

https://dspace.org/

Investigating the structural and functional effects of naturally occurring single nucleotide polymorphisms in hGSTP1-1

Makeleni, Nelisa

2018
https://hdl.handle.net/10539/25014

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

Investigating the structural and functional effects of naturally occurring single nucleotide polymorphisms in hGSTP1-1

Nelisa Makeleni



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

Johannesburg, February 2018

Supervisor: Prof H. W Dirr

agriculture,
forestry & fisheries

Department:
Agriculture, Forestry and Fisheries
REPUBLIC OF SOUTH AFRICA

CERTIFICATE OF REGISTRATION

Registration Number: **39.2/UNIVERSITY OF WITWATERSRAND -18/103**

It is hereby certified that the facility situated at:

UNIVERSITY OF WITWATERSRAND
SCHOOL OF MOLECULAR AND CELL BIOLOGY
EAST CAMPUS
GATE HOUSE ROOM 431
JOHANNESBURG
2050

has been registered in the name(s) of: **PROF H. W DIRR**

on behalf of: **UNIVERSITY OF WITWATERSRAND**

in terms of Regulation 8 of the Genetically Modified Organisms Act, 1997 (Act No. 15 of 1997) in respect of the following-

type of facility: **LABORATORY**

containment level: **ONE (1)**

for period: **31 JANUARY 2018 – 31 JANUARY 2021**

Date issued: **31 JANUARY 2018**

Registrar for Genetically Modified Organisms

Date

PSFRU: Wits IBC Clearance Certificate

Research Office

INSTITUTIONAL BIOSAFETY COMMITTEE
(R 14/16)

CLEARANCE CERTIFICATE **PROTOCOL NUMBER: 20150651Lab**

BRIEF DESCRIPTION OF APPLICATION:

Laboratories situated in rooms 411-434 in Gate House

APPLICANT: Professor HW Dirr

SCHOOL/DEPARTMENT : Molecular & Cell Biology/

DATE CONSIDERED: 22 June 2015

DECISION OF COMMITTEE:

These laboratories are considered to meet the standards of the Institutional Biosafety Committee (IBC), as specified in the IBC's published Standard Operating procedures.
Any change to the type of biohazardous agents being handled should be reported to the IBC

1. This clearance certificate expires on 27 Feb 2018 and may be renewed on application; the expiry date is co-terminous with that of the Dept of Agriculture, Forestry and Fisheries
2. An annual report must be provided on the anniversary date of this certificate, for as long as the project continues
3. Notification of any proposed modifications must be submitted on the attached form

DATE: 25/06/15 **CHAIRPERSON:** _____ (Mr J Larkin)

DECLARATION OF APPLICANT:

To be completed in duplicate and **one copy** returned to the Secretary, Room 10005, 10th floor, Senate House, University.

1. I have read, understood and accepted the approval conditions above
2. I agree to submit a yearly progress report to the Committee and to submit an interim report on the form provided, in the event of any significant unforeseen event, *e.g.* suspension of a drug trial, temporary closure or relocation of my laboratory, etc
3. I note that the University Safety Officer, or his/her representative, may at any reasonable time inspect my laboratory or trial site to ensure compliance with current Health and Safety legislation. I undertake to offer my full co-operation in any such inspection.
4. I have read, understood and will comply with the *recommended standard operating procedures for the handling of biohazardous materials* posted at http://web.wits.ac.za/Academic/Research/Biosafety.htm
5. I declare (delete as appropriate) that:
   a. I have all the approvals required by statute or regulation and by the funding agencies supporting this work, or
   b. that I will not begin work until such approvals are obtained

Signed: _____ **Date:** 30 June 2015

MSWorks2000/lain0015/IBCClear.wps

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

DECLARATION

I, **NELISA MAKELENI (957978),** am a student registered for the degree of Master in Science in the academic year 2017.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at his or any other university.
- The information used in the Thesis/Dissertation/Research Report has been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conversions in referencing the thoughts and ideas of others.
- I understand that the University of Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature__  _____ day of _19_February___Year ___2018___

Research Outputs

Conference Presentations

Name	Date	Location	Title	Authors
Cancer Research Symposium conference sponsored by CANSA	9th February 2017	The Adler Museum, Wits Medical School	Investigating the structural and functional effects of naturally occurring single nucleotide polymorphisms in hGSTP1-1	Nelisa Makeleni Prof H. W Dirr

ABSTRACT

SNPs are DNA sequence variations occurring when a nucleotide in the genome differs between members of a species. These are the most frequent types of genetic variations, accounting for about 90 % of the genetic variations in humans. Some changes in coding regions of DNA tend to alter the amino acid sequence and therefore affect the produced protein. The human glutathione *S*-transferase class pi (hGSTP1-1) has been studied for its involvement in a broad range of functions such as detoxification. It has recently drawn interest because of its increased levels in tumours. In order to incorporate mutations into the hGSTP1-1 DNA sequence, site directed mutagenesis was performed to obtain the SNP forms and their effect on the structure, stability and function were studied. The activity was measured using the 1-chloro-2,4-dinitrobenzene (CDNB-GSH) enzyme assay. The SNP forms L159M, located in the hydrophobic core of domain 2, and T110S positioned next to the H-site of the protein did not affect the secondary structure when compared to the wild-type, while the L159M SNP form had an effect on the protein's tertiary structure when determined using CD and fluorometer. Both of the SNP forms showed a difference in the exposure of hydrophobic surfaces as indicated by ANS binding studies. The SNPs had no effect on the thermal stability of the protein although the enzyme activity of SNP L159M and the SNP T110S is decreased by about 57% and 18 %, respectively, when compared with the activity of the wild-type. The SNP T110S does not have a significant effect on the structure and activity of hGSTP1-1 while the L159M showed significant changes suggesting possible effects on both the structure and activity. This may be attributed to its location in a conserved hydrophobic core of domain 2 of hGSTP1-1. The decrease in activity of the L159M hGSTP1-1 may have a negative role on the detoxification function of the protein.

DEDICATION

This work is dedicated to my Mother, MaDongo and Father, Bhele, for believing in me, for their support and putting their trust in my decisions. And mostly of all their motivation and prayers.

ACKNOWLEDGMENTS

I wish to thank God for making this all possible:

My Mother and Father for their support and motivation.

My supervisor, Prof. H. Dirr, for giving me the opportunity, his support, guidance, all the knowledge he has selflessly provided and for being patient with me throughout the course of my studies.

All current and past members of the Protein-Structure Function Research Unit, with special

thanks to Dr S. Mathura for all her tips and advice and Dr Vijay Kumar for helping with computational studies.

The University of the Witwatersrand and the National Research Foundation of South Africa for financial assistance

I would like to thank all my siblings for being there to support me and pushing me forward when I felt I couldn't go further: Ayanda, Unathi, Viwe, Siyabulela, Nceba and Noyolo Makeleni.

Finally, all my friends and family

TABLE OF CONTENTS

DECLARATION I

RESEARCH OUTPUTS II

ABSTRACT III

DEDICATION IV

ACKNOWLEDGEMENTS VIII

LIST OF FIGURES X

LIST OF TABLES XI

LIST OF ABBREVIATIONS XII

CHAPTER 1: INTRODUCTION

1.1 Gene mutations in humans 1

1.2 Single nucleotide polymorphism (SNPs) and their types 2

1.2.1 Effects of SNPs on DNA coding regions 3

1.3 Glutathione *S*-transferases (GSTs) 5

1.3.1 Functions of hGSTP1-1 6

1.3.2 Structure of hGSTP1-1 7

1.3.3 Effects of SNPs on hGSTP1-1 9

1.4 Human hGSTP1-1 and its relation to cancer 11

1.5 Population distribution of human diseases associated with SNPs found in hGSTP1-1 12

1.6 Aims 14

1.6.1 Objectives 14

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials and methods 15

2.2 Experimental procedures 16

2.3 Identification and construction of hGSTP1-1 SNPs 16

2.4 Expression of hGSTP1-1 wild-type and SNP forms 17

2.5 Purification of hGSTP1-1 wild-type and SNP forms 18

2.6 Protein concentration determination 19

2.7 Sodium dodecyl sulfate- polyacrylamide gel electrophoresis (SDS-PAGE) 19

2.8 Structural characterisation of hGSTP1-1 wild-type and SNP forms 20

2.8.1 Secondary structure characterisation by far-UV circular dichroism 20

2.8.2 Tertiary structure characterisation by fluorescence spectroscopy 20

2.8.3 Size-exclusion chromatography 21

2.9 Stability studies 21

2.10 Functional characterisation of hGSTP1-1 wild-type and SNP forms 22

2.10.1 GST/CDNB activity assay 22

2.10.2 8-Anilinonaphthalene-1-sulfonic acid (ANS) binding studies 22

CHAPTER 3: RESULTS

3.1 Identification and construction of hGSTP1-1 SNP forms by site directed mutagenesis 25

3.2 Site directed mutagenesis 25

3.3 Expression and purification of hGSTP1-1 wild-type and SNP forms 28

3.4 SDS-PAGE 30

3.5 Protein concentration determination 33

3.6 Structural characterisation 34

3.6.1 Secondary structure characterisation using far-UV circular dichroism 34

3.6.2 Tertiary structure characterisation using fluorescence spectroscopy 35

3.6.3 Size exclusion chromatography: Quaternary structure analysis 36

3.7 Thermal stability studies 38

3.8 Functional characterisation of hGSTP1-1 wild-type and SNP forms 39

3.8.1 GST activity assay for hGSTP1-1 wild-type and SNP forms 39

3.8.2 ANS binding studies 41

3.8.3 Dimer docking studies of ANS to hGSTP1-1 42

3.8.4 ANS displacement studies 43

CHAPTER 4: DISCUSSION

4.1 Site directed mutagenesis, expression and purification 45

4.2 Influence of mutations on structure: SNPs in the hydrophobic core of hGSTP1-1 likely to affect structure 45

4.3 Stability of hGSTP1-1 48

4.4 Function of hGSTP1-1 49

4.5 Influence of mutations on ligand binding 50

4.6 Conclusion 52

REFERENCES 53

LIST OF FIGURES

Figure 1.1: A diagram illustrating the homodimeric quaternary structure of hGSTP1-1 7

Figure 1.2: Structure of hGSTP1-1showing the active site in complex with S-hexylglutathione 8

Figure 1.3: Structure of hGSTP1-1 showing positions of identified SNPs 13

Figure 3.1: Pairwise sequence alignment of hGSTP1-1 wild-type nucleotides with L159M and T110S SNP forms 26

Figure 3.2A: Immobilised metal-ion chromatography (IMAC) elution profiles of hGSTP1-1 wild-type 28

Figure 3.2B: Immobilised metal-ion chromatography (IMAC) elution profiles of hGSTP1-1 L159M SNP forms 29

Figure 3.2C Immobilised metal-ion chromatography (IMAC) elution profiles of hGSTP1-1 T110S SNP forms 29

Figure 3.3: SDS-PAGE purification gel for hGSTP1-1 wild-type 30

Figure 3.4: SDS-PAGE purification gel for the hGSTP1-1 L159M SNP form 31

Figure 3.5: SDS-PAGE purification gel for the hGSTP1-1 T110S SNP form 32

Figure 3.6: A) Absorbance spectrum of the hGSTP1-1 wild-type, L159M and T110S. B) Absorbances of serial dilutions of hGSTP1-1 wild-type measured at 280 nm 33

Figure 3.7: Far-UV CD spectra for the hGSTP1-1 wild-type, L159M and T110S SNP forms 34

Figure 3.8: Fluorescence spectra of the hGSTP1-1 wild-type, L159M and T110S SNP forms 35

Figure 3.9: Elution profiles of hGSTP1-1 wild-type, L159M and T110S SNP forms 36

Figure 3.10: A) SDS-PAGE purification gel showing hGSTP1-1 wild-type, L159M and T110S after size exclusion chromatography. B) Analysis of wild-type, L159M and T110S SNP forms shoulder peaks using SDS-PAGE 37

Figure 3.11: Far-UV CD spectra of the hGSTP1-1 wild-type, L159M, T110S SNP forms as a function of temperature at a single wavelength 38

Figure 3.12A: Specific activity of hGSTP1-1 wild-type 39

Figure 3.12B: Specific activity of hGSTP1-1 L159M SNP forms 40

Figure 3.12C: Specific activity of hGSTP1-1 T110S SNP forms 40

Figure 3.13: Fluorescence emission spectra of ANS bound to hGSTP1-1 wild-type, L159M and T110S SNP forms 42

Figure 3.14: Diagram illustrating the binding region of ANS in hGSTP1-1 43

Figure 3.15: Diagram showing displacement of ANS by ethacrynic acid 44

Figure 4.1: Locations of the SNPs in hGSTP1-1 hydrophobic core 46

Figure 4.2: Diagram showing the position of the SNP in position 159 which is substituted by a methionine 47

Figure 4.3: ANS binds to the dimer interface of hGSTP1-1 51

LIST OF TABLES

Table 1.1: SNPs found in hGSTP1-1 and their implicated diseases 10

Table 2.1: Materials used at high purity and grade 15

Table 2.2: Forward and reverse primers used for mutagenesis 17

Table 3.1: Percentage activities of hGSTP1-1 wild-type, L159M and T110S SNP forms 41

LIST OF ABBREVIATIONS

ANS	8 Anilino-1-naphthalenesulfonic acid
AP-1	Activator protein-1
BCNU	N,N'-bis (2-chloroethyl)-N-nitrosourea
CD	Circular dichroism
COPD	Chronic obstructive pulmonary disease
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic
EA	Ethacrynic acid
GSH	Glutathione
GST	Glutathione *S*-transferase
hGSTP1-1	human glutathione *S*-transferase class pi
HPLC	High performance liquid chromatography
IMAC	Immobilized ion affinity chromatography
JNK	c-Jun N-terminal kinase
NBDHEX	6-(7-nitro-2,1,3-benzoxadiazol-4-ylthio) hexanol
OSCC	Oral squamous cell carcinoma
PCR	Polymerase chain reaction
ROS	Reactive oxygen species
SOD	Superoxide dismutase
TLK117	γ-glutamyl-(S-benzyl) cysteinyl-D-phenylglycine
TNF	Tumor necrosis factor
UV	Ultraviolet light
CDNB	1-chloro-2,4-dinitrobenzene
MDR1 (ABC)	MDR1 (ABC) Multi-drug resistance gene1 (ATP binding cassette transporter
Prdx6	Peroxiredoxin6
MAPEG	Membrane-associated proteins in eicosanoid and glutathione metabolism

CHAPTER 1: INTRODUCTION

1.1 Gene mutations in human

Gene mutations are changes formed permanently in the DNA sequence of the human genome causing the sequence to be different from that of other people (Cooper *et al*., 1998). Different types of mutations can occur in a genome and these mutations vary in size such that a single base or a region of DNA is changed. There are different types of mutations and these include insertions, deletions and substitution mutations (Tian *et al*., 2008). Insertion mutations occur when an extra base pair is incorporated in a gene, while deletion mutations occur when a base or a section of DNA is deleted. Substitution mutations are those in which a nucleotide is substituted with a different nucleotide (Duret, 2009). They often occur during transcription of DNA and the change in a nucleotide may result in an altered amino acid codon, which may have an effect on the newly synthesized protein (Tian *et al.*, 2008). This modification in an amino acid may produce a protein with an abnormal structure. Occasionally the structure may not be altered but its function may be affected.

Mutations that affect the protein structure or function are called missense mutations while those mutations which change the DNA sequence without any functional change in the overall protein produced are called silent mutations (Tian *et al*., 2008). The types of mutations mentioned above may be induced or spontaneous. When humans are exposed to certain environmental factors such as chemical carcinogens and ultraviolet light the mutations resulting from these factors are regarded as induced while those that occur naturally constitute spontaneous mutations (Lynch *et al.,* 2008). Spontaneous mutations occur at lower frequencies as compared to induced ones and this can be attributed to the chemical instability of purine and pyrimidine bases and copying errors during DNA replication, most often occurring when there is a foreign chemical or toxin interacting with the DNA. These spontaneous mutations may be recognized across human populations. Some of them are infrequent while some may occur more frequently in a human population (Housman, 1995).

When studied in the context of human population, those mutations that occur most frequently than one percent in the population are referred to as polymorphisms (Housman, 1995). These polymorphisms are very common and because of this reason they are thought to be normal in human DNA (Marth *et al.,* 1999). They may be large in size or smaller, meaning that multiple or single nucleotides can be altered in a sequence. When only one nucleotide has been

mutated and has shown significance across a population, those polymorphisms are called single nucleotide polymorphism (SNPs) (Taillon-Miller *et al.*, 1998).

1.2 Single nucleotide polymorphisms (SNPs) and their types

SNPs are the most frequent type of genetic variations in humans. They account for about 90% of sequence variations with an overall frequency of approximately one per 1000 bases (Tailon-Miller *et al*., 1998). SNPs have been identified to be hereditary or acquired. Some hereditary SNPs cause phenotypic variations in humans and the traits get inherited within generations. On the contrary, acquired SNPs cannot be passed on to the next person and thus have no influence on the change in generations of human populations (Cooper *et al*., 1998).

SNPs have been classified according to where they occur in the gene. The known types of SNPs are non-coding and coding region SNPs. Those on coding regions are made up of synonymous and non-synonymous SNPs (Sunyaev *et al*., 2001). Synonymous SNPs are also called silent substitutions because they do not change amino acid sequence and therefore do not cause phenotypic variations. Wang and Moult (2001) discovered that on average every gene has approximately four SNPs located on coding regions and nearly half of them lead to missense mutations in the matching protein while the additional half of these mutations are silent. These SNPs typically occur on the exon of a gene coding for a protein (Schork *et al.*, 2010). Silent mutations modify an amino acid sequence to a different codon but the function and structure of the protein remains the same. This is because the genetic code is redundant, denoting that some amino acids are coded by more than one codon. Some codons which code for a certain amino acid differ by only one base pair from other bases therefore when a mutation occurs; the usual base is substituted by another base resulting in the same amino acid (Sunyaev *et al*., 2001). Synonymous mutations therefore do not have a negative effect on the codon and overall function of the coded protein. Non-synonymous SNPs, however, may result in non-functional proteins or an incorrect structure.

They alter the amino acid sequence of a protein resulting in biological changes in living organisms and they also cause changes in DNA resulting in the change of the structure and function of proteins in the biological system. These changes may then be the basis of diseases and disorders in humans (Schork *et al.*, 2010). Depending on where the SNP is located in the protein, these SNPs may affect the proteins stability, structure and or its function. The

number of possible damaging SNPs in a human individual or population is of major concern because if the number is high, the welfare of humans can be affected.

1.2.1 Effect of SNPs on DNA coding regions

Non-synonymous SNPs alter the sequence of the protein resulting in changes of the structure which may affect the function of the protein and eventually causing a disease or a disorder. Proteins play a very important role in the biological system of humans such as getting rid of toxins, viruses and bacteria that are a threat to the body. They play a role in maintaining functional harmony of the biological system. Proteins also play distinct roles. The change in the protein sequence affects the protein negatively, sometimes mutations can improve the function of the protein or it can reduce its activity or abrogate it altogether.

The biological system of humans is exposed to a large number of toxins daily from food, drugs or the air in the atmosphere. This is why proteins responsible for detoxification are critical. Cells defend themselves from foreign particles by pumping them out of the system. Many cells are involved in the removal of these foreign particles. Some of the proteins involved in such functions are drug transporters. Drug transporters are mostly found in cell membranes and are responsible for protecting the cell against xenobiotics. The MDRI (ABCBI) gene encodes a membrane bound transporter that actively clears a wide range of compounds from the cell. Although it is a conserved gene, there is growing evidence that polymorphisms found in this gene affect substrate specificity (Gottesman and Pastan., 1993). MDRI (P-glycoprotein, ABCBI) became the first ABC-transporter identified. It falls under the ATP-binding cassette (ABC) transporter superfamily (Chen *et al.,* 1986). The C3435T SNP found in the gene displayed reduced MDRI expression and the C3435T allele was linked with two- fold reduction of MDRI expression in the duodenum (Hoffmeyer *et al*., 2000).

Enzymes contributing to detoxification play a role in protecting the body against toxins initiated by smoke from tobacco, vehicle fumes and dust. Amongst other processes and enzymes that act against the destruction of cells by toxins, is one enzyme that has been of critical importance. It destroys reactive oxygen species (ROS) and provides the balance between the production and removal of ROS. Superoxide dismutase (SOD) has been a very important target enzyme for improving the efficiency of the ROS regulation. It constitutes the first line of defense against ROS (Alscher *et al*., 2002). This metalloenzyme catalyses the conversion of superoxide radical anion (O_2^-) to two less damaging species which are oxygen

(O_2) and hydrogen peroxide (H_2O_2). This mechanism occurs in the first series of defence reactions against cytotoxic free radicals which are generated during reduction of molecular oxygen (McCord and Fridovich., 1969). H_2O_2 is also harmful to the cell and thus gets hydrolysed to water and oxygen by catalase (Alscher *et al*., 2002). The formation and removal of superoxide is balanced under normal conditions and an increase in O_2 overwhelms the cells resulting in the production of SOD under abnormal conditions (Alscher *et al.,* 2002). Most human metabolic reactions occur under aerobic conditions and it is important that oxygen is regulated and SOD will be produced anywhere in the cell where there are aerobic metabolic reactions taking place maintaining the production of these O_2 molecules. Genotypes of SOD and associations with levels of plasma SOD activity have been studied in Caucasian (German), Asian (Japanese) and African (Xhosa) populations. The implications of SOD2 A16V SNP, which is a dimorphism that leads to substitution in the mitochondrial targeting sequence was shown to significantly influence SOD plasma activity and SOD R231G. The SOD R123G causes substitution in the heparin-binding domain of SOD3 influenced total SOD activity (Lida *et al*., 2008).

A large number of studies have linked the effects that SNPs have on coding regions and the proteins produced from these DNA sequences. Amongst proteins involved in detoxification, is glutathione *S*-transferase (GST) which plays a very important role in prevention of cancers by removing toxins in the body before they interact with DNA (Abdall *et al*., 2002).

The damage caused by smoking to genes has been shown to cause susceptibility to chronic obstructive pulmonary diseases (COPD) (Ishii *et al.,* 1999). Investigation of SNPs on GST was performed in subjects with COPD and the amount of Ile105 homozygotes was discovered to be considerably higher when compared to subjects without the disease. Up to 79% of subjects with the disease were found to have acquired the SNP (Ile105). It was then concluded that these genetic polymorphisms may have a role in the formation of COPD. The dominant Ile105 polymorphism has led to this finding and to a conclusion that the human GST class P1-1(hGSTP1-1) Ile105 genotype may have less protective ability against tobacco smoke. The A105 SNP is also implicated in early-onset of coronary artery disease. This has been reported in young South African Indians (Phulukdaree *et al.,* 2012). Investigation of SNPs in GSTs therefore gives better understanding of diseases that may arise in the presence of these mutations.

