the fold. The Trx-like domain has also evolved by domain addition; the added domain can either be inserted between β2 and α2 as in DsbA proteins (E) or forms an independent domain as in GST proteins (F). The image was adapted from Pan and Bardwell (2006).

In DsbA the α-helical domain is inserted between the N-subdomain and α2 whereas in cytosolic GSTs it does not form part of the Trx-like fold and forms an independent domain. PDIs present another evolution of the Trx-like fold which is duplication of the Trx fold resulting in multidomain proteins (Gruber *et al.*, 2006; Atkinson and Babbitt, 2009a; Pedone *et al.*, 2010). PDI contains four Trx-like folds, a linker region and an insertion in the C-terminus; however, only two of the four Trx-like domains contain the C-x-x-C catalytic motif. Within the PDI family, the Trx-like fold has evolved further by increasing the number of Trx-like domains to diversify their function (Gruber *et al.*, 2006).

The solved structures of Trx-like superfamily members bound to ligands or substrates has demonstrated that modification exist in the highly conserved C-x-x-C catalytic motif. The modification of this motif includes the number and/or position of the cysteine residues. For example, in monothiol Grxs, the C-terminal cysteine can be substituted by any residue (Berndt *et al.*, 2008), whilst in Sco1, the two cysteines can be separated by three or four residues (Banci *et al.*, 2008; Pedone *et al.*, 2010). PDI homologs reveal another form of active site modification where both cysteines can be substituted by other residues resulting in the loss of redox activity (Thorpe *et al.*, 2002; Thorpe and Coppock, 2007; Heckler *et al.*, 2008). It also has been observed from the solved structure of Trx-like members that the location of the catalytic motif differs within the Trx-like superfamily (Qi and Grishin, 2005; Atkinson and Babbitt, 2009b). The active site in the Trx-like fold may be located in the N-termini of α1 and the nearby loop region or along the edges of β2 and β4. The positioning of the active site determines which secondary elements of the fold play a role in stabilising the active site conformation or the nucleophile which is used in the catalytic mechanism of these proteins. This, again, highlights the versatile nature of the Trx-like fold.

A bioinformatics study performed by Atkinson and Babbitt, (2009b) which used over four thousand representatives of the Trx-fold like family showed that 22% of Trx-like proteins lack the cysteine-based catalytic motif. Furthermore, the proteins that lacked this motif were dominated by GSTs (69%). It is, however, worth mentioning that the essential catalytic residues (cysteine, tyrosine and serine) used by the majority of the Trx-like superfamily are also used by GST to bind its substrate, glutathione (GSH) (Atkinson and Babbitt, 2009a, 2009b). Another difference between the majority of the Trx-like superfamily members and GSTs, is the role that the N-terminus of α1 plays to stabilise the nucleophile in the catalytic

mechanism of Trx-like proteins. In cytosolic GSTs, α1 is not in close proximity with nucleophile and is, therefore, unlikely to play a role in its stabilisation (Ladner *et al.*, 2004; Atkinson and Babbitt, 2009a). These results presented GST as an unique family when viewed in context of the catalytic mechanism of Trx-like superfamily.

Also, whilst Trx-like proteins share a common fold, some of these proteins have overlapping functions and others have very specialised functions (Table 1). For example, thioredoxin, glutaredoxin and DsbA are all redox proteins catalysing thiol-disulfide exchange reaction. On the other hand, CLICs and glutathione peroxidase have specialised roles (Table 1). Another interesting feature about GSTs is that whilst their major role is catalysing the conjugation of GSH to a wide variety of electrophilic compounds, they can still catalyse other reactions which are mainly performed by other Trx-like proteins. For example, Kappa GST has been shown to have peroxidase activity which is the main function of GPX proteins despite low sequence similarity between these two proteins (Morel *et al.*, 2004; Yu *et al.*, 2005). Both Kappa GST and GPX use GSH as a reducing substrate; however, these enzymes use different catalytic residues in their catalytic mechanism. Kappa employs Ser16 as the catalytic residue (Ladner *et al.*, 2004; Wang *et al.*, 2011), in contrast GPX uses selenocysteine (Yu *et al.*, 2005). Crystal structures of Kappa GST and GPX shows that the relative positions of these catalytic residues in the Trx-like fold in the two enzymes (Ser16 in hGSTK and a selenocysteine in the GPX) are well conserved (Ren *et al.*, 1997; Yu *et al.*, 2005; Wang *et al.*, 2011). Yu and co-workers have shown that incorporation of selenocysteine into the glutathione binding site (G-site) present in GSTs results in enhanced peroxidase activity, comparable with that of natural GPX (Yu *et al.*, 2005). Recently, in another study, the activity of Grx was modified to peroxidase activity using protein engineering (Ge *et al.*, 2009). Ge and co-workers also substituted cysteine residues in the active site of Grx to selenocysteine and the resultant Grx mutant displayed a similar kinetic mechanism and catalytic properties as the GPX enzyme. The findings of Ge *et al.* (2009) and Yu *et al.* (2005) not only shed light onto modification of the Trx-like domain to broaden its function but also demonstrate that (a) a single amino acid residue can modulate the activity of Trx-like proteins and (b) strengthens the general belief that Trx-like proteins evolved from a common ancestor to expand its function.

In further expanding our knowledge and understanding in Trx-like superfamily, GST will be used to probe how the Trx-like domain has evolved as to diversify its function and yet maintain its structure. GSTs are one of the ideal candidates for such a study because they are distant to Trx proteins in terms of amino acid sequence and function despite sharing the overall Trx-like fold.

1.4. Evolution across the GST family

The function of GST is very diverse across the family as well as at class level. For example, a comparison of two isozymes from Alpha class (GSTA1-1 and GSTA4-4) which share 54% sequence identity shows striking difference in terms of CDNB catalytic activity (Bruns *et al.*, 1999; Hou *et al.*, 2007). Interestingly, GSTA4-4 lacks activity towards CDNB which is generally used to determine GST activity. It displays high activity towards alkenals such as hydroxynonenal and isoprostanes which is consistent with the view that GSTA4-4 may have evolved to play a primary role in the metabolism of lipid peroxidation products (Bruns *et al.*, 1999; Hou *et al.*, 2007). In contrast, GSTA1-1 has high activity towards CDNB and lacks activity against 4-hydroxynonenal and isoprostanes. The difference in substrate activity has been attributed to differences in catalytic residues, the overall conformation of the active site and structural dynamics of the proteins (O'Brien, 2006; Hou *et al.*, 2007). In order to get insights as to how GST has evolved to diversify its function it is, therefore, important to look at sequence, structure and functional data collectively. Cytosolic GSTs have been widely studied and therefore will be used to highlight evolutionary events that have occurred within Trx-like domain of GST.

Crystal structures of cytosolic GSTs show that these enzymes may share common catalytic features and an overall fold but they also have distinct structural differences. It was initially believed that Theta GSTs were the ancestors from which the GST family originated (Pemble and Taylor, 1992; Sheehan *et al.*, 2001). However, with the increasing number of solved structures and functional characterisation of GST enzymes, a new evolutionary pathway for cytosolic GSTs have been proposed (Ladner *et al.*, 2004; Robinson *et al.*, 2004; Frova, 2006) and is shown in Figure 5. The evolutionary pathway of GST is divided into phases I and II.

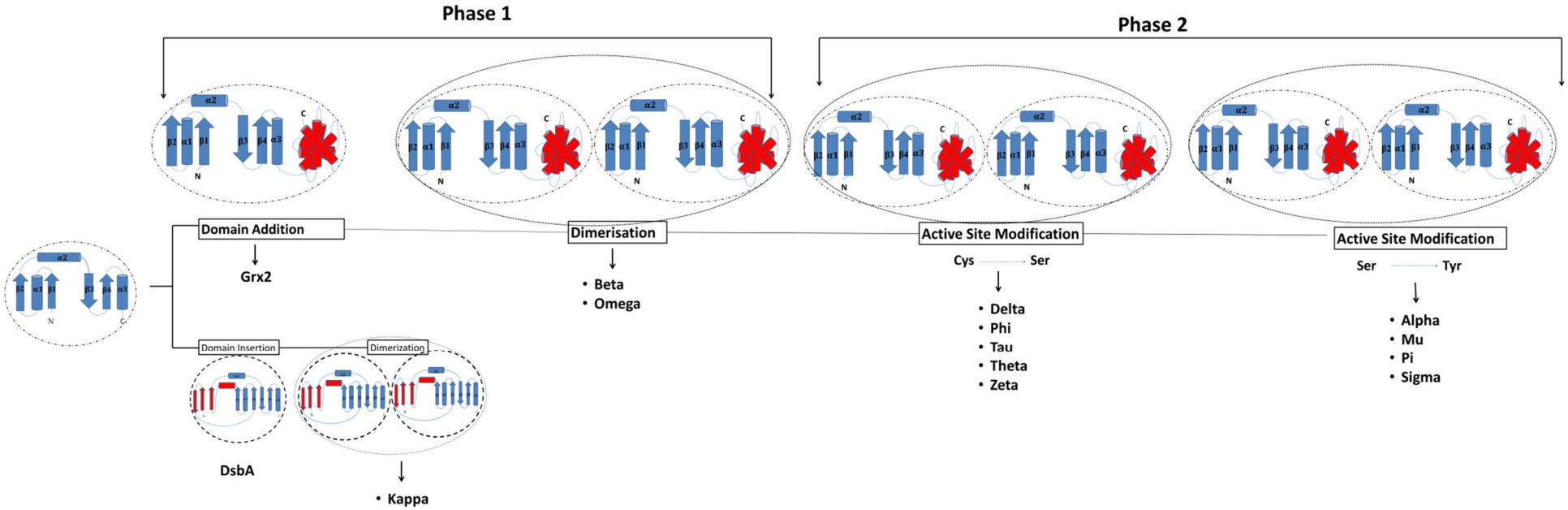


Figure 5: Evolution of cytosolic GSTs

Proposed evolution model of GSTs based on their structural and functional data. The key steps in the proposed pathway include domain addition, dimerisation and active site modification. The image was adapted from Frova *et al.* (2006). .

Phase I encompasses earlier events of the Trx-like fold evolution and it is now generally accepted that Trxs and Grxs are ancient progenitors of all cytosolic GSTs. In phase I, the Trx-like fold was modified with the addition of an α-helical domain via two parallel events. The α-helical domain is either inserted in the Trx-like fold as in DsbA or it forms an independent domain as in Grx2 (Xia *et al.*, 2001; Ladner *et al.*, 2004; Frova, 2006) It is also possible that DsbA evolved from Grx2 via domain repositioning (Ladner *et al.*, 2004; Wang *et al.*, 2011). After α-helical domain addition to the Trx-like fold in phase I, a dimerisation step took place which led to Kappa GST and cytosolic GSTs. Initial studies on Kappa GSTs reported that this GST class does not share any significant amino acid sequence with other GST classes (Pemble *et al.*, 1996; Morel *et al.*, 2004). However, Li and co-workers showed that when the alignment of Kappa GST with other GST proteins is limited to Trx-like domain, Kappa GST shares 19% sequence similarity with Theta GST (Li *et al.*, 2005). Secondary structural topology of Kappa GST is unique in the context of GST superfamily and it resembles that of the DsbA family in the Trx-like superfamily (Figure 4). The overall dimeric structure of Kappa GST also adopts a different structural fold in comparison to other GSTs (Figure 3). Despite these differences, Kappa GSTs have conserved features of the GST family including the conserved *cis*-proline residue involved in a hydrogen bonding network that plays a role in catalysis and stability of GSTs and Trx-like proteins (Kelley and Richards, 1987; Allocati *et al.*, 1999; Charbonnier *et al.*, 1999; Nathaniel *et al.*, 2003; Ladner *et al.*, 2004). Also, Kappa GSTs use a hydroxyl residue (Ser16) as the catalytic residuein a catalytic mechanism commonly used by other GSTs such the Beta and Omega Classes (Ladner *et al.*, 2004; Li *et al.*, 2005).

In the cytosolic GSTs, structural and functional data suggest that Beta and Omega GSTs were the first to evolve from Grx2. This is strengthened by both Beta and Omega class GSTs using cysteine as an active residue as in the Grx2 enzymes. The crystal structure of PmGSTB1-1 bound to GSH shows that Cys10 plays a role in binding GSH (Rossjohn *et al.*, 1998); however, mutation of this residue does not affect GST activity (Casalone *et al.*, 1998). Most GSTs use a hydroxyl residue (Tyr or Ser) in the N-subdomain of the enzyme for stabilising the activated form of GSH (Armstrong, 1991, 1994, 1997; Atkinson and Babbitt, 2009b) however in the Beta and Omega classes this is not the case (Casalone *et al.*, 1998; Nishida *et al.*, 1998; Rossjohn *et al.*, 1998; Board *et al.*, 2000). Instead, the N-terminal cysteine residue located in the G-site (Cys10) is proposed to play a role as a hydrogen-bond

donor that stabilises the negatively charged thiolate ion of GSH during catalysis (Nishida *et al.*, 1998). The Beta and Omega class GSTs also contain oxidoreductase activity involving formation of a mixed disulfide with the active-site cysteine residue (Cys10) (Caccuri *et al.*, 2002a, 2002b; Whitbread *et al.*, 2005). The multiple roles of Cys10 in Beta and Omega class GSTs illustrate how the Trx-like fold evolved to enable GSTs to diversify its function.

In the phase II evolutionary pathway, the step that followed dimerisation was modification of the active site residue (Figure 5). The Y-GSTs are believed to have evolved last in the GST family and this is supported by their taxonomy distribution (Armstrong, 1997). Comparison of Y-GST and S/C-GST structures shows overall similarity with distinctive structural features between classes (Figure 3). The Alpha class has an additional helix (α9) hanging over the Trx fold which is crucial for catalysis (Le Trong *et al.*, 2002; Mosebi *et al.*, 2003). Micaloni *et al.* (2003) further demonstrated the role of helix 9 in GST catalytic activity by showing that catalytic properties of Pi class GST could be modified to those similar to Alpha class GST by appending α9 to the C-terminus of Pi GST. The class Mu contains an extended and mobile loop between β2 and α2. This loop is called the Mu loop and is important for the determination of enzyme affinity for its substrates (Ji *et al.*, 1992; Board *et al.*, 2000; Hearne and Colman, 2006). The unique structural feature of Omega class GST is the extension of the N-terminal with 20 residues which lies in close proximity to the C-terminus (Board *et al.*, 2000).

1.5. Conformational stability of Trx-like fold

In the Trx-like family, Trx proteins have been extensively studied to establish structure-stability and structure-function relationships. This is due to the Trx protein being small (about 107 residues), made up of a single domain, highly stable and well characterised. Various techniques have been employed to study the equilibrium and kinetic unfolding of Trxs. The unfolding of Trx (equilibrium and kinetic) is characterised by a two-state mechanism (Kelley and Stellwagen, 1984; Bhutani and Udgaonkar, 2003). On the other hand, the refolding of this protein is complex with multiple phases (Kelley and Stellwagen, 1984; Kelley *et al.*, 1986). Mutagenesis studies have shown that there are key or topologically conserved residues within Trx-like fold whose main role is to drive folding and/or maintain the structural integrity of the Trx-like fold. Pro56 (numbered according to hGSTA1-1) is an example of a highly conserved residue within the Trx-like family and is

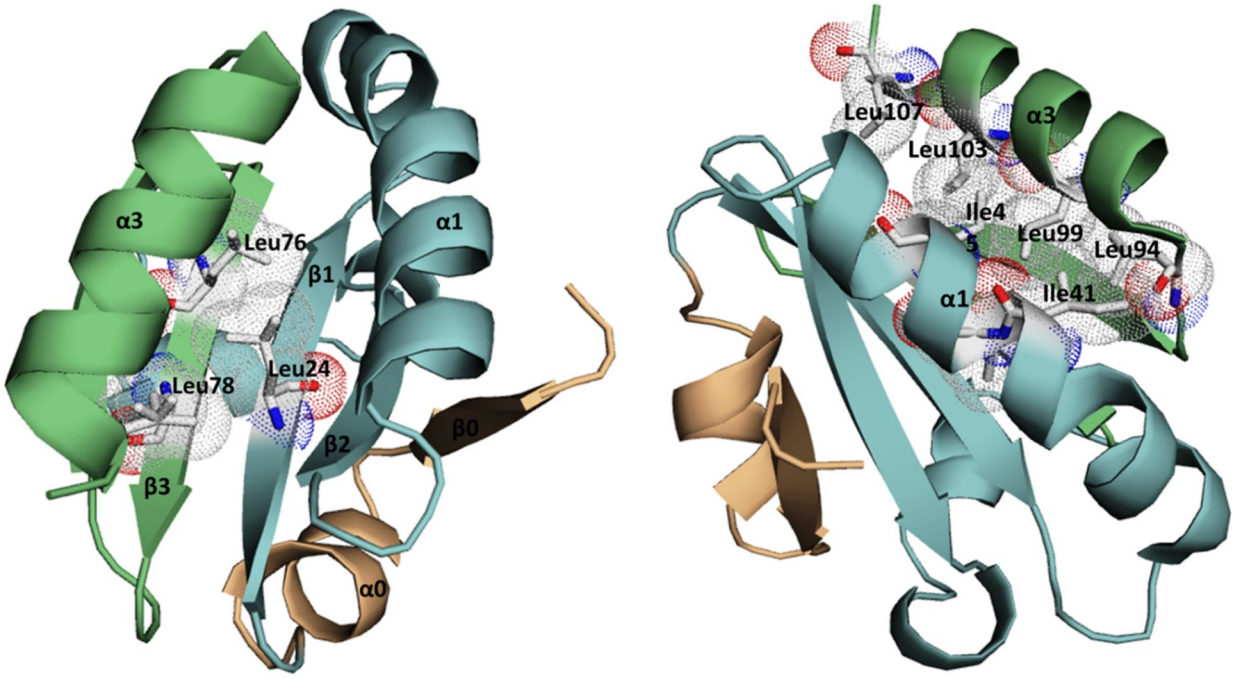
present in most GST sequences. Pro56 is located in the loop region connecting α2 and β3 in the N-terminal domain of GST in a *cis* conformation. This residue has been shown to play an important role in the catalysis and stability of Trx-like proteins such as Trx, DsbA, and GST (Kelley and Richards, 1987; Allocati *et al.*, 1999; Charbonnier *et al.*, 1999; Nathaniel *et al.*, 2003). In hGSTA1-1, Pro56 plays a role in maintaining the correct G-site conformation by packing against two important catalytic residues (Tyr9 and Arg15) and Met16 (Stenberg *et al.*, 1991; Dirr *et al.*, 1994b; Bjornestedt *et al.*, 1995). Mutation of Pro56 to glycine (hGSTA1-1), to alanine (Trx and DsbA) or to serine (GST from *Proteus mirabilis*) results in the loss of catalytic activity and stability of the enzyme (Kelley and Richards, 1987; Gleason, 1992; Charbonnier *et al.*, 1999; Nathaniel *et al.*, 2003). The loss in activity could be attributed to the change in G-site conformation and/or loss of favourable interactions between GSH and the hydroxyl group of Tyr9 (Allocati *et al.*, 1999; Nathaniel *et al.*, 2003). Urea-induced equilibrium studies of Pro56 mutants results in low $\Delta G(H_2O)$ and *m*-values which are indicative of a loss in protein stability and possible deviation from two-state to a three-state unfolding mechanism (Kelley and Richards, 1987; Charbonnier *et al.*, 1999; Nathaniel *et al.*, 2003). The Trx-like fold in proteins such GPX which does not have a conserved proline residue in the loop region connecting α2 and β3 have a longer loop which does not require adopting a sharp conformation in connecting β3 (Nathaniel *et al.*, 2003). This, again, demonstrates the flexibility of the Trx-like fold to modify itself to enable function.

In dissecting the pathway of Trx-like domain folding, Louis *et al.* (2001) probed the structure and stability of a Trx chimeric protein. This hybrid protein was made up of the N-and C-subdomain from the human and *Escherichia coli* sequence, respectively. The overall structural topology of the Trx chimeric protein is similar to that of human and *Escherichia coli* proteins (Louis *et al.*, 2001; Dangi *et al.*, 2002). Thermal and chemical-induced unfolding studies showed that the stability of Trx chimeric proteins is reduced in comparison to the human and *Escherichia coli* proteins (Louis *et al.*, 2001). The hydrophobic interactions and hydrogen bonding network on the interface of βαβ and ββα motifs in Trx chimeric protein were similar to those in parental protein. However, subtle changes in the side chain packing were present in the Trx chimeric protein due to non-optimised complementarity of the N- and C-subdomain (Dangi *et al.*, 2002). The conclusion drawn from the studies of the Trx chimeric protein were: (a) the Trx-like fold is very tolerant to modification of the packing density in the hydrophobic core and (b) the stability of the Trx-

like domain can be reduced without significantly inducing changes in the overall structure of fold (Louis *et al.*, 2001; Dangi *et al.*, 2002).

A number of site-directed mutagenesis studies in Trxs have identified conserved hydrophobic residues in the Trx-like fold tha,t when mutated, the stability of fold is compromised (Hellinga *et al.*, 1992; Godoy-Ruiz *et al.*, 2005). The location of these residues in the Trx-like fold is shown in Figure 6. However, individual mutations of above-mentioned residues fail to show the key interactions which results in the formation of the native Trx-like fold because their mutation did not affect the global structure of Trx (Hellinga *et al.*, 1992; Godoy-Ruiz *et al.*, 2005). In further investigating the key interactions that drive folding or maintain the Trx-like domain, Trx mutants were created which expressed various fragments of the fold. Tasayco *et al.* (2000) suggested that the initial step of Trx-like domain folding involves hydrophobic interactions between β1 and β3 of the N- and C-subdomain, respectively (Figure 6A). This suggestion was strengthened by observing that protein isolated from mutants expressing N- or C-subdomain fragments of the Trx-like domain are disordered without any native secondary or tertiary structure. However upon combining the two proteins a Trx-like complex is formed which has low level of Trx enzymatic activity (Slaby and Holmgren, 1979; Jeng and Dyson, 1995). The protein from N- or C-subdomain fragments lacks any form of structure individually but when these proteins interact a native Trx-like structure is formed. This highlights importance of molecular recognition in folding of the subdomains and the Trx-like domain.

Santos *et al*. (2007) showed that the C-terminal helix of the ββα motif (α3) plays an important role in the final steps of Trx-like domain folding. The Trx mutant, with truncated α3, adopts a molten globule conformation. However, when this mutant interacts with peptide encoding for α3, it adopts a native Trx-like structure. The key interactions that leads to the final folding steps and stabilisation of the Trx-like fold are shown in Figure 6B and involves hydrophobic interactions between residues in α1 and α3 (Santos *et al.*, 2009).



A **B**

Figure 6: The key interactions that initiate and consolidate the folding of Trx-like fold

The basic structure of the Trx-like fold is shown using Trx (PDB code, 2TRX) which is characterised by βαβαββα topology. The nomenclature of the secondary elements of the Trx-like fold is the same as the one used for Grx proteins. The β0 and α0 (shown in brown) are the additional secondary elements of Trx inserted in the N-terminus of the fold. The N- and C-subdomains are shown in cyan and green, respectively. The dots represent the van der Waals radii space filling model. Shown in A, are the hydrophobic interactions between N- and C-subdomains which initiates the folding of the Trx-like domain whereas B shows a leucine rich α3 packing against α1 resulting in the folding of Trx-like domain. The images were generated using PyMol v0.99 (DeLano Scientific, 2006).

In hGSTA1-1, Ile71 is a highly conserved residue in the GST family and is located on α3 at the hydrophobic core of the N-terminal domain in is close proximity to the G-site. The Ile71 side chain is involved van der Waals interactions with residues form both βαβ and ββα motifs of Trx-like domain (Achilonu *et al.*, 2010). Van der Waals interactions in the hydrophobic core are essential in directing the protein folding process and in maintaining the stability of the hydrophobic cores which in turn stabilize the native protein structure via correct and optimised packing (Dill, 1990a, 1990b). Mutation of Ile71 to alanine or valine does not affect the overall structure of hGSTA1-1, however, the Trx-like domain of hGSTA1-1 in I71A is solvent accessible. It contains two water molecules which assist in the packing of this domain thereby maintaining its structural integrity (Achilonu *et al.*, 2010). Urea-induced equilibrium unfolding studies of both mutants show that the Ile71 mutation reduces the stability of the protein in comparison to the wild-type. Furthermore, unfolding data from the mutants also suggest possible deviation from two-state unfolding observed in the wild-type to a three-state demonstrating that the topologically conserved Ile71 is essential in the stability of the N-terminal domain of GST.

The α1 of the Trx-like fold also plays an important role in the structure, function and stability of proteins in the GST family. In hGSTA1-1, Trp21 in α1 is located within the hydrophobic core of the Trx-like fold and is involved in the interdomain interactions with residues from α6 and α8 in domain 2. In monomeric GST homologues, hCLIC1 and Grx2 a methionine residue (Met32, numbered according to hCLIC1) is in an equivalent position to Trp21 (Parbhoo *et al.*, 2011). Trp21 forms part of the lock-and-key motif stabilising the structure of GST. The side chain of Trp21 protrudes from domain 1 and fits into the hydrophobic pocket in domain 2B (Wallace *et al.*, 2000; Balchin *et al.*, 2010). Mutation of Trp21 or Met32 to alanine does not affect the overall structure of the protein. The analysis of W21A structure shows cavity formation within the interdomain of hGSTA1-1 (Wallace *et al.*, 2000; Balchin *et al.*, 2010). This cavity reduces the packing density and the van der Waals interactions at the domain interface leading to destabilisation of hGSTA1-1 structure and reduction in the catalytic activity of the enzyme (Wallace *et al.*, 2000; Balchin *et al.*, 2010). Trp21 mutations also affected the conformation of the H-site which is 15 Å away from mutation site demonstrating that conserved residues may be involved in long-range interactions which are crucial to maintaining the native structure of proteins. The H-site became more solvent accessible in Trp21A in comparison to the wild-type (Balchin *et al.*, 2010). Equilibrium

unfolding of the M32A deviated from a two-state (N ↔ U) seen in the wild-type protein, to a three-state (N ↔ I ↔ U) (Fanucchi *et al.*, 2008; Parbhoo *et al.*, 2011). The deviation from a two-state to a three-state in the M32A signifies a loss in the cooperative folding of domain 1 highlighting a role of Met32 in the stability of the Trx-like fold. The fact that mutation of Trp21 or its equivalent in dimeric and monomeric forms in Trx-like proteins results in similar effects demonstrates the role of topologically conserved residues in maintaining the structure and hence the function and stability of the Trx-like fold (Wallace *et al.*, 2000; Balchin *et al.*, 2010; Parbhoo *et al.*, 2011). Findings from Balchin *et al.* (2010) and Wallace *et al.* (2000) correlate with other GST studies which demonstrated that interdomain interactions also impact on the intersubunit interaction at the dimeric level (Sayed *et al.*, 2000; Hornby *et al.*, 2002). Luo *et al.* (2002) created two domain-exchanged chimeric isoenzymes from class Mu. The domains of the class Mu isozymes M1-1 and M2-2 were exchanged resulting in chimeras (M12 and M21) with one domain from M1 and one domain from M2. The equilibrium unfolding and conformational stability of two domain-exchanged chimeric isoenzymes showed that GSTM12 and M21 monomers were less stable than the wild-type monomer (Luo *et al.*, 2002). This study indicated that domain interface complementarity is critical for correct domain-domain interaction which in turn impacts on protein stability.

Studies of the of N-terminal domain of GST demonstrate that this domain plays an essential role in the binding of GSH and it forms part of the hydrophobic core of GST proteins. Furthermore, studies, using both site-directed mutagenesis and X-ray crystallography, have demonstrated that conserved residues play an important role in catalysis and stability of this protein. In most cases, mutation of conserved residues does not adversely affect the Trx-like fold of GSTs indicating that the Trx fold adapts to overcome the effects of these mutations. This is consistent with extensive modifications that have occurred in Trx-like fold proteins to diversify its function. The conserved residues that are located in the protein core mainly play a role in maintaining the stability of the fold via correct packing, domain-domain interactions as well as subunit-subunit interactions (Mirny and Shakhnovich, 1999; Friedberg and Margalit, 2002; Schueler-Furman and Baker, 2003). Alteration of the protein's core may or may not affect the catalysis of GST. In cases where the catalysis of the protein is affected, the conserved residues are directly involved in the formation of the functional conformation of the active site and/or in the binding of GSH. Moreover, the conserved residues also

influence binding of ligands at the dimer interface by maintaining class specific subunit interactions.

1.6. Objective

Most residues that are conserved across the entire Trx-like proteins in the N-terminal domain of GST, which have been studied, are located in the ββα motif (C-subdomain) of GST proteins. Trp21, which is located in the βαβ motif (N-subdomain) of GST proteins, is the only residue conserved in the GST family that has been studied. Currently, not much is known about conserved residues that are essential to the formation of the βαβ motif in the GST family. These residues could play a major role in the structure, function and stability of the GST since the βαβ forms part of the active site and hydrophobic core of this protein.

The primary aim of this work was to identify topologically conserved residues in the Trx-like domain and seek possible reason/s for the conservation of the identified residues with regards to structure, function and stability of the GST family. The conserved residues were identified using bioinformatics tools and subsequently systematically replaced using site-directed mutagenesis. This study was, therefore, conducted as a comparison to the wild-type and its variants in terms of structure, function and stability in order to broaden our knowledge of Trx-like proteins. The impact of mutations on the structure of the enzyme was studied using spectroscopic tools and X-ray crystallography. Far-UV CD and fluorescence spectroscopy were used to detect whether the mutations introduced any conformational changes at the secondary and tertiary levels, respectively. As mentioned earlier, the hydrophobic core of GST is mainly made up of residues from the Trx-like domain. Therefore, the solved mutant structures were used to determine how changes in the core packing of the N-terminal domain of GST affects the structure of the protein. Also, knowing that the Trx-like fold is highly adaptive (Figure 4), the mutant structures would provide insight and increase our understanding as to how the Trx-like fold adapts to maintain its formation. The effects of the mutations on the overall functional properties of the enzyme were determined by obtaining the specific activity and ligand binding properties of the mutant proteins. The impact of the mutations on hGSTA1-1 structural stability was assessed using thermal unfolding, pulse proteolysis and urea-induced equilibrium unfolding studies.

2. Material and Methods

2.1. Materials

The expression vector, pKHA1, encoding wild-type hGSTA1-1 was a kind donation from Prof B. Mannervik (Department of Biochemistry, University of Uppsala, Sweden). The pKHA1 vectors encoding the sequence for L7A and L23A hGSTA1-1 were from Dr. S. Gildenhuys (Protein Structure-Function Research Unit, University of the Witwatersrand). The sequencing to confirm the plasmid was performed by Inqaba Biotech (Pretoria, South Africa). Dithiothreitol (DTT), isopropyl-β-D-thio-galactoside (IPTG) and molecular weight marker (SM0431) were purchased from Fermentas Life Sciences (Ontario, Canada). 8-Anilino-1-naphthalenesulfonic acid (ANS), thermolysin and reduced L-glutathione (GSH) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Ampillicin was purchased from Roche Diagnostics (Mannheim, Germany). Ultrapure (99.5%) urea was purchased from Merck chemicals (Darmstadt, Germany). All other reagents used were of analytical grade. All solutions were prepared using distilled water and filtered through a 0.45 µm filter.

2.2. Identification of conserved residues in the Trx-like domain of GST

Proteins belonging to the GST family were identified from Structural Classification of Proteins database (SCOP) (Murzin *et al.*, 1995). The protein crystal structures belonging to the GST family were downloaded from the Protein Data Bank (PDB) (http://www.rcsb.org.pdb). The crystal structures of dimeric proteins were modified, using SwissPDB viewer (Guex and Peitsch, 1997), so that each file consisted of only one subunit. Each subunit was used to perform structural alignments to identify conserved residues in the Trx-like domain of GSTs using 3DCoffee (O'Sullivan *et al.*, 2004).

2.3. Expression and purification of wild-type, L7A and L23A hGSTA1-1

2.3.1. Verification of plasmid DNA

In order to propagate and verify the plasmid (pKHA1) containing cDNA (inserts) which encodes wild-type, L7A and L23A hGSTA1-1, their plasmid DNA was used to transform *Escherichia coli* ECLONI cells (Lucigen, Middleton, WI, USA). The competent cells were transformed using the heat shock method (Chung *et al.*, 1989). Briefly, 1 µl of plasmid DNA was added to the competent cells and incubated on ice for 30 minutes. The cells were heat shocked at 42 °C for 45 seconds and rapidly cooled on ice for 2 minutes. This was followed by adding 800 µl of SOC medium (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 250 mM KCl, 1 M glucose, 2 M $MgCl_2$) and incubated at 37 °C for 1 hour. The cells were then plated on Luria-Bertani Agar (LBA) plates (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1.0% (w/v) NaCl, 1.5% (w/v) agar) supplemented with ampicillin (100 $\mu g.ml^{-1}$) and chloramphenicol (30 $\mu g.ml^{-1}$) and incubated overnight at 37 °C to select successfully transformed cells.

The transformed cells were picked and used to inoculate 5 ml of LB broth (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1.0% (w/v) NaCl, 1.5% (w/v) supplemented with (100 $\mu g.ml^{-1}$) ampicillin) and incubated overnight at 37 °C with shaking at 250 rpm. Plasmid DNA was then isolated from the transformed cells using the GeneElute plasmid isolation kit from Fermentas Life Sciences (Ontario, Canada) as per the manufacturer's instructions. The wild-type, L7A and L23A isolated plasmid DNA were sent to Inqaba Biotech (Pretoria, South Africa) for sequencing.

2.3.2. Expression of wild-type, L7A and L23A hGSTA1-1

In order to express the wild-type, L7A and L23A hGSTA1-1, *Escherichia coli* BL21 (DE3)/ pLysS cells (Lucigen, Middleton, WI, USA) were transformed, as described in section 2.3.1 with plasmid DNA. The transformed cells were used to inoculate 10 ml LB broth. The cells were grown at 37 °C with agitation at 250 rpm. The overnight culture at, 100 fold dilution, was used to inoculate fresh LB broth and grown again at 37 °C with shaking at 250 rpm. The cells were grown until mid-exponential phase (0.3 – 0.6 OD_{600}) and at this point the expression of hGSTA1-1 was induced with the addition of 1 mM IPTG. The cells were further grown at 37 °C with agitation at 250 rpm to achieve optimum conditions for protein

expression. The cells were then pelleted by centrifugation at 6500 x g for 20 minutes at room temperature and stored at -80 ^{o}C for at least 2 hours to promote cell lysis or until they were used for hGSTA1-1 purification.

In order to test for expression of wild-type, L7A and L23A hGSTA1-1, the harvested cells were resuspended in 20 ml of 10 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5. The resuspended cells were sonicated and centrifuged at 6500 x g for 30 minutes at 4 °C. The supernatant and insoluble fractions were analysed for hGSTA1-1 expression using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) as described in section 2.4.

2.3.3. Purification of wild-type, L7A and L23A hGSTA1-1

The harvested cells that were stored at -80 °C were then thawed in ice at 4 °C and resuspended in 20 ml of 10 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5. DNAse I and lysozyme were added to a final concentration of 10 mg.ml^{-1} to the cell suspension and incubated for 30 min at 4 °C. The cells were then lysed by pulse sonication on ice and centrifuged at 6000 x g for 30 minutes. The cleared cell lysate was purified using CM-Sepharose cation exchange chromatography (Wallace *et al.*, 1998a). Prior to loading the supernatant onto the column, it was diluted to 50 ml with 10 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5) to optimise protein binding to the column. The diluted cell supernatant was loaded onto a CM-Sepharose cation exchange column pre-equilibrated with 5 column volumes of 10 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5. Once all the supernatant entered the resin, the unbound protein was washed off with at least 10 volumes of 10 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5 until the absorbance of the column eluent at 280 nm was zero. The protein eluting off the column was detected by measuring the absorbance of the eluent, at 280 nm in real-time, using the on-board spectrophotometer on the ÄKTAprime automated chromatographic system coupled to a computer with PrimeView 1.0 software (GE Healthcare Life Sciences, Uppsala, Sweden. The bound proteins were eluted using a 300 ml linear salt gradient of 0 - 300 mM NaCl prepared in 10 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5. The elution profile of the eluent was monitored at 280 nm and 3 ml fractions of the eluent were collected. Eluted fractions were analysed using SDS-PAGE (section 2.4). The fractions that contained hGSTA1-1 were pooled and concentrated using

ultrafiltration through an Amicom PM-10 membrane to a volume of less than 10 ml. This was followed by dialysing the protein in storage buffer (20 mM sodium phosphate, 1 mM EDTA, 0.02% sodium azide, pH 6.5) at 4 °C. The purity and homogeneity of the concentrated proteins were assessed by SDS-PAGE.

An additional purification step (affinity chromatography) was performed on the mutant proteins as a precautionary measure to ensure high levels of purity were achieved since the mutants were to be crystallised. The concentrated eluent from the CM-Sepharose column was further dialysed at 4 °C into 100 mM Tris-HCl, 5 mM EDTA, 0.02% sodium azide, pH 7.5 to optimise protein binding to the *S*-hexyl glutathione affinity column. The dialysed protein was then loaded onto the affinity column. The non-specifically bound proteins were washed off the column using at least 10 column volumes of 100 mM Tris-HCl, 5 mM EDTA, 0.02% sodium azide, pH 7.5 until the absorbance of the column eluent at 280 nm was zero. The bound protein was eluted off the column with 50 mM glycine-NaOH buffer, pH 10. The eluent was monitored at 280 nm and 4 ml fractions were collected. The pH of the eluted proteins was quickly neutralised by adding an equal volume of Mes-NaOH buffer, pH 6.5 and the protein fractions were assessed using SDS-PAGE.

