

https://dspace.org/

Exploring the Structure, Function and Stability of Glutathione Transferases Engineered from Intra- and Inter-class Consensus Sequences: How Forgiving is Nature?

Mulenga, Thabelo

2024-10

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/45436

Mulenga, Thabelo. (2024). Exploring the Structure, Function and Stability of Glutathione Transferases Engineered from Intra- and Inter-class Consensus Sequences: How Forgiving is Nature? [Master's dissertation, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/45436

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



Exploring the Structure, Function and Stability of Glutathione Transferases Engineered from Intra- and Inter-class Consensus Sequences: How Forgiving is Nature?

by

Thabelo Mulenga

(1924159)

MSc dissertation

Submitted in fulfilment of the requirements for the degree

Master of Science

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Prof. Yasien Sayed

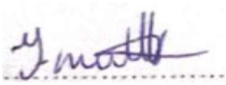
Co-Supervisor: Prof. Ikechukwu Achilonu

October 2024

Declaration

I, Thabelo Mulenga (1924159), am a student registered for the degree of Master of Science in the academic year 2024. 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 this, or any other University.
- The information used in the dissertation has not 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 conventions in referencing the thoughts and ideas of others.
- I understand that the University of the 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: _____  _____ Date: 10-14-2024

Abstract

Protein folding is an enigmatic biochemical process that is foundational to the structural and functional requirements of a cell. The problem of protein folding, in a nutshell, concerns itself with the rate of protein folding as well as the conversion of amino acids from a linear sequence to a fully folded structure. This problem is partly answered by the existence of folding pathways. The folding funnel was conceptualised as a depiction of folding pathways, and it is a framework that illustrates that native proteins naturally favour the lowest energy state, encountering kinetic and thermodynamic barriers as they fold. Consensus protein design, based on this understanding, aims to: (1) enhance stability and (2) navigate the pitfalls of folding by modifying the folding funnel of a protein. This approach can also shed light on the significance of evolutionarily conserved residues. In this study, consensus protein mutants were generated for the Alpha and Mu glutathione transferases (GSTs) classes. The consensus proteins were then benchmarked against the parental proteins that were chosen (hGSTA1-1 and hGSTM1-1). The Alpha consensus mutant had 11 consensus mutations, including a notable M50L mutation, which affects the dynamic behaviour of helices α2 and α9, while the Mu consensus mutant had 13 unique mutations. Protein production and purification showed that the Mu consensus mutant had larger and purer yields. Data from far-UV circular dichroism studies and root-mean-squared-fluctuation (RMSF) from molecular dynamics (MD) simulations showed that the secondary structural components of the Alpha and Mu proteins remained largely the same, although the Alpha consensus mutant displayed a far lower molar residue ellipticity reading than its wildtype counterpart, indicating the disruption of secondary structural elements, likely caused by the M50L mutation. The ANS binding results showed that the M50L mutation in the Alpha consensus protein caused an increase in exposure of the surface area of the H-site, while the Mu consensus protein had a decrease in the solvent accessibility of its H-site. Thermal shift assay results indicated the consensus proteins had increased thermal stability. Enzyme kinetics results showed that the functionality of the proteins was severely diminished in the consensus mutants, particularly the Alpha consensus mutant. MD simulation results showed that there was an overall increase in the rigidity and compactness of the consensus mutant proteins, further affirming the improvement of thermal stability, while signalling the loss in functionality. The results produced herein have the potential to facilitate the proliferation of engineered GSTs for biotechnological applications that require proteins with an increased half-life and greater stability.

Dedication

Forty thousand

and more.

Acknowledgments

Thank you to Professor Sayed who taught me to measure twice and cut once, and to Professor Achilonu whose ambition and insights are boundless.

Thank you to the people of the PSFRU who daily prove that an institution is more than brick and mortar.

Thank you to the National Research Foundation for their financial support.

And, as always, thank you to my mother, Denga Mulenga, without whom these words would not have been written.

Conference output

The Molecular Bioscience Research Thrust (MBRT), University of the Witwatersrand (Johannesburg, South Africa), November 2023. Poster presentation. Exploring the Structure, Function and Stability of Glutathione Transferases Engineered from Intra- and Inter-class Consensus Sequences: How Forgiving is Nature? Mulenga, T, Ikechukwu, I.A., and Sayed Y.

Table of Contents

Declaration i
Abstract ii
Dedication iii
Acknowledgments iv
Conference output v
Abbreviations viii
List of Figures x
List of Tables xii
Chapter 1: Introduction 1
1.1 Protein folding and stability 1
1.1.2 Folding funnel 5
1.2 Consensus protein design 8
1.3 Glutathione transferases 10
1.3.1 Structure 10
1.3.2 Function 10
1.3.3 Reaction mechanism 11
1.3.4 Human hGSTA1-1 13
1.3.5 Human hGSTM1-1 15
1.4 Evolution 17
1.5 Rationale 17
1.6 Aim and objectives 18
1.6.1 Aim 18
1.6.2 Objectives 18
Chapter 2: Methods and Materials 19
2.1 Engineering of consensus proteins 19
2.2 Protein production and purification 19
2.2.1 Transformation 19
2.2.2 Production of recombinant proteins 23
2.2.3 IMAC purification 24
2.2.4 Dialysis 25
2.3 Purity analysis via SDS-PAGE 25
2.4 Spectrophotometric determination of protein concentration and purity 26
2.5 Structural characterisation 27
2.5.1 Far-UV CD 27

2.5.2 ANS binding studies 28
2.6 Enzyme activity and Michaelis-Menten kinetics 28
2.7 Thermal shift assay with SYPRO® orange dye 30
2.8 Molecular dynamics simulations 31
Chapter 3: Results 33
3.1 Engineering of consensus proteins 33
3.2 Purification 33
3.3 Purity analysis via SDS-PAGE 37
3.4 Spectrophotometric determination of protein concentration and purity 37
3.5 Far-UV CD 42
3.6 ANS binding 42
3.7 Thermal shift assays 45
3.8 Enzyme activity and Michaelis-Menten kinetics 45
3.9 Molecular dynamics simulations 51
Chapter 4: Discussion 61
4.1 Protein production and purification 61
4.2 SDS-PAGE purity analysis 62
4.3 Far-UV CD 62
4.4 ANS binding 62
4.5 Thermal shift assays 66
4.6 Enzyme kinetics 67
4.7 MD simulations 68
4.8 Conclusion 70
Chapter 5: References 73

Abbreviations

ΔG – Change in Gibbs free energy

Δ*H* – Change in enthalpy

ΔS_{conf} – Change in conformational entropy

ΔS_{solv} – Change in solvent entropy

ε – Molar extinction coefficient

ANS – 8-anilinonaphthalene-1-sulfonic acid

AU – Arbitrary units

C_{α} – Alpha carbon

c-Jun N-terminal kinase – JNK

CD – Circular dichroism

CDNB – 1-chloro-2,4-dinitrobenzene

C-GST – Cysteine-type GST

Da – Dalton

E.C. – Enzyme commission number

GdnHcl – Guanidinium chloride

G-site – Glutathione binding site

GS-DNB – 1-(*S*-glutathionyl)-2, 4-dinitrobenzene

GSH – Glutathione

GST – Glutathione transferase

hGSTA1-1 – Human glutathione transferase A1-1

hGSTM1-1 – Human glutathione transferase M1-1

H-site – Hydrophobic binding site

IMAC – Immobilised metal affinity chromatography

K_M – Michaelis constant

L-site – Ligand binding site

MD – Molecular dynamics

MSA – Multiple sequence alignment

MW – Molecular weight

NTA – Nitrilotriacetic acid

PDB – Protein data bank

RFU – Relative fluorescence units

Rg – Radius of gyration

rGSTM1-1 – Rat glutathione transferase M1-1

RMSD – Root-mean-squared-deviation

RMSF – Root-mean-squared-fluctuation

S-GST – Serine-type GST

SDS-PAGE – Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SOC – Super optimal broth with catabolite repression

T_m – Melting temperature

Tris – Tris(hydroxymethyl)-aminomethane

UV-Vis – Ultraviolet-visible

V_{max} – Maximal velocity

Y-GST – Tyrosine-type GST

List of Figures

Figure 1. Schematic diagram of the four levels of protein structure. 4
Figure 2. The folding funnel of a predominantly alpha helical protein. 7
Figure 3. The consensus protein folding funnel generated from the MSA of five homologous sequences. 9
Figure 4. The detoxification pathway of xenobiotic substances and reaction mechanism of enzyme-catalysed GSH conjugation reaction. 12
Figure 5. Ribbon structure of hGSTA1-1 in its homodimeric conformation (PDB: 1K3Y). 14
Figure 6. Ribbon structure of hGSTM1-1 in its homodimeric form (PDB: 1GTU). 16
Figure 7. MSA used to generate consensus sequences. 20
Figure 8. MSA used to generate the Mu consensus sequence. 21
Figure 9. MSA used to generate the interclass consensus sequence. 22
Figure 10. Graphical representation of the consensus sequences generated after MSA. 34
Figure 11. The elution profile of the Ni^{2+} IMAC purification of the Alpha wildtype and consensus mutant proteins. 35
Figure 12. The elution profiles of the Ni^{2+} IMAC purification of the Mu wildtype and consensus mutant proteins. 36
Figure 13. SDS-PAGE images depicting the steps of purification for the Alpha and Mu wildtype and consensus mutant proteins. 38
Figure 14. Linear regression models of the corrected absorbance data for the wildtype and mutant proteins of the Alpha and Mu classes. 40
Figure 15. Qualitative assessment of the purity and yield of the proteins after purification. 41
Figure 16. Far-UV CD spectra for the wildtype and mutant proteins of the Alpha and Mu classes. 43
Figure 17. ANS fluorescence spectra of the wildtype and consensus mutant proteins of the Alpha and Mu classes in various conditions. 44
Figure 18. Thermal shift assays graphically represented as negative melt curve derivative plots. 47
Figure 19. Thermal shift assays graphically represented as melt curves in relative fluorescence units.. 48

Figure 20. Michaelis-Menten plots of the wildtype and consensus mutant proteins of the Alpha and Mu classes. 50
Figure 21. Cα-RMSD of the Alpha and Mu wildtype and consensus mutant proteins from the MD simulations conducted over 500 ns. 52
Figure 22. Cα-RMSF of the Alpha and Mu wildtype and consensus mutant proteins per residue.. 53
Figure 23. Radius of gyration of the Alpha and Mu wildtype and consensus mutant proteins from the MD simulations conducted over 500 ns. 54
Figure 24. Atomic distance between the tyrosine 9 (Y9) residues and tyrosine 6 (Y6) residues of both subunits from the MD simulations conducted over 500 ns. 56
Figure 25. Superimposition of the hGSTA1-1 and Alpha consensus mutant structures. 57
Figure 26. Superimposition of the hGSTM1-1 and Mu consensus mutant structures. ..58
Figure 27. Sitemap analysis results presented as a cavity in the surface of the protein. 59

List of Tables

Table 1: Summary of the purification metrics obtained from spectrophotometric measurements at various wavelengths 39
Table 2: Summary of the melting temperatures (T_m) of the wildtype and consensus mutant proteins 46
Table 3: Summary of the enzyme kinetics results from steady-state conditions 49
Table 4: Summary of the sitemap analysis results of the consensus mutant and wildtype proteins 60

Chapter 1: Introduction

Nature exerts selective pressure on organisms, giving rise to new forms. These adaptations need not be as grand as speciation. In fact, it is amongst macromolecules such as proteins where the most diversity is observed. In the case of proteins, these adaptations are manifested as mutations in the amino acid sequence of the protein. Evolution, as observed in proteins, creates large supergene families of proteins. The proteins in these supergene families may have diverse functions, but they possess conserved regions and are all subject to the laws of protein folding.

1.1 Protein folding and stability

Protein folding is a spontaneous event that occurs in a state of equilibrium. The reversibility of protein folding has also been long established [1]. Given the nature of protein folding, the conformational entropy (ΔS_{conf}) renders the process of protein folding thermodynamically unfavourable [2]. This makes intuitive sense as a linear sequence seeks to attain compactness, thereby making the process entropically opposed. However, protein folding does not occur in a vacuum, rather it occurs in a dynamic molecular milieu. The ΔS_{conf} is, therefore, counteracted by various forces that ultimately make protein folding favourable under physiological conditions. The three main forces are solvent entropy (ΔS_{solv}), the hydrophobic effect and the enthalpy (ΔH) of the hydrogen bonds and van der Waal's forces, with the hydrophobic effect being the greatest contributor of the three. These forces ultimately contribute to a favourable Gibbs free energy (ΔG) for the folded protein. Other electrostatic forces and amino acid contacts contribute to the enthalpic term that drives protein folding [2].

The predominating forces in protein folding, namely, ΔS_{solv} and the hydrophobic effect, require hydrophobic constituents in order to take effect [3]. These forces are, therefore, closely related. In the case of proteins, these hydrophobic constituents are the hydrophobic amino acid residues. In the case of the hydrophobic effect, hydrophobic residues minimise interactions with the solvent. As the protein folds, its surface area is greatly reduced in order to increase its compactness, as a result, the hydrophobic residues can be described as forming micellar-like structures [3]. Micelles are amphipathic molecules composed of hydrophobic interiors and hydrophilic exteriors. This analogy to micelles is useful for the visualisation of the molecular composition of proteins. Due to the minimisation of the interactions between the hydrophobic residues and the hydrophilic residues, the hydrophobic amino acid residues

then form a hydrophobic core, to which the solvent has reduced access, leading to the contribution of an increased ΔS_{solv}., while the protein surface is hydrophilic.

In the case of ΔS_{solv}, buried hydrophobic residues are the driving force of protein folding [4]. Non-polar residues are unable to form hydrogen bonds and when they are exposed to the solvent, they cause water molecules to form hydrogen bonds with themselves. This produces a highly ordered structure termed a clathrate, which is a cage-like structure surrounding a non-polar residue. Clathrates are, therefore, thermodynamically unfavourable because their formation is entropically opposed. When a protein folds, the non-polar residues are buried, disrupting the formation of clathrates. This results in the water molecules being disordered, thereby increasing the overall entropy term by increasing the ΔS_{solv} [4]. This phenomenon is the greatest contributor to converting a linear amino acid sequence into a fully folded protein.

Protein folding is a process that can be broadly dissected into four categories: the formation of the primary, secondary, tertiary, and quaternary structures. The process of protein folding may be described as heterogenous cooperativity [5]. After the primary structure has formed, a peptide chain exists in which not all of the molecules are isoenergetic. As protein folding is a highly cooperative process, all interactions that occur between the molecules of the peptide are highly interdependent. Furthermore, there are steric constraints to certain molecules. This places kinetic energy barriers which prevent the budding protein from accessing certain conformations [6]. This phenomenon is a partial answer to Levinthal's paradox, which states that the amount of time it would take for a protein to randomly sample all possible conformations before finding its native structure would be roughly equal to the age of the universe. These constraints direct protein folding onto specific pathways, and thus specific conformations.

The primary structure of a protein is defined by the linear sequence of amino acids that are held together by a peptide bond which joins the alpha carboxyl group of an amino acid to the subsequent alpha amino group of the next amino acid (Figure 1) [7]. The secondary structure, having been determined by the backbone of the polypeptide forms regular repeating patterns which are stabilised by hydrogen bonding [7]. Secondary structural elements are referred to as beta sheets and alpha helices. The tertiary structure then relies heavily on the amino acid side chain interactions, which is where the aforementioned hydrophobic effect and ΔS_{solv} come into play. Within the purview of the tertiary structures, domains are formed. Domains are independent folding units composed of at least two layers of secondary structure [8].

Domains are described as independent because they would seamlessly assume their normal conformation if they were to fold in a vacuum. The quaternary structure concerns itself with the interactions between fully folded protein monomers/polypeptide chains. Hydrophobic interactions and electrostatic interactions stabilise the quaternary structure.

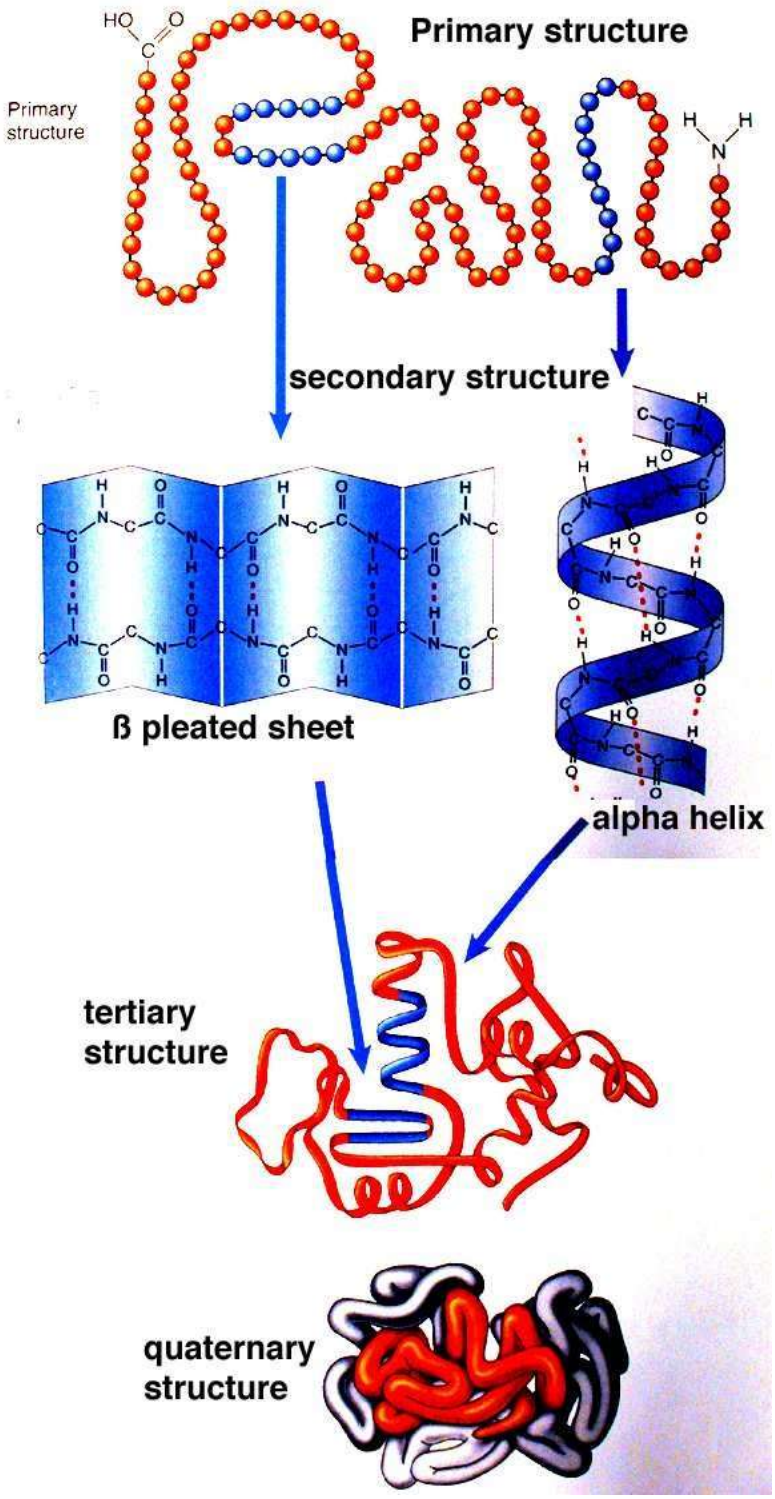


Figure 1. Schematic diagram of the four levels of protein structure. The primary structure is the most rudimentary level of protein structure, and it consists of the amino acid sequence. The secondary structure of a protein is characterised by the folding of the backbone, exemplified by β sheets and α helices. The tertiary structure of a protein is characterised by the interaction between the amino acid sidechains. The quaternary structure of a protein is characterised by the interactions between subunits. Figure taken from Campbell and Reece, 2004 [7].

1.1.2 Folding funnel

Protein folding is a process that has been theorised and conceptualised in the framework of an energy landscape. This conceptual framework is visually represented by what is called a folding funnel [9]. More specifically, the folding funnel largely concerns itself with the kinetics of protein folding, and the accompanying entropy and free energy terms (Figure 2) [10]. In the context of a folding funnel, a molecular ensemble of proteins is conceived to slide down the funnel, until reaching the lowest possible energy state. This tendency of folded proteins having an innate proclivity for lower conformational energy states is what solves Levinthal's Paradox [11][12].

The funnel is conceptualised as a 3-dimensional construct, where in the depth of the funnel represents the conformational free energy, while the width of the funnel represents the conformational entropy [10]. As the ensemble of proteins slides down the funnel, there are kinetic obstacles that confer an additional layer of complexity to protein folding. Within the folding funnel, peaks and troughs, or "mountains" and "valleys", are inaccessible or favourable energetic states for the protein molecules, respectively [13]. The troughs, formally termed as local minima, may be representative of stable intermediate states of the protein [9]. These minima, however, may also be kinetic energy traps, wherein the protein is in a low energy conformation but is not fully and correctly folded, thereby requiring exogenous energy sources to allow it to escape the energy trap and proceed onto the correct folding trajectory.

The valleys themselves are a microcosm of the landscape, and present varying degrees of ruggedness. The ruggedness of the bottom of a valley has long thought to be illustrative of a protein with increased flexibility, which reduces the gap in energy between conformers, thereby allowing the protein to access a higher number of conformers [14]. The bottom of the folding funnel itself is proposed to be a rugged landscape. This implies that the fully folded protein possesses a degree of flexibility that is dictated by the extent of ruggedness at the bottom of the folding funnel.

In higher order oligomeric complexes, the folding funnel gains a further layer of complexity. In the event that fully folded subunits interact to form oligomers, the conceptual implications of this interaction are two-fold. Firstly, the formation of an oligomer entails the theoretical fusion of two or more separate folding funnels, i.e., the folding funnels of each subunit [14].

Secondly, the bottom of the fused funnel becomes a collation of the micro-landscapes of the subunits involved in the higher order complex. The terrain of the energy landscape at the bottom of a funnel is, therefore, ultimately instructive of the functionality of the protein. This has interesting implications for traditional understandings of enzyme-substrate interactions. The conventional lock-and-key and induced-fit models are hugely impacted by this conceptualisation of the folding funnel. The efficiency of enzyme-substrate interactions is, therefore, likely subject to the degree of flexibility of a protein, which in turn is a by-product of the terrain at the bottom of the folding funnel.

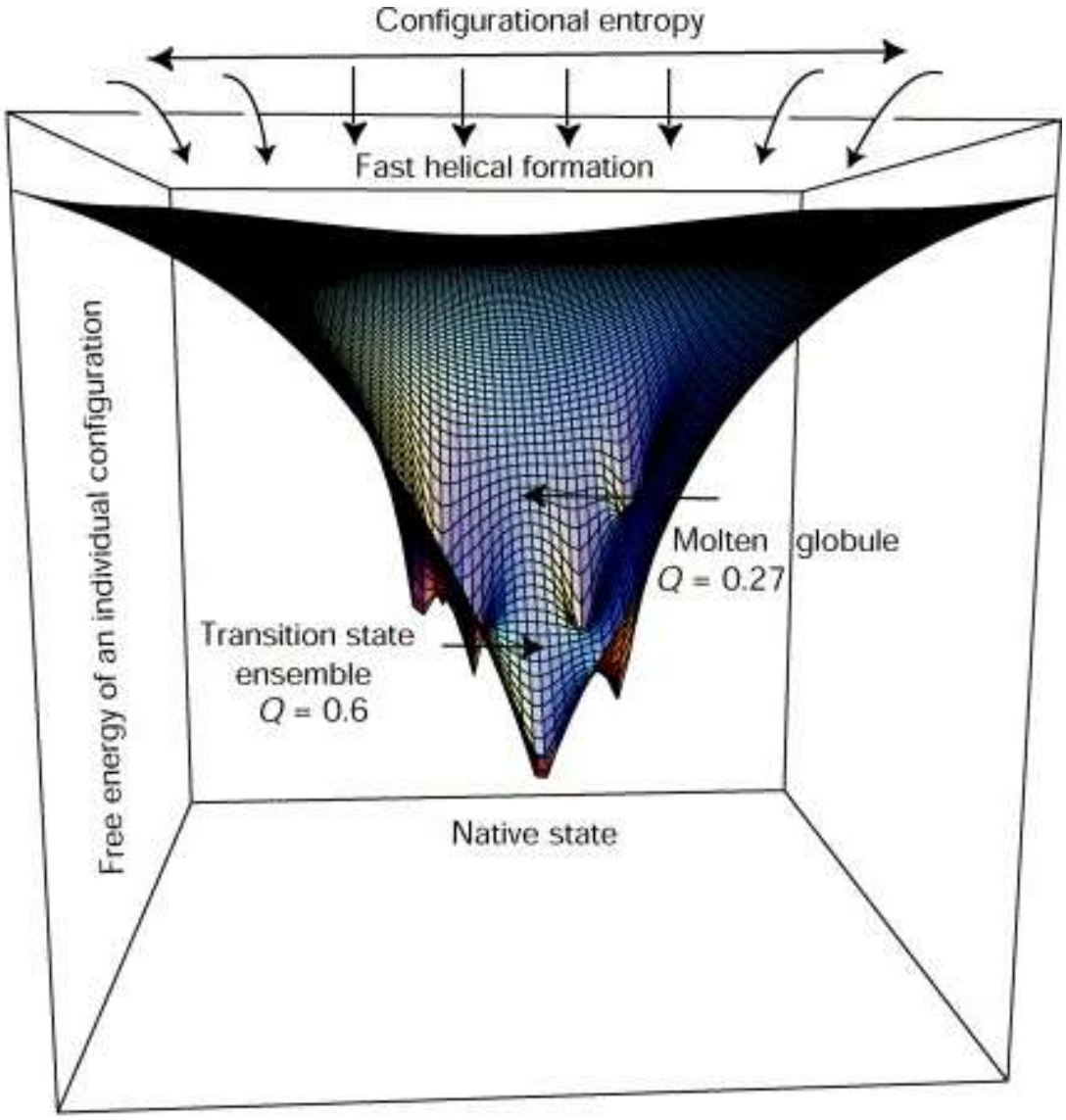


Figure 2. The folding funnel of a predominantly alpha helical protein. The height of the folding funnel represents the conformational free energy, while the width of the funnel represents the conformational entropy. The native state of a protein is the lowest free energy state, with the entropy significantly decreased as a fully folded protein has low entropy. Q is the reaction coordinate used to describe protein folding and it is the fraction of native contacts. Taken from Onuchic *et al.,* 1996.

1.2 Consensus protein design

In the field of protein engineering, a distinct methodology has firmly ensconced itself in a position of legitimacy. Consensus protein design is a so called semi-rational approach to protein engineering. So called because it incorporates computationally derived predictive algorithms with the phylogenetic information of a protein, as well as its structure and function [15]. The process of consensus protein design consists of four primary steps. These steps are: (1) selection of a target protein domain, (2) homologous sequence acquisition, (3) multiple sequence alignment (MSA), and (4) quantifying sequence conservation [15].