1.3 Glutathione *S*-transferases (GSTs)

GSTs are enzymes under the same family that play a role in detoxifying potentially harmful endogenously derived reactive compounds in the biological system (Hayes and Pulford,1995). They achieve this role by catalysing the conjugation of hydrophobic and electrophilic substances that have a reduced glutathione. These proteins are found in living organisms such as microbes, plants, fish, birds, insects and mammals and in cell compartments such as the nucleus, mitochondria, cytoplasm and endoplasm reticulum (Hayes and Pulford, 1995).

Two distinct super families of GSTs possess transferase activities (Hayes and Strange, 2000). The first group of GST enzymes to be characterized was the cytosolic or soluble GSTs (Boyland and Chasseeaud, 1969; Mannervik, 1985). The classification of GSTs is based upon protein sequence and structure, where amino acid sequence similarity between members of the same class is greater than 40 %. At least sixteen members in this superfamily have been found in humans (Board *et al.,* 1997; Hayes and Strange., 2000) and have then been allocated to 8 families or classes based on the degree of their sequence identity. These include Alpha, Theta, Zeta, Omega, Mu, Pi, Kappa, and Sigma (Mannervik, 1985). All the cytosolic GSTs are dimers which are made of two monomers with domains I and II which form lock-and-key structures with each other (Wu and Dong, 2012). The dimeric structure of the GSTs has been suggested to improve protein stability and it stabilizes the active site for efficient catalysis (Armstrong, 2010).

Four additional classes of this superfamily are found in bacteria, insects and plants (Hayes and Strange, 2000). The second superfamily is comprised of microsomal transferases and has selected membrane-associated proteins in eicosanoid and glutathione metabolism or MAPEG in short (Jakobsson *et al*., 1999) with at least six members of this superfamily found in humans (Jakobsson *et al*., 1999). Their function is not detoxification, unlike the soluble GST enzymes. They are involved in biosynthesis of leukotrienes and prostanoids, and endogenous lipid signalling molecules (Jakobsson, 1999).

Soluble enzymes play very important roles in regulating the toxic substances that invade the body. The human GST class pi has been studied for their involvement in a broad range of functions and it has attracted particular interest because of its increased levels in cancer cells (Abdall *et al*., 2002).

1.3.1 Functions of hGSTP1-1

The hGSTP1-1 falls under the cytosolic superfamily and like other GSTs it removes toxic molecules from the body by conjugating electrophilic molecules to the GSH tripeptide resulting in these toxins being less toxic and expelled easily from the body (Armstrong, 1991). It plays a role in regulating the proportion of toxins in the body, especially those which have been associated with neurodegenerative diseases (Brambilla *et al*., 2010). The human GSTP1-1, being a soluble enzyme catalysis the detoxification of electrophilic diol epoxides. These epoxides are produced by the metabolism of polycyclic aromatic hydrocarbons (Adler *et al.,* 1999). GSTP1-1 in human is not only responsible for detoxification but also has many other functions in the cell. Firstly, they can function as molecular carriers of hydrophobic ligands, which are enzymes that enable intracellular transport of non-substrate substances in cells (Listowsky, 1993). The human GSTP1-1 moves between cell compartments such as endoplasm reticulum and cytoplasm carrying ligands (Sheratt *et al*., 1998). It has also been implicated in protein binding and binding of compounds containing iron and nitric oxide (Vasieva, 2011).

Secondly, soluble hGSTP1-1 has been shown to bind other proteins modulating their function. This is observed when hGSTP1-1 interacts with c-Jun kinase (Adler *et al*., 1999). It inhibits c-Jun N-terminal kinase (JNK) via a direct protein-protein interaction. JNK is a mitogen-activated protein kinase that is involved in apoptosis, cellular differentiation and proliferation. It is activated in the presence of protein synthesis inhibitors, ultraviolet radiation and other stress stimuli, JNK in turn phosphorylates c-Jun, a component of the activator protein-1 (AP-1) transcription factor leading to induction of AP-1-dependent target genes which are involved in cell proliferation and cell death (Laborde, 2010). hGSTP1-1 therefore is an inhibitor of signalling pathways leading to apoptosis. They are also involved in functions such as sequestering of carcinogens and modulating signal transduction pathways (Listowsky, 1993; Adler *et al*., 1999 and Cho, 2001).

Thirdly, hGSTP1-1 has been shown to activate peroxidase function in peroxiredoxin V1 (Prdx6). Heterodimerisation of prdx6 with hGSTP1-1 that is loaded with glutathione regulates the glutathione peroxidase activity (Manevich *et al.,* 2002). Prdx6 is a very important antioxidant enzyme that plays a role in protecting biological membranes against damage caused by liquid peroxidation (Manevich *et al.,* 2002). The combination of amino

acids found in the active site of hGSTP1-1 allows it to bind a variety of ligands and substrates. These functions are strongly influenced by the structure of this protein.

1.3.2 Structure of GSTP1-1

The structure of hGSTP1-1 allows it to perform many functions as mentioned above. The human GSTP1-1 is homodimeric and composed of two subunits made of 207 residues each. These subunits are folded into two domains that display different structures (Reinemer *et al*., 1992). The first domain (domain 1: residues 1-74) forms a thioredoxin-like fold while domain 2 (residues 81-207) is exclusively alpha helical, containing five alpha helices (Figure. 1) (Armstrong, 1991).

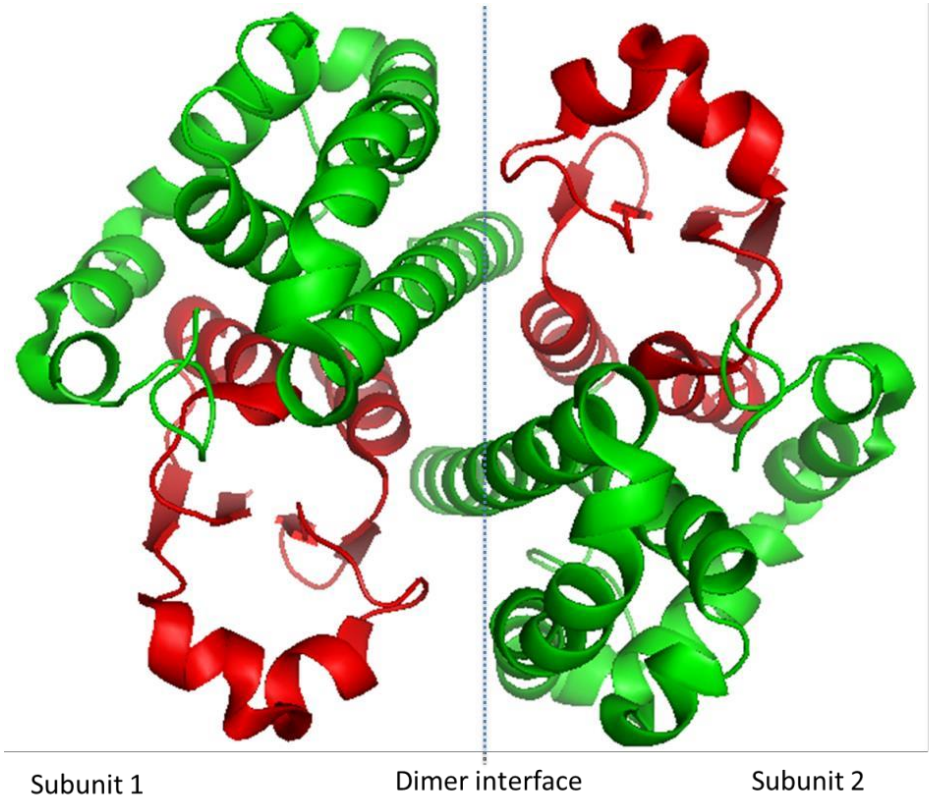


Figure 1.1 A diagram illustrating the homodimeric quaternary structure of hGSTP1-1. The two subunits each comprising of two domains is shown with the dimer interface. Domain 1 shows the theoredoxin-like fold (red) and domain 2 shows the alpha helical fold (green). The image was generated using Pymol coordinates PDB 2A2R.

The hGSTP1-1 subunits are 25 kDa in size with active sites suggested to function independently of the other (Stenberg *et al.*, 2000). Each of these dimers have an active site

suggested to contain a glutathione binding site named the G-site and the hydrophobic binding site called the H-site (Mannervik, 1985; Dirr *et al*., 1994). The hydrophobic amino acid combination in the H-site of the hGSTP1-1 strongly influences the substrates that bind to it (Armstrong, 1997). The location of the binding site for the hydrophobic substrates (H-site) was first discovered by Reinemer and his group in (1992) from the crystal structure of the human GST class pi which was co-crystallized with S-hexyl glutathione. The H-site is adjacent to the G-site and it is formed within a cleft between the N- and the C-terminal domains.

Figure 2 shows where the G-sites and H-sites are located in the structure of hGSTP1-1, the S-hexyl moiety of the S-hexylglutathione binds the H-site of the protein while the glutathione binds to the G-site (Hegazy *et al*., 2008).

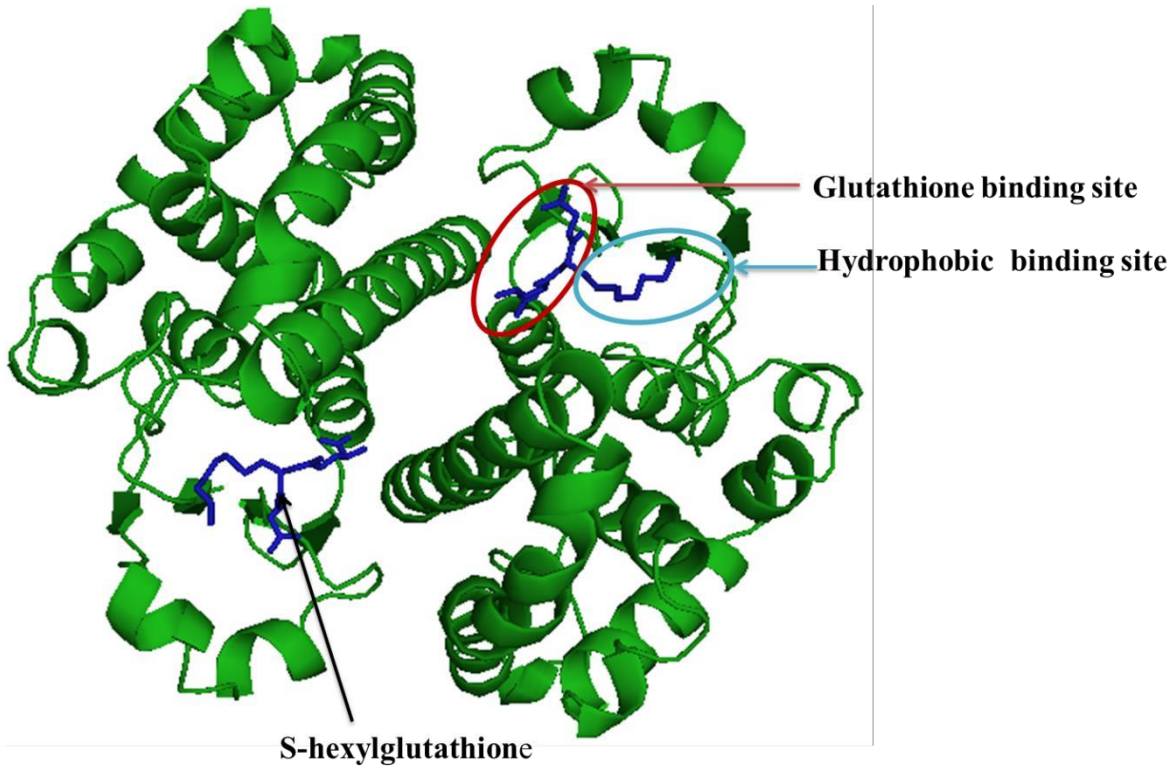


Figure 1.2 Structure of hGSTP1-1 showing the active site in complex with S-hexylglutathione. The active site of the hGSTP1-1 enzyme binds S-hexylglutathione with glutathione binding at the G- site and S-hexyl moiety of S-hexylglutathione binding at the H-site. The image was generated from PyMol using the coordinates for PDB 1GSS.

The H-site is located at the domain interface and made up of structural components from both domain 1 and 2 (Dirr *et al*., 1994). It is made up of three regions, the loop between the first β-strand and the first α-helix, the fourth α-helix and the C-terminal tail (Wu and Dong, 2012). It binds a much broader range of substrates whereas the G- site is highly specific for GSH. The H-site binds inhibitors such as ethacrynic acid (Oakley *et al*., 1997). Even though some substrates can share the H- and the G-site, the ethacrynic acid is bound entirely in the H-site, although in a manner not conducive for conjugation with GSH. The GSH and the ethacrynic acid both interact with catalytic residues in the active site by forming conjugations with the tyrosine in position 108. The hydrogen bonds form between the ketone moiety of the ethacrynic acid stabilising the Michael addition reaction (Lo Bello *et al*., 1997). A water molecule forms hydrogen bonds between Tyr7 and the sulfur atom of the GSH and the second water molecule forms a bond with the Tyr108. The hydrogen bond interaction between the GSH and the catalytic residues play a crucial role in hGSTP1-1 catalysis by stabilising the activated GSH (i.e. the GS^- thiolate) (Armstrong, 2010).

The hGSTP1-1 has SNPs identified in its sequence and certain SNPs have been found to be dominant in patients with diseases such as COPD. How these SNPs cause diseases is still to be investigated.

1.3.3 Effect of SNPs on hGSTP1-1

SNPs can affect both the structure and function of hGSTP1-1 leading to disorders and diseases in human. There are diseases that are thought to be caused by SNPs in DNA level, affecting the production of a protein (Tian *et al*., 2008). The human GSTP1-1 causes susceptibility to triple-negative breast cancer cell metabolism and pathogenicity (Louie *et al.*, 2016) and SNPs in the protein also influence cause of diseases. SNPs found in DNA coding for hGSTP1-1 were found to change the detoxification process leading to elevated accumulation of carcinogenic metabolites and therefore increasing the risk of breast cancer sensitivity in women (Masoodi *et al*., 2012, Onay *et al*., 2006). The protein is also affected by SNPs, for example subjects with chronic obstructive pulmonary diseases (COPD) were found to have high levels of the Ile105 SNP as compared to subjects without the disease (Ishii *et al*., 1999).

Table 1.1 shows certain single nucleotide polymorphism found in hGSTP1-1 and their effects on protein functions and health of humans.

Table 1.1 SNPs found in hGSTP1-1 and their implicated diseases

GSTP1-1 polymorphisms	Effect of SNPs on human health	References
E32V	Occurs more than 1 % in Mexican Americans	Moyer *et al*., 2008
D58N	Effect on humans has not been determined	Moyer *et al*., 2008
I105V	More significant in asthmatic patients and causes susceptibility to asthma	Karam *et al*., 2012 Hanene *et al*., 2006
	Lower enzyme activity and causes less effective detoxification and more susceptibility to neoplasms	Vasieva, 2011
	Occurs more than 10 % in African Americans, Caucasians Americans and Han American Chinese	Moyer *et al.*, 2008
	Higher among diabetic patients and diabetic patients without vascular complications	Zaki *et al.*, 2013
Deletion of Ile105	Was found to abolish enzyme activity and increase susceptibility to oxidative stress	Hanene *et al*., 2006
A114V	Increases the risk of developing oesophageal squamous cell carcinoma. Associated with tobacco smokers	Li *et al.*, 2010
D147Y	Not determined	Moyer *et al.*, 2008

F151L	Substitution of F151L near a conserved folding module has been found to lower stability of GSTP1-1	Lin *et al*., 2003
S185S	Not determined	Lin *et al*., 2003
R187W	Not determined	Moyer *et al*., 2008

1.4 Human GSTP1-1 and its relation to cancer

SNPs in hGTSP1-1 have been linked to several diseases as mentioned above, mostly cancers because SNPs alter the DNA sequence. There are many different cancers that have been implicated in hGSTP1-1, such as cancers of the bladder and testicles (Harries *et al*., 1997) and lung cancer (Ryberg *et al*., 1997). A large number of human tumours such as colon cancers, pancreas, uterine cervix, stomach and breast cancer have shown a raised expression of hGSTP1-1 enzyme (Coles and Ketter, 1990, Tsuchida and Sato, 1992). Human GSTP1-1 methylation and expression has been reported in prostate cancer (Martignano *et al*., 2016). The hGSTP1-1 not only has been shown to increase in patients with certain SNPs but its expression in tumour cells is correlated with resistance of these tumours to anti-cancer drugs (Laborde, 2010). This demonstrates that the protein defends these tumours from being destroyed by chemotherapeutics such as chlorambucil, cyclophosphamide, cisplatin and N,N'-bis (2-chloroethyl)-N-nitrosourea (BCNU). It achieves this by direct detoxification of the drugs and by blocking apoptosis of the cancer cells by interacting with c-JNK kinase. It is therefore necessary to find specific inhibitors of the hGSTP1-1. γ-glutamyl-(S-benzyl) cysteinyl-D-phenylglycine (TLK117) and has been reported to be the most selective hGSTP1-1 inhibitor (Lyttle *et al*., 1994). It has a K_i of 0.4 μM for hGSTP1-1 as compared other classes such as Alpha and Mu (Koehler *et al.*, 1997). The basis for the difference in the K_i values for these classes can be attributed to the D-phenylglycine moiety of this inhibitor which forms contacts with the hydrophobic residues while maintaining the multiple hydrogen-bond interactions that occur between the GSH and the G-site in hGSTP1-1. The binding of the inhibitor to either the Mu or Alpha class, however, causes steric clashes with the C-terminal tail because of the substituted glycyl moiety in GSH by the D-phenylglycine group (Oakley *et al*., 1997). When the inhibitor is bound to the Alpha class, the bulky modification would have a steric clash with the Phe220 and Phe222 residues (Sinning *et al*., 1993) whilst in the case of Mu class the clash would come from interactions with Trp7,

Met43 and Arg42 residues (Ji *et al.,* 1992). A specific inhibitor for the hGSTP1-1 would be of great importance for the cancer patients and those cancers associated with specific SNPs. Several GSTP1-1 inhibitors have been reported lately. The tyrosine-reactive irreversible inhibitor has also been shown to inhibit the protein (Crawford and Weerapana, 2016). It contains a dichlorotriazine that irreversibly inhibits hGSTP1-1. It is selective for GSTP1-1 within cellular proteomes and the Y108 residue in the protein allows for the inhibitor to bind (Craford and Weerapana, 2014). An anti-cancer compound called piperlongumine (PL) also shows inhibition of GSTP1-1 (Harshbarger et al., 2017). The discovery of all these inhibitors will also improve cancer therapy against cancers caused by SNPs across populations.

1.5 Population distribution of human diseases associated with SNPs found in hGSTP1-1

The dominance of a certain SNP in humans can be influenced by the environment they are in, their lifestyles and mostly by genetic heredity. This is also the reason why some diseases affect many people in some areas than others. Cases that show prevalence of diseases in certain areas when compared to others have been documented. For example, in South Africa the hGSTP1-1 SNP substitution, A114V amino acid change has been found to cause an increase in the risk of developing oesophageal squamous cell carcinoma and tobacco smokers are more vulnerable (Li *et al*., 2010). Additionally, a study done by Phulukdaree and his group (2012) in the same country showed higher frequencies of hGSTP1-1 Ala105 in patients with coronary artery diseases.

SNPs such as the hGSTP1 I105V polymorphism has been associated with susceptibility to asthma in Egyptians (Karam *et al*., 2012). The I105V SNP is very dominant and has been implicated in many diseases in different countries. A study done by Moyer and his group (2008), had further shown that the I105V occurs greater than 10 % of the time in African Americans, Caucasian Americans, Han American Chinese and Mexican Americans, while the E32V occurs less than one in Mexican Americans. SNPs are responsible for a number of diseases and there may be more diseases that have not been discovered yet that are caused by them. SNPs can also have a positive effect on the function of the protein by either improving its function or having a stabilising effect on the protein.

Furthermore, Manevich and his group in 2012 did a study from *in silico* modelling of hGSTP1-1 peroxiredoxin (Prdx6) heterodimer, they have shown that the Ile105 and Ala114

polymorphic sites of hGSTP1-1 are in close proximity to the binding interface, and thus these polymorphic variants may have a role in regulating the function of Prdx6 which may further influence human population susceptibility to oxidative stress. Their effect on the structure and function of the affected protein is likely to contribute towards these disorders. It is therefore very important to study these effects on proteins such as hGSTP1-1. SNPs which have been identified in hGSTP1-1 for this study are displayed in Figure 1.3 below.

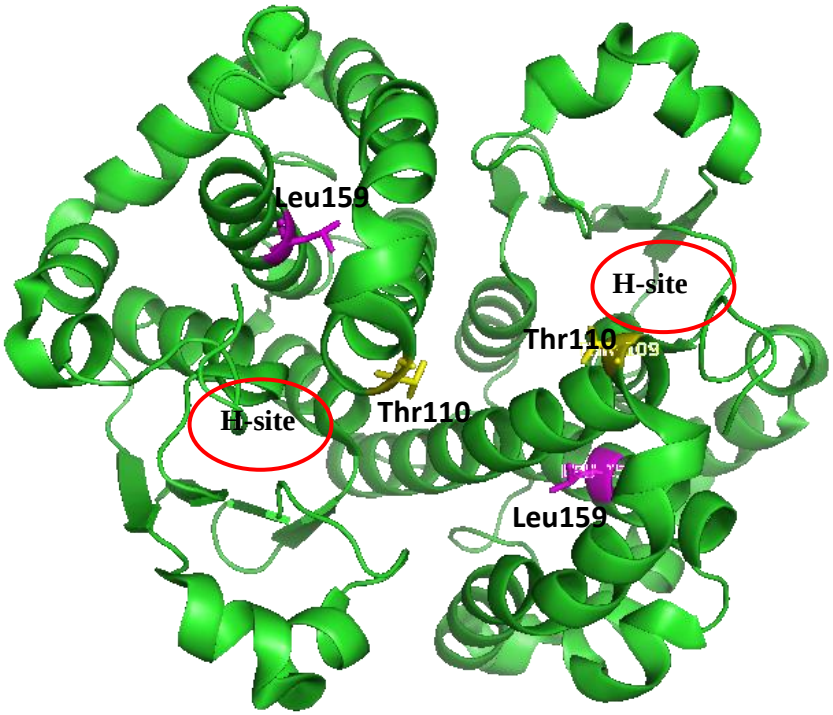


Figure 1.3 Structure of hGSTP1 showing positions of identified SNPs. The positions of amino acids Thr110 (yellow) and Leu159 (magenta) are shown in the diagram above. The T110S is located closely to the H-site (red circle) of the hGSTP1-1 and the L159M is located in the hydrophobic core of domain 2 of hGSTP1-1 in α-helix 6.

In this study, effects caused by SNPs in the human GSTP1-1 are investigated. An SNP in its hydrophobic core and one that is located closest to the H-site of the protein are studied. The hydrophobic core is suspected to play a critical role in the folding process of this protein and contributes to the interaction between the GSH on its active site (Luca *et al.*, 2014, Dirr *et al.*, 2005, Cocco *et al.*, 2001, Dragani *et al.*, 1997). T110S SNP is located in the H-site of the protein and close to a catalytic residue Tyr109. Therefore this study focuses mainly on how the structure and function of this protein is affected by these SNPs.