2.4. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

The purity and homogeneity of the hGSTA1-1 proteins were assessed using 14% SDS-PAGE as described by Laemmli (1970). Prior to loading the protein samples onto the gel, the proteins were incubated with SDS-PAGE sample buffer (0.125 M Tris-HCl, 4 % (w/v) SDS, 20% (v/v) glycerol, 5% (v/v) β-mercaptoethanol and 0.02% (w/v) bromophenol blue, pH 6.8) for 5 minutes at 95 ^{o}C. Samples were electrophoresed for one hour at 180 V. The molecular weight marker (Fermentas Life Sciences, Ontario, Canada) used contained seven proteins, namely, β-galactosidase (116 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), lactate dehydrogenase (35 kDa), restriction endonuclease Bsp98I (25 kDa), β-lactoglobulin (18.8 kDa) and lysozyme (14.4 kDa). The gels were stained for 2 hours in 2% (w/v) Coomassie Blue R250 staining solution containing 13.5% (v/v) glacial acetic acid and 18.75% (v/v) ethanol. The gels were destained with 40% (v/v) ethanol and 10% (v/v) glacial acetic acid until the background was clear.

2.5. Protein concentration determination

The protein concentrations were determined spectrophotometrically by applying the Beer-Lambert Law.

$$\mathbf{A = \varepsilon_{\lambda} cl} \qquad \textbf{Equation 2}$$

where ε_{λ} is the molar absorption coefficient at wavelength λ, *c* is the concentration and *l* is the pathlength of the light through the solution. The molar absorption coefficient of wild-type, L7A and L23A hGSTA1-1 was calculated using equation 2 (Perkins, 1986):

$$\mathbf{\varepsilon(M^{-1}cm^{-1}) = 5500(\sum Trp) + 1340(\sum Tyr) + 150(\sum Cys)} \qquad \textbf{Equation 3}$$

where, ΣTrp is the sum of tryptophan residues, ΣTyr is the sum of tyrosine residues and ΣCys is the sum of cysteine residues within the protein; and the constants indicate the molar absorption coefficients for the respective residues. The molar absorption coefficient at 280 nm for dimeric wild-type, L7A and L23A hGSTA1-1 was estimated to be 38200 $M^{-1}cm^{-1}$.

2.6. Size exclusion-high pressure liquid chromatography (SE-HPLC)

Size exclusion chromatography was performed on the wild-type, L7A and L23A hGSTA1-1 to determine whether the mutations affected the oligomeric state of the protein. The column used was TSK Gel G2000SW_{XL} (TOSOH Corporation, Tokyo, Japan) size exclusion column with a resolution of 5-150 kDa. The column was equilibrated with 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5. The protein concentration used was 20 µM. The protein was eluted from the column at 0.5 $ml.min^{-1}$ and the eluent was monitored at 280 nm with an absorbance detector (Spectra-Physics (model: UV100SP), Fremont, CA, USA).

2.7. Spectroscopic studies

2.7.1. Far-UV circular dichroism (CD)

CD is a technique that is extensively used to study structure of proteins, and it depends on the fact that molecules which are optically active interact differently with right and left polarized

light. In proteins, the polypeptide backbone and aromatic side chains are optically active molecules (Woody, 1995). The CD spectrum in the far-UV region (190 - 250 nm) is dominated by secondary structure features with α-helices giving the two strong negative minima at 222 nm and 208 nm and a positive ellipticity at 190 nm whilst β-sheets display a minimum at 216 nm and positive ellipticity at 195 nm.

The far-UV CD spectra of wild-type, L7A and L23A were measured at 20 °C using a Jasco J-810 spectropolarimeter with Spectra Manager software v1.5.00 (Jasco Inc., Tokyo, Japan). The protein concentration used was 2 µM in 2 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5. The unfolded protein was prepared in 2 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5 with 8 M urea. Far-UV CD spectra were recorded from 190 to 250 nm in a quartz cuvette with a path length of 0.2 cm. The bandwidth and the data pitch were at 1 and 0.2 nm, respectively. The spectrum obtained was an average of ten accumulations. The spectra were normalised by calculating the mean residue ellipticity [θ] deg.cm^2.dmol^{-1}.residue^{-1} using the following equation (Woody, 1995):

$$[\boldsymbol{\theta}] = (\mathbf{100}.\boldsymbol{\theta})/(\mathbf{C}.\mathbf{n}.\boldsymbol{l}) \qquad \textbf{Equation 4}$$

where θ is the CD signal in millidegrees, C is the protein concentration in mM, n is the number of residues, and l is the path length in cm.

2.7.2. Intrinsic fluorescence

The indole-ring of tryptophan and the phenyl-ring of tyrosine fluoresce when excited at 280 nm; however, tryptophan residues have stronger emission fluorescence properties than tyrosine (Lakowicz, 1991). Excitation of protein at 280 nm results in the transfer of fluorescence resonance energy from excited tyrosine to tryptophan which results in a more intense fluorescence emission spectrum (Lackowicz, 1983). Tryptophan residues can be exclusively excited at 295 nm. The fluorescence properties of tryptophan are influenced by the polarity of its environment and this allows them to be used in probing changes in their local environment. Depending on the number and location of the tryptophan residues, they can be used as probes for local or global changes. A single tryptophan residue can be used to monitor local conformational changes whereas multiple tryptophan residues distributed across a three-dimensional structure would monitor global conformational changes. The

fluorescence spectra of wild-type, L7A and L23A were measured using a Perkin Elmer luminescence spectrometer LS50B (Waltham, MA, USA) in a quartz cuvette with a 10 mm path length at 20 °C. The hGSTA1-1 proteins were excited at 280 nm and the intrinsic fluorescence emission was monitored between 290 and 500 nm. The excitation and emission bandwidths were set at 5 nm and the emission spectra collected at 200 nm/min. The protein concentration used was 2 μM in 10 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5. The spectrum obtained is an average of three accumulations per sample. The denatured proteins were prepared in 10 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5 with 8 M urea.

2.8. Functional assays of wild-type, L7A and L23A hGSTA1-1

2.8.1. Specific activity

The standard GSH-CDNB conjugation assay was performed to determine the specific activity of wild-type, L7A and L23A hGSTA1-1 (Habig *et al.*, 1974; Habig and Jakoby, 1981). GSTs catalyse the conjugation of 1-chloro-2,4-dinitrobenzene (CDNB) to GSH, which results in the formation of the chromophoric product 1-(*S*-glutathionyl)-2,4-dinitrobenzene. This compound absorbs light at 340 nm.

Triplicate samples of 1-10 nM wild-type, L7A and L23A hGSTA1-1 were prepared in 0.1 M sodium phosphate, 1 mM EDTA, pH 6.5 in the presence of GSH (1 mM). The reaction was initiated by addition of CDNB to a final concentration of 1 mM. The CDNB was prepared in 100% ethanol thus making final ethanol concentration of 3% v/v in the assay volume of 3 ml. The enzyme activity was measured spectrophotometrically at 20 °C using a Jasco V–630 UV/Vis spectrophotometer (Jasco, Japan) by monitoring the formation of 1-(*S*-glutathionyl)-2,4-dinitrobenze at 340 nm. All reactions were corrected for the non-enzymatic controls. The specific activity of the enzyme was then calculated using an extinction coefficient of 9600 $M^{-1}.cm^{-1}$ for 1-(*S*-glutathionyl)-2,4-dinitrobenze. The specific activity (μmol. $min^{-1}.mg^{-1}$) was determined by linear regression as the slope of a plot between the initial velocity of complex formation ($\mu mol.min^{-1}$) versus protein amount (mg).

2.8.2. ANS binding properties

8-Anilinonaphtalene-1-sulfonic acid (ANS) is an anionic dye which is commonly used as an extrinsic fluorescent probe in the analysis of structural properties of protein due to its high affinity for hydrophobic patches in proteins (Slavik, 1982; Semisotnov *et al.*, 1991; Engelhard and Evans, 1995). The fluorescence properties of ANS are largely dependent on its environment. Free ANS fluoresces at 540 nm when excited at 390 nm. However, when ANS binds to hydrophobic patches in protein, its maximum emission wavelength shifts to a lower wavelength (blue-shift) with an increase in fluorescence intensity (Slavik, 1982; Engelhard and Evans, 1995).

A 20 mM ANS stock solution was prepared in 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5. The concentration of ANS was confirmed using an extinction coefficient of 5000 $M^{-1}.cm^{-1}$ at 350 nm and applying the Beer-Lambert law (section 2.5) (Weber and Young, 1964). To probe the effects of L7A and L23A mutation on hGSTA1-1, ANS fluorescence spectra in the absence and presence of wild-type and its variants were obtained as described previously (Wallace *et al.*, 1998). ANS, at a final concentration of 200 µM, was added to protein solution (2 µM) prepared in the same buffer as ANS. The protein-ANS mixture was incubated for 60 minutes at room temperature. The samples were excited at 390 nm and emission spectra were recorded from 390 to 600 nm. The excitation and emission bandwidths were set at 5 nm and the emission spectra collected at 200 nm/min. The spectrum obtained is an average of three accumulations per sample.

For the urea-induced unfolding studies, the protein solution (2 µM) was incubated with different concentrations of urea (0 - 8 M) at room temperature for an hour to achieve equilibrium. The ANS was then added to protein solutions to final concentration of 200 µM and incubated again for another hour. A series of blanks were generated, each containing 200 µM ANS with the appropriate urea concentration (0 - 8 M). The fluorescence emission intensities at 482 nm were extracted and plotted as a function of urea concentration. All fluorescence measurements were recorded in a quartz cuvette with a 10 mm path length using a Perkin Elmer luminescence spectrometer LS50B and FLwinlab v4.0 software (Waltham, MA, USA).

2.9. Crystallisation of L7A and L23A

Crystals of recombinant L7A and L23A hGSTA1-1 were grown as described previously (Achilonu *et al.*, 2010; Gildenhuys *et al.*, 2010a). Briefly, crystals of L7A and L23A proteins were grown by the hanging-drop vapour-diffusion method at 293 K using a 24-well microplate. Each hanging drop (4, 6 or 8 µl) contained equal volumes of protein stock solution and reservoir buffer. The stock protein concentration for both the mutants was 10 mg ml^{-1} in 0.1 M Tris–HCl, pH 7.5, containing 10 mM DTT and 0.02% sodium azide. The reservoir buffer was PEG 4000 (20% (w/v) in 0.1 M Tris–HCl, pH 7.5, 10 mM DTT and 0.02% sodium azide. The crystals were harvested, briefly soaked in the reservoir buffer and mounted on a cryoloop.

L7A and L23A hGSTA1-1 diffraction data were collected using a Bruker X8 Proteum system with a Microstar copper rotating-anode generator with Montel 200 optics, a PLATINUM 135 CCD detector and an Oxford Cryostream Plus system. Crystals were cooled to 113 K in a stream of nitrogen during data collection and images were collected covering an oscillation angle of 0.5° per image. The data sets were processed using APEX and SAINT software (Bruker AXS Inc., Madison, Wisconsin, USA) for both mutants.

The structures of L7A and L23A hGSTA1-1 were solved by molecular replacement using MOLREP (Vagin and Teplyakov, 2000) as implemented in the CCP4 suite of programs (Collaborative Computational Project, Number 4, 1994), using wild-type hGSTA1-1 (PDB code 1K3L) (Le Trong *et al.*, 2002) as the search model. Model refinement was performed with REFMAC5 (Murshudov *et al.*, 1997) and model building was performed with Coot (Emsley and Cowtan, 2004). Stereochemical validation of the model was performed using PROCHECK (Laskowski *et al.*, 1996) and MolProbity (Chen *et al.*, 2010). PyMOL (DeLano Scientific, 2006) was used to generate images of the structures.

2.10. Conformational stability

2.10.1. Thermal unfolding

The thermal-induced equilibrium unfolding of wild-type, L7A and L23A hGSTA1-1 was determined by monitoring the change in ellipticity at 222 nm. The protein samples (2 µM)

were prepared in 2 mM sodium phosphate, pH 7.5, containing 1 mM EDTA and 0.02% sodium azide. Samples were subjected to increasing temperatures between 20 and 80 °C using a continuous temperature gradient of 1 °C/min, controlled by a Jasco PTC-423S Peltier-Type thermostat control system attached to a water-cooling bath. The far-UV CD measurements were obtained using Jasco J-810 CD spectropolarimeter (Jasco Inc., Tokyo, Japan) in a quartz cuvette with a path length of 0.2 cm. The bandwidth and the data pitch were at 1 and 0.2 nm, respectively. The temperature unfolding profiles were normalised by calculating the mean residue ellipticity [θ] (see section 2.7.1) and plotted against the temperature. The experiment was repeated three times.

2.10.2. Proteolysis

The proteolytic susceptibility of the mutants with respect to wild-type was determined using a published method (Park and Marqusee, 2001, 2005). Briefly, the proteins (10 µM) were prepared in 20 mM Tris-HCl, pH 7.5, 50 mM NaCl 10 mM $CaCl_2$ and varying urea concentrations (0 - 8 M). Prior to the proteolysis reaction, the proteins were incubated at room temperature for an hour to reach equilibrium. The reaction was initiated by addition of thermolysin (0.2 mg/ml) and incubated for a minute at room temperature. After one minute of incubation, EDTA (2.5 mM) was added to quench the reaction and SDS–PAGE loading buffer (0.5 mM Tris-HCl, 20% (v/v) glycerol, 10% (w/v) SDS, 100 mM β-mercaptoethanol, 0.05% bromophenol blue, pH 6.8). The proteolytic susceptibility of the proteins was assessed using 15% SDS-PAGE.

2.10.3. Urea-induced equilibrium studies

The conformational stability of the protein can only be assessed if the unfolding process is reversible and the native state of the protein can be recovered (Pace, 1986). The reversibility of the unfolding process gives an indication whether equilibrium exists between the folded and unfolded states. To test recovery, the proteins (20 µM) in 8 M urea were incubated at room temperature for one hour to allow complete unfolding. After incubation, the unfolded proteins were diluted 10-fold with 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5 and incubated again for another hour to ensure that the proteins return to their native conformation. The control sample was made up of 2 µM protein in 0.8 M urea (representing the final conditions used for refolding). The degree of protein refolding was

assessed using far-UV CD (section 2.7.1) and intrinsic tryptophan fluorescence (section 2.7.2.) as probes. The recovery of the native conformation was assessed on the basis of the extent to which the protein regained its secondary and tertiary structure and was presented as a percentage.

The proteins (2 µM) were incubated with 0-8 M urea for an hour to reach equilibrium. The urea-induced unfolding curves were monitored using far-UV CD, tryptophan fluorescence (excitation 280 nm), ANS binding (excitation 390 nm) and light scattering (excitation and emission at 340 nm). ANS was added to a final concentration of 200 µM. The light scattering was performed to detect the presence of aggregates along the unfolding curve. To check for hysteresis, 20 µM of unfolded protein was diluted with 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5 to a final concentration of 2 µM protein in urea (0-8 M) and allowed to refold for an hour at room temperature. The unfolding and refolding curves were compared to check whether the unfolding process is reversible.

For all the urea-induced equilibrium unfolding experiments, the proteins and urea were prepared in 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7. The pH of the stock urea solution was adjusted to pH 7.5 using phosphoric acid, filtered and the concentration of 10 M confirmed using an Atago R5000 pocket refractometer (Tokyo, Japan).

2.10.4. Fitting of data

The urea-induced unfolding of wild-type hGSTA1-1 is characterised to occur via a two-state mechanism, subsequently the unfolding wild-type data was analysed using a two-state fit (Wallace *et al.*, 1998b). Since the unfolding of mutants (L7A and L23A) has never been characterised before, the unfolding data from both these proteins were analysed using both two-state and multi-state fitting models.

2.10.4.1. Two-state data fit

In a two-state mechanism, equilibrium is reached between the folded native (N) and the unfolded (U) species. Therefore, only the native and the unfolded states are present at significant levels of concentration (Pace, 1986).

$$N_2 \rightleftarrows 2U$$ **Equation 5**

Therefore, it follows that for a 2-state unfolding mechanism:

$$\mathbf{f_N + f_U = 1} \qquad \textbf{Equation 6}$$

where f_N and f_U represent the fraction of the folded and unfolded protein species, respectively. Therefore, the signal obtained along the unfolding process is contributed by both species:

$$\mathbf{y = y_N . f_N + y_U . f_U} \qquad \textbf{Equation 7}$$

where y is the signal obtained from specific probe, y_N represents the *y* value for the native state and can be extrapolated from the linear pre-transition region of the unfolding data, and y_U represents the y value for the unfolded state and can be extrapolated from linear post-transition region of the unfolding data. Combining equation (6) and (7) yields:

$$\mathbf{f_U = (y_N - y)/(y_N - y_U)} \qquad \textbf{Equation 8}$$

$$\mathbf{f_N = (y - y_U)/(y_N - y_U)} \qquad \textbf{Equation 9}$$

the equilibrium constant (K_{eq}) for the unfolding process is expressed as:

$$\boldsymbol{K_{eq} = f_U/f_N} \qquad \textbf{Equation 10}$$

when equation (8) and (9) are substituted into equation (10), we get:

$$\boldsymbol{K_{eq} = (y_N - y)/(y - y_U)} \qquad \textbf{Equation 11}$$

To calculate Gibbs free energy ΔG°:

$$\boldsymbol{\Delta G^\circ = -RT \ln K_{eq}} \qquad \textbf{Equation 12}$$

where ΔG° is the free energy of unfolding, R is the universal gas constant, T is temperature in kelvin and K_{eq} is the equilibrium constant. In order to determine $\Delta G(H_2O)$, it is assumed that ΔG° has a linear dependence on denaturant concentration [D] for all urea concentrations (Tanford, 1968, 1970). Thus;

$$\boldsymbol{\Delta G^\circ = \Delta G(H_2O) - m[D]} \qquad \textbf{Equation 13}$$

where *m* (*m*-value) is determined from the slope of Δ*G* versus denaturant concentration plot and is indicative of the difference of the solvent accessible area between folded and unfolded state of protein (Greene and Pace, 1974)

Combining equations 11, 12 and 13 thus gives:

$$\mathbf{y} = \left[(\mathbf{y_N} + \mathbf{y_U})\left(\mathbf{e}^{-\left(\frac{\Delta G(\mathbf{H_2O}) - m[\mathbf{D}]}{\mathbf{RT}}\right)}\right)\right] \Big/ \left[\mathbf{1} + \mathbf{e}^{-\left(\frac{\Delta G(\mathbf{H_2O}) - m[\mathbf{D}]}{\mathbf{RT}}\right)}\right] \qquad \textbf{Equation 14}$$

The thermodynamic parameters (Δ*G*(H_2O) and *m*-value) were obtained by fitting the equilibrium unfolding data to equation 14 using SigmaPlot version 11.0 (Systat Software Inc; Chicago, IL, USA).

2.10.4.2. Three-state fit

For a three-state unfolding transition, a protein may exist in the native (N), intermediate (I) and unfolded (U) states:

$$\boldsymbol{N_2} \overset{\mathbf{K_1}}{\leftrightarrow} \boldsymbol{I_2} \overset{\mathbf{K_2}}{\leftrightarrow} \mathbf{2U} \qquad \textbf{Equation 15}$$

$$\mathbf{f_N} + \mathbf{f_I} + \mathbf{f_U} = \mathbf{1} \qquad \textbf{Equation 16}$$

where K_1 and K_2 represents the equilibrium constant for the $N \rightleftarrows I_2$ transition and $I_2 \rightleftarrows U$ transition, respectively. The native (f_N), unfolded (f_U) and intermediate (f_I) conformations are the only three species present in the three-state unfolding transitions. The two equilibrium constants can calculated as follows:

$$\mathbf{K_1} = \mathbf{f_I}/\mathbf{f_N} \qquad \textbf{Equation 17}$$

$$\mathbf{K_2} = (\mathbf{f_U})^2/\mathbf{f_I} \qquad \textbf{Equation 18}$$

Thus, the equilibrium constant (K_U) for $N_2 \rightleftarrows 2U$ is

$$\mathbf{K_U} = (\mathbf{f_U})^2/\mathbf{f_N} = \mathbf{K_1} \cdot \mathbf{K_2} \qquad \textbf{Equation 19}$$

where K_1K_2 represents the equilibrium constant for N ⇄ U. The signal (y) obtained for the respective spectroscopic probe is represented as:

$$\mathbf{y = y_N . f_N + y_U . f_U + y_I . f_I} \qquad \textbf{Equation 20}$$

where y_I is signal given by the intermediate. y_N and y_U are estimated from the linear extrapolation of the pre- and post-transition baselines, respectively. Now, by solving f_U in terms of K_1 and K_2 by combining equations (18) and (19)

$$\mathbf{f_U = (K_1 . K_2)/(K_1 . K_2 + 1 + K_1)} \qquad \textbf{Equation 21}$$

Rearranging equation (19) and solving for f_N and f_I in terms of K_1 and K_2:

$$\mathbf{f_N = f_U/(K_1 . K_2)} \qquad \textbf{Equation 22}$$

therefore,

$$\mathbf{f_N = 1/[(K_1 . K_2 + 1 + K_1)/(K_1 . K_2)]} \qquad \textbf{Equation 23}$$

Similarly

$$\mathbf{f_I = K_1/[(K_1 . K_2 + 1 + K_1)/(K_1 . K_2)]} \qquad \textbf{Equation 24}$$

Substituting equations (22), (23) and (24) into equation (20) gives:

$$\mathbf{y = \left(y_N + y_I . K_1 + y_U (K_1 . K_2)\right)/\left(1 + K_1 + (K_1 . K_2)\right)} \qquad \textbf{Equation 25}$$

We know that:

$$\mathbf{\Delta G_1^\circ = -RT \ln K_1} \qquad \textbf{Equation 26}$$

Therefore, rearranging equation 26

$$\mathbf{K_1 = e^{-(\Delta G_1^\circ / RT)}} \qquad \textbf{Equation 27}$$

And

$$\mathbf{\Delta G_1^\circ = \Delta G_1 (H_2O) - m_1 [D]} \qquad \textbf{Equation 28}$$

By combining equations (26) and (27):

$$\mathbf{K_1 = e^{-([\Delta G_1(H_2O) - m_1[D]]/[RT])}} \qquad \textbf{Equation 29}$$

Similarly,

$$\mathbf{K_2 = e^{-([\Delta G_2(H_2O) - m_2[D]]/[RT])}} \qquad \textbf{Equation 30}$$

Substituting equations (29) and (30) into (25) the final equation to fit the data is achieved:

$$\mathbf{y = \frac{\left(y_N + y_I\left[e^{-\left([\Delta G_1(H_2O) - m_1[D]]/[RT]\right)}\right] + y_U\left[e^{-\left([\Delta G_1(H_2O) - m_1[D]]/[RT]\right)}\right]\left[e^{-\left([\Delta G_2(H_2O) - m_2[D]]/[RT]\right)}\right]\right)}{\left(1 + \left[e^{-\left([\Delta G_1(H_2O) - m_1[D]]/[RT]\right)}\right] + \left[e^{-\left([\Delta G_1(H_2O) - m_1[D]]/[RT]\right)}\right]\left[e^{-\left([\Delta G_2(H_2O) - m_2[D]]/[RT]\right)}\right]\right)}} \qquad \textbf{Equation 31}$$

3. Results

3.1. Multiple structural alignment

The sequence identity between proteins in the GST family is very low and yet the overall structure is very similar. This suggests that conserved residues play a key role in the structure, stability and function of this protein family. To identify conserved residues in the Trx-like domain of GSTs, a structural-based alignment of the GST superfamily was performed using hGSTA1-1 as a model. Representatives of the 12 classes in the GST family as well as Grx2, Ure2p and CLIC1 were used in the structural alignment. Figure 7 shows two topologically conserved leucine residues in the N-subdomain of the GST family. The first leucine is at position 7 (Leu7) in β1 and 76% of the sequences aligned has a leucine residue at this position. In the Delta, Omega and Phi class GSTs as well as Grx2, this position is occupied by a similar residue with a similar hydrophobic index (Table 2). The second leucine is at position 23 (Leu23) in α1 with 70% of the sequences aligned containing leucine at this position. In the *Plasmodium falciparum* GST (*pf*GST), and Grx2 this position is occupied by phenylalanine whereas in eEF1 the leucine is replaced by valine. Both phenylalanine and valine have a similar hydrophobic index to leucine; hence, they can be considered as conservative replacements. The Delta and Theta classes have an alanine at an equivalent position to Leu23. Even though both leucine and alanine are hydrophobic residues, there is a difference in the bulk of their side chains as well as in their hydrophobic index. This difference could have an impact in packing of the N-subdomain.

To seek reason/s for conservation of Leu7 and Leu23 in the N-subdomain of GST as well their possible role in the structure and function of GSTs, the local environment of these residues was analysed. Both Leu7 and Leu23 are at the centre of the hydrophobic pocket formed by the hydrophobic core of N-subdomain (Figure 8 and 9). The hydrophobic core of Trx-like domain in GST is made up of β1 and α1 from βαβ motif, as well as β3, β4 and α3 from ββα-motif of the thioredoxin fold (Le Trong *et al.*, 2002). The Leu7 side chain is involved in the packing of α1β1α2 and α3 and forming van der Waals contacts with side chains of with Met16 (α1), Phe30 (loop connecting α1 and β2), Glu32 (β2), and Val58 (β3).

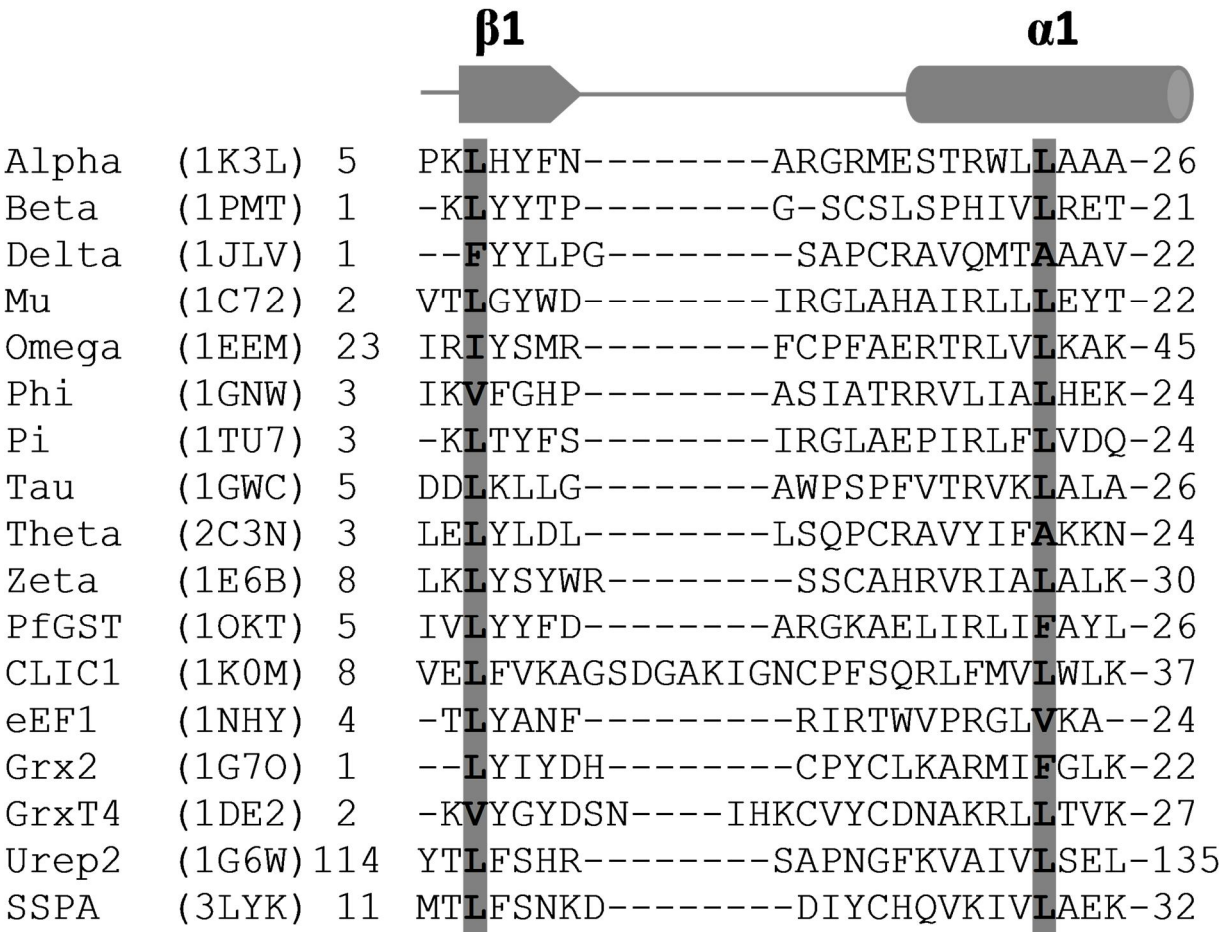


Figure 7: Structure-based sequence alignment of the proteins in the GST family corresponding to βαβ motif of Trx-like fold

The residues in bold indicates the topologically conserved leucine at position 7 (β1) and 23 (α1). Alpha to Zeta represent different gene classes of dimeric GSTs; CLIC1 (chloride intracellular channel 1) and Grx2 (glutaredoxin 2) are monomeric GST homologues. GrxT4 (bacteriophage T4 glutaredoxin), Ure2p (a prion protein in yeast), eEF1 (eukaryotic elongation factor) and SSPA (Stringent starvation protein A) are dimeric GST homologues. The PDB codes are shown in parentheses. The alignment was performed with 3DCoffee (O'Sullivan *et al.*, 2004).

Table 2: Hydrophobicities of the 20 naturally occurring amino acids

The hydrophobicities are based on the solvent transfer free energies from octanol to water (Manavalan and Ponnuswamy, 1978).

Amino acid residue	Hydrophobicity (kcal mol^{-1})
Tryptophan	2.25
Phenylalanine	1.79
Isoleucine	1.80
Leucine	1.70
Cysteine	1.54
Methionine	1.23
Valine	1.22
Tyrosine	0.96
Proline	0.72
Alanine	0.31
Threonine	0.26
Glycine	0.00
Serine	0.04
Histidine	0.13
Glutamine	0.22
Asparagine	0.60
Glutamic acid	0.64
Aspartic acid	0.77
Lysine	0.99
Arginine	1.01

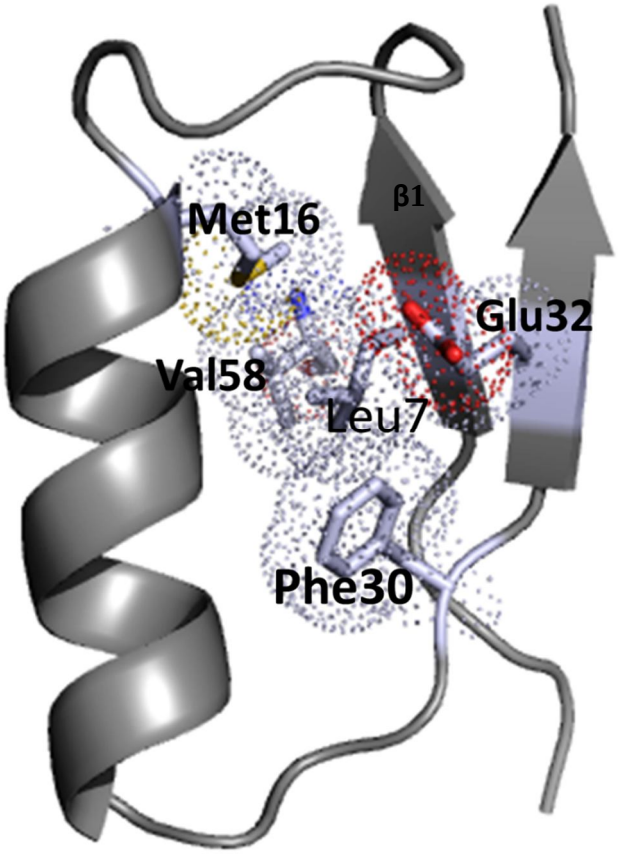


Figure 8: The environment of Leu7 in hGSTA1-1

The βαβ motif of hGSTA1-1 is shown in ribbon form. The amino acid residues that are within 4 Å of Leu7 which makes van der Waals contacts with the Leu7 side chain are represented as sticks. The dots represent the van der Waals radii space filling model. This image was generated using PyMol v0.99 (DeLano Scientific, 2006).

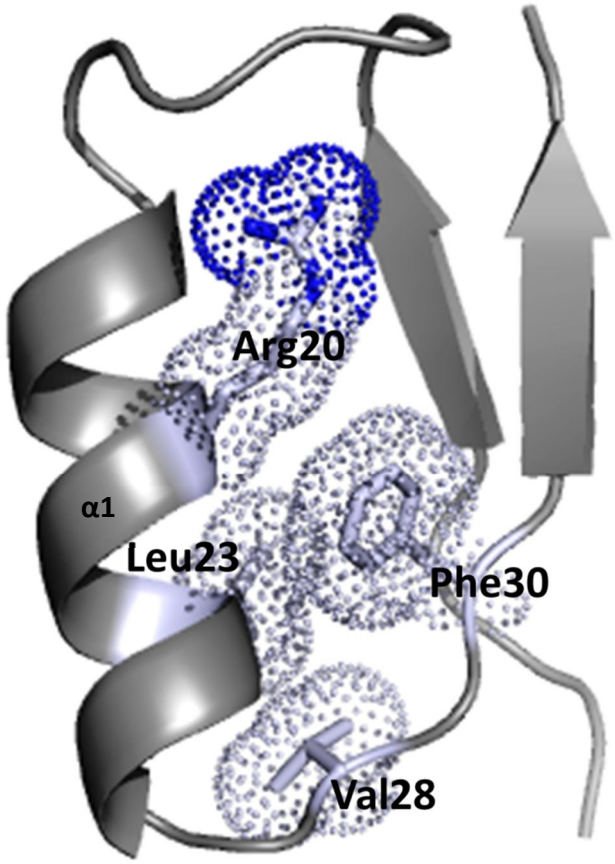


Figure 9: The βαβ of hGSTA1-1 depicting the environment of Leu23

The βαβ motif is shown in ribbon whereas the amino acid residues that make van der Waals contact with Leu23 side chain are represented as sticks. The dots represent the van der Waals radii space filling model. This image was generated using PyMol v0.99 (DeLano Scientific, 2006).

The Leu23 side chain is mainly involved in the van der Waals interactions with Val28 and Phe30 (both residues from the loop region connecting α2 and β2) (Figure 9). The interaction of Leu23 with these residues containing bulky side chains could limit the mobility of this loop region to optimise packing of this domain. Leu7 and Leu23 were then mutated to an alanine residue to study the contribution of the bulky side chain of leucine to the structure, function and stability of the N-subdomain of GST. As mentioned before, the pKHA1 vectors encoding the sequence for L7A and L23A hGSTA1-1 were already prepared by Dr. S. Gildenhuys.

3.2. Over expression and purification of wild-type, L7A and L23A hGSTA1-1

This study is a comparison between the wild-type and its variants (L7A and L23A); it was therefore, necessary to initially confirm whether the plasmids contained inserts with the correct nucleotide sequence. DNA sequencing results verified that the obtained plasmids contained the correct inserts (Figure 10). The nucleotide sequence for the wild-type was submitted for a BLAST search ((http://www.ncbi.nlm.nih.gov/BLAST) and gave a 100% match to a nucleotide sequence of hGSTA1-1. This confirmed that the wild-type nucleotide sequence was correct without any undesirable mutations. Comparison of the wild-type nucleotide sequence with of L7A and L23 showed that the plasmids encoding L7A and L23A contained single mutations at positions 7 and 23 to change leucine to alanine residues, respectively.

The expression of wild-type, L7A and L23A hGSTA1-1 was performed as described in section 2.3.2. Figure 11 shows over-expression of a 26 kDa protein which corresponds to that of the monomeric size of hGSTA1-1. High amounts of wild-type, L7A and L23A were still located in the pellet and this was subsequently prevented by ensuring that the cells were lysed properly during the sonication step (Figure 12). The L7A and L23A mutations did not affect the expression of hGSTA1-1 since expression conditions used for wild-type and mutant proteins were the same. Furthermore, the expression levels of wild-type, L7A and L23A hGSTA1-1 were similar. Purification of wild-type, L7A and L23A hGSTA1-1 was successfully achieved using cation-exchange chromatography. The results obtained during purification steps are shown using L7A hGSTA1-1 as a representative for all proteins. The eluent from CM-Sepharose showed a single symmetric peak (Figure 12B).