In summation, consensus sequence design begins with the identification and selection of target homologous sequences. The homologous sequences, which may be sequences of proteins within the same family, or domains that are related to one another, are aligned in an MSA. Once an MSA is conducted, the most conserved residues are identified. This is then followed by the quantification of residue conservation. Once sequence conservation has been calculated, the results obtained thereafter may be utilised in three ways. Firstly, single to multiple point mutations of the most conserved residues can be done to target sequences. Secondly, novel full-length sequences may be generated, circumventing the need to identify stabilising residues. Thirdly, conserved residues may be deliberately targeted in directed evolution studies.

The underlying thesis posited by proponents of consensus protein design is that by identifying and selecting the most conserved amino acid residue at a given position, the residues give rise to a stabilising effect that creates the most stable protein possible [15]. A corollary of this assumption is that proteins that have been engineered through consensus design have higher melting temperatures (T_m). Due to this direct proportionality, T_m, which is the temperature at which 50% of the protein is denatured, is an accurate indicator of the stability of protein.

Another pelt on the wall that proponents of this technique invoke is the smoothing of the folding funnel of proteins (Figure 3). Consensus protein design has been shown to take a protein of highly complicated folding mechanisms and create a protein that displays reversible two-state folding. It was shown, using the fibronectin type III domain consensus protein (FN3con), that the folding rate of a protein can be drastically increased, and its unfolding rate can be decreased [16]. This illustrates that consensus protein design has the potential to shift the folding equilibrium toward the native state of the protein, an indication of kinetic stability.

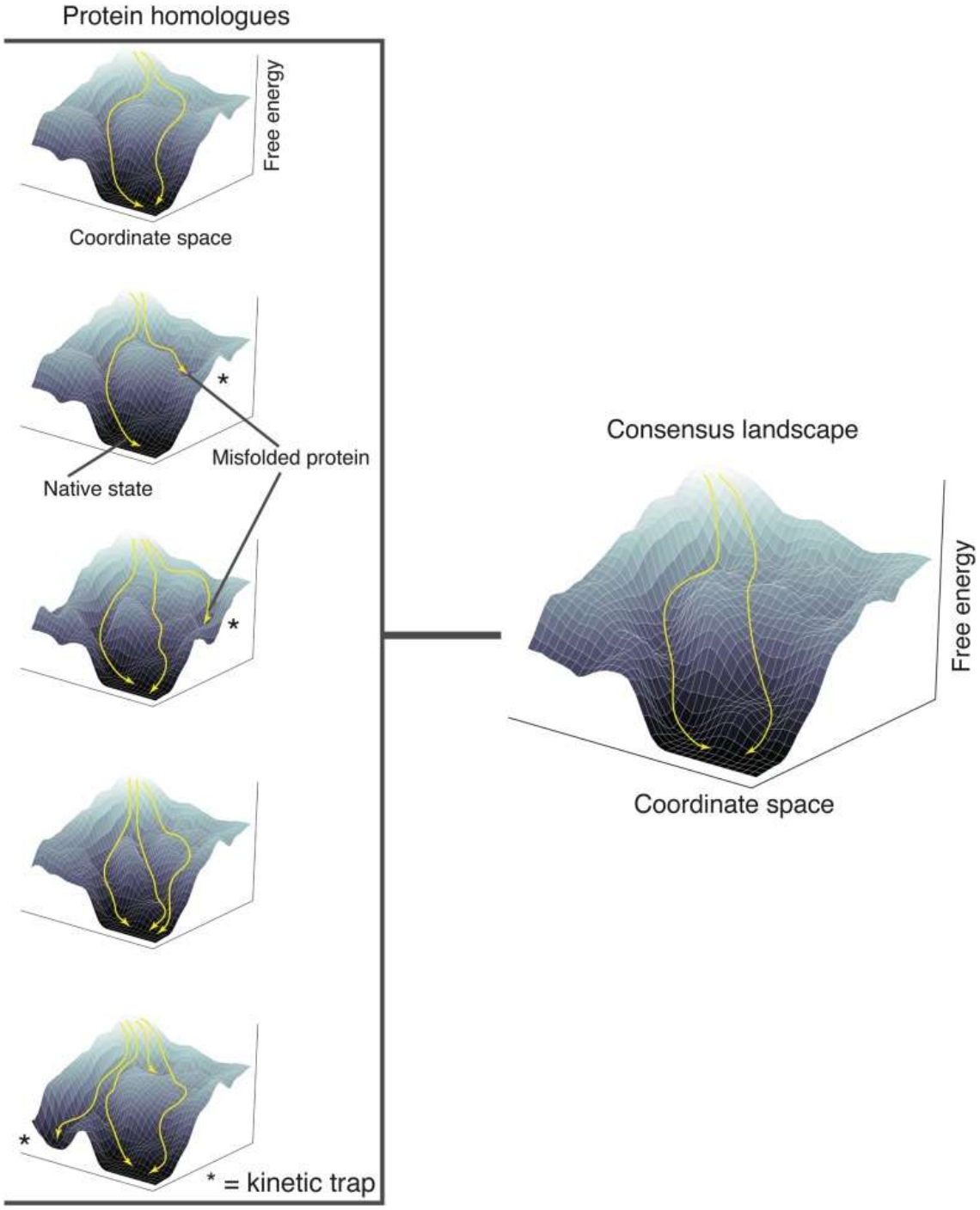


Figure 3. The consensus protein folding funnel generated from the MSA of five homologous sequences. Consensus protein design suggests that the MSA of homologous proteins produces a consensus protein whose folding funnel and energy landscape is rid of the kinetic traps seen in the energy landscapes of the individual proteins. The asterisk (*) indicates kinetic energy traps, wherein irreversible misfolded states or aggregates of the proteins reside. Taken from Porebski *et al.,* 2016 [15].

1.3 Glutathione transferases

1.3.1 Structure

Glutathione transferases (GSTs) (E.C. 2.5.1.18) comprise a supergene family of enzymes involved in phase II detoxification. It has been estimated that 18 GST classes are expressed in humans at the very minimum [17]. GSTs are subdivided into three major families: cytosolic, microsomal, and mitochondrial. Cytosolic GSTs are composed of eight classes, designated by Greek letters; namely, Alpha, Pi, Mu, Kappa, Zeta, Theta, Omega, and Sigma [18]. Arabic numerals are used to describe the subunits, due to the number of alleles for each GST gene class [19]. A monomer of the type 1 gene locus of the Mu class would, therefore, be signified as GSTM1. In the event of dimerisation, the protein would be signified as GSTM1-1 or GSTM1-2, depending on whether it was a homodimer of the type 1 gene locus or a heterodimer of the type 1 and 2 gene loci, respectively.

GSTs are obligate homodimers with the independent subunits being unable to perform any function (catalytic and ligandin) until oligomerisation occurs [20]. Each subunit possesses an active site comprises the hydrophobic binding site (H-site) and the glutathione-binding site (G-site). The G-site, as the name suggests, binds to reduced glutathione (GSH), which is a physiological tripeptide, and is formed by the thioredoxin-like domain [20]. Likewise, the H-site binds to hydrophobic substrates.

These two sites are integral to the conjugation reaction that is catalysed by GSTs. Each monomer that comprises the obligate dimer possesses its own H-site and G-site, which form an active site. This means that each subunit in a GST molecule has its own active site. Through a nucleophilic addition reaction, xenobiotics are metabolised and removed via the mercapturate pathway. Cytosolic GSTs themselves are often characterised according to their catalytic residue. Cytosolic GSTs, therefore, may fall into either the tyrosine GST group (Y-GST) or the serine/cysteine GST group (S/C-GST) [21]. The subjects of this study; namely, the Alpha and Mu classes are both Y-GSTs.

1.3.2 Function

GSTs are integral to phase II detoxification. Phase II detoxification is characterised by the conversion of non-polar electrophilic substances into soluble substances that are easily removed from the body. GSTs have been shown to be involved in the resistance to antibiotics and chemotherapeutic agents, quite notably in the case of Pi GSTs, where increased expression of Pi GSTs led to the resistance of chemotherapy treatments [22]. The synthesis of sex steroids and prostaglandins has been partially attributed to the function of Alpha and

Sigma GSTs, respectively [21]. In an especially unique function, GSTP1 has been shown to inhibit the signal of c-Jun N-terminal kinase (JNK) [23]. In its monomeric form, GSTP1 can form a complex with JNK, leading to the inhibition of JNK signalling. An important functional role that has also been observed in GSTs is the ligandin function. The ligandin function of GSTs is characterised by non-specific binding of a wide range of substances from bilirubin to steroids. Broadly speaking the ligandin binding function confers the ability of sequestration onto GSTs. The ligandin function of GSTs is tied to the shape of the cleft created by the dimer interface. The V-shaped cleft, termed the ligand binding site (L-site), has been identified, through the use of crystal structure data, as a locus of ligand binding in many, but not all GSTs [24]. The V-shaped cleft is specific only to several GSTs such as the Alpha, Mu, and Pi GSTs. Ligandin binding is, however, variable across different classes and species of GSTs [25]. In addition to the L-site, other binding sites such as the N-terminal end of the β2 strand in hGSTP1-1 GSTs perform binding functions [26].

1.3.3 Reaction mechanism

The main function of GSTs is to metabolise xenobiotic substances. This is achieved via a nucleophilic addition reaction, which is catalysed by GST (Figure 4). Briefly, GSH, which binds to the G-site, attacks the electrophilic group that is bound to the H-site, producing a conjugated product. The solubilised conjugated product is discarded through the mercapturate pathway. The interaction between the G-site of the GST and the sulfur of GSH is dependent on the catalytic residue. In the case of Y-GSTs, the active site tyrosine is the catalytic residue.

In the process of the GST-catalysed reaction, a ternary complex is formed, which is composed of GST, GSH, and a hydrophobic substrate. The hydroxyl group of the active site tyrosine is a hydrogen bond donor, donating its hydrogen bond to the sulfur of GSH. The thiol group of GSH as a whole is then endowed with a lower pK_a, which enables GSH to exist as glutathionyl (GS^-) at physiological conditions [20]. The role of the catalytic residue is so integral that site-directed mutagenesis studies of human GSTP1-1 (hGSTP1-1) have shown that a Y7F mutation severely diminishes the enzymatic activity of the enzyme [27].

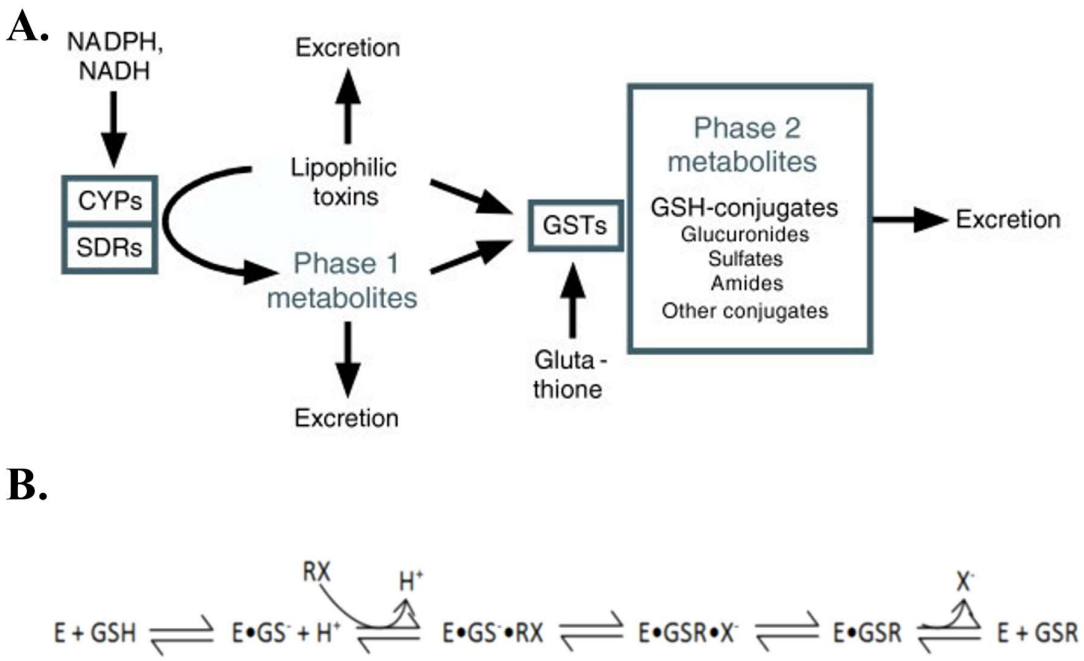


Figure 4. The detoxification pathway of xenobiotic substances and reaction mechanism of enzyme-catalysed GSH conjugation reaction. (A) The detoxification of xenobiotics. Lipophilic xenobiotic substances are converted into non-polar and water-soluble conjugated products. This detoxification process is mediated by the GST-catalysed GSH conjugation reaction. Taken from McElwee *et al.,* 2007 [28]. (B) The reaction scheme E represents GST and R represents the electrophilic group that binds to the H-site. Adapted from Armstrong, 1997 [20].

Additionally, in hGSTP1-1, the active site tyrosine may lower the pKa of the GSH thiol group by up to 2.4 pH units [29]. At this point, a glutathionyl-GST complex has formed, wherein the glutathionyl group is highly reactive. The hydrophobic substrate then binds to the H-site. Through the nucleophilic addition of the thiolate group to the hydrophobic substrate, the leaving group is eliminated from the hydrophobic substrate. This results in the formation of the conjugated product, which is released from the enzyme.

1.3.4 Human hGSTA1-1

One of the primary subjects of inquiry of this study, as it pertains to the Alpha class, is human GSTA1-1 (hGSTA1-1) (Figure 5). The Alpha class of the GST family distinguishes itself with the presence of an additional alpha helix, the α9 helix, in the C-terminal domain [30]. To date, five distinct subunit types have been identified. The α9 helix comprises residues 210-221. The hGSTA1-1 protein comprises 221 amino acids per subunit and has an molecular weight (MW) of 51.26 kDa. The N-terminal domain comprises residues 3-83, while the C-terminal domain comprises residues 85-207. The thioredoxin-like domain in hGSTA1-1 does not deviate from other homologous domains and follows the highly conserved 4 β-sheets (β1-β4) and three α-helices (α1-α4) motif [31]. The C-terminal domain is larger than the N-terminal domain, and is composed of six α-helices, one more than other GST classes.

With regard to the dimer interface molecular interactions, insights can be gained from the crystal structure of hGSTA1-1 (PDB: 1GUH) [32]. Using crystal structure data, it was determined that upon dimerisation, there are five key residues that become buried in the dimer interface. These residues were M50, F51, I85, A89, and M93. Further analysis showed that F51 was the only residue to be completely buried in the dimer interface [33]. The residues that closely followed in terms of the degree to which they were buried were A89 and M93. Moreover, the subunit interface itself is curved and its environment is hydrophobic. Importantly, the α9 helix, residing at the H-site, is positioned between the N- and C-terminal domains, specifically between the α2 and α4 helices [32]. The α2 helix, which comprises residues 38 to 47, interacts with helix α9 via tertiary packing [34]. Additionally, a lock-and-key motif contributes to the formation of the hGSTA1-1 dimer. The lock-and-key motif is highly conserved in the Alpha, Mu, and Pi classes [33].

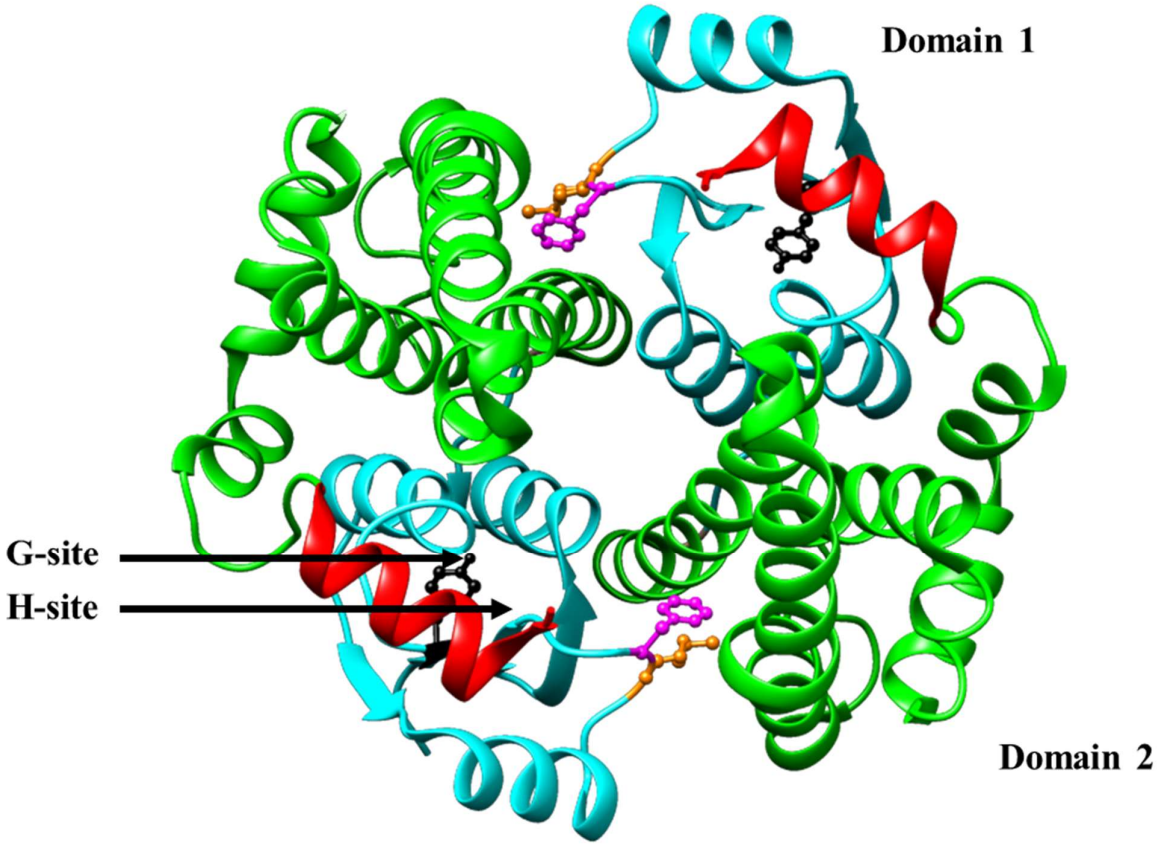


Figure 5. Ribbon structure of hGSTA1-1 in its homodimeric conformation (PDB: 1K3Y). The N-terminal domain is depicted in cyan, the C-terminal domain is depicted in green, the characteristic additional alpha helix is depicted in red, and the inter-domain linker regions are depicted in pink. The key residues are all depicted in ball-and-stick representation. Y9 is depicted in black, M50 is depicted in orange, and F51 is depicted in magenta. M50 and F51 play a crucial role in the hydrophobic lock-and-key motif. Generated using UCSF Chimera developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco [54].

The hGSTA1-1 lock-and-key motif comprises the M50 and F51 residues, which are located on the loop between the second α-helix and the third β-strand of the first domain [32]. The aforementioned residues are lodged between the α4 and α5 helices. The F51 residue of one subunit protrudes into the hydrophobic core of the C-terminal domain of the other subunit. This protrusion forms a hydrophobic interaction with the hydrophobic pocket created by the α4 and α5 helices of the other subunit [34]. As for the composition of the G-site, it is formed, in part, by the α2 helix and the previously mentioned loop. This portion of the G-site enables it to form hydrogen bonds with GSH [34].

1.3.5 Human hGSTM1-1

With regards to the structure of the Mu class, the details shall be recounted here. In the case of the present study, human GSTM1-1 (hGSTM1-1) is the protein chosen from the Mu class (Figure 6). On the front of the number of Mu subunits, six distinct subunits have been identified to date, with the sixth subunit lacking a complete sequencing of its primary structure [35]. Residues 33-43 constitute the Mu loop. The hGSTM1-1 protein is composed of 217 amino acids per subunit and has a molecular weight of 51.42 kDa. The mu loop, which is located between beta strand-2 and alpha helix-3, is unique to the Mu GST class. The N-terminal domain is the thioredoxin-like domain (residues 3-79) [36]. The C-terminal domain is the alpha helical domain, so named due to its abundance of alpha helices (residues 87-207) [36].

The highly conserved thioredoxin-like domain is structurally characterised by the presence of four anti-parallel β-sheets and 3 α-helices. The C-terminal domain, which is hydrophobic in nature is structurally characterised by five α-helices. The Mu class proteins exhibit unique molecular interactions when monomeric members of that class dimerise. As would be expected, the nature of the interactions between subunits of different GST classes differs quite considerably, likely due to the fact that subunits are limited to oligomerisation with subunits of the same GST class. A mixed charge cluster is the most predominant force (after hydrophobic interactions) found at the dimer interface of Mu GSTs [37]. On a monomeric basis, there are no observable charge clusters. When the subunits dimerise, however, there is the emergence of interdomain salt-bridges between interacting residues at the dimer interface that suggest that these salt bridges are integral to the formation of the mixed charge cluster, which, in turn, is integral to the formation of the dimer [37].

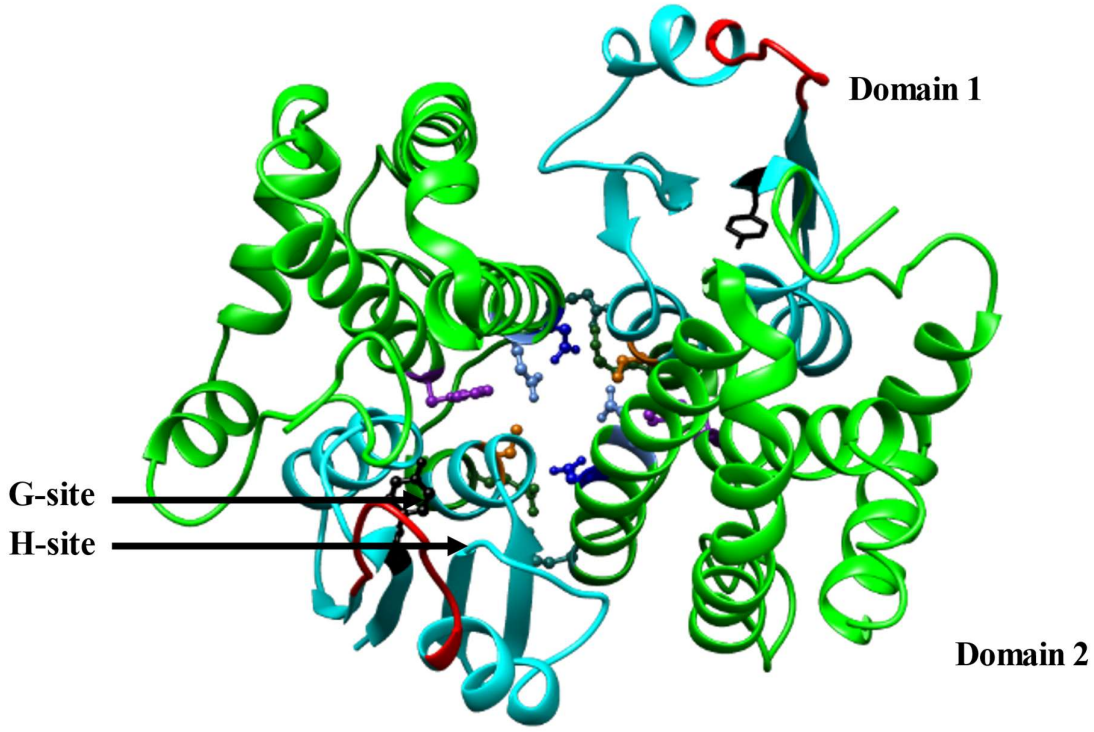


Figure 6. Ribbon structure of hGSTM1-1 in its homodimeric form (PDB: 1GTU). The N-terminal is depicted in cyan, the C-terminal is depicted in green, the distinctive Mu loop is depicted in red, and the inter-domain linker regions are depicted in pink. The key residues are all depicted in ball-and-stick representation. Y6 is depicted in black, C77 is depicted in orange, R81 is depicted in forest green, E90 is depicted in light sea green, D97 is depicted in blue, E100 is depicted in cornflower blue, and F154 is depicted in purple. These residues form the charge cluster that is integral to maintaining the dimer interface interaction. Generated using UCSF Chimera developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco [54].

The mixed charge cluster of hGSTM1-1 consists of six residues. These residues are C77, R81, E90, D97, E100, and F154. The R81 residue of one subunit forms salt bridges with the E90 and D97 residues of the other subunit [37]. Furthermore, there are intra-subunit hydrogen bond formations between E100 and F154 in both subunits [37].

1.4 Evolution

Thought to be evolutionary descendants of glutaredoxin and thioredoxin, the GST family emerged through divergent evolution. As is characteristic of divergent evolution, similar features in the antecedents and descendants are due to homology. This is well evidenced in the case of GSTs. For example, it has already been documented that the N-terminal domain is termed the thioredoxin-like domain for its marked similarity to thioredoxin. With GSTs being thought to have evolved from glutaredoxin and thioredoxin, the diversity within the GST superfamily itself is attributed to further divergent evolution. The prevailing theory on the mechanism of evolution in GSTs is gene amplification and divergence, which led to a wide array of classes within the superfamily [38]. This process is thought to have endowed the respective GSTs with diversified catalytic functions [39].

The mitochondrial, and soluble, Kappa GST class is proposed to be the progenitor of the Theta class. In turn, the cytosolic GSTs Alpha, Mu, and Pi are thought to be descendants of the Theta class, based on the complementary DNA analysis of these classes. A point of evolutionary departure may be marked at the active site of the cytosolic GSTs. The active site differences are used to further cement the fact that the Theta class is an antecedent to the Alpha, Mu, and Pi classes. This is because it has been demonstrated that Theta GSTs are S-GSTs, whereas the latter GSTs are Y-GSTs [21].

A remarkable phenomenon seen in the GST family is that despite the low sequence identity between classes (as low as 25%), the three-dimensional fold of all GSTs is astonishingly similar [25]. Not excepting the presence of unique structural elements, such as the mu loop in Mu GSTs and an extra alpha helix, α9, in the alpha-helical domain of Alpha GSTs, the three-dimensional fold is well conserved. As proteins evolve, they sample amino acid residues that provide stability while conferring a diverse functionality. Consequently, there is a trade-off between function and stability, where in the protein is on a knife-edge as it samples through the permutations of possible sequences and conformations.

1.5 Rationale

As proteins evolved divergently in order to diversify their function, the precarious trade-off between stability and function was a major force for directing mutations in a protein

sequence. The present study aimed to assess the role of conserved sequences across and within the GST family and classes, respectively. The present study also aimed to elucidate the magnitude of the trade-off between protein stability and function. The efficacy of consensus protein design was also assessed. This assessment gave insight into the possibility of engineering proteins of the same function that may be able to avoid misfolded states. GSTs belong to a supergene family that is ubiquitous and has shown variation within and between its classes. The main reason for this study was to assess the potential impact that these observed differences would have on the structure, function, and stability of the GSTs.

1.6 Aim and objectives

1.6.1 Aim

The aim of the present study was to assess the cumulative effect of the combination of the consensus mutations of the *de novo* consensus GST proteins on the structure, function, and stability of their wildtype counterparts. Additionally, the aim encompassed the question as to whether the conservation of residues favoured protein stability or protein function.