1.6 AIM AND OBJECTIVES

1.6.1 Aim

This study is aimed at investigating the effects that naturally occurring single nucleotide polymorphisms found in the hGSTP1-1 have on its structure, stability and function.

1.6.2 Objectives

1.6.2.1 Design reverse and forward DNA primers for site-directed mutagenesis to generate the SNP forms of the wild-type hGSTP1-1.

1.6.2.2 Sequence the SNP forms in the DNA sequence to confirm the presence of codon changes for each SNP.

1.6.2.3 Express and purify the wild-type and SNP hGSTP1-1 proteins.

1.6.2.4 Determine the structural properties of the mutated protein and compare them to those of the wild-type hGSTP1-1.

1.6.2.5 Perform CDNB-GSH assay and ANS binding studies to determine enzymatic functions of the SNP forms and compare them to those of the wild-type hGSTP1-1 protein.

1.6.2.6 Perform computational docking studies to determine the binding sites of the ANS as a hGSTP1-1 ligand.

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials and Methods

The N-terminal His 6-tagged human GSTP1-1 sequence which was used for this study was a gift from S. Y. Blond (Centre for Pharmaceutical Biotechnology, University of Illinois, Chicago, IL). The hGSTP1-1 sequence was cloned into pET-15b vector, and the presence of the hGSTP1-1 sequence in the pET-15b vector was confirmed by cloning the vector into XL10 Gold® ultra-competent (Stratagene, La Jolla, CA, USA) cells and grown to produce more DNA which was extracted using GeneJET™ Plasmid Miniprep Kit (Fermentas, Ontario, Canada) and sent for sequencing at Inqaba Biotech (Pretoria, South Africa). The SNP forms were designed by PCR and identity of all the plasmids was confirmed by sequencing at Inqaba Biotech. The hGSTP1-1 wild-type and SNP forms were expressed using T7 Express Iq *E. coli* competent cells to obtain protein and were purified using the ÄKTA FPLC (fast protein liquid chromatography). The purified wild-type and SNP form were used for functional and structural characterisation. The reagents used were of analytical grade and all solutions were prepared using distilled water and filtered through a 0.45 µm pore size filter. Materials needed at high purity and grades are listed in Table 2.1.

Table 2.1 Materials used at high purity and grade

Material	**Source**	**Location**
Primers for mutagenesis	Inqaba Biotec	Pretoria, South Africa
Quikchange® Lightening Mutagenesis Kit	Stratagene	La Jolla, CA, USA
GeneJET™ Plasmid Miniprep Kit	Fermentas	Ontario, Canada
SDS-PAGE molecular mass markers	Fermentas	Ontario, Canada
DNA Ladders	Fermentas	Ontario, Canada
Isopropyl β-D-thiogalactopyranoside (IPTG)	Sigma-Aldrich	St. Louis, MO, USA
Coomassie Brilliant Blue G-250	Sigma-Aldrich	St. Louis, MO, USA
Dithiothreitol (DTT)	Melford Laboratories Ltd.	Suffolk, UK
5 ml Histrap purification column	GE Healthcare	Buckinghamshire, UK
Gel filtration standards	Bio-Rad	Johannesburg, South Africa
T7 Express Competent Cells	New England Biolabs	Ipswich, MA, USA

Hi Load™ 16/600 Superdex™ 75 pg	GE Healthcare	Buckinghamshire, UK
Ultracel® 10 kDa Ultrafiltration Discs	Amicon Bioseparations	Jeffery, NH USA
8 Anilino-1-naphthalenesulfonic acid (ANS)	Sigma-Aldrich	St. Louise, MO, USA
Reduced L-glutathione (GSH)	Sigma-Aldrich	St. Louis, MO, USA
1-Chloro-2,4-dinitrobenzene (CDNB)	Sigma-Aldrich	St. Louis, MO, USA
XL10 Gold® ultra-competent	Stratagene	La Jolla, CA, USA

2.2 Experimental procedures

To obtain the plasmid containing the hGSTP1-1 sequence, the human GSTP1-1 sequence was cloned into pET-15b vector which was used to transform XL10 Gold ultra-competent *E. coli* (Stratagene) cells. The hGSTP1-1 DNA was synthesized by growing the cells and extracting the DNA which was used to make the SNP proteins using PCR. The wild-type hGSTP1-1 and the SNP DNA were used to transform T7 *E. coli* cells which were used for expression of the proteins. The purified protein was then used to characterize the structural and functional properties of the wild-type protein while comparing them to those of the L159M and T110S SNP forms. The detailed methods of the experiments performed are described below.

2.3 Identification and construction of hGSTP1-1 SNPs

SNPs common in hGSTP1-1 were identified on the dbSNP data base (http://www.ncbi.nlm.nih.gov/projects/SNP) and an SNP located in the hydrophobic region of the hGSTP1-1 (L159M) and the one closest to the H-site (T110S) were selected. These were chosen to investigate how SNPs in the hydrophobic core of hGSTP1-1 affects its structure and possibly affect its function, primarily because the hydrophobic core of hGSTP1-1 has been found to have a role in the folding and stability of the protein (Luca *et al*., 2014, Dirr *et al.*, 2005, Cocco *et al*., 2001, Dragani *et al*., 1997). SNPs closest to the hGSTP1-1 H-site catalytic residue Tyr109 which plays a role in stabilising the binding of substrates to the active site (Prade *et al*., 1997) also raise questions as to whether they affect the binding of substrates and ligands to the protein and also possibly affect its function. T110S SNP, located close to the Tyr109 catalytic residue was selected to investigate its effect on the protein. These SNPs have high frequencies and have been validated by either cluster, frequency or hapmap (http://www/ncbi.nlm.nih.gov/snp). The accession numbers for the L159M and T110S SNPs are rs4986948 and rs199567349, respectively. Primers used for the construction of these SNPs

were designed using the QuikChange primer designer by Agilent Technologies: (http://www.genomic.agilent.com/primerDesignProgram.jsp?&_requestid=103276).

The reverse and forward primer sequences (Table 2.1) were synthesized at Inqaba Biotech.

Table 2.2 Forward and reverse primers used for mutagenesis

Name	**Sequence (5'-3')**
T110S-f	5'-cccgcctcatagttgctgtagatgagggaga-3'
T110S-r	5'-tctccctcatctacagcaactatgaggcggg-3'
L159M-f	5'-tcatggatcagcagcatgtccagcaggttgtag-3'
L159M-r	5'-ctacaacctgctggacatgctgctgatccatga-3'

These primers were used for mutagenesis using a Bio-Rad MyCycler thermal cycler machine and a Quick Change Lightning Site-Directed Mutagenesis kit from Agilent technologies. The experiment was performed as described by Zheng *et al*., (2004). To ascertain that only the desired mutations had occurred during the experiments, the open reading frame of the plasmid encoding hGSTP1-1 wild-type and each SNP form DNA was sequenced at Inqaba Biotech. The pET-15b plasmid is 5.7 kb in size and it has an *AmpR* gene which allows the *E. coli* to be resistant to ampicillin which further allows selection of only cells containing the plasmid.

2.4 Expression of hGSTP1-1 wild-type and SNP forms

The vectors containing the cDNA hGSTP1-1 sequences and those containing the L159M and T110S SNPs were used to transform 50 µl XLGold® Ultra-competent *E. coli* cells and the T7 Express I*q* competent cells. The cells were transformed using an updated transformation method adapted from (Froger and Hall, 2007). The 50 µl competent cells were left to thaw on ice for 15 minutes before adding 2 µl of 90 ng/µl plasmid DNA. The solution was gently mixed before heat shocking the cells for 10 seconds at 42 ºC in a heating block to disrupt the bacterial cell wall so it can take up DNA and the cells were immediately placed on ice for 5 minutes to repair their cell walls from injury caused by heat shock. The competent cells were then added to 950 µl SOC media (2 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 2.5 mM KCl, 10 mM glucose, 20 mM $MgCl_2$, 20 ºC) and then grown in an incubator for an hour at 37 ºC while agitating at 250 rpm. The transformed *E. coli* cells were then plated on agar plates (2.3 % (w/v) plate count agar) (Merck) containing 50 µg/ml carbenicillin and left to grow overnight

(~16 hours). Single colonies were picked from each of the wild-type, L159M, T110S SNP plates and used to perform the expression of hGSTP1-1 protein.

The expression of hGSTP1-1 wild-type protein, L159M and T110S SNP forms was done using *E. coli* T7 cells. The *E. coli* T7 cells containing the vector pET-15b which has the hGSTP1-1 sequence were grown overnight in 2YT media (1.6 % (w/v) tryptone, 1.0 % (w/v) yeast extract and 0.5 % (w/v) NaCl). The 98 ml of media in 500 ml were inoculated in 2 ml overnight culture making 1:50 dilutions and 50 µg/ml of carbenicillin was added. The culture was grown in an incubator operated at 230 rpm at 37 ºC until mid-log growth phase (OD_{600} = 0.6) at which point 0.2 mM IPTG was added to induce expression. The cells were then allowed to express the proteins for 6 hours and the cells were harvested by centrifugation for 20 minutes at 4000 x *g* in Sorvall Lynx 4000 centrifuge (Thermo scientific) and re-suspended in buffer (50 mM tris-HCl, 0.5 M NaCl, 0.02 % (w/v) sodium azide, 50 mM imidazole) at pH 8. The two hGSTP1-1 SNP forms were induced with 0.2 mM IPTG at a starting temperature of 30 ºC and then 0.2 mM IPTG was added every 2 hours. The temperatures were also reduced by 2º C every 2 hours. After 6 hours of incubation the cells were harvested and stored in the –80 ºC freezer until use for purification.

2.5 Purification of hGSTP1-1 wild-type and SNP forms

The cells were removed from the freezer and thawed in ice. In 50 ml of the cell culture, 10 µg/ml DNAse was added, with 10 mM $MgCl_2$ and 1 mg/ml lysozyme to lyse the cells, remove DNA and solubilise the protein. The solution was incubated at 30 °C for 30 minutes in the incubator. Afterwards, the cells were disrupted by sonication using the XL2000 Series Misomix ultrasonic liquid processor (Lasec). The soluble fraction was obtained by subsequent centrifugation at 9000 x *g* for 30 minutes using the Heraeus Megafuge 16 centrifuge (Thermo scientific). The supernatant was filter sterilized using 0.2 µm filters. The supernatant was loaded onto a 5 ml HisTrap purification column (GE Healthcare), pre-equilibrated with equilibration buffer (50 mM Tris-HCl, pH 8, 50 mM imidazole, 0.5 M NaCl and 0.02 % (w/v) sodium azide) and purified with the ÄKTA FPLC (fast protein liquid chromatography) purification system (GE Healthcare). The His-tagged hGSTP1-1 protein binds to the positively charged nickel forming a metal complex between the immobilized transition metal ion bound to the resin and imidazole group of histidine. The low concentrations of imidazole in the equilibration buffer remove weakly bound proteins from the column. The 6-histidine amino

acids allow for binding of the hGSTP1-1 protein to the nickel column. The protein was eluted with elution buffer (50 mM Tris-HCl, 0.5 M NaCl, 0.02 % (w/v) sodium azide, 300 mM imidazole). The high concentration of imidazole in the elution buffer out-competes the tagged protein for a binding space to the nickel ions in the His-Trap column, thus eluting the protein. The fractions collected were run on SDS/PAGE gels and then dialysed in storage buffer (20 mM sodium phosphate, 0.15 mM NaCl, 1 mM EDTA, 2 mM DTT and 0.02 % (w/v) sodium azide, pH 7.4) overnight at 4 ºC then samples were stored in -20 ºC.

The presence of the protein in all purification experiments was determined by measuring the absorbance of the eluent at 280 nm in real time, using the inbuilt spectrophotometer of the AKTA prime automated chromatographic system which is coupled to the computer with PrimeView 1.0 software (GE Healthcare Life Sciences, Uppsala, Sweden).

2.6 Protein concentration determination

The concentration of the hGSTP1-1 wild-type and the two SNP forms (L159M and T110S) were measured using a Jasco FP-6300 UV/VIS absorbance spectrophotometer and calculated using the Beer-Lambert law:

$$A = \varepsilon_{\lambda} CL$$

where A is the absorbance at 280 nm, ελ is the molar extinction coefficient (M^{-1} cm^{-1}) at the specific wavelength. And C is the molar concentration with l being the path length (cm). The maximum absorbance for each of the protein samples, wild-type and SNP forms was used to calculate the concentration using the formula above. Substitution of the corrected absorbance for the wild-type, L159M and T110S absorbance spectra, the extinction coefficient, and the path length (1 cm) into the Beer Lambert law (see above) yields the molar concentration (M) of soluble protein.

The absorbance at 280 nm was also measured for 6 fold serial dilutions of the wild-type, L159M and T110S SNP forms and the concentrations were determined by fitting a linear regression on the graph. All the absorbance readings were buffer corrected by subtracting the blank from the sample.

2.7 Sodium dodecyl sulfate –polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE is a detergent with an anionic charge; it binds to proteins and denatures them into individual polypeptide molecules (Pitt-Rivers and Impiombato, 1968). It separates proteins according to size allowing visualisation of different proteins in a sample upon staining with a dye such as coomasie blue, their solubility and most importantly their purity (Laemmli, 1970). A soluble protein will be in the supernatant while insoluble protein will remain in the pellet. When run on acrylamide gels, the movement of the denatured proteins is based upon charged particles in an electric field by electrophoresis. Proteins are separated solely by charge due to the sieving effect of the pores of the acrylamide and bis-acrylamide (Summers, 1965). The SDS-PAGE gel electrophoresis method consists of two different levels of gels with different pHs and different acrylamide percentages, which is the stacking gel and separating gel. The stacking gel stacks up proteins before being separated by the separating gel and this provides enhancement to the separation of the proteins. The fractions collected after purification were run on SDS-PAGE at 100 V. The stacking gel consisted of (4 % (w/v) acrylamide, 0.36 % (w/v) bis-acrylamide, 0.05 M Tris-HCl pH 6.8, 0.1 % (w/v) SDS, 0.005 % (w/v) ammonium persulfate and 0.2 % (v/v) TEMED) and 16 % acrylamide separating gel. Samples were diluted in a 1:1 ratio with reducing buffer (25 mM Tris-HCl pH 6.8, 4 % (w/v) SDS, 20 % (v/v) glycerol, 10 % (v/v) β-mercaptoethanol and 3.5 μg/mL Coomasie Blue R250) before loading. The gels were then stained with Coomassie Blue R250 staining solution containing (0.1 % (w/v) coomasie dye in 1: 5: 4(v/v/v) acetic acid: methanol: water) and de-stained with (1: 4 (v/v) glacial acetic acid: methanol) until the background was clear.

2.8 Structural characterisation of hGSTP1-1 wild-type and SNP forms

2.8.1 Secondary structure characterization by far-UV circular dichroism (CD)

CD was used to carry out analysis of the secondary structure of hGSTP1-1. The phenomenon determines whether a protein is folded and therefore characterises its secondary structure. It establishes the secondary structure in the far UV spectral region (190-250 nm) where the chromophore in this wavelength is the peptide bond and the signal results when it is in a systematic folded nature (Greenfield, 2006). It measures differences in the absorption of left-handed polarized light from that of right-polarized light resulting from structural asymmetry. The CD spectra were obtained on a Jasco J-710 spectrophotometer. The bandwidth was set at 1.00 nm and a scan speed of 200 nm/min was used. All protein samples were

analysed in 5 mM sodium phosphate buffer at pH 7.4 at 20 ºC. Data were converted to [θ] mrw as described by Schmidt (1989).

2.8.2 Tertiary structure characterization by fluorescence spectroscopy

Fluorescence is an effect of a process with three stages. An appropriate biological specimen must possess a fluorescent probe called fluorophore within a specific region within it. Upon excitation by a photon of energy such as an incandescent lamp or laser, the fluorophore absorbs the photon of energy. This is the first stage of fluorescence called the excitation stage. Then the excited stage exists for a short period of time (1-10 nanoseconds) while a fluorophore undergoes conformation changes and also forms interaction with its molecular environment. Then in the emission stage the fluorophore emits the photon of energy (Gordon *et al*., 1994). There are three known amino acids with intrinsic fluorescent properties; tyrosine, phenylalanine and tryptophan. In the native state of most proteins tyrosine emission tends to be quenched to nearby tryptophan residues as a result of fluorescence resonance energy transfer (FRET). Phenylalanine is not usually used for studies because of lower quantum yields (Lakowicz, 1999). Tryptophan, therefore contributes the biggest fluorescence signal in proteins. Tryptophan fluorescence was used as a probe for the comparison of the tertiary structure of wild-type hGSTP1-1 and the two SNP forms. Tryptophan residues are very useful for analysing the nature of the protein; this is because the fluorescence of the indole ring side chain residue is sensitive to their micro-environment (Lakowicz, 1999). Furthermore, the location of the tryptophan residues affects the fluorescence and therefore can provide for local and global conformational changes of a protein. All fluorescence measurements were made at 20 ºC in a Jasco FP-6300 spec. The excitation band and emission band passes were both set at 5 mM. Fluorescence measurements were made in 5 mM sodium phosphate buffer at pH 7.4. The excitation of tryptophan residues was measured at 295 nm.

2.8.3 Size-exclusion chromatography

The quaternary structure of the hGSTP1-1 wild-type and SNP forms were examined using size exclusion chromatography. This technique separates proteins based on their molecular size (Hagnauer, 1982) and is achieved by differentiation exclusion from the pores of the resin in the column while passing through it. The column is comprised of stationery and a mobile phase. The stationery phase contains porous gel beads which only allow molecules of different sizes

to pass through. The mobile phase consists of a buffer or solvent that maintains favourable conditions for the molecules being analysed (Preneta, 1993). Larger molecules are eluted from the column at the early stages of the experiment because of their limited interaction with the stationery phase while the smaller molecules remain in the column longer (Potschka, 1987). Gel filtration is an excellent tool for separating monomer, oligomer or aggregated forms of a target protein. The Hi Load™ 16/600 Superdex™ 75 pg column was used with KTA protein purification system, the column was equilibrated overnight at 20 ºC with equilibration buffer (20 mM sodium phosphate, 1 mM EDTA and 0.02 % (w/v) sodium azide). The proteins were concentrated with Ultracel® 10 kDa Ultrafiltration Discs. Before loading the samples, the 2 ml superloop was washed with 10 ml volumes of equilibration buffer and 1 ml 60 µl protein samples and 500 µl Bio-rad gel filtration standards consisting of thyroglobulin (from bovine; 670 kDa), γ-globulin (from bovine; 158 kDa), ovalbumin (from chicken; 44 kDa) and myoglobin (from horse; 17 kDa) were loaded and run for 2 hours separately. The AKTA was run at 1 ml/min flow-rate and 0.3 MPa pressure limit. The protein eluents were collected and run on SDS-PAGE.

2.9 Stability studies

The stability of the wild-type and the SNP forms was assessed using CD by varying temperatures to determine the heat in which the proteins are least stable leading to unfolding. The far UV spectra to determine the secondary structure of the proteins was measured to confirm the folded state of the proteins before thermal melting. The hGSTP1-1 has high α-helical content with peaks at 222, 208 and 193 nm. The wavelength at 222 was used because data at higher wavelengths usually have lower absorption and higher signal to noise (Greenfield, 2006). The temperatures were set to vary from 0 – 80 ºC and to determine whether the proteins unfolding were irreversible the reverse scan was set from 80 – 0 ºC. The protein samples were prepared in 5 mM sodium phosphate buffer at pH 7.4. Data were converted to [θ] mrw as described by Schmid (1989).

2.10 Functional characterisation of hGSTP1-1 wild-type and SNP forms

2.10.1 GST/CDNB activity assay

GST activity was assayed spectrophotometrically at 20 ºC using a Jasco FP-6300 absorbance spectrophotometer with a 10 mm UV quartz micro cuvette to measure the rate of conjugation

of GSH to CDNB. Final substrate concentrations were 1 mM glutathione (GSH), 1 mM Chloro-2,4-dinitrobenzene (CDNB) in ethanol and 100 mM sodium phosphate buffer containing 1 mM EDTA and 0.02 % (w/v) sodium azide at pH 6.5. The initial protein concentration was 152 nM and final concentrations in a test tube were varied from 1 to 10 nM. Substrates were added into the 100 mM sodium phosphate buffer pH 6.5 in the order GSH, enzyme then CDNB. The mixture was quickly mixed and measured at 340 nm for 60 seconds. The GST activity was calculated in µmol CDNB conjugated/min using the published extinction co-efficient 9.6 mM^{-1} cm^{-1}. Specific activity was calculated by correcting for protein content and the specific activity was expressed in µmol/min/mg.

2.10.2 8-Anilinonaphthalene-1-sulfonic acid (ANS) binding

ANS is an amphipathic anionic dye which can be used as a probe for the detection of nonpolar surface on proteins (Abdalla *et al.,* 2002). The fluorescence properties of this molecule are largely influenced by the polarity of its surrounding environment. The fluorescence emission maximum of ANS shifts to a lower wavelength when bound to the hydrophobic surface, and this bathochromic shift is accompanied by an increase in fluorescence intensity (Slavik, 1982). A 16 mM ANS stock solution was prepared in 20 mM sodium Phosphate buffer, with 1 mM EDTA and 0.02 % sodium azide, at pH 7.4. The concentration of ANS was confirmed using the Beer-Lambert law by recording the absorbance at 350 nm and using an extinction coefficient of 5000 M-1.cm-1 (Weber and Young, 1964). The ANS fluorescence spectra, in the absence and presence of 2 µM GSTP1-1 were recorded at 20 °C, buffer was corrected by subtracting the scan of buffer from the test sample, and an average of three accumulations at a scan speed of 200 nm.min-1 were recorded. The samples were excited at 390 nm and emission spectra were recorded from 400 to 700 nm. The excitation and emission slit widths were at 5 nm.

The ANS binds to the nonpolar regions of the protein and computational docking studies were employed to determine specific regions in hGSTP1-1 where ANS binds. The crystal structure of the hGSTP1-1 was retrieved from the protein data bank PDB 5J41 and was chosen because of its high resolution of 1.19Å (Harshbarger *et al.,* 2017). It was made of the dimer with water molecules bound. Protein preparation was then performed on the dimeric structure of the hGSTP1-1, this was carried out by removing water molecules, and assigning hydrogen bonds to the structure followed by formal charges, ionization and tautomeric states. The non-

hydrogen bonds were minimized until they reached an RMSD of 0.3A while the amino acids flips were assigned. Induced flexible docking was carried out using Glide software (Schrödinger LLC 2009, USA). This method was chosen because it allows the ligand to move freely, overcoming problems which are usually encountered when the ligand is rigid. A rigid ligand typically gives misleading results in conventional docking methods. The induced flexible docking is the most effective and accurate method used for predicting ligand binding and concomitant structural movements. Induced flexible docking also conforms more closely to the shape and binding mode of the ligand. The computational docking studies were run overnight and the results were retrieved. The energetically favourable docking score, glide energy and glide emodel were retrieved.