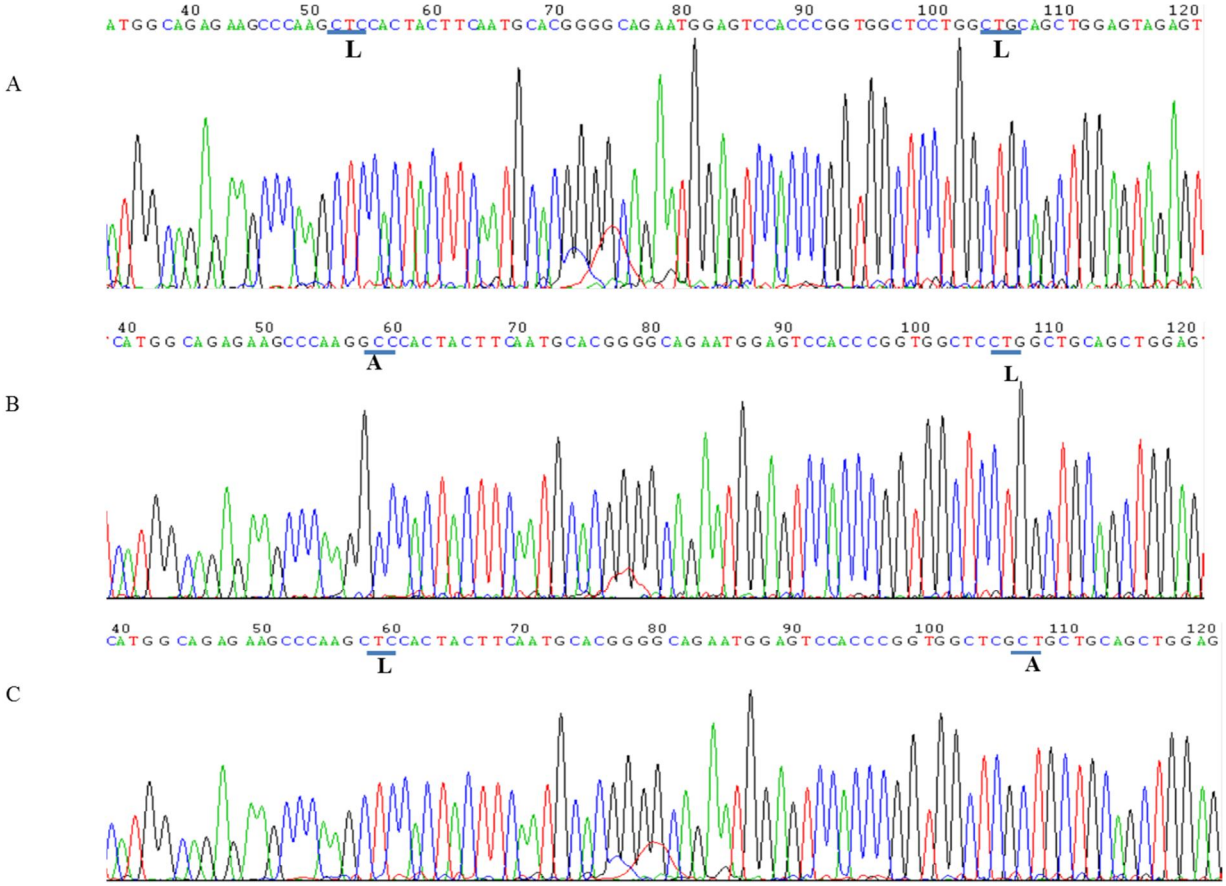


Figure 10: A portion of nucleotide sequencing chromatograms encoding for wild-type, L7A and L23A hGSTA1-1

The nucleotide sequence encoding for wild-type, L7A and L23A hGSTA1-1 is shown in A, B and C, respectively. The mutation of leucine at position 7 and 23 is shown by an alanine codon sequence GCC (B) and GCT (C), respectively. The underlined sequence shows the position of the mutation and letters in bold represent the amino acid code for that sequence. The sequencing results were viewed using Finch TV version 1.4.0 (http://www.geospiza.com/FinchTV Geospiza Inc.).

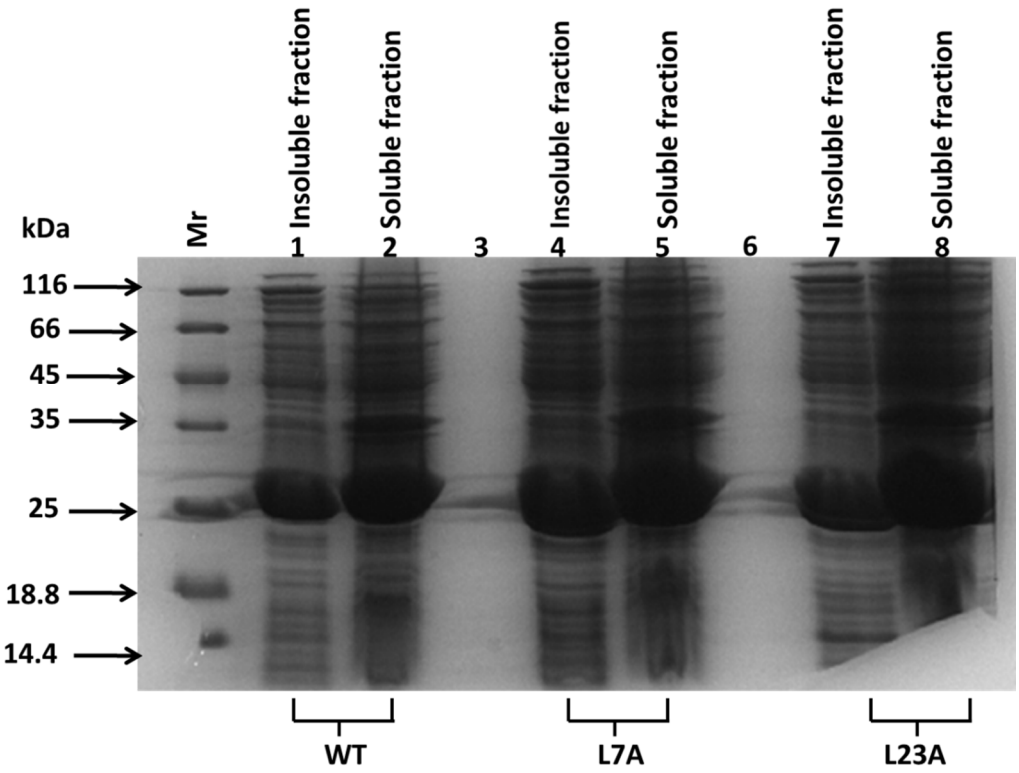


Figure 11: Expression of wild-type, L7A and L23A hGSTA1-1

Lane 3 and Lane 6 were not loaded with any sample. Therefore, proteins shown in these lanes were spilled over from the other lanes during loading.

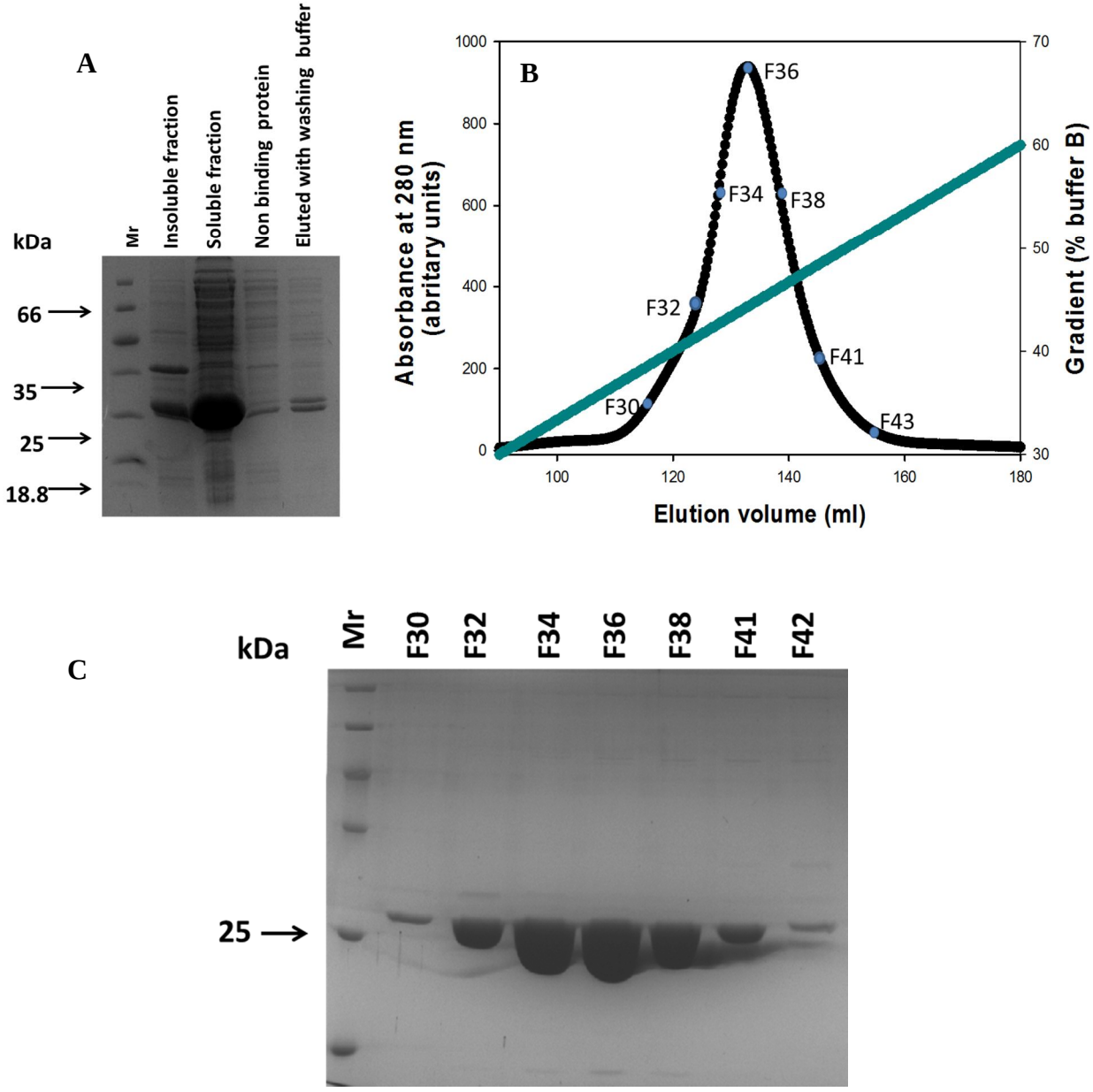


Figure 12: Purification of L7A hGSTA1-1

Purification of L7A protein was performed using a CM-Sepharose cation-exchange column. The proteins that did not bind to the column are shown in A. L7A protein was eluted with a 300 ml linear salt gradient of 0 - 300 mM NaCl prepared in 10 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5 (cyan line on B). L7A eluted as a single peak (B) and selected fractions that make up this peak were analysed by SDS-PAGE and are shown in C. The fractions that were selected for SDS-PAGE analyses are shown by blue symbols on the elution profile on B.

SDS-PAGE analysis of fractions eluted from the CM-Sepharose column shows a single band of proteins indicating successful purification L7A (Figure 12C). The fractions were pooled and concentrated to 5 ml by ultra-filtration with an Amicon membrane that has a molecular weight cut off of 10 kDa. The purity of wild-type, L7A and L23A was assessed by SDS-PAGE (Figure 13). SDS-PAGE analysis shows that wild-type and its variants migrated as a single band on the gel and this band corresponds to an apparent subunit molecular mass of approximately 25 kDa.

SE-HPLC was used to assess the oligomeric state of wild-type, L7A and L23A hGSTA1-1. The results shown in Figure 14 which indicates that all proteins eluted as single peaks confirming the homogeneity of the purified proteins. The retention time of the wild-type and its variant proteins was the same (15 minutes) indicating that the mutations did not affect the hydrodynamic volume of hGSTA1-1.

3.3. Spectroscopic properties of L7A, L23A and wild-type hGSTA1-1 proteins

3.4. Far-UV CD

Far-UV CD was used to (a) determine the secondary structural properties of wild-type, L7A and L23A proteins and (b) probe for any secondary structural changes that may have been induced by the L7A and L23A mutations. The far-UV CD spectral properties of native and chemically denatured wild-type and mutant proteins are shown in Figure 15. The spectra for both wild-type and mutant proteins display a peak at 190 nm and troughs at 208 and 222 nm. These wavelength minima are an indication of a protein with high helical content and which is consistent with the structure of GST proteins (Reinemer *et al.*, 1991; Le Trong *et al.*, 2002; Achilonu *et al.*, 2010). The wild-type and mutant proteins were unfolded in 8 M urea since they lost their secondary structure elements. The spectrum of three proteins was superimposable, indicating that the mutation did not cause any major perturbations on the secondary structure of the mutants compared to the wild-type.

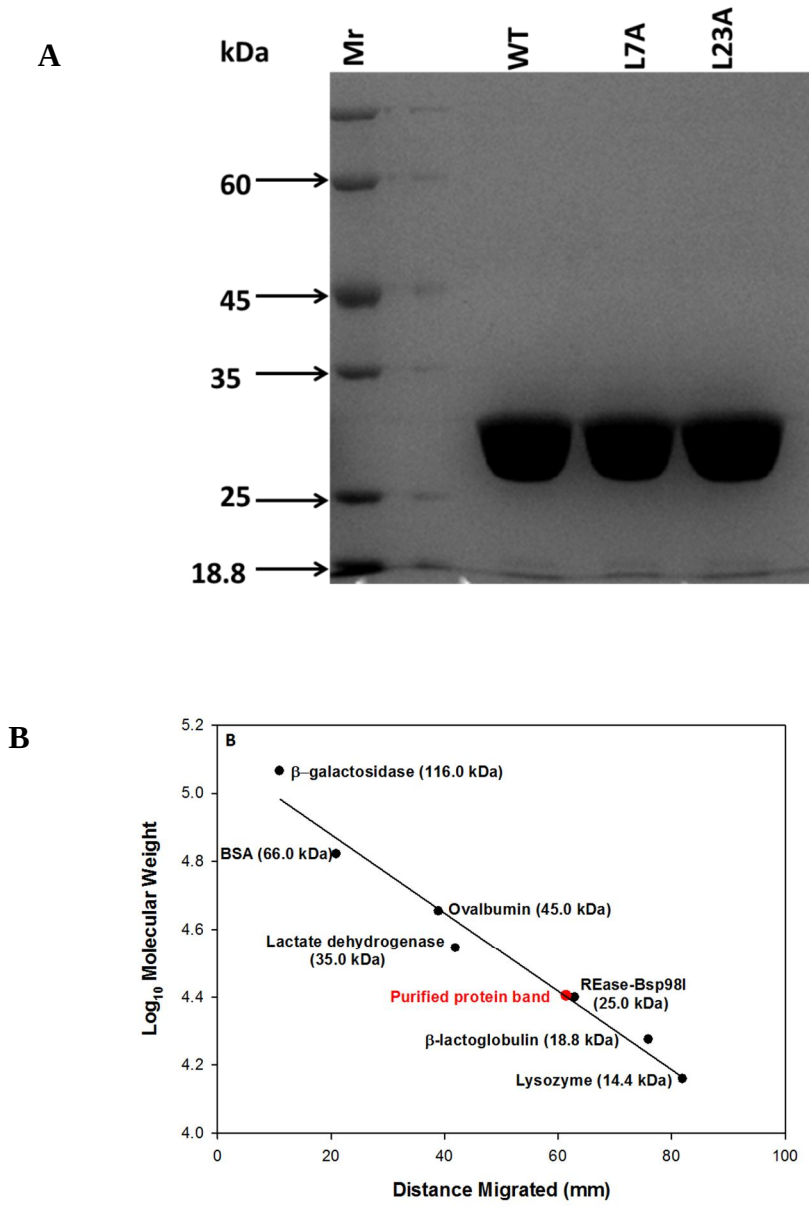


Figure 13: SDS-PAGE analysis of purified wild-type, L7A and L23A hGSTA1-1

The purity of all three proteins was assessed with SDS-PAGE and is shown in A. Lane 1, 2 and 3 represents the purified wild-type, L7A and L23A proteins, respectively. SDS-PAGE molecular weight marker calibration curve is shown in B. The subunit molecular mass for wild-type and its variants is about 26 kDa (represented by the red symbol on the calibration curve).

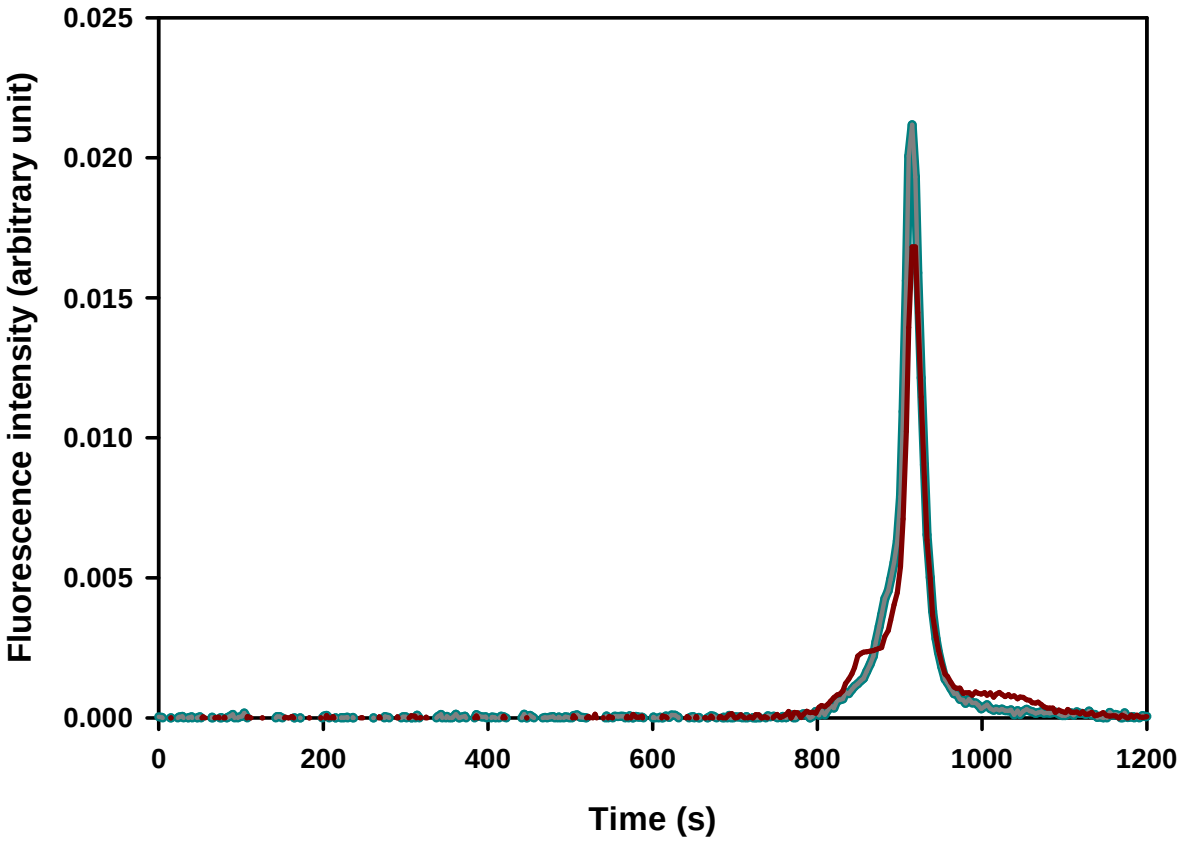


Figure 14: SE-HPLC elution profile of hGSTA1-1

The elution profiles of wild-type, L7A and L23A are shown in cyan, grey and red, respectively. The elution of the proteins was monitored using fluorescence at 280 nm. The column was equilibrated with 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5 at 0.5 ml/min. The protein concentration loaded on the column was 20 µM.

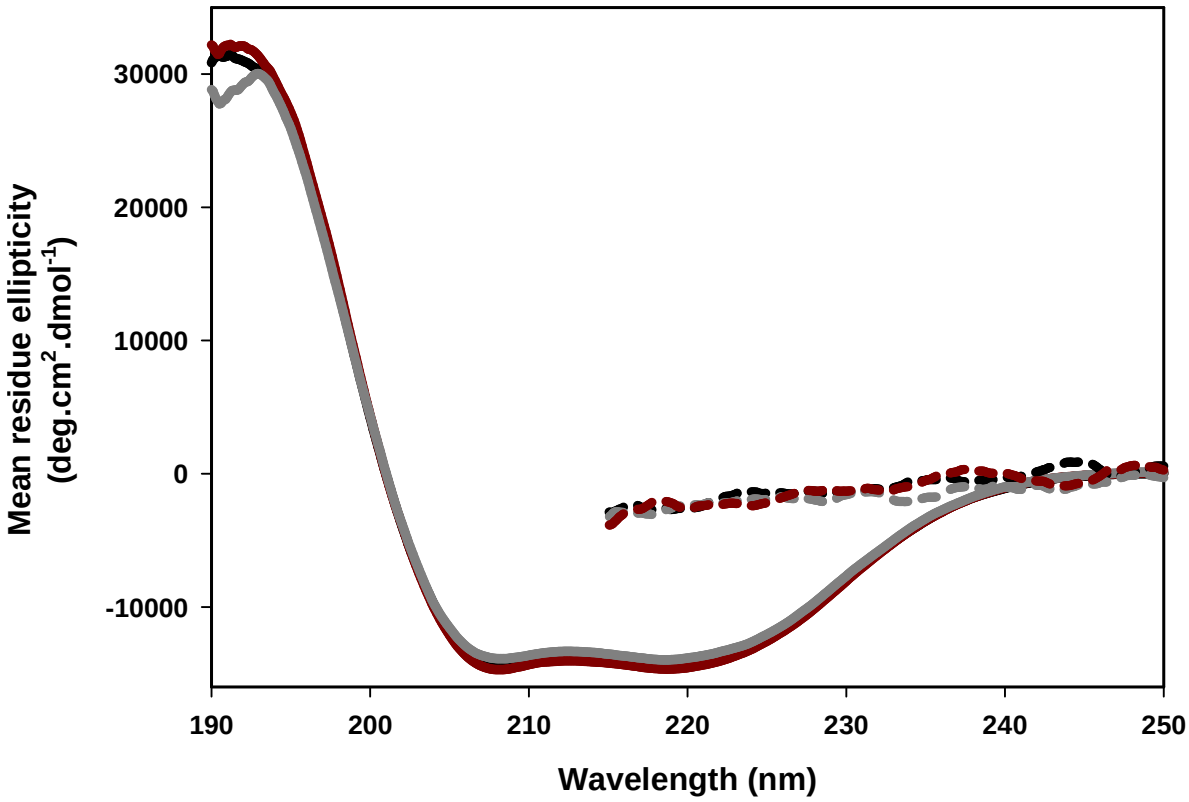


Figure 15: Far-UV CD spectra of wild-type, L7A and L23A hGSTA1-1

The wild-type, L7A and L23A in 2 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5 are shown in black, red and grey lines, respectively. The proteins were unfolded using 8 M urea. The spectra of native and denatured proteins are represented in solid lines and dashed lines, respectively. The protein concentration was 2 µM. The spectrum represents an average of three experiments. The spectrum of the buffer control was subtracted from the experimental spectra.

3.5. Fluorescence spectroscopy

The indole ring side chain of tryptophan residues is sensitive to the polarity of its microenvironment, thus making intrinsic tryptophan fluorescence a useful technique for measuring the local environment of tryptophan residues in a protein (Lakowicz, 1999). The impact of L7A and L23A mutations on their local environment was, therefore, determined using fluorescence as probe.

Each subunit of hGSTA1-1 contains one tryptophan residue (Trp21) and 10 tyrosine residues. Trp21 is located in α1 of the N-subdomain with its indole side chain involved in hydrophobic interactions with residues from α6 and α8 of the C-terminal domain (Wallace *et al.,* 1998b). The intrinsic fluorescence properties L7A and L23A are similar to that of the wild-type, which indicates that the mutation did not cause any major changes to the local environment of Trp21. The fluorescence emission maxima for all three proteins under native conditions is 327 nm (Figure 16) which is consistent with crystallographic studies that showed the indole side chain of Trp21 being solvent inaccessible (Dirr and Reinemer, 1991; Wallace *et al.*, 1998b; Wallace and Dirr, 1999; Wallace *et al.*, 2000; Le Trong *et al.*, 2002). The unfolding of the wild-type and its variants results in the reduction of fluorescence intensity and has a red-shifted emission maximum at 347 nm (Figure 16).

3.6. Crystal structure of L7A and L23A proteins

3.6.1. Crystallisation of L7A and L23A mutants

The L7A and L23A crystals were grown in a range of 18 to 23% (w/v) of PEG 4000 in 0.1 M Tris–HCl pH 7.5, 10 mM DTT and 0.02% sodium azide at 20 °C. The optimal crystal growth conditions for both the mutants was 20% (w/v) PEG 4000 in the above buffer. The crystals appeared overnight and were allowed to grow for 24 hours before being harvested. The hGSTA1-1 crystals appeared as long needles as shown in Figure 17.

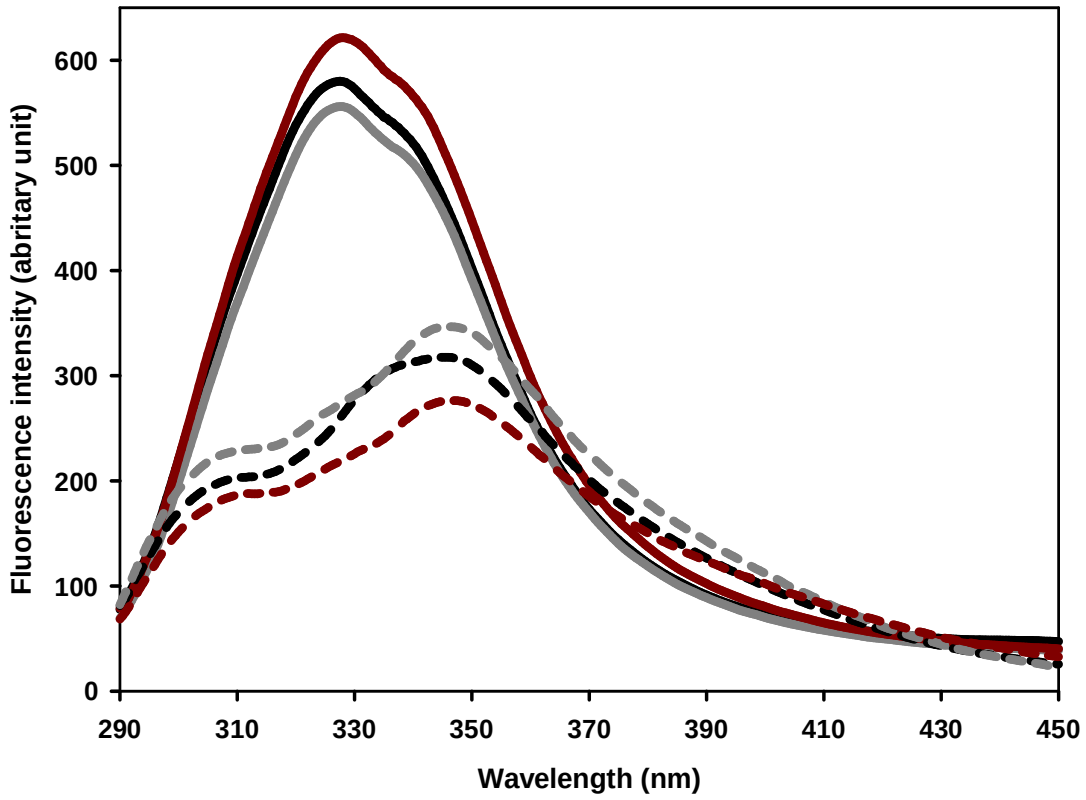
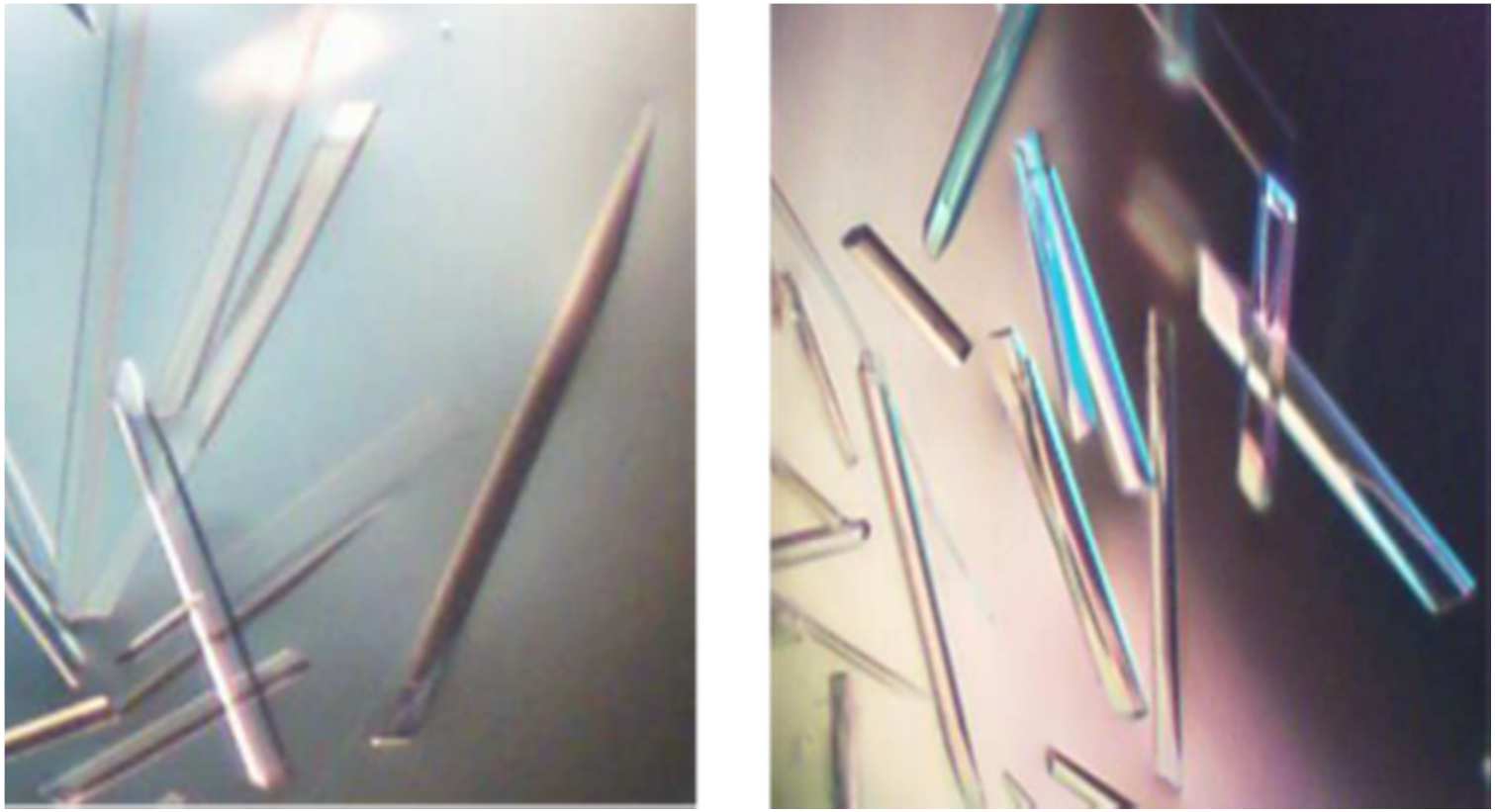


Figure 16: Fluorescence emission spectra for the wild-type, L7A and L23A hGSTA1-1

The wild-type and mutant proteins (2 μM) were excited at 280 nm in 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5 with 0 M or 8 M urea. The wild-type, L7A and L23A proteins are shown in black, red and grey, respectively. The spectra of native and denatured proteins are shown in solid and dashed lines, respectively. The spectrum represents an average of three independent experiments with standard deviation of ±28 nm, ±23nm and ±32 for the wild-type, L7A and L23A proteins at 346 nm, respectively. The spectrum of the buffer control was subtracted from the experimental spectra.



A **B**

Figure 17: L7A and L23A hGSTA1-1 crystals

The L7A (A) and L23A (B) crystals were grown by the hanging drop vapour diffusion method at 20 °C in 20% (w/v) of PEG 4000 in 0.1 M Tris–HCl, pH 7.5, 10 mM DTT and 0.02% sodium azide. The crystals were allowed to grow for 24 hours before being harvested.

3.6.2. Refinement of L7A and L23A models

In order to assess the effects of L7A and L23A mutation on the overall structure of hGSTA1-1-1, the mutants were crystalised in the apo form. The structures of L7A and L23A were determined at 1.79 and 2.2 Å, respectively. The electron densities for residue 2 – 209 (L7A) and 2 – 208 (L23A) for subunit A and (4-210) for subunit B in both mutants were clearly defined. The electron density map at position 7 and 23 in both subunits is consistent with that of alanine in the L7A and L23A structures, respectively (Figure 18). The Matthews coefficient values for the L7A and L23A were 2.29 and 2.37 Da^{-1}, respectively, which indicates that both crystals contained two molecules in the asymmetric unit (Matthews, 1968). The final *R* value and final R_{free} values (Table 3) obtained for both mutant structures gives an acceptable difference of 0.07 (Brunger, 1992, 1993, 1996).

The *R* value is used to determine the quality of a fit model to the diffraction data. It measures the discrepancy between the observed and calculated structure factors. The difference between the *R* value and R_{free} value should be small and it depends on the following factors: (a) quality of the data as measured by R_{sym} completeness, (b) global correctness of the model data, (c) completeness of the model and (d) the degree of over-fitting (Brunger, 1992, 1993, 1996). If any of above factors is incorrect, the refinement of the model will result a discrepancy of *R* and R_{free} values.

The structural alignment of the wild-type and mutants structures using C-alpha gives root mean square deviation (r.m.s.d) values of 0.63 Å (L7A) and 0.67 Å (L23A) between the wild-type and mutant global structures (Figure 19). The r.m.s.d value is used to measure structural similarity between proteins (Maiti *et al.*, 2004). The r.m.s.d values between the wild-type and mutant structures demonstrate that the mutation of Leu7 and Leu23 into alanine did not alter the backbone structure of the mutants (Figure 19). Gln67 is the only non-glycine outlier on Ramachandran analysis and is common to both structures (Figure 20). This residue is located in the loop region connecting β4 and α3 and has been reported as an outlier in other structures of GST (Le Trong *et al.*, 2002; Achilonu *et al.*, 2010). The Gln67 outlier is most likely to be a genuine feature of the crystal structure rather than an error in the model.

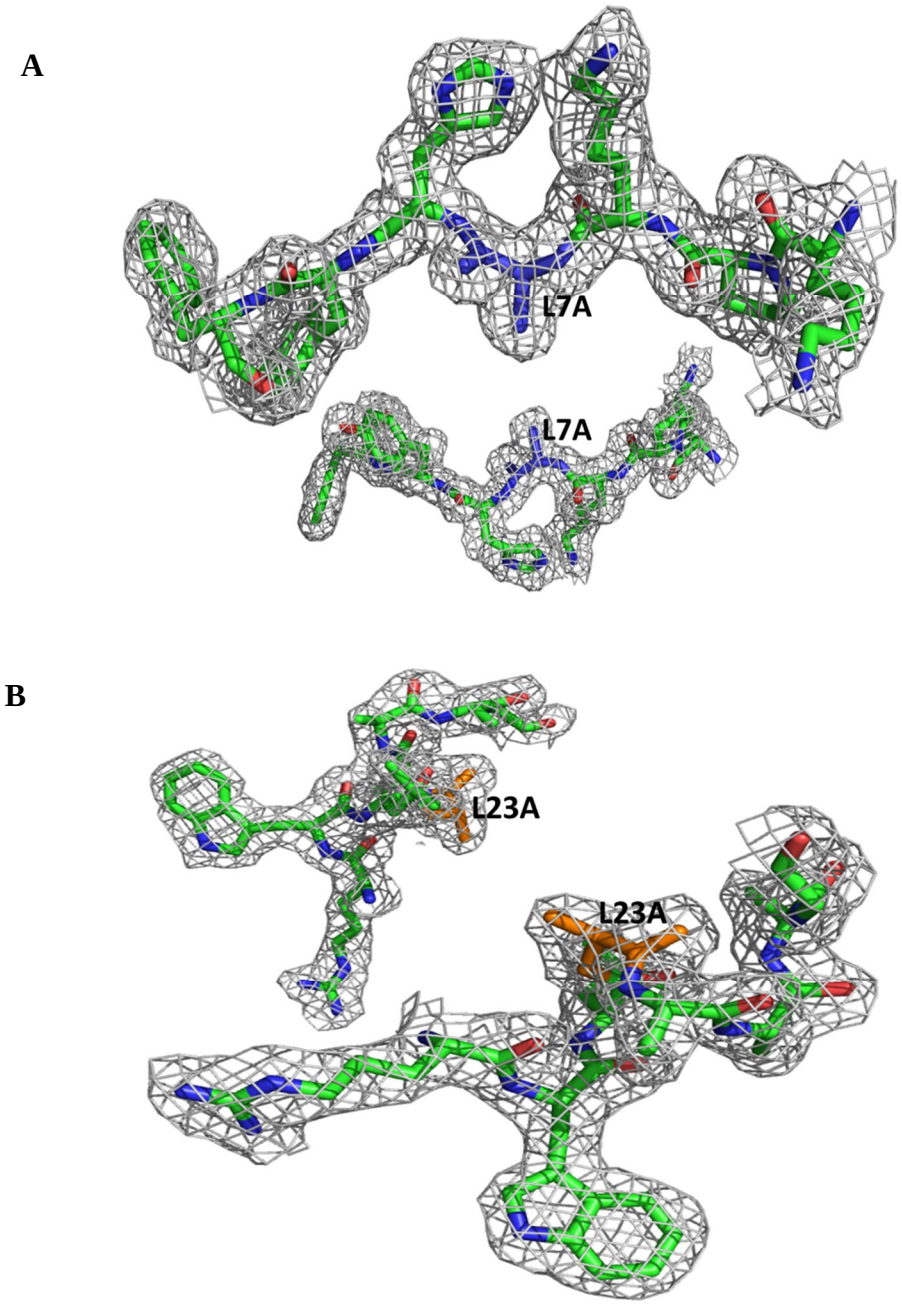


Figure 18: Fit of the L7A and L23A into the $2F_0$-F_c electron density map

Figure A (L7A structure) and B (L23A structure) shows electron density maps at the point of mutation. The electron density map at position 7 (A) and 23 (B) in both subunits is in agreement with that of alanine in the L7A and L23A structures, respectively. The mutated residues are shown in blue and orange in A and B, respectively. This fit confirms a successful mutation of Leu7 and Leu23 to alanine. This image was generated using PyMol v0.99 (DeLano Scientific, 2006).