1.6.2 Objectives

1. Perform MSA to engineer consensus sequences
2. Produce and purify hGSTA1-1, hGSTM1-1, Alpha consensus mutant, and Mu consensus mutant in T7 *E. coli* cells
3. Evaluate the purity of the proteins using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)
4. Assess the secondary structural composition of the respective proteins using far-UV circular dichroism (far-UV CD)
5. Assess the tertiary structure of the respective proteins using ANS binding studies via fluorescence spectroscopy
6. Assess the enzyme specific activity of the respective proteins using the GSH-CDNB conjugation assay
7. Obtain the Michaelis-Menten parameters for the respective proteins using the GSH-CDNB conjugation assay
8. Assess the thermal stability of the proteins relative to each other using a thermal shift assay with the SYPRO® orange dye
9. Assess the impact of the consensus mutations on the dynamics of the proteins using molecular dynamics simulations (MDS)

Chapter 2: Methods and Materials

2.1 Engineering of consensus proteins

The Alpha and Mu classes were chosen as the classes for which the intra-class consensus sequences would be generated. To generate the cytosolic GST mutant protein, the inter-class consensus sequence, an MSA was conducted on the Alpha, Mu, and Pi classes. Cytosolic GSTs are characterised by GSTs that are in the cytosol of the cell. The Alpha, Mu, and Pi classes were chosen due to their prevalence in the pre-existing literature. There was no consensus protein design for the Pi class, as there is only a single protein within the Pi class. The Alpha consensus sequence was generated by performing an MSA of the entire Alpha class, i.e., A1, A2, A3, A4, and A5. The Mu consensus sequence was generated by performing an MSA of the entire Mu class, i.e., M1, M2, M3, M4, and M5. The cytosolic consensus sequence was generated by performing an MSA of the entire Alpha, Mu, and Pi classes (Figures 7, 8, and 9). All MSAs were done using Clustal Omega v1.2.4 [40]. This was followed by reverse translation of the consensus sequences into DNA sequences. These sequences were then codon optimised and designed for heterologous expression. Codon optimisation is a strategy employed to increase the efficiency of heterologous protein production. To achieve codon optimisation for the designed sequences, a proprietary algorithm was used by GenScript. The algorithm considers several factors such as codon usage, GC-content, mRNA secondary structure of the genes, cis-acting mRNA destabilising motifs, RNase splicing sites, and repetitive elements. The above-mentioned factors are optimised for protein expression in a bacterial host.

2.2 Protein production and purification

2.2.1 Transformation

Once the consensus sequences had been designed, the hGSTA1-1, Alpha consensus mutant, hGSTM1-1, Mu consensus mutant, and the cytosolic consensus mutant sequences were all sent for plasmid synthesis, ultimately each being inserted into a separate pET-11a vector (GenScript Biotech, United Kingdom). A transformation protocol was then followed, beginning with pipetting six µl of the vector, which was at a concentration of 100 $ng.\mu l^{-1}$, into 100 µl of competent T7 *Escherichia coli* (*E. coli*) express cells. The cells were then left to chill on ice for a period of 15 minutes, after which they were subjected to a heat shock. The cells were placed in a heating block at 42 °C for 45 s. This allowed for the membrane to become porous and enable the uptake of plasmid DNA. The cells were immediately placed on ice and left to chill for 2 minutes.



Figure 7. MSA used to generate consensus sequences. The parental sequences are arranged in descending order of their percentage sequence homology with the generated consensus sequence. All five members of the Alpha class were aligned to engineer the Alpha consensus sequence. MSAs generated using Clustal Omega v1.2.4.

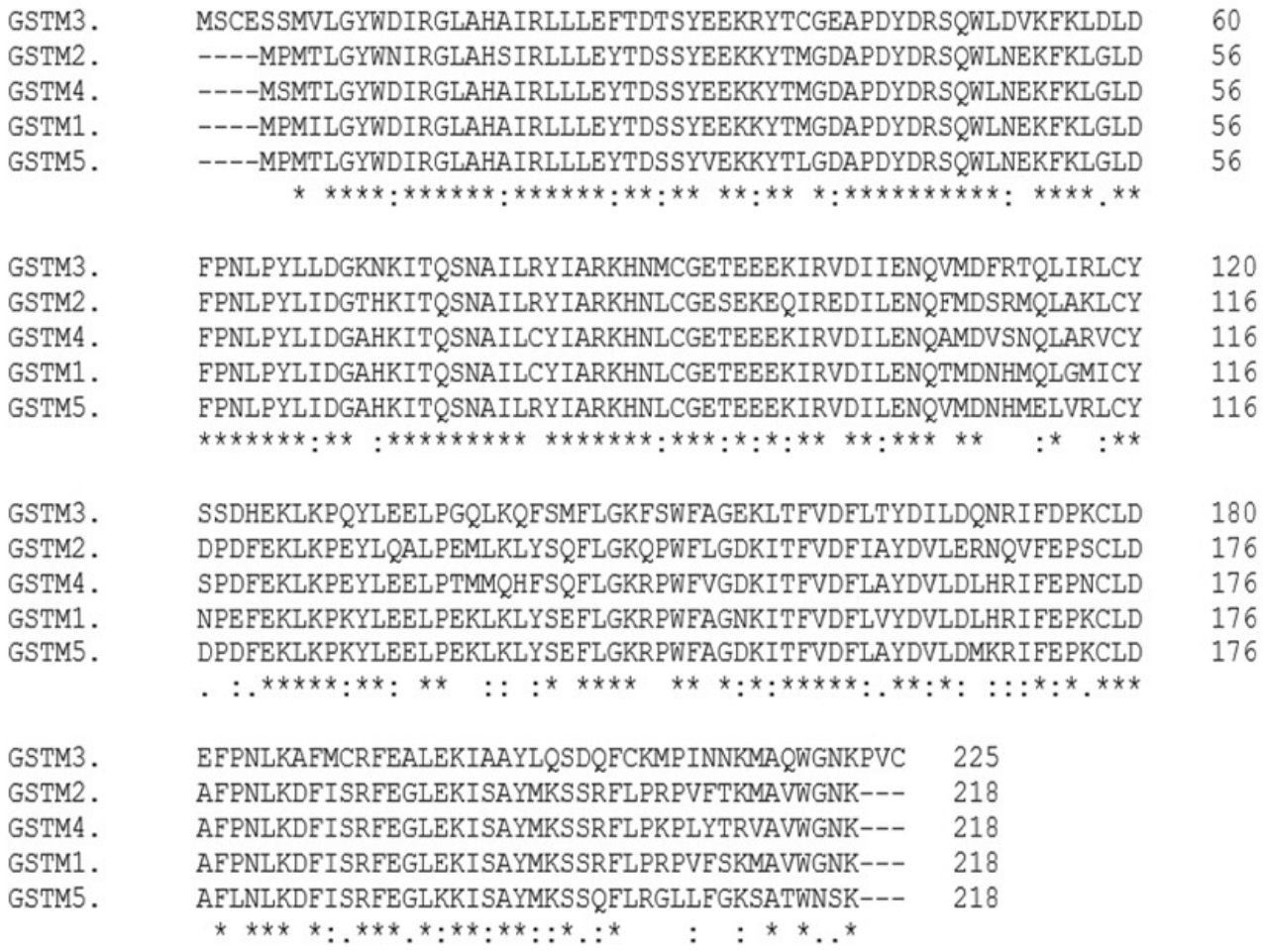


Figure 8. MSA used to generate the Mu consensus sequence. The parental sequences are arranged in descending order of their percentage sequence homology with the generated consensus sequence. All five members of the Mu class were aligned to engineer the Mu consensus sequence. MSAs generated using Clustal Omega v1.2.4.

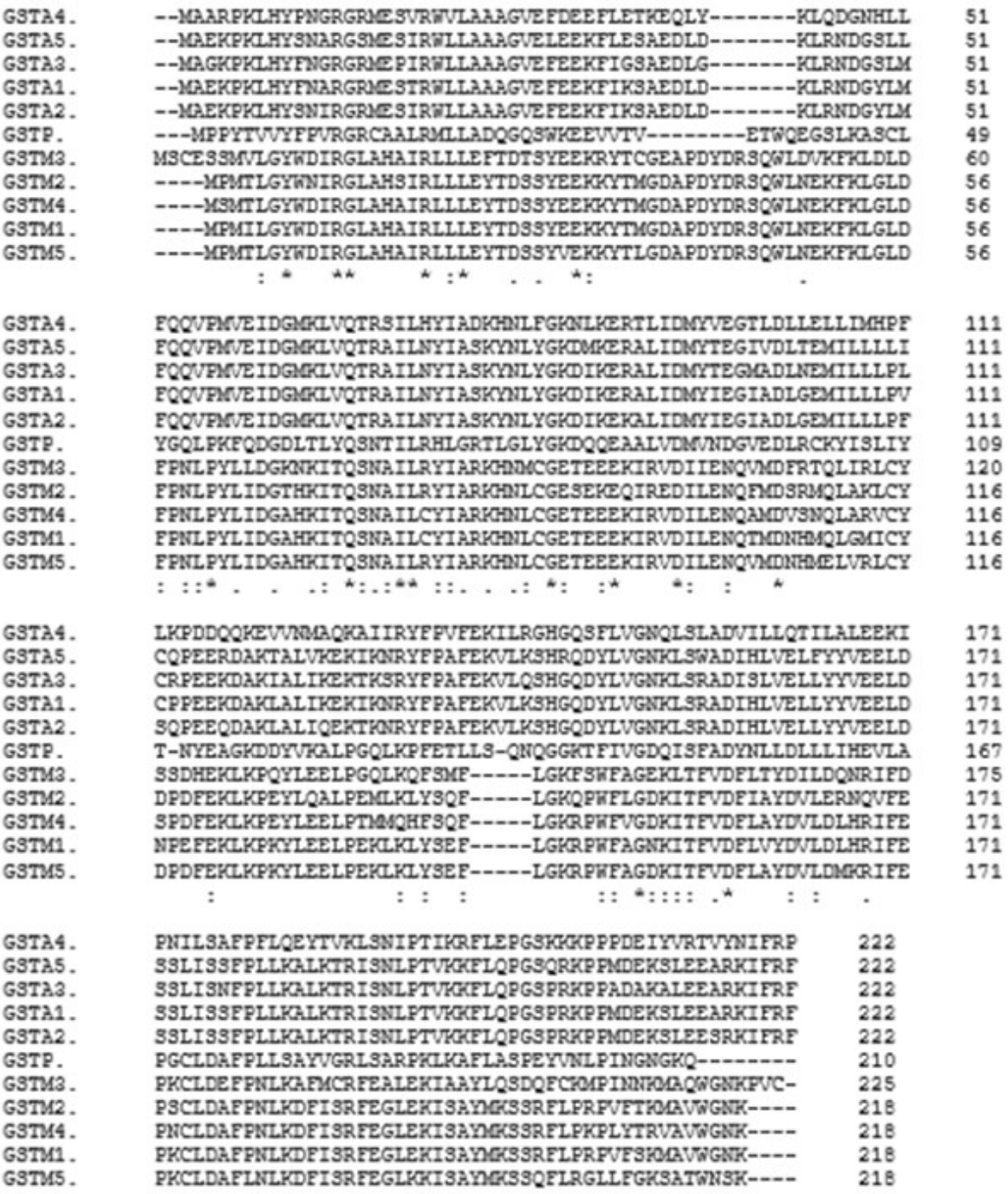


Figure 9. MSA used to generate the interclass consensus sequence. The parental sequences are arranged in descending order of their percentage sequence homology with the generated consensus sequence. All members of the Alpha, Mu, and Pi classes were aligned to engineer the cytosolic consensus sequence. The Pi class consists of only one protein. MSAs generated using Clustal Omega v1.2.4.

The vector conferred ampicillin resistance to the transformed cells. The cells were then transferred to 1 ml of super optimal broth with catabolite repression (SOC) medium (Thermo Fisher Scientific™, Massachusetts, USA). This mixture was then incubated for 1 h at 37 °C, shaking at a speed of 250 rpm. After the 1 h had elapsed, spread plating of the cells on agar plates containing 100 $\mu g.ml^{-1}$ of ampicillin was performed. For each protein, five agar plates were prepared, one of which was a negative control which did not have any cells spread across it, to ensure that the agar was sterile. The cells were then spread across four plates, ensuring a dilution of cells. This would produce lawns with separate colonies. The plates were incubated overnight at 37 °C. Single colonies were selected and 100 ml of 2× YT media (1.6% (w/v) tryptone, 1% (w/v) yeast extract, and 0.5% (w/v) NaCl) was inoculated. The inoculated media was incubated overnight at 37 °C and 250 rpm.

2.2.2 Production of recombinant proteins

The protein production protocol presented here was adapted from Pooe *et al.,* 2022, with minor alterations [41]. After the cell culture had grown overnight and reached appropriate turbidity, a 1:50 dilution was performed, thus inoculating 600 ml of fresh 2× YT media in 3 L flasks. The media contained 100 $\mu g.ml^{-1}$ ampicillin. The flasks were then placed in New Brunswick Scientific incubators and were incubated at 250 rpm and 37 °C until an optical density at 600 nm (OD_{600}) of 0.5 was reached. At an OD_{600} of 0.5, after a 1:50 dilution of the overnight culture was done, the cells in the fresh culture would be at the mid-log growth phase. At the mid-log phase, cell growth was exponential and consistent, making it suitable for the production of recombinant proteins. Once an OD_{600} of 0.5 was reached, the flasks were incubated on ice for 30 minutes to induce the production of cold shock proteins, which have been shown to improve the yield of protein in the soluble fraction. The pET-11a vector, which is under the control of the *T7* promoter, contains the *lacI* repressor gene which expresses a repressor protein. The *lacI* repressor protein inhibits the expression of T7 RNA polymerase. The repressor protein also occupies the *lac* operon. These factors therefore prevented the heterologous production of the proteins. After the flasks had been incubated on ice, 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to each flask to induce heterologous production. IPTG structurally mimics allolactose, thereby inhibiting the repressor protein and enabling the T7 RNA polymerase to bind to the *lac* operon. The flasks were placed back in the incubators and were cultured for 6 h at 30 °C and 250 rpm. Once the time had elapsed, the cells were harvested via centrifugation using a Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fischer Scientific™, Massachusetts, USA). The cells were

spun down for 15 minutes at 4 °C and 5000 x g. Subsequently, the spent media was discarded and the pellet was resuspended in 20 ml of sodium phosphate buffer. The crude whole cell extract was then stored at -20 °C.

2.2.3 IMAC purification

An integral part of the protein purification process is chromatography. Chromatography, broadly speaking, enables one to isolate a target protein with remarkable resolving power. In this study, immobilised affinity chromatography was selected as the technique of choice. Affinity chromatography is fundamentally a technique about exploiting the strong interactions between specific macromolecules in the liquid phase and an immobilised ligand in the stationary phase. In this case, nickel (Ni^{2+}) affinity chromatography was employed. The stationary phase initially consists of nitrilotriacetic acid (NTA) which acts as a chelator. When the column is charged, the NTA sequesters the Ni^{2+} ions onto the stationary phase. The pET-11a vectors contained a hexa-histidine tag sequence at the N-terminal of the protein sequence, resulting in the expression of an uncleavable hexa-histidine tag at the N-terminal of the protein. The hexa-histidine tag enabled the target protein to interact with the immobilised Ni^{2+} ions, thereby allowing contaminants to pass through the column freely.

Prior to IMAC purification, the frozen crude whole cell extract was thawed in room temperature water for 30 min. After thawing, the cell extract was subjected to ten rounds of sonication using a Microson XL-2000® sonicator. The cells were sonicated on ice at an amplitude percentage of 70, with one round of sonication consisting of one minute of sonication and one minute of chilling on ice. Once the viscosity of the lysate was significantly lowered, it was centrifuged at 18 000 x g for 25 min at 4 °C in the Sorvall LYNX 4000 Superspeed Centrifuge (Thermo Fischer Scientific™, Massachusetts, USA). The supernatant was then filtered through a 0.45 µm syringe filter with a cellulose acetate membrane.

For the purification proper, a 5 ml HiTrap® IMAC fast flow column (Sigma Aldrich, Missouri, USA) was used. Nickel sulfate ($NiSO_4$) (100 mM) was used to pack the resin of the column with Ni^{2+} ions. The chromatography system was first washed with one column volume of 0.01% (w/v) sodium azide. Subsequently, the column was connected to the AktaStart™ chromatography system (General Electric Healthcare Life Sciences, Chicago, USA). The flow rate was maintained at 5 $ml.min^{-1}$ and 0.3 kPa throughout the purification process. The column was then equilibrated with five column volumes of equilibration buffer (25 mM imidazole, 1× phosphate buffered saline (PBS), 0.01% (w/v) sodium azide, pH 7.2).

The filtered supernatant was then loaded onto the column and allowed to incubate for 30 min. Subsequently, wash buffer (25 mM imidazole, 1× PBS, 0.05% Tween® 20, 0.01% (w/v) sodium azide, pH 7.2) wash passed through the column. A one-step elution then ensued, wherein the elution buffer (500 mM imidazole, 1× PBS, 0.01% (w/v) sodium azide, pH 7.2) was passed through the column. The eluate was collected in 2 ml fractions until the absorbance units at 280 nm declined to a steady baseline.

2.2.4 Dialysis

Once the fractions from the elution were collected, the collective protein sample was placed into cellulose acetate dialysis tubing with an MW cut-off of 14 kDa (Sigma Aldrich, Missouri, USA). The protein was then dialysed against 2 L of sample buffer for 4 h, followed by another 2 L of sample buffer overnight. Finally, the protein was then dialysed against 1 L of dialysis buffer for 4 h. The same protocol was followed whenever buffer exchange was required in order to perform various downstream experiments.

2.3 Purity analysis via SDS-PAGE

Once purification has been completed, it is important to qualitatively assess the purity of the eluted fractions that have been designated as the fractions carrying the target protein. SDS-PAGE is a technique that enables this assessment. Additionally, SDS-PAGE allows for a quantitative analysis of the purified protein. Its fundamental principle is that proteins are denatured by SDS and are endowed with a negative charge by SDS. Due to this, proteins migrate through a gel based on their molecular weight, and their position in their gel enables one to ascertain whether the target protein has indeed been expressed and purified at its correct molecular weight [42]. Glycine SDS-PAGE was the method selected for resolving and analysing the target proteins of the study.

The SDS-PAGE experiments consisted of 2 gels, a 12.5% (w/v) separating gel, upon which a 4% (w/v) stacking gel was cast. Stock solutions of monomer solution (30% (w/v) acrylamide, 0.8% (w/v) bis-acrylamide), 10% (w/v) SDS, 1.5 M Tris (pH 8.8), 500 mM Tris (pH 6.8), and 10% (w/v) ammonium persulfate (APS) were prepared.

The separating gel consisted of 41.6% (v/v) monomer solution, 26.6% (v/v) stock Tris (1.5 M, pH 8.8), 1% (v/v) stock SDS, 0.7% (v/v) stock APS, and 0.2% (v/v) tetramethyl-ethylenediamine. The stacking gel was composed of 13% (w/v) monomer solution, 20% (v/v) stock Tris (500 mM, pH 6.8), 1% (v/v) stock SDS, 0.7% (v/v) stock APS, and 0.2% (v/v) tetramethyl-ethylenediamine.

Prior to loading the samples into the gel, the samples were placed in reducing sample buffer which contained Tris-HCl (125 mM, pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 10% (v/v) β-mercaptoethanol, and bromophenol blue. At a ratio of 1:1, the samples were mixed with the reducing sample buffer, after which they were placed in boiling water for five minutes. Samples that were predicted to be high in viscosity were briefly sonicated to ensure that loading would not be hindered. The gel was then placed into a tank containing the gel electrophoresis buffer (250 mM Tris-HCl, 192 mM glycine, 0.1% (w/v) SDS, pH 8.3). Loaded into the gel was, firstly, two µl of molecular weight marker. Subsequently, six µl of the samples were loaded into the subsequent wells. Electrophoresis was then carried out using the Mini-PROTEAN® Tetra vertical electrophoresis cell system (Bio-rad, California, USA). Once electrophoresis was complete, the gel was placed into staining solution for two hours. The staining solution contained 0.25% (w/v) Coomassie Brilliant Blue R250, 40% (v/v) methanol, and 10% (v/v) glacial acetic acid. After staining, the gel was placed in destaining solution (40% (v/v) methanol and 10% (v/v) glacial acetic acid) overnight, after which gel imaging occurred.

2.4 Spectrophotometric determination of protein concentration and purity

The quantification of protein concentration is a vital step for downstream experiments, and it is made possible through the use of ultraviolet-visible (UV-Vis) spectrophotometry. UV-Vis spectrophotometry is a technique that entails light that is in the UV-Vis range being emitted and subsequently absorbed by a given substance. When light is absorbed, electrons in the substance are shifted from a ground state to an excited state. Due to the fact that proteins are organic compounds, they possess specific functional groups that absorb light in the UV-Vis range. These functional groups are called chromophores. In addition to the use of spectrophotometry, the Beer-Lambert Law is complementarily used to obtain the protein concentration:

$$A = \varepsilon c l \qquad (1)$$

where A is absorbance in arbitrary units (AU), ε is the extinction coefficient ($M^{-1}.cm^{-1}$), c is the concentration (M), and l is the path length of the cuvette traversed by the light (cm). Thus, chromophores are doubly important because they are able to absorb light in the UV-Vis range and contribute to the ε term in the Beer-Lambert equation. In order to calculate the ε of proteins, the individual ε for all of the relevant chromophores are taken into account in the following equation, as per Perkins, 1986 [43]:

$$\varepsilon_{280}\ (\text{hGSTA1-1}) = 5550\ \Sigma\text{Trp} + 1340\ \Sigma\text{Tyr} + 150\ \Sigma\text{Cys} \quad (2)$$

The absorbance measurements for each protein were obtained using the JASCO V-630 UV-Vis absorbance spectrophotometer (Jasco, Pfungstadt, Germany). The requisite dilutions were obtained by mixing 970 µl of sample buffer with 30 µl of protein. A doubling dilution series was then performed, resulting in five samples in the dilution series. The samples were then pipetted into a Suprasil® quartz cuvette with a 10 mm light path (Hellma®, Müllheim, Germany). Readings were taken at 280 nm and 340 nm. Buffer and light scattering contributions were then accounted for by subtracting the 340 nm readings from the 280 nm readings, thereby producing corrected readings. After the fixed wavelength measurements had been performed, a UV scan from 240 nm to 340 nm was recorded for each protein.

2.5 Structural characterisation

2.5.1 Far-UV CD

Far-UV CD is a technique employed for the purposes of determining the predominating secondary structural elements of a protein. The basic premise of the technique is that when circularly polarised light interacts with a chromophore, a unique spectrum is produced. The inner workings of the premise are that circularly polarised light is divided into two vectors which rotate in opposite directions; namely, clockwise and anticlockwise. When the vectorised light interacts with the chromophore, the two vectors of light are discriminately absorbed. After a buffer exchange, the protein was in far-UV CD buffer (10 mM sodium phosphate, pH 7.0), at a dimeric concentration of 4 µM. The scan speed was set to 100 nm.min^{-1}, the data pitch set at 0.5 nm, with the temperature set to 20 °C, and the bandwidth set to 1 nm. The spectra were measured from 180-240 nm, using a Suprasil® quartz cuvette with a 0.2 cm light path (Hellma®, Müllheim, Germany) and a Jasco J-1500 spectropolarimeter (Jasco, Pfungstadt, Germany). The experiments were performed in triplicate. Once the raw data were collected, they were converted to molar residue ellipticity (MRE) using the following equation, as per Woody, 1995 [44]:

$$MRE = \frac{(100)(\theta)}{cnl} \quad (3)$$

where θ = ellipticity raw data in millidegrees (Mdeg), c = protein concentration in molar (M), n = number of amino acids, and l = path length of the cuvette (cm). Buffer contributions were accounted for.

2.5.2 ANS binding studies

Examining the global tertiary structure of a protein is possible using fluorescence spectroscopy. Intrinsic tryptophan fluorescence is often the default method for examining the structural changes a protein undergoes in various molecular environments. Another technique, which is just as probing, is extrinsic fluorescence. While intrinsic tryptophan fluorescence relies on the fluorescence spectra of a tryptophan residue, extrinsic fluorescence utilises an environmentally sensitive dye to obtain information about the conformation of a protein in a specific molecular milieu. The most commonly used extrinsic fluorescence dye is 8-anilinonaphthalene-1-sulfonic acid (ANS). ANS binds to hydrophobic surfaces and when bound, it produces a spectrum that exhibits higher quantum yields and blue shifts. Its fluorescence is quenched by water and other polar solvents in general. This property is very useful especially in the context of GSTs, as they possess the G- and H-site. This means that if ANS is placed in the same environment as GSH, one can assess the stabilising effects induced by GSH binding by assessing the extent to which ANS fluoresces. Seeing as the H-site of a GST is ostensibly its ANS binding site, the degree of exposure of the H-site before and after GSH binding can be ascertained.

The protein was placed in a fluorescence buffer (50 mM Tris, 1 mM EDTA, 1 mM DTT, 0.01% (w/v) sodium azide, pH 7) using buffer exchange. A stock of 2 mM ANS was prepared. A final concentration of 200 μM of ANS was used in the ANS fluorescence experiments, which was in a mixture with 2 μM protein. For the samples that combined ANS, protein, and GSH, a final concentration of 1 mM GSH was used. Due to the light sensitivity of ANS, the mixtures were incubated in the dark at room temperature for 30 min to an 1 h. The samples were then placed in a Suprasil® quartz cuvette with a 1 cm light path (Hellma®, Müllheim, Germany) and the cuvette, in turn, was placed in a Jasco FP-6300 fluorimeter (Jasco, Pfungstadt, Germany). The excitation wavelength was set to 395 nm, and the spectra were measured from 400-650 nm. Additionally, the spectra were measured with a scan speed of 200 $nm.min^{-1}$, a data pitch of 1 nm, an excitation bandwidth of 2.5 nm, an emission bandwidth of 5 nm, and a temperature of 20 °C. The spectra were corrected for buffer, ANS, and GSH contributions appropriately.

2.6 Enzyme activity and Michaelis-Menten kinetics

Along with the structural aspects of the study, a functional assessment of the proteins was conducted. This took place in the form of enzyme kinetics. As previously stated, GSTs catalyse a conjugation reaction wherein GSH reacts with a hydrophobic substrate. For the

purposes of this study, 1-chloro-2,4-dinitrobenzene (CDNB) was used as the co-substrate. This reaction leads to the formation of the conjugate product 1-(*S*-glutathionyl)-2, 4-dinitrobenzene (GS-DNB) (ε = 9600 $M^{-1}.cm^{-1}$) [45]. Due to the thioether bond that is formed after conjugation, the product becomes a chromophore, and a chromogenic reaction can then be monitored at 340 nm. The conjugation reaction thus formed the basis for the specific activity and Michaelis-Menten kinetics parameters. Specific activity, measured in $\mu mol.min^{-1}.mg^{-1}$, is an accurate summation of the purity of a protein sample [46].

For a greater understanding of the interaction between an enzyme and its substrate, Michaelis-Menten kinetics are indispensable. The two main parameters obtained from Michaelis-Menten kinetics, the Michaelis constant (K_M) and maximal velocity (V_{max}), quantify the nature of the interactions between enzyme and substrate. K_M is defined as the substrate concentration required to reach 50% of V_{max}. V_{max}, in turn, is the highest rate at which the enzyme-catalysed reaction will proceed under *in vitro* assay conditions.