Molecular docking studies have shown ANS to bind the dimer interface of hGSTP1-1. This is when the ANS is left to freely bind the region of hGSTP1-1 with highest affinity (Induced fit docking). While studies have also shown a possibility of ANS binding to the active site, these studies were confirmed untrue with ANS displacement studies. The ANS ligand has hydrophobic aniline and naphthyl rings which have been shown to occupy the non-polar H-site in some GST proteins such as GSTA1-1 (Dirr *et al*., 2005). The displacement of ANS by a ligand also binding to the H-site from the hGSTP1-1 will give insight on where the ANS binds. Ethacrynic acid has been shown to also bind the H-site of hGSTP1-1 and therefore was used for the displacement and determination of the binding site of ANS. Fluorescence measurements of ANS were performed at 20 ºC and exciting at 390 nm.

Free ANS has lower fluorescence intensity and an increase is observed when it is bound to the hydrophobic regions of hGSTP1-1 (Ghisaidoobe and Chung, 2014). When ANS is displaced from the proteins active site a decrease in fluorescence intensity is observed. Fluorescence emission was observed at 485 nm. A control run was done where absolute ethanol was titrated in a 1ml solution containing 100 µM ANS in a quartz cuvette and 2 µM hGSTP1-1 protein, titrations were done from 0 to 6 % ethanol without the presence of ethacrynic acid. The test run contained 100 µM ANS and increasing concentrations of ethacrynic acid dissolved in ethanol up to 8 mM with 2 µM protein.

CHAPTER 3.0: RESULTS

The hGSTP1-1 wild-type, L159M and T110S SNP recombinant protein forms obtained were used for the structural and functional characterisation studies. The purity of the proteins was stained with coomasie brilliant blue and visualised using SDS-PAGE and concentrations were determined using UV/VIS spectrophotometer followed by structural and biochemical assays.

3.1 Identification and construction of hGSTP1-1 SNP forms by site directed mutagenesis

Single nucleotide polymorphisms found in hGSTP1-1 were identified from the NCBI website (www.ncbi.nlm.nih.gov) (August 2015). The L159M SNP is found in the hydrophobic region of domain 2 of the protein in α-helix 6. The T110S SNP is also found in domain 2 and is closely positioned next to the Tyr108 found in the active site mainly the H-site of the protein which stabilizes the binding of ligands to the active-site. The SNPs were identified on the dbSNP (NCBI) data base and the designed T7 primers which were used to incorporate these SNPs in hGSTP1-1 sequence.

3.2 Site directed mutagenesis

Site-directed mutagenesis was performed using polymerase chain reaction (PCR) to incorporate the selected nucleotides to form hGSTP1-1 SNP forms: L159M and T110S. The diagrams in Figure 3.1A demonstrate sequence alignment of the hGSTP1-1wild-type aligned with the L159M and T110S SNP forms retrieved from Blast sequence alignment (NCBI). The sequences were visualised using FinchTV application after sequencing at Inqaba Biotech. Inqaba Biotech conducted all sequencing to confirm identity of the pET-15b open reading frame which encodes for the hGTSP1-1 wild-type, L159M and T110S SNP forms and the presence of any other mutations. The resultant hGSTP1-1 sequence showed 99 % similarity for both L159M and T110S SNP forms (Figure 3.1A 1 and 2). A gap was observed which could have formed from the fragments which were not sequencing.

Constructs expressing wild-type hGSTP1-1 and its SNP forms were used to transform cells towards recombinant protein production and purification.

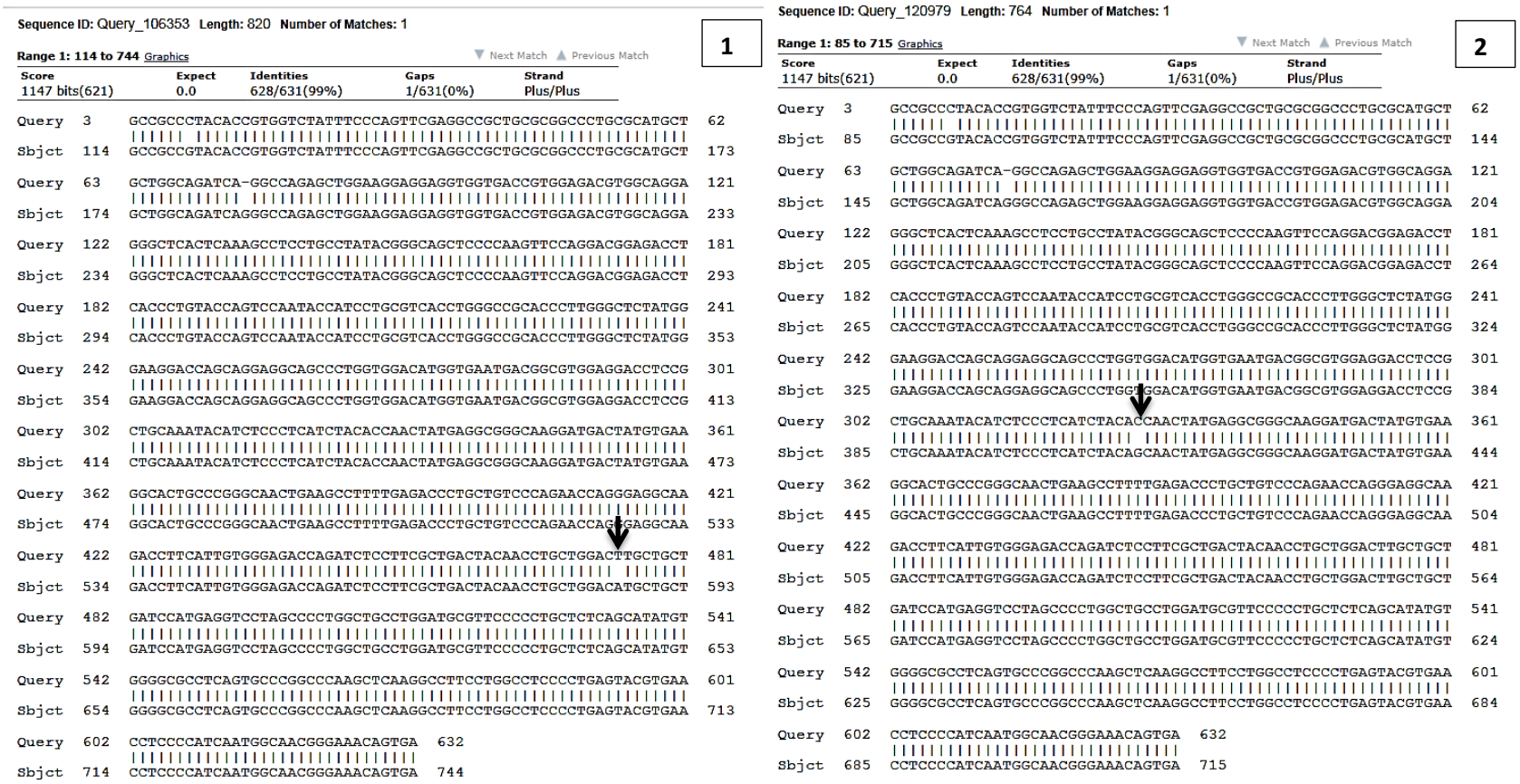


Figure 3.1A Pairwise sequence alignment of hGSTP1-1 wild-type nucleotides with L159M and T110S SNP forms. Alignment of the L159M (1) and T110S (2) with the hGSTP1-1 wild-type yielded 99 % identity. The results show that the incorporation of the mutations into the hGSTP1-1 wild-type DNA sequence was successful (shown in arrows).

The DNA sequences of the hGSTP1-1 SNP forms were translated into protein sequences and aligned with the wild-type. The aligned protein sequences gave identities of 95 % for the L159M SNP form and 98 % for the T110S SNP form. The results were run using the Protein Basic local alignment search tool (BLAST) (NCBI). The obtained results are shown below in Figure 3.1 B and C.

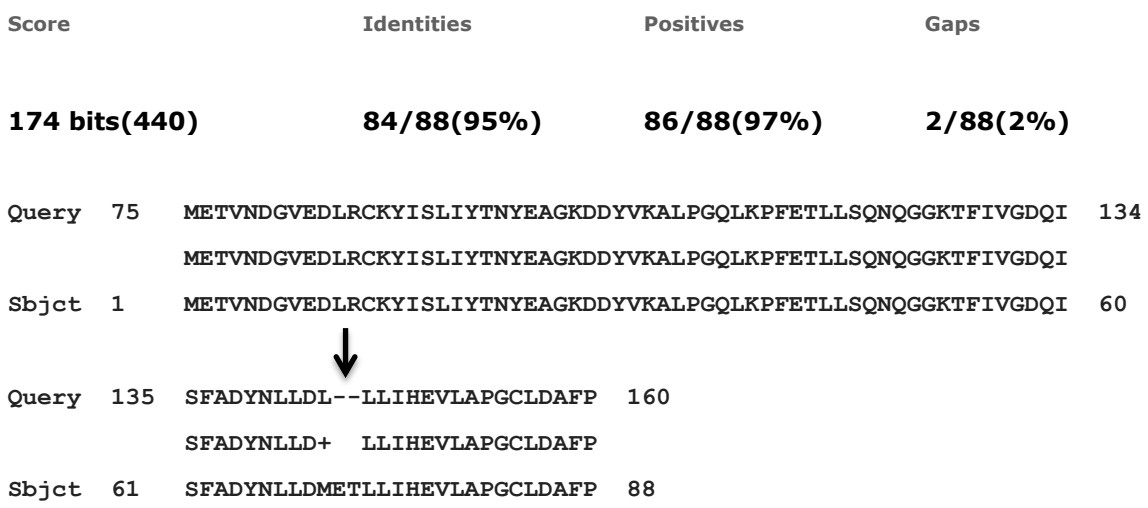


Figure 3.1 B Protein sequence alignment of the wild-type and the L159M SNP. The arrow shows the mutation of the Leucine (L) to a Methionine (Met). The methionine is written in a three characters while all other proteins are written in one character symbols.

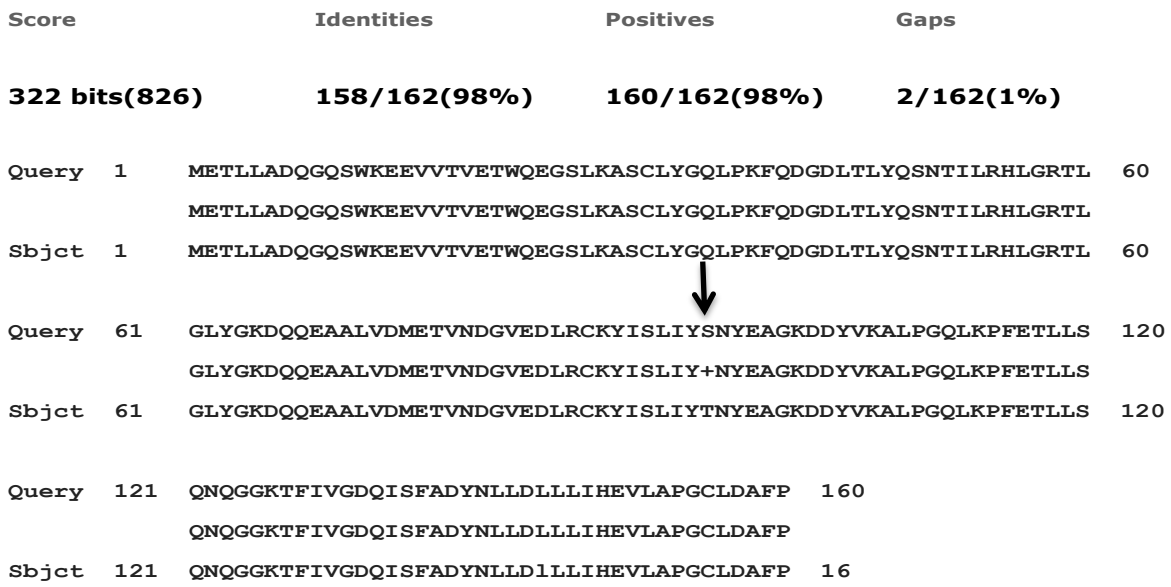


Figure 3.1C Protein sequence alignment of the wild-type and the T110S SNP. The arrow shows the mutation of a protein from a threonine to a serine.

3.3 Expression and purification of hGSTP1-1 wild-type and SNP forms

The proteins were purified using the immobilized metal ion chromatography (IMAC), a technique that uses the binding of his-tagged proteins to the positively charged chelating agents. Nickel was used as a chelating agent and imidazole was used to elute the protein of interest from the column. The elution profiles of hGSTP1-1 wild-type, L159M and T110S SNP form UV scans were all observed at A280 nm (blue scans) (Figure 3.2 A,B and C). The purification conditions used for wild-type in Figure 3.2A were also used for the two SNP forms in Figure 3.2 B and C. The fractions in peaks shown in red circles were collected and stained before visualising on SDS-PAGE (Figure 3.3, 3.4 and 3.5) to confirm if the proteins collected were the proteins of interest.

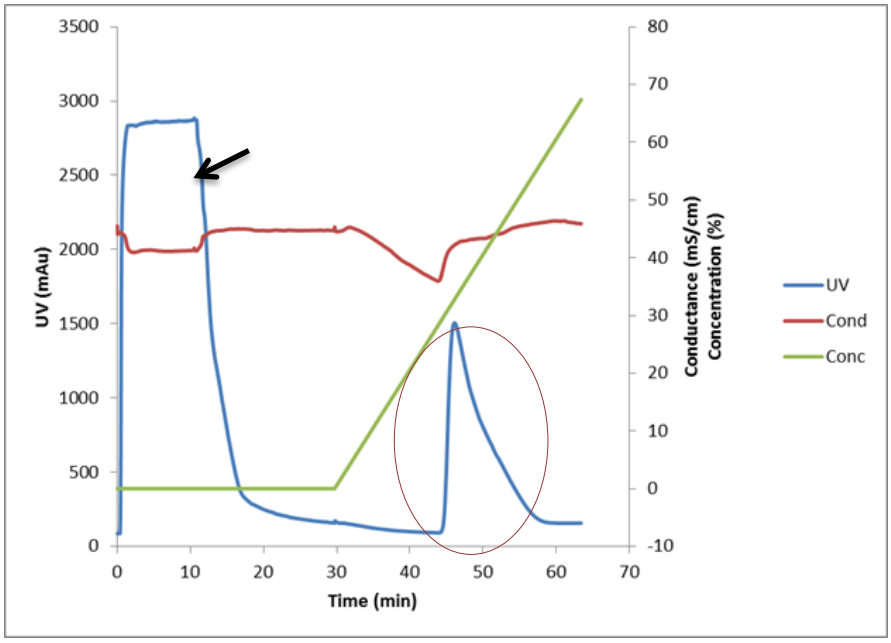


Figure 3.2A Immobilised metal-ion chromatography (IMAC) elution profiles of hGSTP1-1 wild-type. The elution profile of hGSTP1-1 wild-type UV scan was observed at A280 nm (blue scan). The arrow shows a broad peak containing a cocktail of proteins loosely bound to the column while the second peak contains the protein eluted as the concentrations of imidazole increases.

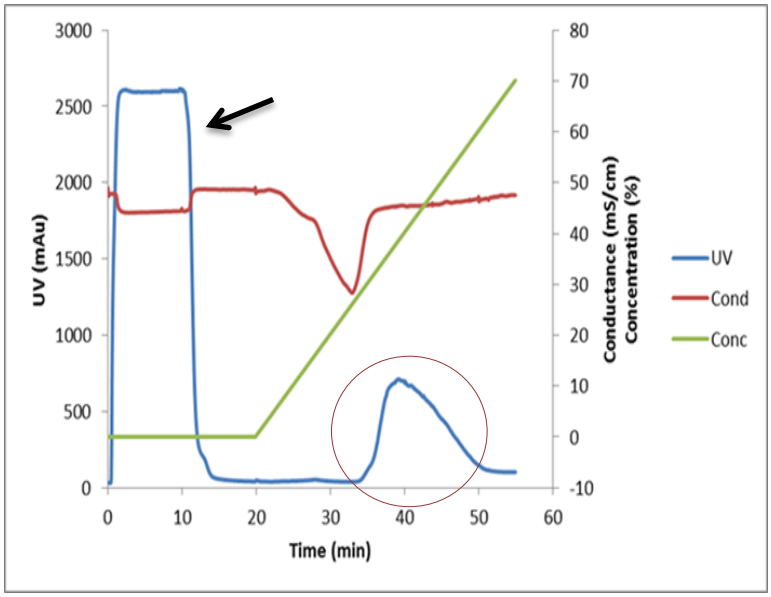


Figure 3.2B Immobilised metal-ion chromatography (IMAC) elution profiles of hGSTP1-1 L159M SNP form. The first broad UV peak (blue scan) shown by an arrow contains protein that have not bound or loosely bound to the column which is washed off with lower concentrations of imidazole (green scan) and the peaks shown in red circles represent the target protein eluting at higher concentrations of imidazole.

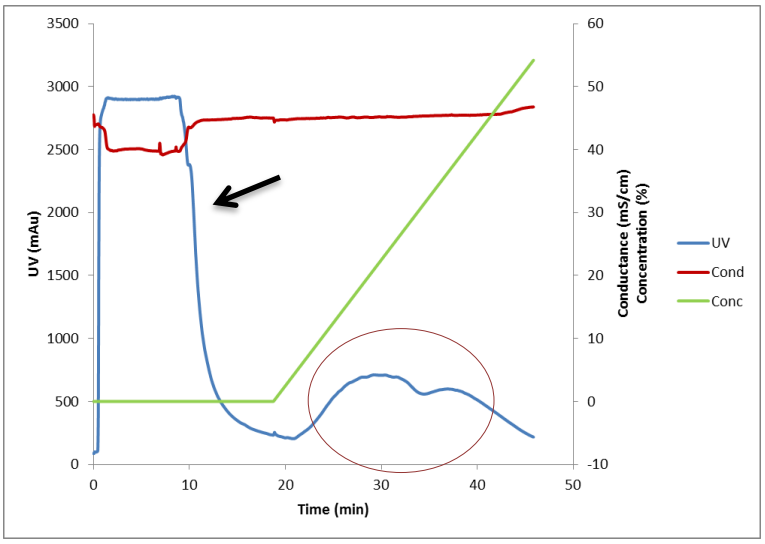


Figure 3.2C Immobilised metal-ion chromatography (IMAC) elution profile of hGSTP1-1 T110S SNP form. The protein was eluted from the nickel column using an imidazole gradient and the collected fractions were analysed using SDS-PAGE.

3.4 SDS-PAGE

The elution profiles in Figure 3.2 showed presence of proteins which were then analysed further by staining with coomasie blue and visualising on SDS-PAGE to confirm purity. The SDS-PAGE gel electrophoresis was run with 4 % stacking gel and 16 % separating gel and the protein samples were run under reducing conditions. The cells were lysed by sonication before purification, however the method used was not very efficient, as there was still a lot of protein left in the pellet as shown in the gels below (Figure 3.2, 3.3 and 3.4). The protein fractions were run individually and combined after visualisation (Figure 3.3, 3.5). The combined fractions (Figure 3.3 B) showed a presence of low concentrations of contamination twice the size of the hGSTP1-1 wild-type monomer.

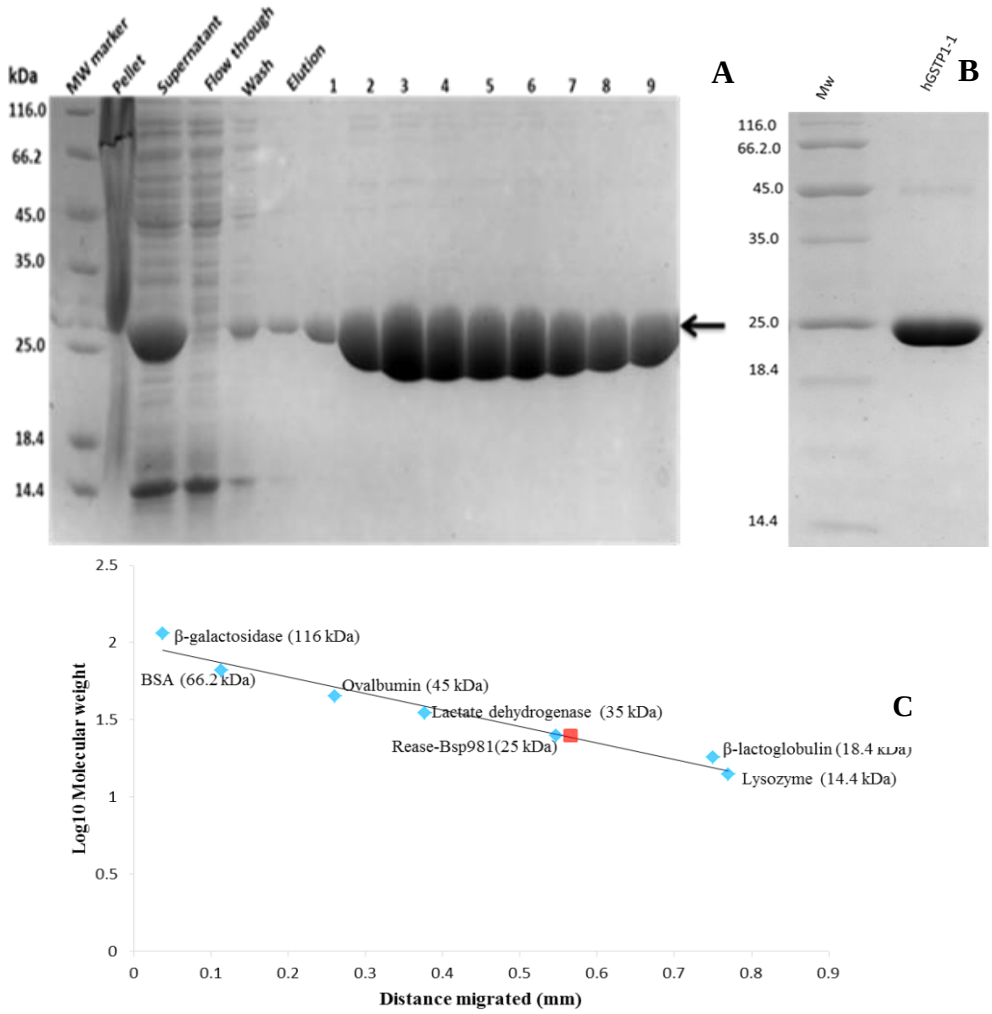


Figure 3.3 A) SDS-PAGE purification gel for hGSTP1-1 wild-type. The first lane shows the Molecular mass marker from Thermo Fisher Scientific, then the Pellet, Supernatant, the Flow through and the Wash. Elution contains the protein of interest and the Lanes labelled 1-9 show the fractions collected during elution. B) Combined protein fractions obtained from A. The arrow in Figure 3.3A

shows the position of the hGSTP1-1 monomer determined from the graph to be 24.32 kDa, shown with a red dot (C).