Table 3: Crystallographic X-ray data collection and refinement statistics L7A and L23A hGSTA1-1

Parameter	**L7A**	**L23A**
Buffer/salt	0.1 M Tris HCl, pH 7.5	0.1 M Tris HCl, pH 7.5
PEG4000 (% w/v)	20%	20%
PDB ID	3Q74	3U6V
Data Collection:	CCD	CCD
Wavelength (Å)	1.5418	1.5418
Space group	*C*2 (*C*121)	*C*2(*C*121)
Unit cell dimensions (Å)	a =99.35	a=90.00
	b =91.32	b=92.53
	c =51.37	c=90.00
Asymmetric unit content	Dimer	Dimer
Resolution range (Å)	51.30 – 1.79	68.6 – 2.2
No. of reflection(obs)	169075	171031
No. of unique reflections	42718	24385
Completeness (%)	93.27	99.6
R_{sym} (%)	12 (64.7)	31.7 (91.6)
I/σ (I)	1.45/12.85	1.69/13.14
Refinement:		
Resolution (Å)	51.3 – 1.79	51.5 – 2.20
R_{cryst}, R_{free}	0.203 (0.271)	0.217 (0.288)
Reflections	42075	24378
Protein atoms	3352	3339
Solvent molecules	509	249
Rmsd bond lengths (Å)	0.024	0.018
Rmsd angles (deg)	1.9021	1.704
Rmsd ΔB (mm/ss)	1.011/3.014	0.795/2.471
$\langle B \rangle$ protein (Å^2)	18.02	30.20
Ramachandran plot:		
Most favoured (%)	97.1	96.8
Generously allowed (%)	2.4	2.5
Disallowed (%)	0.5	0.7

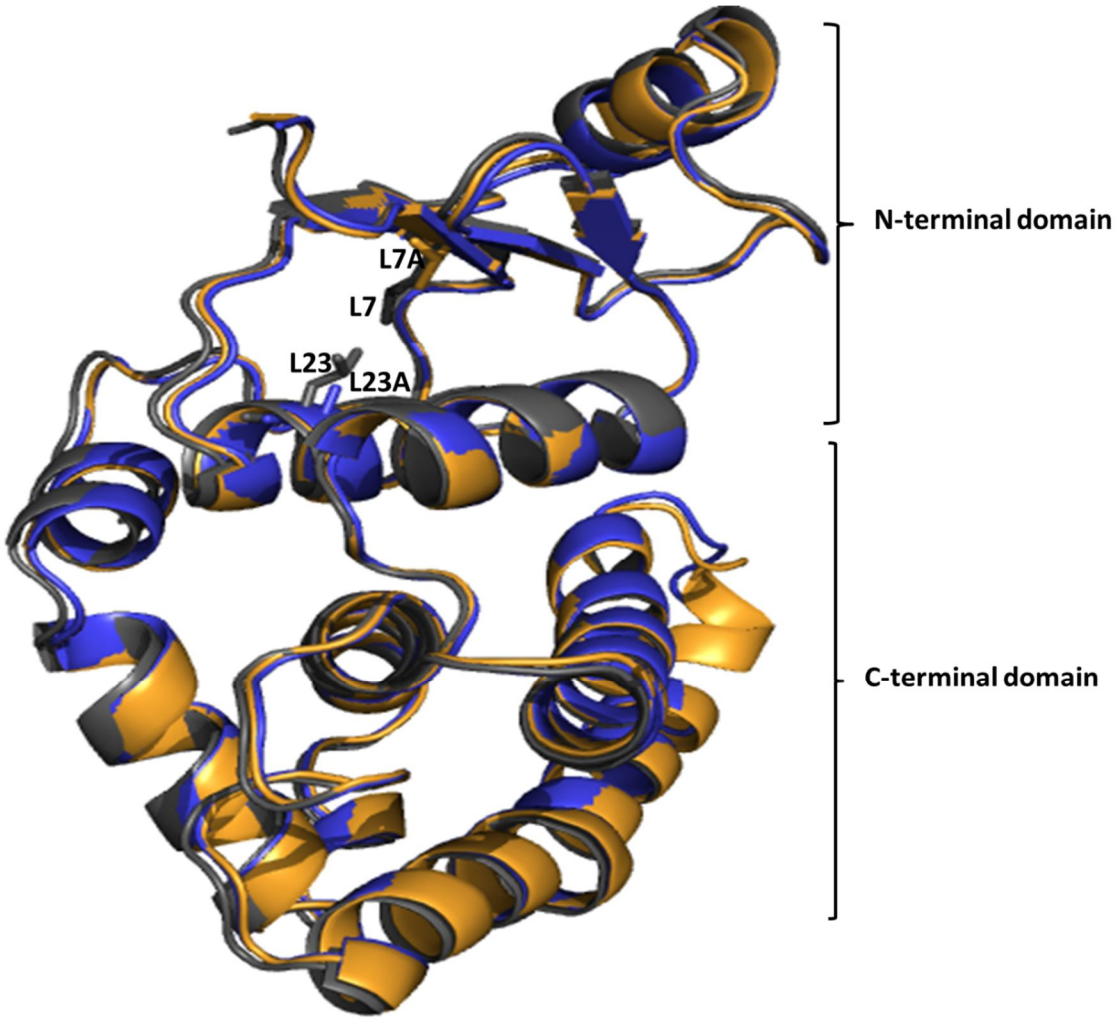


Figure 19: Structural alignment of wild-type and its variants

Structural alignment of wild-type (grey) with L7A (yellow) and L23A (blue) was performed using Swiss PDB Viewer to determine r.m.s.d values (Guex and Peitsch, 1997). The r.m.s.d values of 0.63 Å (L7A) and 0.67 Å (L23A) between wild-type and mutant structures were obtained. The figure was generated using PyMol v0.99 (DeLano Scientific, 2006).

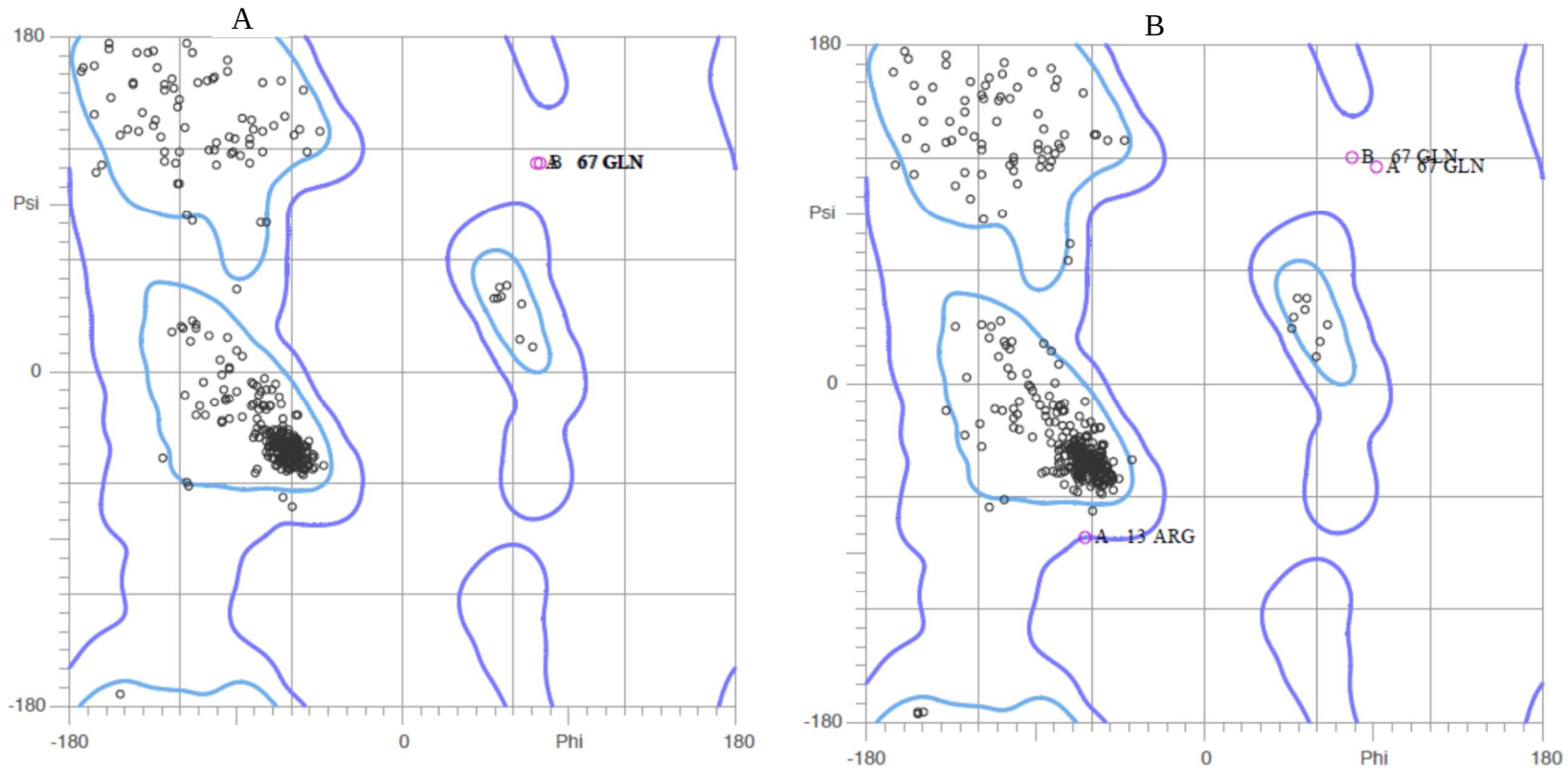


Figure 20: Ramachandran plots of L7A and L23A structural models

The data in allowed regions consisted of 99.5 and 99.3% for L7A and L23A, respectively. Gln67 was a common outlier to both L7A and L23 structures. This image was obtained from MolProbity Ramachandran analysis using PBD codes 3Q74 (L7A) and 3U6V (L23A) (Lovell *et al.*, 2003).

3.6.3. Effects of L7A mutation on hGSTA1-1 structure

Leu7 is located in β1 of the N-subdomain. Its bulky side chain is buried within the hydrophobic core of the Trx-like fold. A structural alignment of the βαβ motif of wild-type and L7A was performed to determine whether any structural changes that occurred within the local environment of L7A mutation site. The alignment gave an r.m.s.d. value of 0.57 Å indicating no significant changes within the mutation site of the L7A structure. The Leu7 side chain makes van der Waals contact with Met16 (α1), Phe30 (loop connecting α1 and β2), Glu32 (β2), and Val58 (β3). These interactions are maintained in the L7A mutant structure, with the exception of Phe30 and Val58 (Figure 21A). The L7A structure also shows that the L7A mutation resulted in the formation of a cavity (58 $Å^3$) within the local environment of the mutation (Figure 21B).

Phe30 in the L7A mutant also adopts a different conformation when compared to that of the wild-type, resulting in the phenyl ring being drawn into the site vacated by the bulky side chain of leucine (Figure 22). The slight change in the orientation of Phe30 results in two new H-bonds being formed with backbone of Phe30 and two H_2O molecules. The loss of van der Waals contacts between L7A and Val58 does not seem to affect the Val58 conformation since it maintains all its interaction as in the wild-type.

3.6.4. Effects of L23A mutation on hGSTA1-1 structure

Like Leu7, Leu23 forms part of the hydrophobic core of N-subdomain. Leu23 is located on α1 with its side chain buried within the hydrophobic core. Figure 23A shows the amino acid residues that make hydrophobic interactions with Leu23 and these interactions are maintained in the L23A structure. This is consistent with a r.m.s.d. value of 0.71 Å obtained when the βαβ motif of the wild-type and L23A are aligned. Examination of the L23A local environment shows that the removal of the bulky side chain of leucine resulted in the formation of a 40 $Å^3$ cavity (Figure 23B).

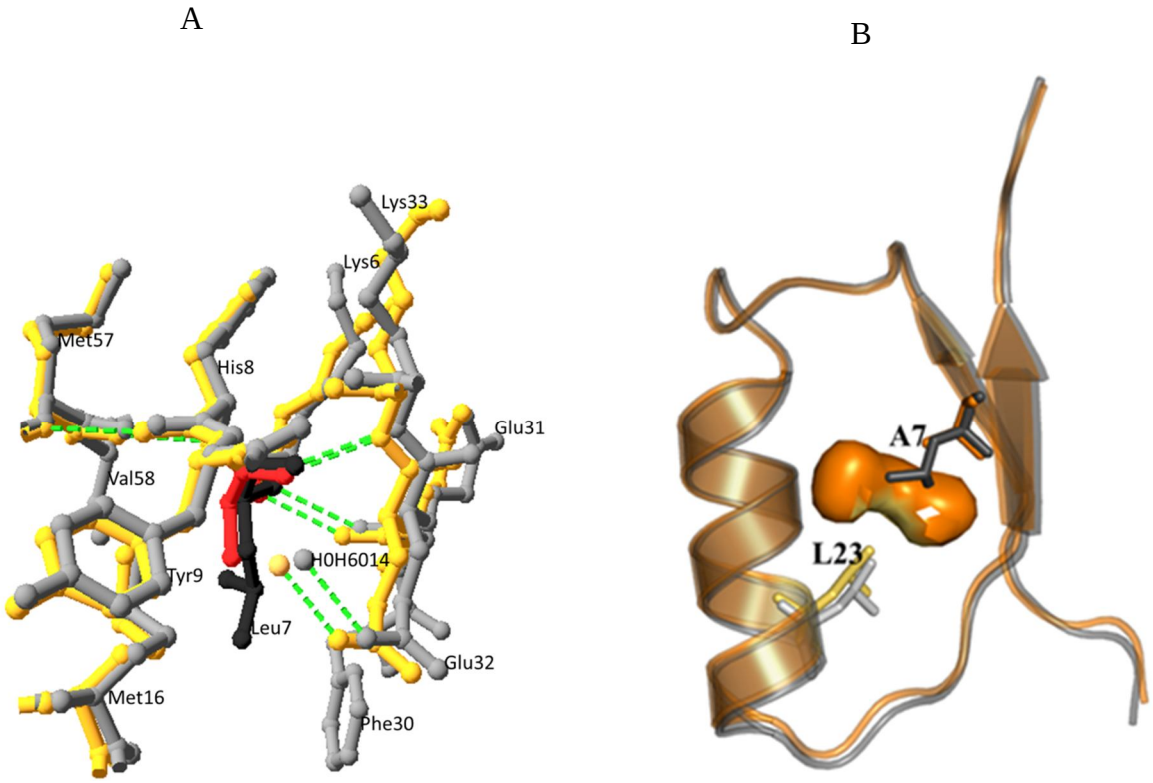


Figure 21: Effect of L7A mutation on the Trx-like domain of hGSTA1-1

The local environments of Leu7 (wild-type) and L7A (mutant) were superimposed to determine if the L7A mutation induced any structural changes. Shown in A are the residues that are within 4 Å from Leu7 and Ala7 residues. These residues are represented in ball-and-stick format. The wild-type is shown in grey with Leu7 coloured in black whereas the mutant is shown in yellow with the Ala7 residue coloured in red. The water molecules located in this region are shown as spheres. Shown in B is the cavity (represented as an orange bulb) created as a result of L7A mutation. The βαβ motif of wild-type (grey) and L7A (yellow) is represented as ribbons. The amino acids, Leu7 and Leu23, from the wild-type (grey) and Ala7 and Leu23 from L7A (yellow) are shown as sticks. The images were generated with PyMOL (DeLano Scientific, 2006).

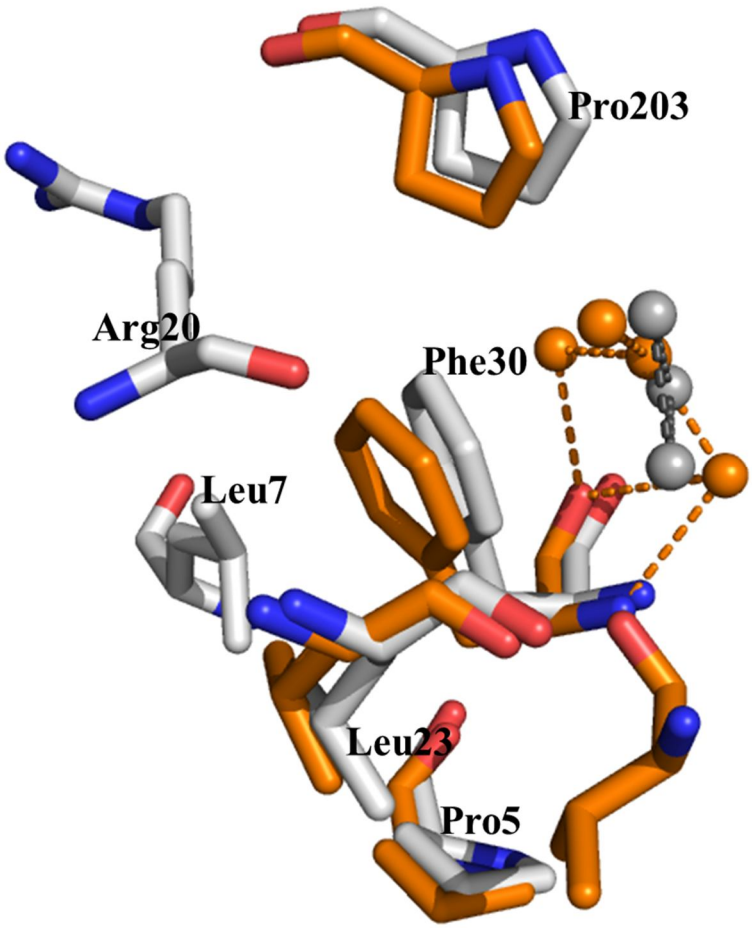


Figure 22: Phe30 side chain adopts a new conformation as a result of the L7A mutation

The rearrangement of the Phe30 side chain in response to L7A mutation is shown above. The amino acid residues that are within 4 Å of Phe30 and which make van der Waals contacts with Phe30 side chain are represented as sticks. The amino acid residues are numbered according to the wild-type and are coloured in grey (for the wild-type) and orange (for the L7A). The surface water molecules that interact with Phe30 (L7A) via hydrogen bonding (represented in dashed orange lines) are shown by orange spheres. The grey spheres represent surface water molecules for the wild-type. The image was generated with PyMOL (DeLano Scientific, 2006).

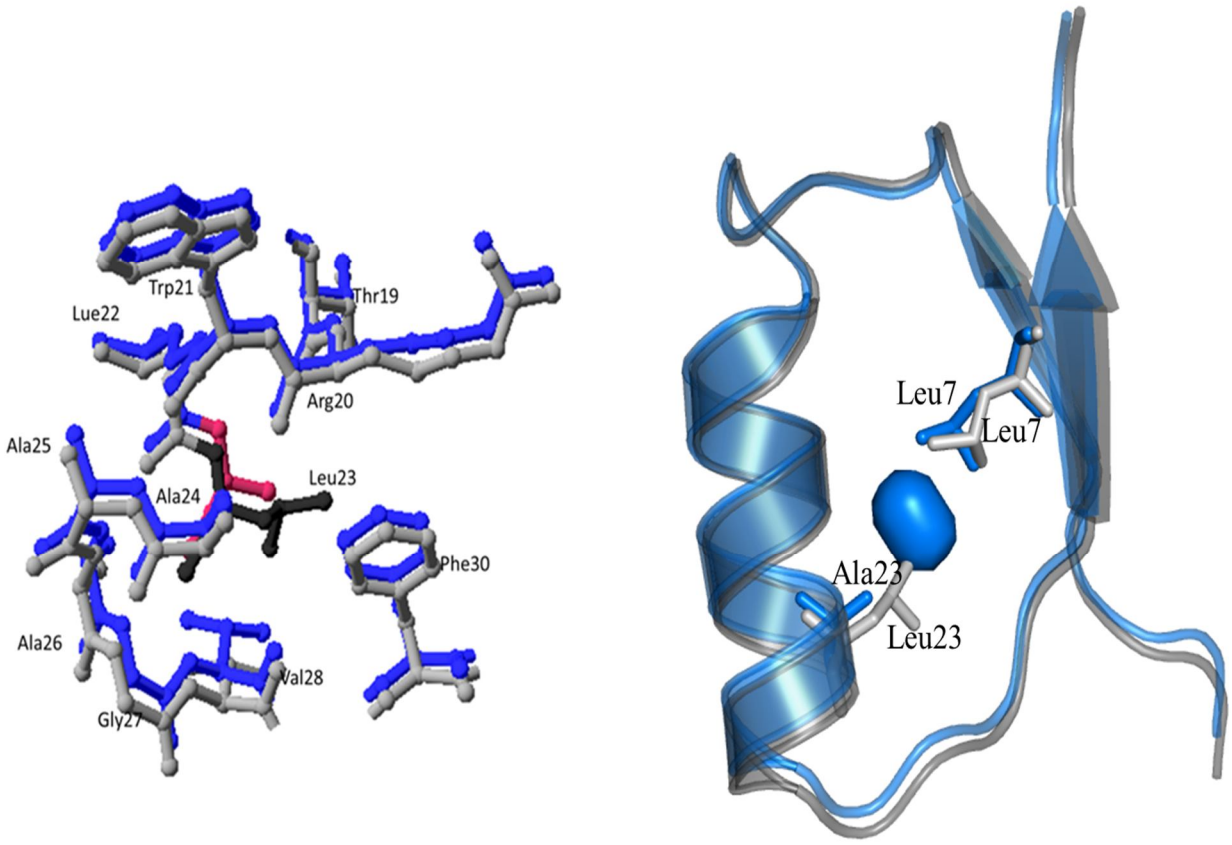


Figure 23: Effect of L23A mutation on the local environment of Leu23

Structural alignment of the wild-type (grey) and L23A (blue) to determine if L23A introduced any changes in the βαβ motif. The amino acid residues that are within 4 Å from Leu23 (wild-type) and Ala23 (mutant) are shown as ball-and–sticks format in A. In B, the βαβ motif is represented as ribbons and the cavity created as result of L23A mutation within motif is represented by a blue blob. The amino acids Leu7 and Leu23 from the wild-type (grey) and Leu7 and Ala23 from the L23A (blue) mutant are shown as sticks. The Leu7 side chain in the mutant structure (blue) adopts a different orientation to that seen in the wild-type structure. The image was generated with PyMOL (DeLano Scientific, 2006).

3.7. Functional properties of wild-type, L7A and L23A hGSTA1-1 proteins

GSTs catalyse the nucleophilic conjugation of glutathione to a variety of electrophilic compounds (Habig and Jakoby, 1981; Sheehan *et al.*, 2001). The G-site is part of the thioredoxin-like domain and binds the thiol substrate, GSH. The binding of the electrophilic substrate occurs at the H-site which is located in the C-terminal domain. Molecular docking experiments have shown that ANS binds GSTs with the hydrophobic anilino and naphthyl rings occupying the H-site and the sulfonate moiety located at the interface of the G- and H-sites (Dirr *et al.*, 2005).

Crystal structures of GSTs bound to the ligand such as *S*-benzyl glutathione and *S*-hexyl glutathione show that L7 and L23 residues are not directly involved in the binding of GSH (Sinning *et al.*, 1993; Le Trong *et al.*, 2002). They do, however, form part of the secondary structure elements that make up the G-site (Figure 24). The Leu7 and Leu23 residues are unlikely to be involved in the direct binding of the ligand in the H-site due to their location. Having said that, proteins are dynamic molecules, and the possible effects of L7A and L23A mutations on the G- and H-site conformation cannot be overlooked. The effects of L7A and L23A mutation on the overall functional properties of the enzyme were determined by comparing the specific activity and ligand binding properties of the wild-type and mutant proteins.

3.7.1. Specific activity

The enzyme activity of wild-type and mutants was determined using the standard GSH-CDNB-conjugation assay (Habig and Jakoby, 1981). Figure 25 shows the specific activity of wild-type, L7A and L23A hGSTA1-1. It was measured spectrophotometrically at 340 nm by monitoring the formation of 1-(*S*-dinitrophenyl)-2,4-dinitrobenzene (Habig and Jakoby, 1981). The specific activity of wild-type, L7A and L23A proteins was obtained from the slope of a linear fit through the experimental data (Figure 25). The specific activity values obtained for L7A (8.01 $\mu mol.min^{-1}.mg^{-1}$) and L23A (6.99 $\mu mol.min^{-1}.mg^{-1}$) proteins were slightly reduced compared to that of wild-type (13.45 $\mu mol.min^{-1}.mg^{-1}$). The result obtained demonstrates that both L7A and L23A are mutants are catalytically active. This implies that Leu7 and Leu23 resides are not involved in catalytic activity of hGSTA1-1, and also that the L7A and L23A mutations do not significantly affect the native G-site conformation.

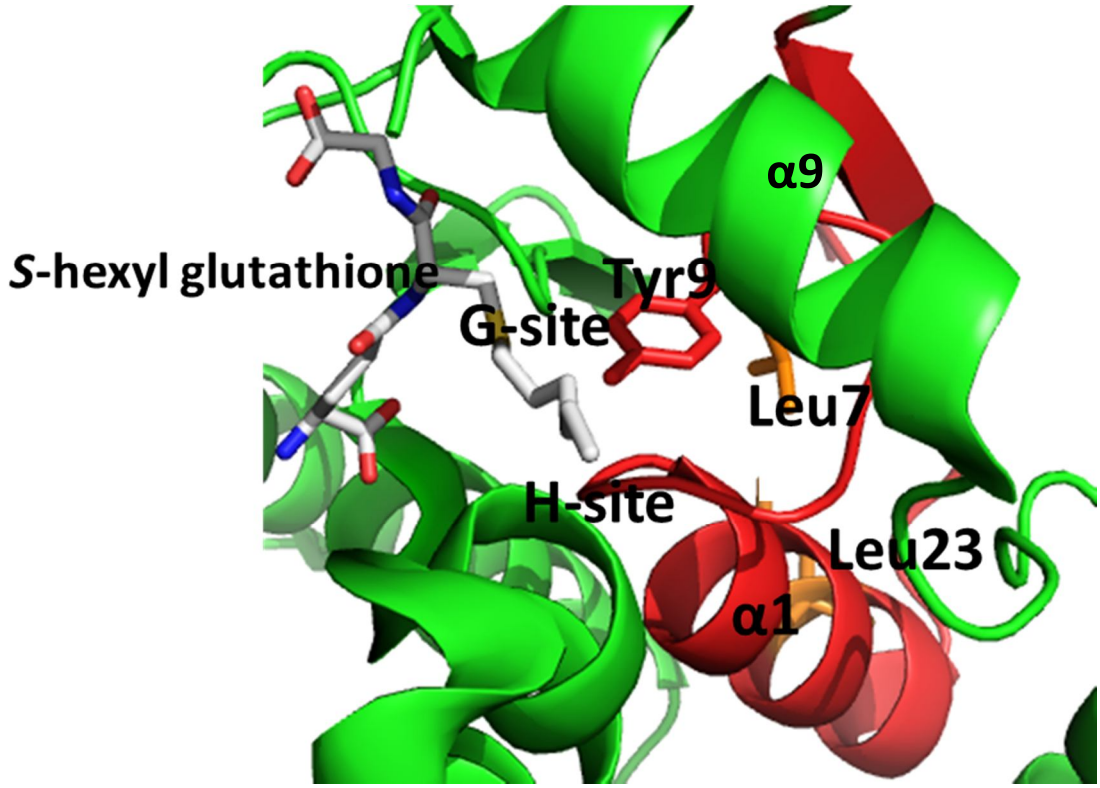


Figure 24: The location of Leu7 and Leu23 residues with regards to the active site location.

A section of hGSTA1-1 showing the G- and H- active site. The location of Leu7 and Leu23 residues in the N-Subdomain of hGSTA1-1 with regards to the active site are shown as stick model. The ligand *S*-hexyl glutathione is represented as stick model. The figure was generated from PDB file 1K3L (Le Trong *et al.*, 2002) using PyMOl v0.99 (DeLano Scientific, 2006).

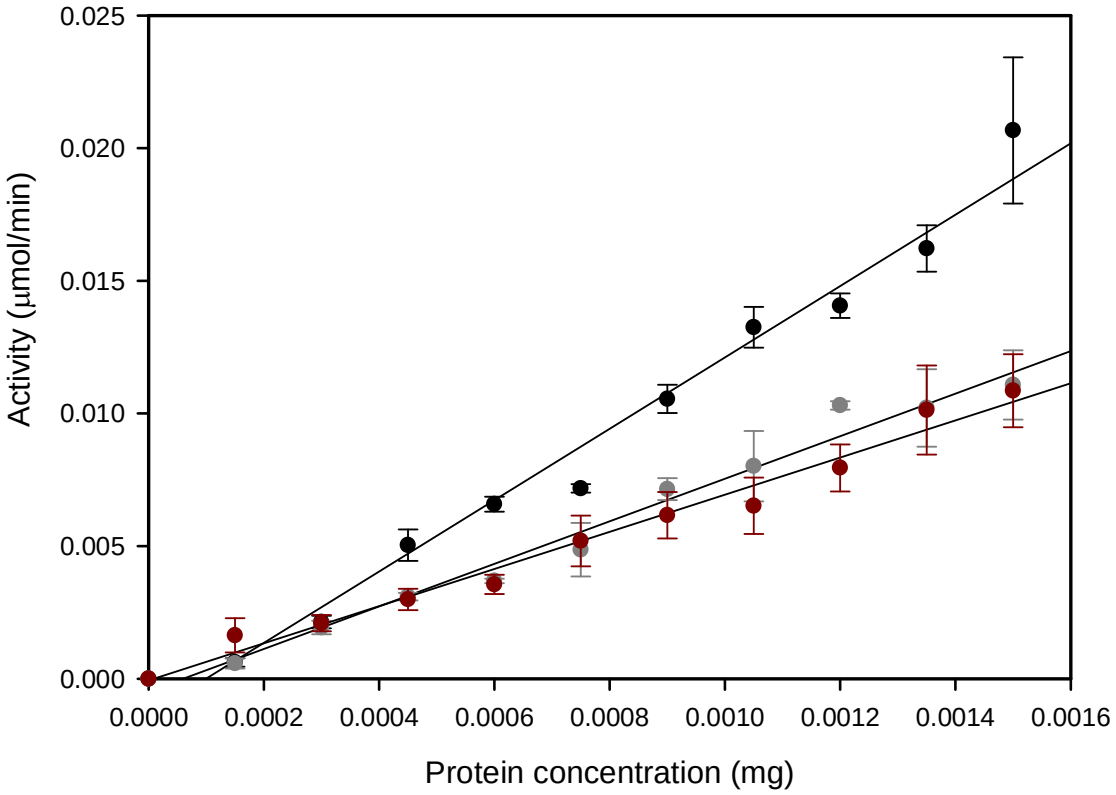


Figure 25: Specific activity of the wild-type, L7A, L23A hGSTA1-1 using the standard CDNB conjugation assay

The CDNB conjugation assay was performed in 0.1 M sodium phosphate buffer, pH 6.5, containing 1 mM EDTA and 0.02% sodium azide in the presence of 1 mM GSH and 1 mM CDNB. The final concentration of ethanol was 3% (v/v). The formation of the 1-(*S*-glutathionyl)-2,4-dinitrobenzene complex was monitored at 340 nm. The enzyme activity obtained for wild-type, L7A and L23A protein is shown in black, grey and red, respectively. The specific activity for each enzyme was determined from the slope of the fit and is shown as solid lines in the experimental data. The correlation coefficient for the fits is 0.9646, 0.9550 and 0.9359 for wild-type, L7A and L23A protein, respectively. The experiment was performed in triplicate (3 independent times) and the error bars show the standard deviation. The non-enzymatic activity was subtracted from the experimental data.

3.7.2. Ligand binding properties

ANS is a versatile fluorescence probe that has been extensively used to characterise the hydrophobic regions in proteins (Slavik, 1982; Semisotnov *et al.*, 1991; Engelhard and Evans, 1995). The change in the hydrophobic nature of the protein could be due to various factors including protein unfolding, protein-ligand interaction and mutations that have an overall effect on protein structure. The fluorescence properties of free ANS in water is characterised by an emission wavelength at 530 nm (Slavik, 1982; Engelhard and Evans, 1995). However, upon ANS binding to hydrophobic patches on proteins, a blue shift in emission wavelength is observed. This wavelength shift is accompanied by an increase in fluorescence intensity (Slavik, 1982; Engelhard and Evans, 1995). Therefore, the extensive use of ANS as a fluorescence probe is due to its fluorescence properties which are largely dependent on the polarity of its environment.

The effects of L7A and L23A mutations on the H-site were examined using ANS as probe. The fluorescence emission spectra of ANS bound to wild-type, L7A and L23A hGSTA1-1 is shown in Figure 26. Free ANS in 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5 has an emission maximum of 518 nm with low fluorescence intensity. As expected, the ANS emission maximum showed a blue shift when bound to hGSTA1-1 with an increase in fluorescence intensity. The emission maximum of ANS in the presence of both wild-type and L7A was 482 nm. This result implies that the hydrophobic nature of H-site is not affected by the mutations. Therefore, the L7A mutation did not alter the native conformation of the H-site. The emission properties of ANS bound L23A are slightly different to that of ANS bound to wild-type protein. It showed an emission maximum of 479 nm with similar fluorescence intensity. Even though the obtained results show a small shift (3 nm) between the emission properties of ANS bound to wild-type and L23A it is unlikely that H-site conformation has changed significantly.

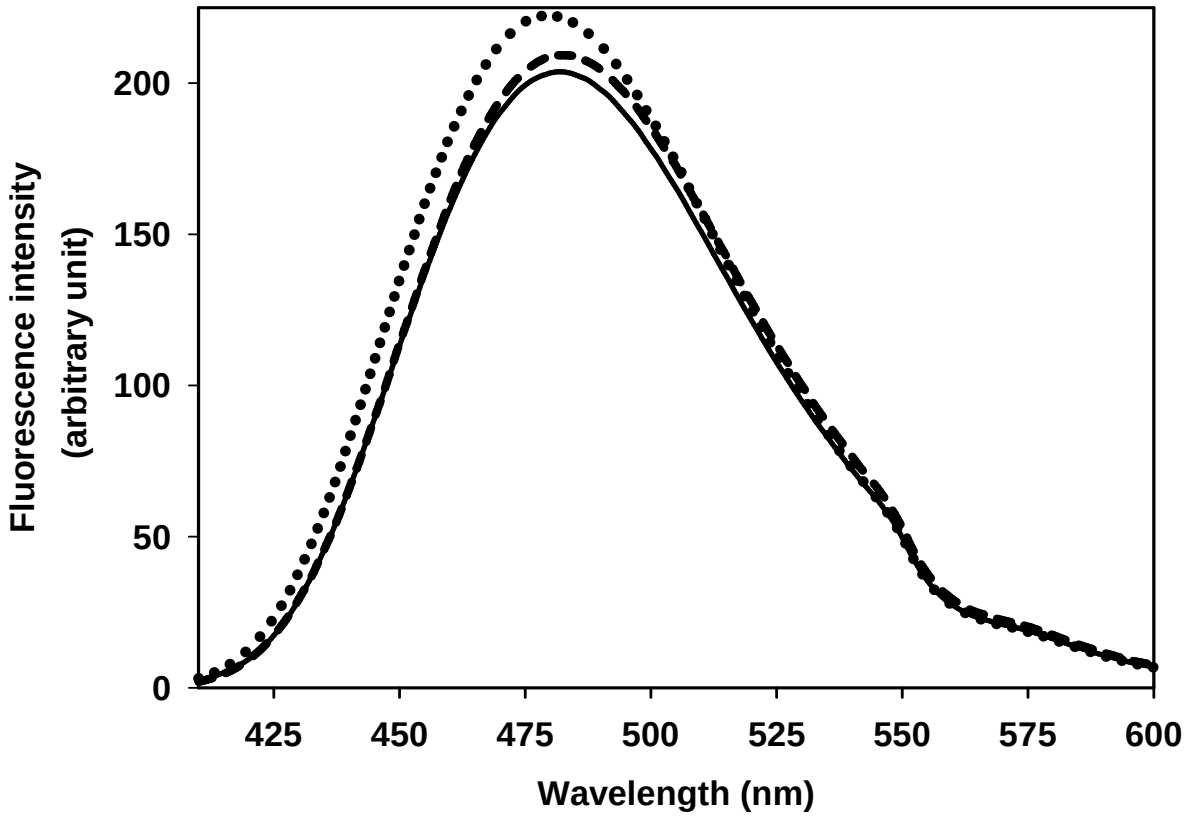


Figure 26: Fluorescence emission spectra of ANS bound to wild-type, L7A and L23A hGSTA1-1

ANS (200 μM) was incubated with wild-type and the mutant proteins (2 μM) in 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5. The data for wild-type, L7A and L23A proteins are shown in black, red and grey solid lines, respectively. The spectrum of free ANS was subtracted from the experimental spectra. The ANS was selectively excited at 390 nm and the emission wavelength was monitored between 400 and 600 nm.