The specific activity experiments were conducted using the POLARstar Omega Microplate Reader (BMG Labtech, Ortenberg, Germany). A kinetics buffer (100 mM sodium phosphate, 1 mM EDTA, 1 mM DTT, 0.01% (w/v) sodium azide, pH 6.5) was used. The experiments were carried out at protein concentrations of 500 nM, 250 nM, 125 nM, 62.5 nM, and 31.25 nM. The final concentration of CDNB, after being dissolved by 100% (v/v) ethanol and being placed in the wells, was 1 mM, with the final concentration of ethanol being 2% (v/v). Due to CDNB being dissolved in ethanol, the reaction was initiated by 5 µM GSH, which was introduced into each well via 20 µl injections. This was done to prevent leakage of CDNB from the microplate syringe.

The Michaelis-Menten kinetics were conducted using the JASCO V-630 UV-Vis absorbance spectrophotometer (Jasco, Pfungstadt, Germany). The experiments were carried out in the aforementioned Tris buffer at 15 different GSH concentrations, while the concentration of CDNB remained constant at 1 mM. CDNB was used to initiate the reaction. Non-enzymatic reaction rates were accounted for by performing blank corrections to the enzyme-catalysed reactions. The experiments were carried out in triplicate.In both cases, the formation of GS-DNB was monitored at 340 nm over the course of 1 minute.

2.7 Thermal shift assay with SYPRO® orange dye

Thermal shift assays are reliable means that enable one to determine the thermal stability of proteins. Although equilibrium and reversibility are not established in the process of thermal unfolding, thermal shift assays produce a valuable parameter, the T_m. Founded on the basis that high temperatures cause the denaturation of proteins, thermal shift assays quantify the degree of denaturation at a given temperature, through the use of a fluorescent probe. This probe binds to hydrophobic patches. Naturally, as the protein is increasingly denatured, more hydrophobic patches are exposed, thus increasing the instances of binding of the probe to the protein. In turn, the probe increases its quantum yield, thereby increasing the detected fluorescence intensity. In this way, the fluorescence maxima are integral for the determination of the T_m. After a certain temperature, the proteins become extremely disordered and begin forming aggregates. This leads to the dissociation of the probes from the protein, and the fluorescent intensity then decreases.

A single 96-well qPCR plate was used. Each well contained a total of 25 µl, with the mixture being composed of 20 µl of 10 µM protein and 5 µl of SYPRO® orange dye. The assays consisted of a SYPRO® orange dye blank, a dye and GSH blank, protein and SYPRO® orange dye sample, and a protein with GSH and SYPRO® orange dye sample. The qPCR plate was then placed in a CFX96 Touch-Real-Time PCR detection system (Bio-Rad Laboratories, Inc., Hercules, California, USA). Each of the stated conditions were performed in triplicate for all of the proteins.

The thermal shift assays, as deployed in the current study, present several limitations. These limitations are due to the fact that reversibility, and therefore equilibrium, of protein unfolding cannot be established in most cases of thermal unfolding. Chemical denaturation utilises chaotropic agents such as urea and guanidinium chloride (GdnHCl) to achieve unfolding of a protein. Unfolding induced by these chaotropic agents establishes the two major criteria for thermodynamic studies which are reversibility and equilibrium [47]. When these criteria are established, key parameters such as ΔG, an indicator of protein stability in the absence of a denaturant (ΔG_{H2O}) and the degree to which the ΔG of protein unfolding is dependent upon the denaturant (m-value) are obtainable [47]. Chemically induced unfolding diminishes the ΔS_{solv} created by the hydrophobic effect. Essentially, the introduction of urea or GdnHCl makes it favourable for hydrophobic residues to be exposed to the solvent [47]. Without the two main criteria that have been stated, however, no conclusions regarding the thermodynamic properties of the concerned properties may be made. Thermal shift assays

are, therefore, mostly utilised for the purposes of obtaining information regarding the temperature at which thermal unfolding occurs in various conditions. This is adequate for the scope of this study.

2.8 Molecular dynamics simulations

Experimental work is integral to the scientific method. As much as it is key, however, results from wet lab experiments are often simply snapshots of a protein at any given moment. Performing MD simulations is, therefore, an integral part of the study of protein stability. This is because MD simulations represent the structural variations exhibited by a specific protein as a function of time or as a function of the number of residues in that protein. One can therefore compare the degree of variation, as well as other metrics, to qualitatively assess the stability of a protein. MD simulations on hGSTA1-1, hGSTM1-1, and their corresponding consensus mutants was done using the crystal structures 1K3Y (hGSTA1-1) and 7BEU (hGSTM1-1). These structures, obtained from the PDB, were of the highest resolution and they were solved to a resolution of 1.3 Å and 1.59 Å, respectively. The wildtype sequences were used for hGSTA1-1 and hGSTM1-1, while residue replacement was performed in Schrödinger Maestro v12.2 in order to generate the sequences of the consensus mutant proteins. Four systems were created, one for each protein. The construction of the systems involved solvating and ionising the protein using the Systems Builder model in Desmond. The protein model was placed in a water box. Counter ions were then added in order to neutralise the charge of the solvated system, with 36 chloride ions neutralising 36 sodium ions. A total concentration of 51.317 mM of NaCl was added to the system, in order to create physiological conditions within the system. The constructed system was then subjected to an equilibration phase which consisted of two steps. The first step was the energy minimisation of the systems using the NVT (moles (N), volume (V), and temperature (T)) ensemble. Once the systems reached a temperature of 300 K, the NPT (moles (N), pressure (P), and temperature (T)) ensemble was used to reach equilibrium. The systems were subsequently subjected to 500 ns of MD simulations.

Once the simulations had been done, data were obtained for several key parameters. The first of these parameters is the root-mean-squared-deviation (RMSD) of the alpha carbon (C_α). The second is the root-mean-squared-fluctuation (RMSF) of the C_α. The RMSD is indicative of the flexibility and movement of the global structure of the protein, while, in contrast, the RMSF is indicative of the fluctuation of a single amino acid residue over the course of an MD simulation run. Both the RMSD and RMSF simulations must converge before the end of

the simulation run in order to show that equilibration has been achieved. If equilibration has not been achieved, the results may be subject to slightly tenuous analysis. The RMSF results are also greatly influenced by the secondary structural elements of a protein. Proteins with ordered α-helices or β-strands show a lesser degree of fluctuation in their RMSF trajectories. The N- and C-terminals of proteins are also much more likely to be highly flexible.

The third metric is the radius of gyration (Rg), which is a metric that is indicative of the degree of compactness of a protein. Given the importance of the active site tyrosine (tyrosine 9 for the Alpha class and tyrosine 6 for the Mu class) the distance between the tyrosine residues of the two subunits was measured and the metric was abbreviated as Y9-Y9 and Y6-Y6 distance for the Alpha and Mu proteins, respectively. Site map analysis was conducted in order to observe the major cavities of the molecular surfaces of the proteins. The cavities with a Dscore equal to or greater than 0.75 were selected for analysis. This is because cavities that display a Dscore that is 0.75 or greater are excellent targets for drugabillity, indicating their importance in the functional and conformational status of a protein. Lastly, the poses of the wildtype structures were superimposed on the consensus mutants. This yielded an RMSD value of the complete structures. In sum, these metrics are important for assessing and determining the structural rigidity and compactness of a protein, which will, in turn, be highly informative of the flexibility and stability of the wildtype and consensus mutant proteins.

Chapter 3: Results

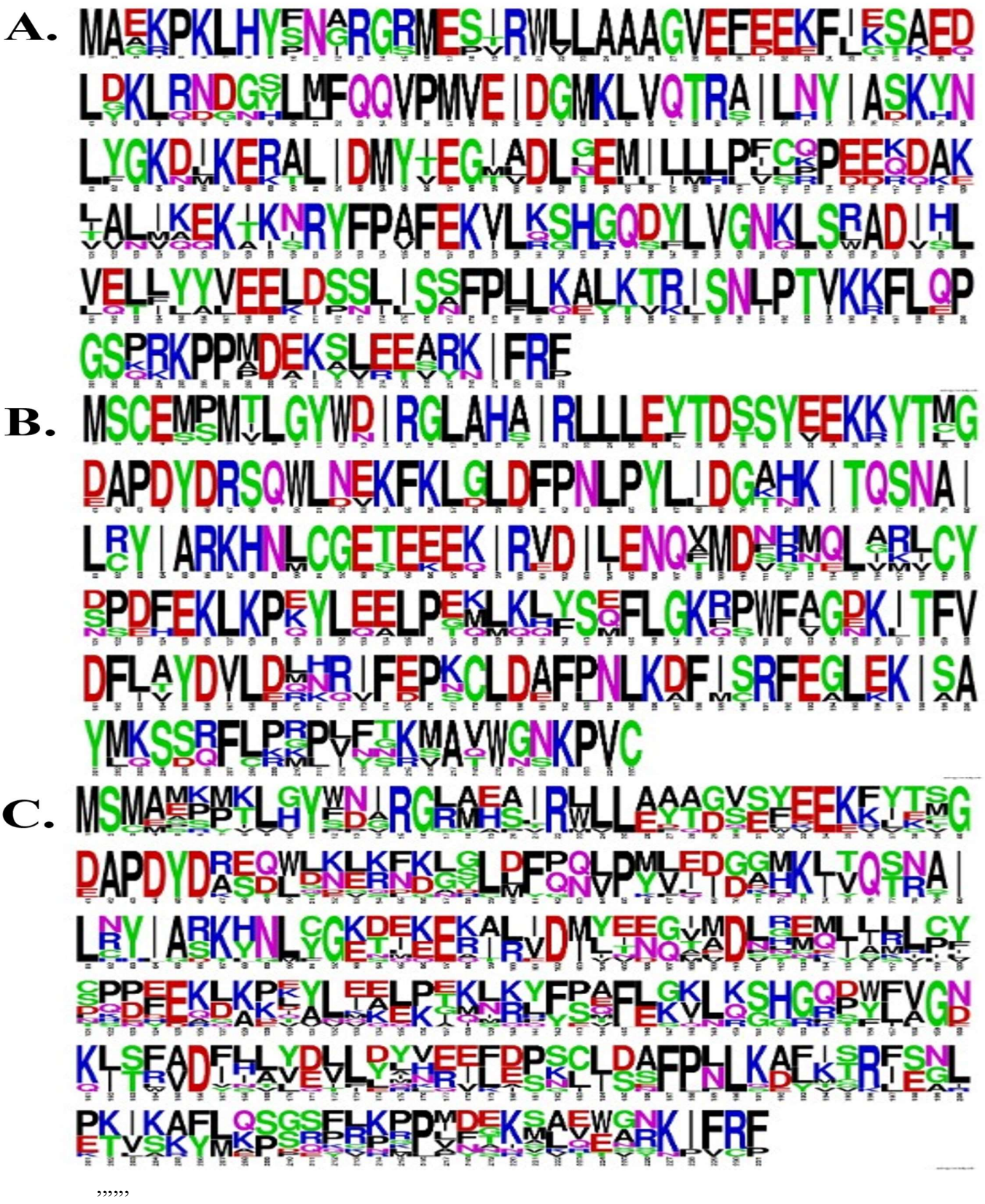
3.1 Engineering of consensus proteins

The consensus sequences of the Alpha, Mu, and Cytosolic classes were generated using consensus protein design in Clustal Omega v1.2.4. The majority of consensus mutations for both consensus mutant proteins occurred in the N-terminal thioredoxin-like domain. A key mutation in the Alpha consensus mutant protein, in comparison to hGSTA1-1, was the M50 residue being replaced by L50. The M50 residue is important because it is located within the G-site and has a considerable effect on the reduction of the dynamic behaviour of helix α2. There was a total of eleven consensus mutations in the Alpha consensus protein, in relation to hGSTA1-1, with five consensus mutations occurring in the N-terminal domain and six consensus mutations occurring in the C-terminal domain. The Alpha consensus mutations are all in relation to hGSTA1-1.

The total number of mutations in the Mu consensus protein was 13. There were ten mutations in the C-terminal domain of the Mu consensus mutant protein, and one in the N-terminal domain. WebLogo was used to generate the images in Figure 10 [48]. The Mu consensus mutations are all in relation to hGSTM1-1. Figure 10 visually represents the quantitative assessment of the conservation of residues. The parental proteins that were chosen, i.e., hGSTA1-1 and hGSTM1-1, were chosen on the basis of their low sequence homology in order to fully appreciate the stabilising role of conserved residues. They were also chosen because of the relative abundance of the literature regarding the parental proteins, therefore making analysis of the results more scientifically robust.

3.2 Purification

The proteins were purified using a Ni^{2+} IMAC one-step elution method. The filtered supernatant was loaded onto the column and subsequently passed through the column. This resulted in a major increase of the 280 nm absorbance readings to 2000 mAU. The Tween® 20 wash step soon followed, wherein the wash buffer was passed through the column. This step saw a marked increase of the absorbance readings at 280 nm to approximately 300 mAU, indicating that no significant amount of any protein was unbound and removed from the column. The equilibration buffer was used to remove the vestiges of detergent from the column and return the 280 nm absorbance readings to baseline. The elution of hGSTA1-1, Alpha consensus mutant, hGSTM1-1, and Mu consensus mutant produced a sharp peak of 2000 mAU, with a volume of approximately 10-15 ml per elution (Figures 11 and 12).



,,,,,,

Figure 10. Graphical representation of the consensus sequences generated after MSA. The height of each amino acid represented by a single letter indicates the frequency of that given residue in that position across members of the family or class. (A) The Alpha consensus protein sequence. (B) The Mu consensus protein sequence. (C) Cytosolic consensus protein sequence. Images were generated from weblogo.berkeley.edu.

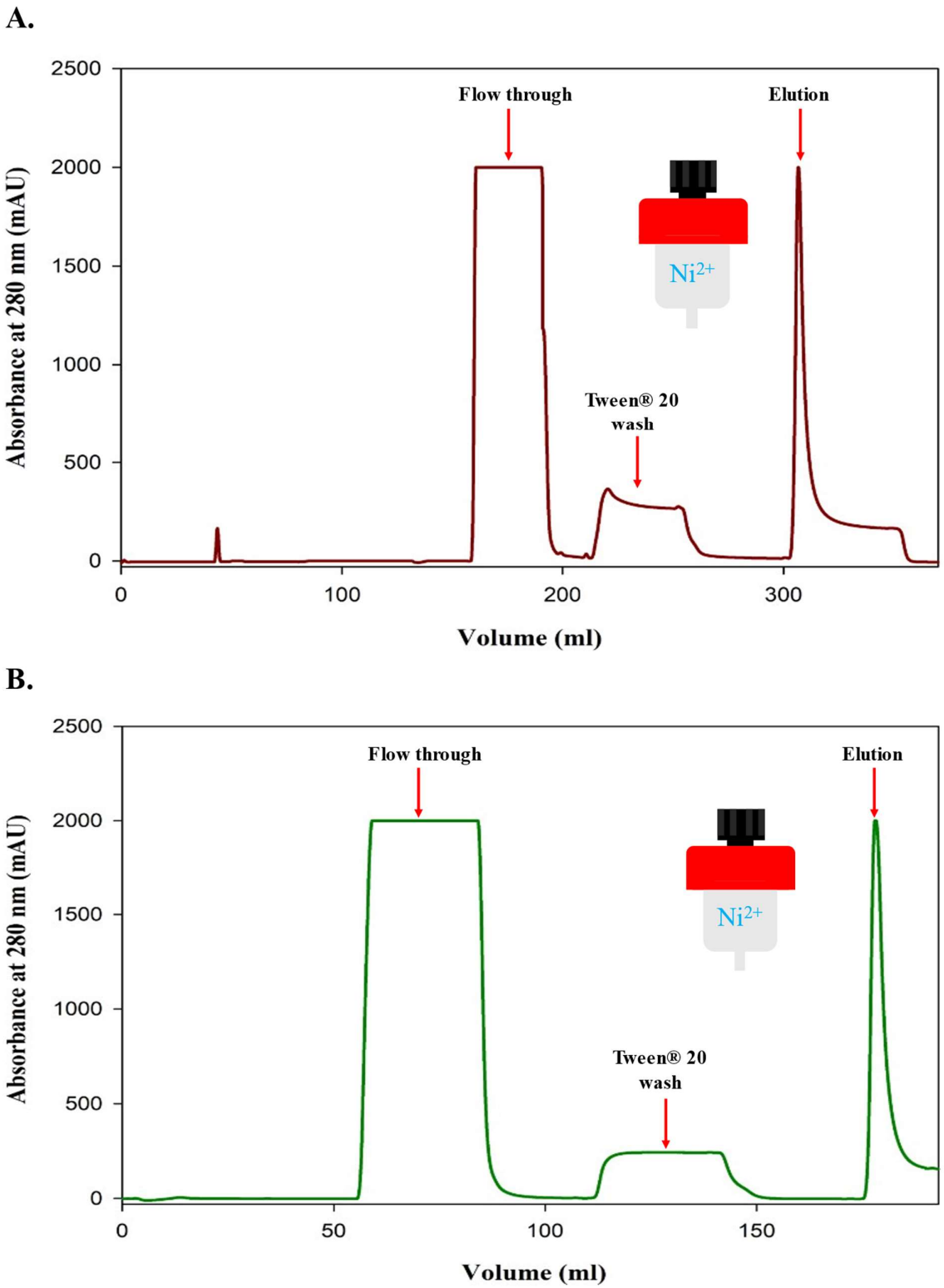


Figure 11. The elution profile of the Ni^{2+} IMAC purification of the Alpha wildtype and consensus mutant proteins. The filtered supernatant was loaded onto the column, thus inducing a peak of 2000 mAU labelled flow through. The binding step is indicated by the flow through. The flow through contained non-specific proteins and other contaminants. Equilibration buffer was passed through the column until the absorbance reached baseline. After incubation, the column was subjected a Tween® 20 wash. Subsequently equilibration buffer was used to allow the absorbance to reach baseline. The elution buffer was then passed through the column, inducing a sharp peak reaching 2000 mAU, and through a one-step elution, fractions of eluate were collected. (A) The elution profile of hGSTA1-1 purification. (B) The elution profile of the Alpha consensus mutant purification.

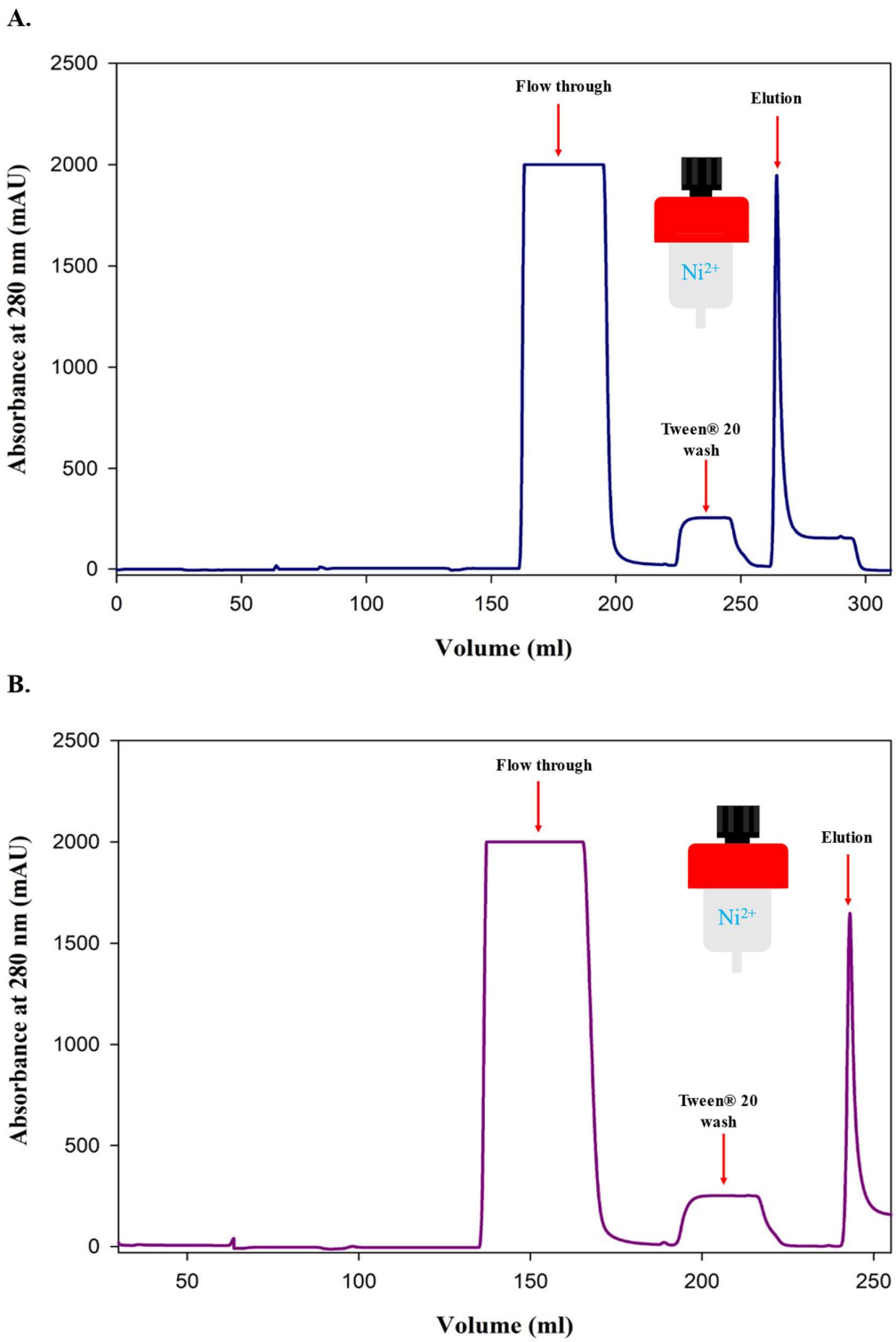


Figure 12. The elution profiles of the Ni^{2+} IMAC purification of the Mu wildtype and consensus mutant proteins. The filtered supernatant was loaded onto the column, thus inducing a peak of 2000 mAU labelled flow through. The binding step is indicated by the flow through. The flow through contained non-specific proteins and other contaminants. Equilibration buffer was passed through the column until the absorbance reached baseline. After incubation, the column was subjected a Tween® 20 wash. Subsequently equilibration buffer was used to allow the absorbance to reach baseline. The elution buffer was then passed through the column, inducing a sharp peak reaching 2000 mAU, and through a one-step elution, fractions of eluate were collected. (A) The elution profile of hGSTM1-1. (B) The elution profile of the Mu consensus mutant.

3.3 Purity analysis via SDS-PAGE

In the case of the consensus mutant proteins, a significant amount of the protein was produced in the insoluble fraction (Figure 13). However, a burdensome procedure of retrieving the protein in the insoluble fraction did not need to be followed. The yield of the protein in the supernatant was adequate for all the subsequent downstream applications. It was observed that the early fractions of eluate (the first and second eluate fractions), for all proteins, were lower in purity than fractions collected at a later stage. The yield of the protein fractions collected later in the elution stage, however, proved to be adequate for all downstream techniques. These impurities were more pronounced in cases where the supernatant did not incubate in the column for at least 30 min. The cytosolic consensus protein was determined to be insoluble. Therefore, downstream techniques were not performed for the cytosolic consensus protein.

3.4 Spectrophotometric determination of protein concentration and purity

After purification, a doubling dilution series of each protein was performed. The absorbance values of the diluted samples were then measured spectrophotometrically. Once standard curves were obtained, the Beer-Lambert Law was deployed to obtain the concentration of protein. Protein yields were consistently within the 30-85 μM range (dimeric concentration), in an average volume of 10-15 ml per purification. As shown Figure 14, the coefficient of determination (R^2), which is a measure of the goodness of fit of a given model, was greater than 0.99 in all cases, indicating a strong positive linear correlation. To assess the purity of the samples, a UV-Vis spectrum was performed (Figure 15), wherefrom insightful readings were obtained (Table 1). The A_{260}/A_{280} ratio was consistently lower than one, indicating that there was insignificant nucleic acid contamination. It is important to note that the A_{260}/A_{280} values were slightly skewed by the phenylalanine contributions to the absorbance at 260 nm, given that phenylalanine has an absorbance maximum at 257 nm. This is also evidenced in the spectrum itself, as there is a rise in absorbance in this region of the spectrum. The A_{340} readings of hGSTA1-1, hGSTM1-1, and the Alpha consensus mutant were of the requisite levels to alleviate any concerns of protein aggregation and precipitation.

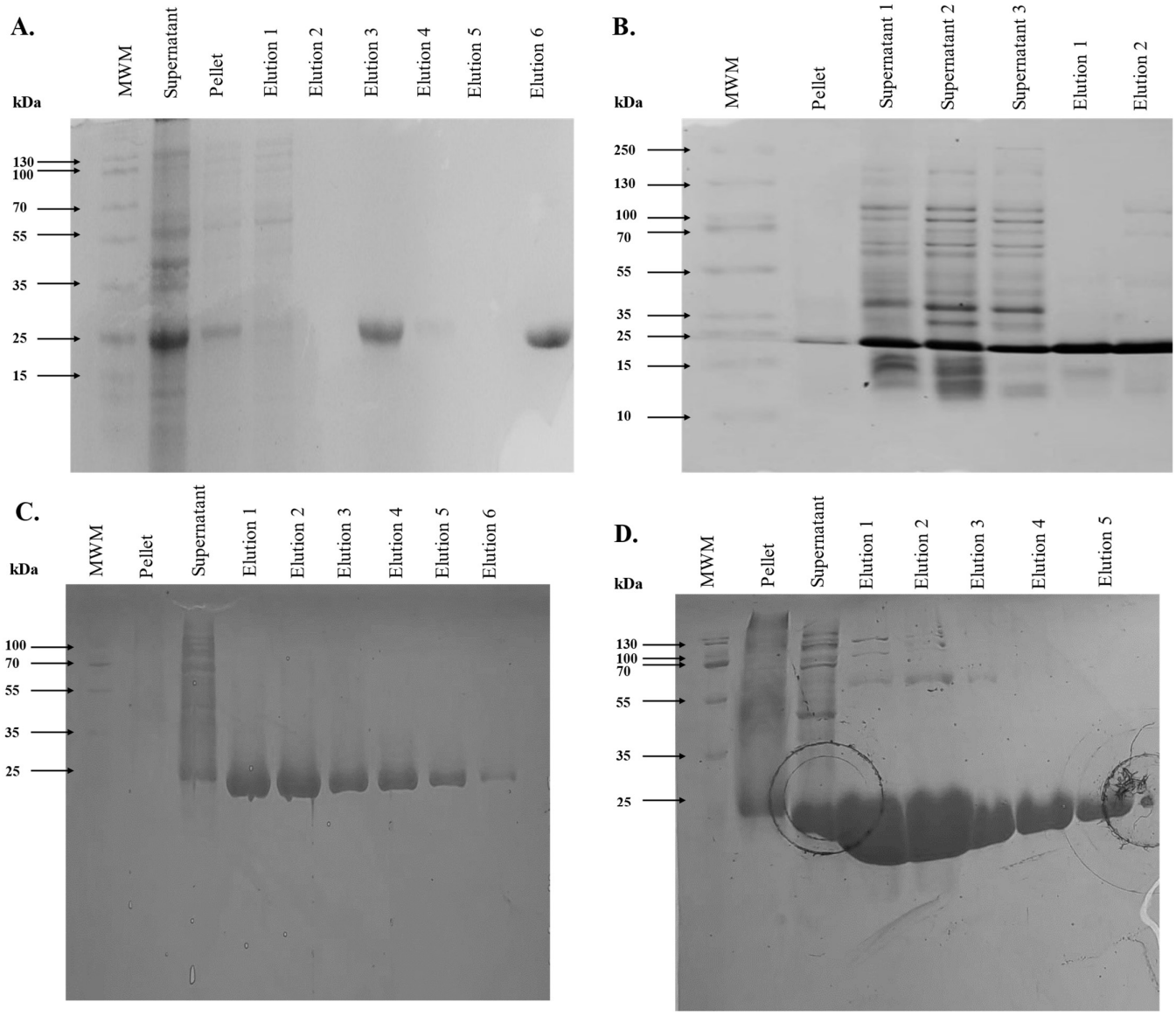


Figure 13. SDS-PAGE images depicting the steps of purification for the Alpha and Mu wildtype and consensus mutant proteins. The gels represent proteins that were produced for six hours at 30 °C and were purified via a one-step Ni^{2+} IMAC purification. MWM: molecular weight marker. (A) The hGSTA1-1 gel depicts a pellet with a marked portion of the protein being produced in the insoluble fraction. Quantitatively, the one-step elution did not produce an exceptionally high yield of protein. (B) The hGSTM1-1 gel exhibits three supernatant samples, which are representative of different sonication conditions. A considerable amount of protein was produced in the insoluble fraction, but the soluble protein was adequate. The elution samples were early collected fractions; therefore, contamination may be seen in these lanes. (C) The Alpha consensus gel exhibits six elution samples, which were collected progressively as the one-step elution was performed. Contamination is markedly lower in the later elution fractions. (D) The Mu consensus gel exhibits considerable loss of the protein in the insoluble fraction. However, the amount of overall produced protein was substantial. Additionally, the fractions are relatively less contaminated than the other proteins.