.

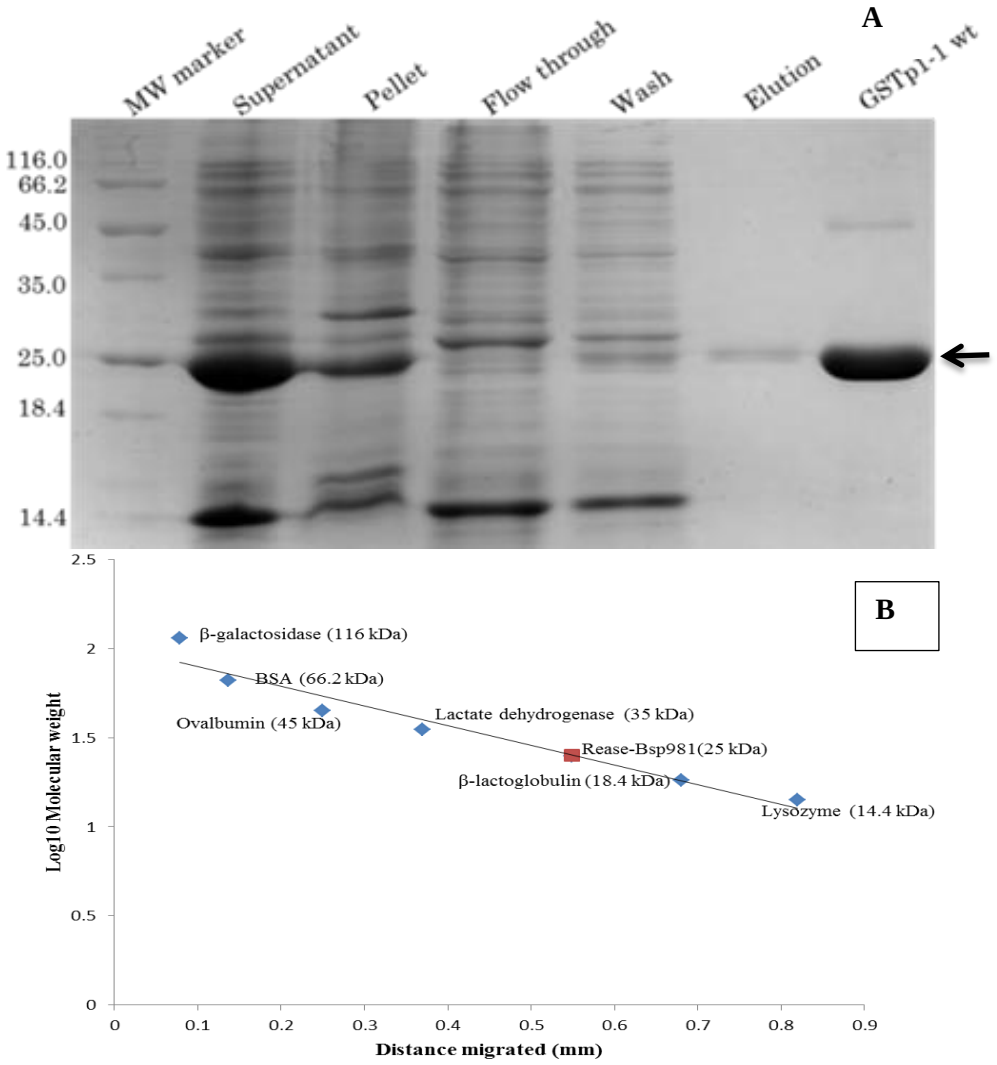


Figure 3.4 A) SDS-PAGE purification gel for the hGSTP1-1 L159M SNP form. The first lane shows the Molecular mass marker, then the Supernatant, and the Pellet as labelled above. The lanes following the pellet contains the Flow through, the Wash and the Elution, with the protein of interest containing the monomer determined from the graph (B) to be 25.11 kDa.

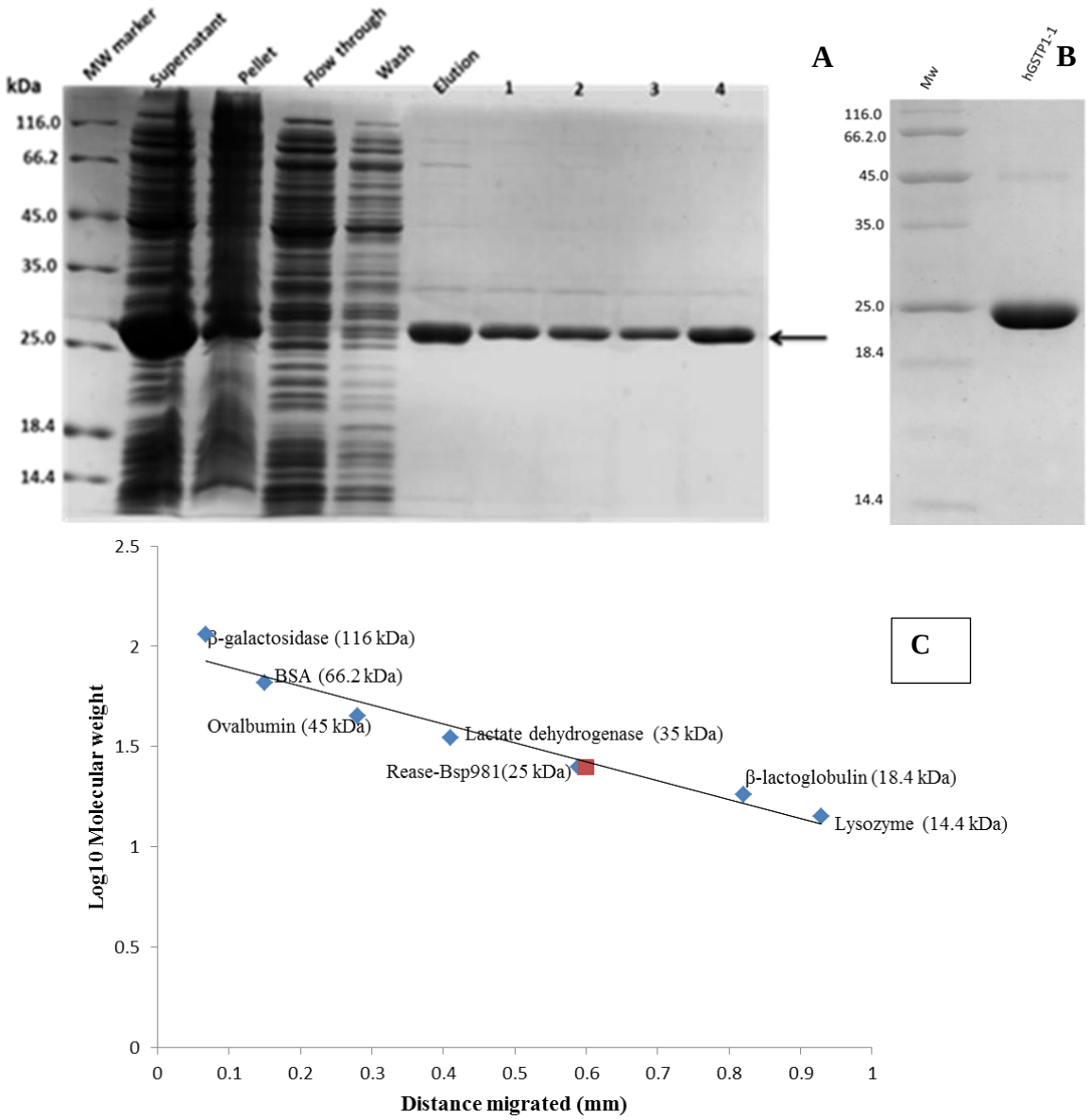


Figure 3.5 A) SDS-PAGE purification gel for hGSTP1-1 T110S SNP form. The first lane shows the Molecular mass marker, then the Supernatant, Pellet, Flow through and the Wash. Elution contains the protein of interest and the Lanes labelled 1-4 show the fractions collected during elution. B) Shows hGSTP1-1 after combining the fractions. The arrow in (A) shows the position of the hGSTP1-1 monomer which is 25.7 kDa in size, shown with a red dot in (C).

3.5 Protein concentration determination

The concentration of the purified protein was determined using a UV/VIS spectrophotometer. The proteins were diluted to obtain the same protein concentrations for all the proteins. Proteins display a characteristic UV absorption spectrum at approximately 280 nm resulting

from the aromatic amino acid residues such as tyrosine and tryptophan. With the use of the known molar extinction co-efficient of the hGSTP1-1, the Beer-Lambert law was use to quantitate the amount of the protein by UV absorbance, assuming, from the results obtain from the SDS-PAGE that the protein is pure and that there are no UV absorbing non-protein components such as bound nucleotide co-factors, heme or iron- sulfur centres. The peak at 280 nm suggests the presence of the proteins (Figure 3.6 A) and the serial- dilutions were also measured at 280 nm to confirm the concentrations using a different method (Figure. 3.6 B). The wild-type, L159M and T110S SNP forms were diluted in such a way that their concentrations are the same so that any differences in concentration may be neglected.

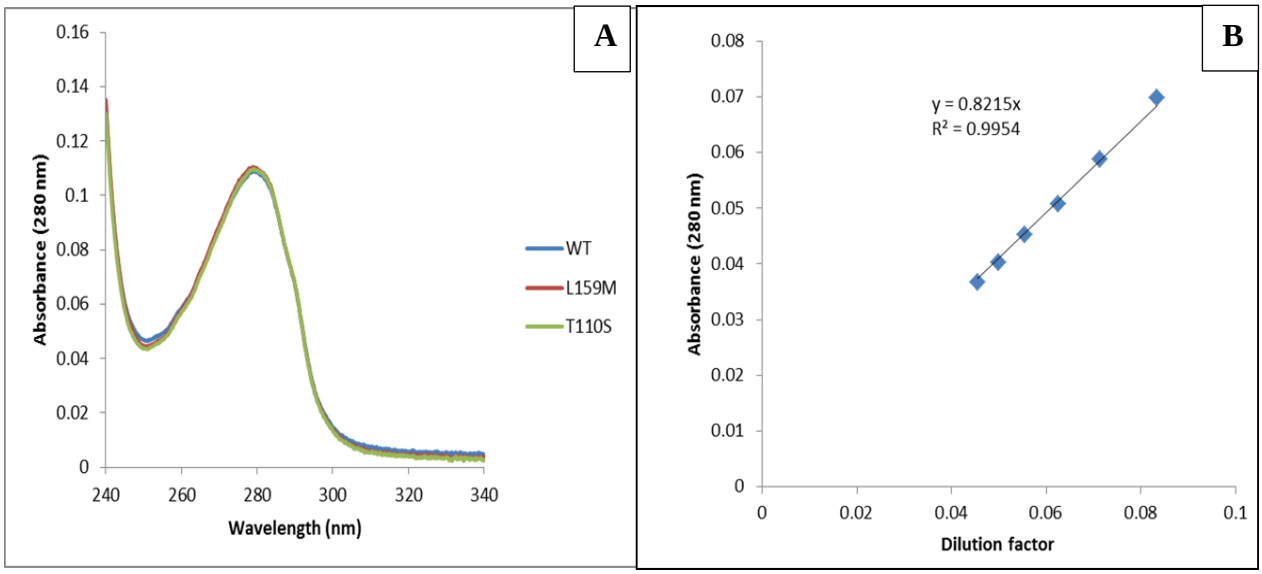


Figure 3.6 A) Absorbance spectrum of the hGSTP1-1 wild-type, L159M and T110S. B) Absorbances of serial dilutions of hGSTP1-1 wild-type measured at 280 nm. The absorbance scans of wild-type (blue), L159M (red) and T110S (green) SNP forms were of same concentrations as observed at 280 nm. The experimental data was collected in 5 mM sodium phosphate, pH 7.4.

The proteins were then used for functional studies and structural studies.

3.6 Structural characterisation

3.6.1 Secondary structure characterisation using far-UV circular dichroism

The secondary structure of the purified hGSTP1-1 wild-type protein and SNP forms was characterized using circular dichroism spectroscopy. Far-UV circular dichroism (CD)

measurements showed that the secondary structure of the two SNP forms were similar to that of wild-type GSTP1-1 measured under identical conditions indicating no structural effects on the wild-type hGSTP1-1 protein. The protein remains predominantly alpha-helical even in the presence of the SNPs. The experimental data was measured using 2 µM concentrations of the protein in 5 mM sodium phosphate, pH 7.4.The results obtained are shown in Figure 3.7 below.

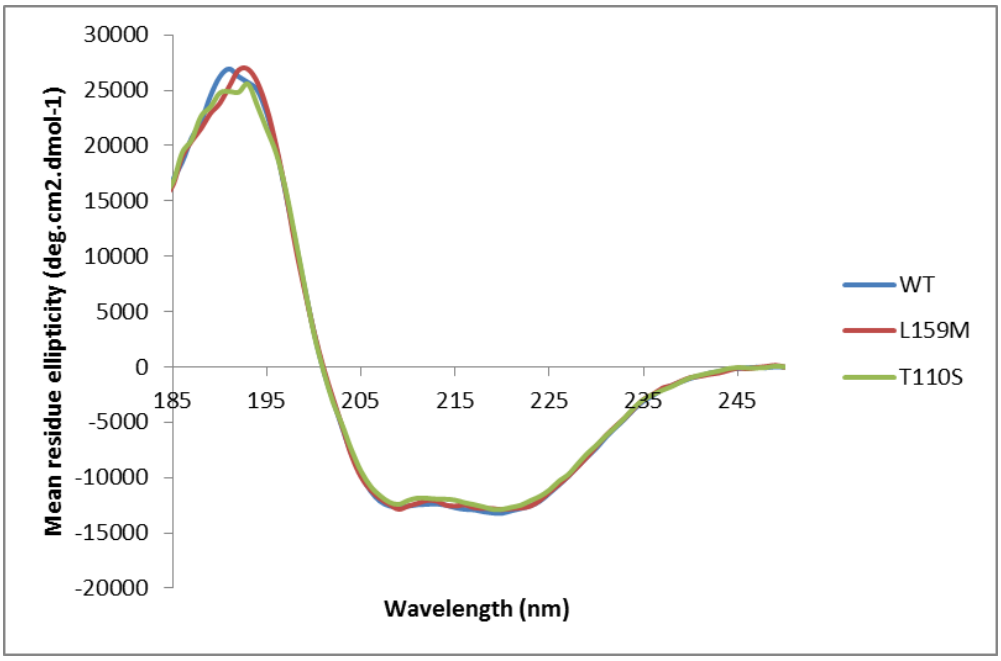


Figure 3.7 Far-UV CD spectra for the hGSTP1-1 wild-type, L159M and T110S SNP forms. Circular dichroism for the hGSTP1-1 wild-type (blue), the L159M (red) and T110S (green) SNP forms showed comparable spectra. They all show troughs at 210 and 220 nm wavelength.

3.6.2 Tertiary structure characterisation using fluorescence spectroscopy

The tertiary structure of the hGSTP1-1 wild-type and the two SNP forms was characterised by measuring the fluorescence of the tryptophan residues at 295 nm. Tryptophans were used as a fluorophore for the measurement of fluorescence of the wild-type and SNP forms. This is because it has a higher quantum yield as compared to other residues used. The intensity in fluorescence of the SNP forms as compared to wild-type will indicate a change in the tertiary structure of the protein. The L159M SNP form showed a slightly significant increase in fluorescence as compared to the T110S, which showed very low increase in intensity (Figure

3.8). The fluorescence spectra results suggest that the two SNPs have an effect on the tertiary structure of the hGSTP1-1 protein.

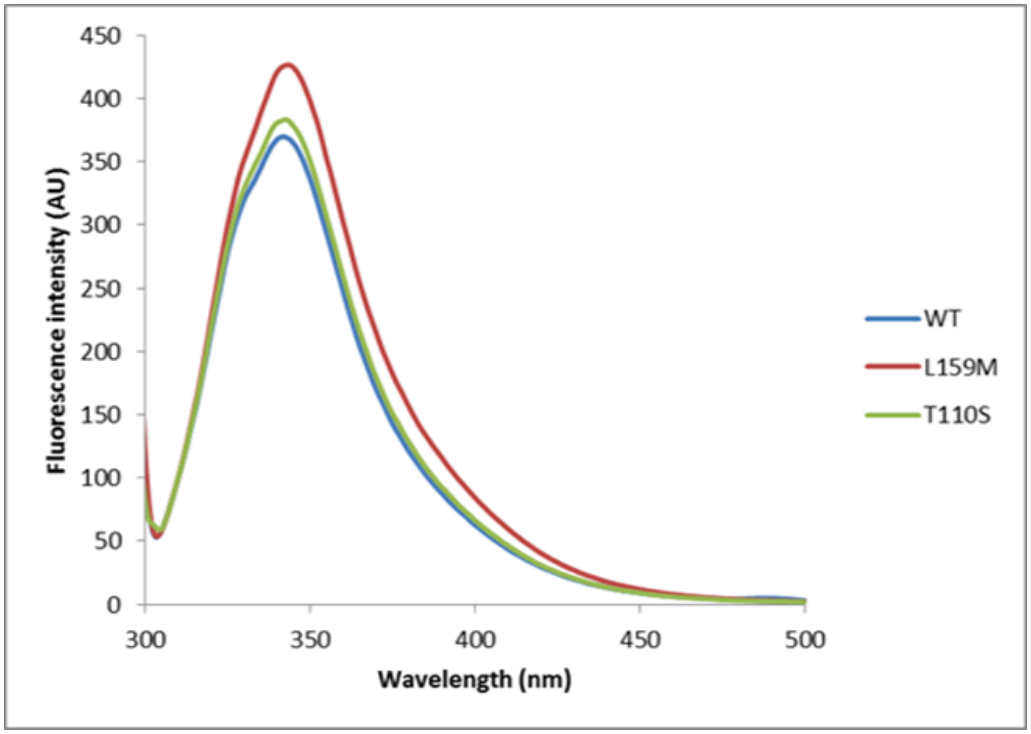


Figure 3.8 Fluorescence spectra of the hGSTP1-1 wild-type, L159M and T110S SNP forms. The wild-type (blue), L159M (red) and T110S (green) SNP forms all display emission at 350 nm. The L159M SNP form displayed higher fluorescence intensity than the wild-type and the T110S SNP form.

3.6.3 Size exclusion chromatography: Quaternary structure analysis

Size exclusion chromatography was used to investigate the oligomeric state of wild-type GSTP1-1, L159M and T110S SNP forms. This technique separates proteins based on their molecular size and the separation is achieved as the molecules pass through a gel filtration medium packed in a column. Unlike other methods like ion exchange or affinity chromatography, molecules do not bind the column so the composition of the buffer does not directly affect the degree of separation between peaks and is dependent on the size of the molecules and the pores of the matrix in the column. Larger molecules are eluted earlier than smaller molecules. The elution profiles of the proteins displayed no difference in retention times suggesting no change in size. Elution profile of wild-type hGSTP1-1 and two SNP forms were eluted from the Hi Load™ 16/600 Superdex™ 75 pg column with 20 mM sodium

phosphate, pH 7.4, containing 1 mM EDTA and 0.02 % sodium azide. The elution profile of the wild-type and the two SNP forms: L159M and T110S are shown in Figure 3.9 below.

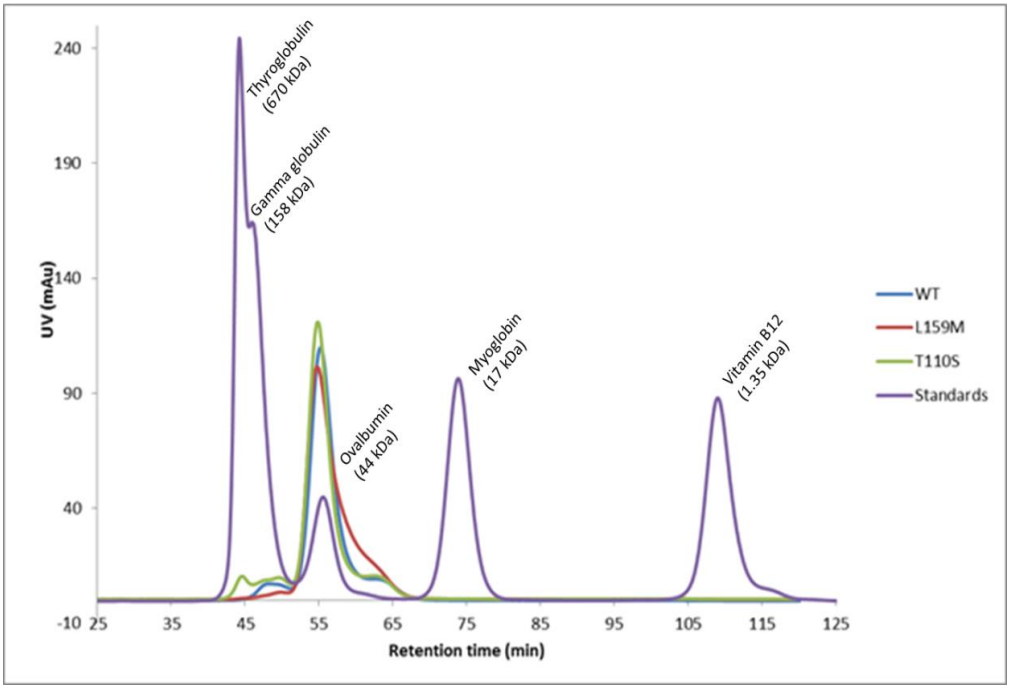


Figure 3.9 Elution profiles of wild-type hGSTP1-1, L159M and T110S SNP forms. Wild-type profile (blue) shows a peak at 55 minutes, with the L159M SNP (red) and T110S SNP forms (green) also eluting at 55 minutes. The wild-type and the two SNP forms showed shoulders on both sides of the peaks and these peaks were analysed using SDS-PAGE (Figure 3.10). The wild-type and the ovalbumin peaks also appear to be marginally to the right of the L159M and the T110S SNP forms.

To investigate what the shoulders revealed on SEC are, elution samples collected from the SEC (Figure 3.10A) were run on SDS-PAGE (B), each peak collected separately.