3.8. Conformational stability of wild-type, L7A and L23A hGSTA1-1

The L7A and L23A structures when compared to the wild-type structure, showed a structural modifications of the N-subdomain as a result of the L7A and L23A mutations. The structural modification for both mutants were (a) presence of cavities in the hydrophobic core, and (b) Leu7 (in L23A) and Phe30 (in L7A) and adopting new conformations; these modification did not however significantly affect the functional properties of hGSTA1-1. Spectroscopic methods and pulse proteolysis were used to determine how these structural changes affect the thermodynamic stability of hGSTA1-1. The thermodynamic stability of a protein is defined by the Gibbs free energy difference between the folded and the unfolded states. The key requirement to determine protein stability is that the protein unfolding reaction needs to be thermodynamically reversible.

3.8.1. Thermal-induced unfolding

Thermal unfolding of wild-type GST has been shown to be an irreversible process, due to GST aggregation upon heat-induced unfolding (Kaplan *et al.*, 1997; Wallace *et al.*, 1998b; Stenberg *et al.*, 2000). The thermal-induced unfolding profiles were therefore qualitatively used to establish whether the L7A and L23A mutations affected the thermal stability of hGSTA1-1. hGSTA1-1 is characterised by a large ellipticity trough at 222 nm and unfolded proteins lack this signal (Figure 27). The thermal denaturation profiles of L7A and L23A were, therefore, determined by monitoring the change in ellipticity at 222 nm in response to constant changes in temperature from 20 - 80 °C in comparison with that of wild-type. The change in ellipticity at 222 nm, which is proportional to the α-helical protein content, was monitored as a function of temperature. The melting temperature (T_m), defined as the midpoint of the denaturation process, was used as a parameter to compare the transition curves of wild-type and mutants. The thermal unfolding profiles of wild-type and mutants show that hGSTA1-1 unfolds in a co-operative manner as characterised by a sigmoidal shape.However, none of the curves were superimposable (Figure 27). The T_m for the wild-type was 57 °C whereas the T_m for the L7A and L23A mutants were both significantly reduced to 49 °C. Even though the thermodynamic parameters cannot be determined from thermal unfolding transitions of hGSTA1-1 since the unfolding curve is irreversible, the reduced T_m values obtained for both mutants suggest that L7A and L23A mutation resulted in destabilisation of hGSTA1-1.

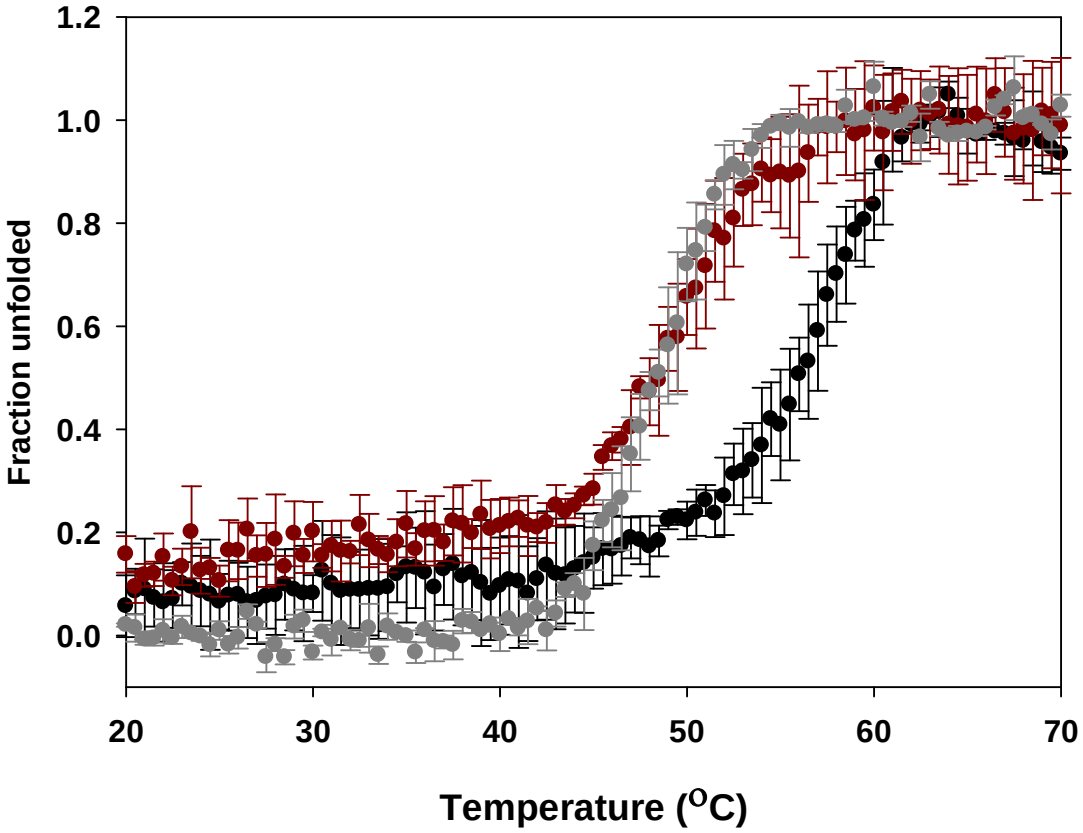


Figure 27: Thermal-induced unfolding curves for wild-type, L7A and L23A hGSTA1-1

Thermal-induced unfolding curves of the wild-type (black) and L7A (red) and L23A (grey) were monitored 222 nm from 20 and 80 °C using far-UV CD as a probe. The concentration of wild-type and mutants was 2 μM in 2 mM sodium phosphate buffer, pH 7.5, containing 1 mM EDTA and 0.02% sodium azide. The experiment was performed in triplicates (3 independent times). The error bars show the standard deviation.

3.8.2. Proteolytic susceptibility of wild-type, L7A and L23A hGSTA1-1 to thermolysin digestion

Pulse proteolysis, using thermolysin as a proteolytic enzyme, has emerged as a recent alternative technique to the traditional spectroscopic methods used to study protein stability (Park and Marqusee, 2005). Like spectroscopic probes, pulse proteolysis can efficiently distinguish the unfolded from folded protein conformations. It exploits the difference in proteolytic susceptibility between these two states. Thermolysin hydrolyses peptide bonds on the N-terminal side of valine, leucine, isoleucine and phenylalanine residues (Ambler and Meadway, 1968).

In proteins under native conditions, these hydrophobic residues are buried within structure thus precluding the native protein from proteolytic susceptibility. When a native state (N) of protein exists in equilibrium with the unfolded state (U), the protein species in the unfolded state are susceptible to proteolysis due to exposure of hydrophobic residues. This allows quantification of the native protein after selective digestion of the unfolded proteins by pulse proteolysis because it may be assumed that the digestion of folded proteins is negligible. Proteins that are digested significantly in their native conformation, pulse proteolysis should not be used to determine the protein's stability (Park and Marqusee, 2005, 2006; Na and Park, 2009). The conformational stability of wild-type, L7A and L23A was studied using thermolysin proteolysis since the digestion of hGSTA1-1 is negligible under native conditions (Figure 28). To create an equilibrium between native and unfolded state of hGSTA1-1, the proteins were prepared in varying concentrations (0 – 8 M) of urea and subjected to pulse proteolysis using thermolysin. The robustness of thermolysin allows it to retain its activity in 8 M urea (Oshima *et al.*, 1977).

The proteolytic susceptibility of L7A and L23A was determined by quantifying fractions of folded and unfolded protein in comparison with that of wild-type. The difference in the proteolytic susceptibility between wild-type and mutants is shown in Figure 29A, B and C. The digestion of L7A and L23A starts at 2 M and 1.5 M urea, respectively whereas the wild-type digestion starts at 4 M urea. The obtained proteolytic susceptibility profiles of L7A and L23A demonstrates that both mutants are more prone to thermolysin digestion when compared to the wild-type.

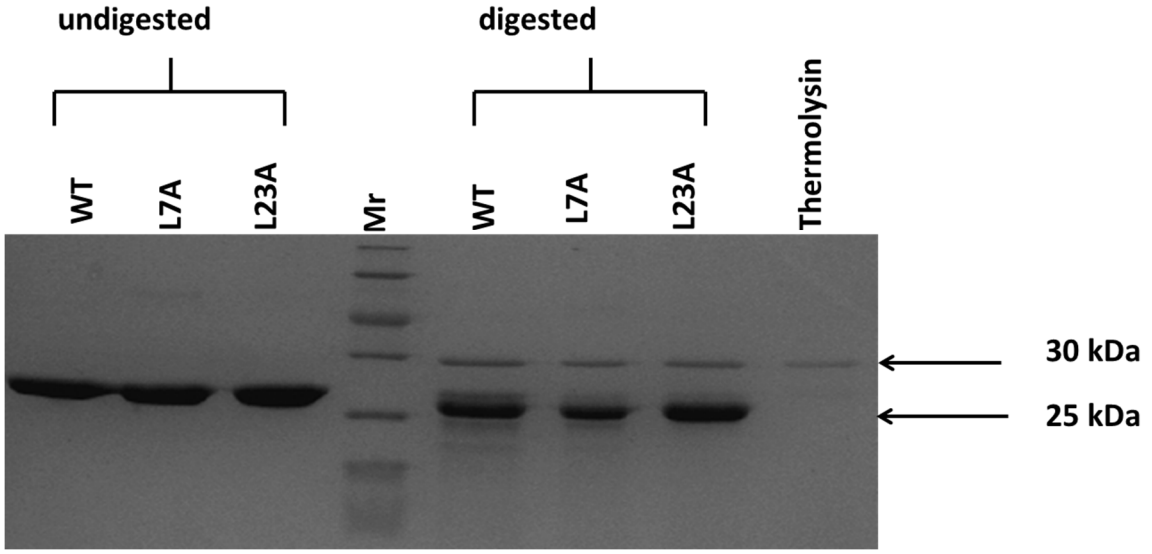


Figure 28: The proteolytic susceptibility of wild-type, L7A and L23A hGSTA1-1 under native conditions

The proteins (10 µM) were subjected to thermolysin digestion for 1 minute in 20 mM Tris-HCl, 50 mM NaCl, 10 mM CaCl2, pH 7.5. The amount of intact protein after pulse proteolysis was determined by quantifying the band intensities using gel densitometry.

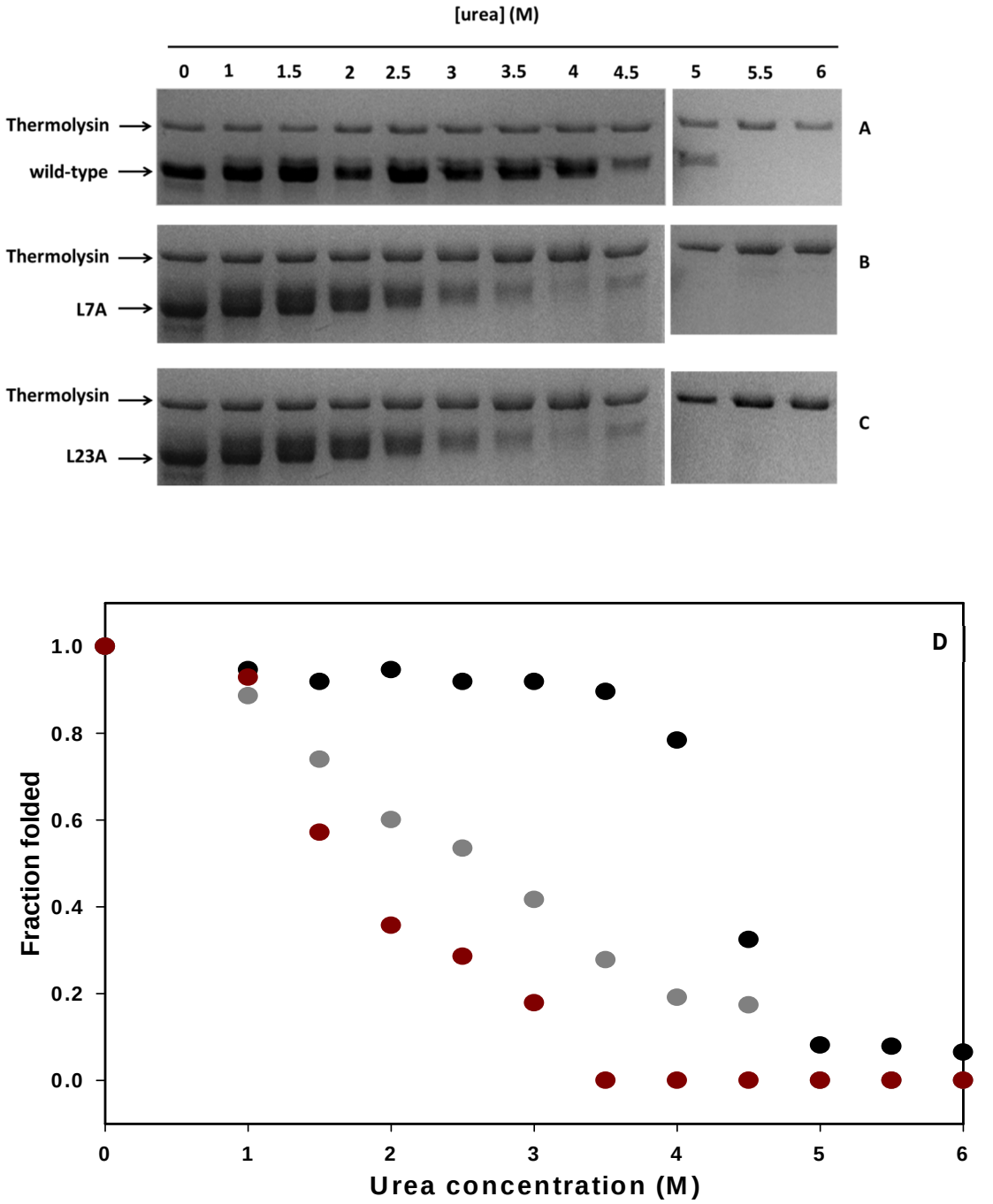


Figure 29: Proteolytic susceptibility of wild-type, L7A and L23A hGSTA1-1 to thermolysin digestion in the presence of urea

The proteolytic susceptibility of wild-type, L7A and L23A proteins in varying urea concentration (0 – 6 M) is shown in A, B, and C, respectively. The amount of intact proteins after pulse proteolysis was determined by quantifying the band intensities using gel densitometry and presented as fraction unfolded in D.

To further evaluate the effect of the L7A and L23A mutations on the stability of hGSTA1-1, the midpoint of wild-type and mutants unfolding transition (C_m) in urea was determined (Figure 29D). The C_m values for the wild-type was 4.2 M which is much higher than of L7A (2.5 M) and L23A (1.8 M).

3.8.3. Urea-induced equilibrium unfolding studies of wild-type, L7A and L23A hGSTA1-1 proteins

Equilibrium unfolding studies of hGSTA1-1 were performed in the presence of 0 – 8 M urea. Prior to performing the unfolding experiments, the reversibility of the unfolding pathway and recovery of the native conformation of the protein needs to be established so as to allow determination of the thermodynamic parameters for the unfolding process (Pace, 1986). The recovery of the native conformation was assessed on the basis of the extent to which the protein regained its secondary and tertiary structure using far-UV CD and the intrinsic fluorescence as probes, respectively. Denaturation of hGSTA1-1 with urea has been shown to be reversible (Wallace *et al.*, 1998b), and Figure 30 shows that urea-induced unfolding of L7A and L23A is also reversible with recoveries of above 90%. The formation of aggregates and hysteresis during the unfolding pathway interferes with the recovery and reversibility of the unfolding reaction, therefore, resulting in a process that is no longer considered to be in equilibrium.

The unfolding and refolding transitions of the wild-type and mutants were superimposable, implying no hysteresis observed during the unfolding pathway (Figure 31). The unfolding transition of wild-type and mutants also showed no protein aggregation when monitored with light scattering (Figure 32). The urea-induced unfolding of L7A and L23A was confirmed to be an equilibrium process. The equilibrium unfolding studies of L7A and L23A proteins were subsequently performed to determine thermodynamic parameters of their unfolding process.

The urea-induced unfolding of L7A and L23A was probed with far-UV CD, fluorescence spectroscopy and ANS binding and compared to that of wild-type. The obtained unfolding profiles are shown in Figure 33. The unfolding transition for all the proteins displayed monophasic sigmoidal transition between the native and unfolded state when probed with far-UV CD and fluorescence spectroscopy.

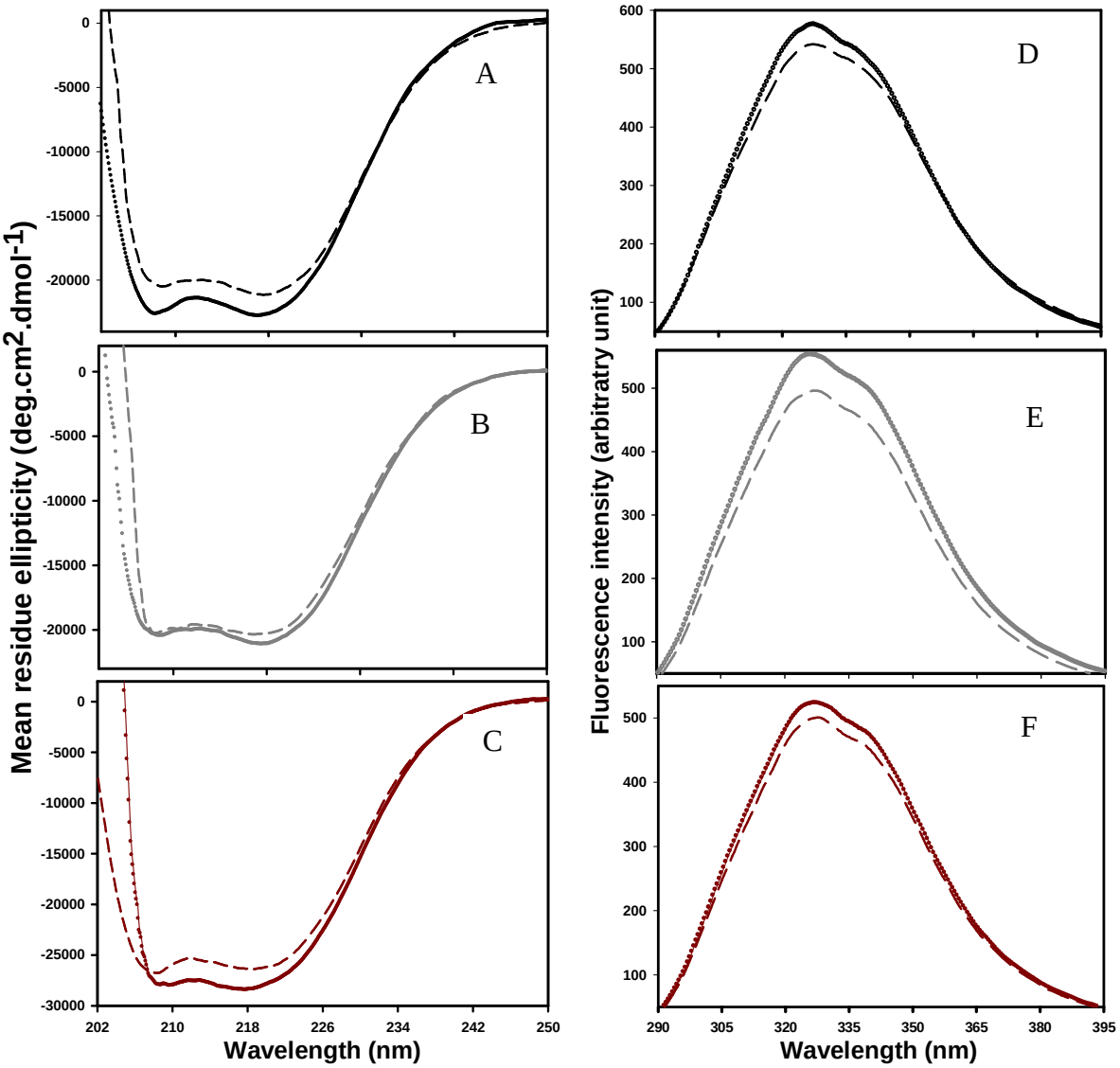


Figure 30: Recovery of wild-type, L7A and L23A hGSTA1-1 unfolding

The recovery of wild-type and its variants was monitored with far-UV CD (A, B, C) and fluorescence spectroscopy (D, E, F). The wild-type (black; A, D), L7A (red; C, F) and L23A (grey, B. E) were unfolded in the presence of 8 M urea. The reaction was incubated for an hour to reach equilibrium and then the proteins were diluted 10-fold in 20 mM sodium phosphate pH 7.5, 1 mM EDTA, and 0.02% sodium azide. The residual concentration of urea for the refolded protein was 0.8 M. The native (control) and refolded proteins are shown as dotted and dashed lines, respectively.

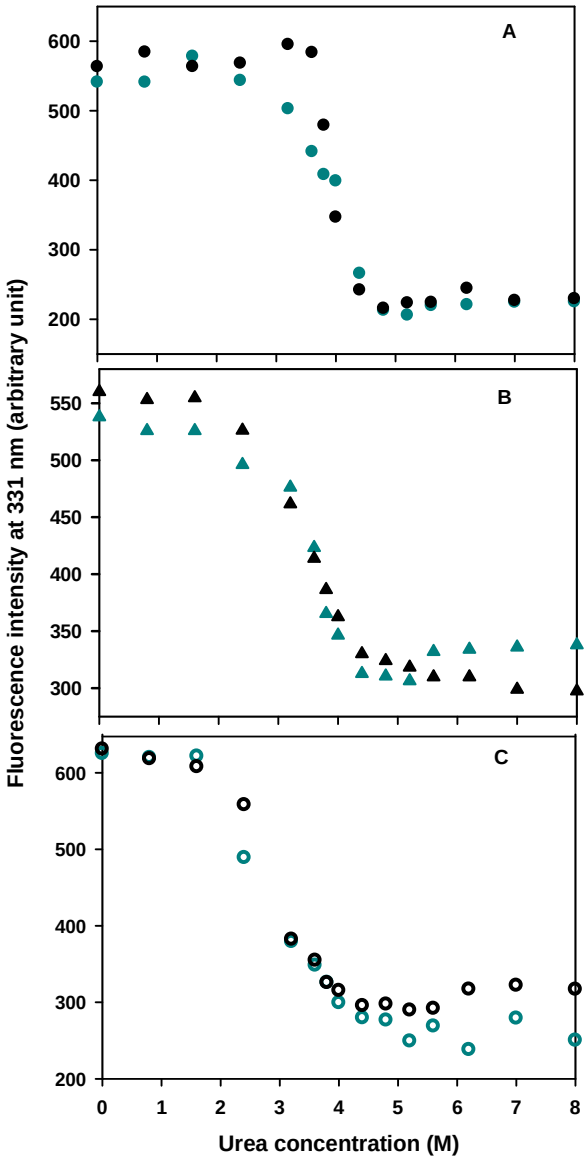


Figure 31: Reversibility of urea-induced equilibrium unfolding of wild-type, L7A and L23A hGSTA1-1

The reversibility of urea-induced equilibrium unfolding wild-type (A), L7A (B) and L23A (C) was monitored with fluorescence spectroscopy. The proteins were unfolded in the presence of 8 M urea and allowed to reach equilibrium for an hour. The proteins were then diluted to a specific urea concentration along the unfolding curve and further incubated for an hour. The proteins were prepared and diluted in 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5. The black and cyan symbols represent unfolding and refolding of the proteins, respectively. The proteins were excited at 280 nm and the emission wavelength was monitored at 331 nm.

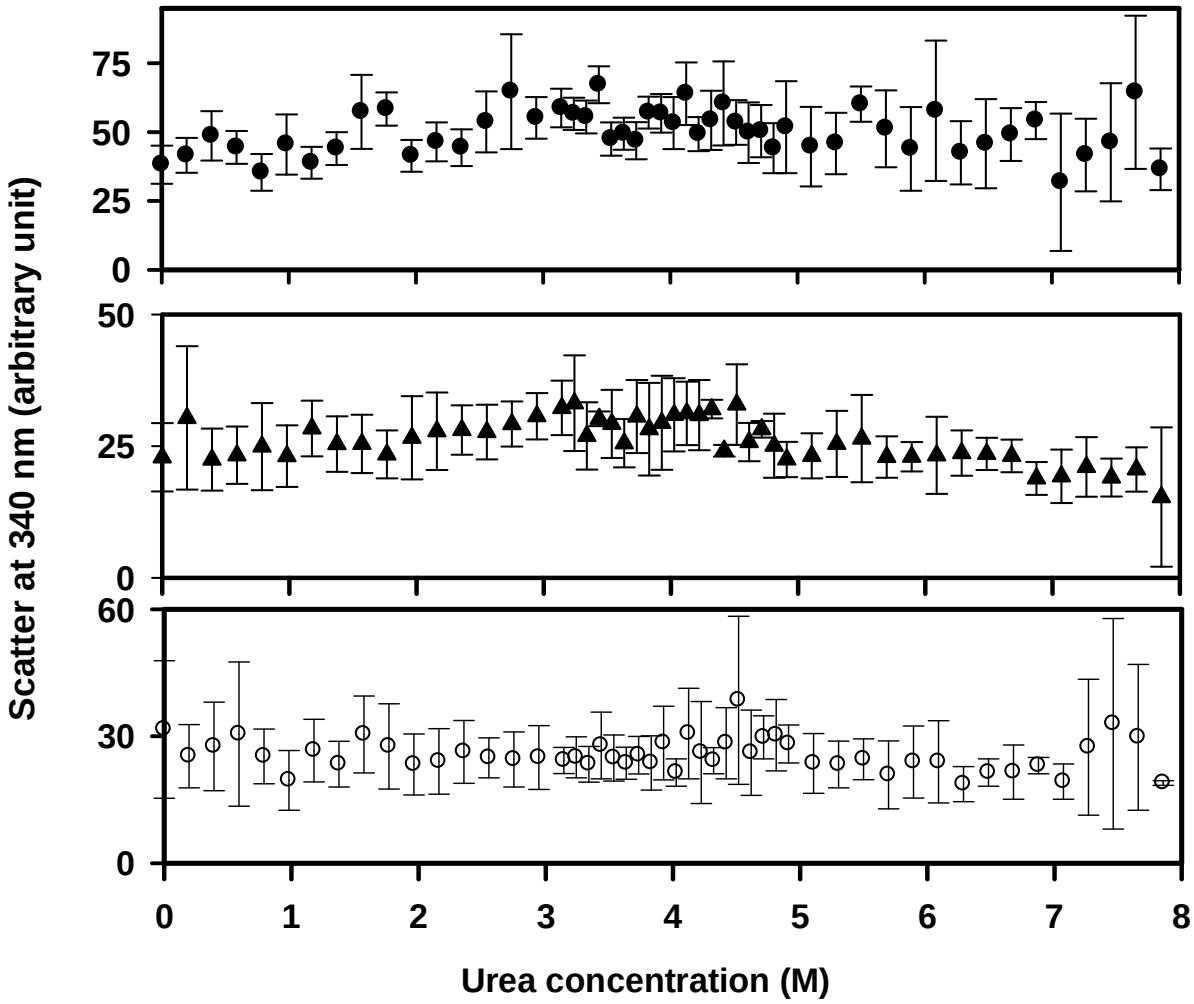


Figure 32: Rayleigh scatter of wild-type, L7A and L23A hGSTA1-1 during urea-induced equilibrium unfolding

Protein aggregation during the unfolding transition of wild-type (●), L7A (▲) and L23A (○) was monitored with light scattering by setting both the excitation and emission wavelengths to 340 nm. The experiment was performed three independent times and the error bars show standard deviation.

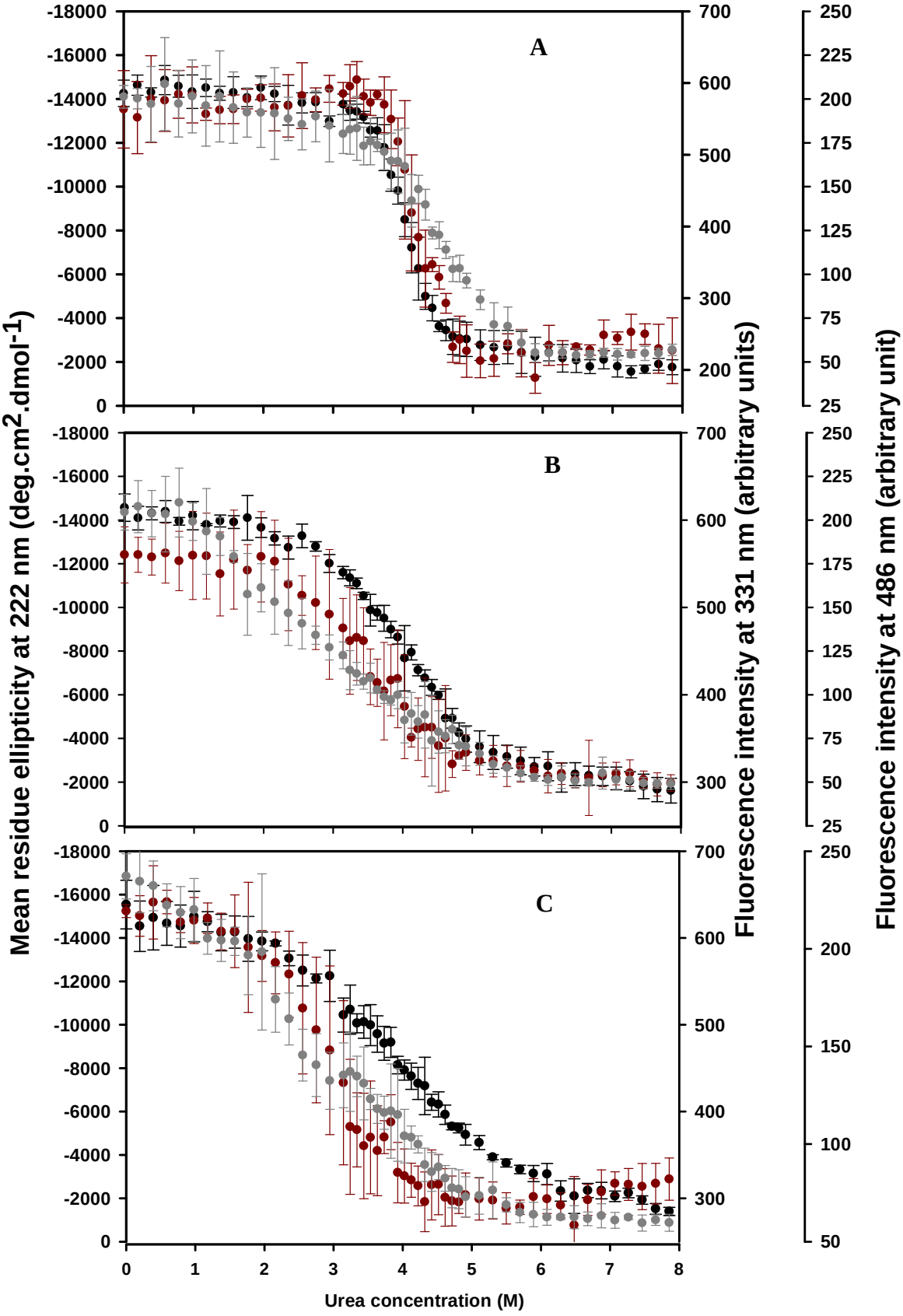

A
B
C
Mean residue ellipticity at 222 nm (deg.cm2.dmol-1)
Fluorescence intensity at 331 nm (arbitrary units)
Fluorescence intensity at 486 nm (arbitrary unit)
Urea concentration (M)
-18000
-16000
-14000
-12000
-10000
-8000
-6000
-4000
-2000
0
700
600
500
400
300
200
250
225
200
175
150
125
100
75
50
25
0
1
2
3
4
5
6
7
8

Figure 33: Urea-induced equilibrium unfolding of wild-type, L7A and L23A hGSTA1-1

Urea-induced unfolding curve of 2 µM wild-type (A), L7A (B) and L23A (C) in 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5 monitored by far-UV CD (black), tryptophan fluorescence (red), and ANS binding (grey). For the far-UV CD measurement, the change in the helical structure of the proteins was monitored at 222 nm. Changes in the tertiary structure were determined by exciting the protein at 280 nm and recording the change in the fluorescence intensity at 331 nm.

The wild-type unfolding data obtained using far-UV CD and fluorescence are superimposable implying that both secondary and tertiary structure are lost simultaneously as the protein unfolds (Figure 33A). This data also suggests a two-state transition model for the wild-type unfolding, which is in agreement with published hGSTA1-1 unfolding data (Wallace *et al.*, 1998b; Wallace and Dirr, 1999). Far-UV CD and fluorescence data obtained for L7A and L23A do not coincide (Figure 33B and C) but both traces show single transitions. The difference in the unfolding profile could be due to the fluorescence probe being more senstive to local structural changes since it is in close proximity to the site of mutation.

Figure 34 shows the differences in the unfolding transitions of wild-type and its variants. The far-UV CD (global probe) data demonstrate that the wild-type begins to unfold at 3.4 M urea whereas both mutants start unfolding below 2 M urea. By visual inspection, the midpoint of the wild-type, L7A and L23A unfolding is the same (4.2 M), however the slope of the transition region is different (Figure 34A). The far-UV CD unfolding profile of L7A and L23A is in contrast to data obtained using fluorescence probe (Figure 34B). None of the three unfolding transitions overlay and the midpoint of L7A and L23A are 3.8 and 3.4 M, respectively which is much lower than that of the wild-type (4.2 M).

The unfolding transition of L7A and L23A was further probed with ANS binding to detect presence of intermediates during the unfolding process. ANS binding reports conformational changes that are induced in proteins during the unfolding process since it displays a higher emission when bound to a molten globule state than when bond to a protein in its native state (Semisotnov *et al.*, 1991; Ptitsyn *et al.*, 1995). The binding of ANS to hGSTA1-1 displays high fluorescence intensity with emission maximum at 482 nm, which is diminished in the unfolded protein. The change in fluorescence intensity was, therefore, used as a function of ANS binding.

Figure 34C shows the different effects of ANS binding to wild-type, L7A and L23A unfolding transition. The binding of ANS to wild-type shows a clear monophasic sigmoidal transition between the native and unfolded state. The transition region is between 3.8 and 4.8 M urea. In contrast, the transition region for the mutants starts before 1 M urea indicating a structural difference between wild-type and its variants (Figure 34C). The absence of enhanced fluorescence intensity during the unfolding transition of L7A and L23A indicates no significant population of intermediate species or aggregates along the unfolding pathway.

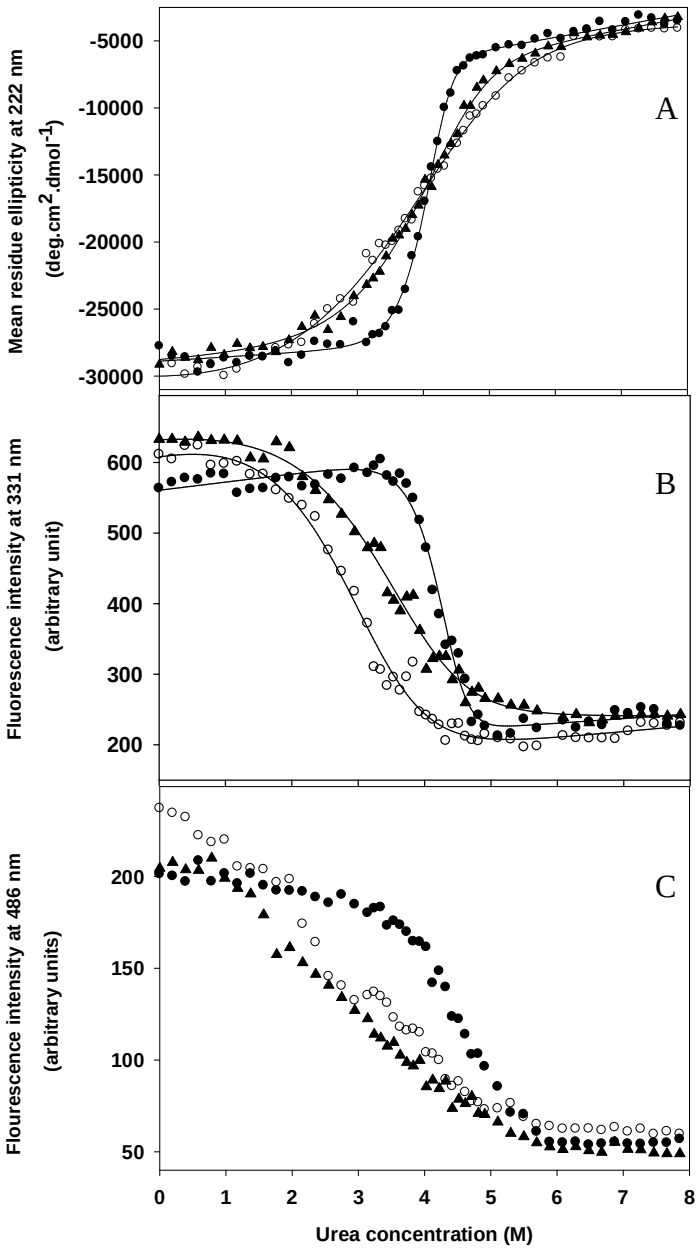


Figure 34: Comparison of urea-induced equilibrium unfolding of wild-type, L7A and L23A hGSTA1-1

Urea-induced unfolding curve of wild-type (●), L7A (▲) and L23A (○) monitored by far-UV circular dichroism (A), tryptophan fluorescence (B) and ANS binding (C). Protein concentration is 2 μM in 20 mM sodium phosphate, 1 mM EDTA, and 0.02% sodium azide, pH 7.5. Each datum point is an average of three independent experiments. The data in A and B was fitted according to a 2-state model (solid lines).