Table 1: Summary of the purification metrics obtained from spectrophotometric measurements at various wavelengths

Absorbance wavelength	**hGSTA1-1**	**Alpha consensus**	**hGSTM1-1**	**Mu consensus**
A_{280}	0.15	0.14	0.14	0.27
A_{340}	0.01	0.02	0.01	0.07
A_{260}/A_{280}	0.86	0.77	0.89	0.81

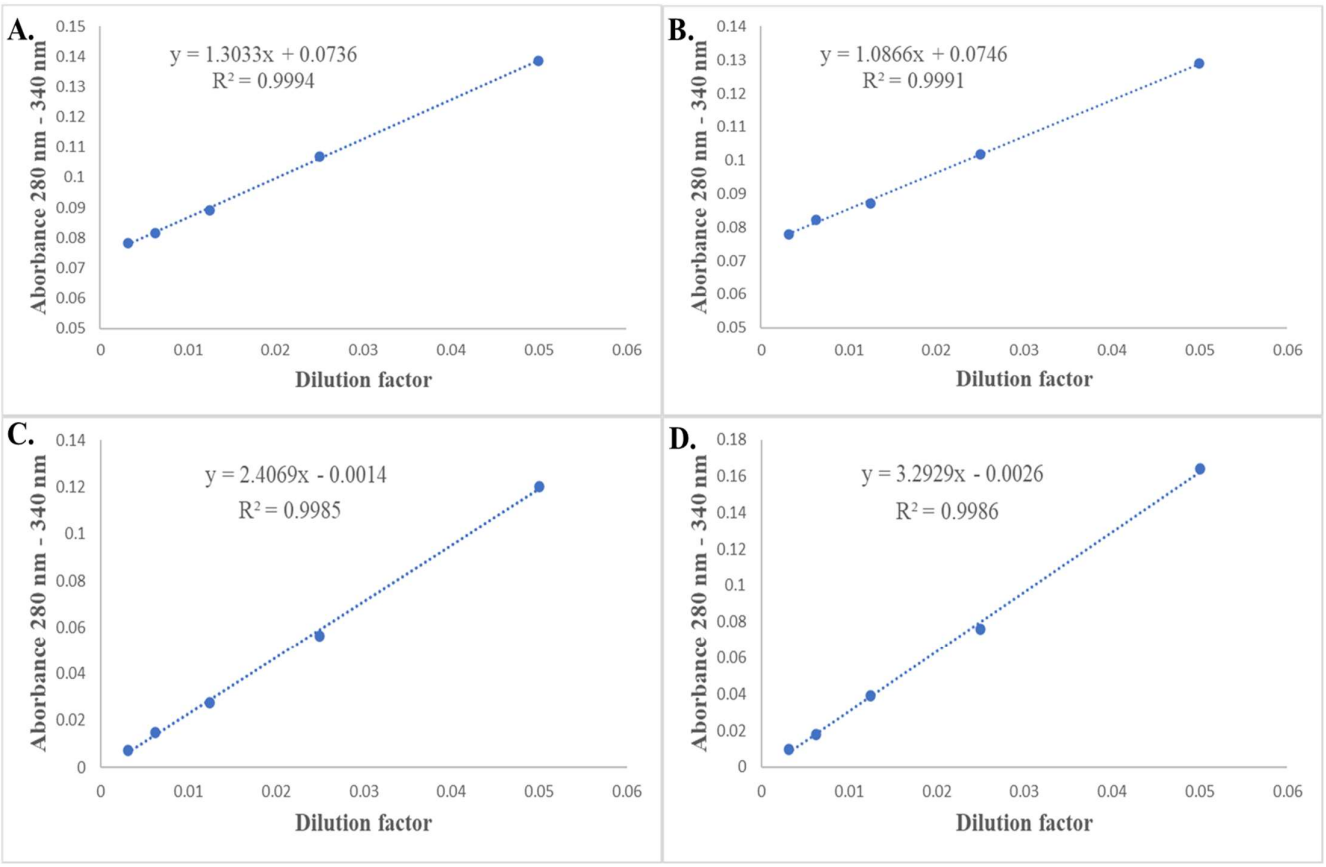


Figure 14. Linear regression models of the corrected absorbance data for the wildtype and mutant proteins of the Alpha and Mu classes. The corrected absorbance readings at 280 nm (after Rayleigh scattering effects have been accounted for) were plotted against the dilution factors for the corresponding absorbance values. A linear regression model was fitted to the data and was used to obtain the equations of the respective standard curves. The equations are accompanied by coefficient of determination (R^2) values. The equations were used in conjunction with the Beer-Lambert Law to obtain the concentration of the proteins in each sample. Concentrations of protein after purification were in the 30-85 µM range. The standard curves for (A) hGSTA1-1, (B) hGSTM1-1, (C) Alpha consensus, and (D) Mu consensus show a strong positive linear correlation.

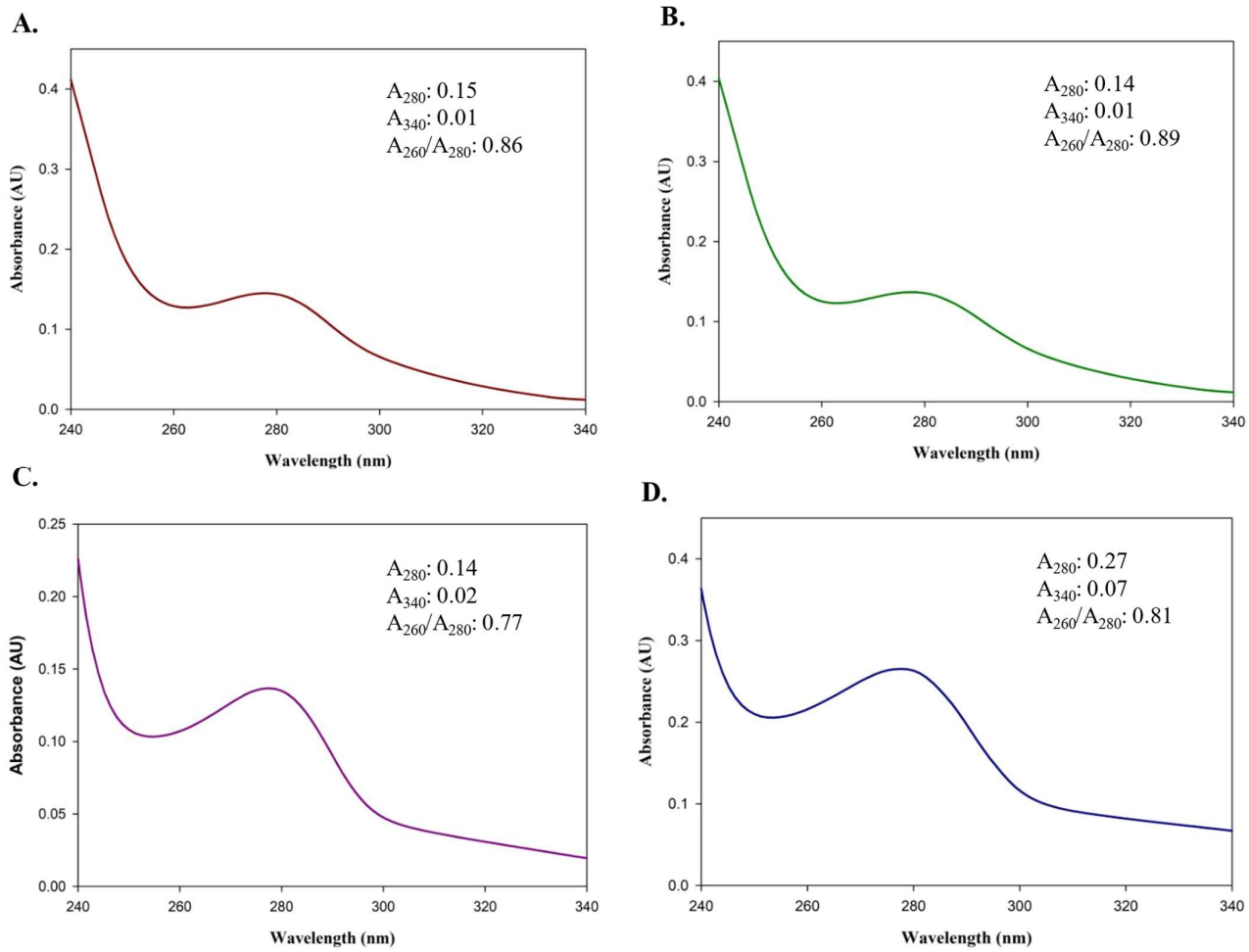


Figure 15. Qualitative assessment of the purity and yield of the proteins after purification. The absorbance spectra were measured from 240-340 nm for (A) hGSTA1-1, (B) hGSTM1-1, (C) Alpha consensus, and (D) Mu consensus. The A_{260}/A_{280} readings, which are representative of DNA contamination, were all below 1, indicating that nucleic acid contamination was not an issue in the purification process. The A_{340} readings were negligible, consistently falling in the 0.0101-0.067 AU range. The consensus mutant proteins of both classes exhibited the highest A_{340} values.

3.5 Far-UV CD

The far-UV CD spectra that were produced for hGSTA1-1, Alpha consensus mutant, hGSTM1-1, and Mu consensus mutant all displayed the spectrum of a protein that is predominantly alpha-helical (Figure 16). The spectra all showed the characteristic troughs at approximately 208 nm and 222 nm. There was, however, a large discrepancy between the MRE values of hGSTA1-1 and the Alpha consensus mutant. The MRE values of the Mu wildtype and consensus mutant protein were within sufficient range of each other, with the discrepancy between ellipticity of the hGSTM1-1 and Mu consensus mutant troughs being slightly more pronounced. It was observed that hGSTM1-1 showed a slightly lower ellipticity at the 208 nm and 222 nm troughs.

3.6 ANS binding

The properties of the extrinsic fluorescence dye ANS were exploited in order to gain a deeper understanding of the global structural composition of the proteins. Due to the increased quantum yield and considerable blue-shift that occurs when ANS binds to a hydrophobic surface and its fluorescence is quenched by water, deductions regarding the tertiary structure of the proteins may be made with some confidence. ANS binding studies were, therefore, performed on the respective proteins (Figure 17).

In the spectra that showed the measurement of the fluorescence signal of the protein and ANS samples, a significant blue-shift of the protein-ANS spectra was observed, in relation to the spectrum produced by the free ANS sample. This established that ANS binds to the proteins in their native form in the absence of any ligands. This confirmed that all of the proteins under investigation possess exposed hydrophobic sites. Although all the spectra (ANS-bound protein) exhibited a blue-shift and increased quantum yield, the blue-shifts and quantum yields were larger for the consensus mutant proteins of both classes.

When in the presence of ANS and GSH, the magnitude of the blue-shifts and quantum yields of hGSTA1-1 and the Alpha consensus mutant spectra (ANS-bound protein) were more pronounced in their increase. Relative to hGSTA1-1, the Alpha consensus mutant exhibited a greater quantum yield and a larger blue-shift. The same trend was observed in the hGSTM1-1, wherein the Mu consensus mutant had a considerably larger blue-shift and a higher quantum yield when in the presence of GSH.

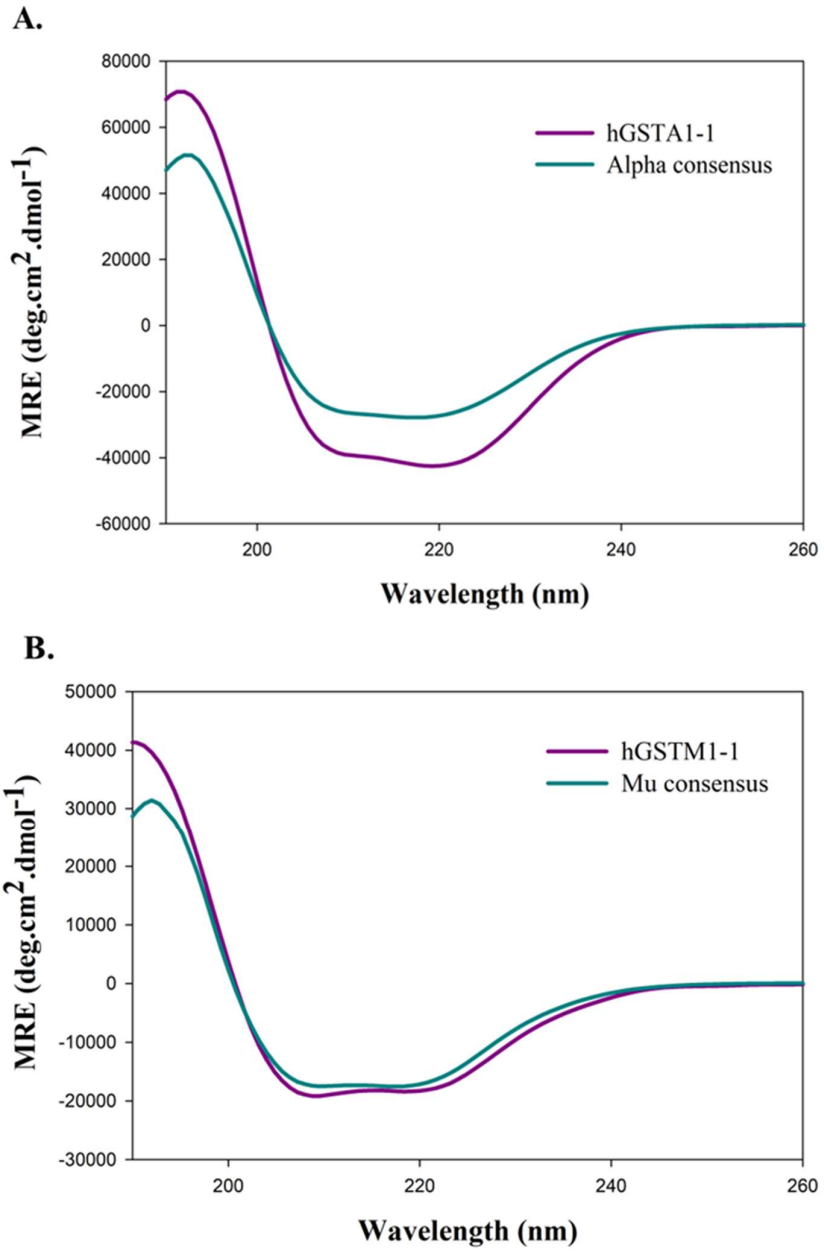


Figure 16. Far-UV CD spectra for the wildtype and mutant proteins of the Alpha and Mu classes. Far-UV CD was used as a means to assess the secondary structural composition of the. The experiments were performed in triplicate at 20 °C. (A) The far-UV CD spectra of hGSTA1-1 and Alpha consensus mutant. (B) The far-UV CD spectra of hGSTM1-1 and Mu consensus mutant. The two troughs at ~208 nm and ~222 nm are characteristic of a predominantly alpha-helical secondary structure. A discrepancy in the MRE readings of hGSTA1-1 and the Alpha consensus mutant was observed, while the spectra of hGSTM1-1 and the Mu consensus mutant displayed insignificant differences.

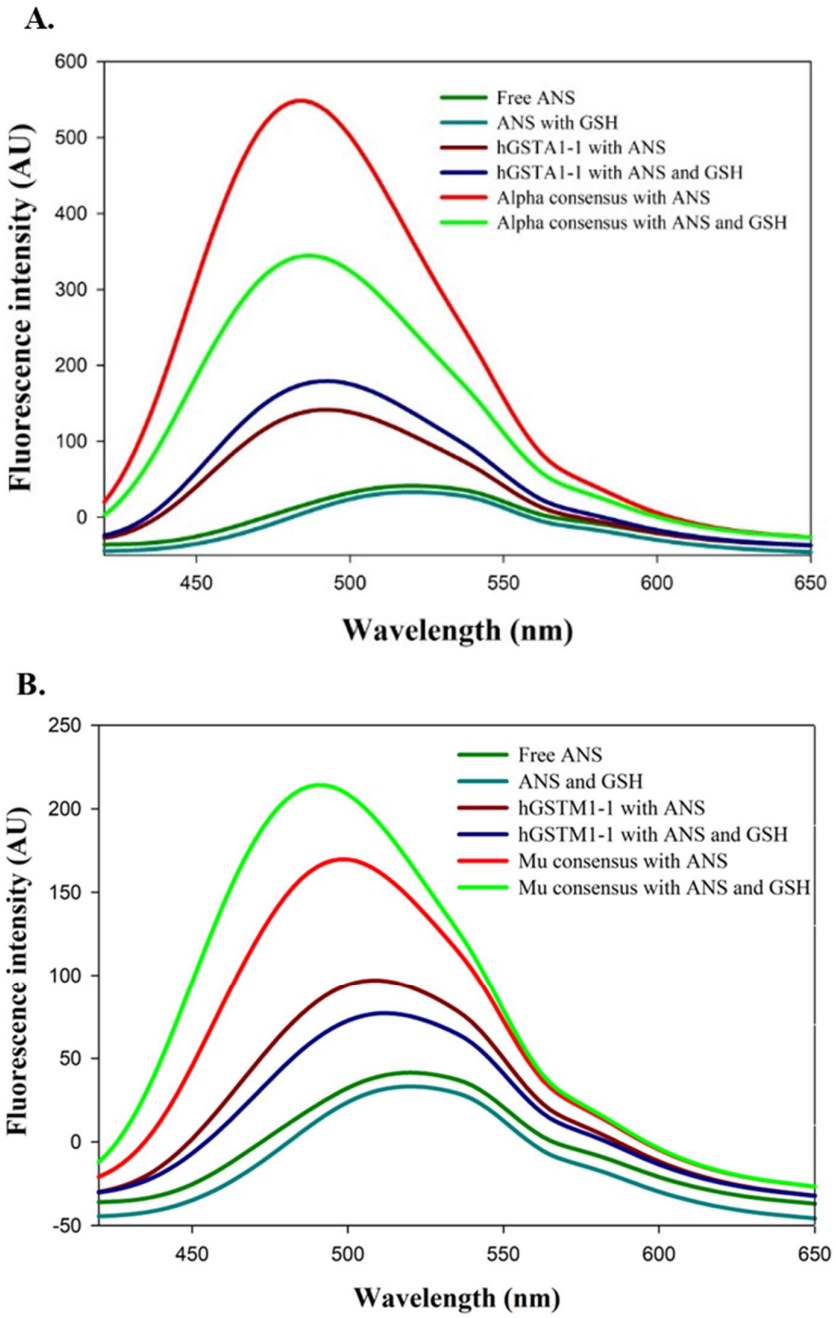


Figure 17. ANS fluorescence spectra of the wildtype and consensus mutant proteins of the Alpha and Mu classes in various conditions. ANS fluorescence was conducted at an excitation wavelength of 395 nm. (A) Fluorescence spectra of hGSTA1-1 and Alpha consensus mutant. (B) Fluorescence spectra of hGSTM1-1 and Mu consensus mutant. Free ANS is shown to have a fluorescence maximum of 540-550 nm. In (A), significant blue shifts are observed. The spectra of the Alpha consensus mutant in the presence of ANS and ANS with GSH produce are far greater quantum yield and a relatively larger blueshift, in comparison to the wildtype spectra of the same conditions. In (B), significant blue shifts were also observed in the spectra. The quantum yield of the ANS when bound to the Mu consensus protein is far greater than when ANS is bound to hGSTM1-1.

3.7 Thermal shift assays

Once the assays were completed, the raw data were extracted and analysed using CFX Maestro™ software (CFX Maestro™, Version 1.0, Bio-Rad Laboratories). The plots of the negative first derivative of the melt curves (Figure 18) and the melt curves in relative fluorescence units (RFU) (Figure 19) were obtained. These curves and plots are the averages of the replicates. The T_m was derived from the midpoint of the sigmoidal curve. The results, as shown in Table 2, indicate that the consensus mutations caused an increase in the thermal stability of the proteins. In the case of all the proteins, the presence of GSH showed some degree of increased conformational stability, as the melting temperatures slightly increased.

3.8 Enzyme activity and Michaelis-Menten kinetics

Once the GSH-CDNB conjugation assay results were obtained, a non-linear regression model was fitted to the data, generating hyperbolic functions using GraphPad Prism v10.0.2, thereby graphically representing the Michaelis-Menten kinetics (Figure 20). Thereafter, the Michaelis-Menten parameters were calculated and statistically analysed (Table 3). The K_M values for hGSTA1-1, Alpha consensus, hGSTM1-1, and Mu consensus were 394.20 µM, 190.80 µM, 894.90 µM, and 538.30 µM, respectively. Likewise, the V_{max} values were 0.27 µmol.min^{-1}, 0.02 µmol.min^{-1}, 0.30 µmol.min^{-1}, and 0.10 µmol.min^{-1}, respectively. In the instances of both the Alpha and Mu classes, the consensus mutants were significantly less active than their wildtype counterparts. As it pertains to the specific activity of the proteins, the Mu consensus mutant showed an increase by a factor of three, relative to hGSTM1-1. The Alpha consensus mutant showed a decrease by a factor of two, relative to hGSTA1-1.

Table 2: Summary of the melting temperatures (T_m) of the wildtype and consensus mutant proteins

	hGSTA1-1		**A consensus**		**hGSTM1-1**		**Mu consensus**	
	No GSH	With GSH	No GSH	With GSH	No GSH	With GSH	No GSH	With GSH
T_m **(°C)**	49.5	50	62	62.5	47	47.5	60	60.5

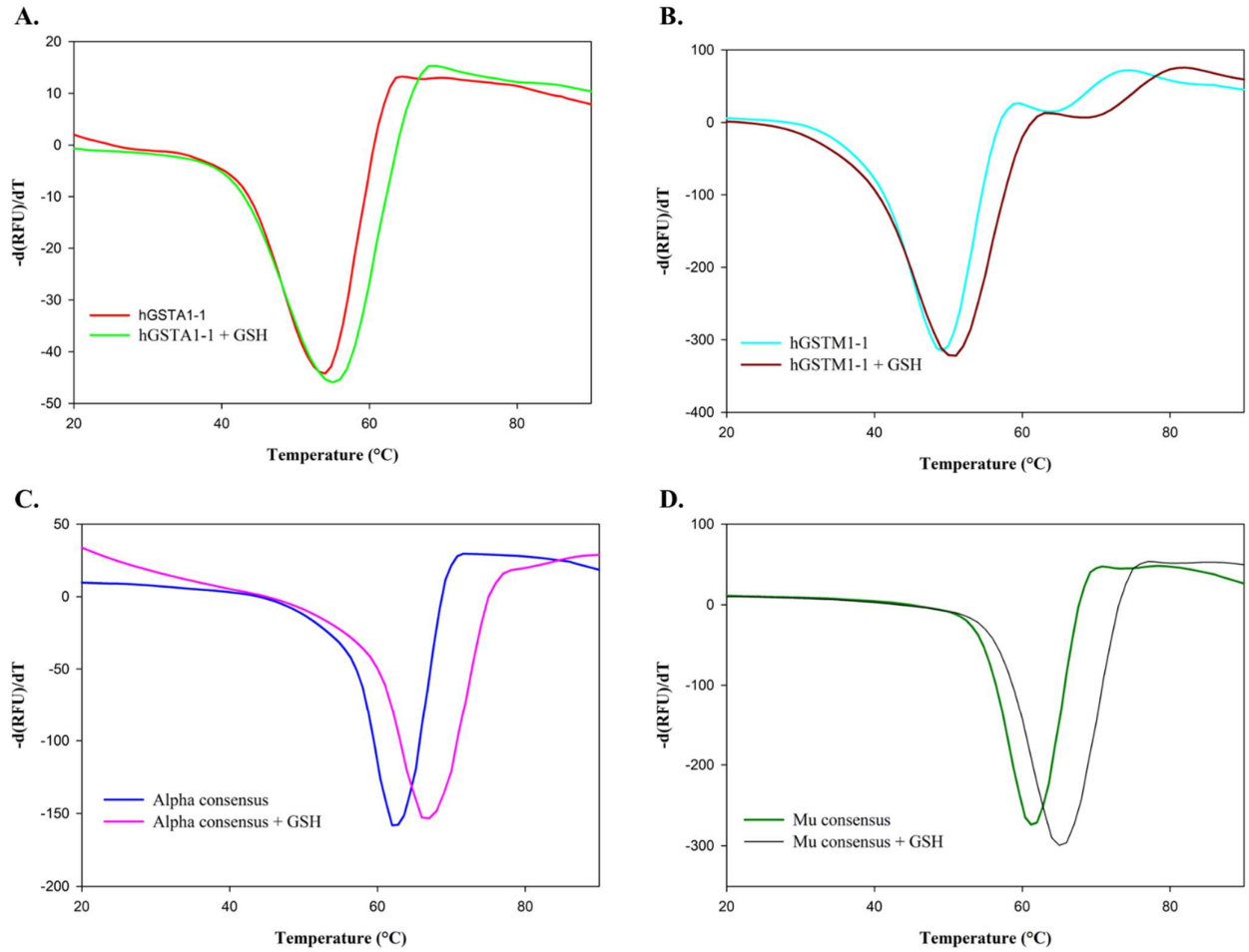


Figure 18. Thermal shift assays graphically represented as negative melt curve derivative plots. Negative derivative plots may often yield a more accurate T_m, hence their inclusion. The fluorescence signal given by the bound SYPRO® orange dye was measured from 0-90 °C. The melting temperatures ostensibly increase when the protein is bound to GSH. (A) The negative derivative melt curve of hGSTA1-1 in the presence (●) and absence (●) of GSH. (B) The negative derivative melt curve of hGSTM1-1 in the presence (●) and absence (●) of GSH. (C) The negative derivative melt curve of the Alpha consensus mutant protein in the presence (●) and absence (●) of GSH. (D) The negative derivative melt curve of the Mu consensus mutant protein in the presence (●) and absence (●) of GSH.