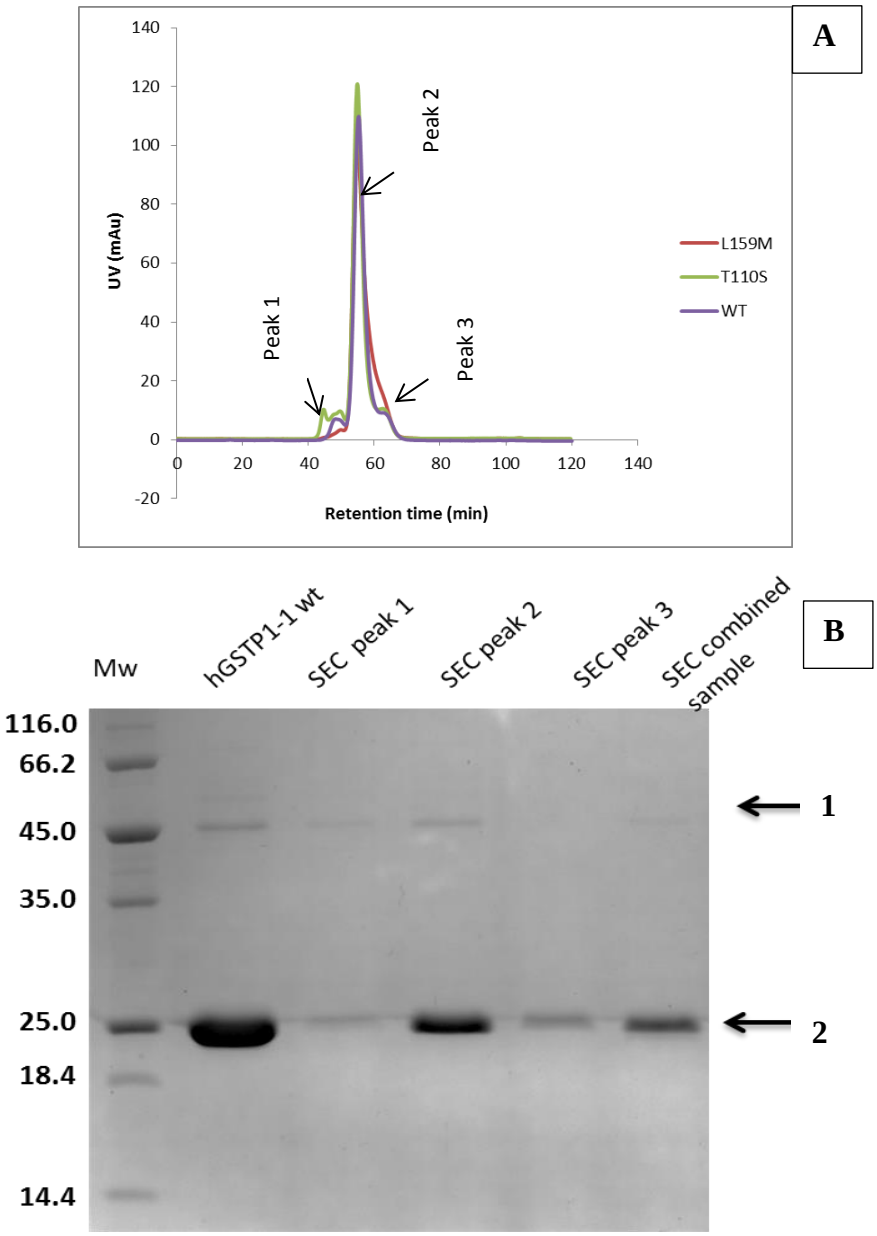


Figure 3.10 A) SDS-PAGE purification gel showing hGSTP1-1 wild-type, L159M and T110S after size exclusion chromatography. **B) Analysis of wild-type, L159M and T110S SNP forms shoulder peaks using SDS-PAGE.** The hGSTP1-1 samples (peak 1, 2 and 3) were collected from the SEC column (A) and run on SDS-PAGE (B). The hGSTP1-1 wild-type shows two bands, a very faint band of size 45.0 (1) and a visible band of size 25.0 kDa (2) which corresponded to the sizes of the wild-type protein monomer and the band at 45 kDa which represents a GST-like contaminant.

3.7 Thermal stability studies

Analysis of CD spectra of proteins as a function of temperature was measured using the circular dichroism to determine the thermodynamics of hGSTP1-1, L159M and T110S protein folding. The hGSTP1-1 wild-type and the two SNP forms do not refold and therefore the thermodynamics could not be determined. The CD spectra of the wild-type and the two SNP forms were measured at 222 nm with temperatures varying from 20 to 80 ºC. Experiments were performed in 5 mM sodium phosphate buffer at pH 7.4. Concentrations of protein used were 2 µM for all proteins.

The unfolding scan of the proteins is shown in Figure 3.11 below. The proteins all show melting at 55 ºC, this result suggests that the SNPs have no influence on the stability of the hGSTP1-1.

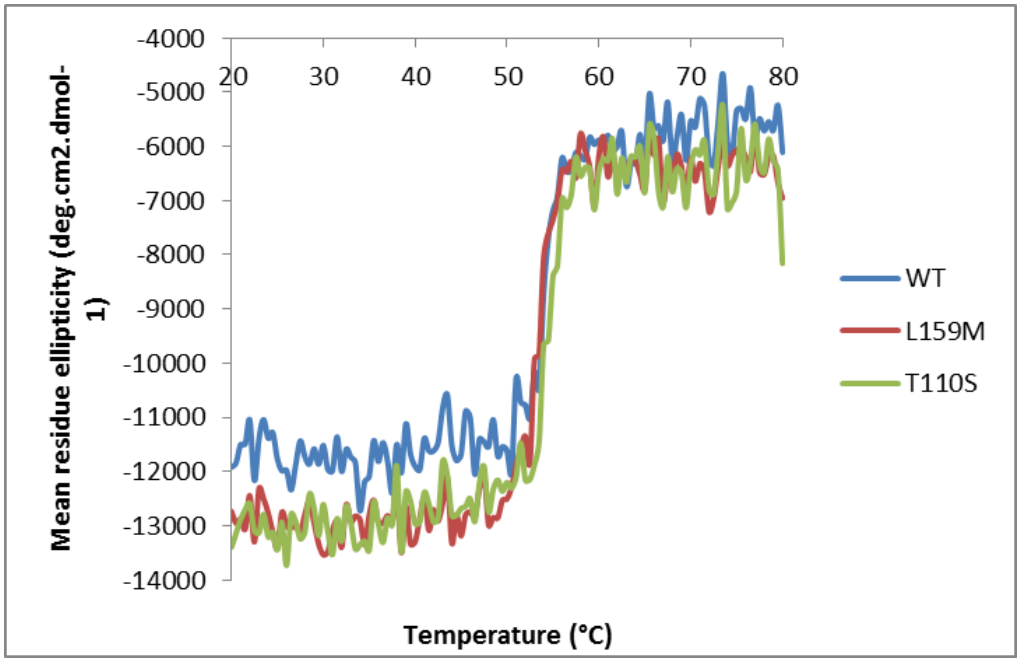


Figure 3.11 Far-UV CD spectra of the hGSTP1-1 wild-type, L159M, T110S SNP forms as a function of temperature at a single wavelength. The wild-type (blue), L159M (red) and T110S (green) all showed unfolding at 55 ºC. The stability of the wild-type was not affected by the presence of the L159M and T110S SNPs.

3.8 Functional characterisation of the hGSTP1-1 wild-type and SNP forms

3.8.1 GST activity assay for hGSTP1-1wild-type and SNP forms

The catalytic activity of hGSTP1-1 was analysed using the CDNB (1-chloro-2,4-dinitrobenzene) assay as described by (Habig *et al.*, 1974). The assay is based upon the GST

catalysed reaction between the GSH and the substrate, CDNB. Cytosolic GSTs catalyse the nucleophilic aromatic substitution reaction of CDNB and GSH resulting in the formation of 1-(s-glutathionyl)-2,4-dintrobenzene (also known as S-2,4 dinitrophenyl glutathione) which absorbs light maximally at 340 nm (Habig and Jakoby, 1981). The solid lines on the graphs represent a linear fit to the experimental data. The specific activity for each enzyme is determined from the slope of the fit with the correlation coefficient of 0.98 for wild-type, 0.97 and 0.98 for the L159M and T110S SNP forms, respectively. The experimental data was collected with varying amounts of protein for the wild-type and the two SNP forms in 100 mM sodium phosphate buffer containing 1 mM EDTA and 0.02 % sodium azide at pH 6.5.

The results obtained showed a decrease in specific activity for T110S SNP form and for L159M SNP form (Figure 3.12B and C) when compared to the wild-type (Figure 3.12 A).

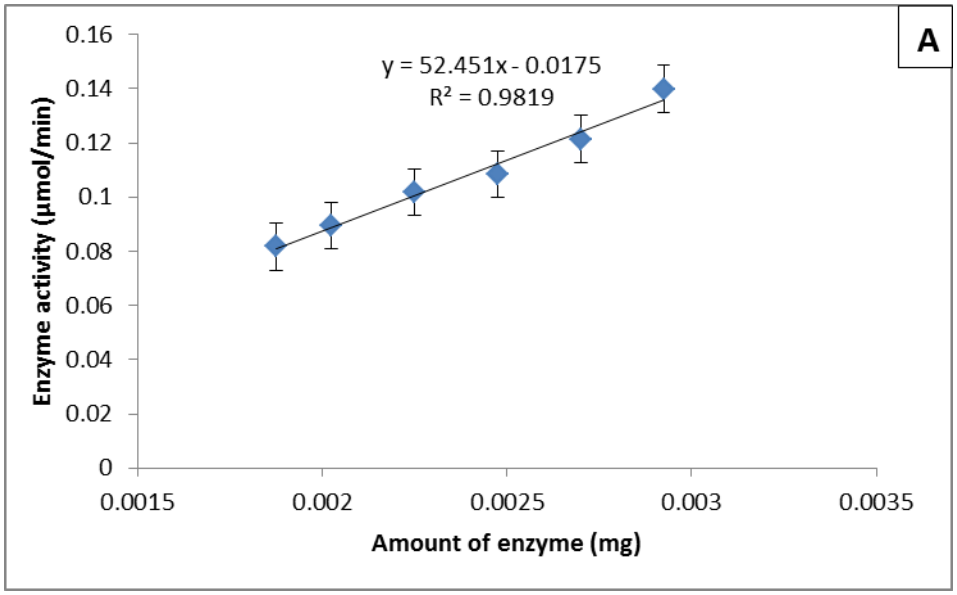


Figure 3.12A Specific activity of hGSTP1-1 wild-type. The wild-type shows a specific activity of 52.4 µmol/min/mg.

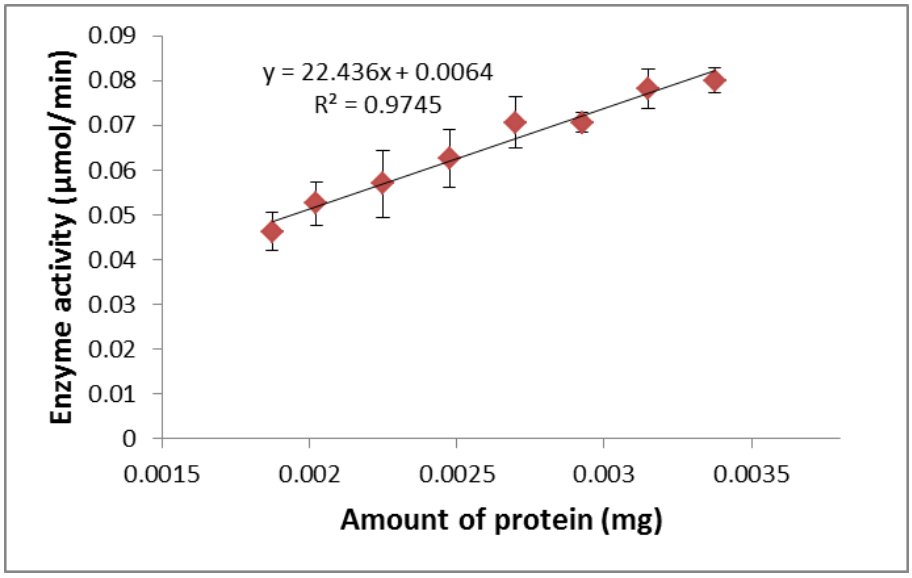


Figure 3.12B Specific activity of hGSTP1-1 L159M SNP forms. The presence of the L159M SNP in the hGSTP1-1 protein caused a decrease in its activity leading to 22.43 µmol/min/mg. The L159M SNP influences the activity of the hGSTP1-1 wild-type protein

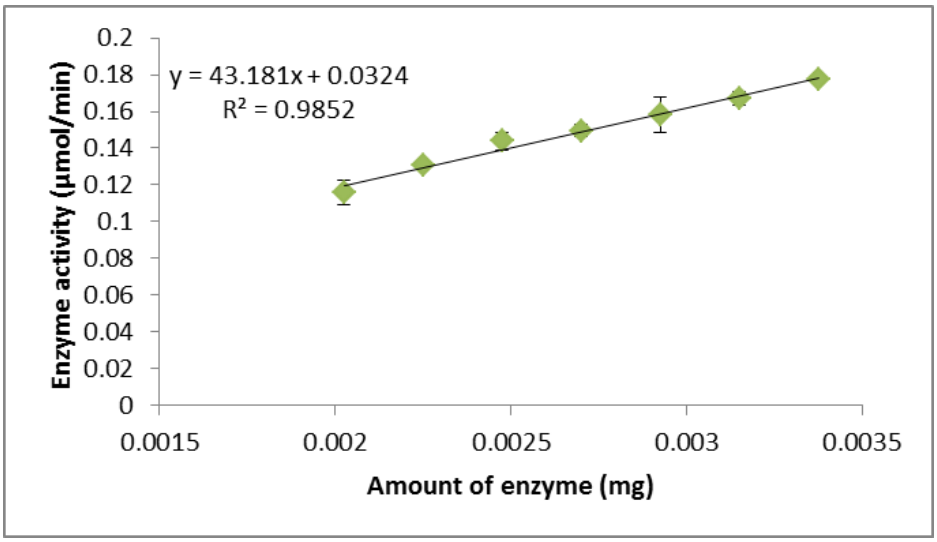


Figure 3.12C Specific activity of hGSTP1-1 T110S SNP forms. The T110S SNP form was determined to have a specific activity of 43.2 µmol/min/mg which is lower than that of the wild-type.

The percentage activities of the SNP forms were calculated and compared to that of the wild-type in Table 3.1.

Table 3.1 Percentage activities of hGSTP1-1 wild-type, L159M and T110S SNP forms

Enzyme	Specific activity (µmol/min/mg)	Percentage activity (%)
hGSTP1-1 wild-type	52.4 ± 2.73	100
hGSTP1-1 L159M	22.4 ± 0.79	42.8
hGSTP1-1 T110S	43.2 ± 0.42	82.3

The specific activity percentages of the wild-type as compared to those of SNP forms differ by 18 % and 57 % for the T110S and L159M SNP forms, respectively. The T110S does not show a significant difference in activity as compared to the wild-type while the L159M SNP form shows over 50 % difference.

The activity of hGSTP1-1 is affected by the presence of SNPs and the binding of ligands to the wild-type was investigated to observe how these SNPs affect the structure of the protein and further affect ligand binding which may be the reason for the decrease in activity. The results obtained from the structural studies, which are the change in tertiary structure presented by the increase in fluorescence of the L159M SNP form also indicate an effect caused by the SNP.

3.8.2 ANS binding studies

The L159M and T110S SNP forms have an effect on the activity of the hGSTP1-1. This led to the investigation of whether these SNPs affect ligand binding and the shape or structure of the wild-type. ANS was used in this study as a probe to measure the overall surface hydrophobicity of the SNP forms relative to the wild-type. The fluorescence properties of ANS are fundamentally influenced by the polarity of its surrounding. A shift to a lower wavelength of the fluorescence emission maxima is observed when ANS is bound to a hydrophobic surface of the protein; this bathochromic shift is accompanied by an increase in fluorescence intensity (Slavik, 1982). Experimental data was obtained by incubating 2 µM GSTP1-1 wild-type and SNP forms: L159M and T110S separately in 20 mM sodium phosphate buffer, pH 7.4 with ANS at a final concentration of 200 µM. ANS was then selectively excited at 390 nm, and

emission spectra were measured from 400 to 700 nm. Fluorescence emission spectra of ANS bound to hydrophobic regions of hGSTP1-1 wild-type, L159M and T110S SNP forms are shown in Figure 3.13.

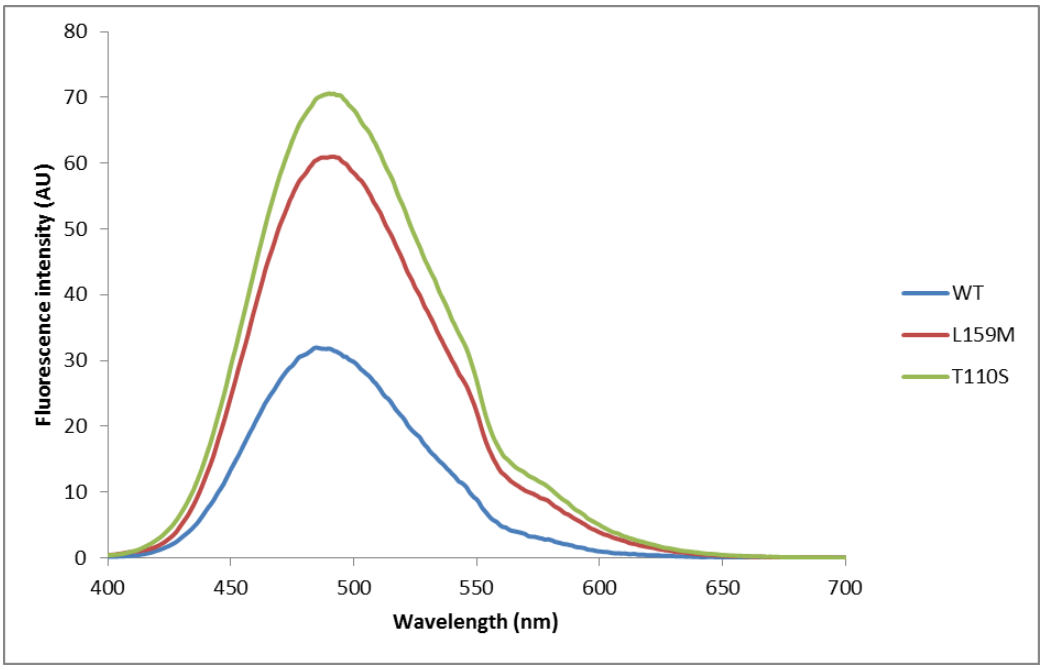


Figure 3.13 Fluorescence emission spectra of ANS bound to GSTP1-1 wild-type, L159M and T110S SNP forms. The hGSTP1-1 wild-type, L159M and T110S had maximum fluorescence intensities of 32, 61 and 70 AU, respectively and shifts in wavelength at 486 nm for the wild-type (blue), 489 nm for the L159M SNP form (red) and 490 nm for the T110S SNP form (green). The fluorescence intensities of the L159M and T110S SNP forms were higher as compared to the wild-type.

3.8.3 Dimer docking studies of ANS to hGSTP1-1

The regions of protein where the ANS binds may affect the intensity in fluorescence. To investigate the possible binding site of the hGSTP1-1, the ANS was allowed to bind to the possible regions in the protein using induced fit docking (flexible docking), and the region of the protein with the highest binding affinity for ANS was the dimer interface. The flexible docking methodology was employed using Glide software (Shrodinger LLC 2009 USA). ANS was allowed to bind freely to the crystal structure of hGSTP1-1. The two protein subunits bind together and form the hGSTP1-1 dimeric complex. The ANS binds to the dimer interface in between the subunits, which is the region of the protein with the highest binding affinity for ANS as obtained from the studies. The structure of the hGSTP1-1 protein was retrieved from the PDB 5J41 along with the 3D structure of the ANS, the results in Figure 3.14 were obtained.

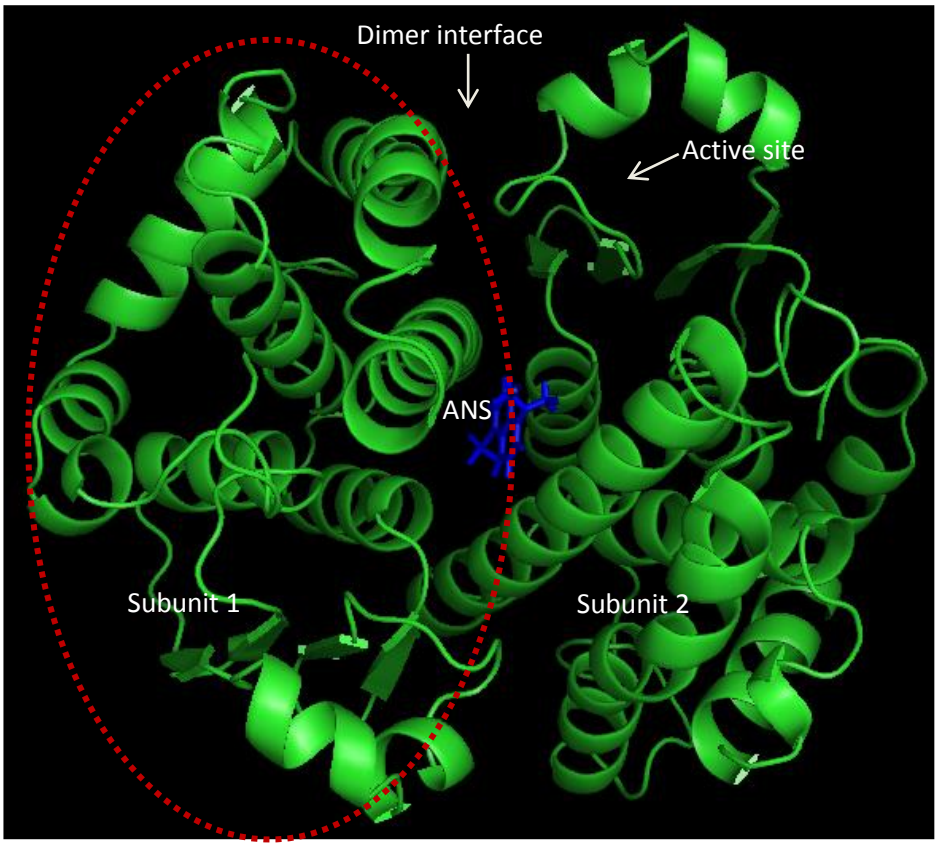


Figure 3.14 Diagram illustrating the binding region of ANS in hGSTP1-1. The ANS binds in the dimer interface of the hGSTP1-1.

The ANS did not bind to the active site as expected during computational studies, and was shown to bind the dimer interface as shown above. This finding was further proved by performing ANS displacement studies in Figure 3.15 below. The image was generated from PyMol using the PDB 5J41coordinates.

3.8.4 ANS Displacement studies

Computational studies showed that ANS binds to the dimer interface of the protein and thus a displacement approach for competitive binding studies was performed, this was investigated by performing ANS displacement studies. This method utilizes the potentiometric 1-anilino-8-naphthalene-sulfonate (ANS). The fluorometer was used to monitor the displaced free ANS probe from the binding site on the protein by stepwise addition of the ethacrynic acid. Free ANS has low fluorescence intensity and an increase is observed when it is bound to the hydrophobic regions of hGSTP1-1. When ANS is displaced from the active site of the protein a decrease in fluorescence intensity is observed. Fluorescence measurements were performed at 20 ºC and the ANS was exciting at 390 nm. Fluorescence emission was observed at 485 nm.