The results for wild-type and its variants are consistence with published data of hGSTA1-1 unfolding monitored with ANS binding (Wallace *et al.*, 1998b; Alves *et al.*, 2006).

Obtaining accurate thermodynamic parameters ($\Delta G(H_2O)$, *m*-value and C_m) for the unfolding experiments requires that the unfolding reaction must be reversible and be in equilibrium. Furthermore, the unfolding data has to analysed using an appropriate model. The model could be two-state or multi-state depending on the unfolding data. The published thermodynamic parameters of hGSTA1-1 were obtained using a two-state model (Wallace *et al.*, 1998a, 1998b, 2000; Wallace and Dirr, 1999). hGSTA1-1 meets the criteria that have been established in order for a protein to be considered to unfold via a two-state unfolding transition (Whitten and Garcia-Moreno, 2000). These criteria are (1) the protein must have identical ΔG values when monitored by different probes, (b) the quality of the two-state fits to the data must be good, (3) the far-UV CD and fluorescence unfolding curves for the protein must be superimposable, and (4) the shape of the unfolding curves must be a single sigmoidal. The L7A and L23A were subjected to the same criteria and subsequently analysed using both a two-state and multi-state since the far-UV CD and fluorescence were not superimposable (Figure 34B and C). Fitting the L7A and L23A data to two-state model showed a good fit and residual plots from the fit are shown in Figure 35. The C_m value, $\Delta G(H_2O)$ and *m*-value obtained for L7A and L23A using the two-state fit are shown in Table 4 and compared to that of the wild-type. The multi-state fits of L7A and L23A unfolding data were poor and are not shown.

The unfolding curves of L7A and L23A do not overlay with those of the wild-type. They were shifted to a lower urea concentration. The C_m value for the L7A and L23A unfolding curve showed a decrease of 0.2 and 0.8 M urea in comparison to the wild-type. The C_m value (transition midpoint) is defined as the point at which half of the population of the protein is unfolded. The lower C_m values obtained for the mutant suggest a loss in hGSTA1-1 stability as a result of the L7A and L23A mutation.

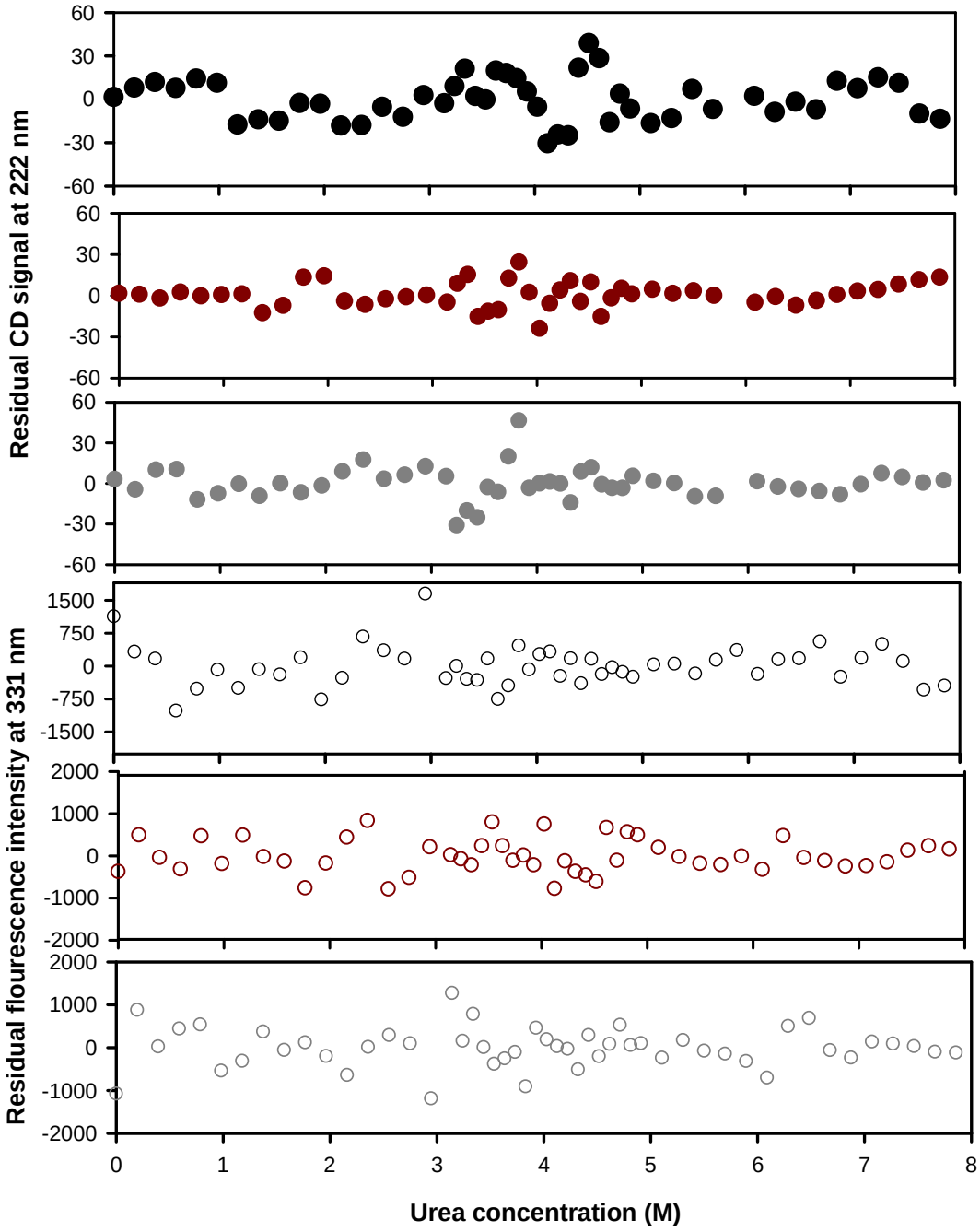


Figure 35: Residual plot of the data fitted to a two-state model

The residuals obtained from fitting far-UV CD and fluorescence data from unfolding curves of wild-type and its variants are shown in filled and unfilled symbols, respectively. The data from wild-type, L7A and L23A proteins are shown in black, red and grey, respectively.

Table 4: Urea-induced equilibrium unfolding thermodynamic parameters of wild-type, L7A and L23A hGSTA1-1

Urea-induced equilibrium unfolding of wild-type and its variants was monitored by tryptophan fluorescence (F331) and far-UV CD (E 222). The thermodynamic parameters were calculated by fitting the unfolding data to a two-state model

Parameter	wild-type			L7A			L23A		
	E 222 nm	F331 nm	Average	E 222 nm	F 331 nm	Average	E 222 nm	F331 nm	Average
$\Delta G(H_2O)$ (kcal.mol^{-1})	23.48 (± 0.02)	24.28 (± 0.10)	23.88	13.25 (±0.42)	11.95 (±0.81)	12.6	10.29 (±0.46)	10.69 (±0.65)	10.49
m (kcal.mol^{-1}.M^{-1}urea)	4.02 (±0.05)	4.1 (±0.05)	4.06	1.53 (±0.09)	1.39 (±0.20)	1.46	0.86 (±0.10)	1.27 (±0.16)	1.06
C_m (M urea)	4.2	4.2	4.2	4.2	3.8	4.0	4.2	3.4	3.8

The parameters obtained for all three proteins using both far-UV CD and fluorescence probes are in agreement with one another (Table 4). The $\Delta G(H_2O)$ values obtained for the L7A (12.6 kcal.mol^{-1}) and L23A (10.49 kcal.mol^{-1}) are lower by 48% and 57%, respectively when compared to the wild-type. The reduction in $\Delta G(H_2O)$ values associated with L7A and L23A mutations in hGSTA1-1 suggest that these topological conserved residues play a significant role in its stability. The *m*-values of 1.46 and 1.06 kcal.mol^{-1}.M^{-1} urea obtained for L7A and L23A, respectively are reduced by over 60% when compared to that of the wild-type (4.06 kcal.mol^{-1}.M^{-1} urea). The *m*-value is defined by dependence of the free energy of unfolding on denaturant concentration (Shirley, 1995). It is the slope in the transition region in the unfolding curve. The *m*-value therefore indicates the degree of surface area exposed during denaturant induced unfolding (Greene and Pace, 1974; Myers *et al.*, 1995; Soulages, 1998). The low *m*-values obtained for the mutants indicate that the cooperativity of hGSTA1-1 unfolding is diminished in both mutants (Soulages, 1998). The difference in the unfolding transition of wild-type and variants is further supported by different effects of ANS binding to these proteins (Figure 34).

hGSTA1-1 is a dimeric protein which unfolds via a two-state model through a cooperative mechanism without a significant population of intermediate species using urea as a denaturant (Wallace *et al.*, 1998b). It is known that for all dimeric proteins, the equilibrium unfolding curve is dependent on the protein concentration due to the law of mass action (Ragone, 2000; Rumfeldt *et al.*, 2008). The shape of the curve remains unchanged with an increase in protein concentration; however, the midpoint of the transition (C_m value) shifts to a higher urea concentration. The reduced *m*-values obtained for the L7A and L23A suggest a presence or accumulation of the intermediate at equilibrium (Myers *et al.*, 1995; Shirley, 1995). With this in mind, the intermediate species formed during the urea-induced unfolding of L7A and L23A hGSTA1-1 was further characterised by performing protein concentration dependence studies. Figure 36 shows concentration shift dependence in the unfolding transition region of wild-type and its variants which is consistent with unfolding of dimeric proteins (Wallace *et al.*, 1998; Ragone, 2000; Rumfeldt *et al.*, 2008).

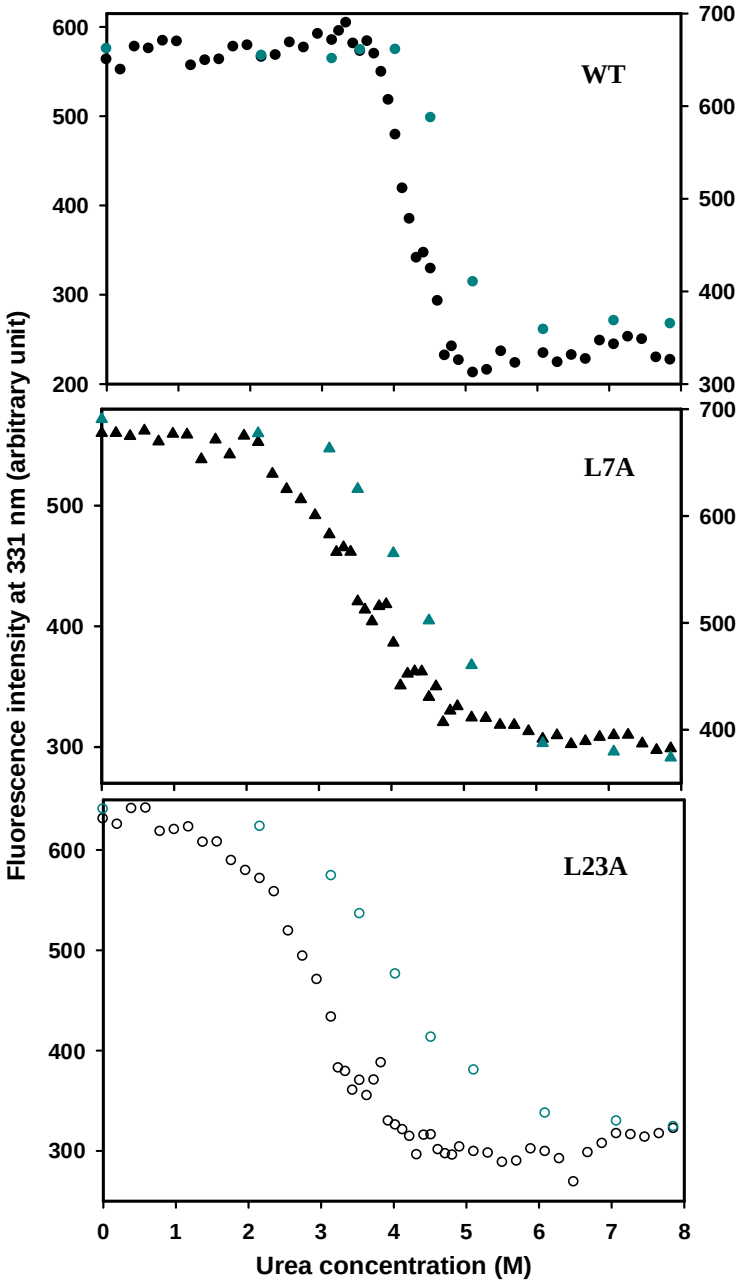


Figure 36: Protein-concentration dependence of urea-induced equilibrium unfolding of wild-type, L7A and L23A hGSTA1-1

Urea-induced unfolding curve of wild-type (●), L7A (▲) and L23A (○) monitored by tryptophan fluorescence. The black and cyan symbols represent 2 and 20 μM protein, respectively, in 20 mM sodium phosphate pH 7.5., 1 mM EDTA, and 0.02% sodium azide.

4. Discussion

The thioredoxin fold is preserved in various protein families with diverse functions despite their low sequence identity (Pan and Bardwell, 2006). The N-terminal domain of proteins in the GST family is characterised by a thioredoxin-like fold, which forms part of the active site and the hydrophobic core of hGSTA1-1 (Figure 2). An interesting feature about the thioredoxin fold is that it has adapted in various ways to enable different types of catalytic reactions and yet its structure has been maintained (Martin, 1995; Pan and Bardwell, 2006; Atkinson and Babbitt, 2009a, 2009b; Pedone *et al.*, 2010). This perhaps indicates that a few key residues are essential to drive folding and the maintenance of this fold. This phenomenon is evident in functionally or structurally-related proteins, with low sequence similarities across the family. With this in mind, structure-based alignments of the thioredoxin-like superfamily, using hGSTA1-1 as model identified leucine residues at position 7 and 23 as being topologically conserved.

Leu7 and Leu23 are located in β1 and α1 of the N-subdomain, respectively, with their side chains buried in the hydrophobic core of hGSTA1-1. The Leu7 and Leu23 side chains make a number of van der Waals contacts with various residues in N-subdomain (Figure 8 and Figure 9). Formation of favourable van der Waals interactions between hydrophobic residues directs the folding process to form hydrophobic cores around which the surrounding secondary structures assemble to form a native protein structure (Dill, 1990a, 1990b). These van der Waals interactions are also important in maintaining the stability of hydrophobic cores that stabilise the native protein structure. Conserved residues could also be involved in long-range interactions which are crucial to maintenance of the global structure in proteins. These interactions have an impact on the catalytic or binding function of the protein. When conserved residues are located at or near the active site, they play an important role in the activity of the enzyme (Mirny and Shakhnovich, 1999; Friedberg and Margalit, 2002; Schueler-Furman and Baker, 2003). To establish the role of topologically conserved Leu7 and Leu23 in Trx-like domain, Leu7 and Leu23 were mutated to alanine. The leucine to alanine mutations resulted in removal of the bulky side chain of leucine without altering the hydrophobicity within the βαβ motif. This would allow to probe the contribution of the bulky side chain of Leu7 and Leu23 to the structure, stability and function of the hGSTA1-1.

The implication of L7A and L23A mutations on the structure and function of hGSTA1-1 were investigated using X-ray crystallography and conformational stability studies. The mutants were crystallised in the apo form to assess the effects of L7A and L23A mutations on the overall hGSTA1-1 structure and the impact of the mutations on the local environment precluding any effects that may be due to ligand binding. The conformational stabilities of the mutant proteins were assessed by pulse proteolysis and urea-induced equilibrium unfolding studies. The crystal structures of L7A and L23A were used to verify and interpret the spectroscopic characterisation of mutant structures and stability data. The properties of L7A and L23A were then compared to those of the wild-type.

4.1. hGSTA1-1 structure tolerates L7A and L23A mutation in the βαβ subdomain

Prior to determining the role of Leu7 and Leu23 conservation in hGSTA1-1, the mutant structures (L7A and L23A) were examined using far-UV CD and fluorescence spectroscopy to detect if the mutations introduced any conformational changes. Leu7 and Leu23 are located in β1 and α1 of hGSTA1-1, respectively. With hGSTA1-1 being mainly alpha helical, it is unlikely that changes in β1 will be detected with far-UV CD due to its low resolution (Woody, 1995). Furthermore, this change will be masked by high signals at 208 and 222 nm. Therefore, a significant change in the mutant structure would be representative of global change rather than local change. In contrast to the far-UV CD probe, Trp21 fluorescence is a local probe, which reports conformational changes in the N-terminal domain of hGSTA1-1 (Wallace *et al.,* 1998b) due to hGSTA1-1 containing a lone Trp21. Fluorescence spectroscopy is, therefore, a more sensitive tool since Trp21 is in close proximity to the site of the mutations. Both the secondary (Figure 15) and tertiary (Figure 16) structural properties of the wild-type enzyme are maintained in L7A and L23A structures, which correlates with the crystal structures of L7A and L23A (Figure 19). The L7A and L23A mutations are, therefore, non-disruptive and do not alter the backbone structure in the mutants. This is confirmed by overall C-alpha r.m.s.d values of 0.63 Å (L7A) and 0.67 Å (L23A) between the wild-type and mutant global structures. The non-disruptive nature of L7A and L23A mutations is further supported by both mutant proteins being catalytically active (Figure 25). This is consistent with the general observation of the structure seen in the wild-type being preserved in the mutant proteins (Matthews, 1991; Xu *et al.,* 1998; Achilonu *et al.,* 2010; Balchin *et al.,* 2010).

The initial step of Trx-like domain folding involves hydrophobic interactions between β1 and β3 of the N- and C-subdomain, respectively (Tasaysco *et al.*, 2000). Even though Leu7 is located in β1, mutation of this residue does not seem to affect the interactions of N- and C-subdomain to drive the folding of Trx-like fold. Examining refolding rates of the mutants will give a clearer indication of the involvement of Leu7 and/or Leu23 in the initial folding events of hGSTA1-1.

It is known, from other crystallographic studies, that neighbouring residues in the region of mutation are likely to be affected by point mutations (Shortle, 1992; Eyal *et al.*, 2003). Since the global structures of L7A and L23A remained unchanged, the local environments of mutation were examined to see if they also remained unchanged. If they changed, how did hGSTA1-1 respond to the mutation? Mutagenesis studies have showed that leucine-alanine mutations can have different effects on structural properties of the protein and is influenced by the local environment of the mutation site (Lim *et al.*, 1992; Shortle, 1992; Jackson *et al.*, 1993; Baldwin and Matthews, 1994; Gabellieri *et al.*, 2008). This type of mutation could create a cavity in the protein core without any significant change in the local environment of the mutation. The cavity that is created can also be occupied by water molecules depending on the volume of the cavity. Another structural change that can be caused by leucine to alanine mutation is the modification of side chain conformations of the neighbouring residues to occupy the cavity caused by the mutation. Changes in the side chain orientation of nearby residues can cause side chain rotation leading to side chains adopting non-ideal torsion angles (Shortle, 1992; Eyal *et al.*, 2003).

An overlay of the βαβ region (which represents the local environment of both mutation sites) between the wild-type and mutant structures demonstrates that the L7A and L23A mutation did not cause any major changes in their respective local environments. Mutation of leucine to alanine resulted in the loss of van de Waals contacts between L7A and Phe30, and L7A and Val58 in the L7A structure (Figure 21). The loss of van der Waals contact between the above mentioned residues was expected, because the shorter side chain of alanine is no longer within van der Waals radius with Phe30 and Val58. The mutation of L7A resulted in a change in the side chain orientation of Phe30. The phenyl plane in Phe30 is flipped when compared to that of the wild-type. The change in Phe30 orientation is due to a change in the bond angle. This new Phe30 orientation is stabilised by two new hydrogen bonds formed

between H_2O and Phe30 backbone. In the wild-type structure, these water molecules are on the surface and they do not make any contacts with Phe30 (Figure 22). It is unlikely that the change in the phenyl ring of the mutant when compared to wild-type is due to the difference in crystal packing in the unit cell since they were both crystallized in same space group.

In the L23A structure, subtle rearrangements in Val28 and Leu7 side chains are evident where their orientation is slightly different when compared to the wild-type structure (Figure 23). Val28 is within 4 Å of L23A and it is located in the loop region connecting α1 and β2. In the mutant structure (L23A), the side chain of Val28 is drawn towards L23A. It is unclear whether this side chain movement of Val28 occurred to optimize van der Waals interaction with nearby residues or whether it was caused by its location in the loop region. The loop regions are known to be highly flexible. Another interesting observation is the change in side chain conformation of Leu7 (β1). This residue is 5.98 Å from L23A and does not make any contact with L23A, however, proteins are dynamic molecules and the effects of mutation are not only limited to nearby residues.

The change in the Phe30 and Leu7 side change conformations in L7A and L23A, respectively, suggest that the packing density of βαβ subdomain of hGSTA1-1 has been modified highlighting the role of the topological conserved Leu7 and Leu23 residues. Despite both mutant structures showing side chain movements within the hydrophobic core in response to point mutations, this is not always the case with all proteins. The degree of side chain movement is also influenced by the packing density of the hydrophobic core.

In the GST superfamily, position 7 in β1 is reserved for an aliphatic residue; in most cases, a leucine residue is located at position 7 (Figure 7). In cases where position 7 is not occupied by a leucine residue, as in Delta, Omega, and Phi class GSTs as well as in Grx2, this position is occupied by an amino acid with a similar hydrophobicity scale (Table 2). This does not necessarlly mean that the behaviour of the naturally occurring substitutions will be same as the wild-type. Residues in the core have been shown to be highly conserved where in some cases, a position is optimised for stabilising interactions for a specific residue. For example, Godoy–Ruiz *et al*. (2005) showed that in thioredoxin proteins an isoleucine to valine substitution, which can happen naturally in evolution, can cause destabilisation. Lin and co-workers have also shown that a naturally occurring substitution of Phe151 to a leucine residue lowers the stability of GSTP1-1 (Lin *et al.*, 2003). Residue 151 is located at the N-

terminus of helix 6 in GST domain 2 and it is completely buried in the hydrophobic core of this domain. Mutation studies also demonstrate that regardless of how conservative the amino acid substitution can be, it can have an effect on stabilisation. Mutation of topologically conserved residue, Ile71 to valine or alanine (note the difference in the hydrophobicity index, (Table 2) which is located in α3 of the hGSTA1-1 showed, that mutation of this residue alters the packing density of the hydrophobic core (Achilonu *et al.*, 2010).

Delta (1JLV) and Theta (2C3N) classes have an alanine at position 19 (Ala19) and 21 (Ala21), respectively, topologically equivalent positions to Leu23 in hGSTA1-1 thus making them natural mutants (Figure 7). Structural alignment of Alpha and Delta class using C^{α} gives an r.m.s.d. of 1.7 Å indicating overall structural similarity between these classes. This value is increased to 5.9 Å when the alignment is limited to the βαβ a motif (Figure 37A) demonstrating an overall difference in βαβ conformation of this motif in the Alpha and Delta classes. Furthermore, the local environment of Ala19 is populated with hydrophobic residues (Table 5) as in the hGSTA1-1, however it is solvent accessible (Figure 37B). It has four water molecules which assist in maintaining the fold via hydrogen bonding with residues within βαβ motif. This suggests that the packing of βαβ motif seen in Alpha class is diminished. The Theta class also has alanine (Ala21) in the position of Leu23 (hGSTA1-1), however it is not an appropriate protein for direct comparison with Alpha class. This is due to a difference in the composition and conformation of βαβ motif between these classes. Structural alignment of βαβ motifs in Alpha and Theta classes gives an r.m.s.d. value of 4.4 Å between these two classes (Figure 38A). This value is much higher than the one obtained (1.7 Å) when aligning the overall structure of Alpha and Theta class. In addition, α1 of Theta class is solvent accessible (Figure 38B) and populated with charged residues (Table 5).

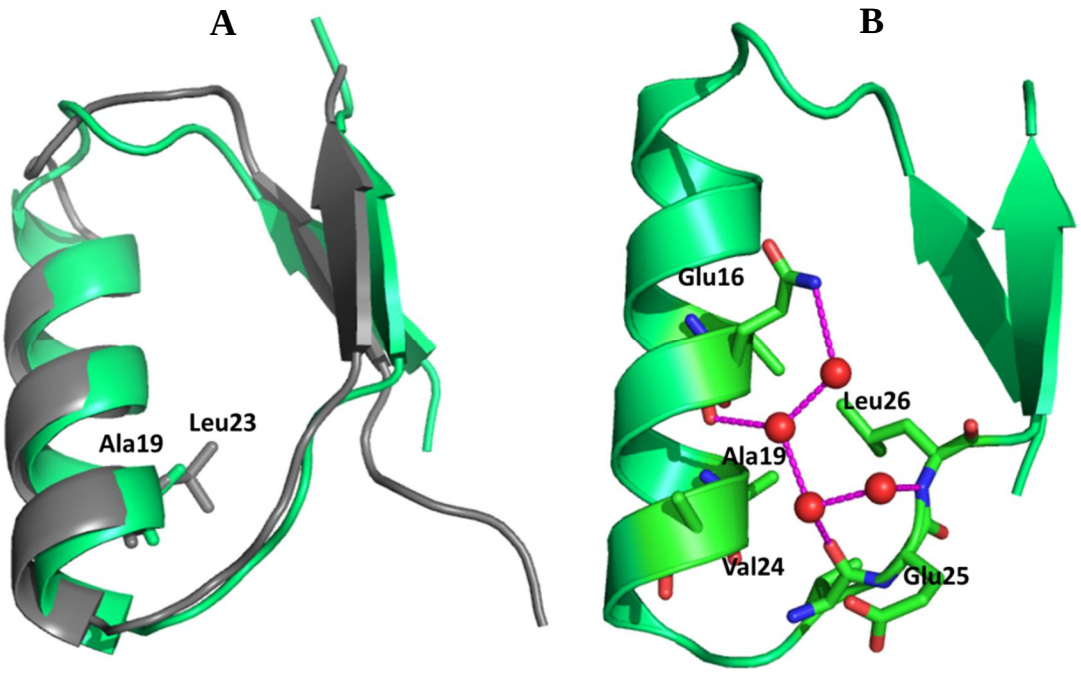


Figure 37: Differences in the βαβ motif of Alpha and Delta class GST.

Shown in A is the alignment of Alpha (grey) and Delta (green) βαβ which gives an r.m.s.d value of 5.9 Å. The Delta class has an alanine at position 19 (Ala19) which is topologically equivalent to Leu23 in the Alpha class. The βαβ motif in the Delta class is solvent accessible with four water molecules (red dots) playing a role in the packing of this motif via hydrogen bonding (pink dashed lines) (B). The coordinates used for Alpha and Delta were 1K3O and 1JLV, respectively. The images were generated using PyMol v0.99 (DeLano Scientific, 2006).

Table 5: Comparison of amino acid residues that are within 4 Å of Leu7, Ala19 and Ala21 side chains of class Alpha, Delta and Theta class GSTs, respectively.

Amino acid side chains that are interacting with Leu23 or topologically equivalent residues via van der Waals interaction are shown in red font.

Residues that are within 4 Å of Leu23 in IK3O (Alpha class)	**Residues that are within 4 Å of Ala19 IJLV(Delta class)**	**Residues that are within 4 Å of Ala21 in 2C3N (Theta class)**
Thr19	Val15	Val17
Arg20	Gln16	Trp18
Leu22	Met17	Phe20
Ala24	Thr18	Lys22
Ala25	Ala20	Lys23
Ala26	Val22	Asn24
Gly27	Gly23	Asp25
Val28	Val24	**Ile26**
Phe30	**Leu26**	**Phe28**
-	H_2O 220	H_2O 2017
-	H_2O 228	H_2O 2019
-	H_2O 230	-
-	H_2O 232	-

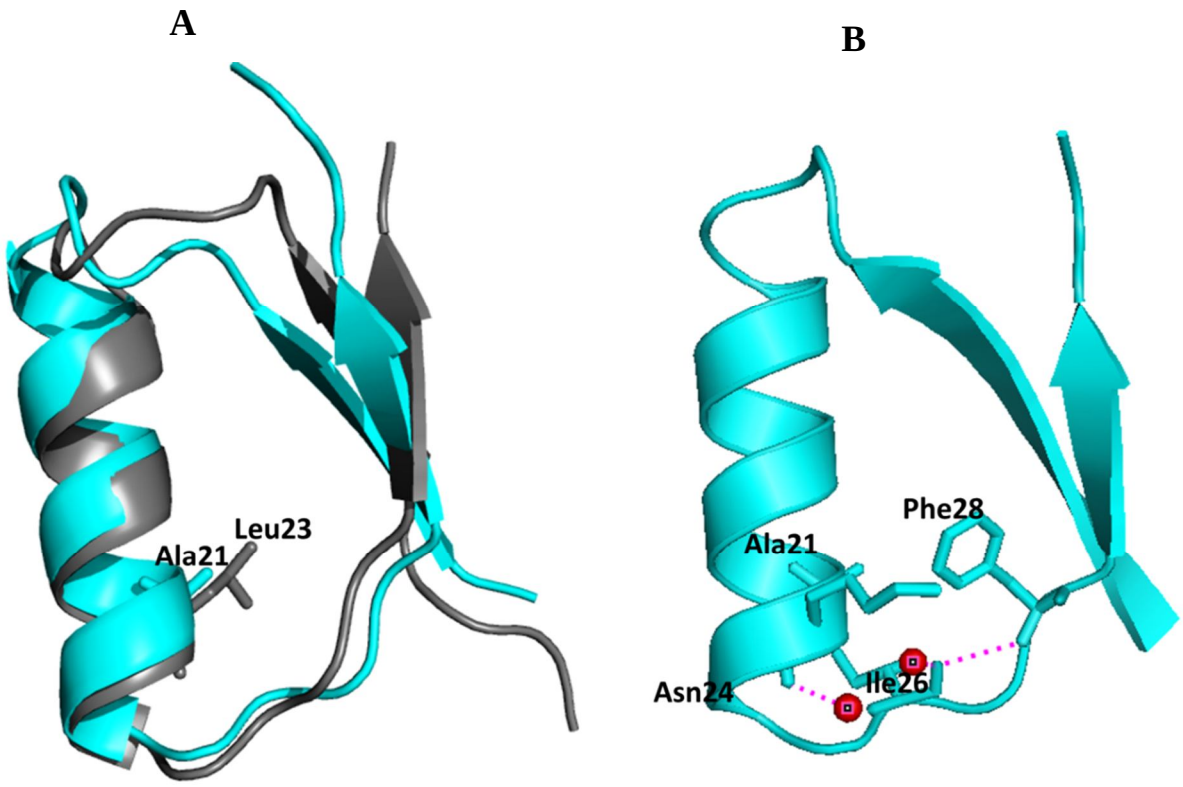


Figure 38: The βαβ motif of Theta class GST

Shown in A is the structural alignment of βαβ motif in Alpha (grey) and Theta (cyan) class. The r.m.s.d. value of 4.4 Å was obtained between Alpha and Theta class. The topological equivalent of Leu23 is Ala21 in Theta class (A). The hydrophobic core Theta class is solvent accessible as shown in B by two molecules (red dots) involved in the packing of the core via hydrogen boding (pink dashed lines). The coordinates used for the Alpha and Theta class are 1K3O and 2C3N, respectively. The images were generated using PyMol v0.99 (DeLano Scientific, 2006).

Another structural difference between Theta and Alpha class is the β2 conformation. It adopts an extended conformation despite being made up of the same number of residues. The loop region connecting β1 and α1 is longer in both Delta and Theta classes compared to the Alpha class. This could have different dynamic effects on the βαβ motif. Therefore, even if in Delta and Theta classes the topologically equivalent of Leu23 is alanine one should be very careful in assuming that replacing Leu23 with alanine in hGSTA1-1 may not affect its structure, function and stability.

Truncation of L7 and L23 side chains was predicted to create a cavity in the core of hGSTA1-1. Knowing that proteins are flexible molecules leads one to wonder how hGSTA1-1 is going to adapt to minimise cavity formation. Formation of cavities in the protein core disrupt the packing density hence can be unfavourable. If a cavity is indeed present as predicted, is it empty or occupied by solvent? As expected, both L7A and L23A structures showed cavities of 58 $Å^3$ and 40 $Å^3$, respectively, in the βαβ subdomain. This was not surprising since the local environments of L7A and L23A showed no major change in side chain conformation which is in agreement with similar r.m.s.d values between wild-type and its variants.

The cavities in the mutant structures were not occupied by solvent. A water molecule has a radius of 1.4 Å and could occupy a volume of 15 $Å^3$. There is controversy whether or not unoccupied cavities in protein cores are indeed empty. Wade and co-workers argue that removing a water molecule from the bulk of solvent to the protein cavity is energetically unfavourable due to loss in hydrogen bonding (Wade *et al.*, 1991) whereas Ernst *et al.* argues that "Nature abhors a vacuum" (Ernst *et al.*, 1995). A water molecule in a strictly non polar environment will be highly mobile making it almost impossible to be detected (Ernst *et al.*, 1995). Findings from Denisov and co-workers have challenged both arguments by showing that water molecules found in the interior of a protein are highly flexible regardless of whether they are involved in hydrogen bonding or not. They can exchange with the bulk solvent on a time scale of 15 ns (Denisov *et al.*, 1997).

The sizes of the cavities created by Leu7 and Leu23 mutations are large enough to accommodate water molecules and yet both structures show "empty" cavities. It is, therefore, likely that the "empty" cavity in the L7A and L23A structures may contain a highly mobile water molecule and its electron density is not detectable using X-ray diffraction. Achilonu

and co-workers have showed inclusion of two highly ordered water molecules in domain 1 of hGSTA1-1 when a hydrophobic core residue (Ile71) was mutated to alanine using crystallographic studies (Achilonu *et al.*, 2010). These water molecules contribute to the packing density of the I71A domain 1 and minimise the destabilisation effects caused by the I71A mutation. The presence of an "empty" cavity in the L7A and L23A structure is also supported by stability data of L7A and L23A which shows that both mutants are destabilised despite their native structure being maintained. With the local environments of both L7A and L23A not being strictly polar, the presence of ordered water molecules could compensate for the destabilisation due to the loss of van der Waals and hydrophobic interactions (Takano *et al.*, 2003).

4.2. The packing of βαβ motif is optimised for functional properties of hGSTA1-1

The glutathione binding site is located in the βαβ motif, with Leu7 and Leu23 residues being part of the hydrophobic core located below this site. The βαβ hydrophobic core in Delta class GST has been shown to play a significant role in maintaining the active site conformation and altering the packing of this hydrophobic core directly affects GSH binding. The effects of altering the packing of hydrophobic core are not limited to the G-site, but also impacts on the adjacent H-site conformation. It impacts on specificity and catalysis of Delta class GSTs (Vararattanavech *et al.*, 2006). In contrast to Delta GSTs, enzyme activity (Figure 25) and ligand binding properties (Figure 26) of Alpha GST suggests that the conformation of the G- and H-site does not seem to be effected by altering the packing on βαβ in hGSTA1-1.

Two observations have been made with regards to Leu7 and Leu23 residues when comparing the crystal structures of GSTs bound and unbound to the ligand. Firstly, both Leu7 and Leu23 residues are not directly involved in binding GSH. Secondly, Leu7 and L23 upon ligand binding forms van der Waals interactions indicating optimised packing for ligand binding (Le Trong *et al.*, 2002). The current study work demonstrates that the specific activity of the mutant enzymes in conjugating CDNB to GSH is reduced to 60% (for L7A) and to 52% (for L23A) in comparison to the wild-type and it is in agreement with Leu7 and Leu23 residues not being directly involved in GSH binding. Even though the structures of the mutants in complex with ligands such as *S*-hexyl glutathione were not determined, the interaction formed between Leu7 and Leu23 residues upon ligand binding may play a role in optimising enzyme activity.

4.3. L7A and L23A play a role in the stability of hGSTA1-1

Arrangements of hydrophobic residues in a polypeptide sequence are one of the key conserved features of protein with the same fold such as thioredoxin. This arrangement has been proposed to be as a resultant of specific secondary and tertiary structures (Bowie *et al.*, 1990a, 1990b; Baldwin and Matthews, 1994; Lim *et al.*, 1994). For studies such as this one, which aims to determine the effects of protein core packing by substituting one non polar residue for another and then measure the resulting change in protein stability, the choice of amino acid for mutation is crucial. The substitution of amino acid of different shape and size leads to steric hindrance which tends to complicate interpretation of experimental data. To overcome such a scenario, Leu7 and Leu23 were replaced with alanine.