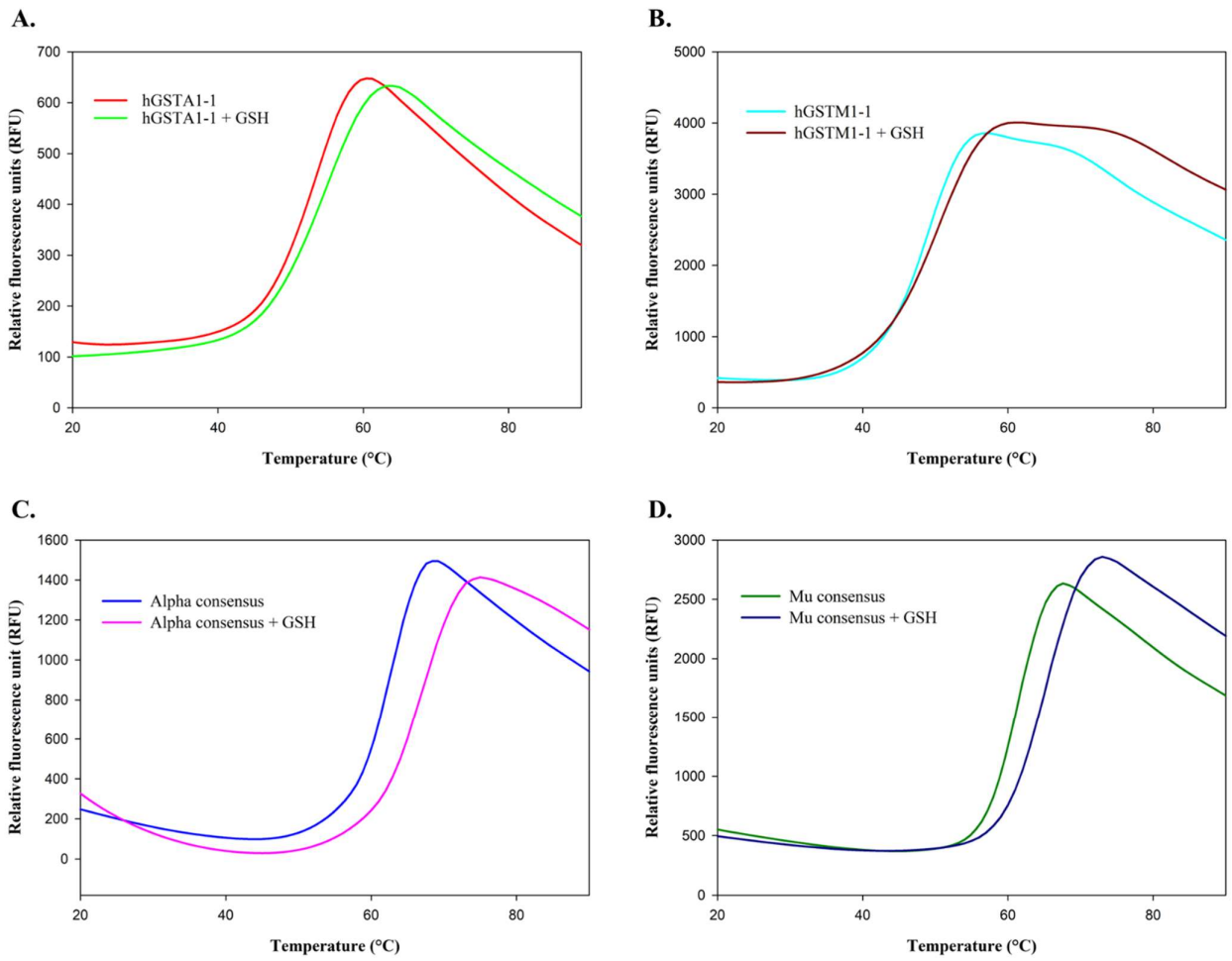


Figure 19. Thermal shift assays graphically represented as melt curves in relative fluorescence units. The fluorescence signal given by the bound SYPRO® orange dye was measured from 0-90 °C. The melting temperatures ostensibly increase when the protein is bound to GSH. (A) The melt curve of hGSTA1-1 in the presence (●) and absence (●) of GSH. (B) The melt curve of hGSTM1-1 in the presence (●) and absence (●) of GSH. (C) The melt curve of Alpha consensus in the presence (●) and absence (●) of GSH. (D) The melt curve of Mu consensus in the presence (●) and absence (●) of GSH.

Table 3: Summary of the enzyme kinetics results from steady-state conditions

Kinetic parameter	**hGSTA1-1**	**A consensus**	**hGSTM1-1**	**Mu consensus**
Specific activity ($\mu mol.min^{-1}.mg^{-1}$)	4.22	1.83	4.78	15.49
K_M (μM)	394.20	190.80	894.90	538.30
V_{max} ($\mu mol.min^{-1}$)	0.27	0.02	0.30	0.10

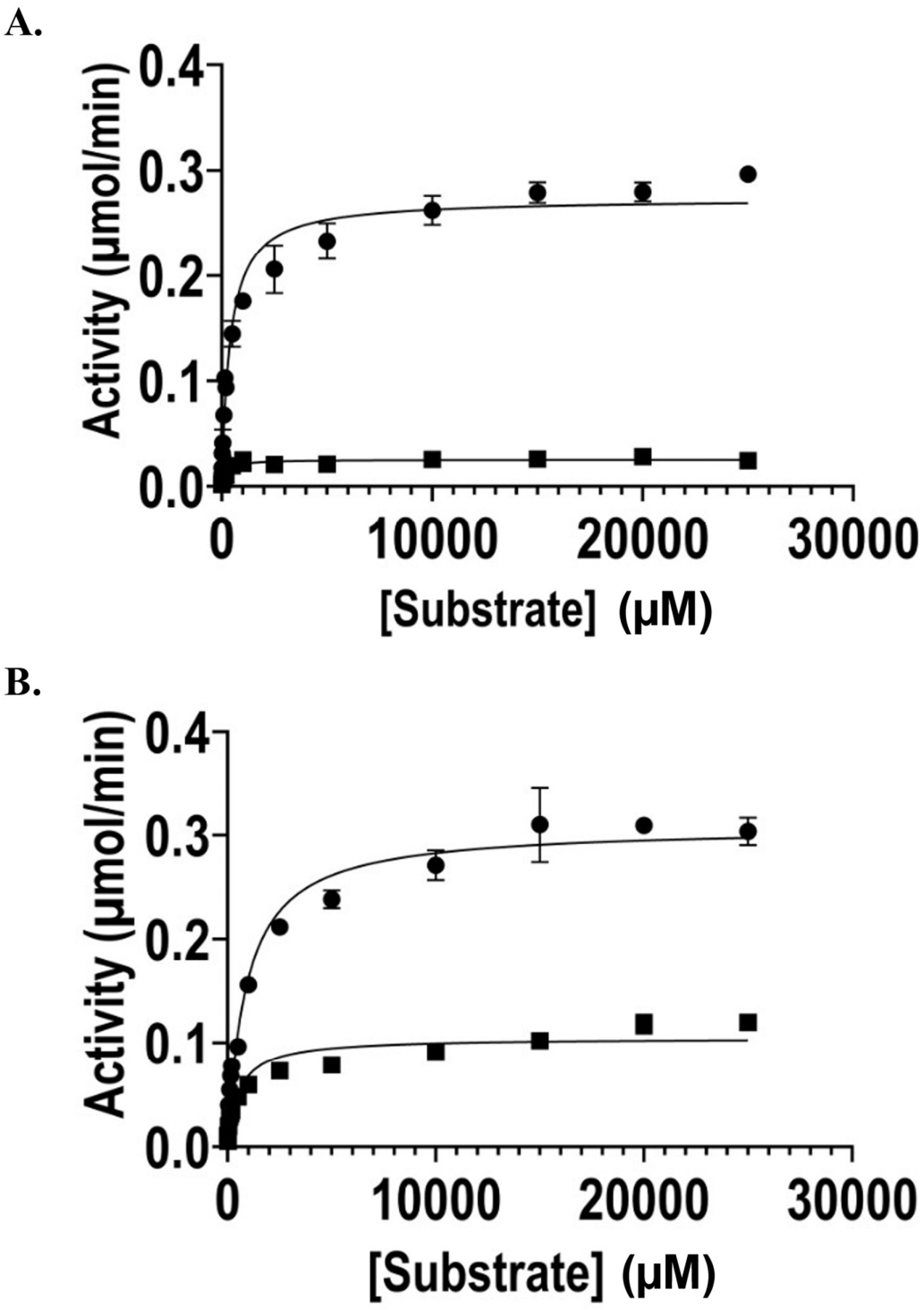


Figure 20. Michaelis-Menten plots of the wildtype and consensus mutant proteins of the Alpha and Mu classes. (A) hGSTA1-1 (●) and Alpha consensus (■). The activity of the Alpha consensus mutant is significantly lower than hGSTA1-1. (B) hGSTM1-1 (●) and Mu consensus (■). The activity between the wildtype and consensus mutant proteins once again showed a large discrepancy. In summary, the activity of the wildtype proteins was observed to be several orders of magnitude higher than the consensus mutant proteins. The Mu consensus mutant was markedly more active than the Alpha consensus mutant. Generated using GraphPad Prism v10.0.2

3.9 Molecular dynamics simulations

Once the MD simulations were run, the systems were analysed. To assess the stability of the proteins, the root-mean-square-deviation (RMSD) and the root-mean-square-fluctuation (RMSF) were used. More specifically the Cα-RMSD and Cα-RMSF values were used. Observing the results between the Alpha consensus mutant system and the hGSTA1-1 system, the RMSD results indicate that the Alpha consensus mutant equilibrated and converged at around 350 ns, while hGSTA1-1 seemed to have equilibrated and converged at around 100 ns (Figure 21). In the instance of the Mu consensus mutant and hGSTM1-1, hGSTM1-1 equilibrated and converged at 300 ns, while the Mu consensus mutant converged and equilibrated at approximately 100 ns.

The RMSD results indicated there was a larger global structural deviation between the Alpha proteins than there was in the Mu protein. Amongst the RMSD simulations of the Alpha proteins, hGSTA1-1 simulations had a median of 2.30 Å while the Alpha consensus mutant had a median of 2.58 Å. For the Mu proteins, the RMSD medians were 1.79 Å and 1.54 Å for hGSTM1-1 and the Mu consensus mutant, respectively. The RMSF results are a function of the number of residues in the protein. RMSF, which is a metric that represents the degree of flexibility of a residue over the course of a simulation, is vital for understanding protein stability. For the Alpha proteins, the RMSF was calculated per residue, for a total of 442 residues. In relation to its wildtype counterpart, the Alpha consensus mutant displayed a higher degree of fluctuation, as evidenced by the median in the box-and-whisker plot (Figure 22). The hGSTA1-1 RMSF median was 0.99 Å and the Alpha consensus mutant median was 1.16 Å. The region with the greatest degree of fluctuation was the region from residues 41 to 51. This region overlaps with a significant portion of the α2 helix. Residues 207 to 219, which partially include the residues of the α9 helix, show greater perturbation than the same region in hGSTA1-1. For the Mu proteins, the RMSF was calculated per residue, for a total of 434 residues. The Mu consensus mutant, which had an RMSF median of 0.79 Å, exhibited a lower degree of fluctuation, as is illustrated by the median of the box-and-whisker plot. The RMSF median of hGSTM1-1 was 1.01 Å.

The Rg results illustrated the disparity between the compactness of the wildtype proteins and the mutant proteins (Figure 23). It was observed that the Rg of hGSTA1-1 was, on average, slightly higher than the Rg of the Alpha consensus mutant. The median values of the proteins attest to this, with the hGSTA1-1 Rg simulations median at 22.04 Å and the Alpha consensus simulations median at 21.99 Å.

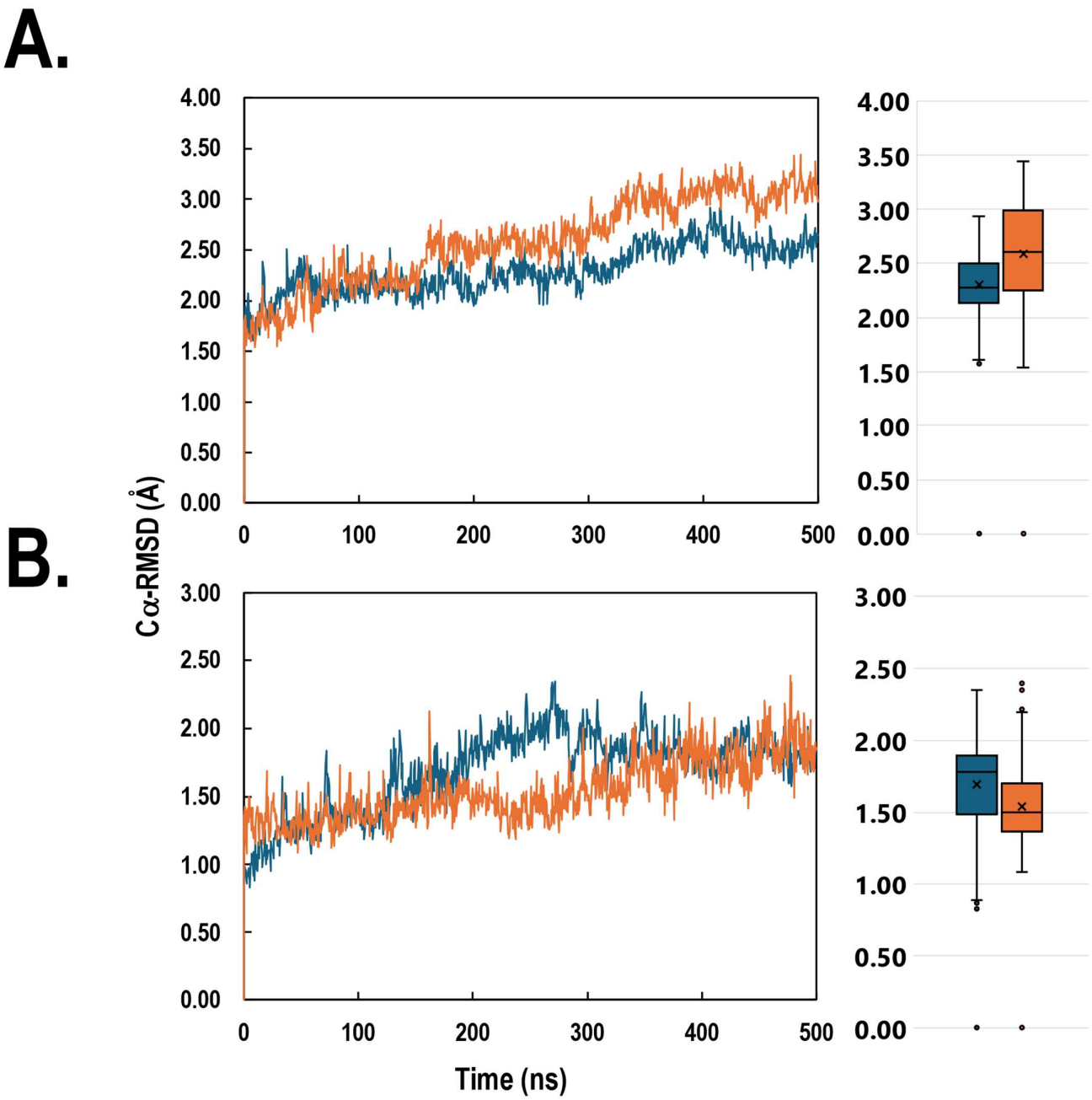


Figure 21. Cα-RMSD of the Alpha and Mu wildtype and consensus mutant proteins from the MD simulations conducted over 500 ns. (A) hGSTA1-1 (●) had a simulation median of 2.30 Å and Alpha consensus (●) had a simulation median of 2.58 Å. (B) hGSTM1-1 (●) had a simulation median of 1.79 Å and Mu consensus (●) had a simulation median 1.54 Å. The blue box-and-whisker plot represents the data from the wildtype GST simulations while the orange box-and-whisker plot represents data from the consensus mutant GST simulations.

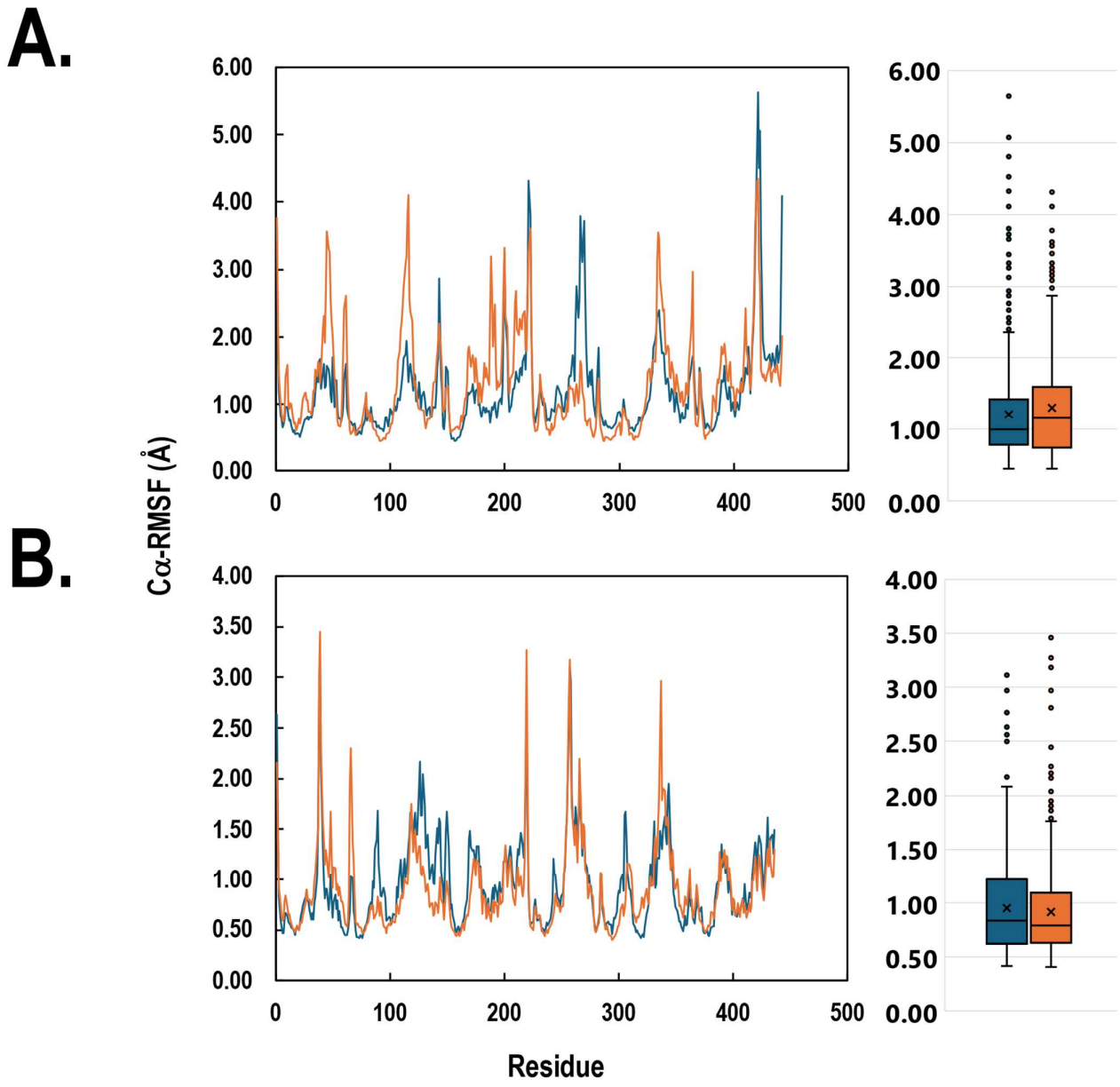


Figure 22. Cα-RMSF of the Alpha and Mu wildtype and consensus mutant proteins per residue. (A) hGSTA1-1 (●) had a simulation median of 0.99 Å and Alpha consensus (●) had a simulation median of 1.16 Å. (B) hGSTM1-1 (●) had a simulation median of 1.01 Å and Mu consensus (●) had a simulation median of 0.79 Å. The respective box and whisker plots are on the right-hand side of the image. The blue box-and-whisker plot represents the data from the wildtype GST simulations while the orange box-and-whisker plot represents data from the consensus mutant GST simulations.

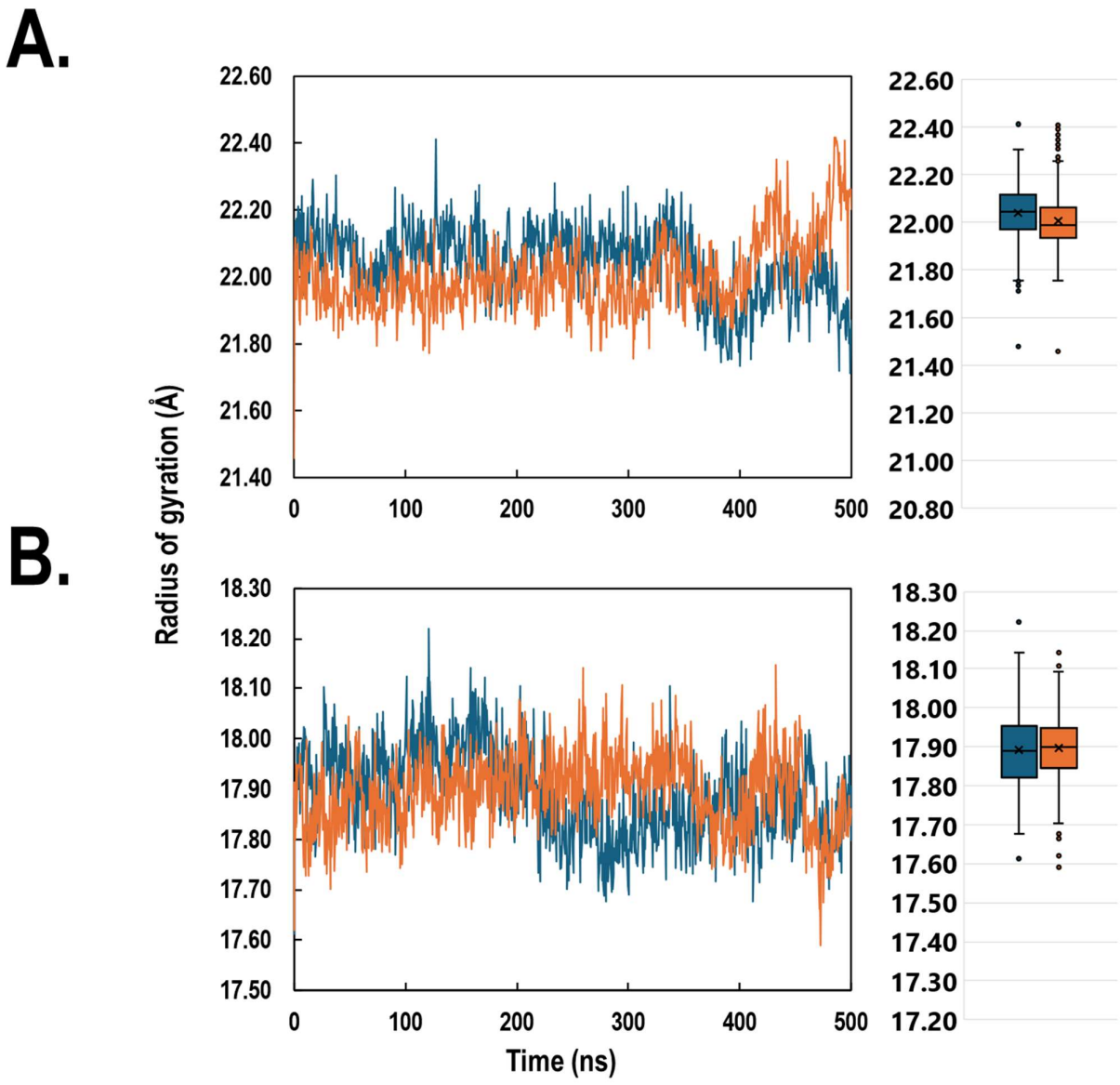


Figure 23. Radius of gyration of the Alpha and Mu wildtype and consensus mutant proteins from the MD simulations conducted over 500 ns. (A) hGSTA1-1 (●) had a simulation median of 22.04 Å and Alpha consensus (●) had a simulation median of 21.99 Å. (B) hGSTM1-1 (●) had a simulation median of 17.89 Å and Mu consensus (●) had a simulation median 17.90 Å. The blue box-and-whisker plot represents the data from the wildtype GST simulations while the orange box-and-whisker plot represents data from the consensus mutant GST simulations.

The Mu proteins, however, exhibited a significant degree of overlap, thus producing only a marginal difference in the respective Rg simulations median values, with the hGSTM1-1 median at 17.89 Å and Mu consensus mutant median at 17.90 Å. The Y9-Y9 and Y6-Y6 distance results further corroborated the Rg results (Figure 24). The discrepancy between the average Y9-Y9 simulation distances of hGSTA1-1 and the Alpha consensus mutant was more pronounced than the Rg results, with the medians being 38.17 Å and 36.62 Å, respectively. The Y6-Y6 results of the Mu proteins once again showed a great overlap between the wildtype and the consensus mutant, with the difference in the average Y6-Y6 distance being insignificant, with the median values being 36.88 Å and 36.75 Å, respectively.

The superimposition of the Alpha consensus mutant structure on the hGSTA1-1 structure showed that the RMSD between these structures was 1.85 Å (Figure 25). For the Mu consensus mutant structure and its wildtype counterpart, the RMSD was 0.88 Å (Figure 26). The RMSD of the superimposed structures was important for determining structural deviation. The results of the superimposition greatly informed the degree to which structural similarity was preserved.

The site map analysis further elucidated the disparities between the characteristics of the global structure of the wildtype proteins and mutant proteins (Figure 27). Per the results of the sitemap analysis and the imposed criterion (a Dscore greater than 0.75), the cavities with the largest volumes were the subunit interfaces of the respective proteins (Table 4). As it pertains to the Alpha proteins, hGSTA1-1 had a dimer interface cavity with a volume of 1377.831 $Å^3$ while the Alpha consensus mutant had a dimer interface cavity with a volume of 1638.511 $Å^3$. As it pertains to the Mu proteins, hGSTM1-1 had a dimer interface cavity with a volume of 1183.007 while the Mu consensus mutant had a dimer interface cavity with a volume of 529.249 $Å^3$.