As a control, absolute ethanol was titrated in a 1ml solution containing 100 µM ANS in a quartz cuvette and 2 µM hGSTP1-1 protein in 5mM sodium phosphate buffer, pH 6.5. Titrations were done from 0 to 6 % ethanol percentage without the presence of ethacrynic acid (Figure 3.15A). The test run contained 100 µM ANS and titrations of ethacrynic acid dissolved in ethanol to up to 8 mM were done with 2 µM protein (Figure 3.15B). An increase in ethanol percentage showed no change and therefore the test run was performed with ethacrynic acid dissolved in ethanol (Figure 3.15B).The results obtained are shown in Figure 3.15 A and B below.

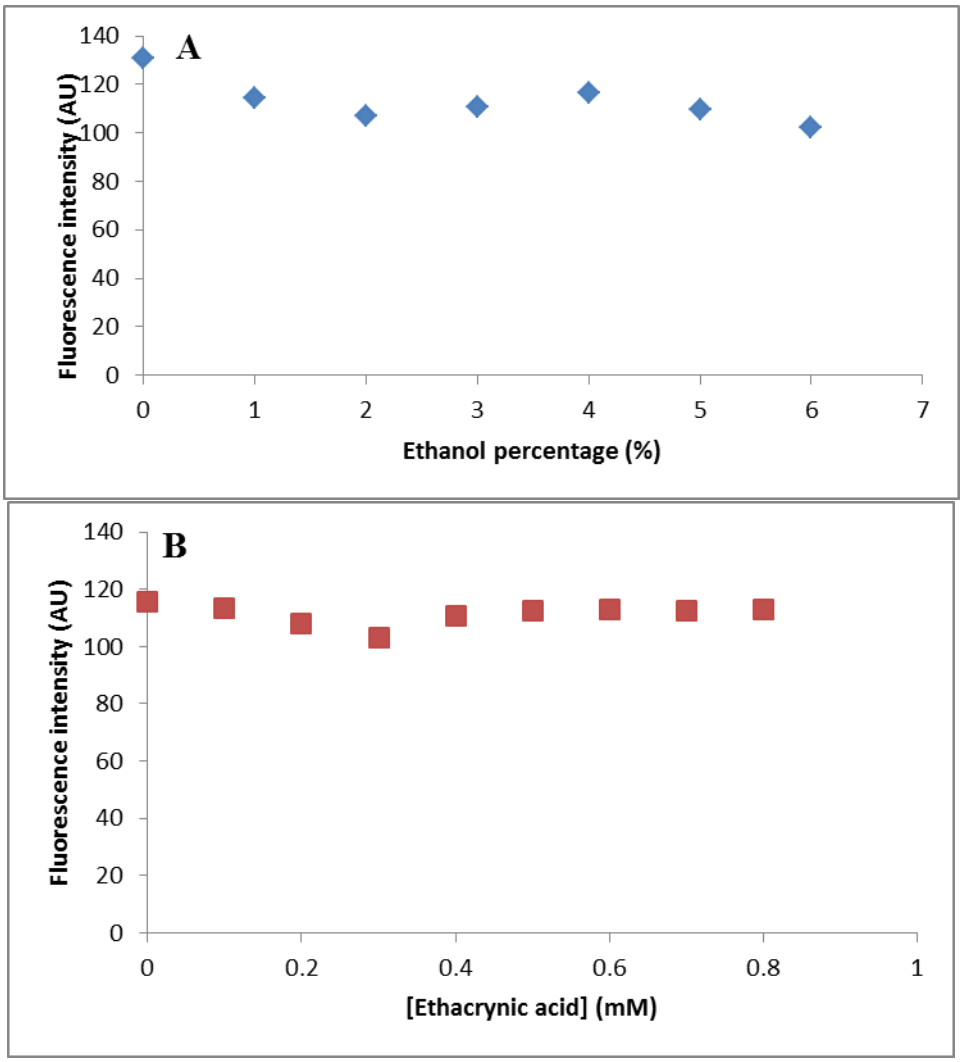


Figure 3.15 Diagram showing displacement of ANS by ethacrynic acid. A control with ethanol showed a straight line (A). There was no displacement of ethacrynic acid from the active site observed as the fluorescence of ANS remained constant (B). This suggests that the ANS did not bind the active site of the protein.

CHAPTER 4: DISCUSSION

The role of hGSTP1-1 in the detoxification of toxins in humans is very important. This is not the only function of the hGSTP1-1 in human but yet the most studied. Other functions of hGSTP1-1 include their function as ligandins and their activation of peroxiredoxin VI (Prdx6) amongst others. However, factors such as changes in temperature, enzyme or substrate concentrations, presence of inhibitors or mutations may affect either its structure or function. SNPs have been studied to better understand their effect on the hGSTP1-1 protein and therefore comprehend if they cause any diseases to this end, there have been a lot of studies done to elucidate the role of SNPs in GSTs and in human health in general. They have been implicated in a lot of diseases, most of which are cancers, such as bladder and testicular (Harries *et al*., 1997) and lung cancer (Ryberg *et al*., 1997).

4.1 Site directed mutagenesis, expression and purification

To incorporate SNPs in the hGSTP1-1 wild-type, site directed mutagenesis was performed. The experiment was successful as it gave up to 99 % sequence similarity for both L159M and T110S SNP forms and the changes in nucleotide were incorporated (Figure 3.1A). A gap which might have formed from a fault during sequencing is observed. The alignment of the protein sequences was performed and the SNPs resulting in changes in proteins from S to T and L to M were incorporated (Figure 3.1 B and C). The expression and purification were also successful (Figure 3.2 A, B and C). This result was also confirmed by running SDS-PAGE gels to analyse the purified proteins (Figures 3.3, 3.4 and 3.5). A band the size of hGSTP1-1 dimer is observed in very low concentrations less than 2 % when observed under densitometry, this cannot represent a hGSTP1-1 dimer because the proteins were run under reducing conditions and could represent a GST like contaminant as it was able to bind the column. The concentrations of the proteins were determined successfully by using absorbance spectrometry (Figure 3.6).

4.2 Influence of mutations on structure: SNPs in the hydrophobic core of hGSTP1-1 likely to affect structure

The structure of proteins is made up of building blocks which are amino acids that join together to form a secondary structure of the protein which is made up of β-sheets and α-helices which are organised to form the tertiary structure. Due to the nature of weak interactions building up the tertiary structure of proteins, proteins are very sensitive molecules and a disruption of its native state by factors such as heat or incorporation of mutations can

result in loss of structural integrity. The SNPs in hydrophobic regions play a very important role in keeping the structure of the protein stable and maintaining its overall shape and the residues making up the active site are important for its function. T110S and L159M SNPs were studied to evaluate if they have any effects on the structure and function of hGSTP1-1, findings from current study show that these SNPs do not present any effect on the secondary structure of hGSTP1-1 (Figure 3.7), as determined in this study by CD. However they show a difference in tertiary structure when compared to the wild-type, as determined by fluorescence spectroscopy, with the L159M SNP form showing the most significant difference, as demonstrated by is increase in fluorescence intensity (Figure 3.8). The positions of the SNPs are shown in Figure 4.1.

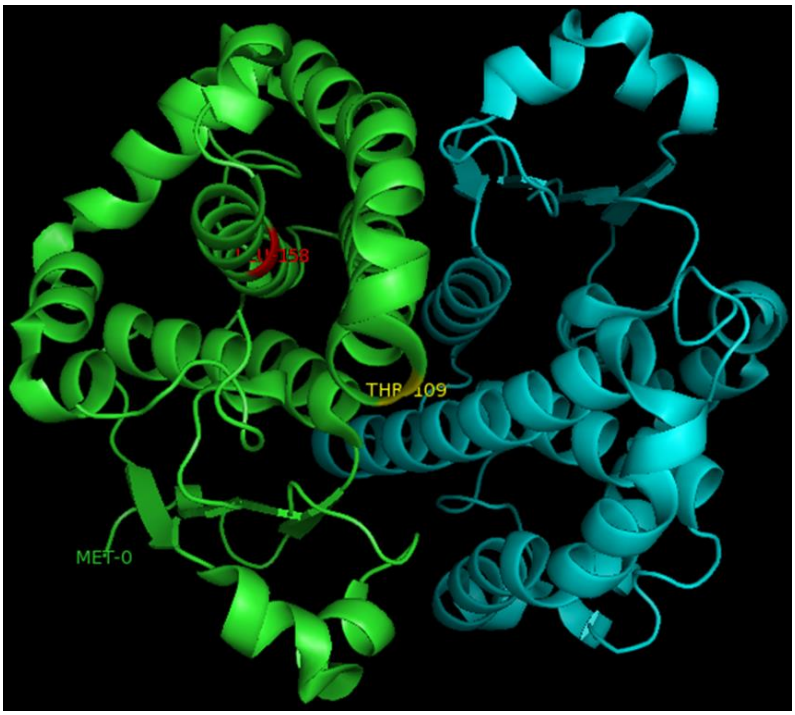


Figure 4.1 Locations of the SNPs in hGSTP1-1 hydrophobic core. The position of the T110S SNP (yellow) is at the end of the α-helix 4 of the protein while the L159M (red) is located inside the hydrophobic core of hGSTP1-1 on α-helix 6.

Tryptophans were used as fluorophores for measuring fluorescence intensities at 295 nm. The hGSTP1-1 contains two tryptophans, Trp28 which has a role in glutathione binding and Trp38 which is located at the beginning of the second β-strand and is closest to the first part of the C-terminal helix (Prade *et al.*,1997). Trp38 is also essential at the G-site of the protein as it functions to sequester glutathione. The change in amino acid sequence from leucine to methionine is likely to have an effect on the structure of the protein because not only is

methionine larger and bulkier it also contains a sulphur atom making it more reactive because of its high electronegativity, although it has lower electronegativity than oxygen, the sulphur is still likely to form disulfide bonds with other residues in close proximity. The substitution of leucine to methionine and the difference in the surface area occupied by methionine is illustrated below, with the methionine shown in white and yellow (Figure 4.2).

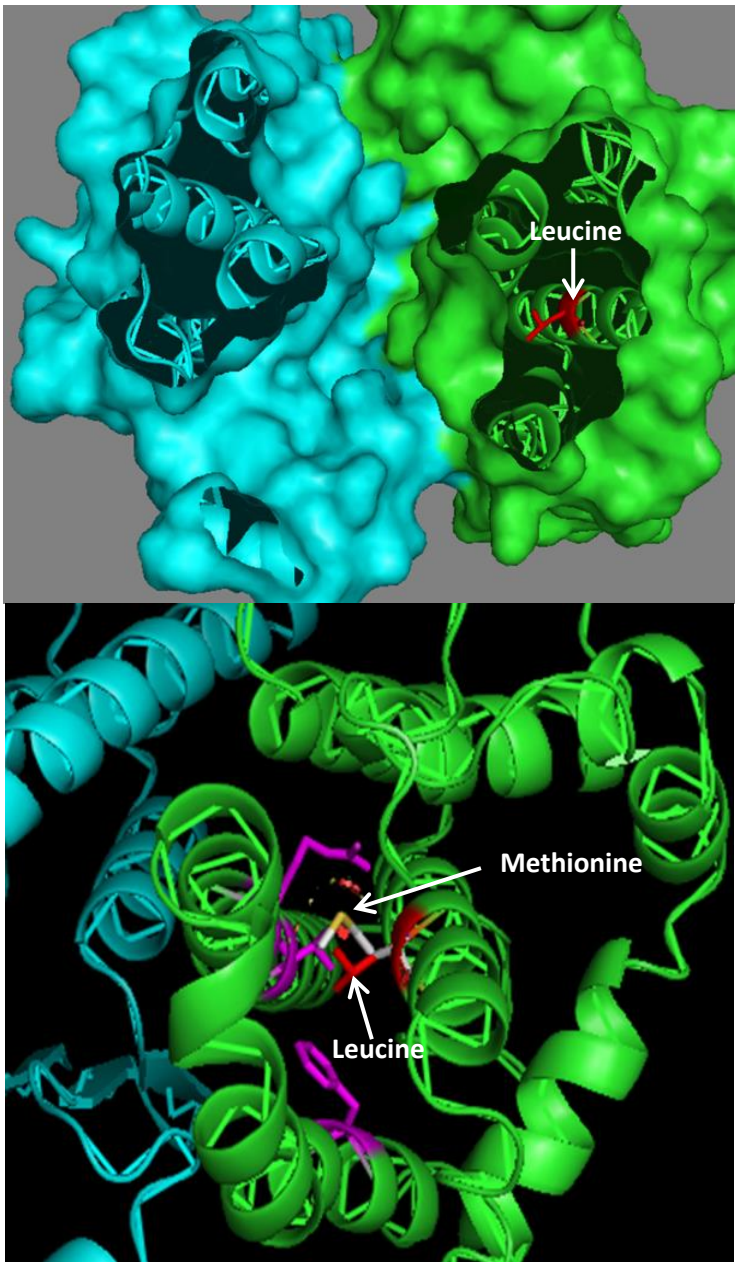


Figure 4.2 Diagram showing the position of the SNP in position 159 which is substituted by a methionine. The methionine is bulkier than the leucine spanning a larger surface area and likely to interact with more residues than leucine because of its possession of a sulphur atom which makes it more electronegative.

The location of methionine, which is the hydrophobic core allows close interaction with other residues. The possible interaction of methionine with other residues may have altered its structure and because fluorescence of tryptophan is strongly affected by the polarity of its surrounding this explains the increased intensity of the L159M SNP form (Figure 3.8). The increase in fluorescence can also be attributed to the tryptophans being hidden as a result of the change in tertiary structure increasing their quantum yield hence the increase in intensity.

The SNPs did not display any impact on the quaternary structure (Figure 3.9). The wild-type and the two SNP forms eluted at the same time suggesting that their sizes were similar and their shape have not been changed. The quaternary interaction of hGSTP1-1 contributes greatly towards maintaining the conformation of the protein. It results in the functional conformation of the active site (Sternberg *et al*., 2000, Sayed *et al*., 2000 and Hornby *et al*., 2002) and the conformational stability of the protein (Dirr and Reinemer, 1991). When the shape or size of the protein is altered this may lead to effects such as decreased stability of the protein.

4.3 Stability of hGSTP1-1

A number of studies have shown SNPs to affect stability of hGSTP1-1 but the two SNP forms T110S and L159M showed no suggestion of an altered stability after evaluation using far-UV CD (Figure 3.11). A change in the stability by the T110S SNP is not expected considering the position in which it is located at the end of α-helix 4 in the hydrophilic region. It may however be expected to have an effect on the function of the protein since it is very closely positioned next to a catalytic residue of the H-site which is the Tyr108. The L159M however, even though it is buried in the hydrophobic region of the protein was expected to have a role in the stability of the protein. The hydrophobic core of hGSTP1-1 has been found to have a role in the folding and further influencing the stability of the protein (Dirr *et al*., 2005). The studies conducted however, the findings do not suggest this to be the case. Despite studies of mutations such as F151L being done; this mutation is located next to a conserved folding module and in the same helix 6 as the L159M SNP form and has been shown to lower stability. The α-helix 6 of hGSTP1-1 is predominantly hydrophobic and contains essential conserved features such as the N-capping box motif at the beginning of the hydrophobic core of α-helix (Dragani *et al*., 1997). The S150A and D153A mutations in this same helix were also shown to be significantly more thermolabile when compared to the wild-type demonstrating that the local destabilization of the α-helix which is determined by a single capping residue mutation affects the overall

stability of the protein (Dragani *et al*., 1997). The stability of the protein is critical for stabilising ligands that bind to it and to the active site; a stable protein will result in efficient catalytic function.

4.4 Function of hGSTP1-1

The hGSTP1-1 plays a very important role in the regulation of toxins in the body. It accomplishes this by conjugating the electrophilic substance to the GSH tripeptide making the molecules less toxic and easily excretable from the body (Armstrong, 1991). The hGSTP1-1 follows the phase 1 drug metabolism reaction which is often catalysed by members of the cytochrome p450 supergene family (Klaasen, 1996). The cytochrome p450 enzymes are responsible for catalysing the introduction of a functional group such as epoxides into a chemically inactive compound. This electrophilic compound allows the reduced GSH which is a nucleophile to attack the electrophilic centre of the molecule, and the hGSTP1-1 catalyses this reaction.

The active site of the hGSTP1-1 is made of the G-Site and the H-site, the G-site binds the glutathione and the H-site which is half hydrophobic and hydrophilic binds the hydrophobic substrates (Reinemer *et al*., 1992). The H-site is a pocket structure with sidewalls made by residues Phe8 and Tyr108. The other two sidewalls are formed by Val38 and Ile104, with the bottom built by Val10 and Gly205 (Prade *et al*., 1997). The active site also spans the domain interface with the thioredoxin and C-terminal domains providing better binding for GSH and the xenobiotic compound, respectively (Armstrong 1997, Sheehan *et al*., 2001). In this study, the function of the hGSTP1-1 was analysed using the CDNB assay as mentioned above with the addition of GSH. The addition of the GSH to the protein allows for the xenobiotic-conjugates to be noticed and further removed from the cell during phase III drug metabolism. This method measures the specific activity of the protein and is highly reliant on an active protein. The specific activity of the wild-type was found to be 52.4 µmol/min/mg (Figure 3.12 A). The T110S SNP form was found to be 43.2 µmol/min/mg (Figure 3.12C) and had 17.7 % decrease in activity when compared to the wild-type. The L159M SNP form had an activity of 22.4 µmol/min/mg which decreased the activity of the protein by 57.2 % when compared to the wild-type (Figure 3.12B). The specific activity of the hGSTP1-1 in literature has been shown to be around 40 - 70 µmol/min/mg (Zimniak *et al*., 1994). The wild-type and the T110S SNP form fall in this range but not the L159M SNP form which is significantly low. The activity and stability of the protein is very important and the combination of the residues that

make it up influences its function. When the GSH binds the active site it is hydrogen bonded via the sulphur atom of the molecule which is located at the N-terminal end of the helix 1. The hydrogen bond is donated by the Tyr7 residue and this interaction plays a crucial role in GST catalysis by stabilising the activated GSH (GS^-) (Armstrong, 2002; Abdalla *et al*., 2002). The water molecule that forms the hydrogen bond between the sulfur atom and the Tyr7 also has two additional H-bonds to a second water molecule and the main chain nitrogen of Arg13. Additionally, the NH-amide of Arg13 is also H-bonded to the Tyr7 (Prade *et al*., 1997).

The GSH is also held in place by a second water molecule which is H-bonded to the OH oxygen of Tyr108 and this residue further has H-bond contacts to the main chain nitrogen of Gly205 and the nitrogen of Asn204. All these h-bonds have distances ranging from 3.0 -3.5 Å which represent weak electrostatic bonds.

Although not known whether there is any interaction between the Tyr108 and the Thr109 (residue counted excluding the start amino acid), there is a possibility that the residues close to the tyrosine also stabilise the tyrosine which also stabilise the binding of the GSH and therefore allowing for efficient catalysis. When the threonine is substituted with the serine, even though both these amino acids are hydrophilic, denoting that they easily interact with surrounding molecules but the threonine still has a larger molecular mass because in has an additional hydroxyl making it even more hydrophilic as compared to the serine. This may suggest that the threonine has a higher chance of stabilising the surrounding environment than the serine hence not much effect is presented on its activity by T110S.

The L159M SNP form showed an effect on the tertiary structure of the wild-type (Figure 3.8) and this may have affected the stability of the residues around the mutation affecting the stability of the active site and the residues around it. Such mutations may affect the stability of the protein which results in weak interactions of the protein with ligands.

4.5 Influence of mutations on ligand binding

When ANS is bound to the hydrophobic regions of the protein the fluorescence intensity increases. The ligand has been shown to bind the dimer interface of the protein (Figure 3.13). Free ANS has lower fluorescence intensity which increases upon binding to hydrophobic regions of proteins. The results obtained showed an increase in fluorescence intensity of the SNP forms when ANS is bound suggesting that the structure of the SNP forms is slightly

changed exposing more hydrophobic regions for ANS to bind. It could also be that the ANS binds to the dimer interface (Figure 4.3) of the protein and makes direct contact mostly with helix 4 residues.

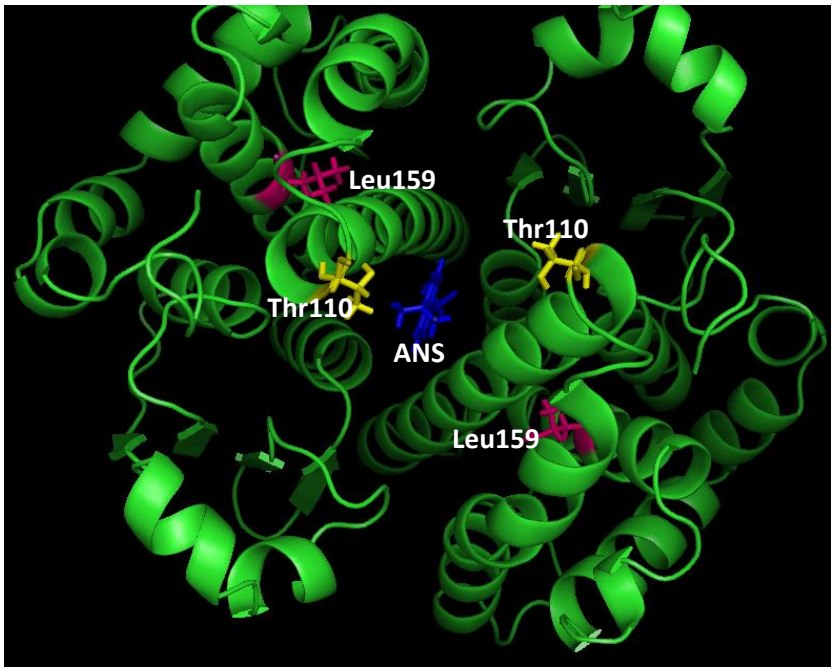


Figure 4.3 ANS binds to the dimer interface of hGSTP1-1. ANS shows high binding affinity for the hydrophobic dimer interface than the active site. The T110S (yellow) and L159M (magenta) are positioned in the α-helices with residues closest to the dimer interface and likely to interact with the ligand.