Mutagenesis studies have shown that a single amino change acid may modulate protein stability (Matthews, 1991; Jackson *et al.*, 1993; Wallace *et al.*, 1998a, 2000; Sayed *et al.*, 2000; Nathaniel *et al.*, 2003; Bueno *et al.*, 2006a; Gabellieri *et al.*, 2008; Achilonu *et al.*, 2010; Balchin *et al.*, 2010). The extent of single amino acid mutation effects is dependent on its location, and the type of interaction that the given residue is involved in. Structural analysis of hGSTA1-1 shows that both Leu7 and Leu23 are involved in hydrophobic and van der Waals interactions with the surrounding residues in the βαβ motif of Trx fold. It is well documented that these types of interactions play essential roles in stabilising the protein cores. The possibility of the Leu7 and Leu23 residues being conserved for GST activity has been ruled out (Figure 25). To establish the role which Leu7 and Leu23 play with regards to Trx fold stability, urea-induced equilibrium unfolding studies of mutants were performed and the thermodynamic parameters were compared to that of the wild-type (Table 4).

It has been proposed that the folding pathway of structurally and functionally related proteins is conserved despite these proteins sharing low sequence similarity (Hollecker and Creighton, 1983; Plaxco *et al.*, 2000a, 2000b). This hypothesis has been shown to be true for a number of protein families (Gunasekaran *et al.*, 2001; Forsyth and Matthews, 2002). The equilibrium unfolding of GST superfamily has been extensively characterised with the class Alpha (Wallace *et al.*, 1998b), class Pi (Dirr and Reinemer, 1991; Erhardt and Dirr, 1995; Gildenhuys *et al.*, 2010b) and *Schistosoma japonicum* GST (Kaplan *et al.*, 1997) displaying a simple two-state mechanism. However, other members of the GST family such class Mu (Hornby *et al.*, 2000), class Beta (Sacchetta *et al.*, 1999; Abdalla and Hamed, 2006) Sigma

(Stevens *et al.*, 1998) and Ure2p (Galani *et al.*, 2002) display a more complex multi-state unfolding mechanism. The two-state unfolding mechanism involves a native dimer to unfolded monomers without significant population of an intermediate. So, with the two-state model, the unfolding curve should be monophasic and superimposable using variety of spectroscopic tools. The transition should be concentration dependent. This is the case for wild-type hGSTA1-1 (Figure 33). The unfolding curves obtained from L7A and L23A data are also monophasic, concentration dependent but the spectroscopic probes do not overlay. There could be two possible causes for these curves not to overlay. The first reason could be due to the fluorescence probe being more sensitive to structural changes in the Trx-like fold since it is a local probe whereas the far-UV CD probe is a global probe. If this possibility is correct, then the unfolding of L7A and L23A would meet the criteria set out to analyse the data using a two-state model. The second possibility for data obtained using multiple spectroscopic probes in L7A and L23A not to overlay could be a result of equilibrium unfolding of hGSTA1-1 deviating from a two-state to a multi-state in these proteins. If this is the case, fitting the L7A and L23A to a two-state model will yield incorrect thermodynamic parameters. Subsequently, the L7A and L23A unfolding data was therefore analysed using both two-state and multi-state model. The two-state fit on the wild-type, L7A and L23A resulted in good fit (Figure 35) with R^2 value of 0.99. The obtained thermodynamic parameters for the mutants are compared to that of wild-type hGSTA1-1.

The chemically-induced three-state unfolding of dimeric proteins may be described in two models (Rumfeldt *et al.*, 2008). The native dimer may dissociate into 2 monomers before unfolding where the intermediate is assumed to be monomeric ($N_2 \leftrightarrow 2I \leftrightarrow 2U$). In the second model, native dimer unfolding precedes dissociation into monomers giving rise to a dimeric intermediate as shown in equation 15 (see section 2.10.4). Fabrini *et al.* (2009) showed that the likelihood of monomeric intermediates in GSTs unfolding is scarce and may occur at low (nM) concentration levels. Taking into account that the current hGSTA1-1 unfolding study was performed at μM concentration range and Figure 35 shows protein concentration dependence of the unfolding transition region, equation 15 was used to analyse the three-state unfolding of L7A and L23A proteins. The three-state fit did not fit the L7A and L23A data obtained in this study as demonstrated by poor R^2 values (less than 0.5 for both the proteins). This does not conclusively mean that the urea-induced unfolding of L7A and L23A does not occur via three-state, it could also imply that the probes used in this work

did not detect a significant population of an intermediate species. Studies on urea-induced unfolding of GSTs have previously shown failure of far-UV CD and fluorescence spectroscopy to distinguish dimeric form monomeric intermediate species (Alves *et al.*, 2006; Parbhoo *et al.*, 2011).

The determination of the Gibbs free energy change in the absence of denaturant $\Delta G(H_2O)$ is used to measure the conformational stability of a protein. The $\Delta G(H_2O)$ value (23.88 kcal.mol^{-1}) obtained for the wild-type is in agreement with that of published hGSTA1-1 (Wallace *et al.*, 1998a, 1998b, 2000; Sayed *et al.*, 2000; Wallace *et al.*, 2000; Mosebi *et al.*, 2003). The $\Delta G(H_2O)$ for the wild-type is also consistent with $\Delta G(H_2O)$ calculated for dimeric proteins (26.6 kcal.mol^{-1}) (Neet and Timm, 1994; Rumfeldt *et al.*, 2008). This value is almost halved in L7A and L23A hGSTA1-1. It corresponds to experimentally determined values for monomeric CLIC1 and Grx2 proteins (Fanucchi *et al.*, 2008; Parbhoo *et al.*, 2011; Achilonu *et al.*, 2012) and to values calculated for monomeric proteins (6 – 14 kcal.mol^{-1}) (Neet and Timm, 1994; Rumfeldt *et al.*, 2008). The reduced $\Delta G(H_2O)$ of the L7A and L23A proteins correlates with reduced T_m values obtained using temperature as a denaturant (Figure 27). The mutants were also more prone to thermolysin digestion when compared to the wild-type further demonstrating lower stability against urea. The reduced $\Delta G(H_2O)$ value obtained for both mutants in comparison to those of the wild-type demonstrate that both Leu7 and Leu23 plays a role in the stability of hGSTA1-1.

Another important parameter that is used to determine the role of a particular residue with regards to its contribution in protein stability is the *m*-value. The *m*-value for the wild-type (4.06 kcal.mol^{-1}.M^{-1}urea) is in agreement with published values for hGSTA1-1 (Wallace *et al.*, 1998a, 1998b, 2000; Sayed *et al.*, 2000; Mosebi *et al.*, 2003). The *m*-values obtained for both mutants are low; 1.46 and 1.06 (kcal.mol^{-1}.M^{-1}urea) for L7A and L23A proteins, respectively. The *m*-value is dependence of the free of energy unfolding of denaturant and is related to the difference in the accessible solvent area (ΔASA) between the native and unfolded states (Myers *et al.*, 1995; Shirley, 1995). The theoretical value for ΔASA of the wild-type and mutants is the same (40199 Å^2) when calculated using the equation by Myers *et al.* (1995):

$$\Delta ASA = (n \times 93) - 907 \qquad \textbf{Equation 32}$$

where n is the number of residues that make up the protein molecule. With the *m*-value being proportional to the ΔASA, the theoretical *m*-values (*m*) for the wild-type should be the same (40199 $Å^2$) as for the L7A and L23A proteins when calculated using the equation below:

$$\boldsymbol{m = 374 + (0.11 \times \Delta ASA)} \qquad \textbf{Equation 33}$$

The correlation between the theoretical and experimentally determined *m*-values for the wild-type further infers two-state unfolding transition for hGSTA1-1. The difference between calculated and experimentally obtained *m*-values for L7A (70%) and L23A (78%) is very large. Reduced *m*-values could suggest (a) a difference in the exposed surface area between the native and unfolded state, or (b) changes in the interaction of a denaturant with the protein molecules in the denatured state or (c) a deviation from two-state to multi-state unfolding transition (Arakawa and Timasheff, 1984; Pace, 1986; Shortle, 1992; Shirley, 1995; Soulages, 1998). It is unlikely that the reduced *m*-values for the L7A and L23A could be due to suggestion (a), because (i) both mutations were non-disruptive (Figure 19) with the hydrodynamic volume similar to that of wild-type under native conditions (Figure 14) and (ii) ANS binding properties of the wild-type and mutants in the native state are similar suggesting similar hydrophobic surface exposure in both wild-type and mutant structures (Figure 26). This would mean that the solvent accessible area between wild-type and its variants is also unchanged. Also, the possibility of suggestion (b) to be true for this study is unlikely because urea completely unfolds the wild-type and mutants with all the proteins displaying similar properties in the denatured state (Figure 15 and Figure 16). Furthermore, unfolding GST with urea or guanidinium chloride results in similar $\Delta G(H_2O)$ values despite both chemicals using different mechanisms to unfold proteins. This suggests that there are no specific binding sites as a these denaturants (Dirr and Reinemer, 1991). The first two suggestions for possible explanation of the low *m*-values obtained for the mutants were ruled out. The possibility of L7A and L23A urea-induced unfolding deviating from two-state to multi-state was examined.

The experimentally determined *m*-values for mutants are reduced to 36% (for L7A) and to 26% (for L23A) in comparison to that of wild-type (Table 4). The next question is does the reduced *m*-values indicate deviation from a two-state to multi-state unfolding transition in the mutants indicating a loss in cooperative folding of hGSTA1-1? This led to the need to question the validity of two-state model used to analyse L7A and L23A unfolding data. A

deviation from a two-state to multi-state unfolding transition is consistent with reduced *m*-values obtained for L7A and L23A. This deviation should result in the accumulation of intermediate species in the transition region of the unfolding curve. The inability to distinguish the monomeric and dimeric intermediates is one of the major limitations of far-UV CD and fluorescence spectroscopy in studying GST unfolding. Previous studies have shown that both these species have similar spectroscopic properties (Stevens *et al.*, 1998; Hornby *et al*., 2000, 2002). Monitoring unfolding curves of the wild-type and its variants with ANS binding and light scattering did not suggest accumulation of intermediate species. The noticeable difference in the unfolding curves when monitored by ANS binding was that both the mutants displayed a loss in ANS binding in the pre-transition region whereas the wild-type shows loss at a much higher urea concentration. The loss in ANS binding at low urea concentration has been attributed to the highly flexible α helix 9 unfolding prior to global unfolding of the protein (Wallace and Dirr, 1999; Alves *et al.*, 2006). This occurrence is evident in other mutants of hGSTA1-1 that showed stabilisation under similar conditions when compared to the wild-type.

The protein concentration dependence of the transition region for all the proteins suggests a significant population of dimeric species in the transition region (Figure 36). So, even if the pathway changed from a two-state to multi-state, the intermediate species are unlikely to be monomeric. The presence of a monomeric intermediate will be interesting given the rigidity of the Trx fold and additionally the interactions at the dimer interface of the wild-type are maintained in both mutants. The spectroscopic data (Figure 33) and low *m*-values (Table 4) obtained using two-state fit suggested a deviation from two-state to multi-state unfolding for the mutants. Recent studies suggest the exercise of caution when interpreting unexpectedly low *m*-values obtained from two-state unfolding analysis (Spudich and Marqusee, 2000; Parbhoo *et al.*, 2011) because the use of an incorrect model leads to a large underestimation of the thermodynamic parameters (Soulages, 1998). In this work, attempts were made to analyse L7A and L23A unfolding data using a three-state fit however, this model gave poor fits. Other studies have also reported unexpectedly low *m*-values with an unacceptable three-state correlation coefficient fit (Wallace *et al.*, 1998a, 2000; Hornby *et al.*, 2000; Nathaniel *et al.*, 2003; Alves *et al.*, 2006; Achilonu *et al.*, 2010). At this point, it is uncertain whether the L7A and L23A mutations resulted in the hGSTA1-1 deviating from two-state to three-state. However, it is clearly evident that these mutations have affected the stability of hGSTA1-1.

This is strengthened by the correlation of various methods (thermal- and urea-induced unfolding and pulse proteolysis) used in this study which all indicate that the stability of both the L7A and L23A proteins is reduced.

Site directed mutagenesis studies on Trx-like domain in hGSTA1-1 and Trx proteins have demonstrated that Van der Waals interactions in the hydrophobic core are essential in directing the protein folding process and in maintaining the stability of the hydrophobic cores that stabilise the native protein structure via correct and optimised packing (Wallace *et al.*, 1998, 2000, Tasayco *et al*. 2000, Santos *et al.*, 2007, 2009; Achilonu *et al.*, 2010; Balchin *et al.*, 2010, Parbhoo *et al.*, 2011). Furthermore, these studies have also shown that hydrophobic residues in the protein cores impacts on the protein folding via hydrophobic interaction network and disrupting this network can adversely affect protein structure, function and/or stability. Tasayco *et al*. (2000) and Santos *et al.* (2009) showed that both α1 and β1 plays a key role in the folding of the Trx-like domain. The folding of this domain depends on the interactions between N- and C-subdomain (Figure 6). The initial the folding of Trx-like domain is driven by hydrophobic network between β1 and β3, whereas the final step involves key interactions between α1 and α3. Therefore, the folding and stability of Trx domain is maintained via key hydrophobic interactions between N- and C-subdomain.

The L7A structure shows that mutation of Leu7 results in the loss of van der Waals interaction between Leu 7 (β1) and Val58 (β3) (Figure 21). The loss of this interaction between the two subdomains, together with altered packing of N-subdomain, have diminished the cooperativity of unfolding hGSTA1-1 as indicated by reduced *m*-values. The L23A structure only shows the altered packing of the N-subdomain and not loss in the interaction between the N- and C-subdomian. However, the unfolding cooperativity of L23A h GSTA1-1 is also greatly diminished. Therefore, the topologically conserved leucine residues in N-subdomain of GST contributes to the cooperative folding of this protein via hydrophobic interaction network between N- and C-subdomain and/or interaction within each subdomain. Disruption on this network impacts of the cooperativity of hGSTA1-1 folding.

Taking into the account that the current study suggests a deviation of L7A and L23A unfolding from a two-state to a multi-state state in comparison to the wild type without evidence of monomeric intermediate formation (Figure 36), it is likely that as the dimeric L7A and L23A protein unfolds, the N- and C-subdomain unfolds in an independent manner.

This suggestion is further supported by Figure 34 which shows great loos in ANS binding in both L7A and L23A protein at lower urea concentration in comparison to the wild-type.

A possible explanation of destabilisation of the L7A and L23A mutants is obtained in the local environment of the mutation site. Mutation of these residues resulted in 58 $Å^3$ (L7A) and 40 $Å^3$ (L23A) cavities within βαβ motif. The burial of non-polar side chains contributes to protein stability in two ways: (a) they contribute to the hydrophobic effect and (b) in packing interactions through van der Waals forces stabilising the native state of the protein (Matthews, 1991; Kono *et al.*, 2000; Bueno *et al.*, 2006). Molecular dynamic simulation and protein engineering studies have shown that it is difficult to separate the contribution of hydrophobic effect from packing interactions (Kono *et al.*, 2000; Bueno *et al.*, 2006a, 2006b). The L7A mutation resulted in the L7A structure losing van der Waals interaction of Ala7 with Phe30 and Val58. In contrast, the L23A mutation did not result in loss of any van der Waals interactions within the βαβ motif. Both structures showed "empty" cavities within the local environment of the mutation site (Figure 21 and Figure 23). It, therefore, seems mostly likely that the destabilisation is due to the creation of these cavities, which resulted in disrupting the energetically favoured packing within this motif. In cases where the local environment of the mutation site is loosely packed, formation of a cavity in such an environment induces extensive structural rearrangements which in turn minimize the destabilisation effects (Jackson *et al.*, 1993; Gabellieri *et al.*, 2008). Even though the core packing was disrupted by cavity formation in both mutant structures, the Trx domain seems tolerant to the changes in the hydrophobic core since very little movement of the amino acid residues occurred (Figure 21A and Figure 23A). This view is supported by acceptable B-factors (Table 3) obtained for L7A and L23A structures which signify confidence in the position of their atoms. B-factors are determined by spatial vibration in the local packing density in proteins and they reflect the uncertainty in atom positions in the model (Radivojac *et al.*, 2004; Halle, 2002). B-factor values greater than 60 $Å^2$ implies disorder or uncertainty in position of the atom.

The role of hydrophobic cores in Trx-like domains has been studied in adGSTD4-4 and hGSTP1-1. In adGSTD4-4, Leu6, Leu33 and Ile52 are in the core of the Trx-like domain located in the highly flexible loop region between β1-α1, β2-α2 and α2-β3, respectively. These amino acids were mutated to alanine. The Leu6 mutation affected the active site

conformation, which altered the substrate specificity, decreased the stability and the refolding rate of the protein. On the other hand, Leu33 and Ile52 altered the intrinsic fluorescence properties without loss in enzyme activity (Vararattanavech *et al.*, 2006). When Leu33 was mutated to phenylalanine, the protein stability was increased; however, enzyme activity was lost (Vararattanavech and Ketterman, 2003). Phe6 is topologically equivalent of Leu7 in hGSTP1-1 and currently the role of this hydrophobic residue in terms of structure, function or stability in Pi class has not been investigated. It would be interesting to know whether the effects of mutating Phe6 will be similar to L7A, given the difference in the topology of the βαβ motif between these two proteins.

In hGSTP1-1, Leu21 is located in α1 and is involved in inter-domain interactions with α6, and the L21A mutation destabilised the protein without affecting its refolding rates (Stenberg *et al.*, 2000). Leu21 is not a topological equivalent of Leu23 in hGSTA1-1. The thermodynamic profiles obtained for L7A and L23A show a similar trend when compared to other mutants when their hydrophobic core residues within domain 1 and even those located in the hydrophobic core of domain 2 were mutated. These findings highlight the role of hydrophobic cores in modulating the properties of the proteins via correct core packing which is optimised for enzyme function.

The driving force in protein folding is the hydrophobic effect, which results in hydrophobic residues being located in protein cores (Dill, 1990; Mirny and Shakhnovich, 1999). This does not exclude the presence of polar residues in protein cores. Polar residues may be located in the protein core for functional, structural and stability purposes. The unfavourable energy associated with the burial of polar residues in a protein core is compensated by the buried polar residues forming intermolecular hydrogen bonds in the folded protein. Structural alignment of Trx and hGSTA1-1 proteins leads to an interesting observation regarding position 7 in β1 of the βαβαββα fold (Figure 39). In the GST family, this position is reserved for leucine residue (Figure 7) whereas in Trx family, position 7 is occupied by a topologically conserved aspartic acid residue (Asp26) (Eklund *et al.*, 1991; Gleason, 1992).

The role of Asp26 in Trx proteins has been investigated using site-directed mutagenesis, crystallographic and molecular dynamic simulation studies (Langsetmo *et al.*, 1991; Gleason, 1992; Chivers and Raines, 1997; Bolon and Mayo, 2001; Roos *et al.*, 2009). Mutation of Asp26 to alanine or isoleucine does not disrupt the global protein structure and results in

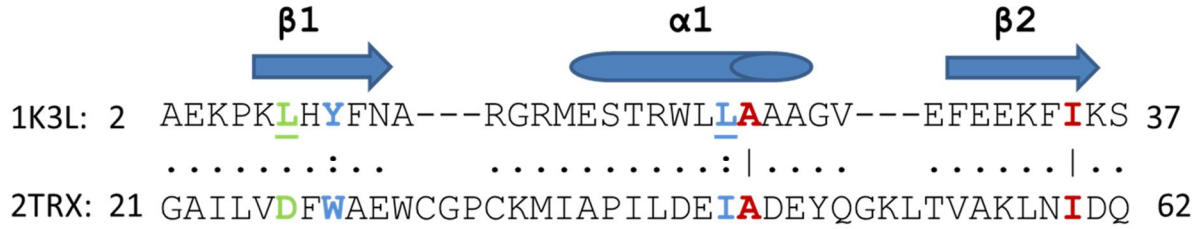


Figure 39: Structural alignment of hGSTA1-1 (1K3L) and Trx (2TRX)

The Leu7 and Leu23 residues are shown residue in bold underlined font. Residues that are structurally equivalent and identical between the two proteins are shown red font whereas residues that are structurally equivalent and similar residues are shown in light blue font. The structurally equivalent of Leu7 in Trx protein is shown in green font. The PDB codes are shown in parentheses. The alignment was performed with 3DCoffee (O'Sullivan *et al.*, 2004).

increased stability relative to wild-type Trx (Langsetmo *et al.*, 1991; Gleason, 1992; Bolon and Mayo, 2001). Furthermore, the amide hydrogen-deuterium exchange of D26I and wild-type Trx showed that the protons in D26I were more protected than that of the wild-type Trx inferring increased stability in D26I (Bolon and Mayo, 2001). The increased stability of D26A or D26I mutants is attributed to the removal of the charge in the protein core in which its hydrogen bonding was unsatisfied.

Asp26 has been proposed to play a key role in the catalytic function of Trx by activating the C-terminal cysteine (Cys32) of the C-G-P-C catalytic motif (Qin *et al.*, 1995; Chivers and Raines, 1997; Menchise *et al.*, 2001). However, Gleason (1992) and Roos *et al.* (2009) demonstrated that D26A mutant is able form a complex with its substrate and catalyse its reduction. The catalytic efficiency (k_{cat}/K_M) of D26A for its substrate (thioredoxin reductase) is reduced in comparison to the wild-type. This is mainly due to D26A displaying 10-fold increase in the K_M value and thus the author suggested that Asp26 is conserved in most Trxs to provide a unique charge distribution in the active site region which results in optimised interactions with other proteins (Gleason, 1992).

Site-directed mutagenesis of Asp26 in Trx also demonstrates a stability-function trade-off in proteins at the level of individual mutation. This phenomenon has been shown for a number of proteins where a charged or polar residue located in the hydrophobic core of proteins is substituted for a hydrophobic residue results in increased stability at the expense of activity (Meiering *et al.*, 1992; Schreiber *et al.*, 1994; Shoichet *et al.*, 1995; Beadle and Shoichet, 2002; Tokuriki *et al.*, 2008). The explanation for stability-function trade-off in proteins is that active site formation in proteins may not be thermodynamically favourable.

A polar or charged residue in the active site has more to do with a functional role, such as maintaining buried charges or cavities in a protein's interior. This, in turn, makes buried polar or charged residues poorly optimised for stability (Bolon and Mayo, 2001; Schueler-Furman and Baker, 2003). It is widely accepted that Trxs are progenitors of GST (Martin, 1995; Pan and Bardwell, 2006; Pedone *et al.*, 2010). The sequence identity between Trx and GST proteins is very low and it is consistent with the highly or ever evolving Trx-like domain. Furthermore, the GST family, when viewed in context Trx-like family is considered an outlier in terms of function and active site construction. It is attractive to suggest that Asp26 was mutated to a leucine residue (Leu7) during evolution to enhance stability of the

N-subdomain in GST. However Asp26 is not the only residue that became substituted in the GST amino acid sequence (Figure 39). In fact, GSTs lack the C-G-P-C catalytic motif and has a G-site in its equivalent position (N-terminus of α1 and nearby loop region) which employs a serine, tyrosine or cysteine residue as a nucleophile to attack the GSH bound to the G-site. With β1 located within the vicinity of the active site (Trx) or G-site (GST), its conformation may directly affect the functioning of the enzyme. Furthermore, composition of amino acid residue in β1 should complement the function of the enzyme. Taking into account that position 7 of β1 is reserved for aspartic residue in the Trx family and for leucine residue in the GST family, it will be interesting to observe how hGSTA1-1 will respond to L7D mutation in terms of structure, function and stability. In Trx proteins, replacement of a leucine residue for aspartic acid has been tolerated despite this residue introducing a charge in the protein core which can be thermodynamically unfavourable (Hellinga *et al.*, 1992).

In other Trx-like proteins such as DsbA and the PDI family, position 7 is also occupied by a topologically conserved glutamic acid (Glu24 numbered according to DsbA) (Martin *et al.*, 1993; Jacobi *et al.*, 1997; Ellgaard and Ruddock, 2005; Hatahet and Ruddock, 2009). Glu24 (or its equivalent) forms a salt bridge with a conserved Lys58; however, mutation of Glu24 or Lys58 does not affect the stability or function of DsbA (Jacobi *et al.*, 1997). Jacobi and co-workers proposed that since the Glu24-Lys58 interaction is conserved across the family, it may have a role in directing the folding pathway without stabilising the native structure of DsbA (Jacobi *et al.*, 1997). Such a role has been confirmed for Asp13-Lys18 salt bridge in Trx proteins (Huber *et al.*, 2010). Mutation of Asp13 to asparagine disrupts the salt bridge but does not destabilise or affect the refolding rates of the enzyme in the oxidised form. In contrast, the reduced form of the enzyme was slightly stabilised (0.5 kcal/mol) with reduced folding rates (Huber *et al.*, 2010). These results suggest the Asp13-Lys18 salt bridge plays a role in initial folding events by restricting the number of possible conformations of early folding intermediates resulting in acceleration of folding. Furthermore, the Asp13-Lys18 salt bridge may not be essential to stabilise the fully folded state (Huber *et al.*, 2010). The reduced form of Trx is the physiologically relevant form of the enzyme. Therefore Huber and co-workers highlight the importance of studying the correct form of the enzyme (Huber *et al.*, 2010).

PDI consists of four domains, each with a Trx-like fold but not all domains are catalytically active. In the PDI family, the equivalent of Leu7 is Glu47 and this residue forms a salt bridge with Lys81. Interestingly, the Glu47-Lys81 salt bridge is only present in the catalytically active domain. Mutation of Glu47 or Lys81 reduces the activity of the enzyme (Ellgaard and Ruddock, 2005; Appenzeller-Herzog and Ellgaard, 2008; Hatahet and Ruddock, 2009). Non-catalytic domains of PDI lack the conserved Glu47-Lys81 salt bridge; therefore, the Glu47-Lys81 salt bridge plays a role in substrate binding during the isomerisation reaction (Ellgaard and Ruddock, 2005). It, therefore, appears that position 7 of the Trx-like fold is conserved. However, the type of a residue differs from family to family. Therefore, the current work, when viewed together with other studies on β1 of the Trx-like fold provide the basis for β1 playing a role in the architecture of the Trx-like fold, and that its role is versatile depending on the specific function of the enzyme. To further support this view, Figure 39, shows that in β1 position 9 is occupied by an conserved catalytic residue (Tyr9) in Y-GSTs (Atkinson and Babbitt, 2009a). The equivalent of Tyr9 in Trx family is a conserved tryptophan residue (Trp28). In Y-GSTs, Tyr9 catalyses the activation of GSH in the GST reaction mechanism (Armstrong, 1991, 1994, 1997). On the other hand, Trp28 plays an important role in the stability of Trx proteins despite being part of the active site. It is not involved in any intramolecular hydrogen bonds with active site residues (Roos *et al.,* 2007, 2010). Mutation of Trp28 to alanine largely affects the structure of the Trx protein. Trp28A monomeric structure becomes partially unfolded and molten globule–like and its quaternary structure shows swapped domains (Garcia-Pino *et al.*, 2009). The partial unfolding of the Trp28A structure also reduces its redox potential in comparison to the wild-type protein (Roos *et al.,* 2007, 2010).

The current study has shown that the topologically conserved Leu23 in α1 plays a role in stabilising the N-subdomain of GST. The mutation of this residue to alanine leads to cavity formation in the βαβ motif which in turn affects the stability of the protein without compromising the structure of the Trx-like fold. The equivalent of Leu23 in Trx proteins is Ile45. Ile45 is located within a hydrophobic cluster in the core of the Trx protein. Mutation of this residue to valine which is believed to be conservative also largely impacts of the stability (Godoy-Ruiz *et al.,* 2005). In addition, mutation of other residues (Ile38 and Ile41) which are also located within the Ile45 hydrophobic cluster also gave similar results despite not being conserved. The crystal structures of isoleucine mutants were not determined. It

will be interesting to know whether the structural effects on the surrounding environment of the mutation will be similar to that of the L23A structure. Other studies on hGSTA1-1 have shown that isoleucine substitution to valine or alanine affects the stability of the Trx-like domain but the structural response to mutation may vary depending on the residue used to replace the isoleucine (Achilonu *et al.*, 2010). Findings from the aforementioned studies infer that the hydrophobic core can be highly optimised for specific interactions and disruption of these interactions may affect the stability but not the structure of the Trx-like fold.

4.4. Conclusions

The present study highlights the role of two topologically conserved leucine residues (Leu7 and Leu23 in hGSTA1-1) in the stability of the N-subdomain of GSTs. These bulky hydrophobic residues are important for the optimised packing of the βαβ motif. When Leu7 and Leu23 are replaced with a less bulky residue (Ala), the packing of the βαβ motif is disrupted as indicated by the presence of cavities in the mutant structures. It is interesting to note that although the Trx-like domain is highly adaptable, both L7A and L23A structures show very little modification of side chain conformations of the neighbouring residues of the mutation site to overcome cavity formation. The modification of the N-subdomain in L7A and L23A proteins significantly affected the stability of hGSTA1-1. The low *m*-values obtained for the L7A and L23A proteins strongly suggest deviation from a two-state unfolding mechanism reported for hGSTA1-1 to a multi-state unfolding mechanism and diminished cooperativity in unfolding of hGSTA1-1. Other studies on conserved hydrophobic residues located in the N-subdomain of Trx-like fold, showed that mutation of conserved hydrophobic residues may affect the stability and/or functioning of the enzyme (Stenberg *et al.*, 2000; Vararattanavech and Ketterman, 2003; Vararattanavech *et al.*, 2006; Roos *et al.*, 2007, 2010; Garcia-Pino *et al.*, 2009; Balchin *et al.*, 2010). In the case of Leu7 and Leu23A mutation in hGSTA1-1, the specific activity was reduced implying that the core packing in the N-subdomain is also optimised for functioning of hGSTA1-1. Taking into account that conserved hydrophobic core residues may also be part of the folding nucleus or involved in the early folding events, future work may involve determining whether Leu7 and Leu23 is involved in such a role.

5. References

Abdalla, A. M., and R.R. Hamed. 2006. Multiple unfolding states of glutathione transferase from Physa acuta (Gastropada: Physidae). *Biochem Biophys Res Commun* **340**:625-632.

Achilonu, I., S. Gildenhuys, L. Fisher, J. Burke, S. Fanucchi, B.T. Sewell, M. Fernandes, and H.W. Dirr. 2010. The role of a topologically conserved isoleucine in glutathione transferase structure, stability and function. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**:776-780.

Achilonu, I., S. Fanucchi, M. Cross, M. Fernandes, and H.W. Dirr. 2012. Role of individual histidines in the pH-dependent global stability of human chloride intracellular channel 1. *Biochemistry* **51**:995-1004.

Allocati, N., E. Casalone, M. Masulli, I. Ceccarelli, E. Carletti, M.W. Parker, and C. Di Ilio. 1999. Functional analysis of the evolutionarily conserved proline 53 residue in Proteus mirabilis glutathione transferase B1-1. *FEBS Lett* **445**:347-350.

Alves, C.S., D.C. Kuhnert, Y. Sayed, and H.W. Dirr. 2006. The intersubunit lock-and-key motif in human glutathione transferase A1-1: role of the key residues Met51 and Phe52 in function and dimer stability. *Biochem J* **393**:523-528.

Ambler, R.P., and R.J. Meadway. 1968. The use of thermolysin in amino acid sequence determination. *Biochem J* **108**:893-895.

Appenzeller-Herzog, C., and L. Ellgaard. 2008. The human PDI family: versatility packed into a single fold. *Biochim Biophys Acta* **1783**:535-548.

Arakawa, T., and S.N. Timasheff. 1984. Protein stabilization and destabilization by guanidinium salts. *Biochemistry* **23**:5924-5929.

Armstrong, R.N. 1991. Glutathione *S*-transferases: reaction mechanism, structure, and function. *Chem Res Toxicol* **4**:131-140.

Armstrong, R.N. 1994. Glutathione S-transferases: structure and mechanism of an archetypical detoxication enzyme. *Adv Enzymol Relat Areas Mol Biol*. **69**:1-44.

Armstrong, R.N. 1997. Structure, catalytic mechanism, and evolution of the glutathione transferases.*Chem Res Toxicol* **10**:2-18.

Atkinson, H.J., and P.C. Babbitt. 2009a. An atlas of the thioredoxin fold class reveals the complexity of function-enabling adaptations. *PLoS Comput Biol* **5**:e1000541.

Atkinson, H.J., and P.C. Babbitt. 2009b. Glutathione transferases are structural and functional outliers in the thioredoxin fold. *Biochemistry* **48**:11108-11116.

Balchin, D., S. Fanucchi, I. Achilonu, R.J. Adamson, J. Burke, M. Fernandes, S. Gildenhuys, and H.W. Dirr. 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.

Baldwin, E.P., and B.W. Matthews. 1994. Core-packing constraints, hydrophobicity and protein design. *Curr Opin Biotechnol* **5**:396-402.

Banci, L., I. Bertini, S. Ciofi-Baffoni, T. Hadjiloi, M. Martinelli, and P. Palumaa. 2008. Mitochondrial copper(I) transfer from Cox17 to Sco1 is coupled to electron transfer. *Proc Natl Acad Sci U S A* **105**:6803-6808.

Bashton, M., and C. Chothia. 2007. The generation of new protein functions by the combination of domains. *Structure* **15**:85-99.

Beadle, B.M., and B.K. Shoichet. 2002. Structural bases of stability-function tradeoffs in enzymes. *J Mol Biol* **321**:285-296.

Berndt, C., C.H. Lillig, and A. Holmgren. 2008. Thioredoxins and glutaredoxins as facilitators of protein folding. *Biochim Biophys Acta* **1783**:641-650.

Bhutani, N., and J.B. Udgaonkar. 2003. Folding subdomains of thioredoxin characterized by native-state hydrogen exchange. *Prot Sci* **12**:1719-1731.

Bjornestedt, R., G. Stenberg, M. Widersten, P.G. Board, I. Sinning, T.A. Jones, and B. Mannervik. 1995. Functional significance of arginine 15 in the active site of human class Alpha glutathione transferase A1-1. *J Mol Biol* **247**:765-773.

Board, P.G., M. Coggan, and D.M. Woodcock. 1992. The human Pi class glutathione transferase sequence at 12q13-q14 is a reverse-transcribed pseudogene. *Genomics* **14**:470-473.

Board, P.G., R.T. Baker, G. Chelvanayagam, and L.S. Jermiin. 1997. Zeta, a novel class of glutathione transferases in a range of species from plants to humans. *Biochem J* **328**:929-935.

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

Bolon, D.N., and S.L. Mayo. 2001. Polar residues in the protein core of *Escherichia coli* thioredoxin are important for fold specificity. *Biochemistry* **40**:10047-10053.

Bowie, J.U., J.F. Reidhaar-Olson, W.A. Lim, and R.T. Sauer. 1990b. Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* **247**:1306-1310.

Bowie, J.U., N.D. Clarke, C.O. Pabo, and R.T. Sauer. 1990a. Identification of protein folds: matching hydrophobicity patterns of sequence sets with solvent accessibility patterns of known structures. *Proteins* **7**:257-264.

Bruns, C.M., I. Hubatsch, M. Ridderström, B. Mannervik, and J.A. Tainer. 1999. Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. *J Mol Biol* **288**:427-439.

Bueno, M., L.A. Campos, J. Estrada, and J. Sancho.2006a. Energetics of aliphatic deletions in protein cores. *Protein Sci* **15**:1858-1872.

Bueno, M., N. Cremades, J.L. Neira, and J. Sancho. 2006b. Filling small, empty protein cavities: structural and energetic consequences. *J Mol Biol* **358**:701-712.

Buetler, T.M., and D.L. Eaton.1992. Complementary DNA cloning, messenger RNA expression, and induction of alpha-class glutathione S-transferases in mouse tissues. *CancerRes* **52**:314-318.

Caccuri, A.M., G. Antonini, N. Allocati, C. Di Ilio, F. De Maria, F. Innocenti, M.W. Parker, M. Masulli, M. Lo Bello, P. Turella, G. Federici, and G. Ricci. 2002a. GSTB1-1 from Proteus mirabilis: a snapshot of an enzyme in the evolutionary pathway from a redox enzyme to a conjugating enzyme. *J Biol Chem*. **277**:18777-18784.

Caccuri, A.M., G. Antonini, N. Allocati, C. Di Ilio, F. Innocenti, F. De Maria, M.W. Parker, M. Masulli, F. Polizio, G. Federici, and G. Ricci. 2002b. Properties and utility of the peculiar mixed disulfide in the bacterial glutathione transferase B1-1. *Biochemistry*. **41**:4686-4693.

Casalone, E., N. Allocati, I. Ceccarelli, M. Masulli, J. Rossjohn, M.W. Parker, and C. Di Ilio. 1998. Site-directed mutagenesis of the *Proteus mirabilis* glutathione transferase B1-1 G-site. *FEBS Lett* **423**:122-124.

Chae, H.Z., S.J. Chung, and S.G. Rhee. 1994. Thioredoxin-dependent peroxide reductase from yeast. *J Biol Chem* **269**:27670-27678.

Charbonnier, J.B., P. Belin, M. Moutiez, E.A. Stura, and E. Quemeneur. 1999. On the role of the *cis*-proline residue in the active site of DsbA. *Protein Sci* **8**:96-105.

Chen, V.B., W.B. Arendall, 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, and D.C. Richardson. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**:12-21.