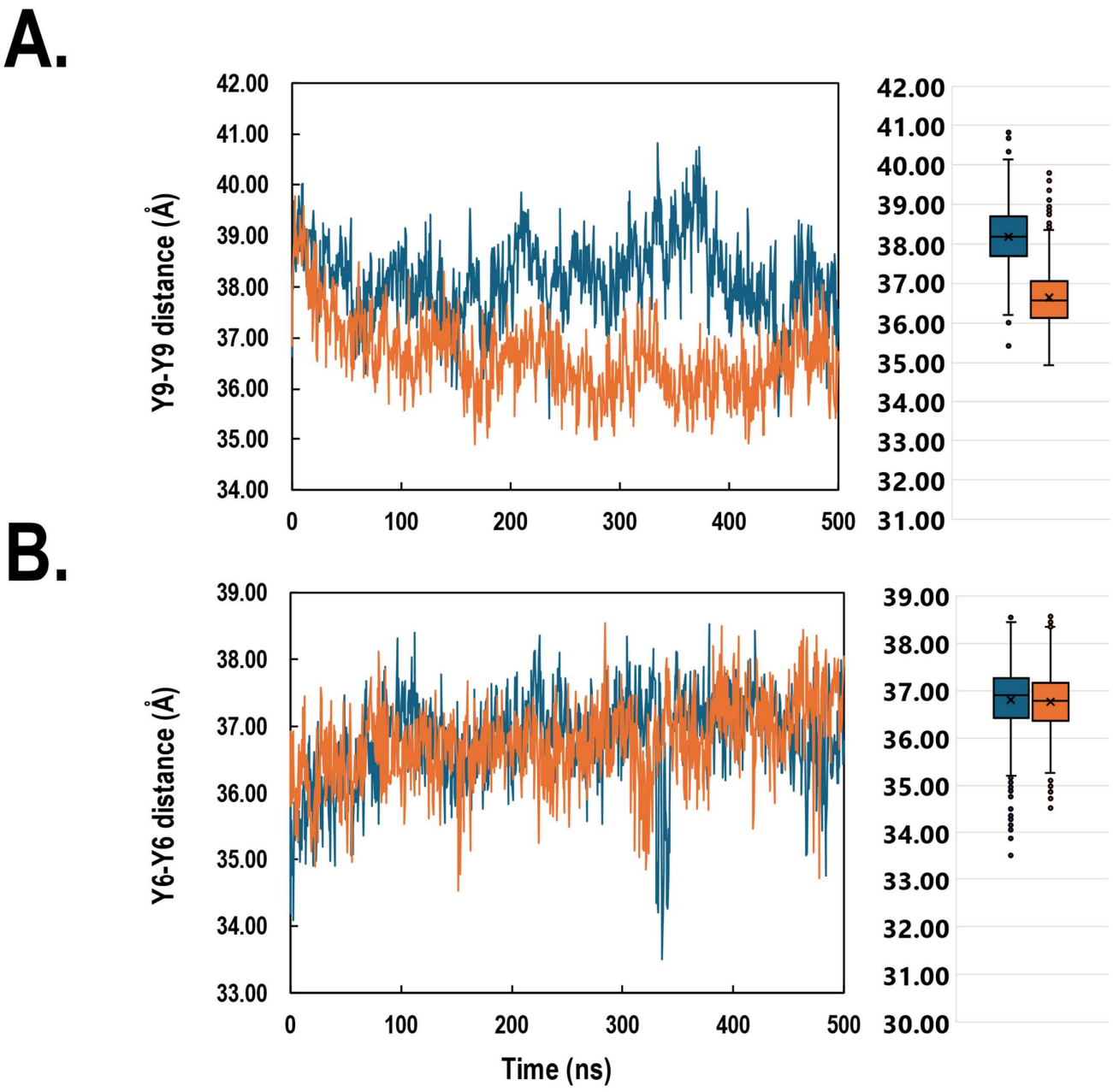


Figure 24. Atomic distance between the tyrosine 9 (Y9) residues and tyrosine 6 (Y6) residues of both subunits from the MD simulations conducted over 500 ns. (A) The Y9-Y9 distance of the Alpha subunits is measured. hGSTA1-1 (●) had a simulation median of 38.17 Å and Alpha consensus (●) had a simulation median of 36.62 Å. (B) The Y6-Y6 distance of the Mu subunits is measured. hGSTM1-1 (●) had a simulation median of 36.88 Å and Mu consensus (●) had a simulation median of 36.75 Å. The blue box-and-whisker plot represents the data from the wildtype GST simulations while the orange box-and-whisker plot represents data from the consensus mutant GST simulations.

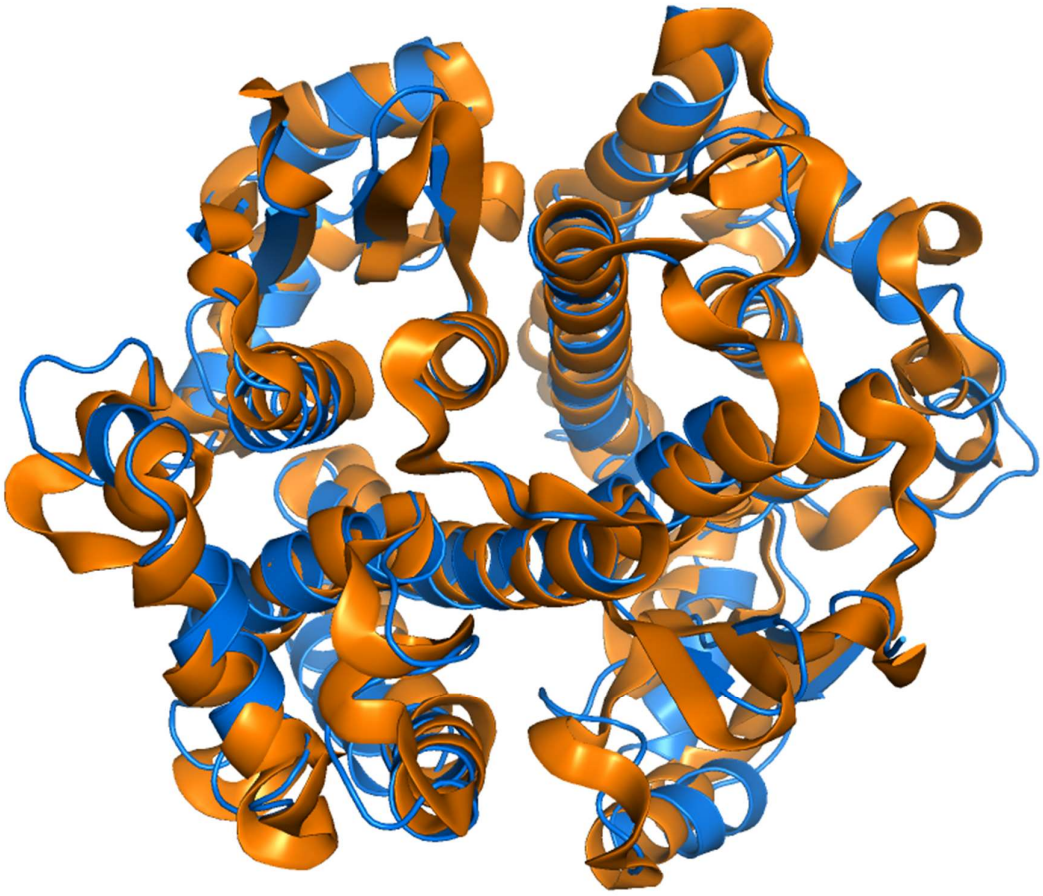


Figure 25. Superimposition of the hGSTA1-1 and Alpha consensus mutant structures. The structures are depicted in ribbon form, and hGSTA1-1 is depicted in blue (●), while the Alpha consensus mutant is depicted in orange (●). The RMSD between the wildtype and the mutant is 1.85 Å. Generated using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

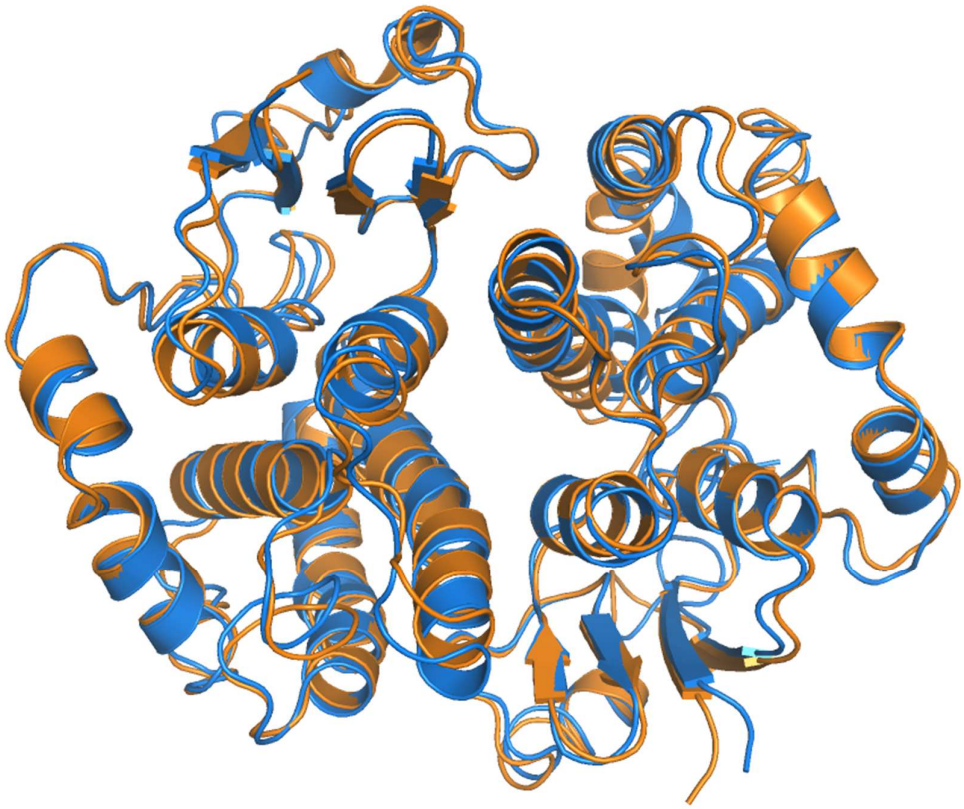


Figure 26. Superimposition of the hGSTM1-1 and Mu consensus mutant structures. The structures are depicted in ribbon form, and hGSTM1-1 is depicted in blue (●), while the Mu consensus mutant is depicted in orange (●). The RMSD between the wildtype and the mutant is 0.88 Å. Generated using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

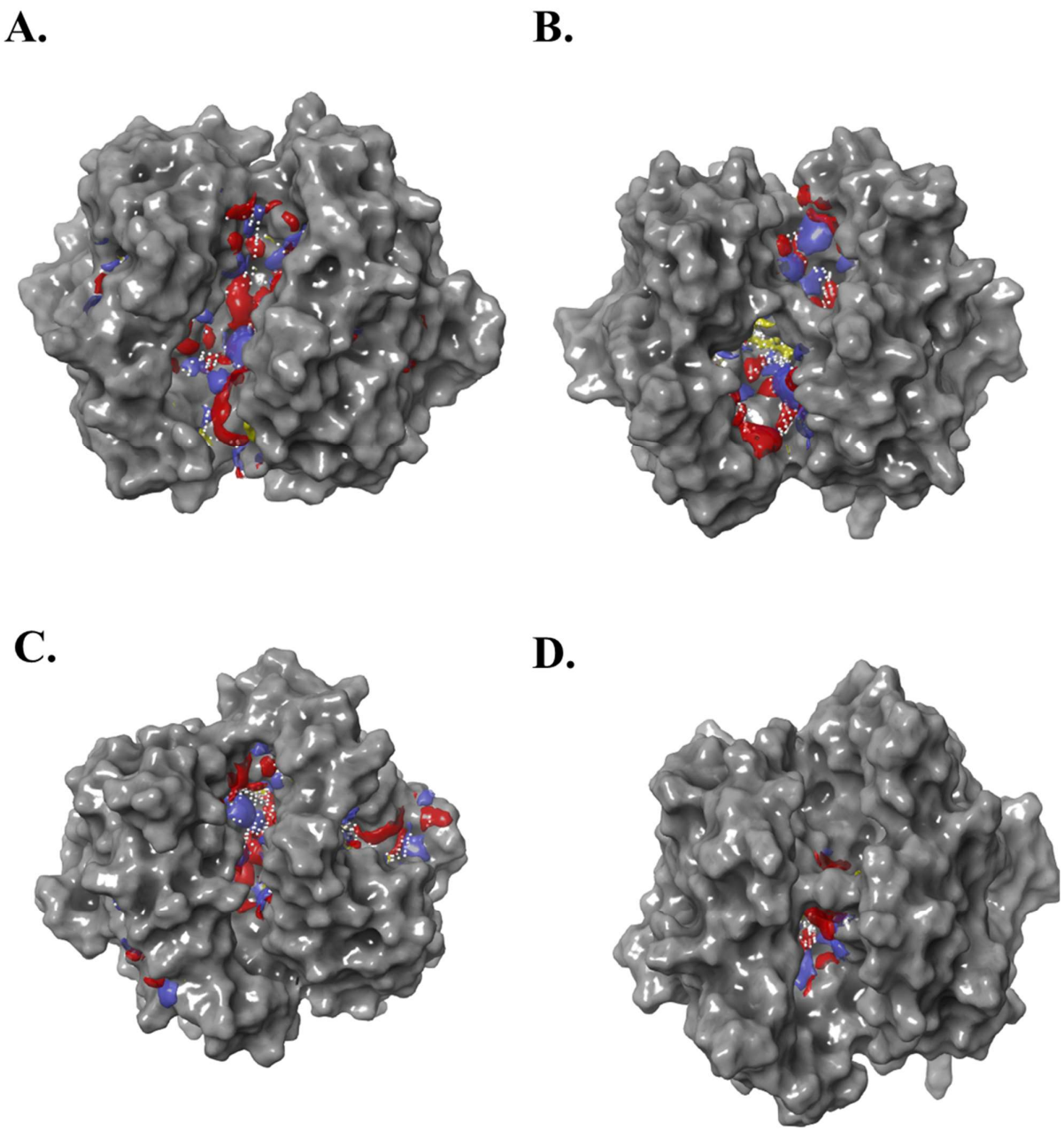


Figure 27. Sitemap analysis results presented as a cavity in the surface of the protein. The cavity with the largest volume and a Dscore greater than 0.75 was selected and displayed. All of the cavities depicted are located at the dimer interface. (A) hGSTA1-1 had a dimer interface cavity volume of 1377.831 $Å^3$. (B) Alpha consensus mutant had a dimer interface cavity volume of 1638.511 $Å^3$. (C) hGSTM1-1 had a dimer interface cavity volume of 1183.007 $Å^3$. (D) Mu consensus mutant had a dimer interface cavity volume of 529.249 $Å^3$.

Table 4: Summary of the sitemap analysis results of the consensus mutant and wildtype proteins

Site	**A1**		**A-mutant**		**M1**		**M-mutant**	
	Dscore	Vol ($Å^3$)	Dscore	Vol ($Å^3$)	Dscore	Vol ($Å^3$)	Dscore	Vol ($Å^3$)
1	0.967	354.319	1.004	1638.511	0.990	390.677	1.099	529.249
2	1.027	1377.831	0.999	562.520	1.055	1183.007	0.872	206.143
3	0.713	173.558	0.884	190.365	0.817	227.752	1.036	466.137
4	0.735	159.152	0.862	167.384	0.707	138.572	0.839	175.616
5	0.766	250.390	0.701	112.847	0.603	119.707	0.660	122.108
Total		1982.540		2558.780		1940.008		1377.145
Average		660.847		639.695		485.002		344.286

Chapter 4: Discussion

Due to the unlikelihood that the results of unsuccessfully designed consensus proteins would be published, assessing the comparative literature regarding consensus protein design will not lead to overly robust analyses and conclusions. After the MSA was performed, a key residue mutation was found in the Alpha consensus mutant protein; the M50L mutation. The replacement of the M50 residue in hGSTA1-1 is likely to have affected the lock-and-key motif that is vital for dimer interface interactions, as well as G-site stability. As shown in a study that assessed a M50A mutation, the replacement of the M50 residue affects the G- and H-sites in hGSTA1-1. This is because the M50 residue itself is located on the loop that forms part of the G-site, and also because the M50 residue maintains interactions with the α2 helix, which, in turn, maintains interactions with the α9 helix. Such key mutations were not found in the Mu consensus mutant protein. Due to the number of mutations, it is not possible to draw direct correlations between the stability and function of proteins with specific residues, with the notable exception of the M50L mutation. The mutations may have a cumulative effect on the stability and function of the proteins. Conversely, single mutations may be solely responsible for the effects observed. The comparison, however, between the effects of the M50A and the M50L mutations is apt, given that both residues are aliphatic.

4.1 Protein production and purification

The heterologous production of the Mu consensus mutant protein produced the largest yield of all the proteins discussed hitherto. Its A_{280} reading of 0.26 AU, accompanied by its specific activity of 15.5 $\mu mol.min^{-1}.mg^{-1}$ rendered it the most pure and abundant protein. The Mu consensus protein exhibited such a high and pure yield, that it may be intimated that the consensus mutations are conducive for better heterologous protein production and purification. The Mu consensus mutant displayed an A_{340} value that was larger than the other wildtype and mutant proteins, indicating a larger proportion of aggregated protein. The production and purification of the Alpha consensus mutant protein showed low purity and low yield, relative to hGSTA1-1. The Alpha consensus mutant did, though, have a better A_{340} value than the Mu consensus mutant.

This result (of both increased and decreased protein production yields) has precedence in the literature regarding consensus protein design. The monomeric Azami green (mAG) consensus protein was shown to be produced in copious amounts, with respect to the wildtype protein [49]. Another protein in the same study, consensus green protein (CGP) had 89.8% sequence

similarity with mAG but had poorer protein production yields [49]. Production of recombinant proteins is multicausal, which would mean that trying to attribute the protein yield and purity to single characteristics of the proteins is counterproductive and fruitless.

4.2 SDS-PAGE purity analysis

The cytosolic consensus mutant protein proved to be completely insoluble. Therefore, no downstream experiments were carried out. The SDS-PAGE results, interpreted together with the purification results, show that the Mu consensus mutant was the protein with highest yield and purity. The yield of the Alpha consensus mutant was not considerably higher than hGSTA1-1.

4.3 Far-UV CD

The far-UV CD results largely corroborated the fact that the secondary structure was retained after the introduction of consensus mutations. There was, however, a marked discrepancy between the MRE measurements of hGSTA1-1 and the Alpha consensus mutant. This discrepancy may be attributed to the Alpha consensus mutant being partially unfolded under the experimental conditions (10 mM sodium phosphate, pH 7.0). Additionally, this discrepancy may be related to the nature of the α-helices in the consensus mutant protein. It has been shown that when isolated helices are greater in number than coiled-coil helices, the spectrum produced will be unlike when the number of isolated helices are fewer than coiled-coil helices [50]. This, however, is unlikely as coiled-coil helices produce larger MRE values than isolated helices and given that hGSTA1-1 is not composed of coiled-coil helices, the Alpha consensus protein cannot be composed of coiled-coil helices. Furthermore, studies that have investigated the consequences of replacing the M50 residue have shown that the additional α-helix, α9, is destabilised when the M50 residue is mutated [34]. This destabilisation may be attributed to the increased dynamism and flexibility of the α2 helix, which affects the α9 helix. When coupled with the RMSF results, it is more likely that the disruption of α9 occurred, as opposed to partial unfolding of the protein.

4.4 ANS binding

ANS binding was deemed to be the appropriate technique to experimentally examine the global tertiary structures of the proteins, as it provides information on the global structure of the GSTs [51]. To establish that ANS did in fact bind to the proteins, the respective proteins were incubated with 200 μM ANS and subsequently subjected to fluorescence spectroscopy. The overarching pattern observed from the ANS fluorescence spectroscopy results is that the

consensus mutant proteins of both classes showed increased quantum yields and significant blue shifts, relative to their wildtype counterparts. These changes may be ascribed to structural changes at the H-site. Mutations in GSTs have been shown to affect the available surface area of the H-site, thereby affecting ANS binding to the H-site, and affecting the degree of ANS fluorescence from that region [34]. Importantly, with regard to ANS binding studies, a key aspect of analysis is understanding the interplay between increased accessible hydrophobic surface area and increased solvent accessibility. The former increases ANS quantum yield, while the latter decreases ANS quantum yield through solvent quenching. These two phenomena counteract each other to a degree.

When the Mu wildtype and consensus mutant proteins were in the presence of ANS and GSH, blue-shifts and increased quantum yields in the spectra were observed, relative to the spectra of the Mu proteins solely bound to ANS. This is indicative of the possibility that GSH binding to the G-site likely stabilises the conformation of these proteins, priming them for ligand binding at the H-site. Although the Mu consensus protein did not possess a mutation that was as impactful as the M50L mutation seen in the Alpha consensus protein, the mutations in unison created a global structure of the Mu consensus protein that was increasingly compact. This is seen in Table 4, wherein the total volume of the cavities of the Mu consensus protein is 562.863 Å smaller than hGSTM1-1. This indicates that solvent accessibility of the Mu consensus mutant H-site is lower that of hGSTM1-1, reducing the quenching of the ANS fluorescence. Although the Alpha consensus protein cavities are larger than the hGSTA1-1 cavities, the hydrophobicity of the H-site, as will be explained below, is greater in the Alpha consensus protein.

To begin with, it is important to note that the Alpha consensus protein has a larger cavity volume than hGSTA1-1. This means that a larger surface area of the H-site is exposed to the solvent. Secondly, after analysing the spectrum of the Alpha consensus mutant in the presence of ANS and GSH, a decrease in fluorescence intensity was observed, in comparison to the Alpha consensus mutant spectrum of the protein binding solely with ANS. This is indicative of decreased hydrophobic surface area when GSH is bound. This is despite the M50L mutation found in the G-site. The M50 residue is a part of the loop found in the G-site between helix α2 and strand β3. The M50L mutation increases the motion of the α2 helix, which in turn, increases the motion of the α9 helix. This is because the α2 and α9 helices interact through tertiary packing interactions. Upon GSH binding, however, there is a stabilising effect that occurs, resulting in the increased rigidity of the α2 helix. This means

that increased α2 helix rigidity, in turn, causes the α9 helix to become more rigid. The increased α9 rigidity reduces the H-site hydrophobic surface area to which ANS can bind, thereby decreasing the ANS quantum yield.

When there is no GSH binding, the M50L mutation causes a change in the nature of the tertiary packing interactions between helices α2 and α9. The M50L mutation causes an increase in the dynamism of the α2 helix. This is because the loop between helix α2 and strand β3 interacts in a way that increases the motion of the α2 helix, thereby decreasing the rigidity of the α2 helix. Also, with no binding to the G-site, the are no forces that impose rigidity on the α2 helix. Due to the tertiary packing interactions between helices α2 and α9, the flexibility of the α9 helix is affected. This, in turn, leads to a more flexible α9 helix, increasing the hydrophobic surface area accessible for ANS. This causes an increase in the ANS fluorescence signal that is emitted. In both spectra, the increased hydrophobic surface area outweighs the increased solvent accessibility, largely due to the additional α9 helix, which displaces water molecules near the H-site. The larger cavity volumes of the Alpha consensus mutant also mean that there is an increase in the exposure of the H-site, which corresponds to a larger ANS quantum yield. Indeed, site directed mutagenesis studies have showed that the quantum yield of the spectrum of hGSTA1-1 with an M50 mutation is higher than the wildtype spectrum, owing to the increase in H-site hydrophobic surface area [34]. Even more specifically, the amino acid residue used as a substitute in that study was alanine, an aliphatic residue just as leucine is.

When this is taken into consideration together with the MD simulation results, this becomes the likelier explanation. As seen in the Alpha consensus mutant RMSF results (Figure 22), the region of residues 41 to 51, which corresponds to the α2 helix, exhibited greater perturbation than hGSTA1-1 due to the consensus mutations. Additionally, the region consisting of residues 207 to 219 showed greater flexibility than hGSTA1-1, a region which overlaps greatly with helix α9. This provides further evidence that the M50L mutation causes perturbation of the α2 helix, which causes increased α2 flexibility. Since the α2 and α9 helices are in contact due to tertiary packing interactions, increased flexibility in the α2 helix causes increased flexibility of the α9 helix, which, in turn, causes the increase of the accessible H-site surface area.

The same principles, however, do not apply to hGSTA1-1, due to the retainment of the M50 residue. Firstly, hGSTA1-1 has a smaller total cavity volume than the Alpha consensus

protein. This means that solvent accessibility to the H-site will be minimised. The additional α9 helix likely gains stabilising interactions with helix α2 upon GSH binding because hydrogen bond formation with the loop occurs and helix α2 displacement is not augmented by any mutations. This increases the rigidity of helix α2, which, in turn, decreases flexibility of the α9 helix, thus decreasing the H-site solvent accessibility. This corresponds to an increase ANS fluorescence. In the case of hGSTA1-1 when it is solely bound to ANS, the helices are not affected by the M50L mutation. Therefore, the interactions between the α2 and the α9 helix are maintained. There is no stabilising force in the form of GSH binding, however, so the solvent accessibility is increased. The fluorescence intensity, therefore, is comparatively lower. Increased solvent accessibility leads to a more polar H-site environment, which ultimately quenches the ANS fluorescence signal, whereas decreased solvent accessibility facilitates high ANS quantum yields. As opposed to the Alpha consensus protein, hGSTA1-1 maximises the reduction of solvent accessibility to the H-site in order to cause an increase in ANS fluorescence intensity. The discrepancy between the hGSTA1-1 spectra and the Alpha consensus protein spectra is, therefore, due to the M50L mutation. The M50L mutation also ensures that accessible H-site surface area is higher in the Alpha consensus protein than in hGSTA1-1.

Due to the α9 helix residing at the H-site, the Alpha consensus mutant and hGSTA1-1 possess an additional element of secondary structure at the H-site which other GST classes do not possess, which leads to decreased solvent accessibility and increased hydrophobicity of the H-site. The ultimate explanation of the consistent discrepancy between the Mu proteins and the Alpha proteins would, therefore, be the presence of the additional α9 helix situated at the H-site of the Alpha proteins. It is also important to note that ligand binding to the H-site itself increases the stability of the α9 helix. This, in turn, means that the hydrophobicity of the H-site under these conditions is increased. The larger cavities of the Alpha mutant and wildtype proteins also explain why their fluorescence intensity is higher than that of their Mu counterparts.

The Mu proteins overall, possessed smaller cavities than the Alpha proteins, thereby producing ANS spectra that were considerably lower in fluorescence intensity than the spectra of the Alpha proteins. Due to the smaller cavity volumes, the Mu proteins relied on minimising solvent accessibility to achieve higher ANS quantum yields. The Mu consensus mutant protein had a smaller total cavity volume than hGSTM1-1. This increased structural compactness likely contributed to the Mu consensus protein having lower solvent

accessibility and a higher degree of hydrophobicity at the H-site. This caused the Mu consensus protein H-site to be more hydrophobic than that of hGSTM1-1, allowing ANS to fluoresce at a higher intensity.