Binding of the ANS to the dimer interface (Figure 4.3) could slightly alter the structure of the protein while the protein stabilises its binding and for the reason that the T110S is located in the helix 4. The hydrophobic aniline and naphthyl rings of the ANS have been shown to occupy the non-polar H-site in GSTA1-1 (Dirr *et al*., 2005) and this result was also proved to be true for the hGSTP1-1 when the ANS was made to dock specifically to the active site, but when the flexible docking is performed, where the ligand is freely allowed to bind anywhere on the protein. It binds the dimer interface proving that the ANS has higher affinity for the dimer interface of the protein than the active site. The possibility that the ANS can bind to the active site of the hGSTP1-1 as it does to that of the Alpha class cannot be ruled out. This idea arises from the studies done by Cameron and his group in 1995 on class-Alpha (Cameron *et al*., 1995) and class-Pi studies done by Oakley and his group in 1997 which show these proteins in complex with ethacrynic acid drug and it's GSH-conjugate (Oakley *et al.,* 1997). This study demonstrated that the H-site of the Pi- and Alpha-class GSTs bind similar

compounds distinctly. The ANS displacement studies which were performed to prove that the ANS does not bind to the hGSTP1-1 active site were done so using ethacrynic acid. Binding of the non-substrate ligand to the GSTs often but not always occur at the dimer interface. The possibility of an interaction between the non-substrate ligand binding site and the active site of the dimeric protein is interesting

Studies showed no displacement of ANS from the active site (Figure 3.15) and this could be because the ethacrynic acid is a weaker competitor, however binding of ethacrynic acid to the H-site has been shown to be favourable for the hGSTP1-1 because it forms a much more efficient conjugation reaction in class Pi (Oakley *et al*., 1997). It binds more deeply in the H-site forming a direct or indirect hydrogen bond with its acetone oxygen and the H-site, causing an increase in electrophilicity of ethacrynic acid β-alkene carbon and therefore more efficient conjugation. This also proves how important it is for the protein in its dimeric structure for ligandin binding.

It is not known whether the dimeric structure of hGSTP1-1 is important to the biological function of the enzyme, compelling evidence suggests that the dimeric structure improves the protein stability and provides the active site with a proper structure for efficient catalysis (Armstrong, 2010 & Abdalla *et al*., 2002, Dirr *et al*., 1994).

Conclusion

The decreased activity of hGSTP1-1 by L159M SNP form could cause diseases in human because the protein is not functioning fully when the SNP is incorporated to its sequence leading to accumulation of toxins in the body, on the other hand it could serve as an advantage for cancer patients as it will not cause resistance to cancer drugs. The T110S has no significant impact on the protein structure and function and therefore could be a silent mutation without any disadvantages.

REFERENCES

Abdalla, A., Bruns, M., Trainer, J. A., Mannervik, B and Stenberg, G (2002). Design of monomeric human GSTP1-1, a structurally stable but catalytically inactive protein. *Protein Eng.* 15: 827-834

Adler, V., Yin, Z., Fuch, S. Y., Benezra, M., Rosario, L., Tew, K. D., Pincus, M. R., Sardana, M., Henderson, C. J., Wolf, C. R., Davis, R. J and Ronai, Z (1999). Regulation of JNK signaling by GSTpi. *EMBO J.* 18: 1321-1334

Alscher, R., Erturk, N and Heath, L (2002). Role of superoxide dismutase in controlling oxidative stress in plant. *J. Exp. Bot.* 53:1331-1341

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

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

Armstrong R. N (2010). Glutathione *S*-transferases, in comprehensive toxicology (2nd edn) *Elsevierscience.* pp295-321

Barrow, P. A and Soothill, J. S (1997). Bacteriophage therapy and prophylaxis: rediscovery and renewed assessment of potential. *Trends Genet*. 5:268–271

Brambilla, D., Verpillot, R., Kimpe, L., Tarvena, M., Droumaguet, D., Nicolas, J., Canovi, Gobbi, M., Salmona, M., Nicolas, V., Scheper, W., Couvreur, P and Andrieux K (2010). Nanoparticles against Alzheimer's disease: PEG–PACA nanoparticles are able to link the aβ peptide and influence its aggregation kinetic. *J. Control. Release*. 148: 112–124

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

Boyland, E and Chassead, L. F (1969). The role of glutathione and glutathione *S*transferases in mercapturic acid biosynthesis. *Adv. Enzymol.* 32: 173-219

Cocco, R., Stenberg, G., Dragani, B., Rossi, P. D., Paludi, D., Mannervik, B and Aceto, A (2001). The folding and stability of human Alpha class glutathione transferase A1-1 depend on distinct roles of a conserved N-capping box and hydrophobic staple motif. *J. Biol. Chem*. 276: 32177–32183

Chen, C. J., Chin, J. E., Ueda, K., Clark, D. P., Pastan, I., Gottesman, M. M., Roninson, I. B (1986). Internal duplication and homology with bacterial transport proteins in the mdr1 (Pglycoprotein) gene from multidrug-resistant human cells. *Cell Biol*. 47: 381-389

Cho, S. G., Lee, Y. H., Park, H. S., Ryoo, K., Kang, K. W., Park, J., Eom, S. J., Kim, M. J., Chang, T. S., Choi, S. Y., Shim, J., Kim, Y., Dong, M. S., Kim, S. G., Ichijo, H and Choi, E. J (2001). Glutathione S-transferase Mu modulates the stress-activated signals by suppressing apoptosis signal-regulating kinase 1. *J. Biol. Chem*. 279: 12749-12755

Coles, B and Ketterer, B (1990). The role of glutathione and glutathione transferases in chemical carcinogensis. *Crit. Rev. Biochem. Mol. Biol*. 25: 47-70

Cooper, D. N., Ball, E. V and Krawczak, M (1998). The human gene mutation database. *Nucleic Acids Res.* 26: 285–287

Crawford, L.A and Weerapana, E (2016). A tyrosine-reactive irreversible inhibitor for glutathione *S*-transferase Pi (GSTP1). *Mol Biosyt*. 12(6):1768-1771

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

Dirr, H. W., Reinemer, P and Huber, R (1994). Refined crystal structure of porcine class pi glutathione *S*-transferase (pGSTP1-1) at 2.1 Å resolution. *J. Mol. Biol.* 243: 72-92.

Dirr, H. W., Little, T., Kuhnert, D. C and Sayed, Y (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

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

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

Duret, L (2009). Mutation patterns in the human genome: More variable than expected. *PLoS Biol.* 7(2): 0217-0219

Froger, A and Hall, J. E (2007). Transformation of plasmid DNA into *E. coli* using the heat shock method. *J Vis Exp*. 10: 253

Ghisaidoobe, A.B.T and Chung S.J (2014). Intrinsic tryptophan fluorescence in the detection and analysis of proteins: A focus on forster resonance energy transfer techniques. *Int.J.Mol.Sci.* 15:22518-22538

Gordon, D. B., **Wilson, K and Walker, J. M** (1994). Spectroscopic techniques: Pricnciples and techniques of practical biochemistry. Cambridge, New York, Cambridge University Press. 4th ED. 357-363

Gottesman, M. M and Pastan, I (1993). Biochemistry of multi-drug resistance mediated by the multi-drug transporter. *Annu Rev Biochem*. 62: 385-427

Graminski G. F, Zhang, P, Sesay M. A, Ammon H. L and Armstrong R. N (1989). Formation of the 1-(S-glutathionyl)-2,4,6-trinitrocyclohexadienate anion at the active site of glutathione *S*-transferase: Evidence for enzymatic stabilization of σ-complex intermediates in nucleophilic aromatic substitution reactions. *Biochem. J.* 28:6252-6258.

Greenfield, N. J (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc*. 1: 2876-2890.

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

Habig, W. H., Pabst, M. J and Jakoby, W. B (1974) Glutathione *S*-transferases. The first enzymatic step in mercapturic acid formation. *J Biol Chem*. 249: 7130–7139

Habig, W. H., Pabst, M. J., Fleishner, G., Gatmaitan, Z., Arias, I. M and Jakoby, W. B (1974). The identity of glutathione *S*-transferase B with ligandin, a major binding protein of liver. *Proc. Natl.Acad.Sci.* 71:3879-3883

Hagnauer, G. L (1982). Size Exclusion Chromatography. *Anal. Chem*. 54:265-276

Hanene, C., Jihene, L., Jamel, A., Kamel, H and Agn`es, H (2007). Association of GST Genes Polymorphisms with Asthma in Tunisian Children. *Mediat. Inflamm*. 19564:1-6

Harries, L. W., Stubbins, M. J., Forma, D., Howard, G. C. W and Wolf, C. R (1997). Identification of genetic polymorphisms at the glutathione S-transferase pilocus and association with susceptibility to bladder testicular and prostage cancer. *Carcinogenesis.* 18: 641-644

Harshbarger, W., Gondi, S., Ficarro, S. B., Hunter, J., Udayakumar, D., Gurbani, D., Singer, W. D., Liu, Y., Li, L., Marto, J. A and Westover, K. D (2017). Structural and Biochemical Analyses Reveal the Mechanism of Glutathione *S*-Transferase Pi 1 Inhibition by

the Anti-cancer Compound Piperlongumine. *J. Biol. Chem*. 292: 112-120

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

Hayes, J. D and Strange, R. C (2000). Glutathione *S*-transferase polymorphisms and their biological consequences. *J. Pharmacol*. 61: 154-166

Hegazy, U. M., Tars, Z., Hellman, U and Mannervik, B (2008). Modulating Catalytic Activity by Unnatural Amino Acid Residues in a Gsh-Binding Loop of Gst P1-1. *J.Mol.Biol.* 376: 811-826

Hoffmeyer, S., Burk, O and von Richter, O (2000). Functional polymorphisms of the human multi-drug resistance gene: multiple sequence variations and correlation of one allele with Pglycoprotein expression and activity in vivo. *Proc Natl Acad Sci.* 97:3473-3478

Hornby, J. A., Codreanu, S. G., Armstrong, R. N., and Dirr, H. W (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

Housman, D (1995). Molecular medicine: Human DNA polymorphisms. *N. Engl. J. Med*. 332:318-320

Ishii, T., Matsuse, T., Teramoto, S., Matsui, H., Miyao, M., Hosoi, T., Takahashi, H., Fukuchi, Y and Ouchi, Y (1999). Glutathione *S*-transferase P1 (GSTP1) polymorphism in patients with chronic obstructive pulmonary disease. *Thorax*. 54:693–696

Jakobsson, P. J., Morgenstern, R., Mancini, J., Ford-Hutchinson, A and Persson, B (1999). Common structural features of MAPEG- a widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism.

Prot. Sci. 8: 689-692

Ji, X., Zhang, P., Armstrong, R. N and Glliland, G. L (1992). The three dimensional structure of a glutathione *S*-transferase from the Mu gene class. Structural analysis of the binary complex of iso-enzyme 3-3 and glutathione at 2.2 Å resolution. *Biochemistry*. 34: 5317-5328

Karam, R. A., Pasha, H. F., El-Shal, A., Rahman, M. A and Gad, D (2012). Impact of glutathione *S*-transferase gene polymorphisms on enzyme activity, lung function and bronchial asthma susceptibility in Egyptian children. *Gene.* 497: 314–319

Koehler, R. T., Villar, H. O., Bauer, K. E., and Higgins, D. L. (1997). Ligand based protein alignment and isozyme specificity of glutathione S-transferase inhibitors. Proteins: Struct. Funct. Genet. 28:202-216

Laborde, E (2010). Glutathione transferases as mediators of signalling pathways involvedin cell proliferation and cell death. Cell Death Differ. 17: 1373-1380

Laemmli, U. K (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227(5259): 680-685

Lakowicz, J. R (1999). Principles of fluorescence spectroscopy, Protein fluorescence. Singapore, Springer. 3rd ED. 445-465

Li, D., Dandara, C and Parker, M (2010). The 341 C/T polymorphism in the GSTP1 gene is associated with increased risk of oesophageal cancer. *BMC Genetics.* 11(47): 1-9

Lida, R., Tsubota, E., Takeshita, H and Yasuda, T (2008). Multiplex single base extension method for simultaneous genotyping of non-synonymous SNP in the three human SOD genes. *Electrophoresis*. 29:4788-4794

Lin, H. J., Johansson, A. S., Stenberg, G., Materi, A. M., Park, J. M., Dai, A., Zhou, H.,

Gim, J. S., Kau, I. H., Hardy, S. I., Parker, M. W and Mannervik, B (2003). Naturally occurring Phe151Leu substitution near a conserved folding module lowers stability of glutathione transferase P1-1. *Biochem Biophys Acta*. 1649(1):16-23

Listowsky, I (1993). High capacity binding by glutathione S-transferase and glucocorticoid resistance. *Struct. Funct. Glut. Transf.* 199-209

Lo Bello, M., Oakley, A. J., Battistoni, A., Mazzeti, A. P., Nuccetelli, M., Mazzarese, G., Rossjon, J., Parker, M. W and Ricci, G (1997). Multifunctional role of Tyr 108 in the catalytic mechanism of human glutathione trasferase P1-1. Crystallographic and kinetic studies on the Y108F mutant enzyme. *Biochemistry.* 36: 6207-6217

Luca, A., Me, G., Rosato, N., Nicolai, E., Federici, L., Palumbo, C., Pastore, A., Serra, M and Caccuri, A. M (2014). The fine-tuning of TRAF2–GSTP1-1 interaction: effect of ligand binding and *in situ* detection of the complex. *Cell Death. Dis.* 5:1-11

Lynch, M., Sung, W., Morris, K., Coffey, N and Landry, C. R (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci*. 105: 9272- 9277

Lyttle, M. H., Hocker, M. D., Hui, H.C., Caldwell, C. G., Aaron, D. T., Engqvist-Goldstein, A., Flatgaard, J. E., and Bauer, K. E (1994). Isozyme specific glutathione Stransferase inhibitors: desing and sysnthesis. *J. Med. Chem*. 37: 189-194

Manevich, Y., Hutchens, S., Tew, K. D and Townsend, D. M (2013). Allelic variants of glutathione *S*-transferase P1-1 differentially mediate the peroxidase function of peroxiredoxin VI and alter membrane lipid peroxidation. *Free. Radic Biol. Med*. 54: 62– 70

Manevich, Y., Sweitzer, T., Ho Pak, J., Feinstein, S. I., Muzykantov, V and Fisher B. (2002). 1-Cys peroxiredoxin overexpression protects cells against phospholipid peroxidation mediated membrane damage. *PNAS*. 99:11599-11604

Mannervik, B (1985). The isoenzymes of glutathione *S*-transferase. *Adv. Enzymol. Relat. Areas Mol. Biol.* 57: 357-417

Marth, G. T., Korf, I., Yandell M. D., Yeh, T. R., Gu, Z., Zakeri, H., Stitziel, N. O., Hillier, L., Kwok, P and Gish, W. R (1999). A general approach to single nucleotide polymorphism discovery. *Nat. Genet*. 23:452-456

Martignano, F., Gurioli, G., Salvi, S., Calistri, D., Costantini, M., Gunelli, R., Giorgi, U., Foca. F and Casadio, V (2016). GSTP1 Methylation and protein expression in prostate cancer: Diagnostic Implications. *Disease Markers*. 1-6

Masoodi, T. A., Talluri, V. R., Shaik, N. A., Al-Aama, J. Y and Hasan, Q (2012). Functional genomics based prioritization of potential nsSNPs in EPHX1, GSTT1, GSTM1 and GSTP1 genes for breast cancer susceptibility studies. *Genomics.* 99: 330-339

McCord, J and Fridovich, L (1969). Superoxide Dismutase: an enzymatic function for erythrocuprein (hemocuprein). *J. Biol. Chem*. 244:6049-6055

McPherson, A (2004). Introduction to protein crystallization. *Nat. Methods.* 34: 254-265.

Moyer, A. M, Salavaggione, E. O., Wu, T., Moon, I., Eckloff, B. W., Hildebrandt, A. T., Schaid, D. J., Wieben, E. D and Weinshilboum, R (2008). Glutathione *S*-Transferase P1: Gene sequence variation and functional genomic studies. *Cancer Res*. 68 (12): 4791- 4801

Oakley, A., Bello, M. L., Battistoni, A., Ricci, G., Rossjohn, J., Villar, H. O and Parker, M. W (1997). The structures of human GSTP1-1 in complex with glutathione and various inhibitors at high resolution. *J. Mol.Biol.* 274:84-100

Onay, V. U., Briollais, L., Knight, J. A., Shi, E., Wang, Y., Wells, S., Hong, Li., Rajendram, I., Andrulis, I and Ozelik, H (2006). SNP-SNP interactions in breast cancer susceptibility. *B.M.C.* 2006:6-114

Phulukdaree, A., Khan, S., Moodley, D and Chuturgoon, A (2012). GST polymorphisms and early-onset coronary artery disease in young South African Indians. *S. Afr. Med. J.* 102: 628-630

Porath, J., Carlsson, J., Olsson, I and Belfrage, G (1975). Metal chelate affinity chromatography: a new approach to protein fractionation. *Nature*. 258: 598–599.

Potschka, M. (1987). Universal calibration of gel permeation chromatography and determination of molecular shape in solution. *Anal Biochem*. 162(1): 47-64.

Pitt-Rivers, R and Impiombato, F. S (1968). The binding of sodium dodecyl sulfate to various proteins. *Biochem J* 109(5): 825-830

Prade, L., Huber, R., Manoharan, T. H., Fahl, W. E and Reuter (1997). Structures of class pi glutathione *S*-transferase from human placenta in complex with substrate, transition- state analogue and inhibitor. *Struct. Curr. Biol*. 5:1287-1295

Preneta, A. Z (1993). Separation on the basis of size: gel permeation chromatography. Protein purification methods: a practical approach. New York, Oxford University Press. 293-306.

Reinemer, P., Dirr, H. W., Ladenstein, R., Huber, R., Lo Bello, M., Federici, G and Parker, M. W (1992). Three-dimensional structure of class pi glutathione *S*-transferase from human placenta in complex with *S*-hexylglutathione at 2.8 Å resolution. *J. Mol. Biol*. 227:214-226

Robichon, C., Luo, J., Causey, T. B., Benner, J and Samuelson, J. C (2011). Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. *coli* protein contamination after purification by immobilized metal affinity chromatography. *Appl. Environ. Microbiol*. 77: 4634–4646

Ryberg, D., Skang, V., Hewer, A., Philips, D. H., Harries, L. W., Wolf, C. R., Ogreid, D., Ulvik, A., Vu, P and Hangen, A (1997). Genotypes of glutathione transferase M1 and P1 and their significance for lung DNA adduct levels and cancer risk. *Carcinogenesis.* 18:1285-1289

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M and Lincoln, D. S (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 409:928-933

Sayed, Y., Wallace, L. A., and Dirr, H. W (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.

Schmidt, M. F. G (1989). Fatty acylation of proteins. *Biochim. Biophys. Acta*. 988: 411–426

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

Sherrat, P. J and Hayes, J. D (2001). 9 Glutathione *S*-transferases. *Enzym. Met. Drugs. Xenobio*. 319-352

Sherrat, P. J., Manson, M. M., Thomson, A. M., Hissink, E. A. M., Neal, G. E., van Bladeren, P. J., Green, T and Hayes, J. D (1998). Increase bio-activation of dihaloalkanes in rat liver due to induction of class Theta glutathione S-transferase T1-1. *Biochem. J.* 335: 619-630

Sinning, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H.W., Huber, R., Gilliland, G. L., Armstrong, R. N., Ji, X., Board, P. G., Olin, B., Mannervik, B and Jones, T. A (1993). Structure determination and refinement of human Alpha class glutathione transferase A1-1, and a comparison with the Mu and Pi class enzymes. *J.Mol.Biol.* 232: 192-

212

Slavik, J (1982) Anilinonaphthalene sulfonate as a probe of membrane composition and function. *Biochem. Biophys. Acta*. 694:1-2

Stenberg, G., Abdalla, A. M., and Mannervik, B (2000) Tyrosine 50 at the subunit interface of dimeric human glutathione transferase P1-1 is a structural key residue for modulating protein stability and catalytic function. *Biochem. Biophys. Res. Commun*. 271, 59–63

Stenberg, G., Dragani, B., Cocco, R., Mannervik, B and Aceto, R (2000). A conserved "Hydrophobic staple motif" plays a crucial role in the refolding of GSTP1. *J. Biol. Chem.* 272: 10421-10428

Sunyaev, S., Ramensky, V., Koch, I., Cathe, Ill. W., Kondrashov, A and Bork, P (2001). Prediction of deleterious human alleles. *Hum. Mol. Gen*. 10: 591-597

Summers, D. F., Maizel, J. V. Jr and Darnell, J. E Jr (1965). Evidence for virus-specific non-capsid proteins in poliovirus-infected HeLa cells. *Proc Natl Acad Sci*. 54(2): 505-513

Taillon-Miller, P., Gu, Z., Li, Q., Hillier, L and Kwok, P. Y (1998). Over-lapping genomic sequences: a treasure trove of single nucleotide polymorphisms. *Genome Res.* 8:748-754

Tian, D., Wang, Q., Zang, P., Araki, H and Yang, S (2008). Single nucleotide mutations rate increases close to insertions/deletions in eukaryotes. *Nature.* 455: 105-108

Tsuchida, S and Sato, K (1992). Glutathione transferases and cancer. *CRC Crit. Rev. Biochem. Mol. Biol*. 27: 337-384

Vasieva, O (2011). The Many Faces of Glutathione Transferase Pi. *Curr. Mol. Med.* 11:129-139

Wu, B and Dong (2012). Human cytosolic glutathione *S*-transferases: structure, function and drug discovery. *Trends Pharmacol Sci*. 33:656-668

Wang, Z and Moult, J (2001). SNPs, protein structure and disease. *Hum. mutation*. 17: 263-270

Weber, G and Young, L. B (1964). Fragmentation of bovine serum albumin by pepsin I. The Origin of the Acid Expansion of the Albumin Molecule. *J Biol Chem*. 239:1415-1423

Zaki, M. A., Moghazy, T. F., Kamal El-Deeb, M., Mohamed, A. H and Mohamed, N (2015). Glutathione S-transferase M1, T1 and P1 gene polymorphisms and the risk of developing type 2 diabetes mellitus in Egyptian diabetic patients with and without diabetic vascular. *Alexandria Med J*. (2015) 51, 73–82

Zheng, L., Baumann, U., and Reymond, J. L (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucl. Acids Res*. 32(e115):1-5

Zimniak, P., Nanduri, B., Pikula, S., Bandorowics-Pikula., Singhal, S., Srivastava, S and Awasthi C.Y (1994). Naturally occurring human glutathione S-transferase GSTP1-1 isoforms with isoleucine and valine in position 104 differ in enzymic properties. *Eur. J. Biochem.* 224: 893-899

Turnitin2017.docx

ORIGINALITY REPORT

13%	3%	7%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	[redacted]	6%
2	[redacted] Reviews, 2011	2%
3	[redacted]	1%
4	[redacted]	1%
5	[redacted]	1%
6	[redacted]	1%

7	Submitted to Abdullah Gul University	<1 %
8	Vlaykova, Tatyana, Maya Gulubova, Yovcho Yovchev, Dimo Dimov, Denitsa Vlaykova, Petjo Chilingirov, and Nikolai Zhelev. "Glutathione-S-Transfereses in Development, Progression and Therapy of Colorectal Cancer", Colorectal Cancer Biology - From Genes to Tumor, 2012.	<1 %
9	[illegible]	<1 %
10	[illegible]	<1 %
11	researchspace.auckland.ac.nz	<1 %
12	[illegible]	<1 %
13	R Laurindo. "Glutathione transferases as mediators of signaling pathways involved in cell proliferation and cell death", Cell Death & Differentiation, 07/02/2010	<1 %