Chern, M.-K., T.-C.Wu, C.-H.Hsieh, C.-C.Chou, L.-F.Liu, I.C. Kuan, Y.-H.Yeh, C.-D.Hsiao, and M.F. Tam. 2000. Tyr115, Gln165 and Trp209 contribute to the 1,2-epoxy-3-(p-nitrophenoxy) propane-conjugating activity of glutathione *S*-transferase cGSTM1-1. *J Mol Biol* **300**:1257-1269.

Chivers, P.T., and R.T. Raines. 1997. General acid/base catalysis in the active site of *Escherichia coli* thioredoxin. *Biochemistry***36**:15810-15816.

Chothia, C. 1992. Proteins.One thousand families for the molecular biologist. *Nature* **357**:543-544.

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

Collet, J.F., and J. Messens. 2010. Structure, function, and mechanism of thioredoxin proteins. *Antioxid Redox Signal* **13**:1205-1216.

Copley, S.D., W.R. Novak, and P.C. Babbitt. 2004. Divergence of function in the thioredoxin fold suprafamily: evidence for evolution of peroxiredoxins from a thioredoxin-like ancestor. *Biochemistry* **43**:13981-13995.

Dangi, B., A.V. Dobrodumov, J.M. Louis, and A.M. Gronenborn. 2002. Solution structure and dynamics of the human−*Escherichia coli* thioredoxin Chimera: Insights into thermodynamic stability. *Biochemistry* **41**:9376-9388.

DeLano, W. L. (2006) The PyMOL Molecular Graphics System, DeLano Scientific, LLC, San Carlos, CA.

Denisov, V.P., K. Venu, J. Peters, H.D. Hörlein, and B. Halle. 1997. Orientational disorder and entropy of water in protein cavities. *J Phys Chem B* **101**:9380-9389.

Desmots, F., P. Loyer, M. Rissel, A. Guillouzo, and F. Morel. 2005. Activation of C-Jun N-terminal kinase is required for glutathione transferase A4 induction during oxidative stress, not during cell proliferation, in mouse hepatocytes. *FEBS Lett* **579**:5691-5696.

Dill, K.A. 1990a.The meaning of hydrophobicity.*Science* **250**:297-298.

Dill, K.A. 1990b. Dominant forces in protein folding. *Biochemistry.* **29**:7133-7155

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

Dirr, H., P. Reinemer, and R. Huber. 1994a. Refined crystal structure of porcine class Pi glutathione *S*-transferase (pGST P1-1) at 2.1 Å resolution. *J Mol Biol* **243**:72-92.

Dirr, H., P. Reinemer, and R. Huber. 1994b. X-ray crystal structures of cytosolic glutathione *S*-transferases. Implications for protein architecture, substrate recognition and catalytic function.*Eur J Biochem* **220**:645-661.

Dirr, H.W., T. Little, D.C. Kuhnert, and Y. Sayed. 2005. A conserved N-capping motif contributes significantly to the stabilization and dynamics of the C-terminal region of class Alpha glutathione *S*-transferases. *J Biol Chem* **280**:19480-19487.

Dixon, D.P., B.G. Davis, and R. Edwards. 2002. Functional divergence in the glutathione transferase superfamily in plants. Identification of two classes with putative functions in redox homeostasis in *Arabidopsis thaliana*.*J Biol Chem* **277**:30859-30869.

Dourado, D.F., P.A. Fernandes, and M.J. Ramos. 2008. Mammalian cytosolic glutathione transferases. *Curr Protein Pept Sci* **9**:325-337.

Eklund, B.I., S. Gunnarsdottir, A.A. Elfarra, and B. Mannervik. 2007. Human glutathione transferases catalyzing the bioactivation of anticancer thiopurine prodrugs. *Biochem Pharmacol* **73**:1829-1841.

Eklund, H., C. Cambillau, B.M. Sjoberg, A. Holmgren, H. Jornvall, J.O. Hoog, and C.I. Branden. 1984. Conformational and functional similarities between glutaredoxin and thioredoxins. *EMBO J* **3**:1443-1449.

Eklund, H., F.K. Gleason, and A. Holmgren. 1991. Structural and functional relations among thioredoxins of different species. *Proteins* **11**:13-28.

Ellgaard, L., and L.W. Ruddock. 2005. The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep* **6**:28-32.

Emsley, P., and K. Cowtan. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**:2126-2132.

Engelhard, M., and P.A. Evans. 1995. Kinetics of interaction of partially folded proteins with a hydrophobic dye: evidence that molten globule character is maximal in early folding intermediates. *Protein Sci* **4**:1553-1562.

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

Ernst, J., R. Clubb, H. Zhou, A. Gronenborn, and G. Clore. 1995. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science* **267**:1813-1817.

Eyal, E., R. Najmanovich, M. Edelman, and V. Sobolev. 2003. Protein side-chain rearrangement in regions of point mutations. *Proteins: Struct Funct and Bioinf* **50**:272-282.

Fabrini, R., A. De Luca, L. Stella, G. Mei, B. Orioni, S. Ciccone, G. Federici, M. Lo Bello, and G. Ricci. 2009. Monomer−dimer equilibrium in glutathione transferases: A critical re-examination. *Biochemistry* **48**:10473-10482.

Fanucchi, S., R.J. Adamson, and H.W. Dirr. 2008. Formation of an unfolding intermediate state of soluble chloride intracellular channel protein CLIC1 at acidic pH. *Biochemistry* **47**:11674-11681.

Forsyth, W.R., and C.R. Matthews. 2002. Folding mechanism of indole-3-glycerol phosphate synthase from *Sulfolobus solfataricus*: a test of the conservation of folding mechanisms hypothesis in β/α barrels. *J Mol Biol* **320**:1119-1133.

Friedberg, I., and H. Margalit. 2002. Persistently conserved positions in structurally similar, sequence dissimilar proteins: Roles in preserving protein fold and function. *Protein Sci* **11**:350-360.

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

Furge, L.L., and F.P. Guengerich. 2006. Cytochrome P450 enzymes in drug metabolism and chemical toxicology: An introduction. *Biochem Mol Biol Educ*. **34**:66-74.

Gabellieri, E., E. Balestreri, A. Galli, and P. Cioni. 2008. Cavity-creating mutations in Pseudomonas aeruginosa azurin: effects on protein dynamics and stability. *Biophys J* **95**:771-781.

Galani, D., A.R. Fersht, and S. Perrett. 2002. Folding of the yeast prion protein Ure2p: kinetic evidence for folding and unfolding intermediates. *J Mol Biol* **315**:213-227.

Garcia-Pino, A., S. Martinez-Rodriguez, K. Wahni, L. Wyns, R. Loris, and J. Messens. 2009. Coupling of domain swapping to kinetic stability in a thioredoxin mutant. *J Mol Biol* **385**:1590-1599.

Ge, Y., Z. Qi, Y. Wang, X. Liu, J. Li, J. Xu, J. Liu, and J. Shen. 2009. Engineered selenium-containing glutaredoxin displays strong glutathione peroxidase activity rivaling natural enzyme. *Int J Biochem Cell Biol* **41**:900-906.

Gildenhuys, S., L.A. Wallace, J.P. Burke, D. Balchin, Y. Sayed, and H.W. Dirr. 2010b. Class Pi glutathione transferase unfolds via a dimeric and not monomeric intermediate: Functional implications for an unstable monomer. *Biochemistry* **49**:5074-5081.

Gildenhuys, S., M. Dobreva, N. Kinsley, Y. Sayed, J. Burke, S. Pelly, G.P. Gordon, M. Sayed, T. Sewell, and H.W. Dirr. 2010a. Arginine 15 stabilizes an SNAr reaction transition state and the binding of anionic ligands at the active site of human glutathione transferase A1-1. *Biophys Chem* **146**:118-125.

Gleason, F.K. 1992. Mutation of conserved residues in Escherichia coli thioredoxin: Effects on stability and function. *Protein Sci* **1**:609-616.

Godoy-Ruiz, R., R. Perez-Jimenez, B. Ibarra-Molero, and J.M. Sanchez-Ruiz. 2005. A stability pattern of protein hydrophobic mutations that reflects evolutionary structural optimization. *Biophys J* **89**:3320-3331.

Greene, R.F., Jr., and C.N. Pace. 1974. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J Biol Chem* **249**:5388-5393.

Gruber, C.W., M. Cemazar, B. Heras, J.L. Martin, and D.J. Craik. 2006. Protein disulfide isomerase: the structure of oxidative folding. *Trends Biochem Sci* **31**:455-464.

Guex, N., and M.C. Peitsch. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**:2714-2723.

Gunasekaran, K., S.J. Eyles, A.T. Hagler, and L.M. Gierasch. 2001. Keeping it in the family: folding studies of related proteins. *Curr Opin Struct Biol* **11**:83-93.

Habig, W.H., and W.B. Jakoby. 1981. Assays for differentiation of glutathione *S*-transferases. *Methods Enzymol* **77**:398-405.

Habig, W.H., M.J. Pabst, G. Fleischner, Z. Gatmaitan, I.M. Arias, and W.B. Jakoby. 1974. The identity of glutathione *S*-transferase B with ligandin, a major binding protein of liver. *Proc Natl Acad Sci U S A* **71**:3879-3882.

Hall, A., P.A. Karplus, and L.B. Poole. 2009. Typical 2-Cys peroxiredoxins--structures, mechanisms and functions. *FEBS J* **276**:2469-2477.

Halle, B. 2002. Flexibility and packing in proteins. *Proc Natl Acad Sci U S A* **99**:1274-1279.

Handy, D.E., E. Lubos, Y. Yang, J.D. Galbraith, N. Kelly, Y.Y. Zhang, J.A. Leopold, and J. Loscalzo. 2009. Glutathione peroxidase-1 regulates mitochondrial function to modulate redox-dependent cellular responses. *J Biol Chem* **284**:11913-11921.

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

Hatahet, F., and L.W. Ruddock. 2009. Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. *Antioxid Redox Signal* **11**:2807-2850.

Hayes, J.D., and D.J. Pulford. 1995. The glutathione *S*-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit Rev Biochem Mol Biol* **30**:445-600.

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

Hearne, J.L., and R.F. Colman. 2006. Contribution of the mu loop to the structure and function of rat glutathione transferase M1-1. *Protein Sci* **15**:1277-1289.

Heckler, E.J., P.C. Rancy, V.K. Kodali, and C. Thorpe. 2008. Generating disulfides with the Quiescin-sulfhydryl oxidases. *Biochim Biophys Acta* **1783**:567-577.

Hellinga, H.W., R. Wynn, and F.M. Richards. 1992. The hydrophobic core of *Escherichia coli* thioredoxin shows a high tolerance to non-conservative single amino acid substitutions. *Biochemistry* **31**:11203-11209.

Hollecker, M., and T.E. Creighton. 1983. Evolutionary conservation and variation of protein folding pathways. Two protease inhibitor homologues from black mamba venom.*J Mol Biol* **168**:409-437.

Holmgren, A., B.O. Soderberg, H. Eklund, and C.I. Branden. 1975. Three-dimensional structure of *Escherichia coli* thioredoxin-S2 to 2.8 Åresolution. *Proc Natl Acad Sci U S A***72**:2305-2309

Holmgren, A. 1985.Thioredoxin.*Annu Rev Biochem* **54**:237-271.

Hornby, J.A., J.K. Luo, J.M. Stevens, L.A. Wallace, W. Kaplan, R.N. Armstrong, and H.W. Dirr. 2000. Equilibrium folding of dimeric class Mu glutathione transferases involves a stable monomeric intermediate. *Biochemistry* **39**:12336-12344.

Hornby, J.A., S.G. Codreanu, R.N. Armstrong, and H.W. Dirr. 2002. Molecular recognition at the dimer interface of a class Mu glutathione transferase: role of a hydrophobic interaction motif in dimer stability and protein function. *Biochemistry* **41**:14238-14247.

Hou, L., M.T. Honaker, L.M. Shireman, L.M. Balogh, A.G. Roberts, K.-c.Ng, A. Nath, and W.M. Atkins. 2007. Functional promiscuity correlates with conformational heterogeneity in Alpha-class glutathione *S*-transferases. *J. Biol. Chem*:M700868200.

Huang, Y.C., S. Misquitta, S.Y. Blond, E. Adams, and R.F. Colman. 2008. Catalytically active monomer of glutathione *S*-transferase Pi and key residues involved in the electrostatic interaction between subunits. *J Biol Chem* **283**:32880-32888.

Huber, D., A. Chaffotte, M. Eser, A.-G.l. Planson, and J. Beckwith. 2010. Amino acid residues important for folding of thioredoxin are revealed only by study of the physiologically relevant reduced form of the protein. *Biochemistry* **49**:8922-8928.

Inaba, K. 2009. Disulfide bond formation system in *Escherichia coli*.*J Biochem* **146**:591-597.

Jackson, S.E., M. Moracci, N. elMasry, C.M. Johnson, and A.R. Fersht. 1993. Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry* **32**:11259-11269.

Jacobi, A., M. Huber-Wunderlich, J. Hennecke, and R. Glockshuber. 1997. Elimination of all charged residues in the vicinity of the active-site helix of the Disulfide Oxidoreductase DsbA. *J Biol Chem* **272**:21692-21699.

Jang, H.H., K.O. Lee, Y.H. Chi, B.G. Jung, S.K. Park, J.H. Park, J.R. Lee, S.S. Lee, J.C. Moon, J.W. Yun, Y.O. Choi, W.Y. Kim, J.S. Kang, G.W. Cheong, D.J. Yun, S.G. Rhee, M.J. Cho, and S.Y. Lee. 2004. Two enzymes in one; two yeast peroxiredoxins display oxidative stress-dependent switching from a peroxidase to a molecular chaperone function. *Cell* **117**:625-635.

Jeng, M.F., and H.J. Dyson. 1995. Comparison of the hydrogen-exchange behavior of reduced and oxidized *Escherichia coli* thioredoxin. *Biochemistry* **34**:611-619.

Ji, G., and S. Silver. 1992. Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid pI258. *Proc Natl Acad Sci U S A* **89**:9474-9478.

Ji, X., P. Zhang, R.N. Armstrong, and G.L. Gilliland. 1992. The three-dimensional structure of a glutathione *S*-transferase from the Mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2-Å resolution.*Biochemistry* **31**:10169-10184.

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

Kelley, R.F., and E. Stellwagen. 1984. Conformational transitions of thioredoxin in guanidine hydrochloride. *Biochemistry* **23**:5095-5102.

Kelley, R.F., and F.M. Richards. 1987. Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry* **26**:6765-6774.

Kelley, R.F., J. Wilson, C. Bryant, and E. Stellwagen. 1986. Effects of guanidine hydrochloride on the refolding kinetics of denatured thioredoxin. *Biochemistry* **25**:728-732.

Kono, H., M. Saito, and A. Sarai. 2000. Stability analysis for the cavity-filling mutations of the Myb DNA-binding domain utilizing free-energy calculations. *Proteins* **38**:197-209.

Ladenstein, R., O. Epp, K. Bartels, A. Jones, R. Huber, and A. Wendel. 1979. Structure analysis and molecular model of the selenoenzyme glutathione peroxidase at 2.8 Åresolution. *J Mol Biol* **134**:199-218.

Ladner, J.E., J.F. Parsons, C.L. Rife, G.L. Gilliland, and R.N. Armstrong. 2004. Parallel evolutionary pathways for glutathione transferases: structure and mechanism of the mitochondrial class Kappa enzyme rGSTK1-1. *Biochemistry* **43**:352-361.

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. pp 11-14, 188, 237-249, 447-

Langsetmo, K., J.A. Fuchs, and C. Woodward. 1991. The conserved, buried aspartic acid in oxidized *Escherichia coli* thioredoxin has a pKa of 7.5. Its titration produces a related shift in global stability. *Biochemistry* **30**:7603-7609.

Laskowski, R.A., J.A. Rullmannn, M.W. MacArthur, R. Kaptein, and J.M. Thornton. 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**:477-486.

Le Trong, I., R.E. Stenkamp, C. Ibarra, W.M. Atkins, and E.T. Adman. 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 Bioinf* **48**:618-627.

Li, J., Z. Xia, and J. Ding. 2005. Thioredoxin-like domain of human Kappa class glutathione transferase reveals sequence homology and structure similarity to the Theta class enzyme. *Protein Sci* **14**:2361-2369.

Lim, W.A., A. Hodel, R.T. Sauer, and F.M. Richards. 1994. The crystal structure of a mutant protein with altered but improved hydrophobic core packing. *Proc Natl Acad Sci U S A.* **91**:423-427.

Lim, W.A., D.C. Farruggio, and R.T. Sauer. 1992. Structural and energetic consequences of disruptive mutations in a protein core. *Biochemistry* **31**:4324-4333.

Lin, H.J., A.-S. Johansson, G. Stenberg, A.M. Materi, J.M. Park, A. Dai, H. Zhou, J.S.Y. Gim, I.H. Kau, S.I. Hardy, M.W. Parker, and B. Mannervik. 2003. Naturally occurring Phe151Leu substitution near a conserved folding module lowers stability of glutathione transferase P1–1. *Biochimica et Biophysica Acta* **1649**:16-23.

Louis, J.M., R.E. Georgescu, M.L. Tasayco, O. Tcherkasskaya, and A.M. Gronenborn. 2001. Probing the structure and stability of a hybrid protein: The Human−*E. coli* thioredoxin Chimera. *Biochemistry* **40**:11184-11192.

Lovell, S.C., I.W. Davis, W.B. Arendall, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, and D.C. Richardson. 2003. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins: Struct Funct Bioinf* **50**:437-450.

Lubos, E., J. Loscalzo, and D.E. Handy. 2011. Glutathione peroxidase-1 in health and disease: from molecular mechanisms to therapeutic opportunities. *Antioxid Redox Signal* **15**:1957-1997.

Luo, J.K., J.A. Hornby, L.A. Wallace, J. Chen, R.N. Armstrong, and H.W. Dirr. 2002. Impact of domain interchange on conformational stability and equilibrium folding of chimeric class micro glutathione transferases. *Protein Sci* **11**:2208-2217.

Maiti, R., G.H. Van Domselaar, H. Zhang, and D.S. Wishart. 2004. SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res* **32**:W590-594.

Manavalan, P., and P.K. Ponnuswamy. 1978. Hydrophobic character of amino acid residues in globular proteins. *Nature* **275**:673-674.

Mannervik, B., Y.C. Awasthi, P.G. Board, J.D. Hayes, C. Di Ilio, B. Ketterer, I. Listowsky, R. Morgenstern, M. Muramatsu, and W.R. Pearson. 1992. Nomenclature for human glutathione transferases. *Biochem J* **282** :305-306.

Mannervik, B., P. Alin, C. Guthenberg, H. Jensson, M.K. Tahir, M. Warholm, and H. Jornvall. 1985. Identification of three classes of cytosolic glutathione transferase common to several mammalian species: correlation between structural data and enzymatic properties. *Proc Natl Acad Sci U S A* **82**:7202-7206.

Mannervik, B., P.G. Board, J.D. Hayes, I. Listowsky, and W.R. Pearson. 2005. Nomenclature for mammalian soluble glutathione transferases. *Methods Enzymol* **401**:1-8.

Margis, R., C. Dunand, F.K. Teixeira, and M. Margis-Pinheiro. 2008. Glutathione peroxidase family - an evolutionary overview. *FEBS J* **275**:3959-3970.

Martin, J.L., J.C. Bardwell, and J. Kuriyan. 1993. Crystal structure of the DsbA protein required for disulphide bond formation in vivo. *Nature* **365**:464-468.

Martin, J.L. 1995. Thioredoxin--a fold for all reasons.*Structure* **3**:245-250.

Martin, P., S. DeMel, J. Shi, T. Gladysheva, D.L. Gatti, B.P. Rosen, and B.F.P. Edwards. 2001. Insights into the Structure, Solvation, and Mechanism of ArsC Arsenate Reductase, a Novel Arsenic Detoxification Enzyme. *Structure* **9**:1071-1081.

Matthews, B.W. 1968. Solvent content of protein crystals.*J Mol Biol***33**:491-497.

Matthews, B.W. 1991. Mutational analysis of protein stability : Current Opinion in Structural Biology 1991, 1:17-21. *Curr Opin Struct Biol* **1**:17-21.

Meiering, E.M., L. Serrano, and A.R. Fersht. 1992. Effect of active site residues in barnase on activity and stability. *J Mol Biol* **225**:585-589.

Menchise, V., C. Corbier, C. Didierjean, M. Saviano, E. Benedetti, J.P. Jacquot, and A. Aubry. 2001. Crystal structure of the wild-type and D30A mutant thioredoxin of *Chlamydomonas reinhardtii* and implications for the catalytic mechanism. *Biochem J* **359**:65-75.

Messens, J., J.C. Martins, K. Van Belle, E. Brosens, A. Desmyter, M. De Gieter, J.-M. Wieruszeski, R. Willem, L. Wyns, and I. Zegers. 2002. All intermediates of the arsenate reductase mechanism, including an intramolecular dynamic disulfide cascade. *Proc Natl Acad of Sci USA* **99**:8506-8511.

Meyer, Y., B.B. Buchanan, F. Vignols, and J.P. Reichheld. 2009. Thioredoxins and glutaredoxins: unifying elements in redox biology. *Annu Rev Genet* **43**:335-367.

Micaloni, C., G.K.W. Kong, A.P. Mazzetti, M. Nuccetelli, G. Antonini, L. Stella, W.J. McKinstry, G. Polekhina, J. Rossjohn, G. Federici, G. Ricci, M.W. Parker, and M. Lo Bello. 2003. Engineering a New C-terminal Tail in the H-site of Human Glutathione Transferase P1-1: Structural and Functional Consequences. J Mol Biol. **325**:111-122.

Mirny, L.A., and E.I. Shakhnovich. 1999. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *J Mol Biol* **291**:177-196.

Morel, F., C. Rauch, E. Petit, A. Piton, N. Theret, B. Coles, and A. Guillouzo. 2004. Gene and protein characterization of the human glutathione *S*-transferase Kappa and evidence for a peroxisomal localization. *J Biol Chem* **279**:16246-16253.

Mosebi, S., Y. Sayed, J. Burke, and H.W. Dirr. 2003. Residue 219 impacts on the dynamics of the C-terminal region in glutathione transferase A1-1: implications for stability and catalytic and ligandin functions. *Biochemistry* **42**:15326-15332.

Murshudov, G.N., A.A. Vagin, and E.J. Dodson. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**:240-255.

Murzin, A.G., S.E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**:536-540.

Myers, J.K., C.N. Pace, and J.M. Scholtz. 1995. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **4**:2138-2148.

Na, Y.R., and C. Park. 2009. Investigating protein unfolding kinetics by pulse proteolysis. *Protein Sci* **18**:268-276.

Nathaniel, C., L.A. Wallace, J. Burke, and H.W. Dirr. 2003. The role of an evolutionarily conserved *cis*-proline in the thioredoxin-like domain of human class Alpha glutathione transferase A1-1. *Biochem J* **372**:241-246.

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

Nishida, M., S. Harada, S. Noguchi, Y. Satow, H. Inoue, and K. Takahashi. 1998. Three-dimensional structure of Escherichia coli glutathione *S*-transferase complexed with glutathione sulfonate: catalytic roles of Cys10 and His106. *J Mol Biol* **281**:135-147.

Oakley, A. 2011. Glutathione transferases: a structural perspective. *Drug Metab Rev* **43**:138-151.

O'Brien, P.J. 2006. Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chem Rev* **106**:720-752.

Ojeda, L., G. Keller, U. Muhlenhoff, J.C. Rutherford, R. Lill, and D.R. Winge. 2006. Role of glutaredoxin-3 and glutaredoxin-4 in the iron regulation of the Aft1 transcriptional activator in *Saccharomyces cerevisiae*. *J Biol Chem* **281**:17661-17669.

Oshima, H., T. Abe, and K. Takahashi. 1977. Studies on thermolysin II. effects of chemical modifications on the activity of thermolysin. *J Biochem* **81**:65-70.

O'Sullivan, O., K. Suhre, C. Abergel, D.G. Higgins, and C. Notredame. 2004. 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *J Mol Biol* **340**:385-395.

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

Pan, J.L., and J.C. Bardwell. 2006. The origami of thioredoxin-like folds. *Protein Sci*. **15**:2217-2227.

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

Park, C., and S. Marqusee. 2005. Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding. *Nat Meth* **2**:207-212.

Park, C., and S. Marqusee. 2006. Quantitative determination of protein stability and ligand binding by pulse proteolysis. *Curr Protoc Protein Sci* Chapter 20:Unit 20 11.

Pedone, E., D. Limauro, K. D'Ambrosio, G. De Simone, and S. Bartolucci. 2010. Multiple catalytically active thioredoxin folds: a winning strategy for many functions. *Cell Mol Life Sci* **67**:3797-3814.

Pemble, S.E., and J.B. Taylor. 1992. An evolutionary perspective on glutathione transferases inferred from class-theta glutathione transferase cDNA sequences. *Biochem J* **287**:957-96.

Pemble, S.E., A.F. Wardle, and J.B. Taylor. 1996. Glutathione *S*-transferase class Kappa: characterization by the cloning of rat mitochondrial GST and identification of a human homologue. *Biochem J* **319**:749-754.

Plaxco, K.W., K.T. Simons, I. Ruczinski, and D. Baker. 2000b. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry* **39**:11177-11183.

Plaxco, K.W., S. Larson, I. Ruczinski, D.S. Riddle, E.C. Thayer, B. Buchwitz, A.R. Davidson, and D. Baker. 2000a. Evolutionary conservation in protein folding kinetics.*J Mol Biol* **298**:303-312.

Ptitsyn, O.B., V.E. Bychkova, and V.N. Uversky. 1995. Kinetic and equilibrium folding intermediates. *Philos Trans R Soc Lond B Biol Sci* **348**:35-41.

Qi, Y., and N.V. Grishin. 2005. Structural classification of thioredoxin-like fold proteins. *Proteins* **58**:376-388.

Qin, J., G.M. Clore, W.M.P. Kennedy, J.R. Huth, and A.M. Gronenborn. 1995. Solution structure of human thioredoxin in a mixed disulfide intermediate complex with its target peptide from the transcription factor NFκB. *Structure* **3**:289-297.

Radivojac, P., Z. Obradovic, D.K. Smith, G. Zhu, S. Vucetic, C.J. Brown, J.D. Lawson, and A.K. Dunker. 2004. Protein flexibility and intrinsic disorder. *Protein Sci* **13**:71-80.

Ragone, R. 2000. How the protein concentration affects unfolding curves of oligomers. *Biopolymers* **53**:221-225.

Reinemer, P., H.W. Dirr, R. Ladenstein, J. Schaffer, O. Gallay, and R. Huber. 1991. The three-dimensional structure of class Pi glutathione *S*-transferase in complex with glutathione sulfonate at 2.3 Åresolution. *EMBO J* **10**:1997-2005.

Ren, B., W. Huang, B. Åkesson, and R. Ladenstein. 1997. The crystal structure of seleno-glutathione peroxidase from human plasma at 2.9 Å resolution. *J Mol Biol* **268**:869-885.

Robinson, A., G.A. Huttley, H.S. Booth, and P.G. Board. 2004. Modelling and bioinformatics studies of the human Kappa-class glutathione transferase predict a novel third glutathione transferase family with similarity to prokaryotic 2-hydroxychromene-2-carboxylate isomerases. *Biochem J* **379**:541-552.

Roos, G., A. Garcia-Pino, K. Van belle, E. Brosens, K. Wahni, G. Vandenbussche, L. Wyns, R. Loris, and J. Messens. 2007. The conserved active site proline determines the reducing power of *Staphylococcus aureus* thioredoxin. *J Mol Biol* **368**:800-811.

Roos, G., N. Foloppe, K. Van Laer, L. Wyns, L. Nilsson, P. Geerlings, and J. Messens. 2009. How thioredoxin dissociates its mixed disulfide. *PLoS Comput Biol* 5:e1000461.

Roos, G., P. Geerlings, and J. Messens. 2010. The conserved active site tryptophan of thioredoxin has no effect on its redox properties. *Prot Sci*. **19**:190-194.

Rossjohn, J., G. Polekhina, S.C. Feil, N. Allocati, M. Masulli, C. Di Illio, and M.W. Parker. 1998. A mixed disulfide bond in bacterial glutathione transferase: functional and evolutionary implications. *Structure* **6**:721-734.

Rumfeldt, J.A.O., C. Galvagnion, K.A. Vassall, and E.M. Meiering. 2008. Conformational stability and folding mechanisms of dimeric proteins. *Prog Biophys Mol Biol* **98**:61-84.

Sacchetta, P., A. Pennelli, T. Bucciarelli, L. Cornelio, F. Amicarelli, M. Miranda, and C. Di Ilio. 1999. Multiple unfolded states of glutathione transferase bbGSTP1-1 by guanidinium chloride. *Arch Biochem Biophys* **369**:100-106.

Sandalova, T., L. Zhong, Y. Lindqvist, A. Holmgren, and G. Schneider. 2001. Three-dimensional structure of a mammalian thioredoxin reductase: Implications for mechanism and evolution of a selenocysteine-dependent enzyme. *Proc Natl Acad Sci* **98**:9533-9538.

Santos, J., Marino-Buslje, C., Kleinman, C., Ermacora, M. R., and Delfino, J. M. (2007) Consolidation of the thioredoxin fold by peptide recognition: interaction between E. coli thioredoxin fragments 1-93 and 94-108. *Biochemistry* **46**, 5148–5159.

Santos, J., M.P. Sica, C. Marino Buslje, A.M. Garrote, M.R. Ermácora, and J.M. Delfino. 2009. Structural selection of a native fold by peptide recognition. Insights into the thioredoxin folding mechanism.*Biochemistry* **48**:595-607.

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

Schreiber, G., A.M. Buckle, and A.R. Fersht. 1994. Stability and function: two constraints in the evolution of barstar and other proteins. *Structure* **2**:945-951.

Schroder, E., and C.P. Ponting. 1998. Evidence that peroxiredoxins are novel members of the thioredoxin fold superfamily. *Protein Sci* **7**:2465-2468.

Schroder, E., J.A. Littlechild, A.A. Lebedev, N. Errington, A.A. Vagin, and M.N. Isupov. 2000. Crystal structure of decameric 2-Cys peroxiredoxin from human erythrocytes at 1.7 Åresolution. *Structure* **8**:605-615.

Schueler-Furman, O., and D. Baker. 2003. Conserved residue clustering and protein structure prediction. *Proteins* **52**:225-235.

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

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

Shirley, B.A. 1995. Urea and guanidine hydrochloride denaturation curves. *Methods Mol Biol* **40**:177-190.

Shoichet, B.K., W.A. Baase, R. Kuroki, and B.W. Matthews. 1995. A relationship between protein stability and protein function. *Proc Natl Acad Sci U S A* **92**:452-456.

Shortle, D. 1992. Mutational studies of protein structures and their stabilities. *Q Rev Biophys* **25**:205-250.

Sinning, I., G.J. Kleywegt, S.W. Cowan, P. Reinemer, H.W. Dirr, R. Huber, G.L. Gilliland, R.N. Armstrong, X. Ji, P.G. Board, B. Olin, B. Mannervik, T.A. Jones. 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.

Skopelitou, K., P. Dhavala, A.C. Papageorgiou, and N.E. Labrou. 2012a. A glutathione transferase from *Agrobacterium tumefaciens* reveals a novel class of bacterial GST Superfamily. *PLoS ONE*.**7**:e34263.

Skopelitou, K., A. Muleta, O. Pavli, G. Skaracis, E. Flemetakis, A. Papageorgiou, and N. Labrou. 2012b. Overlapping protective roles for glutathione transferase gene family members in chemical and oxidative stress response in *Agrobacterium tumefaciens*.*Funct Integr Genomics* **12**:157-172.

Slaby, I., and A. Holmgren. 1979. Structure and enzymatic functions of thioredoxin refolded by complementation of two tryptic peptide fragments. *Biochemistry* **18**:5584-5591.

Slavik, J. 1982. Anilinonaphthalene sulfonate as a probe of membrane composition and function.*Biochim Biophys Acta* **694**:1-25.

Soulages, J.L. 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.

Spudich, G., and S. Marqusee. 2000. A change in the apparent m value reveals a populated intermediate under equilibrium conditions in *Escherichia coli* ribonuclease HI. *Biochemistry* **39**:11677-11683.

Stenberg, G., P.G. Board, and B. Mannervik. 1991. Mutation of an evolutionarily conserved tyrosine residue in the active site of a human class Alpha glutathione transferase. *FEBS Lett* **293**:153-155.

Stenberg, G., B. Dragani, R. Cocco, B. Mannervik, and A. Aceto. 2000. A conserved "hydrophobic staple motif" plays a crucial role in the refolding of human glutathione transferase P1-1. *J Biol Chem* **275**:10421-10428.

Stevens, J.M., J.A. Hornby, R.N. Armstrong, and H.W. Dirr. 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.

Stevens, J.M., R.N. Armstrong, and H.W. Dirr. 2000. Electrostatic interactions affecting the active site of class Sigma glutathione *S*-transferase. *Biochem J* **347**:193-197.

Takano, K., Y. Yamagata, and K. Yutani. 2003. Buried water molecules contribute to the conformational stability of a protein. *Protein Eng* **16**:5-9.

Tasayco, M. L., Fuchs, J., Yang, X. M., Dyalram, D., and Georgescu, R. E. (2000) Interaction between two discontiguous chain segments from the beta-sheet of Escherichia coli thioredoxin suggests an initiation site for folding. *Biochemistry* **39**, 10613-10618.

Thorpe, C., K.L. Hoober, S. Raje, N.M. Glynn, J. Burnside, G.K. Turi, and D.L. Coppock. 2002. Sulfhydryl oxidases: emerging catalysts of protein disulfide bond formation in eukaryotes. *Arch Biochem Biophys* **405**:1-12.

Thorpe, C., and D.L. Coppock. 2007. Generating disulfides in multicellular organisms: emerging roles for a new flavoprotein family. *J Biol Chem* **282**:13929-13933.

Tokuriki, N., F. Stricher, L. Serrano, and D.S. Tawfik. 2008. How protein stability and new functions trade off. *PLoS Comput Biol* **4**:e1000002.

Vagin, A., and A. Teplyakov. 2000. An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **56**:1622-1624.

Vararattanavech, A., and A.J. Ketterman. 2003. Multiple roles of glutathione binding-site residues of glutathione *S*-transferase. *Protein Pept Lett* **10**:441-448.

Vararattanavech, A., P. Prommeenate, and A.J. Ketterman. 2006. The structural roles of a conserved small hydrophobic core in the active site and an ionic bridge in domain I of Delta class glutathione *S*-transferase. *Biochem J* **393**:89-95.

Wade, R.C., M.H. Mazor, J.A. McCammon, and F.A. Quiocho. 1991. A molecular dynamics study of thermodynamic and structural aspects of the hydration of cavities in proteins. *Biopolymers* **31**:919-931.

Wallace, L.A., G.L. Blatch, and H.W. Dirr. 1998a. A topologically conserved aliphatic residue in alpha-helix 6 stabilizes the hydrophobic core in domain II of glutathione transferases and is a structural determinant for the unfolding pathway. *Biochem J* **336**:413-418.

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

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

Wallace, L.A., J. Burke, and H.W. Dirr. 2000. Domain-domain interface packing at conserved Trp-20 in class Alpha glutathione transferase impacts on protein stability. *Biochimica et Biophysica Acta* **1478**:325-332.

Wang, B., Y. Peng, T. Zhang, and J. Ding. 2011. Crystal structures and kinetic studies of human Kappa class glutathione transferase provide insights into the catalytic mechanism. *Biochem J* **439**:215-225.

Weber, G., and L.B. Young. 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.

Whitbread, A.K., A. Masoumi, N. Tetlow, E. Schmuck, M. Coggan, and P.G. Board. 2005. Characterization of the omega class of glutathione transferases. Methods Enzymol. 401:78-99

Whitten, S.T., and E.B. Garcia-Moreno. 2000. pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state. *Biochemistry* **39**:14292-14304.

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

Williams, J.C., C. Sue, G.S. Banting, H. Yang, D.M. Glerum, W.A. Hendrickson, and E.A. Schon. 2005. Crystal Structure of Human SCO1.*J Biol Chem* **280**:15202-15211.

Wongsantichon, J., and A.J. Ketterman. 2006. An intersubunit lock-and-key 'clasp' motif in the dimer interface of Delta class glutathione transferase. *Biochem J* **394**:135-144.

Wood, Z.A., E. Schroder, J. Robin Harris, and L.B. Poole. 2003. Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem Sci* **28**:32-40.

Woody, R.W. 1995. Circular dichroism.*Methods Enzymol* **246**:34-71.

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

Xu, C., C.Y. Li, and A.N. Kong. 2005. Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Arch Pharm Res* **28**:249-268.

Xu, J., W.A. Baase, E. Baldwin, and B.W. Matthews. 1998. The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci* **7**:158-177.

Yu, H.-j., J.-q.Liu, A. Böck, J. Li, G.-m.Luo, and J.-c. Shen. 2005. Engineering glutathione transferase to a novel glutathione peroxidase mimic with high catalytic efficiency. *J Biol Chem* **280**:11930-11935.

Zhou, H., J. Brock, D. Liu, P.G. Board, and A.J. Oakley. 2012. Structural insights into the dehydroascorbate reductase activity of human Omega-Class glutathione transferases. *J Mol Biol* **420**:190-203.