4.5 Thermal shift assays

The thermal shift assay results indicated a notable increase in the thermal stability of the consensus mutant proteins of both the Alpha and Mu classes, relative to their wildtype counterparts (Figure 19). The MD simulation results, particularly the Rg results (Figure 23), show that the overall structures of the consensus mutant proteins were more compact than their wildtype counterparts. Although this fact means that the ΔS_{conf} is unfavourable for protein folding of the consensus mutant proteins, the ΔS_{solv} term is much greater than that of the wildtype. This means that the overall ΔG for the consensus mutant proteins is lower than that of the wildtype proteins.

This larger thermodynamic barrier of the consensus mutant proteins means that unfolding of these proteins is more unfavourable than the unfolding of the wild types. This explains the additional heat energy required to unfold 50% of the consensus mutant proteins. Additionally, the thermal shift assay results indicated that the binding of GSH marginally increased the stability of all the proteins. This may be attributed to the fact that upon GSH binding, GSTs become more conformationally rigid. The increased compactness causes an increase in protein contacts, which decreases the ΔH, thereby requiring more energy to make protein unfolding thermodynamically favourable. This is because protein contacts, and bond formation in general, are enthalpically favourable, while bond cleavage is enthalpically unfavourable. High thermostability of consensus proteins is well established, as seen in the case of the Azami green fluorescent protein. It was shown that the consensus design of the Azami green fluorescent protein led to an increase of 5.5 °C of its T_m [49].

Although the T_m is a valuable parameter, greater insight into the thermodynamics of the proteins may be attained. To derive substantive thermodynamic parameters, reversibility and equilibrium would have to be established for protein unfolding, wherein 90% of the protein refolds after an unfolding experiment. This could be done thermally or chemically. For the purposes of this study, however, thermal stability was the benchmark set for assessing consensus protein stability.

4.6 Enzyme kinetics

The CDNB-GSH conjugation reaction was the basis for determining the role of conserved residues in the preservation of the function of an enzyme. The Mu consensus mutant had a specific activity that was higher than its wildtype counterpart by a factor of three. However, this is not necessarily indicative of the Mu consensus mutant possessing a greater affinity for the substrates. As is evidenced by the Michaelis-Menten results, the Mu consensus mutant possesses a significantly lower affinity for GSH, which had lower K_M than hGSTM1-1 by a factor of 1.6. This means that in all probability, the higher specific activity seen in the Mu consensus mutant is due to the consensus mutant being part of a purer batch than its wildtype counterpart. This is suggestive of the fact that the consensus mutations in the Mu protein create a protein that is optimal for higher and purer yields of protein. The Alpha consensus mutant exhibited negligible enzymatic activity. This is in contrast to the Mu consensus mutant. The Alpha consensus mutant may, therefore, be seen as a more poorly expressed protein. The Alpha consensus protein also had an affinity for GSH that was lower than hGSTA1-1 by a factor of two, indicating diminished enzyme function.

On the basis of the analysis of the Michaelis-Menten kinetics, the consensus mutant proteins lost a significant amount of their catalytic capacity, with respect to the wildtype proteins. The significant reduction of the enzyme activity may be attributed to the fact that the protein products of full-length consensus sequences have been shown to have a decrease in catalytic rates [52]. More specifically, the Alpha consensus mutant protein may have lost a significant portion of its functionality due to the M50L mutation. The replacement of the M50 residue in hGSTA1-1 has been shown to reduce enzyme activity by 20%. As previously noted, the G-site is partially formed by the α2 helix as well as the loop between the α2 helix and β3 strand. Considering the fact that the M50 residue is located in the loop between the α2 helix and β3 strand (which forms part of the G-site), the compromised activity of the Alpha consensus mutant is likely caused by this alteration. This is because this portion of the G-site is responsible for the formation of hydrogen bonds between the G-site and GSH. The M50L mutation eliminates the formation of hydrogen bonds because leucine does not have any hydrogen bond acceptor or donor groups. Additionally, the M50L mutation affects the molecular state of the H-site. It does this through the aforementioned relationship between helices α2 and α9. The lack of stabilising interactions on the a9 helix due to the increase in flexibility of the α2 helix causes the H-site to be more solvent accessible, affecting the hydrophobicity of the H-site. Despite all of this, the M50L mutation does not account for the

near eradication of the enzymatic activity seen in the Alpha consensus mutant protein. The other mutations likely have a cumulative effect on the enzymatic activity of the Alpha consensus mutant. It is, therefore, not possible to attribute a single residue mutation, except for the M50L mutation, for the diminishment of the Alpha consensus mutant activity. The dissipation of enzyme activity in the Mu consensus mutant is not attributable to a single mutation. It is more likely that the structural rigidity of the Mu consensus protein decreased its conformational dynamism, which reduced the favourability of substrate binding.

4.7 MD simulations

Overall, the MD simulations indicated that the introduction of consensus mutations resulted in an increase of the compactness of the consensus mutant proteins of both classes (Figures 21-23). The increased compactness that is observed in the consensus mutant proteins may be an explanation as to why there is a marked decrease in the enzymatic function of the consensus mutant proteins. It has been theorised that proteins that resist conformational flexibility are unfavourable binding partners for their cognate substrates [53]. The C_α-RMSF medians of the wildtype and consensus Alpha proteins showed that the consensus mutations caused a considerable perturbation of the residues (Figure 22). The results showed that there were regions in the Alpha consensus mutant which displayed a higher degree of flexibility. The specific regions that showed increased flexibility were residues 41 to 51 and residues 207 to 219. These regions correspond with the α2 and α9 helices respectively. This is likely due to the fact the M50L mutation causes the α2 helix to become increasingly flexible. This causes the α9 helix to become more flexible due to its interactions with the α2 helix through tertiary packing interactions. This ultimately causes the perturbation of helix α9. The *in vitro* implications of this may be that these two helices are partially unordered due to excessive flexibility. This may explain the discrepancy seen in the MRE readings of the far-UV CD results of the Alpha wildtype and consensus mutant proteins. It is important to note that increased RMSF of specific regions in a protein does not imply overall increased flexibility and dynamism for the global protein in its entirety. For the Mu proteins, both wildtype and mutant, the C_α-RMSF results in Figure 22 show that the consensus mutations did not significantly change the amino acid residue flexibility. In fact, the residues were slightly less flexible in the Mu consensus mutant. This means that the number of available conformers decreased, thereby decreasing the binding capacity.

The MD simulations, particularly the C_α-RMSD (Figure 21) and Rg results (Figure 23), indicate that the smoothening of the bottom of the folding funnel for GSTs did occur, but

produced undesirable outcomes. The smoothing of the bottom of the folding funnel can be said to have taken place because the increased compactness seen in Figure 21 and 23 indicates that the consensus proteins do not access a large number of conformers. This finding is in line with the lack of a rugged folding funnel minimum, as a rugged folding funnel minimum indicates a higher number of different energy-state conformers available to a native protein. Coupling the RMSD results with the thermal assay results, the increased compactness caused an increase in the T_m values of the consensus proteins. Thus, the smoothing of the folding funnel can be postulated to have occurred due to the increased conformational stability in the form of a lower energy native state. As previously stated, increased ruggedness is directly correlated with an increase in protein conformers because more conformations are available to a protein in a more rugged folding funnel minimum. The increased compactness seen in the MD simulation results illustrates that there was possibly a considerable reduction in the ruggedness of the bottom of the folding funnel of the consensus proteins, thereby limiting the flexibility of the consensus mutant proteins, which, in turn, limited the number of conformers accessible to the consensus mutant proteins. The lower number of available conformations for the consensus mutants meant that they would be less suitable binding partners for the traditional GST substrates. This is highly instructive, as it explains, in part, the significant reduction of the enzymatic function seen in the consensus mutants. Although both consensus mutant proteins exhibited increased compactness, the Alpha consensus mutant is, by all indications the more compact consensus mutant.

The Y9-Y9 and Y6-Y6 distance simulation results (Figure 24) further add to the rich tapestry of the molecular dynamics of the proteins. The median Y9-Y9 distance of the Alpha consensus mutant was 1.55 Å less than the hGSTA1-1 median. This further indicated that the overall movement and flexibility of the Alpha consensus mutant was severely repressed by the introduction of the consensus mutations. The Y6-Y6 distance results for the Mu protein showed that the consensus mutations did not have any significant impact on the flexibility of the protein, as evidenced by the 0.13 Å difference in the median Y6-Y6 distance of both Mu proteins.

The sitemap analysis results indicated that the largest surface cavities for all of the proteins was the dimer interface (Figure 27). The sitemap analysis, however, also showed massive disparities between the mutant and wildtype protein cavity volumes. As previously discussed, the dimer interface of Alpha and Mu GSTs, broadly speaking, consists of a hydrophobic lock-and-key motif as well as charge clusters that increase the stability of the quaternary structure

of the protein. The results of the sitemap analysis are, therefore, insightful. Firstly, the Alpha consensus mutant has a cavity at its dimer interface whose volume is 260.68 $Å^3$ more than hGSTA1-1. This may be attributed to the M50L mutation, which has the role of maintaining electrostatic and van der Waal's forces within the hydrophobic lock pocket of the hydrophobic lock-and-key motif. Secondly, the Mu consensus mutant has a dimer interface volume that is 653.758 $Å^3$ less than hGSTM1-1. This major decrease in dimer interface cavity volume may explain the increased compactness of the Mu consensus mutant. It also coincides with increased rigidity seen the Mu consensus C_α-RMSD and Rg results. Furthermore, the cavity sizes were foundational to the ANS fluorescence spectroscopy results, as discussed in section 4.4.

The superimposition of the poses added further context to the functional and structure behaviour of the consensus mutants. The RMSD from the superimposition of the Alpha consensus mutant on the hGSTA1-1 structure was 1.85 Å (Figure 25). This indicated major structural variation between the Alpha wildtype and mutant. The same RMSD parameter for the Mu proteins was 0.88 Å (Figure 26), indicating significant structural variation, though not on the scale of the Alpha proteins. These results further cements that the M50L mutation was more powerful than the other mutations in the Mu consensus protein, due to its role in the hydrophobic lock-and-key motif.

In summary, the consensus mutations increased the compactness and rigidity of the proteins. The Mu consensus mutant, however, was closer to the wildtype in terms of the overall conformational flexibility than the Alpha consensus mutant was to the Alpha wildtype. It may be said, therefore, that the consensus mutations limited the number of conformers available to the consensus mutant proteins. This likely decreased the binding capacity of the consensus mutants. In the case of the Alpha consensus protein, although the RMSF median was higher than that of the wildtype, the increased RMSF pointed to a likely disruption of the secondary structure of the Alpha consensus mutant, in particular the α2 and α9 helices. This would lead to a decrease in enzyme functionality.

4.8 Conclusion

In the final analysis, it would seem that nature is not all that forgiving, as there is a sort of compromise equilibrium between protein stability and function. As it is in nature, so in the process of consensus protein design. Each consensus mutation must be painstakingly examined in order to see its effect on the function and stability of the protein. This would

enable one to eliminate destabilising mutations. Overall, in this study, the consensus mutations may be deemed as detrimental to the functionality of the proteins, while marginally increasing the thermal stability of the proteins. This is despite the fact that the majority of the consensus mutations appear distal to the active site. This means that the decreased enzymatic function of the consensus mutant proteins is due to the increased compactness observed in the C_α-RMSD and C_α-RMSF results.

The secondary structural components of the Mu consensus mutant protein are, for all intents and purposes, identical to its wildtype counterpart. In the case the Alpha proteins, a marked change was observed in the secondary structure of the consensus mutant in comparison to the wildtype. This change, likely caused by the M50L mutation, is due to the α2 and α9 helices being more dynamic and possibly less ordered under *in vitro* conditions. Indeed, the far-UV CD results, in tandem with the RMSF results indicate that the highly ordered α-helical structure of hGSTA1-1 becomes slightly disordered with the introduction of the consensus mutations. The ANS binding studies showed that the consensus mutant proteins possessed higher hydrophobicity at their respective H-sites, relative to the wildtype proteins. In the case of the Alpha consensus mutant, this is primarily due to the M50L mutation, which played a role in the packing interactions between the α2 and α9 helices. For the Mu consensus protein, the lower cavity volume reduced solvent accessibility to the H-site, thereby preventing solvent quenching of ANS.

The enzymatic function of both consensus proteins was severely diminished. No single mutation may be solely attributed to this reduction of activity. However, the M50L mutation in the Alpha consensus mutant affected GSH binding through the lack of hydrogen bond formation. As for the Mu consensus mutants, the rigidity of the protein may be the ultimate cause of reduction of enzymatic activity. This may be remedied by limiting the consensus mutations to specific domains of the protein. Though the Alpha consensus mutant exhibited high volumes of surface cavities, the Y9-Y9 distance simulations indicated that the subunits were not far apart from each other. Seemingly, the Mu consensus protein did not suffer any significant loss of dimer interface interactions. The thermal shift assay results concluded that the consensus proteins were more thermally stable than the wildtype proteins. This is supported by the Y9-Y9 and Y6-Y6 atomic distances, which are indicative of compressed global structures of the mutant proteins of both classes.

Ultimately, generating a *de novo* consensus protein likely introduced stabilising and destabilising mutations, offsetting potential benefits of increasing thermal stability and maintaining the enzymatic function. Coupled with the fact that the most conserved region of GSTs is the thioredoxin-like domain, it is highly probable that from an evolutionary standpoint, the conservation of residues aims to primarily preserve protein stability rather than protein function. For posterity, obtaining the 3D structures of the consensus mutant proteins will conclusively elucidate the impact of the mutations on the conformations of the proteins. This will entail crystallisation and structure determination.

The implications of this study are mostly applicable to the field of biotechnology. With these results, it is possible to engineer GSTs that are innately conducive to producing higher yields during recombinant protein production. By fine-tuning the stability of the enzymes, GSTs can also be used to generate more suitable binding partners for molecules that require high specificity for their delivery. Additionally, increasing the stability of GSTs enables the potential for engineering GSTs that have a longer half-life. This increased GST half-life would be highly useful for strenuous biotechnological applications. Functionally, more optimisation needs to be done in order to achieve GSTs that catalyse the GSH conjugation reaction at a greater and more efficient rate.

Chapter 5: References

[1] C.B. Anfinsen, The formation and stabilization of protein structure., Biochem. J. 128 (1972) 737. https://doi.org/10.1042/BJ1280737.

[2] P.L. Privalov, Thermodynamics of protein folding, J. Chem. Thermodyn. 29 (1997) 447–474. https://doi.org/10.1006/JCHT.1996.0178.

[3] L. Lins, R. Brasseur, The hydrophobic effect in protein folding, FASEB J. 9 (1995) 535–540. https://doi.org/10.1096/FASEBJ.9.7.7737462.

[4] J.A. Caro, K.W. Harpole, V. Kasinath, J. Lim, J. Granja, K.G. Valentine, K.A. Sharp, A.J. Wand, Entropy in molecular recognition by proteins, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) 6563–6568. https://doi.org/10.1073/PNAS.1621154114/-/DCSUPPLEMENTAL.

[5] H.S. Chan, S. Bromberg, K.A. Dill, Models of cooperativity in protein folding, Philos. Trans. R. Soc. Lond. B. Biol. Sci. 348 (1995) 61–70. https://doi.org/10.1098/RSTB.1995.0046.

[6] K.A. Dill, Polymer principles and protein folding., Protein Sci. 8 (1999) 1166. https://doi.org/10.1110/PS.8.6.1166.

[7] N.A. Campbell, J.B. Reece, Biology, 7th ed., Benjamin Cummings, San Francisco, 2002.

[8] P.L. Privalov, S.J. Gill, Stability of protein structure and hydrophobic interaction, Adv. Protein Chem. 39 (1988) 191–234. https://doi.org/10.1016/S0065-3233(08)60377-0.

[9] P.E. Leopold, M. Montal, J.N. Onuchic, Protein folding funnels: a kinetic approach to the sequence-structure relationship., Proc. Natl. Acad. Sci. 89 (1992) 8721–8725. https://doi.org/10.1073/PNAS.89.18.8721.

[10] J.N. Onuchic, N.D. Socci, Z. Luthey-Schulten, P.G. Wolynes, Protein folding funnels: the nature of the transition state ensemble, Fold. Des. 1 (1996) 441–450. https://doi.org/10.1016/S1359-0278(96)00060-0.

[11] C. Levinthal, Are there pathways for protein folding?, J. Chim. Phys. 65 (1968) 44–45. https://doi.org/10.1051/JCP/1968650044.

[12] A. Šali, E. Shakhnovich, M. Karplus, How does a protein fold?, Nat. 1994 3696477.

369 (1994) 248–251. https://doi.org/10.1038/369248a0.

[13] C. Tsai, S. Kumar, B. Ma, R. Nussinov, Folding funnels, binding funnels, and protein function, Protein Sci. 8 (1999) 1181–1190. https://doi.org/10.1110/PS.8.6.1181.

[14] B. Ma, S. Kumar, C.J. Tsai, R. Nussinov, Folding funnels and binding mechanisms, Protein Eng. Des. Sel. 12 (1999) 713–720. https://doi.org/10.1093/PROTEIN/12.9.713.

[15] B.T. Porebski, A.M. Buckle, Consensus protein design, Protein Eng. Des. Sel. 29 (2016) 245–251. https://doi.org/10.1093/protein/gzw015.

[16] B.T. Porebski, A.A. Nickson, D.E. Hoke, M.R. Hunter, L. Zhu, S. McGowan, G.I. Webb, A.M. Buckle, Structural and dynamic properties that govern the stability of an engineered fibronectin type III domain, Protein Eng. Des. Sel. 28 (2015) 67. https://doi.org/10.1093/PROTEIN/GZV002.

[17] D.W. Nebert, V. Vasiliou, Analysis of the glutathione S-transferase (GST) gene family, Hum. Genomics. 1 (2004) 460. https://doi.org/10.1186/1479-7364-1-6-460.

[18] B. Mannervik, U. Helena Danielson, B. Ketterer, Glutathione transferases--structure and catalytic activity, CRC Crit. Rev. Biochem. 23 (1988) 283–337. https://doi.org/10.3109/10409238809088226.

[19] B. Mannervik, Y.C. Awasthi, P.G. Board, J.D. Hayes, C. Di Ilio, B. Ketterer, I. Listowsky, R. Morgenstern, M. Muramatsu, W.R. Pearson, C.B. Pickett, K. Sato, M. Widersten, C.R. Wolf, Nomenclature for human glutathione transferases., Biochem. J. 282 (1992) 305. https://doi.org/10.1042/BJ2820305.

[20] R.N. Armstrong, Structure, catalytic mechanism, and evolution of the glutathione transferases, Chem. Res. Toxicol. 10 (1997) 2–18. https://doi.org/10.1021/TX960072X/ASSET/TX960072X.FP.PNG_V03.

[21] H.J. Atkinson, P.C. Babbitt, Glutathione Transferases Are Structural and Functional Outliers in the Thioredoxin Fold, Biochemistry. 48 (2009) 11108. https://doi.org/10.1021/BI901180V.

[22] E. Laborde, Glutathione transferases as mediators of signaling pathways involved in cell proliferation and cell death, Cell Death Differ. 17 (2010) 1373–1380. https://doi.org/10.1038/CDD.2010.80.

[23] V. Adler, Z. Yin, S.Y. Fuchs, M. Benezra, L. Rosario, K.D. Tew, M.R. Pincus, M. Sardana, C.J. Henderson, C.R. Wolf, R.J. Davis, Z. Ronai, Regulation of JNK signaling by GSTp., EMBO J. 18 (1999) 1321. https://doi.org/10.1093/EMBOJ/18.5.1321.

[24] M.A. Vargo, R.F. Colman, Affinity labeling of rat glutathione S-transferase isozyme 1-1 by 17beta -iodoacetoxy-estradiol-3-sulfate, J. Biol. Chem. 276 (2001) 2031–2036. https://doi.org/10.1074/JBC.M008212200.

[25] A. Oakley, Glutathione transferases: a structural perspective, Drug Metab. Rev. 43 (2011) 138–151. https://doi.org/10.3109/03602532.2011.558093.

[26] A.J. Oakley, M. Lo Bello, A.P. Mazzetti, G. Federici, M.W. Parker, The glutathione conjugate of ethacrynic acid can bind to human pi class glutathione transferase P1-1 in two different modes, FEBS Lett. 419 (1997) 32–36. https://doi.org/10.1016/S0014-5793(97)01424-5.

[27] K.H. Kong, M. Nishida, H. Inoue, K. Takahashi, Tyrosine-7 is an essential residue for the catalytic activity of human class PI glutathione S-transferase: chemical modification and site-directed mutagenesis studies, Biochem. Biophys. Res. Commun. 182 (1992) 1122–1129. https://doi.org/10.1016/0006-291X(92)91848-K.

[28] J.J. McElwee, E. Schuster, E. Blanc, M.D. Piper, J.H. Thomas, D.S. Patel, C. Selman, D.J. Withers, J.M. Thornton, L. Partridge, D. Gems, Evolutionary conservation of regulated longevity assurance mechanisms, Genome Biol. 8 (2007). https://doi.org/10.1186/GB-2007-8-7-R132.

[29] K.H. Kong, K. Takasu, H. Inoue, K. Takahashi, Tyrosine-7 in human class Pi glutathione S-transferase is important for lowering the pKa of the thiol group of glutathione in the enzyme-glutathione complex, Biochem. Biophys. Res. Commun. 184 (1992) 194–197. https://doi.org/10.1016/0006-291X(92)91177-R.

[30] H. Dirr, P. Reinemer, R. Huber, X-ray crystal structures of cytosolic glutathione S-transferases, Eur. J. Biochem. 220 (1994) 645–661. https://doi.org/10.1111/J.1432-1033.1994.TB18666.X.

[31] P. Reinemer, H.W. Dirr, R. Ladenstein, R. Huber, M. Lo Bello, G. Federici, M.W. Parker, Three-dimensional structure of class π glutathione S-transferase from human

placenta in complex with S-hexylglutathione at 2.8 Å resolution, J. Mol. Biol. 227 (1992) 214–226. https://doi.org/10.1016/0022-2836(92)90692-D.

[32] I. Sinning, G.J. Kleywegt, S.W. Cowan, P. Reinemer, H.W. Dirr, R. Huber, G.L. Gilliland, R.N. Armstrong, X. Ji, P.G. Board, B. Olin, B. Mannervik, T.A. Jones, 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 (1993) 192–212. https://doi.org/10.1006/JMBI.1993.1376.

[33] Y. Sayed, L.A. Wallace, H.W. Dirr, The hydrophobic lock-and-key intersubunit motif of glutathione transferase A1-1: implications for catalysis, ligandin function and stability, FEBS Lett. 465 (2000) 169–172. https://doi.org/10.1016/S0014-5793(99)01747-0.

[34] C.S. Alves, D.C. Kuhnert, Y. Sayed, H.W. Dirr, The intersubunit lock-and-key motif in human glutathione transferase A1-1: role of the key residues Met51 and Phe52 in function and dimer stability, Biochem. J. 393 (2006) 523–528. https://doi.org/10.1042/BJ20051066.

[35] D. Sheehan, G. Meade, V.M. Foley, C.A. Dowd, Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily., Biochem. J. 360 (2001) 1. https://doi.org/10.1042/0264-6021:3600001.

[36] S. Lu, J. Wang, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, G.H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, M. Yang, D. Zhang, C. Zheng, C.J. Lanczycki, A. Marchler-Bauer, CDD/SPARCLE: the conserved domain database in 2020, Nucleic Acids Res. 48 (2020) D265–D268. https://doi.org/10.1093/NAR/GKZ991.

[37] Z.Y. Zhu, S. Karlin, Clusters of charged residues in protein three-dimensional structures., Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 8350. https://doi.org/10.1073/PNAS.93.16.8350.

[38] R.N. Armstrong, Mechanistic imperatives for the evolution of glutathione transferases, Curr. Opin. Chem. Biol. 2 (1998) 618–623. https://doi.org/10.1016/S1367-5931(98)80093-8.

[39] L.O. Hansson, R. Bolton-Grob, T. Massoud, B. Mannervik, Evolution of differential substrate specificities in Mu class glutathione transferases probed by DNA shuffling, J. Mol. Biol. 287 (1999) 265–276. https://doi.org/10.1006/JMBI.1999.2607.

[40] F. Sievers, D.G. Higgins, Clustal Omega for making accurate alignments of many protein sequences, Protein Sci. 27 (2018) 135–145. https://doi.org/10.1002/PRO.3290.

[41] K. Pooe, M. Thulo, H. Makumbe, B. Akumadu, O. Otun, C. Aloke, I. Achilonu, Biophysical description of Bromosulfophthalein interaction with the 28-kDa glutathione transferase from Schistosoma japonicum, Mol. Biochem. Parasitol. 252 (2022) 111524. https://doi.org/10.1016/J.MOLBIOPARA.2022.111524.

[42] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature. 227 (1970) 680–685. https://doi.org/10.1038/227680a0.

[43] S.J. PERKINS, Protein volumes and hydration effects, Eur. J. Biochem. 157 (1986) 169–180. https://doi.org/10.1111/J.1432-1033.1986.TB09653.X.

[44] R.W. Woody, Circular dichroism, Methods Enzymol. 246 (1995) 34–71. https://doi.org/10.1016/0076-6879(95)46006-3.

[45] W.H. Habig, M.J. Pabst, W.B. Jakoby, Glutathione S-Transferases: The first enzymatic step in mercapturic acid formation, J. Biol. Chem. 249 (1974) 7130–7139. https://doi.org/10.1016/S0021-9258(19)42083-8.

[46] H. Bisswanger, Enzyme assays, Perspect. Sci. 1 (2014) 41–55. https://doi.org/10.1016/J.PISC.2014.02.005.

[47] J.M. Scholtz, G.R. Grimsley, C.N. Pace, Solvent denaturation of proteins and interpretations of the m value., Methods Enzymol. 466 (2009) 549–565. https://doi.org/10.1016/S0076-6879(09)66023-7.

[48] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190. https://doi.org/10.1101/GR.849004.

[49] M. Dai, H.E. Fisher, J. Temirov, C. Kiss, M.E. Phipps, P. Pavlik, J.H. Werner, A.R.M. Bradbury, The creation of a novel fluorescent protein by guided consensus engineering, Protein Eng. Des. Sel. 20 (2007) 69–79. https://doi.org/10.1093/PROTEIN/GZL056.

[50] S.Y.M. Lau, A.K. Taneja, R.S. Hodges, Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded α-helical coiled-coils, J. Biol. Chem. 259 (1984) 13253–13261. https://doi.org/10.1016/s0021-9258(18)90686-1.

[51] J.N. Ketley, W.H. Habig, W.B. Jakoby, Binding of nonsubstrate ligands to the glutathione S-transferases., J. Biol. Chem. 250 (1975) 8670–8673. https://doi.org/10.1016/S0021-9258(19)40723-0.

[52] M. Lehmann, D. Kostrewa, M. Wyss, R. Brugger, A. D'Arcy, L. Pasamontes, A.P.G.M. Van Loon, From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase, Protein Eng. Des. Sel. 13 (2000) 49–57. https://doi.org/10.1093/PROTEIN/13.1.49.

[53] G. Hernández, F.E. Jenney, M.W.W. Adams, D.M. LeMaster, Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 3166–3170. https://doi.org/10.1073/PNAS.97.7.3166/ASSET/A02C9D4B-60A5-437C-B128-CBA221E77B94/ASSETS/GRAPHIC/PQ0405696003.JPEG.

[54] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612. https://doi.org/10.1002/JCC.20084.